Referee 2

The manuscript by Ren et al. presents a machine learning (ML) model based on Random Forest Regression (RFRM) to predict cloud condensation nuclei number concentration (N_{CCN}) at typical supersaturations in the North China Plain (NCP), which they then demonstrate that their model reduces the prediction bias from 39% (WRF-Chem) to 8%. The study also analyzes the importance of different input factors, evaluates the model's performance in simulating spatiotemporal variability, and quantifies the reduction in cloud radiative forcing uncertainty achieved by mitigating N_{CCN} simulation biases. While the topic is highly relevant to GMD and represents a potential advancement in climate modeling, I have significant concerns regarding the methodological approach and validation logic, as well as the clarity of the manuscript. I recommend returning the manuscript to the authors with encouragement to revise and resubmit after addressing the following concerns.

Major Concerns

The training and testing datasets use WRF-Chem simulations as the output variable (target), but model performance is evaluated against observations. From a machine learning perspective, if the target variable (output) in both training and testing is WRF-Chem simulations, the model's objective should be to approximate WRF-Chem's output—not observations. Thus, model performance should be assessed based on how well it replicates WRF-Chem's results, not observations. The authors report that WRF-Chem exhibits a significant bias (~39%) compared to observations, yet their ML model, trained to emulate WRF-Chem, shows a much smaller bias (~8%). This discrepancy is counterintuitive and requires a thorough explanation, which is currently missing.

Re: The reviewer put forward an insightful comment. However, the reviewer may misunderstand the part where we compared the CCN predicted by the RFRM model with both the observed CCN and that simulated by WRF-Chem. This could be attributed to that we do not elaborate on this clearly in the methodology section. Due to the scarcity of a large spatial scale CCN observations (regional and global), in this study, the RFRM model was trained using simulated CCN concentrations from WRF-Chem as the target variable, which can capture general temporal variations in ambient CCN concentrations (Fig. 4) despite some biases.

When constructing the model, all the aforementioned species data (predictor and target variables) were processed to match the same spatiotemporal resolution, and then split into 7:3 ratio for model training and testing respectively. Fig. 2 shows the performance of our developed RFRM models by comparing the testing dataset (sample size: N=117585) of CCN simulated by WRF-Chem with the predicted CCN. It showed that the estimated N_{CCN} at S=0.2% are highly correlated with the values from WRF-Chem, with the correlation R^2 of ~0.89-0.91 and slopes of 0.83 and 0.86 in our developed RFRM model. This suggests that our model works well in estimating N_{CCN} with a high aerosol loading environment

However, given that the ultimate goal of model development is to more accurately predict the actual atmospheric concentration of CCN, we further conducted a comparative analysis of the prediction results against the in-situ observation data in this

region, spanning from the hourly scale to the interannual scale (see Figure 3-6). For this comparison, we also plotted and included the WRF-Chem simulated CCN in the figures. As a result, we can get how the developed model has improved the WRF-Chem simulations.

We have added detailed descriptions in the revised text or see follows (**Lines 117-173**):

"… **2.2 Model construction and validation**

Here we develop the ML-based N_{CCN} prediction model by employing the Random Forest Regression method (RFRM) that has been demonstrated and can solve multivariate and nonlinear regression problems (Nair and Yu, 2020; Liang et al., 2022). The diagram of the model construction and the N_{CCN} prediction is shown in Figure 1. Due to lack of a large spatial scale observed N_{CCN} data, we use simulated N_{CCN} by WRF-Chem model as the targeted variable that can basically capture the ambient temporal variability of CCN concentration despite a deviation of ~40% by comparing with our six field observations (Figure 4). The input parameters include the chemical components of $PM_{2.5}$ (organic, sulfate, nitrate, ammonium, black carbon) from the Tsinghua University Tracking Air Pollution in China dataset (Liu et al., 2022) and gas and particulate pollutants (nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3) and $PM_{2.5}$) collected from the China National Environmental Monitoring Centre network. Meteorological parameters are from the European Centre for Medium-range Weather Forecasts Reanalysis version 5 (ERA-5) and include temperature, relative humidity (RH), total precipitation (TP), wind speed (WS), wind direction (WD), planetary boundary layer height (BLH), surface pressure (SP) and surface net solar radiation (SNSR). These datasets have undergone validation and been proven highly suitable for developing machine learning models for atmospheric applications (Nair and Yu, 2020; Wei et al., 2023). Cartesian coordinates were also added as input due to the spatiotemporal nature of the input data (Yang et al., 2022). Supplemental Table 1 provides more details about the input parameters.

When constructing the model, all the aforementioned species data (predictor and target variables) were processed to match the same spatiotemporal resolution, and then split into 7:3 ratio for model training and testing respectively. As a result, a total of 274365 samples is included in the training datasets. In order to assure a stronger generalization ability of the N_{CCN} prediction model, the 10-fold cross-validation is adopted (Wei et al., 2023). The optimization parameters of RF model were examined by varying hyperparameters (Fig. S1). In addition, cross-validation (CV) is applied to select the hyperparameters during the data preprocessing (Yang et al., 2022). The CV results showed that when the number of trees (n_estimators) was less than 200, the prediction accuracy increased rapidly with the increase of the number of trees, and then gradually stabilized. According to the CV score and the number of data sample, the number of trees was set to 500 in this study. The impact of max depth on the CV score showed that, with the increase of depth, the complexity of the model increases. Thus, the max depth is set to 28. Also, the model generalization error was larger when the minimum sample number of the leaf and branch node are large, indicating that the model itself is close to the optimal model complexity level. Therefore, a higher value was set given the large sample size in this case. The influence of the maximum selection

feature number on CV score showed a trend of increasing first and then decreasing, so the maximum value of CV curve was set to 16.

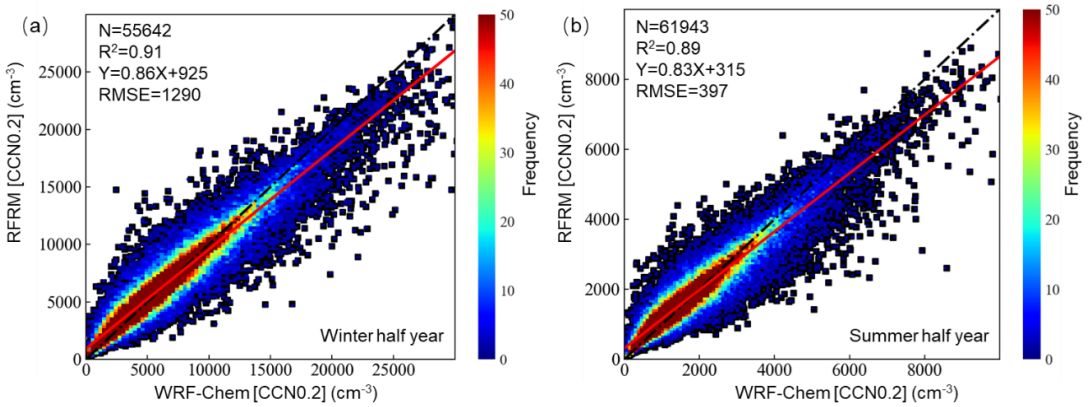

Figure 2. Comparison of RFRM retrieval and WRF-Chem simulated N_{CCN} at S=0.2%. (a and b) Density plots of retrieval N_{CCN} at S=0.2% as a function of the simulations from WRF-Chem on the testing dataset.

Fig. 2 shows the performance of our developed RFRM models by comparing the testing dataset (sample size: N=117585) of CCN simulated by WRF-Chem with the predicted CCN. Here the quality metrics for model performance are based on the correlation coefficient (R^2), root mean square error (RMSE) and the slope of the RFRM predicted and the WRF-Chem simulated CCN concentrations. It showed that the estimated N_{CCN} at S=0.2% are highly correlated with the values from WRF-Chem, with the correlation R^2 of ~0.89-0.91 and slopes of 0.83 and 0.86 in our developed RFRM model. This suggests that our model works well in estimating N_{CCN} with a high aerosol loading environment. We also found that the accuracy of the CCN prediction will deteriorate slightly if not including the information of chemical compositions (Fig. S2), or if using XGBoost algorithm (Fig. S3) when constructing the model.

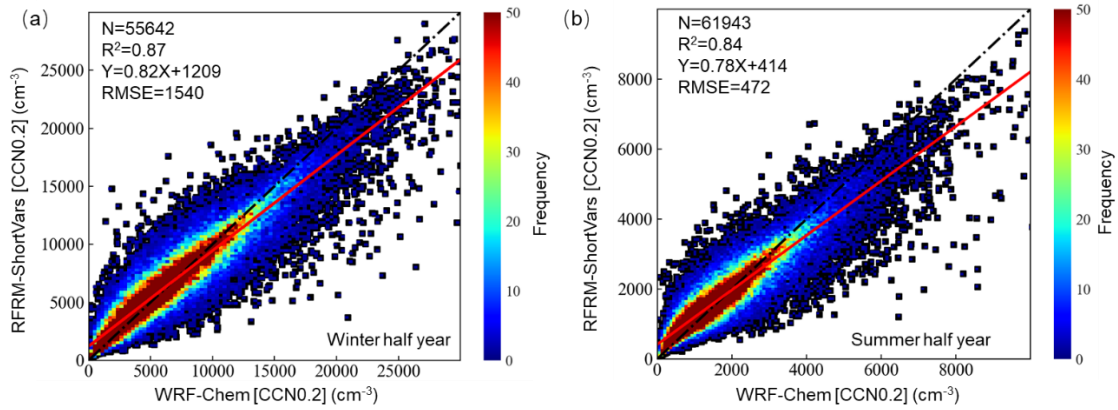

Figure S2. Comparison of RFRM-ShortVars model retrieval and WRF-Chem simulated N_{CCN} at S=0.2%. (a and b) Density plots of retrieval N_{CCN} at S=0.2% as a function of the simulations from WRF-Chem on the testing dataset.

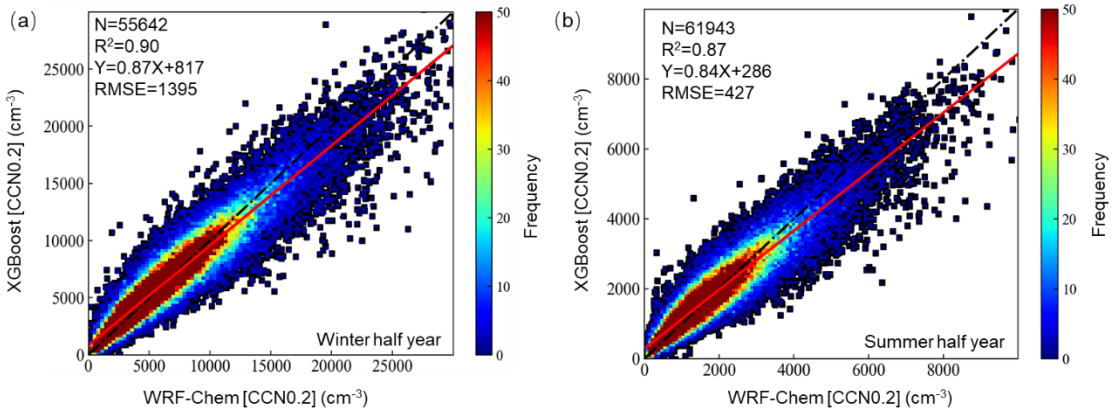

Figure S3. Same as Figure S2 but from XGBoost model.

Given that the ultimate goal of model development is to more accurately predict the actual atmospheric concentration of CCN, we further conducted a comparative analysis of the prediction results against the in-situ observation data in this region, spanning from the hourly scale to the interannual scale (see Figure 3-6), which will be presented and discussed in detail in Section 3.2 and 3.3 …"

Use multisource datasets as input and WRF-Chem simulations as output to make training and test sets. Table S1 indicates that all input variables used in the ML model are also available from WRF-Chem outputs. However, the author uses datasets from different sources and different spatiotemporal resolutions to construct the data set through interpolation. The authors provide no rationale and advantage of this approach. Conventionally, one would train ML models using consistent model outputs and later test with observational inputs to assess potential gains.

Re: Although the input feature variables used in model training can indeed be derived from the output of WRF Chem. However, they exhibit varying degrees of bias. Assuming the non-linear relationship between predictor features and target variables is accurate, train the RFRM model. So, choosing more accurate feature factors to train machine learning models is crucial for improving model accuracy. It is also consistent with the standard practice of environmental data modeling. In machine learning modeling, we used $PM_{2.5}$ chemical composition from the Tsinghua University Tracking China Air Pollution Dataset (Liu et al., 2022), gas and particulate pollutants from the China National Environmental Monitoring Center network, and meteorological parameters from the European Centre for Medium Range Weather Forecast Reanalysis 5th edition (ERA-5) data as input variables. These are commonly used datasets for building atmospheric application machine learning models (Nair and Yu, 2020; Wei et al., 2023). The results indicate that the model trained on carefully selected predictive factors can reliably predict CCN. We have revised as follows and see **Lines 125-138**:

"…The input parameters include the chemical components of $PM_{2.5}$ (organic, sulfate, nitrate, ammonium, black carbon) from the Tsinghua University Tracking Air Pollution in China dataset (Liu et al., 2022) and gas and particulate pollutants (nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3) and $PM_{2.5}$) collected from the China National Environmental Monitoring Centre network. Meteorological parameters are from the European Centre for Medium-range Weather Forecasts Reanalysis version 5 (ERA-5) and include temperature, relative humidity (RH), total

precipitation (TP), wind speed (WS), wind direction (WD), planetary boundary layer height (BLH), surface pressure (SP) and surface net solar radiation (SNSR). These datasets have undergone validation and been proven highly suitable for developing machine learning models for atmospheric applications (Nair and Yu, 2020; Wei et al., 2023). Cartesian coordinates were also added as input due to the spatiotemporal nature of the input data (Yang et al., 2022). Supplemental Table 1 provides more details about the input parameters …"

The model is trained on data from 378 monitoring sites across the NCP (2014–2018) but evaluated using three sites (Beijing, Xingtai, Gucheng) within the same region and time period. This does not test generalizability. True validation requires spatiotemporal extrapolation: e.g., evaluation outside the training period or region. Performance on training-era/training-region data is expected to be favorable and does not demonstrate robust predictive skill.

Re: Thank you. In this study, the RFRM model is trained on data from 378 sites across in NCP from 2014 to 2018. The six observations were excluded from the set of 378 monitoring sites across NCP, and the corresponding time periods were eliminated in the model training process. Therefore, the evaluation is conducted outside of the training period or region.

Besides, observations at these sites could represent the average polluted and background conditions in the NCP (eg., $PM_{2.5}$ shown in Fig. 4b). Compared with the observed values, the better performance of the model can indicate its good predictive ability. See **Lines 188-225:**

"…**2.3.2 Gound-based measurements and datasets**

Ground measurements of atmospheric gaseous precursors, fine particles chemical compositions, and CCN number concentration (at supersaturations of 0.2% and 0.4%) were collected during six field campaigns at three sites in the NCP (Fig. 4), used to assess the performance of the developed ML-based model in predicting N_{CCN}. The six campaigns were conducted as follows: at the Beijing (BJ) site from 8–30 November 2014, 20 August to 6 October 2015, 16 November to 20 December 2016, and 28 May to 27 June 2017; at the Xingtai (XT) site from 17 May to 14 June 2016; and at the Gucheng (GC) site from 23 January to 3 February 2018. They are accordingly named BJ2014_WIN, BJ2015_AUT, BJ2016_WIN, BJ2017_SUM, XT2016_SUM, and GC2018_WIN (Fig. 4a).

The BJ site (Longitude: 116.37° E; Latitude: 39.97° N) is located at the meteorological tower station of the Institute of Atmospheric Physics, Chinese Academy of Sciences. It is representative of the general emission conditions in urban areas of the northern NCP. The primary pollution sources here are surrounding traffic and residential emissions. The XT site (Longitude: 114.37° E; Latitude: 37.18° N) is situated at a national weather station. It is primarily influenced by emissions from surrounding towns and factories (e.g., coal-fired power plants, coking, steel, cement, and chemical industries) and thus reflects polluted suburban conditions in the southern NCP. The GC site (Longitude: 115.74° E; Latitude: 39.15° N) is located at the Integrated Ecological-Meteorological Observation and Experiment Station of the Chinese Academy of Meteorological Sciences. Surrounded mainly by nearby villages,

farmland, and transportation networks, this site represents the regional background pollution in the northern NCP.

The CCN number concentrations were measured by using the Droplet Measurement Technologies CCN counter (model CCNC-100, DMT Inc. Lance et al., 2006) at BJ and XT site. The supersaturation (S) levels set for each CCN measurement cycle were 0.1%, 0.2%, 0.4%, and 0.8%, respectively. Another measurement at GC site was referred from Zhang et al. (2020). In this study, the comparisons between the measured and predicted N_{CCN} were mostly based on the value at S=0.2% and S=0.4%. The observed N_{CCN} varies from a few hundred to tens of thousands at these sites, and the campaign mean mass concentration of $PM_{2.5}$ ranges from 35.6 to 160 μg m^{-3} (Fig. 4b), indicating that the observations can represent various atmospheric conditions, spanning from clean to polluted in the region. More details about the observations could be found in Fan et al. (2020), Ren et al. (2018), and Zhang et al. (2019). In addition, the long-term measurement of particle number size distribution (PNSD) at a field site in Beijing (Fig. S5, Shang et al., 2022) is also used for deriving the long-term trend of yearly averaged N_{CCN} …"

Other Major Concerns:

The manuscript positions itself as unique by focusing on polluted regions (Lines 88–95), yet only 6 field campaigns are used for evaluation, with no dedicated analysis of heavy-pollution events. Additionally, heavy reliance on "mean prediction bias" is misleading: if RMSE/MSE is the training loss (unstated in the text), ML models inherently bias predictions toward the mean. Therefore, the improvement of the "mean prediction bias" cannot fully prove the performance of the ML model in real scenarios, and it is more meaningful to conduct a detailed evaluation and analysis of a single severe pollution event.

Re: Thanks for your suggestion, we have added some discussion about haze events in the revised text, see as follows or **Lines 290-311:**

"…Compared to WRF-Chem simulations, the RFRM model showed the greatest improvement during the winter campaigns when $PM_{2.5}$ concentrations were usually higher. For example, during the GC2018_WIN campaign, the observed N_{CCN} is underestimated as large as 61% by the WRF-Chem (Fig. S8), while the underestimation is largely improved with the predicted bias of only 3% in the RFRM model (Fig. S8). WRF-Chem simulations for warm seasons noticeably improved, e.g., the uncertainty decreased to 8% during the BJ2015_AUT campaign (Fig. S8). Overall, the RFRM model still performs better than the WRF-Chem model and is with averaged predicted bias of 18% during summer campaigns. Occasionally, the WRF-Chem model overestimated the N_{CCN} apparently, e.g., the episodes of September 21 to 24 during the BJ2015_AUT campaign, and May 28 to 31 during the BJ2017_SUM campaign. Here four pollution events from different seasons have been selected to further examine the capability of RFRM model to predict CCN concentrations (Figure 5). Fig. 5a presents a case from 14th to 18th September, 2015, during which $PM_{2.5}$ levels increased from 50 to 315 μg m^{-3}. As pollution intensified, CCN concentrations also rose. Compared to observations, the RFRM model exhibited lower relative bias. Fig. 5b–d display three additional individual pollution episodes of varying severity with $PM_{2.5}$ ranging from 10

to 660 μg m^{-3}. In all cases, the RFRM model more accurately captures the peak CCN concentrations during pollution events, exhibiting consistently lower relative bias. Especially for the case of 2nd to 5th December in 2016, the RFRM model can better capture the peak N_{CCN} of high pollution, while the WRF-Chem did not simulate the peak on December 4th very well …"

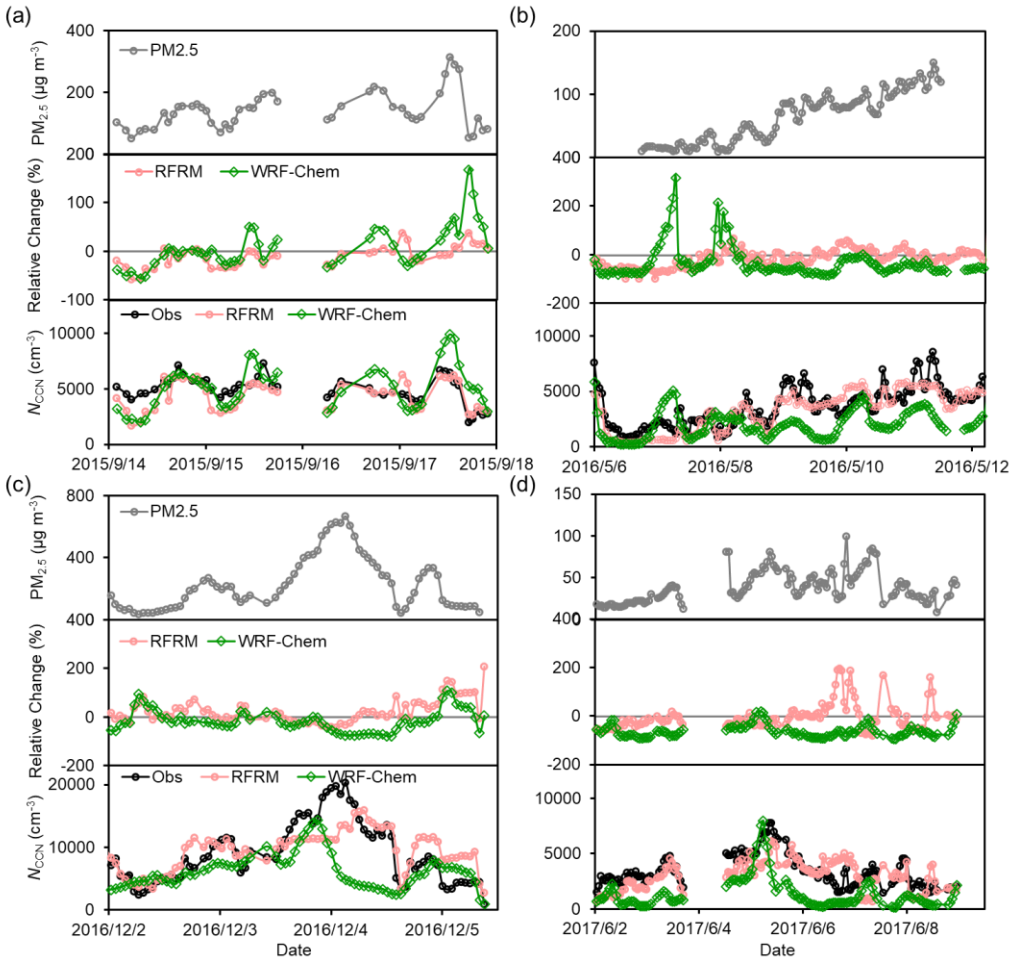

Fig. 5 Performance of the RFRM model in predicting N_{CCN} during haze events. (a) Case of 24 to 18 September in 2015, (b) case of 6 to 18 May in 2016, (c) case of 2 to 5 December in 2016, (d) case of 2 to 8 June in 2017.

A common but concerning trend in ML applications is showcasing successes while neglecting failures. This paper follows that pattern. There is no discussion of scenarios where the model underperforms, its limitations, or potential pitfalls. For instance, an eager graduate student might misuse this model for policy analysis without realizing its constraints (e.g., lack of generalizability), leading to significant wasted effort. A rigorous journal paper must present a balanced view of model capabilities and weaknesses.

Re: Thanks for the suggestion, some discussions about the model limitations were revised in the section of **4.2 Limitations and outlook** or see **Lines 481-500:**

"**4.2 Limitations and outlook**

In this study, the RFRM model was trained using simulated CCN concentrations

from WRF-Chem as the target variable, assuming that the nonlinear relationships between the predictor features and the target variable are accurate. However, as noted earlier, even though WRF-Chem simulations can capture the variation of N_{CCN}, they carry an uncertainty of ~20–40% compared to observations (Fanourgakis et al., 2019). This contributes directly to uncertainty in the RFRM model's predictions. Additionally, note that in this study, observational data from six campaigns at three sites are analyzed. Validating the simulated N_{CCN} through comparisons with observations at more ground sites is thus warranted. In the future, it is crucial to obtain comprehensive monitoring data of CCN and other key aerosol properties (e.g., particle size distribution, chemical compositions) in different environments.

The RFRM framework presented here relies on readily available atmospheric state variables (eg., chemical compositions, gas pollutants, and meteorology elements) and significantly improves the accuracy of N_{CCN} prediction, thereby helping to bridge observational gaps. Our modeling framework could then be used to simulate ground-level CCN data in other regions around the world and even on a global scale. Moreover, this approach may guide the development of machine-learning–based models to predict CCN vertical profiles, which are critical for accurately assessing aerosol–cloud interactions…"

Minor Concerns:

Clarity and Presentation Issues.

The methods section is placed in the Supplement, making the manuscript harder to follow.

Re: Thank you for your efforts and time on handling the paper. We have updated the section of **Methods** and see as follows or **Lines 106-225**:

"**2. Methods**

2.1 Study area

In this work, we select the North China Plain (NCP) (32°-40°N and 114°-121°E) as the study area. Being one of the most polluted areas in China, the aerosol particles in NCP are with more complex composition and mixing state, which leads to great challenge in accurate prediction of cloud concentration nuclei (CCN) concentrations. In recent years, emissions of gas pollutants and fine particles have shown a significant downward trend year by year (Wei et al., 2023) due to the implementation of the vigorous emission reduction in China (Zheng et al., 2018). This also makes changes in aerosols CCN activity in the study area from the point of view in assessment of the climate effect of aerosols.

2.2 Model construction and validation

Here we develop the ML-based N_{CCN} prediction model by employing the Random Forest Regression method (RFRM) that has been demonstrated and can solve multivariate and nonlinear regression problems (Nair and Yu, 2020; Liang et al., 2022). The diagram of the model construction and the N_{CCN} prediction is shown in Figure 1. Due to lack of a large spatial scale observed N_{CCN} data, we use simulated N_{CCN} by WRF-Chem model as the targeted variable that can basically capture the ambient temporal variability of CCN concentration despite a deviation of ~40% by comparing with our six field observations (Figure 4). The input parameters include the chemical

components of $PM_{2.5}$ (organic, sulfate, nitrate, ammonium, black carbon) from the Tsinghua University Tracking Air Pollution in China dataset (Liu et al., 2022) and gas and particulate pollutants (nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3) and $PM_{2.5}$) collected from the China National Environmental Monitoring Centre network. Meteorological parameters are from the European Centre for Medium-range Weather Forecasts Reanalysis version 5 (ERA-5) and include temperature, relative humidity (RH), total precipitation (TP), wind speed (WS), wind direction (WD), planetary boundary layer height (BLH), surface pressure (SP) and surface net solar radiation (SNSR). These datasets have undergone validation and been proven highly suitable for developing machine learning models for atmospheric applications (Nair and Yu, 2020; Wei et al., 2023). Cartesian coordinates were also added as input due to the spatiotemporal nature of the input data (Yang et al., 2022). Supplemental Table 1 provides more details about the input parameters.

When constructing the model, all the aforementioned species data (predictor and target variables) were processed to match the same spatiotemporal resolution, and then split into 7:3 ratio for model training and testing respectively. As a result, a total of 274365 samples is included in the training datasets. In order to assure a stronger generalization ability of the N_{CCN} prediction model, the 10-fold cross-validation is adopted (Wei et al., 2023). The optimization parameters of RF model were examined by varying hyperparameters (Fig. S1). In addition, cross-validation (CV) is applied to select the hyperparameters during the data preprocessing (Yang et al., 2022). The CV results showed that when the number of trees (n_estimators) was less than 200, the prediction accuracy increased rapidly with the increase of the number of trees, and then gradually stabilized. According to the CV score and the number of data sample, the number of trees was set to 500 in this study. The impact of max depth on the CV score showed that, with the increase of depth, the complexity of the model increases. Thus, the max depth is set to 28. Also, the model generalization error was larger when the minimum sample number of the leaf and branch node are large, indicating that the model itself is close to the optimal model complexity level. Therefore, a higher value was set given the large sample size in this case. The influence of the maximum selection feature number on CV score showed a trend of increasing first and then decreasing, so the maximum value of CV curve was set to 16.

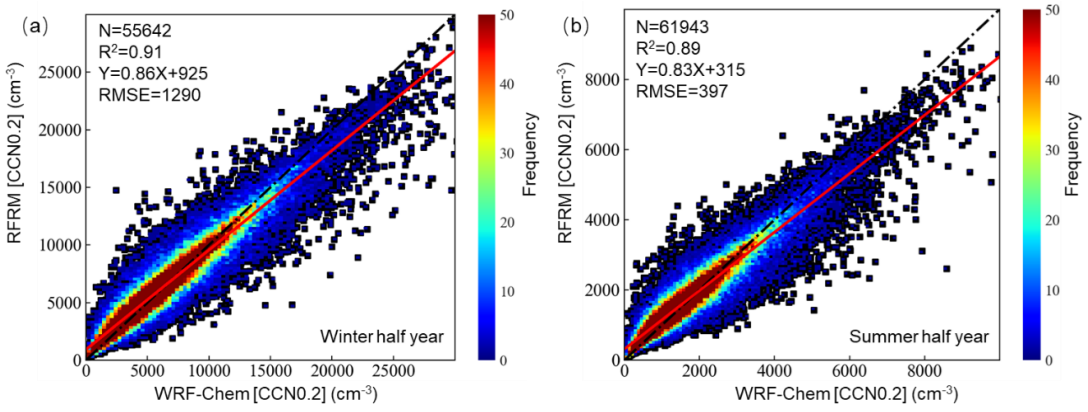

Figure 2. Comparison of RFRM retrieval and WRF-Chem simulated N_{CCN} at S=0.2%. (a and b) Density plots of retrieval N_{CCN} at S=0.2% as a function of the simulations

from WRF-Chem on the testing dataset.

Fig. 2 shows the performance of our developed RFRM models by comparing the testing dataset (sample size: N=117585) of CCN simulated by WRF-Chem with the predicted CCN. Here the quality metrics for model performance are based on the correlation coefficient (R^2), root mean square error (RMSE) and the slope of the RFRM predicted and the WRF-Chem simulated CCN concentrations. It showed that the estimated N_{CCN} at S=0.2% are highly correlated with the values from WRF-Chem, with the correlation R^2 of ~0.89-0.91 and slopes of 0.83 and 0.86 in our developed RFRM model. This suggests that our model works well in estimating N_{CCN} with a high aerosol loading environment. We also found that the accuracy of the CCN prediction will deteriorate slightly if not including the information of chemical compositions (Fig. S2), or if using XGBoost algorithm (Fig. S3) when constructing the model.

Given that the ultimate goal of model development is to more accurately predict the actual atmospheric concentration of CCN, we further conducted a comparative analysis of the prediction results against the in-situ observation data in this region, spanning from the hourly scale to the interannual scale (see Figure 3-6), which will be presented and discussed in detail in Section 3.2 and 3.3.

2.3 Data and other details in the model construction

2.3.1 N_{CCN} simulated by WRF-Chem model

The WRF-Chem version 4.1.5 is used to simulate N_{CCN} in this study, which nested a domain in 10 km×10 km covering the entire NCP (Fig. S4) and contained 181×170 grids. The simulation in WRF-Chem is conducted from 1 January 2014 to 31 December 2018 with an hourly resolution. In the WRF-Chem modeling system, the sectional Model for Simulating Aerosol Interactions and Chemistry (MOSAIC), the Morrison two-moment scheme (Morrison et al., 2009) and the Carbon Bond Mechanism Z photochemical mechanism (Zaveri et al., 1999) are employed. We also compared the simulation using the Regional Acid Deposition Model (Stockwell et al., 1990) and the Lin microphysics scheme (Lin et al., 1983). Considering the calculation efficiency and accuracy with the measurements, the CBMZ-MOSAIC and Morrison 2-monment scheme were finally applied to simulate the long-term CCN concentration. More details about the other parameterizations used for WRF-Chem simulation were given in SI.

2.3.2 Gound-based measurements and datasets

Ground measurements of atmospheric gaseous precursors, fine particles chemical compositions, and CCN number concentration (at supersaturations of 0.2% and 0.4%) were collected during six field campaigns at three sites in the NCP (Fig. 4), used to assess the performance of the developed ML-based model in predicting N_{CCN}. The six campaigns were conducted as follows: at the Beijing (BJ) site from 8–30 November 2014, 20 August to 6 October 2015, 16 November to 20 December 2016, and 28 May to 27 June 2017; at the Xingtai (XT) site from 17 May to 14 June 2016; and at the Gucheng (GC) site from 23 January to 3 February 2018. They are accordingly named BJ2014_WIN, BJ2015_AUT, BJ2016_WIN, BJ2017_SUM, XT2016_SUM, and GC2018_WIN (Fig. 4a).

The BJ site (Longitude: 116.37° E; Latitude: 39.97° N) is located at the meteorological tower station of the Institute of Atmospheric Physics, Chinese Academy

of Sciences. It is representative of the general emission conditions in urban areas of the northern NCP. The primary pollution sources here are surrounding traffic and residential emissions. The XT site (Longitude: 114.37° E; Latitude: 37.18° N) is situated at a national weather station. It is primarily influenced by emissions from surrounding towns and factories (e.g., coal-fired power plants, coking, steel, cement, and chemical industries) and thus reflects polluted suburban conditions in the southern NCP. The GC site (Longitude: 115.74° E; Latitude: 39.15° N) is located at the Integrated Ecological-Meteorological Observation and Experiment Station of the Chinese Academy of Meteorological Sciences. Surrounded mainly by nearby villages, farmland, and transportation networks, this site represents the regional background pollution in the northern NCP.

The CCN number concentrations were measured by using the Droplet Measurement Technologies CCN counter (model CCNC-100, DMT Inc. Lance et al., 2006) at BJ and XT site. The supersaturation (S) levels set for each CCN measurement cycle were 0.1%, 0.2%, 0.4%, and 0.8%, respectively. Another measurement at GC site was referred from Zhang et al. (2020). In this study, the comparisons between the measured and predicted N_{CCN} were mostly based on the value at S=0.2% and S=0.4%. The observed N_{CCN} varies from a few hundred to tens of thousands at these sites, and the campaign mean mass concentration of $PM_{2.5}$ ranges from 35.6 to 160 μg m^{-3} (Fig. 4b), indicating that the observations can represent various atmospheric conditions, spanning from clean to polluted in the region. More details about the observations could be found in Fan et al. (2020), Ren et al. (2018), and Zhang et al. (2019). In addition, the long-term measurement of particle number size distribution (PNSD) at a field site in Beijing (Fig. S5, Shang et al., 2022) is also used for deriving the long-term trend of yearly averaged N_{CCN} …"

Numerous ambiguities and errors hinder comprehension.

For example:

Lines 109–110 describe the study domain as 32°–40°N and 114°–121°E, but Figure S1 shows a different region.

Re: Here Figure S1 has been revised as Figure S4. It shows the simulation domain of WRF-Chem, which nested a domain in 10 km×10 km covering the entire North China Plain and contained 181×170 grids. A region within 32°-40°N and 114°-121°E in the NCP is chosen as the study area. The distance between the study area and the boundary of the simulation domain must be greater than 10 times of the resolution. Our study area is within the range of Fig. S4. The sentence has been revised as follows or see **Lines 108-109** and **176-178**:

"…In this work, we select the North China Plain (NCP) (32°-40°N and 114°-121°E) as the study area …"

"…The WRF-Chem version 4.1.5 is used to simulate N_{CCN} in this study, which nested a domain in 10 km×10 km covering the entire NCP (Fig. S4) and contained 181×170 grids …"

Lines 117–119 incorrectly state that simulated N_{CCN} is an input to the RFRM model (it should be the output).

Re: The sentence has been revised as follows or see **Lines 122-125**:

"…Due to lack of a large spatial scale observed N_{CCN} data, we use simulated N_{CCN} by WRF-Chem model as the targeted variable that can basically capture the ambient temporal variability of CCN concentration despite a deviation of ~40% by comparing with our six field observations (Figure 4) …"

Lines 154–155: The phrase "more to the model's output" is unclear.

Re: The sentence has been revised as follows or see **Lines 241-243**:

"…During the winter, changes in BLH contribute more to CCN predictions than $PM_{2.5}$ (Fig. 3b) and the model's output changes more significantly with this factor (Fig. 3c) …"

Figure 3c: The frequency unit appears to be 1e-8, but this is not explicitly stated.

Re: Note that figure 3 has been revised as figure 4. The figure has been revised. See follows:

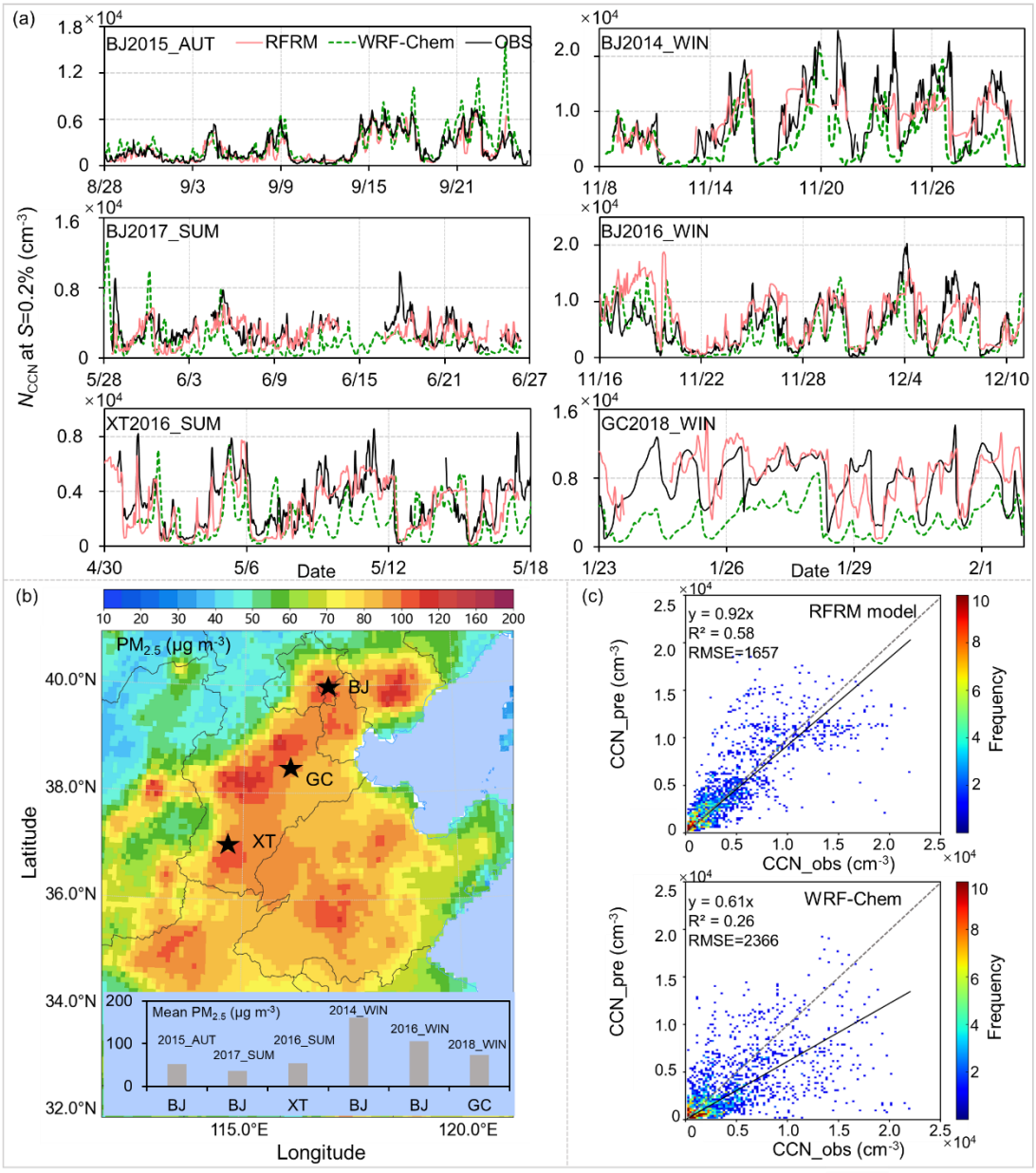

Fig. 4 Performance of the RFRM model in predicting N_{CCN} at field sites in NCP. (a) Time series of the observed and predicted CCN number concentrations at S=0.2% for the six campaigns (BJ2015_AUT, BJ2017_SUM, XT2016_SUM, BJ2014_WIN,

BJ2016_WIN, GC2018_WIN) in the North China Plain; (b) Map for average mass concentration of PM2.5 of 2014 from TAP dataset in NCP (http://tapdata.org.cn/) and field observed average mass concentration of $PM_{2.5}$ during the six field campaigns (see embedded histogram); (c) Scatter plots of the observed N_{CCN} at S=0.2% with the RFRM model predicted (top) and WRF-Chem simulated (bottom) respectively.

Figure 6e shows N_{CCN} uncertainties within 150%, while Figures 6a–d display uncertainties exceeding 500%.

Re: Note that figure 6 has been revised as figure 8. And here figure 8a–d present all-sites data points of the six observation campaigns. The statistical results show that the Random Forest Regression Model (RFRM) errors range from –90 to +600%, whereas the WRF-Chem model exhibits a broader error span of –100 to +1800% when compared with the observations. Figure 8e summarizes the mean values across these campaigns with the N_{CCN} uncertainties within 150%. Some descriptions have been added as follows or see **Lines 446-448:**

"…While, the mean uncertainties for all these parameters are largely reduced when the mean underestimation of ~8±38% in N_{CCN} at S=0.2% that is caused by RFRM model is applied (Fig. 8e) …"

Similar problems appear in many places throughout the article, accompanied by punctuation errors, improper use of terms, etc., making reading extremely difficult.

Insufficient Explanation of Counterintuitive Results.

For instance, Figure 2a shows that sulfate has low permutation importance but high R-Square. The authors do not adequately explain or validate this finding, leaving readers to speculate.

Re: Despite the high correlation between sulfate features and the target variable, their importance scores within the RFRM model remain low. Two main factors explain this:

a. Nonlinear model behavior

Random Forest is a nonlinear algorithm that constructs an ensemble of decision trees; it captures complex, non-additive interactions between predictors and response variables. As a result, even a feature with a strong linear correlation to the outcome may not play a pivotal role in the trees' local splits. Thus, sulfate may exhibit high correlation with CCN concentration but contribute little to the actual partitioning decisions made by the model.

b. Collinearity with other predictors.

Strong inter-feature correlations (e.g. sulfate with nitrate/ammonium at 0.84-0.92/0.92-0.95) lead the model to favor one predictor (e.g. nitrate) over others when building decision splits. Because Random Forest often uses only one variable from a set of highly correlated candidates to optimally partition the data, sulfate's importance score can be artificially diminished, despite sharing information with the target.

Considering that the high hygroscopicity of sulfates is an effective seed for CCN, sulfate features were not removed during model training in our study.

Some explanation has been added in the revised text, see as follows or **Lines 250-259:**

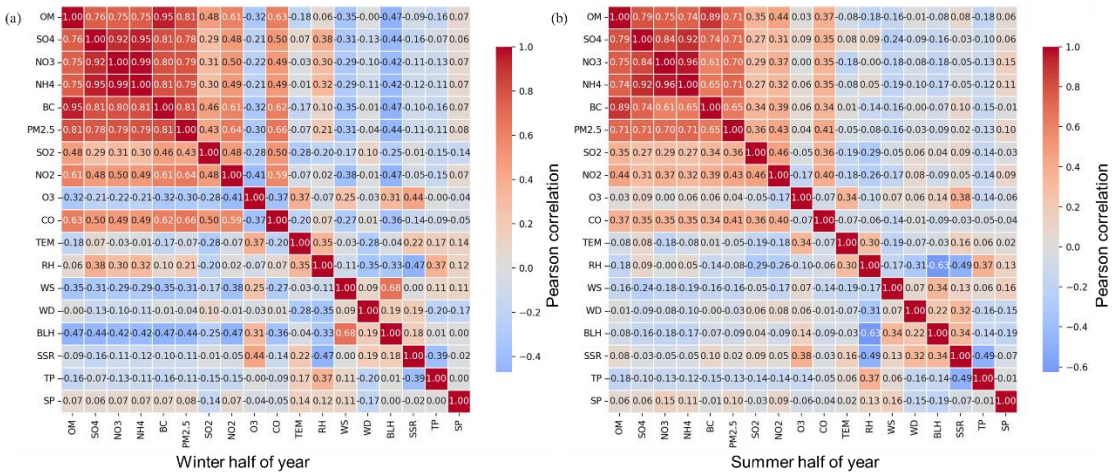

Figure S7. Heatmap of the feature variables in the winter half of year (a) and summer half of year (b).

"… Note that the impact of sulfate aerosols on N_{CCN} prediction is much less important in both summer and winter seasons compared to nitrate particles, with a permutation importance score ranging from ~0.02 to 0.03 but with higher correlation of ~0.31-0.49., This is mainly because the collinearity with nitrate features (~0.84-0.92) as seen in Fig. S7. In general, the machine learning algorithm often chooses one variable from a set of highly correlated candidates to optimally partition the data. Here sulfate's importance score can be artificially diminished, largely due to its decreased proportion in $PM_{2.5}$ in recent years (Liang et al., 2022; Li et al., 2020). As a note, due to the high hygroscopicity of sulfates is an effective seed for CCN, it was not removed in RFRM model…"

Overstated Claims About Model Applicability.

Lines 382–384 suggest that integrating this framework into traditional climate models could reduce aerosol indirect effect uncertainties. However, since the model is only validated within its training spatiotemporal domain, such claims about generalizability are premature. The authors should temper these statements or provide evidence of the model's robustness beyond the tested conditions.

Re: The sentence has been revised in the text, see as follows or **Lines 477-480**:

"…Given the simplified setting in current climate models, this work emphasizes the necessity and urgency to obtain the precise N_{CCN} values, offering a new framework for predicting CCN concentrations based on machine learning algorithms and effectively filling the observation gap of CCN concentrations…"

References:

Fanourgakis, G., Kanakidou, M., Nenes, A. et al.: Evaluation of global simulations of aerosol particle and cloud condensation nuclei number, with implications for cloud droplet formation, Atmos. Chem. Phys., 19(13), 8591–8617, https://doi.org/ 10.5194/acp-19-8591-2019, 2019.

Nair, A. A., Yu, F.: Using machine learning to derive cloud condensation nuclei number concentrations from commonly available measurements, Atmos. Chem. Phys., 20(21), 12853–12869, https://doi.org/10.5194/acp-20-12853-2020, 2020.

Yang, N., Shi, H., Tang, H. et al.: Geographical and temporal encoding for improving the estimation of $PM_{2.5}$ concentrations in China using end-to-end gradient boosting, Remote Sensing of

Environment, 269, 112828, https://doi.org/10.1016/j.rse.2021.112828, 2022.

Wei, J., Li, Z., Wang, J. et al.: Ground-level gaseous pollutants (NO_2, SO_2, and CO) in China: daily seamless mapping and spatiotemporal variations, Atmos. Chem. Phys., 23, 1511–1532, https://doi.org/10.5194/acp-23-1511-2023, 2023.

Morrison, H., Thompson, G., Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes, Monthly Weather Review, 137(3), 991–1007, https://doi.org/10.1175/2008MWR2556.1, 2009.

Zaveri, R. A., Peters, L. K.: A new lumped structure photochemical mechanism for large-scale applications, Journal of Geophysical Research: Atmospheres, 104, 30387–30415, https://doi.org/10.1029/1999JD900876, 1999.

Stockwell, W. R., Middleton, P., Chang, J. S., et al.: The second generation regional acid deposition model chemical mechanism for regional air quality modeling, Journal of Geophysical Research: Atmospheres, 95(D10): 16343-16367, https://doi.org/10.1029/JD095iD10p16343, 1990.

Lin, Y., Farley, R., Orville, H.: Bulk parameterization of the snow field in a cloud model, Journal of Applied Meteorology and Climatology, 22(6): 1065-1092, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2, 1983.

Fan, X., Liu, J., Zhang, F. et al.: Contrasting size-resolved hygroscopicity of fine particles derived by HTDMA and HR-ToF-AMS measurements between summer and winter in Beijing: the impacts of aerosol aging and local emissions, Atmos. Chem. Phys., 20, 915–929, https://doi.org/10.5194/acp-20-915-2020, 2020.

Ren, J., Zhang, F., Wang, Y. et al.: Using different assumptions of aerosol mixing state and chemical composition to predict CCN concentrations based on field measurements in urban Beijing, Atmos. Chem. Phys., 18(9), 6907–6921, https://doi.org/10.5194/acp-18-6907-2018, 2018.

Zhang, F., Ren, J., Fan, T. et al.: Significantly enhanced aerosol CCN activity and number concentrations by nucleation-initiated haze events: A case study in urban Beijing, Journal of Geophysical Research: Atmospheres, 124(24), 14102–14113, https://doi.org/10.1029/2019JD031457, 2019.

Lance, S., Nenes, A., Medina, J. et al.: Mapping the operation of the DMT continuous flow CCN counter, Aerosol Science and Technology, 40(4), 242–254, https://doi.org/10.1080/02786820500543290, 2006.

Zhang, Y., Tao, J., Ma, N. et al.: Predicting cloud condensation nuclei number concentration based on conventional measurements of aerosol properties in the North China Plain, Science of The Total Environment, 719, 137473, https://doi.org/10.1016/j.scitotenv.2020.137473, 2020.

Liang, M., Tao, J., Ma, N. et al.: Prediction of CCN spectra parameters in the North China Plain using a random forest model, Atmospheric Environment, 289, 119323, https://doi.org/10.1016/j.atmosenv.2022.119323, 2022.

Li, S., Zhang, F., Jin, X. et al.: Characterizing the ratio of nitrate to sulfate in ambient fine particles of urban Beijing during 2018-2019, Atmospheric Environment, 237, https://doi.org/10.1016/j.atmosenv.2020.117662, 2020.

Liu, S., Geng, G., Xiao, Q., Zheng, Y., Liu, X., Cheng, J., & Zhang, Q.: Tracking daily concentrations of PM2.5 chemical composition in China since 2000, Environ Sci Technol, 56, 16517–16527, https://doi.org/10.1021/acs.est.2c06510, 2022.

Zheng, B., Tong, D., Li, M., Liu, F. et al.: Trends in China's anthropogenic emissions since 2010 as

the consequence of clean air actions, Atmos. Chem. Phys., 18, 14095-14111, https://doi.org/10.5194/acp-18-14095-2018, 2018.

Shang, D., Tang, L., Fang, X. et al.: Variations in source contributions of particle number concentration under long-term emission control in winter of urban Beijing, Environmental Pollution, 304, 119072, https://doi.org/10.1016/j.envpol.2022.119072, 2022.