ACP2019-173 by Vu et al.

Responses to the reviewers

General response: We thank both reviewers for providing detailed comments. We have addressed all the comments made and revised the manuscript accordingly.

Review 1

General comment: The major issue I see with this manuscript is in the lack of detail in model descriptions, evaluations, and data sources, all of which are lacking throughout the manuscript. I've laid out specific concerns below. Overall, a general lack of detail makes it difficult to trust the results and conclusions about the effectiveness of the various control actions.

Response: We agree with the reviewer that model description, evaluation and data sources are important in a scientific paper.

Exactly for this reason, we evaluated the model extensively in this work. In page 7 of the supplement, we have provided two figures (Figure S2 and S3; note that they are now Figure S3 and S4) to compare the model predicted variables with observed ones (i.e., for the 30% of the dataset that were not used for constructing the model). In page 7 of the supplement, we also provided the correlation coefficients between predicted hourly and observed concentrations for all the parameters. In Figure S3 and Figure 5, we provided the regression equations as well as the correlation coefficients. In page 3, line 109 to 111 of the original main text, we explained that "we firstly construct the RF model from a training data set (e.g., 70% of the all data available) of observed concentrations of a pollutant and its predictor variables and then validate the model by unseen data sets (testing data sets)". Furthermore, in Figure 5 of the original manuscript, we compared the model predicted monthly concentration of $PM_{2.5}$ by the RF model and the WRF-CMAQ model against the observed values. Therefore, the RF model results were evaluated against observations.

We have indeed calculated other parameters for model evaluation, for example RMSE, but we did not report it because the figures and the r2 already showed the good performance of the model. However, we respond in more detail below and have included more parameters in the revised manuscript.

Line 161-168: "Table S2, Figure S3-S4 and Section S3 provided information on the performance of our model using a number of statistical measures including mean square error (MSE)/ root mean square error (RMSE), correlation coefficients (r2), FAC2 (fraction of predictions with a factor of two), MB (mean bias), MGE (mean gross error), NMB (normalised mean bias), NMGE (normalised mean gross error), COE (Coefficient of Efficiency), IOA (Index of Agreement) as suggested in a number of recent papers (Emery et al. 2017, Henneman et al., 2017, and Dennis et al., 2010). These results confirm that the model perform very well in comparison with traditional statistical methods and air quality models (Henneman et al., 2015)".

The reviewer also questioned that there is a lack of detail on the data sources. We have explained in the original text that data were collected from the 12 national air quality monitoring stations in Beijing. In the revised manuscript, we made this clearer: "Hourly air quality data for six key air pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO) was collected by 12 national air quality monitoring stations in Beijing by the China National Environmental Monitoring Network (CNEM). Hourly air quality data were downloaded from the CNEM website - http://106.37.208.233:20035. Since air quality data are removed from the website on a daily basis, data were automatically downloaded to a local computer and combined to form the whole dataset for this paper." All data are now available at https://github.com/tuanvvu/Air_Quality_Trend_Analysis (last access 5 June 2019).

With regards to the model descriptions, we did not generate this algorithm from scratch. We used the Grange et al. (2018) model as a basis. In the revised manuscript, we emphasized that in this work we modified the Grange et al. (2018) algorithm in order to understand the seasonal variation of air pollutants. We have revised our method section to make it clearer as below:

"A weather normalisation technique predicts the concentration of an air pollutant at a specific measured time point (e.g., 09:00 on 01/01/2015) with randomly selected meteorological conditions. This technique was firstly introduced by Grange et al. (2018). In their method, a new dataset of input predictor features including time variables (day of the year, the day of the week, hour of the day, but not the Unix time variable) and meteorological parameters (wind speed, wind direction, temperature and RH) is firstly generated (i.e., re-sampled) randomly from the original observation dataset. For example, for a particular day (e.g., 01/01/2011), the model randomly selects the time variables (excluding Unix time) and weather parameters at any day from the data set of predictor features during the whole study period. This is repeated 1,000 times to provide the new input data set for a particular day. The input data set is then fed to the random forest model to predict the concentration of a pollutant at a particular day (Grange et al., 2018; Grange and Carslaw, 2019). This gives a total of 1,000 predicted concentrations for that day. The final concentration of that pollutant, referred hereafter as weather normalised concentration, is calculated by averaging the 1000 predicted concentrations. This method normalises the impact of both seasonal and weather variations. Therefore, it is unable to investigate the seasonal variation of trends for a comparison with the trend of primary emissions. For this reason, we enhanced the meteorological normalisation procedure.

In our algorithm, we firstly generated a new input data set of predictor features, which includes original time variables and re-sampled weather data (wind speed, wind direction, temperature, and relative humidity). Specifically, weather variables at a specific selected hour of a particular day in the input data sets were generated by randomly selecting from the observed weather data (i.e., 1988-2017 or 2013-2017) at that particular hour of different dates within a four-week period (i.e., 2 weeks before and 2 weeks after that selected date). For example, the new input weather data at 08:00 15/01/2015 are randomly selected from the observed data at 08:00 am on any date from 1st to 29th January of any year in 1988-2017 or 2013-2017. The selection process was repeated automatically 1,000 times to generate a final input data set. Each of the 1,000 data was then fed to the random forest model to predict the concentration of a pollutant. The 1,000 predicted concentrations were then averaged to calculate the final weather normalised concentration for that particular hour, day, and year. This way, unlike Grange et al., (2018), we only normalise the weather conditions but not the seasonal and diurnal variations. Furthermore, we are able to re-sample observed weather data for a longer period (for example, 1998-2017), rather than only the study period. This new approach enables us investigate the seasonality of weather normalised concentrations and compare them with primary emissions from inventories". (Line 171-204).

We provided the R code in the following website so that an experienced statistician will be able to test the model. https://github.com/tuanvvu/Air_Quality_Trend_Analysis

Specific comments and responses

1. **Comment:** abstract- "improved a novel machine learning-based random forest technique". How?

Response: In our study, we enhanced the weather normalisation technique using the random forest technique algorithm of Grange et al. (2018). We explained this in detail in the revised manuscript. Please see response to general comment above.

We have revised the text in the abstract to "applied machine learning-based random forest technique". (line 30 in the revised manuscript).

2. **Comment:** Line 75- "But they usually gave a poor fitting, suggesting a poor performance of the KZ filter model, or did not allow us to investigate the effect of input variables in neural network models (therefore it is referred as a "black- box" model): A poor fit does not necessarily reflect a poor performance; performance is dictated by the goals of the modeling, whereas fit is a measure of the ability to reproduce training data.

Response: The reviewer argued that "fit is a measure of the ability to reproduce training data". In our case, "fit" is a measure of the ability to reproduce testing data, rather than the training data. The training data are used to train the model. We agree that "performance is dictated by the goals of the modelling" but we do not think a model has a good performance if it failed to predict the testing data (e.g., observations). When modelling a time-series data set of pollutants, the performance of the model is usually evaluated by MSE (or RMSE) and R^2. Other parameters are also used, which are now included in a new table - Table S2 in the supplement to show the performance of our RF model.

We changed the sentence to "Among these models, the deep neural network models showed a better performance (i.e., higher correlation coefficient, lower root mean square error – RMSE) but did not allow us to investigate the effect of input variables". (line 84-87)

3. **Comment:** Line 79: Again, "performance" here is not defined. I recommend

Response: The reviewer wrote "I recommend" but we did not find what exactly the reviewer is recommending.

We explained in the revised manuscript that "performance" represents higher correlation coefficient, and lower root mean square error to make this clearer.

4. **Comment:** Line 79: Should mention the increased propensity of over-fitting with these models for completeness

Response: In this study, the over-fitting is checked by the testing data sets. The further investigation of over-fitting problem from the random forest algorithm is out of the scope of this study. We have discussed the over-fitting of decision tree models in the revised main text (Line 94-97): "Also, the decision trees models are prone to over-fitting, especially when the number of tree nodes is large (Kotsiantis, 2013). An over-fitting problem of a random forest model is checked by its performance using an unseen training data set".

5. **Comment:** Line 110: Recommend showing in Figure 1 that you used 70% of the data for training, 30% for model evaluation. In addition, I recommend reading Oreskes et al. (1994) for distinction between evaluation/validation on environmental datasets. Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science, 263(5147), 641–646.

Response: We followed the comment and added the information in the Figure 1. We also change the term "validation" into "evaluation". Thanks for the recommended article. Oreskes et al. (1994) discussed the concept of model evaluation and validation in the Earth Sciences. In our specific case (regression modelling of a time series data sets), the valuation/evaluation of model are on cross-validation based on the out-of-bag technique and evaluation of the predicted concentration using a testing data set. Specifically, in the random forest algorithm that we applied, the algorithm used the out-of-bag technique: each decision tree is trained using a bootstrapped subset of observations. This means that for every tree there is a separate subset of observations (called OOB observations) not being used to train that tree. The model uses OOB observations as a test set to cross-validate the performance of the random forest. This is why we used the testing data set to evaluate the predicted values from models.

6. **Comment:** Line 95: "press." has a period, whereas the other abbreviations do not.

Response: It is changed to pressure. We also removed abbreviations for other parameters.

7. **Comment:** Line 104: it => its

Response: We corrected it.

8. **Comment:** With a holdout analysis, there are many comparisons to be made beyond R^2 that tell us more about model fit. Many of the studies cited in the introduction include detailed evaluations, including with slope, intercept, and root mean square error. These should be included at the very least. There may be still other metrics that are informative for the evaluation in this particular application.

Response: Figure 5 and Figure S3 in the original supplement (now becoming Figure S4) have already showed information on some of the information suggested. In the revised manuscript, we provided more parameters, including the RMSE and other parameters recommended in the papers suggested by the reviewer (comment 17) in the supplement in Table S2.

9. **Comment:** sample => samples

Response: We corrected it.

10. **Comment:** Line 140-150: Was this a separate random forest model from the initial model described in the "Random Forest (RF) model development" section?

Response: No. In the revised manuscript, we re-wrote the section to make this clearer. In our study, we applied the RF which was already built using R codes from Grange et al. (2018). Their codes were originally based on the R package "ranger" by Wright et al. (2018) (https://github.com/imbs-hl/ranger)" Please see response to general comment above.

11. **Comment:** Line 152: This statement ("only either data (MET data) sets were re-sampled") directly contradicts the statement in the paragraph above.

Response: This appears to be a misunderstanding. We have re-written the whole section to make this clear. Please see response to general comment above.

12. **Comment:** Lines 162-8: Please state what you are regressing using the Theil-Sen estimator

Response: It is the concentration of a pollutant after weather normalisation. The Theil-Sen estimator is usually used for long-term trend analysis of a pollutant. We used this estimator to find the slope of the concentration trend of a pollutant. We modified the text to make it clear. (Line 207-208)**: "**The Theil-Sen regression technique was performed on the concentration of air pollutants after meteorological normalisation to investigate the long-term trend of pollutants".

13. **Comment:** Lines 207-210: The conclusion that this evidence indicates a robust model requires more exploration. What about the meteorology from 1998-2013 would result in the 2µg m 3 increase in detrended PM2.5 in 2017?

Response: We are unable to understand the question. We did not mention in any part of our model "2µg m 3". Thus, we cannot directly respond to this comment. We compared the model predicted concentrations against the observations (test dataset) in Figure S3 and S4, which showed the performance/bias of the model. Matrices for model performance are also shown in Table S2. We've revised the section to avoid confusion (Line 279-282):

"When meteorological conditions were randomly selected from 2013-2017 (instead of 1998-2017) in the RF model, the normalised level of $PM_{2.5}$ in 2017 was 60 µg m^{-3}, which is 1 µg m^{-3} difference to that using 1998-2017 data. This difference is due to the variation of the long-term climatology (1998-2017) to the 5 year period (2013-2017)"

14. **Comment:** Line ~220: This could also indicate that formation/deposition/reaction of PM10 and NO2 are affected differently than the other pollutants. From the evidence provided, it is difficult to fully embrace the claim that PM10 and NO2 were affected by sources that were not controlled. Figure 2 presents no evidence relating to dust events that I can see.

Response: We agree and revised this to:
"The Action Plan also led to a decrease in PM_{10} and NO_2 but to a lesser extent than that of CO, SO_2 and $PM_{2.5}$, indicating that PM_{10} and NO_2 were affected by other less well controlled sources or different atmospheric processes". (Line 292-294).

15. **Comment:** Line 223: Figure 3 does present differences between urban/rural/suburban, but there is no information on how many sites and their location. I recommend including a map so that distance to roadways/industries/spatial representativeness can be determined

Response: Site information is given in Shi et al. (2019). However, to make this clearer, we've added a figure and a Table S1 in the supplementary to show in detail the different type of sites (Figure S1).

Figure S1. Map of 12 monitoring stations in Beijing.

We were not sure why the reviewer mentioned industrial sites. There is no industrial site in Beijing so we were unable to include this in the figure.

16. Comment: Line 230: This evaluation is difficult to interpret. Are the average WRF-CMAQ values calculated in the same grid cells as the monitors? Presumably, CMAQ modeling used emissions for year 2017 (state this explicitly if so), what about years 2013 and 2016 make them reasonable comparison years for detrended PM2.5?

Response: WRF-CMAQ modelling has been described in Cheng et al. (2018). The average WRF-CMAQ values were calculated for the whole of Beijing. Yes, the CMAQ modelling used the emissions for year 2017. This is now clarified in the text (Line 119-120): “Monthly emission inventories of air pollutants were from Multi-resolution Emission Inventory for China (http://www.meicmodel.org/), and for the whole Beijing region”.

The 2013 year was chosen because it is the start-year of the Action Plan. 2016 was chosen to see the immediate effect of the 2017 measures in comparison the year before. More detailed explanation is given in Cheng et al. (2018).

17. Comment: Line 241-247: For model evaluation, I recommend including the recommended statistics from extensive publication on appropriate evaluation approaches like in Emery et al. 2017, Henneman et al., 2017, and Dennis et al., 2010. Emery, C., Liu, Z., Russell, A., Talat Odman, M., Yarwood, G., & Kumar, N. (2016). Recommendations on Statistics and Benchmarks to Assess Photochemical Model Performance. Journal of the Air & Waste Management Association. Dennis, R., T. Fox, M. Fuentes, A. Gilliland, S. Hanna, C. Hogrefe, J. Irwin, S.T. Rao, R, Scheffe, K. Schere, D.A. Steyn, and A. Venkatram. 2010. A framework for evaluating regio- nal-scale numerical photochemical modeling systems. J. Environ. Fluid Mech.10:471–89. doi: 10.1007/s10652-009- 9163-2. Henneman, L. R., Liu, C., Hu, Y., Mulholland, J. A., & Russell, A. G. (2017). Air quality modeling for accountability research: Operational, dynamic, and diagnostic evaluation. Atmospheric Environment, 166(2017), 551–565.

Response: Thanks for these recommended articles. We provided an additional table (Table S2) to include the parameters recommended in these publications.

18. Comment: Line 259: Please define the term “based line”

Response: The “baseline” of a pollutant (except for ozone) was the defined as the lowest concentration of air pollutants in the summer (the summer concentrations) – please see line 334-336: “On the other hand, the “baseline” SO_2 concentration – minimum monthly average concentration in the summer (Figure 2) – also reduced somewhat during the same period.”

19. Comment: Line 280: This contradicts the statement above that buffered changes in NO2 are due exclusively to sources that were not controlled

Response: The sentence was changed to: “The different trends between SO_2 and NO_2 indicate that other sources (e.g. traffic emissions, Figure S9) or atmospheric processes have a greater influence on ambient concentration of NO_2 than coal combustion. For examples the chemistry of the $NO/NO_2/O_3$ system will tend to “buffer” changes in NO_2 causing non-linearity in NO_x-NO_2 relationships.” (Line 356-360).

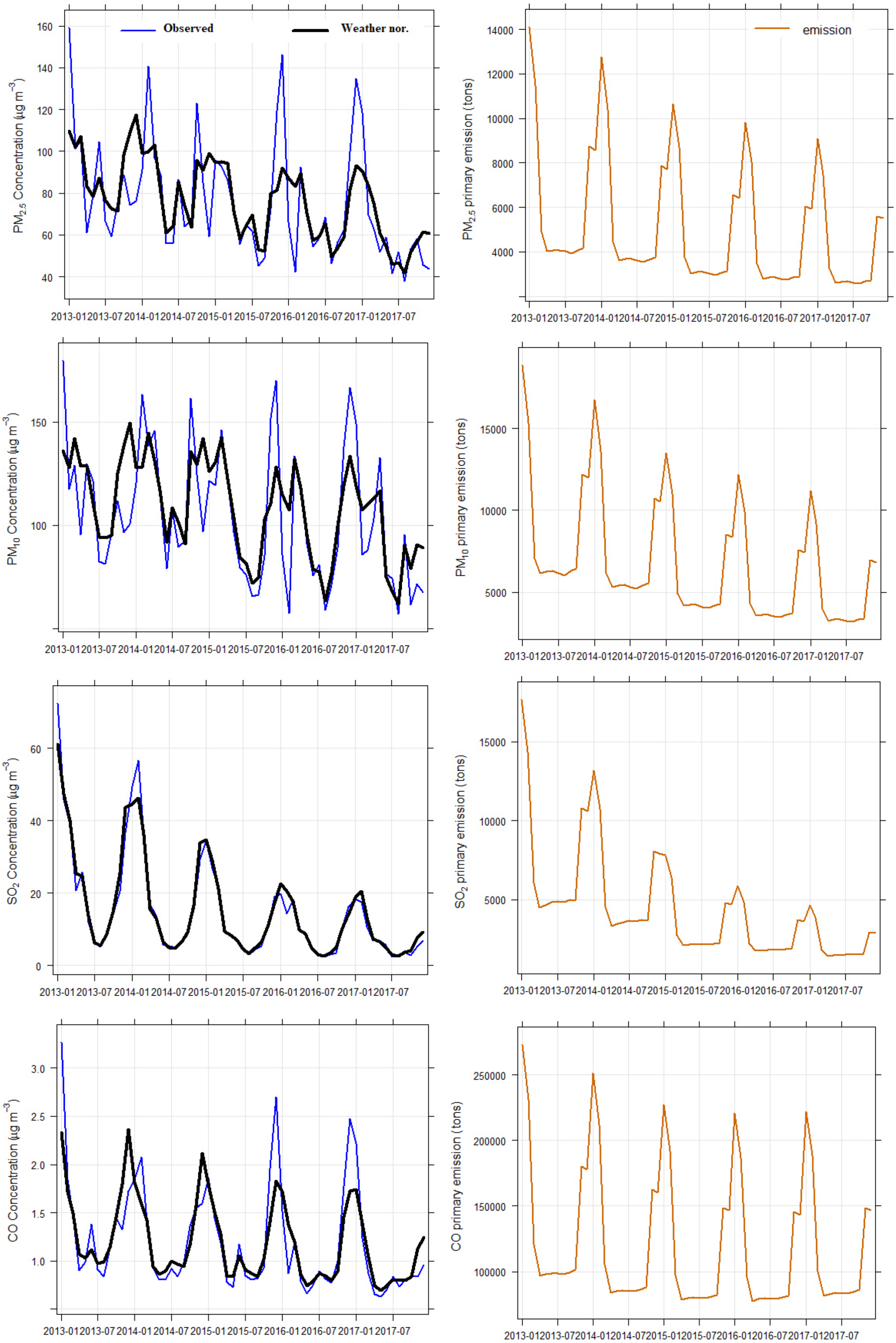
20. Comment: Line 330: Please elaborate on which data would improve this study.

Response: We refer to detailed information on the implemented policies such as the start/end date of air pollution control actions. It is now included in the main text. (Line 413-415).

21. Comment: Figure 2: I recommend including separate plots for emissions and concentrations. Plots with two vertical axes can lead to information manipulation (it is not clear, for instance, why an SO2 concentration of 40ppb corresponds to an emissions level of 2 kilotons). It would be useful to include correlations between detrended emissions and concentrations. Further, I recommend extending all vertical axes to values of 0.

Response:

We plotted the figures (see below) as suggested. We can easily replace the figure with the following ones. However, we felt that it is harder to compare the observed concentration, weather normalised concentration and primary emission in these new figures. Therefore, we suggest that it would be better to plot the primary emissions and concentrations in a single figure for a comparison.



Observed
Weather nor.
PM2.5 Concentration (µg m−3)
emission
PM2.5 primary emission (tons)
PM10 Concentration (µg m−3)
PM10 primary emission (tons)
SO2 Concentration (µg m−3)
SO2 primary emission (tons)
CO Concentration (µg m−3)
CO primary emission (tons)
2013-01 2013-07 2014-01 2014-07 2015-01 2015-07 2016-01 2016-07 2017-01 2017-07

The reviewer asked us to include correlations between detrended emissions and concentrations. We emphasise here that emissions cannot be detrended. They are based on bottom-up estimates which have nothing to do with meteorology. We tried to extend all vertical axes to 0, but they make the figure less readable (e.g., the temporal trends are hard to see).

22. Comment: Figures S4 and S5 require more description. What are Variable Importance and Variable Interactions?

Response: This has been added to the description in Figure captions.

23. Comment: Where is the emissions data from? What locations?

Response: We have added to the revised text: "Monthly emission inventories of air pollutants were from Multi-resolution Emission Inventory for China (http://www.meicmodel.org/), and it is for the whole Beijing region" (Line 119-120). The MEIC emission inventory is internationally recognized as the leading inventory for China.

24. Comment: I recommend moving much of the information on the regulations from the supplement to the main text body. I recommend using consistent language to refer to the weather normalised concentrations. At points in the manuscript, figures, and tables, these values are referred to as detrended, "Nor."

Response: We moved the key information on regulations into the main text. We use the term "weather normalised concentration" and change the "Nor." and "detrend" in Table 1 and Figure 2 to "model".

Review 2:

1. **Comment:** The authors note the use of met data from Beijing Airport. How representative is this data of all sites studied? I'm a little concerned this forms an important factor in determining the general applicability of the model. As the paper by Grange and Carslaw 2019 shows, the selection of wind directions, for example, can have significant impact on model fidelity if a site is affected by specific geography.

Response:

Airport met data are most representative of regional scale meteorology of the whole city. Because the meteorological measurements at each site are seriously affected by very local influences, it is not meaningful to compare the meteorology with that at the airport. Air pollution in the Beijing area is a regional phenomenon (Shi et al. 2019). We found very high correlations between air pollutant concentrations measured from 12 monitoring sites (Shi et al. 2019).

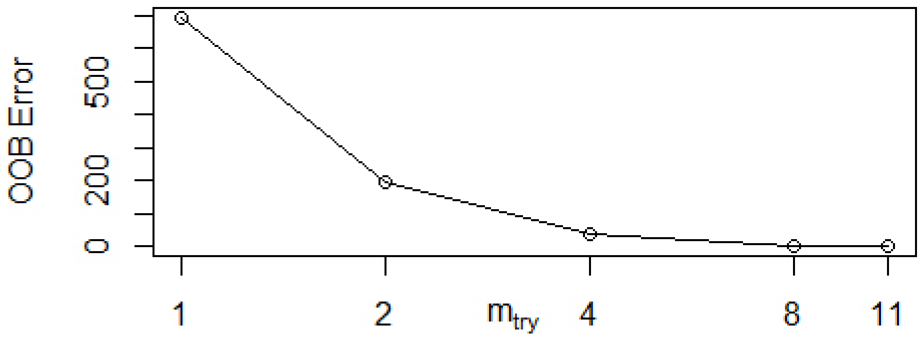
In Grange & Carslaw's paper, they also used the surface met data from the airport using the "worldmet" package. Regarding the selection of wind directions, Grange & Carslaw (2018) also noted that "Interestingly, wind direction was often a relatively unimportant variable (Fig. 4). This may be due to daily wind direction averages not contributing much information gain in the model because the aggregation period results in the metric representing atmospheric motion rather poorly".

2. **Comment:** Rather than referring to variables 'such as', please be specific in all cases.

Response: It is corrected!

3. **Comment:** You state that the 'regression model is an ensemble-model which consists of hundreds of individual decision tree models'. Please clearly state the number and how hyperparameters were derived.

Response: It is given in the SI (Section 3, Figure S1). The number of trees is 200, the minimum size of terminal nodes (Nodesize) is 3 and the variables randomly sampled for splitting (Mtry) the decision tree is 4. Mtry can be estimated based on the OOB error (as in the figure below). The number of trees and modesize was determined by RMSE and R^2. It is found with the tree numbers larger than 150 and the nodesize of 3, the RMSE is minimum and stable. A larger number of trees and nodesizes lead to little improvement in R value and RMSE, but it significantly increases the computation time. Another way we optimize the Mtry and nodesize is by a trial and error method, in which we vary the Mtry from 3 to 10 and number of trees from 20 to 500 to find the dependence of the error on the values of Mtry or number of trees.



4. **Comment:** You state you used ’e.g. 70% of the all data [correct - of all the data]’. Is this an example or is this the actual training portion you used? I think this is clarified later on but please refrain from vague statements in describing any model development workflow.

Response: It is the actual training portion we used. It is now updated in the text.

5. **Comment:** It is customary to combine a single random sampling strategy with K-folds [e.g. 5] validation. Has this been used? If not, why?

Response: No, in our study, we used out-of-bag (OOB) score estimation instead of the K-folds for model cross-validation. In the random forest algorithm which we used: each decision tree is trained using a bootstrapped subset of observations. This means that for every tree there is a separate subset of observations (called OOB observations) not being used to train that tree. The model can use OOB observations as a test set to cross-validate the performance of the random forest. The learning algorithm compares the observation's true value with the prediction from a subset of trees not trained using that observation, and calculates the overall score as a single measure of a random forest's performance.

6. **Comment:** If random sampling, how do you know if using different initial seeds in any random number generator leads to better or worse results? I can’t see any code sharing so can’t check this - please see a further comment on this.

Response: We have already considered this and used the function set.seed before running the RandomForest function to test the reproducibility. The result is almost the same. The code is available on: https://github.com/tuanvvu/Air_Quality_Trend_Analysis/blob/master/R/Air_Quality_Weather_Normalised_Trend.R

7. **Comment:** The authors talk about an ’enhanced’ normalisation procedure. Please explain more clearly how this is different from the original paper by Grange et al 2018. I will admit, that paper isnt as clear as it could be, but they do provide the model base. As far as I can tell, both studies only re-sample weather data.

Response: The concept of weather normalisation is similar and was introduced by Grange et al. (2018). Both studies re-sample the weather data, but we did it in a different way.
In Grange et al. (2018), both the weather and time predictor features (except the Unix date) were randomly generated from the original data set of predictor features as the following code:

```
"# Randomly sample observations
n_rows <- nrow(df) #df is original data set
index_rows <- sample(1:n_rows, replace = replace)
# Transform data frame to include sampled variables
df[variables] <- lapply(df[variables], function(x) x[index_rows])”
```

It means the seasonal, weekend/week, hour and weather data are also re-sampled.
In our study, only weather data were re-sampled. The advantage is that we can now see the seasonal effects. We revised the text to:

“In our algorithm, we firstly generated a new input data set of predictor features, which includes original time variables and re-sampled weather data (wind speed, wind direction, temperature, and relative humidity). Specifically, weather variables at a specific selected hour of a particular day in the input data sets were generated by randomly selecting from the observed weather data (i.e., 1988-2017 or 2013-2017) at that particular hour of different dates within a four-week period (i.e., 2 weeks before and 2 weeks after that selected date). For example, the new input weather data at 08:00 15/01/2015 are randomly selected from the observed data at 08:00 am on any date from 1st to 29th January of any year in 1988-2017.” (Line189-196).

8. **Comment:** Also there is no discussion of classification into back trajectories, for example, or estimated boundary layer heights etc. If these products are not used, how is this study an enhancement?

Response: Thank you for the suggestions. We did add the back trajectories into the model, but it did not improve the model's performance. Therefore, we have not included this in the model. We now added a sentence in the Supplement to make this point clearer (Line 107-108, SI).

We used the hourly data sets as input variables in our study. Estimated hourly boundary layer heights from models, e.g., WRF-Chem are highly uncertain. Using such uncertain data will cause unpredictable uncertainty in our results. Our RF model performed very well already, with existing input variables.

9. **Comment:** In some ways I struggle to see how section 2 'weather normalisation' is significantly different from the Grange et al approach. If they are different, they need clearly stating why - perhaps even with a visual workflow/table for each - and a comparison on findal data products. The title of the paper leads me to believe this is a new technique.

Response: Please find our response to comment 7. We clarified that we did not create a new technique. We applied the random forest model and only enhanced the "weather normalisation technique". However, the key point of this work is that we can now look at applications of the method to evaluate the air quality trends in Beijing, including seasonal variations.

10. **Comment:** line 104 - concentrations of an air pollutant and it[s] predictor variables - please correct

Response: It is corrected.

11. **Comment:** line 116: 'These time variables' - do you mean parameters that vary with time or the time variable?

Response: We mean the time variables (features): date of year, hour of the year and week/weekend. This is now modified.

12. **Comment:** line 119 [equation with no label] - what is the significance of year 'i'? Is this defined on, say, the Unix epoch?

Response: Yes, it is. It is corrected to i^{th} year (i from 2013 to 2017).

13. **Comment:** line 134: 'To validate the model for unseen data sets, a test data set which represents 30% of entire data sets[set] is input into the random forest model which has been constructed from training data sets.' This is a confusing statement. The test and training sets refer to both features and predicted variable. Thus, only features are 'input into the model'? Please re-phrase this. In fact, I would suggest you consider using the term 'features' when referring to variables to which you are fitting the model.

Response: It is re-phrased in the model evaluation line 145-147: "As shown in Figure 1, the whole data sets were randomly divided into: 1) a training data set to construct the random forest model and 2) a testing data set to test the model performance for unseen data sets. The training data set comprised of 70% of the whole data, with the rest as testing data". We changed the "variable" to "predictor features" as suggested.

14. **Comment:** line 140: 'A weather normalisation technique predicts the concentration of an air pollutant at a specific measured time point but with various meteorological conditions

(termed as “weather normalised concentration”).’ Do you mean to state that this technique predicts the concentrations of an air pollutant as a function of meteorological factors alone?

Response: It is not so, because it is also a function of the time variables. If a new weather condition is inputted to the model, it can predict the concentration of a pollutant in a certain time period.

15. Comment: line 142: ’Both time variable (month, hour) and meteorological parameters, except the trend variable were re-sampled randomly and was added into the random forest model as input variables to predict the concentration of a pollutant’. This is a confusing statement when referred to ’adding’. What do you mean by adding? On top of preexisting variables?

Response: “add” here means input. This is now updated: “A weather normalisation technique predicts the concentration of an air pollutant at a specific measured time point (e.g., 09:00 on 01/01/2015) with randomly selected meteorological conditions. This technique was firstly introduced by Grange et al. (2018). In their method, a new dataset of input predictor features including time variables (day of the year, the day of the week, hour of the day, but not the Unix time variable) and meteorological parameters (wind speed, wind direction, temperature and RH) is firstly generated (i.e., re-sampled) randomly from the original observation dataset. For example, for a particular day (e.g., 01/01/2011), the model randomly selects the time variables (excluding Unix time) and weather parameters at any day from the data set of predictor features during the whole study period. This is repeated 1,000 times to provide the new input data set for a particular day. The input data set is then fed to the random forest model to predict the concentration of a pollutant at a particular day (Grange et al., 2018; Grange and Carslaw, 2019). This gives a total of 1,000 predicted concentrations for that day. The final concentration of that pollutant, referred hereafter as weather normalised concentration, is calculated by averaging the 1000 predicted concentrations.”. (Line 171-184).

16. Comment: Section 3.4 Please explain why, in a few cases, normalised values are higher than original.

Response: As we discussed in Figure 4, if the weather during that month is more favourable for the dispersion of air pollutants, the normalised values will be higher than the observed concentration.

17. Comment: Section 3.5 ’Our results confirmed that the “Action Plan” has been highly effective’. Please define ’highly effective’.

Response: We’ve updated this to “’Our results confirmed that the “Action Plan” has led to major improvement in air quality.”

18. Comment: Code/data availability: The current paper has no statement on this. The authors need to meet the current data and code sharing standards provided by Copernicus: https://www.atmospheric-chemistry-and-physics.net/about/data_policy.html https://peerj.com/articles/cs-86/ Indeed, there are currently many uncertain aspects of this study which could be resolved by clear code sharing and documentation.

Response: They are now available at: https://github.com/tuanvvu/Air_Quality_Trend_Analysis

19. Comment: There are a number of grammatical issues throughout the paper:

Response: A senior co-author has re-checked the grammar throughout the manuscript.

Reference:

In-Depth Study of Air Pollution Sources and processes within Beijing and its Surrounding Region (APHH-Beijing), Shi, Z., Vu, T., Kotthaus, S., Harrison, R.M., Grimmond, S., Yue, S., Zhu, T., Lee, J., Han, Y., Demuzere, M., Dunmore, R.E., Ren, L., Liu, D., Wang, Y., Wild, O., Allan, J., Acton, W.J., Barlow, J., Barratt, B., Beddows, D., Bloss, W.J., Calzolai, G., Carruthers, D., Carslaw, D.C., Chan, Q., Chatzidiakou, L., Chen, Y., Crilley, L., Coe, H., Dai, T., Doherty, R., Duan, F., Fu, P., Ge, B., Ge, M., Guan, D., Hamilton, J.F., He, K., Heal, M., Heard, D., Hewitt, C.N., Hollaway, M., Hu, M., Ji, X. Jiang, R. Jones, M. Kalberer, F.J. Kelly, L. Kramer, B. Langford, C. Lin, A.C. Lewis, J. Li, W. Li, D., Liu, H., Liu, J., Loh, M., Lu, K., Lucarelli, F., Mann, G., McFiggans, G., Miller, M.R., Mills, G., Monk, P., Nemitz, E., O'Connor, F., Ouyang, B., Palmer, P.I., Percival, C., Popoola, O., Reeves, C., Rickard, A.R., Shao, L., Shi, G., Spracklen, D., Stevenson, D., Sun, Y., Sun, Z., Tao, S., Tong, S., Wang, Q., Wang, W., Wang, X., Wang, X., Wang, Z., Wei, L., Whalley, L., Wu, X., Wu, Z., Xie, P., Yang, F., Zhang, Q., Zhang, Y., Zhang, Y. and Zheng, M., Atmos. Chem. Phys., 19, 7519–7546, 2019

Assessing the impact of Clean Air Action on Air Quality Trends in Beijing Megacity using a machine learning technique

Tuan V. Vu[1], Zongbo Shi[1,3*], Jing Cheng[2], Qiang Zhang[2],

Kebin He[4,5], Shuxiao Wang[4], Roy M. Harrison[1,6*]

[1] Division of Environmental Health & Risk Management, School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham B1 52TT, United Kingdom.

[2] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China.

[3] Institute of Earth Surface System Science, Tianjin University, Tianjin, 300072, China.

[4] State Key Joint Laboratory of Environment, Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China.

[5] State Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex, Beijing 100084, China.

[6] Department of Environmental Sciences / Center of Excellence in Environmental Studies, King Abdulaziz University, PO Box 80203, Jeddah, Saudi Arabia.

* Correspondence to r.m.harrison@bham.ac.uk and z.shi@bham.ac.uk

ABSTRACT

A five-year Clean Air Action Plan was implemented in 2013 to reduce air pollutant emissions and improve ambient air quality in Beijing. Assessments of this Action Plan is an essential part of the decision-making process to review the efficacy of the Plan and to develop new policies. Both statistical and chemical transport modelling ~~were~~ have been previosuly applied to assess the efficacy of this Action Plan. However, inherent uncertainties in these methods mean that ~~a~~ new and independent methods are required to support the assessment process. Here, we applied a ~~improve a novel~~ machine learning-based random forest technique to quantify the effectiveness of Beijing's Action Plan by decoupling the impact of meteorology on ambient air quality. Our results demonstrate that meteorological conditions have an important impact on the year to year variations in ambient air quality. Further analysis show that the ~~favorable meteorological conditions in winter 2017 contributed to a lower~~ $PM_{2.5}$ mass concentration ~~(58 μg m^{-3})~~ would have broken the target of the Plan (2017 annual $PM_{2.5}$ < 60 μg m^{-3}) were it not for the meteorological conditions in winter 2017 favouring the dispersion of air pollutants~~than predicted from the random forest model (61 μg m^{-3}), which is higher than the target of the Plan (2017 annual $PM_{2.5}$ < 60 μg m^{-3})~~. However, over the whole period (2013 to 2017), ~~impact of meteorological conditions on the trend of ambient air quality are small. It is~~ the primary emission controls ~~, because of~~ required by the Action Plan~~, that has~~ have led to ~~the~~ significant reductions in $PM_{2.5}$, PM_{10}, NO_2, SO_2 and CO from 2013 to 2017~~,~~ ~~which are~~ of approximately 34%, 24%, 17%, 68%, and 33%, respectively, after meteorological correction. The marked decrease in $PM_{2.5}$ and SO_2 is largely attributable to a reduction in coal combustion. Our results indicate that the Action Plan ~~is~~ has been highly effective in reducing the primary pollution emissions and improving air quality in Beijing. The Action Plan offers a

successful example for developing air quality policies in other regions of China and other developing countries.

Keywords: Clean air action plan, Beijing, air quality, emission control, coal combustion

1. INTRODUCTION

In recent decades, China has achieved rapid economic growth and become the world's second largest economy. However, it has paid a high price in the form of serious air pollution problems caused by the rapid industrialization and urbanization associated with its fast economic growth (Lelieveld et al., 2015; Zhang et al., 2012; Guan et al., 2016). According to the World Bank, air pollution costs China's economy $159 billion (~9.9 % of GDP equivalent) in welfare losses and was associated with 1.6 million deaths in China in 2013 (Xia et al., 2016; World Bank and IHME, 2016). Accordingly, air pollution has been receiving much attention from both the public and policymakers in China, especially in Beijing - the capital of China with around 22 million inhabitants- which has suffered extremely high levels of air pollutants (Rohde and Muller, 2015; Guo et al., 2013; Zhu et al., 2012; Cai et al., 2017). To tackle air pollution problems, China's State Council released the action plan in 2013 which set new targets to reduce the concentration of air pollutants across China (CSC, 2013). Within the plan, a series of policies, control and action plans with a focus on Beijing-Tianjin-Heibei, the Yangtze River Delta and the Pearl River Delta regions were proposed. To implement the national Action Plan and further improve air quality, Beijing Municipal Government (BMG) formulated and released the "Beijing 2013-2017 Clean Air Action Plan" (the "Action Plan"), which set a target for the mean concentration of fine particles ($PM_{2.5}$, particulate matter with aerodynamic diameter less than 2.5 μm) to be below 60 μg m^{-3} by 2017

(BMG, 2013). Since then, the five-year period of 2013-2017 has seen the implementation of numerous regulations and policies in Beijing.

It is of great interest to the government, policymakers and the general public to know whether the Action Plan is working to meet the set targets. Research in this area is often termed as an air quality accountability study (HEI, 2003; Henneman et al., 2017; Cheng et al., 2018). This is highly challenging because both the actions taken to reduce the air pollutants ~~as well as~~and the meteorological conditions affect the air quality levels during a particular period (Henneman et al., 2017; Cheng et al., 2018; Liu et al., 2017; Grange et al., 2018; Chen et al., 2019). Therefore, it is essential to decouple the meteorological impact from ambient air quality data to see the real benefits in air quality by different actions.

Chemical transport models are used widely to evaluate the response of air quality to emission control policies (Wang et al., 2014; Daskalakis et al., 2016; Souri et al., 2016; Chen et al., 2019). However, there are major uncertainties in emission inventories and in the models themselves, which inevitably affect the outputs of chemical transport models (Li et al., 2017; Gao et al., 2018). Statistical analysis of ambient air quality data is another commonly used method to decouple the meteorological effects on air quality (Henneman et al., 2017; Liang et al., 2015), including the Kolmogorov-Zurbenko (KZ) filter model and deep neural networks (Wise and Comrie, 2005; Comrie, 1997; Eskridge et al., 1997; Hogrefe et al., 2003; Gardner and Dorling, 2001). Among these models, the deep neural network models showed a ~~greater~~better performance (i.e., higher correlation coefficient, lower root mean square error – RMSE) but ~~But they usually gave a poor~~

~~fitting, suggesting a poor performance of the KZ filter model, or~~ did not allow us to investigate the effect of input variables ~~in neural network models~~ (therefore it is referred as a "black- box" model) (Gardner and Dorling, 2001; Henneman et al., 2015). More recently, new approaches based on regression decision~~classification~~ trees are being developed, which are suitable for air quality weather detrending, including the boosted regression trees (BRT) and random forest (RF) algorithms (Carslaw and Taylor, 2009; Grange et al., 2018). ~~T~~These machine learning based techniques have a better performance ~~compared to~~ than the traditional statistical and air quality models by reducing variance/bias and error in high dimensional data sets (Grange et al., 2018). However, similar to the deep learning algorithms ~~such as~~including neural networks, it is hard to interpret the working mechanism inside these models ~~and~~ as well as the results. ~~Also~~In addition, the decision trees models are prone to over-fitting, especially when the number of tree nodes is large (Kotsiantis, 2013). An over-fitting problem of a random forest model is checked by its ~~performance~~ability to reproduce observations using an unseen training data set. Recently published R-packages can partly explain and visualise random forest models ~~such as~~including the importance of input variables and their interactions (Liaw and Wiener, 2018; Paluszynska, 2017).

Here, we applied~~developed~~ a ~~novel~~ machine learning technique based upon the random forest algorithm and the latest R-packages to quantify the role of meteorological conditions in air quality and thus evaluate the effectiveness of the Action Plan in reducing air pollution levels in Beijing. The results were compared with the latest emission inventory as well as results from previous study which used a chemical transport model - the Weather Research and Forecasting (WRF)-Community Multiscale Air Quality (CMAQ) model (Wong et al., 2012; Xiu and Pleim, 2001).

2. MATERIALS AND METHODS

2.1 Data Sources

Hourly air quality data for six key air pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO) was collected ~~across~~ by 12 national air quality monitoring stations in Beijing by the China National Environmental Monitoring Network (CNEM). Hourly air quality data were downloaded from the CNEM website - http://106.37.208.233:20035. Since air quality data are removed from the website on a daily basis, data were automatically downloaded to a local computer and combined to form the whole dataset for this paper. All data are now available at https://github.com/tuanvvu/Air_Quality_Trend_Analysis (last access 5 June 2019). These sites were classified in three categories (urban, suburban, and rural areas). ~~(t~~The map and categories of the~~se~~ monitoring sites ~~is~~are given in Figure S1~~,~~ and Table S1~~)~~. Hourly meteorological data including wind speed (ws), wind direction (wd), temperature ~~(temp)~~, relative humidity (RH) and pressure ~~(press.)~~ recorded at Beijing International Airport were downloaded using the “worldMet”-R package (Carslaw, 2017b). Monthly emissions ~~inventories~~ of air pollutants were from the Multi-resolution Emission Inventory for China (http://www.meicmodel.org/), and for the whole Beijing region~~s~~. Data was analyzed in R Studio with a series of packages, including the “openair”, “normalweatherr”, and “randomForestExplainer” (Liaw and Wiener, 2018; Carslaw and Ropkins, 2012; Carslaw, 2017a; Paluszynska, 2017).

2.2 Random forest m~~M~~odelling

Figure 1 shows a conceptual diagram of the data modelling and analysis which consists of three steps:

1) Building the r~~R~~andom forest (RF) model ~~development:~~

A decision tree-based random forest regression model describes the relationships between hourly concentrations of an air pollutant and ~~its~~ their predictor features~~variables~~ (including time variables~~variation~~:~~ such as~~ month 1 to 12, day of the year from 1 to 365, hour of a day from 0 to 23, and meteorological parameters: wind speed, wind direction,~~ such as~~ temperature, pressure, and relative humidity). The RF regression model is an ensemble-model which consists of hundreds of individual decision tree models. The RF model ~~was~~is described in detail in Breiman (1996 & 2001).

In the RF model, the bagging algorithm, ~~(~~which uses bootstrap aggregating~~)~~, randomly samples observations and their predictor features with replacement from a training data set. In our study, a single regression decision tree is grown in different decision rules based on the best fitting between the observed concentrations of a pollutant (response variable) and their predictor features. The predictor features are selected randomly to gives the best split for each tree node. The hourly predicted concentrations of a pollutant are given by the final decision as the outcome of the weighted average of all individual decision tree. By averaging all predictions from bootstrap samples, the bagging process decreases variance, thus helping the model to minimize over-fitting.

As shown in Figure 1, ~~T~~the whole data sets were randomly divided into~~ two with a fraction of 0.7~~: 1) a training data set to construct the random forest model and 2) a testing data set to test the model performance ~~for~~ with unseen data sets. The training data set comprised of 70% of the whole data, with the rest as testing data. ~~we firstly construct the RF model from a training data sets (70% of the all data available) of observed concentrations of a pollutant and its featurespredictor variables~~

~~and then evaluatevalidate the model by unseen data sets (testing data sets).~~ The RF model was constructed using R-"normalweatherr" packages by Grange et al. (2018).

The original data sets contain hourly concentrations of air pollutants (response) and their predictor features~~variables~~ that include time variables (t_{trend} - Unix epoch time, the day of the year, week/weekend, hour) and meteorological parameters (wind speed, wind direction, pressure, temperature, and relative humidity). These time predictor features~~variables~~ represent effects upon concentrations of air ~~pollution~~ pollutants by diurnal, weekday/weekend day and seasonal cycles and t_{trend} (Unix epoch time) represents the trend in time which captures the long-term change of air pollutant due to changes in policies/regulations, which was calculated as:

$$t_{trend} = year_i + \frac{t_{JD-1}}{N_i} + \frac{t_H}{24N_i}$$

where, N_i is the number of days in a year i (the year ith from 2013 to 2017), t_H: diurnal hour time (0-23); t_{JD}: day of the year (1-365)) (Carslaw and Taylor, 2009).

Table S2, Figure S3-S4 and Section S3 provided information on ~~T~~the performance of our model to reproduce observations ~~was evaluated based on~~ based on a number of statistical measures including mean square error (MSE)/ root mean square error (RMSE), correlation coefficients (r^2), FAC2 (fraction of predictions with a factor of two), MB (mean bias), MGE (mean gross error), NMB (normalised mean bias), NMGE (normalised mean gross error), COE (Coefficient of Efficiency), IOA (Index of Agreement) ~~for a linear regression between observed and modelled values for both training and testing data sets~~ as suggested in a number of recent papers (Emery et al. 2017, Henneman et al., 2017, and Dennis et al., 2010~~). Furthermore, other model evaluation metrics (FAC2- fraction of predictions with a factor of two, MB mean bias, MGE mean gross~~

~~error, NMB normalised mean bias, NMGE normalised mean gross error, COE Coefficient of Efficiency, IOA Index of Agreement) were also calculated (Table S3, Figure S3-S4, Section S2)~~. These results confirm that the model performs very well in comparison with traditional statistical methods and air quality models (Henneman at al., 2015).

2) Weather normalisation using the RF model

A weather normali~~z~~sation technique predicts the concentration of an air pollutant at a specific measured time point (e.g., 09:00 on 01/01/2015) with ~~various~~ randomly selected meteorological conditions~~ (term as “weather normalised concentration)~~. ~~Meteorological normalization~~ This technique was firstly introduced by Grange et al. (2018). In their method, a ~~A~~ new dataset of input predictor features ~~(~~including ~~Both~~ time variables~~:~~ (~~(month,~~ day of the year, the day of the week, hour of the day, ~~except~~but not the Unix time variable) and meteorological parameters~~:~~ (wind speed, wind direction, temperature and RH) is firstly generated (i.e., re-sampled) randomly ~~based on~~from the original ~~input~~observation dataset. For example, for a particular day (e.g., 01/01/2011), the model randomly selects the time variables (excluding Unix time) and weather –parameters ~~conditions~~ at any day from the data set of predictor features during the whole study period. This is repeated 1,000 times to provide the new input data set for a particular day. ~~And then,~~ The input data set is then fed to~~, except the trend variable were re-sampled randomly and was added into~~ the random forest model ~~will~~as ~~input variables to~~ to predict the concentration of a pollutant at a particular day ~~based on the new input data sets~~ (Grange et al., 2018; Grange and Carslaw, 2019). This gives a total of 1,000 predicted concentrations for that day. The final concentration of that pollutant, referred hereafter as ~~meteorological~~ weather normalised concentration, is calculated by averaging the 1000 predicted concentrations~~predictions from the RF model~~. ~~By this way, the model~~

~~results in a predicted concentration of pollutant by normalization~~ This method normalises ~~of~~ the impact of both seasonal and weather variations. ~~However~~Therefore, it is unable to investigate the seasonal variation of trends for a comparison with the trend of primary emissions. ~~Therefore~~For this reason, we enhanced the meteorological ~~normaliz~~normalisation procedure.

In our algorithm, we firstly generated ~~the~~a new input data set of predictor features~~s, (~~which ~~contains:~~includes original time variables and re-sampled weather data (wind speed, wind direction, temperature, and relative humidity)~~Unix time, day of the year, week/weekend day, hour of the day variables, wind speed, wind direction, temperature, and relative humidity during 2013-2017). with new only weather data (MET data) sets were re-sampled from thirty-year data sets (1988-2017) of weather in Beijing. We also enhanced modified the code to re-sample the MET data for a long term period rather than MET data during the conducted study from 2013-2017. In particular, Tthirty-year MET in Beijing (1988-2017)~~ Specifically, weather variables at a specific selected hour of a particular day in the input data sets were generated by randomly selecting from the observed weather data (i.e., 1988-2017 or 2013-2017) at that particular hour of different dates within a four-week period (i.e., 2 weeks before and 2 weeks after that selected date). For example, the new input weather data at 08:00 15/01/2015 are randomly selected from the observed data at 08:00 am on any date from 1^{st} to 29^{th} January of any year in 1988-2017 or 2013-2017. The selection process was repeated automatically 1,000 times to generate a final input data set. Each of the 1,000 data was then fed to the random forest model to predict the concentration of a pollutant. The 1,000 predicted concentrations were then averaged to calculate the final weather normalised concentration for that particular hour, day, and year. This way, unlike Grange et al., (2018), we only normalise the weather conditions but not the seasonal and diurnal variations.

Furthermore, we are able to re-sample observed weather data for a longer period (for example, 1998-2017), rather than only the study period. This new approach enables us investigate the seasonality of weather normalised concentrations and compare them with primary emissions from inventories.

~~was used to enable a better representation of average meteorological conditions. Specifically, MET data variables at a specific selected hour of a particular day in the input data sets was replaced randomly by the MET data at that hour for a period of 2 weeks before and after that selected data in the 30 year MET data set (1988-2017). For example, the MET data at 8:00 15/01/2015 could be randomly replaced by the MET data at 8:00 am in any date from 1st to 30th January of any year in 1988-2017. Similar to Grange's approach, with each a new input dataset we generated the concentration of a pollutant based on a random forest model which was built in the step one. We repeated this generation process by a thousand times, and the final concentration of a pollutant (weather normalized concentration) was calculated as an average of all values from each generation process.~~

3) Quantifying long-term trend using Theil-Sen estimator~~:~~

The Theil-Sen regression technique was performed on~~estimates~~ the concentrations of air pollutants after meteorological normalis~~z~~ation to investigate the long-term trend of pollutants~~ to calculate their long-term trends~~. The Theil-Sen approach which computes the slopes of all possible pairs of pollutant concentrations and takes the median value, has been commonly used for long-term trend analysis over recent years. By selecting the median of the slopes, the Theil-Sen estimator tends to give us accurate confidence intervals even with non-normal data and non-constant error variance (Sen, 1968). The Theil-Sen function is provided via the "openair" package in R.

2.3. Notices, regulations and policies for air pollution control in Beijing

The five-year period of 2013-2017 saw the implementation of numerous regulations and policies. The "Beijing Clean Air Action Plan 2013-2017" proposed eight key regulations including: (1) Controlling the city development intensity, population size, vehicle ownership, and environmental resources, (2) Restructuring energy by reducing coal consumption, supplying clean and green energy, and improving energy efficiency, (3) promoting public transport, implementing stricter emission standards, eliminating old vehicles and encouraging new and clean energy vehicles, (4) Optimizing industrial structure by eliminating polluting capacities, closing small polluting enterprises, building eco-industrial parks and pursuing cleaner production, (5) Strengthening treatment of air pollutants and tightening environmental protection standards, (6) Strengthening urban management and regulation enforcement, (7) Preserving the ecological environment by enhancing green coverage and water area, and (8) Strengthening emergency response to heavy air pollution. We collected more than 70 major notices and policies on air pollution control ~~during~~ from the Beijing government website (http://zhengce.beijing.gov.cn/library/). Most important regulations were related to energy system re-structuring and vehicle emissions (Section S2). These key measures include: 1) Reform and upgrade Action Plan for coal energy conservation and emission reduction (2014); 2) "no-coal zone" for Beijing-Tianjin-Hebei regions in October 2014; 3) Beijing implemented the fifth phase emission standards for new light-duty gasoline vehicles (LDVs) and heavy-duty diesel vehicles (HDVs) for public transport in 2013; 4) traffic restrictions to yellow-label and non-local vehicles to enter the city within the sixth ring road during daytime since 2015.

3. RESULTS AND DISCUSSIONS

3.1 Observed Levels of Air Pollution in Beijing During 2013-2017

The a~~A~~nnual mean concentration of $PM_{2.5}$ and PM_{10} in Beijing measured from the 12 national air quality monitoring stations declined by 34 and 19 % from 88 and 110 µg m^{-3} in 2013 to 58 and 89 µg m^{-3} in 2017, respectively. Similarly, the annual mean levels of NO_2 and CO decreased by 16 and 33 % from 54 µg m^{-3} and 1.4 mg m^{-3} to 45 µg m^{-3} and 0.9 mg m^{-3} while the annual mean concentration of SO_2 showed a dramatic drop by 68 % from 23 µg m^{-3} in 2013 to 8.0 µg m^{-3} in 2017. Along with the decrease of annual mean concentration, the number of haze days (defined as $PM_{2.5}$ > 75 µg m^{-3} here) also decreased (Figure S7~~6~~). These results confirm a significant improvement of air quality and that Beijing ~~seem~~ appeared to have achieved its $PM_{2.5}$ target under the Action Plan (annual average $PM_{2.5}$ target for Beijing is 60 µg m^{-3} in 2017). On the other hand, the annual mean concentration of $PM_{2.5}$ is still substantially higher than ~~the~~ China's national ambient air quality standard (NAAQS-II) of 35 µg m^{-3} (Table S3~~2~~~~1~~) and the WHO Guideline of 10 µg m^{-3}. While PM_{10}, $PM_{2.5}$, SO_2, NO_2 and CO showed a decreasing trend, the annual average concentration of O_3 increased slightly by 4.9 % from 58 µg m^{-3} in 2013 to 61 µg m^{-3} in 2017. The number of days exceeding NAAQS-II standards for O_3-8h averages (160 µg m^{-3}) during the period 2013-2017 was 329, accounting for 18 % of total days.

3.2 Air Quality Trends After Weather Normali~~z~~sation

A key aspect in evaluating the effectiveness of air quality policies is to quantify separately the impact of emission reduction and meteorological conditions on air quality (Carslaw and Taylor, 2009;Henneman et al., 2017), as these are the key factors regulating air quality. By applying a random forest algorithm, we ~~decoupled the effect of meteorological condition to~~ showed the normalis~~z~~ed air quality parameters, ~~—~~under the ~~condition of the~~ 30-year average (1988-2017) meteorological conditions (Figure 2). The temporal variations of ambient concentrations of monthly average $PM_{2.5}$, PM_{10}, CO, and NO_2 do not ~~offer a clear~~ show a smooth trend from 2013 to 2017 because of the spikes ~~in the winters~~during pollution events. However, after the weather normalis~~z~~ation, we can clearly see the decreasing ~~true~~ real trend (Figure 2). The trends of the normalis~~z~~ed air quality parameters represent the effects of emission control and, in some cases, associated chemical processes (for example, for ozone, $PM_{2.5}$, PM_{10}). SO_2 showed a dramatic decrease while ozone increased year by year (Figure 2). The normalis~~z~~ed annual average levels of $PM_{2.5}$, PM_{10}, SO_2, NO_2, and CO decreased by 7.4, 7.6, 3.1, 2.5, and 94 μg m^{-3} $year^{-1}$, respectively, whereas the level of O_3 increased by 1.0 μg m^{-3} $year^{-1}$.

Table 1 compares the trends of air pollutants before and after normalis~~z~~ation, which are largely different depending on meteorological conditions. For example, the annual average concentration of fine particles ($PM_{2.5}$) after weather normalis~~z~~ation was 61 μg m^{-3} in 2017, which was higher than their observed level of 58 μg m^{-3} by ~~about~~ 5.2%. This suggests that Beijing would have missed its $PM_{2.5}$ target of 60 μg m^{-3} if not for the favorable meteorological conditions in winter 2017 and the emission reduction contributed to 10 μg m^{-3} out of the 13 μg m^{-3} (77%) $PM_{2.5}$ reduction (71 to 58 μg m^{-3}) from 2016 to 2017. Overall, the emission control led to a 34%, 24%, 17%, 68%, and 33% reduction in normalis~~z~~ed mass concentration of $PM_{2.5}$, PM_{10}, NO_2, SO_2 and CO respectively from 2013 to 2017 (Table 1).

When meteorological conditions were randomly selected from 2013-2017 (instead of 1998-2017) in the RF model, the normalis~~z~~ed level of $PM_{2.5}$ in 2017 was 60 μg m^{-3}, which is 1 μg m^{-3} difference to that using 1998-2017 data. This difference is due to the variation of the long-term climatology

(1998-2017) to the 5 year period (2013-2017)~~. This indicates that our modelling results are robust. Additional uncertainty in the meteorological normalised levels of $PM_{2.5}$ obtained from a random forest model is discussed later in Section 3.3.~~

The observed $PM_{2.5}$ mass concentration reduced by 30 µg m^{-3} from 2013 to 2017, whereas the normalis~~z~~ed values reduced by 32 µg m^{-3}. Similarly, the observed PM_{10} and SO_2 mass concentration reduced by 30 and 15.5 µg m^{-3} from 2013 to 2017, whereas the normalis~~z~~ed values ~~by~~ were 33 and 17.9 µg m^{-3}. These results suggest that the effect of emission reduction would have contributed to an even better improvement in air quality (except ozone) from 2013 to 2017 if not for meteorological variations year by year.

Figure 3 shows that the Action Plan has been ~~highly effective~~led to a major improvement ~~in improving~~ in the air quality of Beijing at both the urban, suburban and rural sites, particularly for SO_2 (16-18 % year^{-1}), CO (8-9 % year^{-1}), and $PM_{2.5}$ (6-8 % year^{-1}). The Action Plan also led to a decrease in PM_{10} and NO_2 but to a lesser extent than that of CO, SO_2 and $PM_{2.5}$, indicating that PM_{10} and NO_2 were ~~significantly~~ affected by other less well controlled sources or ~~they are affected differently than the other pollutants due to their~~ different atmospheric processes~~. For example, Figure 2 suggested that the high levels of PM_{10} in spring were mostly affected by the frequent Asian dust events~~. Urban sites showed a bigger decrease in $PM_{2.5}$, PM_{10}, and SO_2 concentrations in comparison to the rural and suburban sites (Figure 3).

3.3 Impact of Meteorological Conditions on $PM_{2.5}$ levels: A Comparison with Results from CMAQ-WRF Model

We compared our RF modelling results with those from an independent method by Cheng et al. (2018) who evaluated the de-weathered trend by simulating the monthly average $PM_{2.5}$ mass concentrations in 2017 by the CMAQ model with meteorological conditions of 2013, 2016 and 2017 from the WRF model. The WRF-CMAQ results ~~show~~ predict that the annual average $PM_{2.5}$ concentration of Beijing in 2017 is 61.8 and 62.4 µg m^{-3} ~~if~~ under the 2013 and 2016 meteorological conditions respectively, both of which are higher than the measured value – 58 µg m^{-3}. Thus, the modelled results are similar to those from the machine learning technique~~s~~, which gave a weather-normalis~~z~~ed $PM_{2.5}$ mass concentration of 61 µg m^{-3} in 2017.

Figure 4 also shows that the $PM_{2.5}$ concentrations would have been significantly higher in November and December ~~in~~ 2017 if under the meteorological conditions of 2016. In contrast, the $PM_{2.5}$ concentrations would have been lower in spring 2017 ~~of~~ under the ~~MET~~ meteorological conditions ~~data~~ of 2016 or the 30-year normalised ~~MET~~ meteorological data. ~~Since severe $PM_{2.5}$ pollution and haze events frequentlyalmost always occur in winter in Northern China (Cai et al., 2017), t~~The more favourable meteorological conditions in the two months contributed appreciably to the lower measured annual average $PM_{2.5}$ level in 2017. It also suggests that the monthly levels of $PM_{2.5}$ strongly depend upon the monthly variation of weather.

Comparison of model uncertainties from the two methods

Figure 5 compares observation and prediction of monthly concentrations of $PM_{2.5}$ by the WRF-CMAQ model and the RF model. The correlation coefficient r^2 between monthly values was 0.82, whereas that from the random forest method is >0.99 for both the training and test data sets. The difference between the monthly observed $PM_{2.5}$ values and those simulated by the WRF-CMAQ model ranged from 3 to 33.6%, resulting in 7.8% difference in the yearly value. ~~By~~ In contrast, the deviation between observed and predicted $PM_{2.5}$ value from the RF model ranges from 0.4-7.9% with an average of 1.5%. In the modelled concentration of $PM_{2.5}$ from the random forest technique,

~~the s~~Standard ~~variation~~ deviation of the 1,000 predicted concentration of $PM_{2.5}$ in 2017 ~~those 1000 predictions by a random forest~~ is only 0.35 μg m^{-3}, accounting for~~ed~~ 0.6% of the observed $PM_{2.5}$ concentration~~s in 2017~~.

3.4 Evaluating the Effectiveness of the Mitigation~~s~~ Measures in the Clean Air Action Plan

The weather normalised air quality trend (Figure 2) allows us to assess the effectiveness of various policy measures to improve air quality to some extent. In particular~~ly~~, the SO_2 normalis~~z~~ed trend clearly shows that the peak monthly concentration~~s~~ in the winter months decreased from 60 μg m^{-3} in January 2013 to less than 10 μg m^{-3} in December 2017 (Figure 2). This indicates that the control of emissions from winter-specific sources was highly successful in reducing SO_2 concentrations. The Multi-resolution Emission Inventory for China (MEIC) shows a major decrease in SO_2 emissions from heating (both industrial and centralized heating) and residential sector (mainly coal combustion) (Figure S8~~7~~), which is consistent with the trend analyses. On the other hand, the “~~based line~~baseline” SO_2 concentration ~~—~~defined as the minimum monthly concentration ~~the lowest ones~~ in the summer (Figure 2) – also reduced somewhat during the same period. ~~The “based line”~~ SO_2 in the summer mainly came from non-seasonal ~~(winter)~~ sources including power plants, industry, and transportation (Figure S9~~7~~). Overall, the MEIC estimated that SO_2 emissions decreased by 71 % from 2013 to 2017 (Figure S8~~7~~), which is close to the 67% decrease in the weather normali~~z~~sed concentration of SO_2 (Table 1). According to the Beijing Statistical Year Books (2012-2017), coal consumption in Beijing declined remarkably by 56 % in 6 years as shown in Figure 6 (Karplus et al., 2018;BMBS, 2013-2017). The slightly faster decrease in SO_2 concentrations relative to coal consumption (Figure S9~~8~~) was attributed to the adoption of clean coal technologies that were enforced by the "Action Plan for Transformation and Upgrading of Coal Energy Conservation and Emission Reduction (2014-2020)" (Karplus et al., 2018; Chang et al., 2016). In summary, energy re-structuringe~~e~~, e.g., replacement of coal with natural gas (Figure 6; Section S2), is ~~the~~ a highly~~most~~ effective measure in reducing ambient SO_2 pollution in Beijing.

Coal combustion is not only a major source of SO_2, but also an important source of NO_x and primary particulate matter (PM) in Beijing (Streets and Waldhoff, 2000; Zíková et al., 2016; Lu et al., 2013; Huang et al., 2014). Precursor gases ~~such as~~including SO_2 and NO_x from coal combustion also contribute to secondary aerosol formation (Lang et al., 2017). The MEIC emission inventory showed that 8.8-29 % of NO_x was emitted from heating, power and residential activities, primarily associated with coal combustion. As shown in Figure S9~~8~~, the normalis~~z~~ed NO_2 concentration is also decreasing, but much slower than that of SO_2. Most notably, the level of SO_2 dropped rapidly in 2014 but the level of NO_2 decrease by a small proportion. The different trends between SO_2 and NO_2 indicate that other sources (e.g. traffic emissions, Figure S9~~8~~) or atmospheric processes have a greater influence on ambient concentration of NO_2 than coal combustion. For examples~~, although~~ the chemistry of the $NO/NO_2/O_3$ system will tend to "buffer" changes in NO_2 causing non-linearity in NO_x-NO_2 relationships (Marr and Harley, 2002). NO_2 concentrations decreased more rapidly from January 2015, ~~particularly~~ specifically by 17%, 18%, 10%, 15% (Figure 2) in the first six months of 2015, which suggests that emission control measures implemented in 2015 were effective. These measures~~, including~~ include regulations on spark ignition light vehicles to meet the national fifth phase standard, and expanded traffic restrictions to certain vehicles, including banning entry of high polluting and non-local vehicles to the city within the sixth ring road during daytime, and phasing out of 1 million old vehicles (Yang ~~Z~~et al., 2015) (Section S2).

Normalis~~z~~ed $PM_{2.5}$ decreased faster than NO_2, but slower than SO_2 (Figure S9~~8~~). Yearly peak ~~normaliz~~normalised $PM_{2.5}$ concentrations decreased from 2013-14 to 2015-2016 but slighted rebounded in 2016-2017. The monthly ~~normaliz~~normalised peak $PM_{2.5}$ concentration reduced from 115 μg m^{-3} in Jan 2013 to 60 μg m^{-3} in Dec 2017. The biggest drop is seen in winter 2017, which decreased by more than half from the peak value in winter 2016, suggesting that the "no coal zone" policy (Section S2) to reduce pollutant emissions from winter specific sources (i.e., heating and residential sectors) ~~were~~ was highly effective in reducing $PM_{2.5}$. The ~~normaliz~~normalised "~~based line~~baseline" concentration – ~~lowest~~ minimum monthly average concentration ~~values~~ in ~~each year~~the summer – also decreased from 71 μg m^{-3} in summer 2013 to 42 μg m^{-3} in summer 2017. This suggests that non-heating emission sources, ~~such as~~including industry, industrial heating and power plants also contributed to the decrease in $PM_{2.5}$ from 2013 to 2017. These are broadly consistent with the $PM_{2.5}$ and SO_2 emission trends in MEIC (Figure S8~~7~~). A small peak in both $PM_{2.5}$ and CO in June/July seen in Figure 2 from 2013 to 2016 attributed to agricultural burning almost disappeared over the period of the measurements and simulations in 2017, suggesting the ban on open burning is effective.

The ~~normaliz~~normalised trend of PM_{10} is similar to that of $PM_{2.5}$, except that the rate of decrease is slower. The trend agrees well with PM_{10} primary emissions for the summer (Figure S8~~7~~). The biggest drop in peak monthly PM_{10} concentration is seen in winter 2017, which decreased by more than half from the peak value in winter 2016, suggesting that "no coal zone" policy (Section S2) to reduce pollutant emission from winter specific sources (i.e., heating and residential sectors) were highly effective in reducing PM_{10}, ~~similar to that of~~as with $PM_{2.5}$. The rate of decrease of peak monthly PM_{10} emission is slower than that of weather normalised PM_{10} concentrations,~~,~~ which may suggest an underestimation of the decrease ~~in~~ by the MEIC. The ~~normaliz~~normalised "~~based line~~baseline" concentration ~~—~~(minimum monthly average concentration, Figure 2)~~lowest values in summer (Figure 2) The "based line" of a pollutant (except for ozone) was the defined as the lowest concentration of air pollutions in the summer (the summer concentrations)~~ ~~—~~ also decreased ~~from~~ substantially from 2013 to 2017. This indicates that non-heating emission sources, ~~such as~~including industry, industrial heating and power plants also contributed to the decrease in PM_{10}. This is consistent with th~~os~~e trends in MEIC (Figure S8~~7~~). The peaks in the spring are attributed to Asian dust events.

The ~~normaliz~~normalised CO trend shows that the peak CO concentration reduced by approximately 50% from 2013 to 2017 with the largest drop from 2016 to 2017 (Figure 2). The decreasing trend in total emission of CO in the MEIC is slower from 2015 to 2017, suggesting that ~~the~~ CO emission in the MEIC may be overestimated in these two years. During 2013-2016, the CO level decreased by 26 % and 34 % for ~~both~~ winter and summer~~ ("baseline")~~. Similar to the ~~normaliz~~normalised $PM_{2.5}$ trend, a small peak of CO concentration occurred in Jun-July during 2013-2016, which is likely associated with open biomass burning around the Beijing region. This peak disappeared in 2017. A major decrease in ~~normaliz~~normalised CO levels in winter 2017 is attributed to the "no-coal zone" policy (see below Section S2; Figure S8~~7~~).

3.5 Implications and Future Perspectives

We have applied a machine learning based model to identify the key mitigation measures contributing to the reduction of air pollutant concentrations in Beijing. However, three challenges remain. Firstly, it is not always straightforward to link a specific mitigation measure to improvement in air quality quantitatively. This is because often more than two measures were implemented ~~at~~ on a similar timescale, making it difficult to disentangle the impacts. Secondly, we were not able to compare the calculated benefit for each mitigation measure with ~~the~~ that intended ~~one designed~~ by the government due to a lack of information~~data~~ about the implemented policies, for example, ~~such as~~ the start/end date of air pollution control actions. If data on the intended benefits are known, this will further enhance the value of this type of study. Thirdly, the ozone level increased slightly during 2013-2017, especially for the summer periods (Table 1). Because ozone is a secondary pollutant, interpretation of the effects of emission changes ~~it is not possible to directly compare the trend with emission~~ of precursor pollutants is ~~. The mechanisms of this increase are~~ complex and ~~out of~~ beyond the scope of this study.

Our results confirm~~ed~~ that the “Action Plan” has been led to a major ~~highly effective in~~ improvement in the~~ing~~ real (~~normaliz~~normalised) air quality of Beijing (Figure 3). However, it would have failed to meet the target for annual average $PM_{2.5}$ concentrations if not for better than average air pollutant dispersion (meteorological) conditions in 2017. This suggests that future target setting should consider meteorological conditions. Major challenges remain in reducing the $PM_{2.5}$ levels to below Beijing’s own targets, as well as China’s national air quality standard and WHO guidelines. Another challenge is to reduce the NO_2 and O_3 levels, which show little decrease or even an increase from 2013 to 2017. The lessons learned in Beijing thus far may prove beneficial to other cities as they develop their own clean air strategies.

ACKNOWLEDGMENTS

Funding: This research is supported by the NERC funding though AIRPOLL-Beijing project within the APHH programme (NE/N007190/1), Met Office CSSP-China (Scoping Study on Air Quality Climate Service) and National Natural Science Foundation of China (41571130032 and 4151130035).

Author contributions: This study was conceived by Z.S. and T.V.. Statistical modelling was performed by T.V. and CMAQ modelling was performed by J.C, Q.Z., S.W. and K.H. T.V, Z.S, and R.M.H drafted the manuscript. All authors revised the manuscript and approved the final version for publication.

Competing interests: The authors declare no competing interests.

REFERECES

BMBS: Beijing Municipal Bureau of Statistics (BMBS): Beijing Statistical Yearbook http://www.bjstats.gov.cn/nj/main/2017-tjnj/zk/indexeh.htm (update 30/08/2018), 2013-2017.

BMG: Beijing Municipal Government (BMG): Clean Air Action Plan (2013-2017). Available online: http://www.bjyj.gov.cn/flfg/bs/zr/t1139285.html, 2013.

Breiman, L.: Bagging predictors, Mach. Learn., 24, 123–140, https://doi.org/10.1007/BF00058655, 1996.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001

Cai, W., Li, K., Liao, H., Wang, H., and Wu, L.: Weather conditions conducive to Beijing severe haze more frequent under climate change, Nature Climate Change, 7, 257, 10.1038/nclimate3249 https://www.nature.com/articles/nclimate3249#supplementary-information, 2017.

Carslaw, D. C., and Taylor, P. J.: Analysis of air pollution data at a mixed source location using boosted regression trees, Atmospheric Environment, 43, 3563-3570, https://doi.org/10.1016/j.atmosenv.2009.04.001, 2009.

Carslaw, D. C., and Ropkins, K.: openair — An R package for air quality data analysis, Environmental Modelling & Software, 27-28, 52-61, https://doi.org/10.1016/j.envsoft.2011.09.008, 2012.

Carslaw, D. C.: Normalweather: R package to conduct meteorological/weather normalisation on air quality, Available on: https://github.com/davidcarslaw/normalweatherr, 2017a.

Carslaw, D. C.: Worldmet: Import Surface Meteorological Data from NOAA Integrated Surface Database (ISD), Available on:http://github.com/davidcarslaw/, 2017b.

Chang, S., Zhuo, J., Meng, S., Qin, S., and Yao, Q.: Clean Coal Technologies in China: Current Status and Future Perspectives, Engineering, 2, 447-459, https://doi.org/10.1016/J.ENG.2016.04.015, 2016.

Chen, D., Liu, Z., Ban, J., Zhao, P., Chen, M.: Retrospective analysis of 2015-2017 wintertime PM2.5 in China: resposne to emission regulations and the role of meteorology, Atmosperic Chemistry and Physics, 19, 7409-7427, 10.5149/acp-19-7409-2019.

Cheng, J., Su, J., Cui, T., Li, X., Dong, X., Sun, F., Yang, Y., Tong, D., Zheng, Y., Li, J., Zhang, Q., and He, K.: Dominant role of emission reduction in PM2.5 air quality improvement in Beijing during 2013-2017: a model-based decomposition analysis, Atmos. Chem. Phys. Discuss., 2018, 1-31, 10.5194/acp-2018-1145, 2018.

Comrie, A. C.: Comparing Neural Networks and Regression Models for Ozone Forecasting, Journal of the Air & Waste Management Association, 47, 653-663, 10.1080/10473289.1997.10463925, 1997.

CSC: China State Council (CSC)'s notice on the Air Pollution Prevention and Control Action Plan, Available online: http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm, 2013.

Daskalakis, N., Tsigaridis, K., Myriokefalitakis, S., Fanourgakis, G. S., and Kanakidou, M.: Large gain in air quality compared to an alternative anthropogenic emissions scenario, Atmos. Chem. Phys., 16, 9771-9784, 10.5194/acp-16-9771-2016, 2016.

Dennis, R., T. Fox, M. Fuentes, A. Gilliland, S. Hanna, C. Hogrefe, J. Irwin, S.T. Rao, R, Scheffe, K. Schere, D.A. Steyn, and A. Venkatram. A framework for evaluating regio- nal-scale numerical photochemical modeling systems. J. Environ. Fluid Mech.10, 471–89, 2010. doi: 10.1007/s10652-009- 9163-2, 2010.

Emery, C., Liu, Z., Russell, A., Talat Odman, M., Yarwood, G., & Kumar, N. Recommendations on ~~S~~statistics and b~~B~~enchmarks to a~~A~~ssess ~~P~~photochemical ~~M~~model ~~P~~performance. J. Air & Waste Manage. Asso., 67, 582-598, doi: 10.1080/10962247.2016.1265027, 2017.

Eskridge, R. E., Ku, J. Y., Rao, S. T., Porter, P. S., and Zurbenko, I. G.: Separating Different Scales of Motion in Time Series of Meteorological Variables, Bulletin of the American Meteorological Society, 78, 1473-1484, 10.1175/1520-0477(1997)078<1473:SDSOMI>2.0.CO;2, 1997.

Gao, M., Han, Z., Liu, Z., Li, M., Xin, J., Tao, Z., Li, J., Kang, J. E., Huang, K., Dong, X., Zhuang, B., Li, S., Ge, B., Wu, Q., Cheng, Y., Wang, Y., Lee, H. J., Kim, C. H., Fu, J. S., Wang, T., Chin, M., Woo, J. H., Zhang, Q., Wang, Z., and Carmichael, G. R.: Air quality and climate change, Topic 3 of the Model Inter-Comparison Study for Asia Phase III (MICS-Asia III) – Part 1: Overview and model evaluation, Atmos. Chem. Phys., 18, 4859-4884, 10.5194/acp-18-4859-2018, 2018.

Gardner, M., and Dorling, S.: Artificial Neural Network-Derived Trends in Daily Maximum Surface Ozone Concentrations AU - Gardner, Matthew, Journal of the Air & Waste Management Association, 51, 1202-1210, 10.1080/10473289.2001.10464338, 2001.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random forest meteorological normalisation models for Swiss PM10 trend analysis, Atmos. Chem. Phys., 18, 6223-6239, 10.5194/acp-18-6223-2018, 2018.

Grange, S. K., and Carslaw, D. C.: Using meteorological normalisation to detect interventions in air quality time series, Science of The Total Environment, 653, 578-588, https://doi.org/10.1016/j.scitotenv.2018.10.344, 2019.

Guan, W.-J., Zheng, X.-Y., Chung, K. F., and Zhong, N.-S.: Impact of air pollution on the burden of chronic respiratory diseases in China: time for urgent action, The Lancet, 388, 1939-1951, 10.1016/S0140-6736(16)31597-5, 2016.

Guo, Y., Li, S., Tian, Z., Pan, X., Zhang, J., and Williams, G.: The burden of air pollution on years of life lost in Beijing, China, 2004-08: retrospective regression analysis of daily deaths, BMJ : British Medical Journal, 347, 2013.

HEI: Assessing health impact of air quality regulations: Concepts and methods for accountability research, Health Effects Institute, Accountability Working Group, Comunication 11, 2003.

Henneman, L. R. F., Holmes, H. A., Mulholland, J. A., and Russell, A. G.: Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000–2012) data in Atlanta, Atmospheric Environment, 119, 201-210, https://doi.org/10.1016/j.atmosenv.2015.08.007, 2015.

Henneman, L. R. F., Liu, C., Mulholland, J. A., and Russell, A. G.: Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks, Journal of the Air & Waste Management Association, 67, 144-172, 10.1080/10962247.2016.1242518, 2017.

Henneman, L. R., Liu, C., Hu, Y., Mulholland, J. A., and Russell, A. G.: Air quality modeling for accountability research: Operational, dynamic, and diagnostic evaluation, Atmospheric Environment, 166, 551–565, https://doi.org/10.1016/j.atmosenv.2017.07.049, 2017.

Hogrefe, C., Vempaty, S., Rao, S. T., and Porter, P. S.: A comparison of four techniques for separating different time scales in atmospheric variables, Atmospheric Environment, 37, 313-325, https://doi.org/10.1016/S1352-2310(02)00897-X, 2003.

Huang, R.-J., Zhang, Y., Bozzetti, C., Ho, K.-F., Cao, J.-J., Han, Y., Daellenbach, K. R., Slowik, J. G., Platt, S. M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S. M., Bruns, E. A., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., An, Z., Szidat, S., Baltensperger, U., Haddad, I. E., and Prévôt, A. S. H.: High secondary aerosol contribution to particulate pollution during haze events in China, Nature, 514, 218, 10.1038/nature13774. https://www.nature.com/articles/nature13774#supplementary-information, 2014.

Karplus, V. J., Zhang, S., and Almond, D.: Quantifying coal power plant responses to tighter SO₂ emissions standards in China, Proceedings of the National Academy of Sciences, 115, 7004, 10.1073/pnas.1800605115, 2018.

Kotsiantis, S. B.: Decision trees: a recent overview, Artif. Intell. Rev., 39, 261–283, https://doi.org/10.1007/s10462-011-9272-4, 2013.

Lang, J., Zhang, Y., Zhou, Y., Cheng, S., Chen, D., Guo, X., Chen, S., Li, X., Xing, X., and Wang, H.: Trends of PM2.5 and Chemical Composition in Beijing, 2000–2015, Aerosol and Air Quality Research, 17, 412-425, 10.4209/aaqr.2016.07.0307, 2017.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525, 367, 10.1038/nature15371, 2015.

Li, M., Liu, H., Geng, G., Hong, C., Tong, D., Geng, G., Cui, H., Zhang, Q., Li, M., Zheng, B., Liu, F., Man, H., Liu, H., He, K., and Song, Y.: Anthropogenic emission inventories in China: a review, National Science Review, 4, 834-866, 10.1093/nsr/nwx150, 2017.

Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., and Chen Song, X.: Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 471, 20150257, 10.1098/rspa.2015.0257, 2015.

Liaw, A., and Wiener, M.: R- Package "ramdom Forest", Available on: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf, 2018.

Liu, T., Gong, S., He, J., Yu, M., Wang, Q., Li, H., Liu, W., Zhang, J., Li, L., Wang, X., Li, S., Lu, Y., Du, H., Wang, Y., Zhou, C., Liu, H., and Zhao, Q.: Attributions of meteorological and emission factors to the 2015 winter severe haze pollution episodes in China's Jing-Jin-Ji area, Atmos. Chem. Phys., 17, 2971-2980, 10.5194/acp-17-2971-2017, 2017.

Lu, Q., Zheng, J., Ye, S., Shen, X., Yuan, Z., and Yin, S.: Emission trends and source characteristics of SO2, NOx, PM10 and VOCs in the Pearl River Delta region from 2000 to 2009, Atmospheric Environment, 76, 11-20, https://doi.org/10.1016/j.atmosenv.2012.10.062, 2013.

Marr, L. C., and Harley, R. A.: Modeling the Effect of Weekday−Weekend Differences in Motor Vehicle Emissions on Photochemical Air Pollution in Central California, Environmental Science & Technology, 36, 4099-4106, 10.1021/es020629x, 2002.

Paluszynska, A.: randomForestExplainer: Explaining and Visualizing Random Forests in Terms of Variable Importance, Available on:https://github.com/MI2DataLab/randomForestExplainer, 2017.

Rohde, R. A., and Muller, R. A.: Air Pollution in China: Mapping of Concentrations and Sources, PLOS ONE, 10, e0135749, 10.1371/journal.pone.0135749, 2015.

Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau AU - Sen, Pranab Kumar, Journal of the American Statistical Association, 63, 1379-1389, 10.1080/01621459.1968.10480934, 1968.

Souri, A. H., Choi, Y., Jeon, W., Li, X., Pan, S., Diao, L., and Westenbarger, D. A.: Constraining NOx emissions using satellite NO2 measurements during 2013 DISCOVER-AQ Texas campaign, Atmospheric Environment, 131, 371-381, https://doi.org/10.1016/j.atmosenv.2016.02.020, 2016.

Streets, D. G., and Waldhoff, S. T.: Present and future emissions of air pollutants in China:: SO2, NOx, and CO, Atmospheric Environment, 34, 363-374, https://doi.org/10.1016/S1352-2310(99)00167-3, 2000.

Wang, S., Xing, J., Zhao, B., Jang, C., and Hao, J.: Effectiveness of national air pollution control policies on the air quality in metropolitan areas of China, Journal of Environmental Sciences, 26, 13-22, https://doi.org/10.1016/S1001-0742(13)60381-2, 2014.

Wise, E. K., and Comrie, A. C.: Extending the Kolmogorov–Zurbenko Filter: Application to Ozone, Particulate Matter, and Meteorological Trends, Journal of the Air & Waste Management Association, 55, 1208-1216, 10.1080/10473289.2005.10464718, 2005.

Wong, D. C., Pleim, J., Mathur, R., Binkowski, F., Otte, T., Gilliam, R., Pouliot, G., Xiu, A., Young, J. O., and Kang, D.: WRF-CMAQ two-way coupled system with aerosol feedback: software development and preliminary results, Geosci. Model Dev., 5, 299-312, 10.5194/gmd-5-299-2012, 2012.

World Bank, and IHME: World Bank and Institue for Health Metrics and Evaluation:The Cost of Air Polllution: Strengthening the Economic Case for Action, World Bank: Washington, DC, USA, 2016.

Xia, Y., Guan, D., Jiang, X., Peng, L., Schroeder, H., and Zhang, Q.: Assessment of socioeconomic costs to China's air pollution, Atmospheric Environment, 139, 147-156, https://doi.org/10.1016/j.atmosenv.2016.05.036, 2016.

Xiu, A., and Pleim, J. E.: Development of a Land Surface Model. Part I: Application in a Mesoscale Meteorological Model, Journal of Applied Meteorology, 40, 192-209, 10.1175/1520-0450, 2001.

Yang Z, W. H., Shao Z, Muncrief R: Review of Beijing's Comprehensive motor vehicle emission Control program, Communication, 2015.

Zhang, Q., He, K., and Huo, H.: Cleaning China's air, Nature, 484, 161, 10.1038/484161a, 2012.

Zhu, T., Melamed, M. L., Parrish, D., Gauss, M., Klenner, L. G., Lawrence, M., Konare, A., and Loiusse, C.: Impacts of megacities on air pollution and climate, World Meteorological Organization Report 205, 2012.

Zíková, N., Wang, Y., Yang, F., Li, X., Tian, M., and Hopke, P. K.: On the source contribution to Beijing PM2.5 concentrations, Atmospheric Environment, 134, 84-95, https://doi.org/10.1016/j.atmosenv.2016.03.047, 2016.

TABLE LEGENDS:

Table 1: A comparison of the annual average concentrations of air pollutants before and after weather ~~normaliz~~normalisation

FIGURE LEGENDS:

Figure 1: A diagram of long-term trend analysis model

Figure 2: Air quality and primary emissions trends

Figure 3: Yearly change of air quality in different area of Beijing

Figure 4: Relative change in monthly $PM_{2.5}$ levels in 2017 under different weather conditions

Figure 5: Comparison of MRF-CMAQ and RF models' performance

Figure 6: Primary energy consumption in Beijing

Table 1. A comparison of the annual average concentrations of air pollutants before and after weather ~~normaliz~~normalisation.

Pollutants	$PM_{2.5}$		PM_{10}		NO_2		SO_2		CO		O_3	
year	Obs.	Model	Obs.	Model	Obs.	Model	Obs.	Model	Obs.	Model	Obs.	Model
2013	88	93	110	123	54	58	23	26.3	1.4	1.5	58	59
2014	84	85	119	121	57	56	20	20	1.2	1.3	55	56
2015	80	75	107	106	50	50	13	13	1.3	1.2	58	59
2016	71	71	98	101	47	48	10	10	1.1	1.1	63	60
2017	58	61	90	93	45	48	7.5	8.4	0.9	1.0	60	61

Note: Obs: observed concentration. Model~~Nor~~.: Modelled c~~C~~oncentration of a pollutant after weather ~~normaliz~~normalisation. Unit: µg m^{-3} for all pollutants, except CO (mg m^{-3})

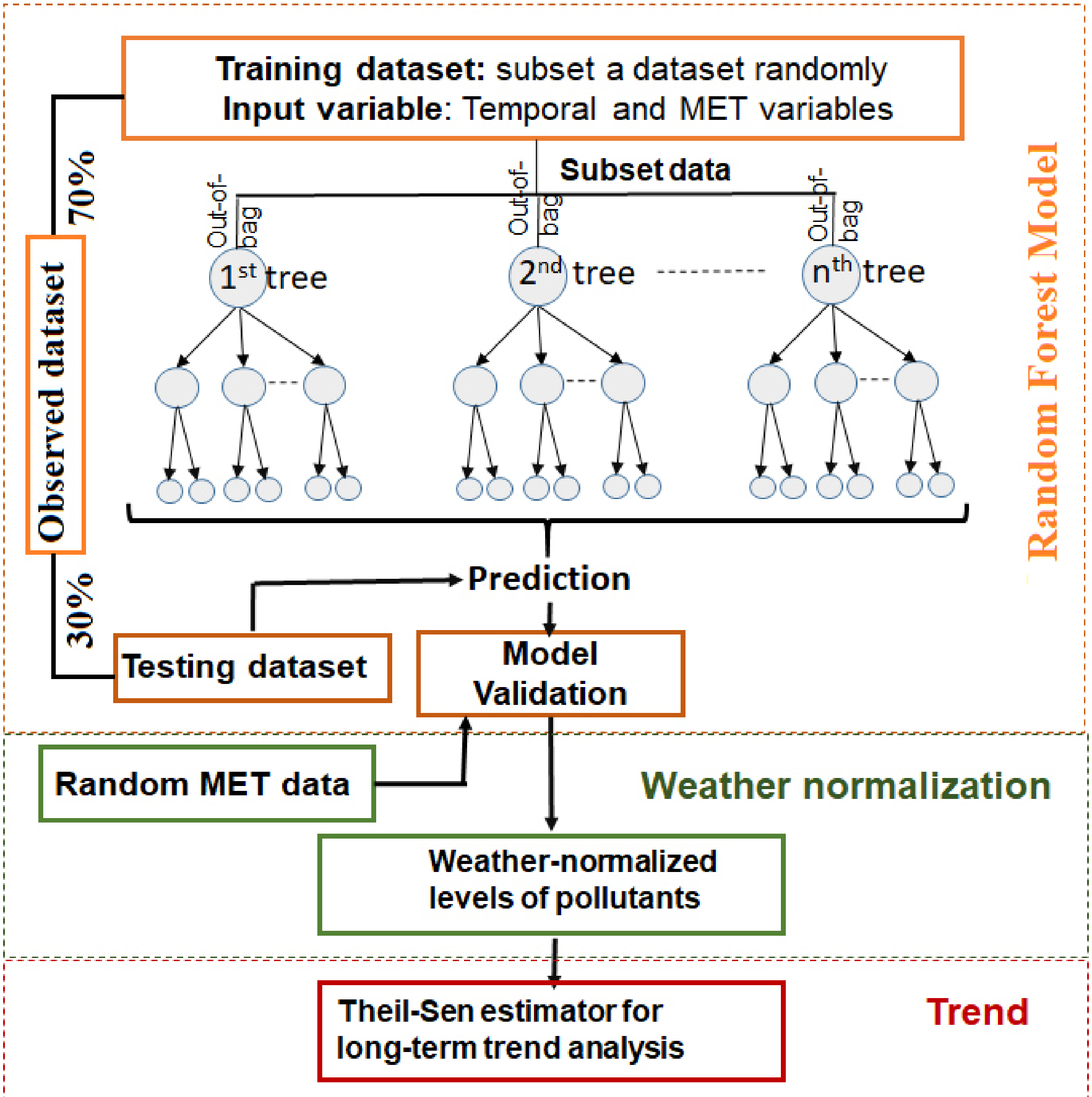
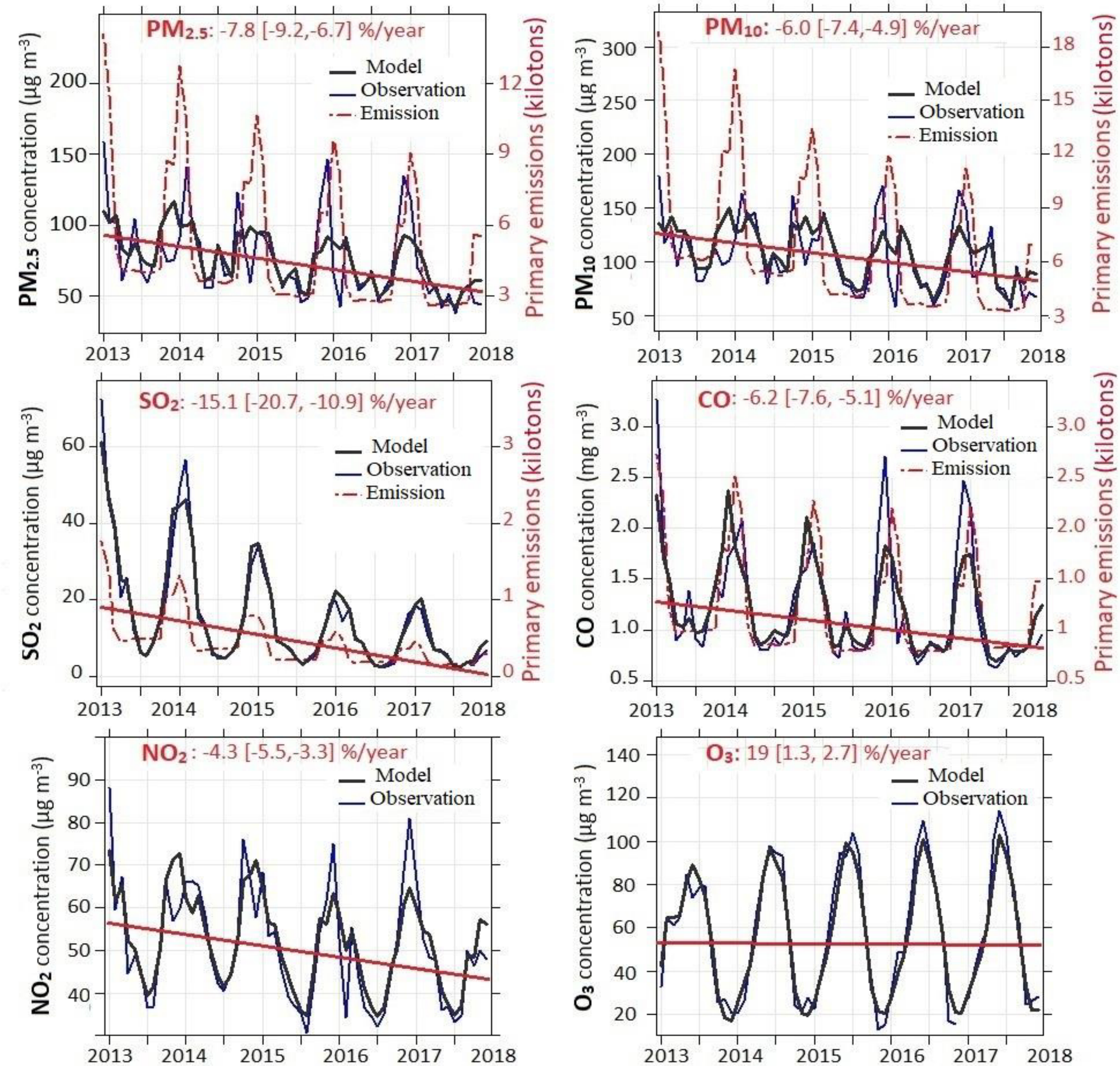


Figure 1: A diagram of long-term trend analysis model



PM2.5: -7.8 [-9.2,-6.7] %/year
Model
Observation
Emission
PM2.5 concentration (μg m-3)
Primary emissions (kilotons)
PM10: -6.0 [-7.4,-4.9] %/year
Model
Observation
Emission
PM10 concentration (μg m-3)
Primary emissions (kilotons)
SO2: -15.1 [-20.7, -10.9] %/year
Model
Observation
Emission
SO2 concentration (μg m-3)
Primary emissions (kilotons)
CO: -6.2 [-7.6, -5.1] %/year
Model
Observation
Emission
CO concentration (mg m-3)
Primary emissions (kilotons)
NO2: -4.3 [-5.5,-3.3] %/year
Model
Observation
NO2 concentration (μg m-3)
O3: 19 [1.3, 2.7] %/year
Model
Observation
O3 concentration (μg m-3)
2013
2014
2015
2016
2017
2018

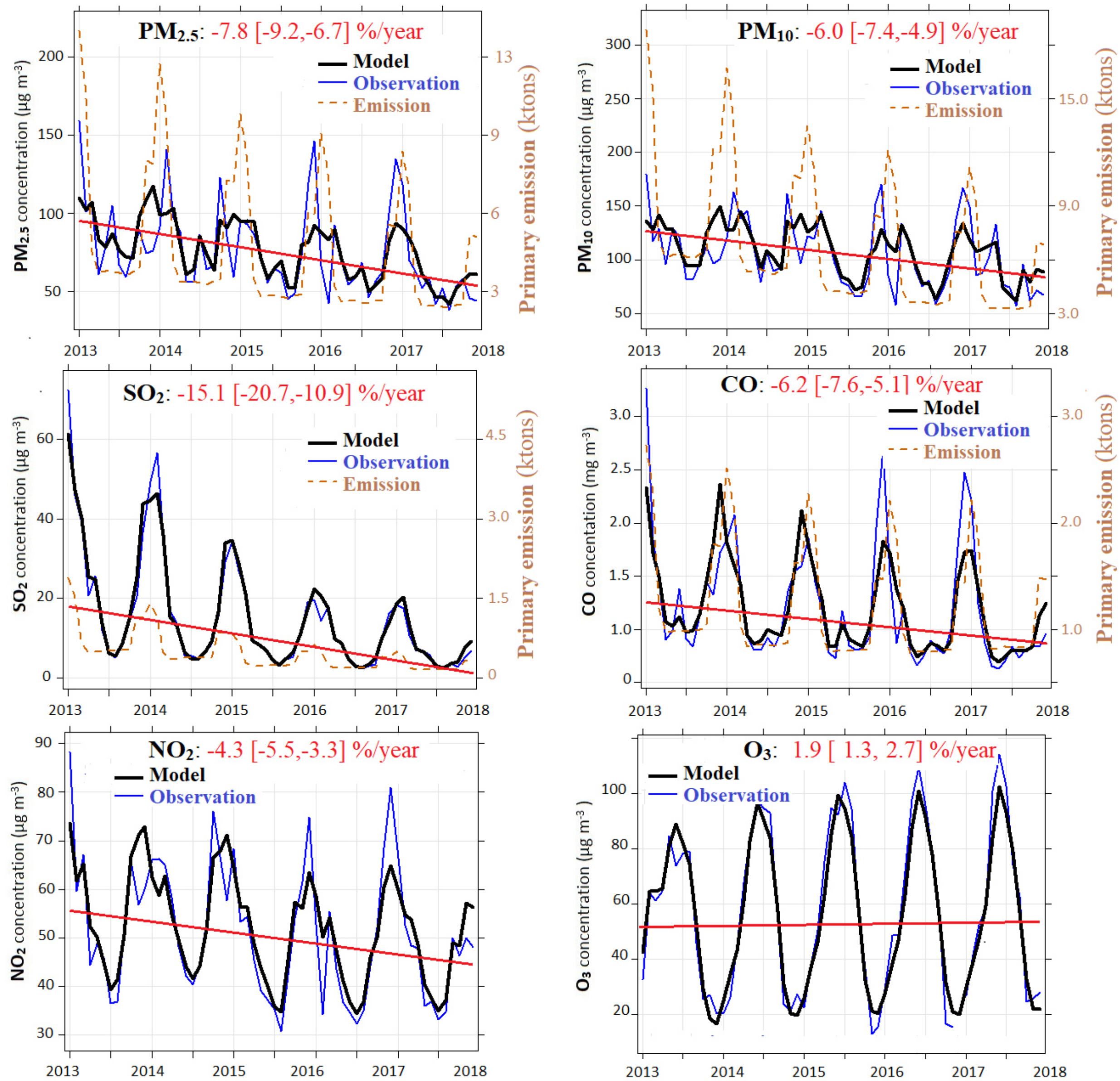


Figure 2. Air quality and primary emissions trends. Trends of monthly average air quality parameters before and after normalis~~z~~ation of weather conditions (first vertical axis), and the primary emissions from the MEIC inventory (secondary vertical axis). "Model" in the figure means the modelled concentration of a pollutant after weather normalisation~~zed~~. The red line shows the Theil-Sen trend after weather normalis~~z~~ation. The black and blue dot lines represent weather normalis~~z~~ed and ambient (observed) concentration of air pollutants. The red dot line represents total primary emissions. The levels of air pollutants after removing the weather's effects decreased significantly with median slopes of 7.2, 5.0, 3.5, 2.4, and 120 µg m^{-3} $year^{-1}$ for $PM_{2.5}$, PM_{10}, SO_2, NO_2, and CO, respectively, while the level of O_3 slightly increased by 1.5 µg m^{-3} $year^{-1}$.

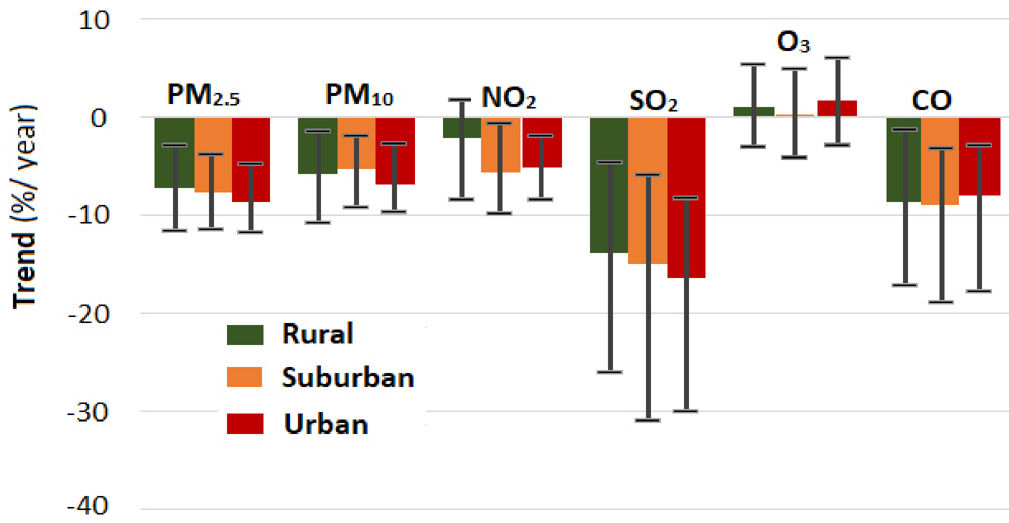


Figure 3. Yearly change of air quality in different area of Beijing. This figure presents yearly average changes of weather normalis~~z~~ed air pollutant concentrations at rural, suburban and urban sites (see Figure S1 for classification) of Beijing from 2013 to 2017. Specifically, average yearly changes are for SO_2 (-14%, -15%, -16 % year^{-1}- for rural, suburban, and urban areas, respectively), CO (-9%, -9%, -8% year^{-1}), $PM_{2.5}$ (-7%, -8%, -9% year^{-1}), PM_{10} (-6%, -5%, -7% year^{-1}), NO_2 (-2%, -6%, -5% year^{-1}) and O_3 (1%, 0.3%, 2% year^{-1}). The error on the bar shows the minimum and maximum yearly change.

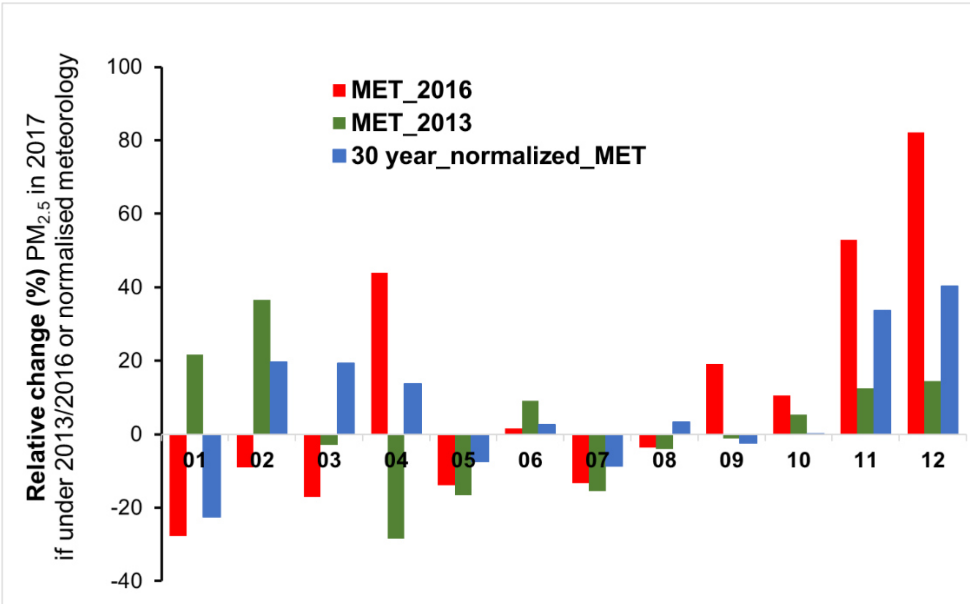


Figure 4. Relative change in monthly $PM_{2.5}$ levels in 2017 under different weather conditions. This figures presents relative changes (%) in monthly average modelled $PM_{2.5}$ concentrations in 2017 if under the 2016 (red) and 2013 (green) meteorological condition using CMAQ model and under averaged 30 years of meteorological condition using the machine learning technique**.** A positive value indicates $PM_{2.5}$ concentration would have been higher in 2017 if under the 2013 or 2016 meteorological conditions. Under the meteorological condition of 2016, monthly $PM_{2.5}$ concentration in 2017 would have been approximately 28% lower in January but 53% to 82% higher in November and December. This suggests that 2017 meteorological conditions were very favourable for better air quality comparing to those in 2016. If under the meteorological condition of 2013, monthly $PM_{2.5}$ concentration in 2017 would have been higher in January (22%) and February (36%) but only slightly higher in November (12%) and December (14%).

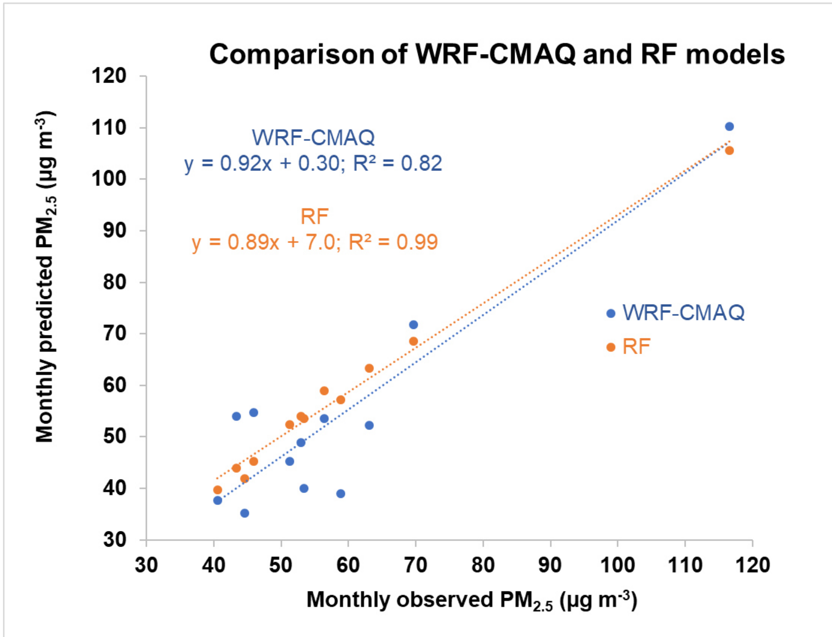


Figure 5. Comparison of predicted monthly average $PM_{2.5}$ mass concentrations by the ~~MRF~~WRF-CMAQ (Cheng et al., 2018) and RF model against observations in Beijing. WRF-CMAQ results are averaged over the whole Beijing region and the observed values refer to the average concentration of $PM_{2.5}$ over the 12 sites.

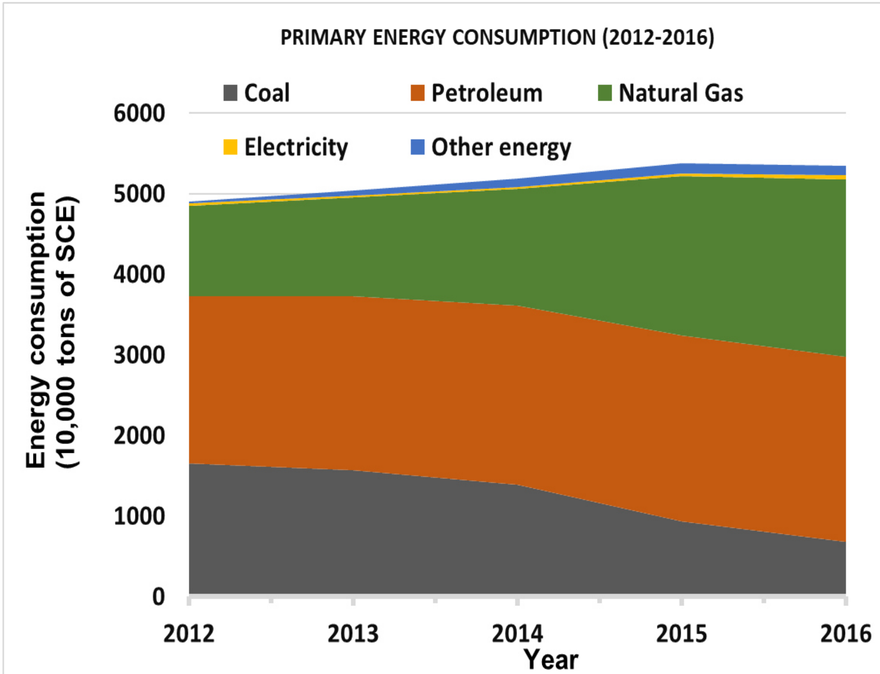


Figure 6. Primary energy consumption in Beijing**.** Petroleum consumption remained stable (21-23 million tonnes coal equivalent (Mtce)) over the years while natural gas and primary electric power increased significantly by 1.8 times and reached 23 Mtce in 2016. Coal consumption declined remarkably by 56.4% from 15.7 Mtce in 2013 to 6.8 Mtce in 2016. The proportion of coal in primary energy consumption in 2016 was 9.8 %, within its target of 10 % set by the Beijing government.

SUPPORTING INFORMATION

CLEAN AIR ACTION AND AIR QUALITY TRENDS IN BEIJING MEGACITY

T.V. Vu, J. Cheng, Z. Shi, Q. Zhang, K. He, S. Wang, R.M. Harrison

Number of pages : 11
Number of tables : 3~~1~~
Number of figures : 5

CONTENTS

Section S1. Data collection and overview of air quality

Hourly air quality data for six air pollutants was collected in Beijing from 17/01/2013 to 31/12/2017 across 12 national air quality monitoring stations which were classified in three categories (urban, suburban, and rural areas) based on hierarchical clustering (Figure S1, Table 1). Specifically, $PM_{2.5}$ levels at urban, suburban and rural sites decreased from 89.8, 78.3, and 67.8 µg m^{-3} in 2013 to 59.6, 54.6, and 47.8 µg m^{-3} in 2017, respectively. In 2017, 23 % of days still exceeded the NAAQS-II. A higher decrease in PM_{10} levels by 20.2 % was found at urban sites compared to those at suburban sites (17.2 %). PM_{10} also shows exceedances of NAAQS-II standards both for daily averages (150 µg m^{-3}) and annual averages (70 µg m^{-3}). It suggests that particulate matter, especially $PM_{2.5}$ is still a critical air pollutant in Beijing. In 2017, SO_2 does not show exceedance of the NAAQS-II standards either for daily averages (150 µg m^{-3}) and annual averages (60 µg m^{-3}). For CO, only 12 days do not meet NAAQS-II standards of 4 µg m^{-3}. In contrast, the annual average concentration of NO_2 in 2017 was slightly higher than the NAAQS-II standard of 40 µg m^{-3}, with 18 days exceeding the NAAQS-II standard for daily averages (80 µg m^{-3}).

Figure S1. Map of 12 monitoring stations

Section S2. Notices, regulation and policies for air pollution control in Beijing

Regulation and policies on energy system re-structuring:

- In October 2013, the government of Huairou district enforced a policy to replace anthracite stoves from 3000 rural households, change coal heating to electricity for 1170 households, supply liquefied petroleum to the countryside for 20,000 households, construct energy-saving residential housing and implement district heating; this reduced the consumption of 47,000 tons of poor quality coal.
- In Oct 2013, the government of Shijingshan, an urban district of Beijing, planned to cut 2800 tons of coal usage from coal-fired boilers in 2013, and reduce coal usage by more than 4500 tons in 2014, and eliminate coal-fired boilers in 2015.
- In November 2013, Miyun government issued an action plan to “Reduce coal for clean air” with a focus on urban transformation, conversion to natural gas, replacement with high quality coal, relocation of mountain communities, conservation of household energy, and removal of illegal constructions.
- In September 2014, the China State government released an important regulation on the “Reform and upgrade Action Plan for coal energy conservation and emission reduction (2014-2020)” that requires Beijing to place strict controls upon energy efficiency. Following that Action Plan, stack gas emissions of SO_2, NO_x, and PM from coal-fired power plants must be limited to below 10, 35, and 50 mg m^{-3} respectively.
- In March 2017, the Ministry of Environmental Protection issued the “2017 Air Pollution Prevention and Control Work Plan for Beijing-Tianjin-Hebei”. According to this plan, before the end of October 2017, Beijing, Tianjin, Langfang and Baoding City of Hebei will become the “no-coal zone”.

Regulations and policies on vehicle emission control: In order to control air pollution from vehicle emissions, during 2013-2017 the city announced a series of policies and regulations focusing on the implementation of stricter standards for new vehicles and vehicle fuels, elimination of yellow-label vehicles (which do not meet basic emission standards), and promotion of public transport. Consequently, Beijing led the nation in improving the fuel quality standards by adopting the desulfurization of gasoline and diesel fuels (sulfur content <10 ppm) in 2012, three years ahead of the surrounding regions (Tianijin and Hebei) and five years before the national deadline. Major policies for air pollution from transportation management:

- In February 2013, Beijing implemented the fifth phase emission standards for new light-duty gasoline vehicles (LDVs) and heavy-duty diesel vehicles (HDVs) for public transport.
- In June 2013, another notice from the Beijing government emphasized that all heavy-duty vehicles sold and registered in Beijing must meet the national fourth-phase emission standards
- In August 2014, a notice from Beijing's government declared that all spark ignition light vehicles must meet the national five phase standard from 1st January 2015.
- In 2014, Beijing Municipal Commission of Transport (BMCT) expanded traffic restrictions to certain vehicles, particularly yellow-label and non-local vehicles to enter the city within the sixth ring road during daytime since 2015.
- In November 2014, the governments of Yanquing and Miyun, two rural districts of Beijing, released regulations to prohibit yellow-label gasoline vehicles entering certain roads.
- In February 2015, the Beijing Municipal government issued a notice to promote elimination and replacement of old motor vehicles with an expectation of 1 million old vehicles/year phased out.
- Other policies which may have contributed to the enhancement of air quality during 2013-2017 included a ban of outdoor biomass burning and improved suppression of dust discharges from construction sites.

Section S3. Model performance and explanation

Variables and hyperparameters: The input variables contain time and MET variables.

Time variables: day_unix (or t_{trend}) represents the emission trend of a pollutant; Julian_day (t_{JD}: the day of the years) represents for the seasonal variation; weekday/weekend represents the difference of pollution between the week and weekend days.

MET variables: wind speed (m s^{-1}), wind direction (°), temperature (°C), relative humidity (%), and atmospheric pressure (mbar). The back-trajectories can be used as a predictor feature, but it does not increase the performance of the model in this case.

Selected parameters in a random forest:

- Mtry=4: variables randomly sampled for splitting the decision tree
- Nodesize=3: minimum size of terminal nodes for model
- Ntree=200, the number of trees to grow. Figure S2~~1~~ shows the dependence of model performance on the number of trees.

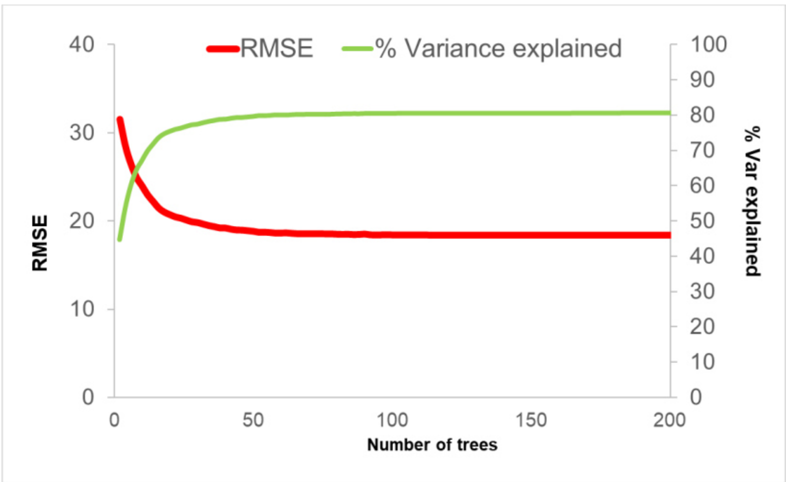


Figure S2~~1~~. The influence of number of trees on the model performance for $PM_{2.5}$.

Model performance's evaluation~~:~~

A random forest shows a good performance with the correlation (r^2) between hourly predicted and observed data for both training and testing data sets. In particular, r^2 value ranged 0.81-0.83, 0.75-0.79, 0.80-0.83, 0.88-0.90, 0.85-0.87, and 0.89-0.90 for $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO and O_3, respectively. Figure S3~~2~~ shows the hourly correlation between observed and predicted data for a testing data. Other model evaluation metrics are shown in Table S2.

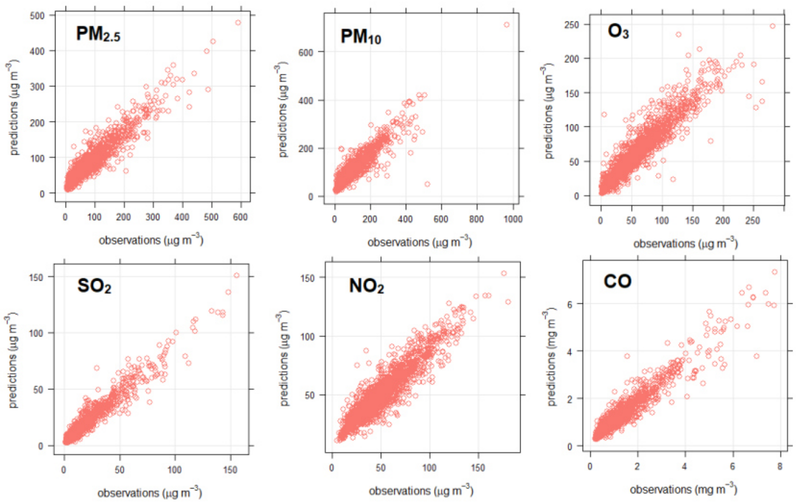


Figure S3~~2~~. Correlations between daily observed and predicted data from testing data sets

As shown in Figure S3~~2~~, it is likely that the model underestimates hourly concentration of air pollutants at th~~os~~e extremely high levels. These errors are reduced when we compare the weekly averaged concentration as shown in Figure S4~~3~~.

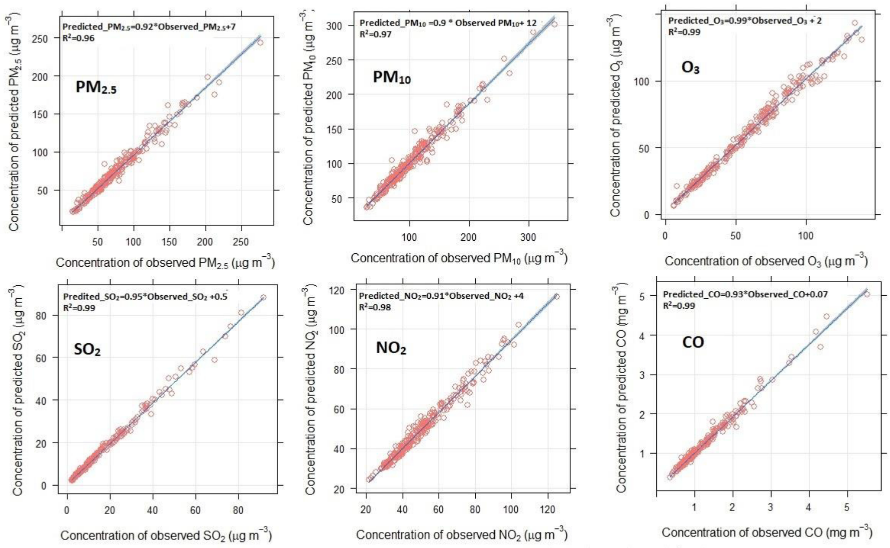
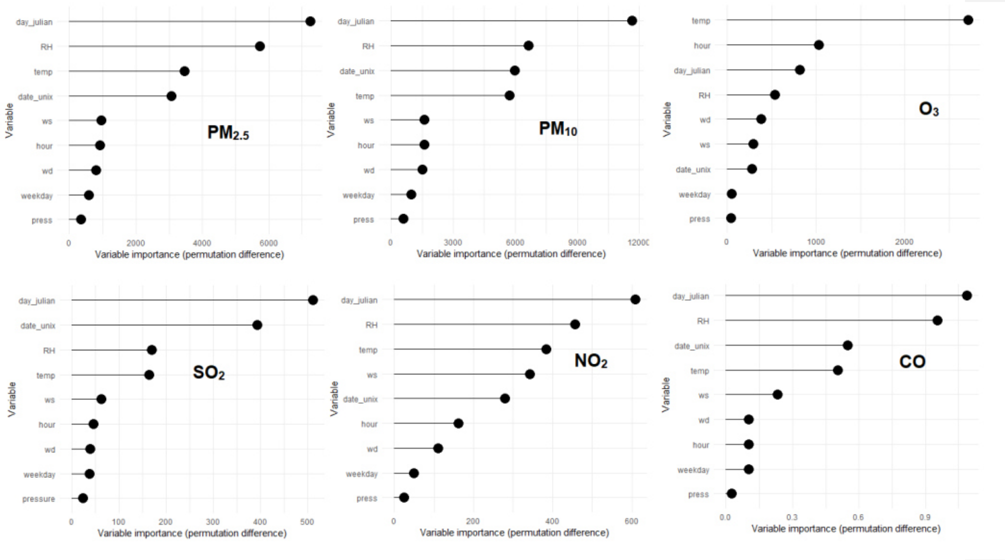


Figure S4~~3~~. The correlation between observed and modelled concentrations is approximately 0.9-0.99 for weekly averaged data. In our study, a RF forest model was trained using a fraction of 0.7 from the datasets.

Variable importance and interactions:

As shown in Figure S44, seasonal variations (day_julian) play the most important variable in the model, except for ozone when temperature and diurnal pattern (hour) mainly control the predicted values. The trend (day_unix) shows more important role in the model of SO_2 and CO, indicating emission control shows most effectiveness on the decrease of SO_2 and CO. Regarding MET variables, humidity and temperature play a more important role in the model of PM while wind speed has a larger impact in the model of NO_2. The variable interaction is shown in Figure S5.



Figuer S4. Importance of predictor features: date_unix, day of the year (day_julian), hour of day (hour), week/weekend, temperature (temp), RH, pressure (press), wind speed (ws), wind direction (wd) in the random forest model. Figure 4 shows the day of the year (seasonal variable) is the most important variable~~s~~ controlling the concentration of the pollutant (except for ozone: the most important is the temperature variable). The trend (date_unix) has a larger effect on SO_2, tha~~e~~n CO and PM, less effect on the NO2 and no significant effect on O3 concentration.

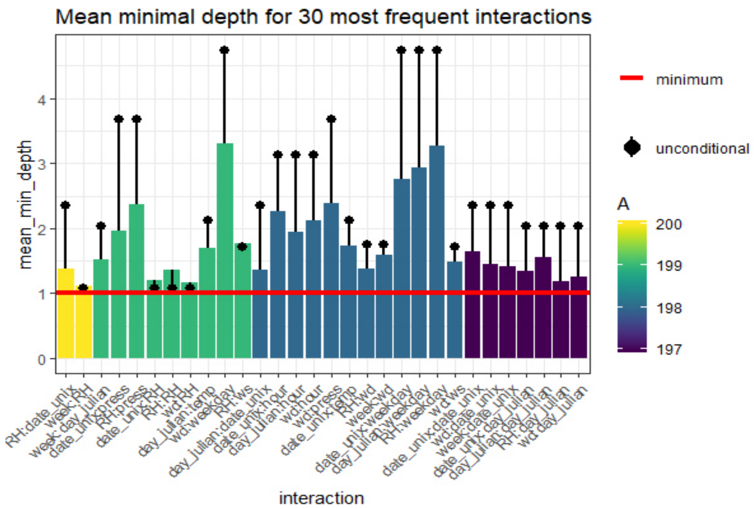


Figure S6~~5~~. Features~~Variation~~ interactions in a random forest model for $PM_{2.5}$. This figure shows the co-occurrence of a pair of variables in a similar tree. For example, in the first node of the tree, RH and date_unix is the most frequent occurrence.

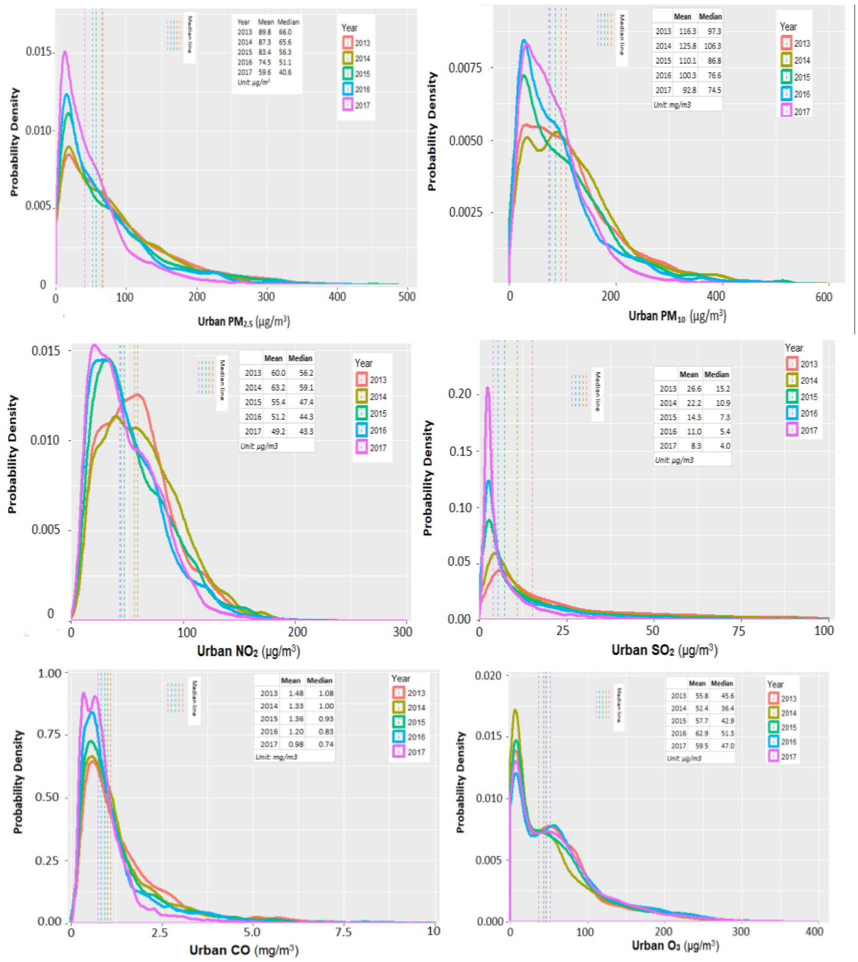


Figure S7~~6~~. Probability density of urban air pollutant concentrations during 2013-2017. Number of heavy polluted events decreases from 2013 to 2017 for all pollutants, except ozone.

Formatted: Font: Times New Roman

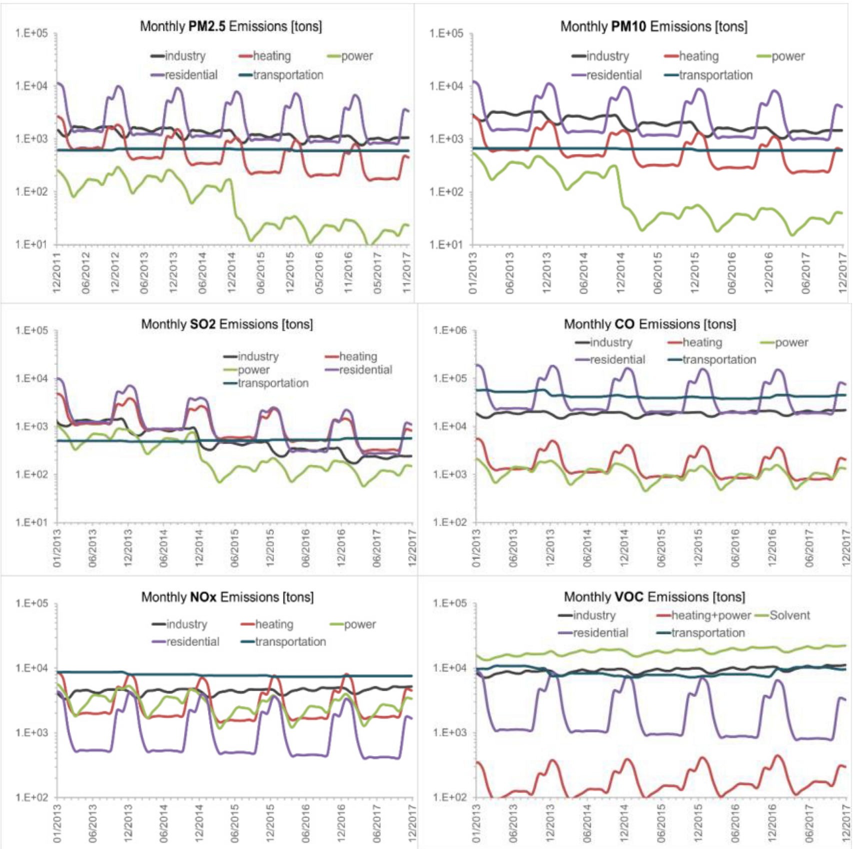


Figure S8~~7~~. Monthly emission inventories of air pollutants in Beijing during 2013-2017. The emissions of $PM_{2.5}$, PM_{10}, NO_x, SO_2, CO in Beijing dropped by 35 %, 44 %, 11 %, 71 %, 17% from 76, 109, 260, 93, 1.7 Gg in 2013 to 49, 61, 231, 27, 1.4 Gg in 2017, respectively. Power sector represents the coal-fired, gas-fired and oil-fired power plants; industry sector includes two subsectors as industrial process and industrial boilers (to offer the mechanical energy); heating includes both industrial heating (to offer the thermal energy) and domestic heating (refers to centralized heating); residential sources are the urban and rural burning with traditional stoves with coal or biomass fuels; transportation includes both on-road and off-road traffic; solvent use contains all the subsectors which would use solvent during production processes, such as paint, ink, pharmaceutical production and household solvent use.

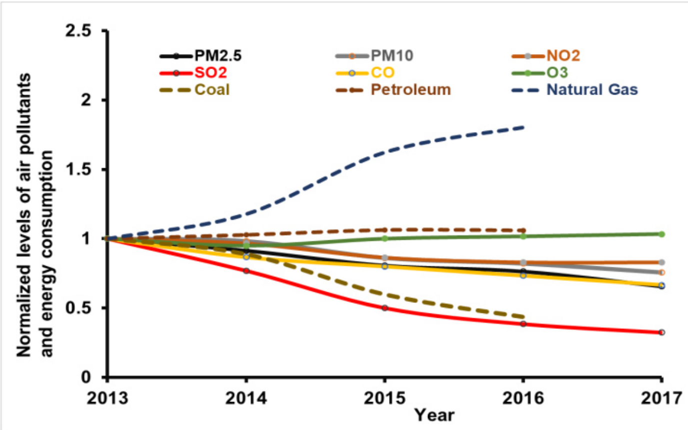


Figure S98. Normalized levels of air pollutants and energy consumption. The trend of SO_2 was very close to the normalized trend of coal consumption, but showed a faster decrease than trends of $PM_{2.5}$ and NO_2.

Table S1. Locations and cate~~o~~gories of monitoring site

Station ID	Name	Category	Longtitude	Latitude
01	Wangshouxigong	Urban	116.37	39.87
02	Dingling	Rural	116.17	40.29
03	Dongsi	Urban	116.43	39.95
04	Tiantan	Urban	116.43	39.87
05	Nongzhanguan	Urban	116.47	39.97
06	Guanyuan	Urban	116.36	39.94
07	Haidianquwanliu	Urban	116.32	39.99
08	Shunyixincheng	Urban	116.72	40.14
09	Huairouzhen	Suburban	116.64	40.40
10	Changpingzhen	Suburban	116.23	40.20
11	Aotizhongxin	Urban	116.40	39.98
12	Gucheng	Suburban	116.26	39.93

Table S2~~2~~: RF model performance for testing data set (in hourly time resolution).

Pollutants	RMSE	r2	FAC2	MB	MGE	NMB	NMGE	COE	IOA
PM2.5	17.9	0.95	0.94	0.62	10.00	0.01	0.14	0.81	0.91
PM10	43.1	0.79	0.87	1.46	27.10	0.01	0.26	0.57	0.79
NO2	14.3	0.78	0.95	-0.01	10.16	0.00	0.20	0.59	0.79
SO2	7.0	0.89	0.89	0.22	3.70	0.02	0.25	0.73	0.87
CO	0.4	0.86	0.96	0.01	0.24	0.01	0.21	0.67	0.84
O3	18.4	0.89	0.82	0.50	12.90	0.01	0.21	0.70	0.85

Note:- FAC2 (fraction of predictions with a factor of two), MB (mean bias), MGE (mean gross error), NMB (normalised mean bias), NMGE (normalised mean gross error), COE (Coefficient of Efficiency), IOA (Index of Agreement) (Emery et al. 2017).

Table S3. Air Quality Standards. China's Air Quality Standards: GB 3095-2012, phase-in 2012-2016; WHO Air Quality Guidelines (2005). The Class 2 standards apply to urban areas.

Pollutants	Averaging time	China standards		WHO	unit
		Class 1	Class 2		
$PM_{2.5}$	annual	15	35	10	μg m^{-3}
	24 hours	35	75	25	μg m^{-3}
PM_{10}	annual	40	70	20	μg m^{-3}
	24 hours	50	150	50	μg m^{-3}
NO_2	annual	40	40	40	μg m^{-3}
	24 hours	80	80	-	μg m^{-3}
	hourly	200	200	200	μg m^{-3}
SO_2	annual	20	60	-	μg m^{-3}
	24 hours	50	150	20	μg m^{-3}
	hourly	150	500	-	μg m^{-3}
	10 min	-	-	500	μg m^{-3}
CO	annual	4	4	-	mg m^{-3}
	24 hours	10	10	-	mg m^{-3}
O_3	8-hour mean, daily max	100	160	100	μg m^{-3}
	hour	160	200	-	μg m^{-3}