Response to the reviewers

Referee #1

This study applies machine learning (random forest) to the spatiotemporal prediction of cloud condensation nuclei (CCN) concentrations in the heavily polluted North China Plain. It significantly reduces the CCN simulation bias from −59% to approximately −31%, and consequently lowers the simulation uncertainty of cloud radiative forcing from 1.89 ± 0.78 W m^{-2} to 0.81 ± 0.63 W m^{-2}. This provides a new method for accurately assessing aerosol–cloud interactions and the climate benefits of pollution control. Moreover, by incorporating observational data such as PM2.5, NO2, and SO2, the model captured the long-term decreasing trend in aerosol concentration from 2014 to 2018 and quantified the reduction in cloud radiative forcing uncertainty achieved by mitigating N_{CCN} simulation biases. This topic is highly relevant to the journal of Geoscientific Model Development. The model incorporates the CCN concentrations simulated by WRF-Chem as an input, with the observed CCN as the target. This represents a reasonable bias-correction approach. The methodology and results are interesting. Thus, I suggest several minor modifications to be made before publication.

Re: We sincerely appreciate the time and effort the reviewers have dedicated to evaluating our manuscript. Based on the reviewers' comments and suggestions, we have addressed each point accordingly, and we believe the revised version has been significantly improved.

Minor Concerns:

Line 160: "TAP" is used without definition; please provide the full name (Tracking Air Pollution in China) at first use.

Re: Revised.

Line 162: The OM and sulfate in the TAP dataset are "underestimated by approximately 50%" and therefore apply a "twofold correction factor." However, it is unclear whether this correction is applied during the training phase, the prediction phase, or both. If the correction is applied only during the prediction phase while the training phase uses uncorrected data, this would lead to a distribution mismatch between training and inference.

Re: Thanks for the comments. During model construction, the TAP dataset was not adopted. Instead, the model was developed solely based on observational data from several monitoring sites. In this study, the TAP dataset was only utilized to derive region-scale CCN number concentrations using the established model.

As described in the Methods section (Section 2.2), all chemical composition data used for model training and testing were derived from ground-based observations. See **Lines 125-130**:

"…The main dataset that are used to construct the ML-based N_{CCN} prediction model were from six field campaigns at three sites in the NCP (more details see section 2.3.1). The observed CCN number concentration (at supersaturations of 0.2% and 0.4%) is as targeted parameter, and the simultaneously measured atmospheric gaseous precursors, fine particles chemical compositions, and meteorological parameters are as model input parameters…"

The TAP dataset was used exclusively for the regional-scale and interannual predictions. Prior to using the TAP data, we first compared its chemical components with collocated ground-based observations. This comparison revealed a systematic underestimation of OM and sulfate in the TAP dataset (approximately 50%). To ensure consistency between the model input during prediction and the observed data used for training, we therefore applied a twofold correction factor only to the OM and sulfate concentrations from the TAP dataset during the prediction phase.

See **Lines 163-166**: "…Compared with the observations at the BJ site, mass concentrations for the OM and sulfate from TAP dataset were largely underestimated by approximately 50% (Fig. S1). Therefore, a twofold correction factor was applied to these components when estimating the regional scale and interannual CCN concentrations in NCP…"

Line 199-201 and 276-277: The sentence "the campaign mean mass concentration of PM2.5 ranges from 35.6 to 160 μg m^{-3}, indicating that the observations can represent various atmospheric conditions" is repetitive with the introduction in Section 2.3.1.

Re: The sentence in **Lines 201-203** has been revised as: "The observed N_{CCN} varies from a few hundred to tens of thousands at these sites, indicating that the observations can represent various atmospheric conditions."

Line 236: The authors interpret the SHAP values as evidence that OM has strong hygroscopicity driving CCN activation. However, SHAP values reflect association rather than causation. It is recommended to add a qualifying statement: "While SHAP indicates a strong association, the causal interpretation should be supported by the known hygroscopicity of OM in the NCP (Liu et al., 2021)."

Re: Thank you for your insightful comment. Yes, the SHAP value explains the potential influence of model predictors on the predicted CCN concentration. The wide range of SHAP values for OM reflects the diversity of its physicochemical properties—OM can act either as an inhibitor of CCN activation or as a strong promoter. Low SHAP values may arise from freshly emitted hydrophobic OM, which suppresses CCN activation and contributes little or even negatively. In contrast, high SHAP values correspond to aged/oxidized OM with enhanced hygroscopicity, which significantly promotes CCN activation. We have added some explanation as follows or see **Lines 236-254**:

"…Organic matter emerges as the most crucial indicator with the highest SHAP value. The wide range of SHAP values for OM reflects the diversity of its physicochemical properties. Specifically, low-concentration or freshly emitted hydrophobic OM contributes negatively to NCCN (suppressing activation), whereas high-concentration or aged/oxidized OM contributes positively (promoting activation). From an overall perspective, SHAP values increase monotonically with OM concentration, and the absolute values of positive SHAP values exceed those of negative ones, demonstrating a synergistic positive effect of OM concentration on the variation of CCN number concentration. This finding differs from the conventional view that inorganic salts contribute more to CCN due to their strong hygroscopicity (Petters and Kreidenweis, 2007). However, in fact, we also note that under conditions of high OM, the concentrations of CN and CCN indeed show an increasing trend (Fig. S5). In addition, previous studies have shown that in the North China region where the proportion and concentration of OM are both high, organic particles affected by strong anthropogenic emission sources was found exhibit strong hygroscopicity, enabling them to serve as more effective CCN (Liu et al., 2021); in addition, the surface tension lowering effect of OM particles in this region can also enhance particle CCN activity (Fan et al., 2024). Therefore, the SHAP analysis results further confirm the conclusions of previous studies…"

Line 289: "During the GC2018_WIN campaign, the observed N_{CCN} is underestimated by as much as 71% by WRF-Chem (Fig. S6). Here, it might be referring to Fig. S5?"

Re: In the revised text, it referred to Fig. S6.

Lin 289-296: It notes that the model's performance improves much more during severely polluted winter conditions than during cleaner summer conditions. It is recommended to emphasize more explicitly in the Conclusions that the improvement is particularly pronounced under highly polluted (cold-season) conditions, which has practical implications for CCN prediction in heavily polluted regions.

Re: Thank you for this suggestion. In response, we have revised the Conclusions section to explicitly state that the model's improvement is particularly pronounced under severely polluted (cold-season) conditions. The sentence has been added as follows or see **Lines 476-481**: "…The results show that the prediction bias of N_{CCN} compared to observations is approximately -31% from the RFRM model. Good accuracy has also been achieved during heavy pollution periods or cold seasons. This improvement is much more pronounced under severely polluted winter conditions, demonstrating the model's particular value for CCN prediction in heavily polluted environments…"

Line 297: "The improvements in RFRM model also demonstrate the effectiveness of the model trained on atmospheric variables to revise the simulation in model". The phrase "in model" at the end of the sentence is unclear in meaning.

Re: Revised. See follows: "The improvements in RFRM model also demonstrate the effectiveness of the model trained on atmospheric variables to revise the simulation in WRF-Chem model."

Line 298: The article has already used another observation at GC site to provide independent spatiotemporal validation. This should be discussed in detail to demonstrate that the model's generalizability has been effectively verified.

Re: Revised. The sentences have been added as follows or see **Lines 311-316**:

"… In addition, we also evaluated the model performance based on another observation at GC site in January (Zhang et al., 2020) (Fig. S7). Compared to WRF-Chem simulations, the RFRM model could more accurately captures the peak and valley CCN concentrations. The mode showed the greatest improvement and the underestimation is largely improved with the predicted bias of only 4% in the RFRM model (Fig. S7) …"

Line 307: "the underestimation of N_{CCN} by the WRF-Chem model is likely due to the overestimation of the organics and BC mass fraction induced by WRF-Chem (Fig. S8), but the underestimation of the hygroscopic parameter of organics, and the simplified prescriptions in particle size distribution". "But the clauses before and after 'but' do not indicate a contrast."

Re: The sentences have been revised as follows or see **Lines 322-325**: "…the underestimation of NCCN by the WRF-Chem model is likely due to the overestimation of the organics and BC mass fraction induced by WRF-Chem (Fig. S9), along with the underestimation of the hygroscopic parameter of organics, and the simplified prescriptions in particle size distribution..."

Line 329: Here they used long-term PNSD measurements and κ-Köhler theory to calculate the "observed" annual mean N_{CCN} (N_{CCN}_obs). However, the κ values themselves are derived from the TAP dataset (which is biased even after correction). Moreover, the authors note that "κ values are much less sensitive to changes in NCCN compared to the PNSD" (Lines 345–347). A sensitivity analysis is needed to quantify the impact of κ uncertainty on the final N_{CCN} _obs.

Re: Thanks for the comments. The sensitivity analysis was shown in Fig. R1. Revised see follows or **Lines 360-371**:

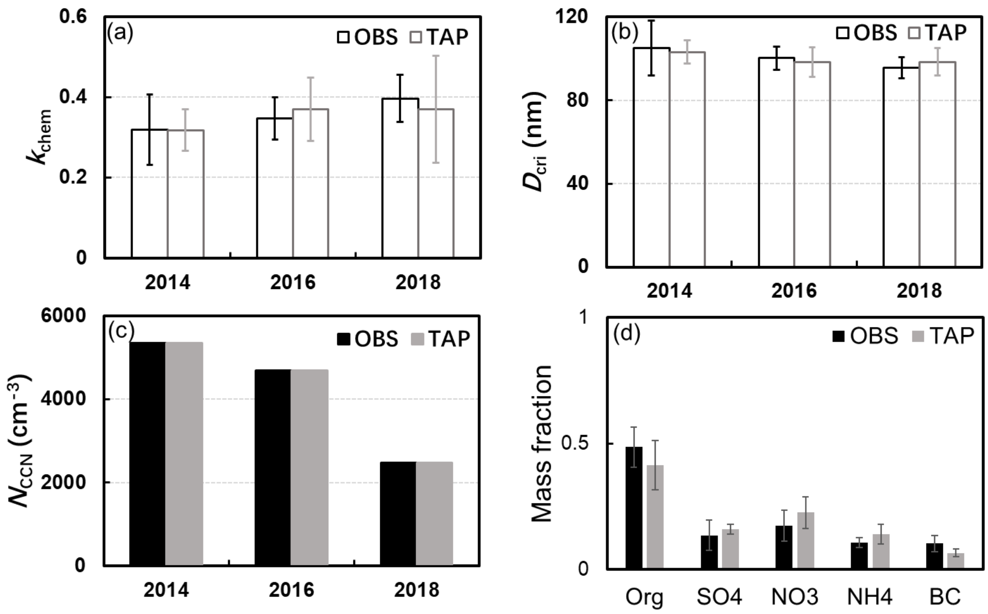

Fig. R1 Average annual value of the κ_{chem} calculated from chemical composition (a), the critical diameter at S=0.2% (b), N_{CCN} at S=0.2% (c), the mass fraction (d) between the observed and TAP dataset in the winter of 2014, 2016, 2018.

"…A comparison of the values of κ and N_{CCN} between that derived using field observations and the TAP dataset shows little differences (Fig. S10); actually, the long-term change of N_{CCN} is much less sensitive to changes in κ values compared to the PNSD (Fig. S10c). Sensitivity analysis showed that a ±20% change in κ leads to a change in N_{CCN} of approximately ±8%. Comparing the κ_{chem} derived from the TAP method and the OBS method, the difference is approximately ±6%, the estimated deviation in the critical activation diameter ranges from -2% to 3%. Although absolute concentrations of components in the TAP dataset deviate from observations, their mass fractions are consistent (Fig. S10d), rendering the impact on the calculated κ negligible. In addition, the method to calculate N_{CCN} at S=0.2% based on κ-Köhler theory would cause an upper-limit uncertainty of 7% (Ren et al., 2018) …"

Line 476: In the "Limitations and outlook" section, the authors honestly acknowledge that the observational data come from only six campaigns at three sites. It is recommended to add a discussion on the spatiotemporal representativeness of these observations.

Re: Thank you for your suggestion. We have added the following discussion in the "Limitations and outlook" section, see **Lines 501-511**:

"…Second, this study analyzes observational data from six campaigns conducted at three sites. Although the number of sites is limited, these sites represent urban, suburban, and regional background conditions, respectively, and the observation periods cover different seasons and years. Therefore, the current dataset can reasonably characterize the overall aerosol and CCN conditions in North China, and may also provide useful implications for other polluted regions with similar emission characteristics. Validating the simulated N_{CCN} through comparisons with observations at more ground sites is warranted in future. Also, 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…"

Line 732: Figure 2 shows an R^2 of 0.86–0.95 for the test set, but Figure 3c shows an R^2 of only 0.86 (RFRM vs. observations). These two R^2 values are calculated for different targets (the former may be RFRM vs. WRF-Chem? The latter is RFRM vs. observations). This should be clearly stated in the figure captions.
Re: Figure 2 presents the results on the test set, while Figure 3 presents the results on the validation set for the selected independent cases of the six campaigns. The caption of Figure 3 has already been revised see follows: "Fig. 3 Performance of the RFRM model in predicting N_{CCN} at field sites in NCP. (a) Time series of the observed and predicted N_{CCN} at S=0.2% for the six periods (BJ2015_AUT, BJ2017_SUM, XT2016_SUM, BJ2014_WIN, BJ2016_WIN, GC2018_WIN) in the North China Plain for the validation set."

Line 732: In Figure 2b, the SHAP plot lacks units or dimensions. Please clarify in the caption: "SHAP values represent the contribution to N_{CCN} (in cm^{-3})."

Re: Revised.

Line 737: Use N_{CCN} consistently throughout the manuscript. For example, in the title of Fig. 3 on page 34, it reads "Time series of the observed and predicted CCN number concentrations", while the axis label reads "N_{CCN} (cm^{-3})"

Re: The caption has been revised.

The manuscript uses both "RFRM" (Random Forest Regression Method) and "RF model" (line 138). It is recommended to use "RFRM model" consistently throughout.

Re: Revised.