https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

Machine learning significantly improves the simulation of hourly-to-yearly scale cloud nuclei concentration and radiative forcing in polluted atmosphere

Jingye Ren[1,2], Songjian Zou[3], Honghao Xu[3], Guiquan Liu[3], Zhe Wang[3], Anran Zhang[3], Chuanfeng Zhao[4], Min Hu[5], Dongjie Shang[5], Lizi Tang[5], Ru-Jin Huang[1], Yele Sun[6], Fang Zhang[3]*

[1]State Key Laboratory of Loess Science, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, 710061, China

[2]Xi'an Institute for Innovative Earth Environment Research, Xi'an, 710061, China

[3]Shenzhen Key Laboratory of Organic Pollution Prevention and Control, School of Eco-Environment, Harbin Institute of Technology Shenzhen, Shenzhen, 518055, China

[4]Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, 100871, China

[5]State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Peking University, Beijing, 100871, China

[6]State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China

Corresponding author: Fang Zhang, *zhangfang2021@hit.edu.cn*

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

Abstract

The accurate prediction of cloud condensation nuclei (CCN) number concentration (N_{CCN}) on a large spatiotemporal scale is challenging but critical to evaluate the aerosol cloud interaction effect. Combining multi-source dataset and the N_{CCN} simulated by the Weather Research and Forecasting coupled with Chemistry (WRF-Chem) model, we have developed a Random Forest Regression method (RFRM) model which achieves well prediction of hourly-to-yearly scale N_{CCN} at typical supersaturations in polluted North China Plain (NCP). We show that the prediction bias of N_{CCN} compared to observations is reduced from -59% with the WRF-Chem model to approximately -31% with the RFRM model (the prediction precision is improved by 1.6 times accordingly) during the campaigns. The greatest improvement is seen in both very polluted and clean cases. The RFRM model captures well the spatial variation and better describes long-term trends of N_{CCN}. More importantly, the prediction reveals a significant long-term decreasing trend of N_{CCN} in NCP due to a rapid reduction in aerosol concentrations from 2014 to 2018, during which a series of strict emission reduction measures were implemented by the Chinese government. This reflects the climate benefit of pollution control. Our study further illustrates that the RFRM model reduces the uncertainty in simulating cloud radiative forcing from an overestimation of 1.89±0.78 W m^{-2} to 0.81±0.63 W m^{-2}, illustrating the high sensitivity of climate forcing to changes in N_{CCN}. This work offers a new modeling framework that guides the way to simulate CCN in other regions around the world and has the potential to effectively filling the observation gap of CCN concentrations.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

1. Introduction

Aerosol indirect radiative effects caused by aerosol-cloud interactions (ACI) are the largest source of uncertainty when assessing climate change (IPCC 2021). A major issue is the lack of an accurate characterization of cloud condensation nuclei (CCN, or cloud nuclei) number concentrations (N_{CCN}) in global climate models (Sotiropoulou et al., 2007; Fanourgakis et al., 2019). This is largely due to the nonlinear interactions between the aerosol physical and chemical properties and CCN, making the quantification of the N_{CCN} remain highly uncertain. Current models tend to underestimate CCN number concentrations by 20–40% on average, based on comparisons between results from 14 models and observations at sites distributed globally (Fanourgakis et al., 2019). Biases were greater in the Northern Hemisphere due to intensive human activities (Sotiropoulou et al., 2007). It has been proposed that ~10–30% changes in cloud droplet concentrations might be associated with the uncertainties in cloud radiative forcing by ~0.1–2 W m^{-2} (Charlson et al., 1987; Sotiropoulou et al., 2007; Yu et al., 2022). Towards improving estimates of the ACI effect in modeling, it is critical to obtain accurate spatiotemporal distributions of CCN.

Predicting N_{CCN} remains challenging because aerosol CCN activity varies greatly in time and space and involves microphysical and chemical processes. Previous studies have underscored that the main uncertainties in simulating N_{CCN} at regional and global scales are the simplified representations of particle size distribution (Menon & Rotstayn, 2006), as well as the lack of detailed treatment of the microphysical and chemical

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

processes in current models (Sha et al., 2019; Yu et al., 2020). Therefore, a considerable number of CCN closure studies have been carried out to predict N_{CCN} in different environments (Moore et al., 2012; Zhang et al., 2014, 2016, 2017, 2019; Xu et al., 2021; Ren et al., 2018, 2023). Although extensive CCN observations and closure studies might help to reconcile this uncertainty, field measurements are relatively sparse and have only been collected during a few campaigns at a few sites (Schmale et al., 2018; Rose et al., 2021) due to the limitations of techniques and cost. Closure studies, however, have mostly focused on investigating the relative importance of the aerosol physical and chemical properties to CCN prediction and have not yet provided a CCN parameterization scheme that is applicable to different regions over the globe. Some studies have attempted to develop a correlation between aerosol optical properties and CCN number concentrations (Rosenfeld et al., 2016). Compared with the measurement of CCN at ground sites, satellite retrieval methods offer global coverage with high spatial and temporal resolutions (Rosenfeld et al., 2016; Liu et al., 2020). However, they are limited to clear-sky conditions. Due to the aerosol swelling effect (Liu et al., 2018), there are typically large deviations of -30% to +90% in the estimation accuracy of N_{CCN} in different environments (Shen et al., 2019).

In recent years, machine learning (ML) has been used for the inversion of atmospheric environmental parameters such as tropospheric ozone and particulate matter with a diameter of 2.5 μm or smaller ($PM_{2.5}$) (Grange et al., 2018; Geng et al., 2021; Wei et al., 2023). To our knowledge, ML-based prediction of CCN properties is few and far between, with studies only focused on analyzing the importance of different

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

variables to estimating CCN spectra at a single field site (Liang et al., 2022) or were conducted in relative clean regions (Nair et al., 2020, 2021). It would be a step forward to use an ML-based method to predict N_{CCN} in polluted areas because it would help to verify the applicability of the method in different regions, but most importantly, it would improve model simulations of the ACI effect in polluted regions where errors in predicting N_{CCN} are greater than in clean regions.

In this study, we have developed a ML-based model for predicting N_{CCN} on hourly-to-yearly scale in the heavily polluted North China Plain (NCP) by using a multi-source dataset of atmospheric variables and CCN concentration outputs from the Weather Research and Forecasting coupled with online chemistry (WRF-Chem) model. We have presented and analyzed the relative importance of the different parameters to CCN prediction. Moreover, we have verified the performance of the RFRM model in predicting the N_{CCN} over different temporal and spatial scales in the NCP. Finally, by incorporating the N_{CCN} prediction biases into the evaluation of cloud parameters and radiative forcing, we investigate the sensitivity of aerosols indirect climate forcing to CCN concentrations changes.

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

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

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. 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. Considering the motivation and main goal of this study is to achieve the prediction of CCN number concentration on a relatively large spatiotemporal scale (e.g. regional scale), especially in polluted areas. Therefore, to better constrain the performance and accuracy of the model in regional-scale predictions, we also incorporated the N_{CCN} simulated by WRF-Chem (more details see section 2.3.2) as an input parameter for model training when constructing the model.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

When constructing the model, the N_{CCN} simulated by WRF-Chem were processed to match the locations of each observation site and their corresponding measurement periods. The data were split into 7:3 ratio for model training and testing respectively. 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. In addition, cross-validation (CV) is applied to select the hyperparameters during the data preprocessing (Yang et al., 2022).

Here the quality metrics for model performance are based on the correlation coefficient (R^2), root mean square error (RMSE), the slope of the RFRM predicted with the observed CCN concentrations and the normalized mean bias (NMB). It showed that the estimated N_{CCN} are highly correlated with the observed values for the testing dataset, with the correlation R^2 of ~0.86-0.95 and slopes of 0.80-0.88 at S=0.2% and S=0.4% (Fig. 2). This suggests that our model works well in estimating N_{CCN} with a high aerosol loading environment.

Given that the primary goal of model development is to more accurately predict the relatively large spatial and temporal CCN concentrations, finally, the RFRM model was employed by integrating with multisource-fusion datasets (more details see 2.3.3) to predict regional and yearly scale CCN concentrations. Supplemental Table 1 also provides more details about the multisource-fusion datasets. According to the results of the RFRM model and SHAP (SHapley Additive exPlanations) analysis (Fig. 2), fine particle chemical components such as organic matter (OM), ammonium ions (NH_4^+),

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

and nitrate ions (NO_3^-) are the top-ranked factors with significant impacts on the prediction results. Particularly, OM is the most important factor among all influencing factors for the predictions. Therefore, prior to conducting predictions based on the multi-source dataset, we compared and validated the reliability of the TAP multi-source dataset against the available observational data (Fig. S1). 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. Cartesian coordinates were also added as input due to the spatiotemporal nature of the input data (Yang et al., 2022).

2.3 Data and other details in the model construction

2.3.1 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, used to construct the ML-based N_{CCN} prediction model. 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 16 November to 15 December 2018. Here it should be noted that this study additionally selected six independent cases to assess the performance of the developed ML-based model in

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

predicting N_{CCN} (Fig. S2). They are accordingly named BJ2014_WIN, BJ2015_AUT, BJ2016_WIN, BJ2017_SUM, XT2016_SPR, and GC2018_WIN (Fig. 3a). These cases were excluded from both the training and testing datasets described in Section 2.2.

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 Tao et al. (2021). 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

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

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}, 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. S3, Shang et al., 2022) is also used for deriving the long-term trend of yearly averaged N_{CCN}.

2.3.2 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.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

2.3.3 Multisource-Fusion datasets

The RFRM model was further employed by integrating with multisource-fusion datasets to estimate regional scale and long-term CCN concentrations (Section 3.3 and 3.4). Those dataset include the N_{CCN} simulated by WRF-Chem, the chemical components of $PM_{2.5}$ (organic, sulfate, nitrate, ammonium) that were taken from the Tsinghua University Tracking Air Pollution in China dataset (Liu et al., 2022), the gas and particulate pollutants (nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3) and $PM_{2.5}$) which were collected from the China National Environmental Monitoring Centre network, meteorological parameters (from the European Centre for Medium-range Weather Forecasts Reanalysis version 5,ERA-5) of temperature (Tem), relative humidity (RH), planetary boundary layer height (BLH), surface pressure (SP) and surface net solar radiation (SNSR).

3. Results

3.1 The relative importance of the input parameters to the prediction of N_{CCN}

The SHAP algorithm is employed to interpret the outputs of the RFRM model, as illustrated in Fig. 2. Organic matter emerges as the most crucial indicator with the highest SHAP value. It indicates a strong, monotonic positive effect of organic matter on N_{CCN}, with low-OM contributing negatively to N_{CCN}, while high-OM contributing positively to N_{CCN}. It reflects the synergistic positive effect of organic matter concentration on the variation of CCN number concentration, which differs from the conventional view that inorganic salts contribute more CCN due to their strong

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

hygroscopicity (Petters and Kreidenweis, 2007). However, in fact, 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. $PM_{2.5}$ concentration is the second most important factor, only after OM. It shows that higher $PM_{2.5}$ mass concentration correspond to positive SHAP contributions, meaning that increase in $PM_{2.5}$ will increase CCN concentrations. Actually, it has been demonstrated that the strong association of $PM_{2.5}$ mass concentration with the CCN concentrations in NCP, due to the synchronously growth in particle size and enhanced hygroscopicity (Zhang et al., 2019). The SO_2 and NO_2 are also among the relatively top-ranked factors, with positive effect on CCN levels according to the SHAP value. This highlights the potential impact of gaseous precursors on CCN activity. Ammonium (NH_4^+) also contributes positively to CCN predictions, with relative larger SHAP values, though occasional negative values might suggest context-dependent effects under certain chemical regimes (Dinar et al., 2008). Nitrate (NO_3^-) exhibited a moderate positive effect in the model, with relatively concentrated SHAP value distributions indicating its stable contribution to the model output. Temperature demonstrated a bidirectional influence, suggesting nonlinear modulation of CCN activity potentially associated with

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

the temperature dependence of nucleation growth and secondary generation of particles (Song et al., 2022).

In contrast, other meteorological parameters (RH, SP, PBL, SNSR) showed lower SHAP values, implying marginal contribution to the model output during the study period. Overall, the high SHAP values of organic matter (OM), $PM_{2.5}$, SO_2 and NO_2 underscore the critical role of chemical constituents and gaseous precursors in CCN predictions, which can be well explained by previous known physical mechanisms of the impact of aerosol particles atmospheric processes on CCN activity.

3.2 Performance of the RFRM model in predicting N_{CCN} at field sites in the NCP

To assess the performance of the RFRM model in predicting N_{CCN}, we compare the predicted N_{CCN} with both the simulations by WRF-Chem and the observations in NCP (Fig. 3). Here, the selected independent cases of the six campaigns (BJ2014_WIN, BJ2015_AUT, BJ2016_WIN, BJ2017_SUM, XT2016_SPR, and GC2018_WIN) used for validating the model performance (Fig. 3a). 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. 3b), indicating that the observations can represent various atmospheric conditions, spanning from clean to polluted in the region. Fig. 3a shows N_{CCN} at a supersaturation of 0.2% (the typical range of supersaturations in clouds). It shows that, for all the six periods, the time series of N_{CCN} predicted by the RFRM model agrees better with the observed N_{CCN} compared to simulations by the WRF-Chem model. Although both the RFRM model and WRF-Chem exhibit underestimations in observed N_{CCN}, the slope between the predicted and observed N_{CCN}

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

ranged from 0.41 by the WRF-Chem model (with the NMB of -40%, RMSE of 4569 cm^{-3}) to 0.68 by the RFRM model (the NMB of -8% and RMSE of 2045 cm^{-3}) (Fig. 3c), corresponding the R^2 increased from 0.33 to 0.86. Compared to WRF-Chem simulations, the RFRM model showed the greatest improvement during the winter campaigns with the NMB decreased from -45% to -14% (eg., slope increased from 0.39 to 0.67, R^2 from 0.32 to 0.83) when $PM_{2.5}$ concentrations were usually higher. For example, during the GC2018_WIN campaign, the observed N_{CCN} is underestimated as large as 71% by the WRF-Chem (Fig. S6), while the underestimation of observed values by the RFRM has been significantly improved, with an enhancement of approximately 36% (Fig. S5). WRF-Chem simulations for warm seasons noticeably improved (the NMB of -29%), e.g., the uncertainty decreased to 25% during the BJ2017_ SUM campaign (Fig. S5). Overall, the RFRM model still performs better than the WRF-Chem model and is with the NMB of 4% during summer campaigns (eg., slope of 0.67 and R^2 of 0.67). Occasionally, the WRF-Chem model overestimated the N_{CCN} apparently, e.g., the episodes of September 24 to 25 during the BJ2015_AUT campaign, and May 7 to 8 during the XT2017_SPR campaign. In addition, we also evaluated the model performance based on another observation at GC site in January (Zhang et al., 2020) (Fig. S6). And Figure S7 shows comparisons of N_{CCN} at S=0.4%. Both exhibit similarly good performance as shown in Fig. 3. Overall, the RFRM model performs well and can accurately capture the observed fluctuations during these episodes. The improvements in RFRM model also demonstrate the effectiveness of the model trained on atmospheric variables to revise the simulation in model.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

In our case, 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 (Fanourgakis et al., 2019). In fact, the much lower and fixed hygroscopicity parameter value in WRF-Chem model does not represent the hygroscopicity of organics throughout the study period (Liu et al., 2021). Neglecting to distinguish between POA and SOA information during the training of the RFRM model may cause the overestimation of N_{CCN} when POA dominates. Uncertainties incurred by the RFRM model could also originate from the lack of physical interpretability in these ML-based models (Wei et al., 2023). Additional input parameters that carry rich and more meaningful information (e.g., particle number size distribution, aerosol sources and other secondary processes) are expected to further improve the predictability of N_{CCN} in future.

3.3 Performance of the RFRM model in predicting hourly-to-yearly-scale N_{CCN}

To further examine the performance of the RFRM model in predicting N_{CCN} at different time scales, we compare the RFRM model-predicted hourly-to-yearly N_{CCN} in Beijing with both WRF-Chem simulations and observed values (Fig. 4). The RFRM model captures well the diurnal cycle (Fig. 4a), while the WRF-Chem model underestimates N_{CCN}, especially at night. Concerning seasonal variations, similarly, the RFRM model performs better with the NMB of ~ 4% compared to observations. While the mean bias can increase to be ~28% by the WRF-Chem (Fig. 4b). Note that, the bias is much greater in the cold seasons than that in the warm seasons for the WRF-Chem.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

This is probably due to the higher wintertime CN and CCN concentrations which are more difficult for models to capture and simulate (Fanourgakis et al., 2019).

Figure 4c shows the long-term trend of yearly averaged N_{CCN}. Here, the real atmospheric long-term trend of N_{CCN} (denoted as N_{CCN_Obs}) is derived using the long-term measurement of particle number size distribution (PNSD) at a field site in Beijing (Fig. S3) (Fig. 4d, Shang et al., 2022) and the κ calculated from the measured chemical compositions based on the κ-Köhler theory (Petters and Kreidenweis, 2007). The results show that the predicted average annual N_{CCN} at S=0.2% based on the RFRM model agrees well with N_{CCN_Obs} in terms of magnitude and long-term trend (Fig. 4c), showing a decreasing trend year by year with the average annual CCN number concentration of about 6216±3624 cm^{-3} in 2014 and 3278±2306 cm^{-3} in 2018; however, although the WRF-Chem simulations also show a similar decreasing trend year by year, it significantly underestimates the average annual N_{CCN} of all years (with the NMB of 43%), resulting in the inter-annual trend lines being parallel but not coincident. The small bias (within ±6%) between the RFRM model predictions and the observations may be due to the uncertainty from how N_{CCN_Obs} is calculated , i.e., using the Tracking Air Pollution in China (TAP) dataset to calculate κ. A comparison of the values of κ and N_{CCN} between that derived using field observations and the TAP dataset shows little differences (Fig. S9); actually, the long-term change of N_{CCN} is much less sensitive to changes in κ values compared to the PNSD (Fig. S9c). Although absolute concentrations of components in the TAP dataset deviate from observations, their mass fractions are consistent (Fig. S9d), rendering the impact on the calculated κ negligible.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

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).

According to Fig. 4d-e, the long-term decreasing trend of N_{CCN} at S=0.2% from 2014 to 2018 is mainly attributed to a significant reduction in aerosol particle number concentrations in the atmosphere. In addition, the peak diameter of the PNSD shows a shift toward the left, decreasing slightly from about 70 nm in 2014 to 30 nm in 2018 due to the enhanced new particle formation events in recent years (Zhu et al., 2021). This would also result in less aerosol particles serving as CCN. Although the κ_{chem} has a slight upward trend from 2014 to 2018 (Fig. 4e), yielding decreases in activation diameter and thereby more CCN, the aerosol particle hygroscopicity, however, plays less significant role in regulating the long-term total N_{CCN} variations compared to the changes in particle number size distribution during the period.

3.4 Spatial variations of N_{CCN} derived by the RFRM model

We further examine the spatiotemporal changes of N_{CCN} at S=0.2% in the NCP derived by the RFRM model (Fig. 5). Regionally, the N_{CCN} predicted by the RFRM model is also generally higher than that simulated by WRF-Chem at most of the sites. The N_{CCN} derived by the RFRM model and WRF-Chem both decrease from 2014 to 2018 but with different decreasing rates (Fig. 5c-e). On average, N_{CCN} derived by the RFRM model and WRF-Chem decrease from 4415±643 to 2910±789 cm^{-3} and from 2834±1366 to 2111±546 cm^{-3} respectively from 2014 to 2018 in the NCP region (Fig. 5c), corresponding to annual decreasing rates of approximately ~ -250 cm^{-3} yr^{-1} for the RFRM model and ~ -167 cm^{-3} yr^{-1} for the WRF-Chem model (Fig. 5d-e).

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

Moreover, N_{CCN} and its changes from 2014 to 2018 predicted by the RFRM model show more significant spatial variations than that simulated by the WRF-Chem model. Differences in RFRM -model-predicted N_{CCN} between 2014 and 2018 (2018 minus 2014) show negative values at ~90% of the sites, i.e., downward trends in N_{CCN} (Fig. 5c1). The sites with apparent N_{CCN} reduction are mainly located in the central and northern of NCP, especially in Beijing-Tianjin-Hebei (BTH) and central Shandong. Sites in southern NCP have slight downward trends in N_{CCN}. The downward trend is consistent with the variations in concentration of gaseous pollutants due to the emission reduction in past years in China (Fig. S10). Interestingly, note a few sites with positive values (upward trends in N_{CCN}) are mainly located along the coast. An increase in the fraction of accumulation-mode particles in coastal areas has been reported contributing more CCN (Zhu et al., 2021). This demonstrates the good performance of the RFRM model in capturing the real-time spatial variations of CCN on a regional scale. While the WRF-Chem model might mask the variations of N_{CCN} among different sites. This will smooth out the true impact of aerosols on weather and climate at regional or local scales, leading to uncertainties in model simulations.

3.5 Sensitivity of the cloud parameters and radiative forcing to CCN prediction biases

To evaluate the effects introduced by N_{CCN} prediction biases to the aerosol indirect effects, we further incorporate the deviations between observed N_{CCN} (denoted as CCN_{OBS}) and N_{CCN} predicted by the RFRM model (denoted as CCN_{RFRM}) and the simulated by the WRF-Chem model (denoted as $CCN_{WRF\text{-}Chem}$) into calculations of the

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

cloud parameters and radiative forcing, as are shown in Fig. 6 (for S =0.2%) and Fig. S11 (for S=0.4%). Typically, aerosol particles serving as CCN could indirectly affect the global climate by the Twomey (Twomey, 1977) and Albrecht effects (Albrecht, 1989). According to Wang et al. (2019), two parameters of cloud optical thickness (τ) and the absorption coefficient (1-ω_0) can be used to estimate the Twomey effects. The process of cloud-to-rain conversion, which can be parameterized by the critical radius (r_c) and the cloud-to-rain conversion threshold function (TA), is critical to estimate the Albrecht effect. Therefore, the r_c and TA is also calculated here. Indirect (cloud) radiative forcing (F_c) is also evaluated based on the deviations in CCN number concentration under the assumption of a constant liquid water content (Charlson, 1992; Wang et al., 2008). Section S2 provides details about the methods used to evaluate aerosol indirect effects.

In general, the results show that these cloud properties are more sensitive to the changes in N_{CCN} when the models underestimate the CCN number concentrations (ΔN_{CCN}<0) compared to the cases with an overestimation (Figure. 6a-d). For example, a ~50% underestimation (overestimation) of N_{CCN} could lead to relative deviations (uncertainties) of -21% (14%) for τ, 27% (-12%) for (1-ω_0), and -11% (7%) for r_c at S=0.2%. Note that, on average, both the RFRM model and WRF-Chem in this study show underestimations in N_{CCN} within the sensitivity zone of the cloud effect (Fig. 6), It is thus expected to cause large uncertainties in evaluating the cloud radiative forcing, a topic worthy of further attention. Given that the uncertainty in N_{CCN} predicted by the WRF-Chem model is much greater than that of RFRM model, the uncertainties and

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

415 variation ranges of these cloud parameters from WRF-Chem simulations are also
416 greater. Specifically, the uncertainties of CCN_{RFRM} and $CCN_{WRF\text{-}Chem}$ lead to the
417 uncertainties of -33% to +78% and -77% to +92% respectively, for the τ (Fig. 6a and
418 a1), -44% to +50% and -48% to +344% respectively, for the $1\text{-}\omega_0$ (Fig. 6b and b1), -
419 18% to +34% and -53% to +38% respectively, for the r_c (Fig. 6c and c1), and -118% to
420 +94% and -258% to +353% respectively, for the *TA* (Fig. 6d and d1).

421 In addition, the underestimation of CCN would lead to underestimations of cloud
422 optical thickness τ and the critical radius r_c of the automatic cloud/rain transformation,
423 but overestimations of $(1\text{-}\omega_0)$ and the threshold function *TA* of the automatic cloud/rain
424 transformation, all of which depend on their physical mechanisms within the realm of
425 aerosol-cloud interactions (Stier et al., 2024) (Fig. S12). This is also the case at the other
426 supersaturation levels considered (Fig. S11).

427 As a result, we derive that the mean underestimation of ~59±19% in N_{CCN} at S=0.2%
428 caused by the WRF-Chem leads to underestimations of 26±11% in the τ, 14±7% in the
429 r_c, and an overestimation of 35±191% in the absorption coefficient $(1\text{-}\omega_0)$ and 93±42%
430 in the *TA*. While, the mean uncertainties for all these parameters are largely reduced
431 when the mean underestimation of ~31±15% in N_{CCN} at S=0.2% that is caused by
432 RFRM model is applied (Fig. 6e). For example, the underestimation of cloud optical
433 thickness τ decreases to ~12%, an improvement compared to the underestimation of
434 about 14% by the WRF-Chem model. Also, the RFRM model reduces the
435 underestimation of the critical radius r_c of the automatic cloud/rain transformation to
436 only ~6%. Ultimately, the uncertainty of cloud radiative forcing F_c has been

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

significantly reduced from an overestimation of 1.89±0.78 W m^{-2} by the WRF-Chem model to only 0.81±0.63 W m^{-2} by the RFRM model, showing the high sensitivity of climate forcing to the uncertainties in CCN number concentrations. Note that a limitation when evaluating the cloud radiative forcing based on the assumption of cloud fraction and the fractional transmission is the approximate analytical expression. Therefore, the results presented here may represent the upper limit, and the sensitivity of the radiative forcing to changes in N_{CCN} would be weaker over continental areas (Wang et al., 2008; Yu et al., 2022).

4. Summary

4.1 Discussion and conclusions

In this study, using a multisource dataset of ground-based observations, atmospheric variables, the N_{CCN} simulations by the WRF-Chem model, we have developed a new machine-learning-based model that predicts well the regional-scale N_{CCN} in the polluted NCP region. 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. In general, the RFRM model better captures the spatial differences of N_{CCN} than the WRF-Chem model. In addition, the prediction reveals a long-term downward trend of N_{CCN} coincident with the observed trend for the period of 2014–2018. By further incorporating the N_{CCN} prediction biases into the evaluation of cloud parameters and radiative forcing, we found that the cloud properties are more sensitive to the changes

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

in N_{CCN} when the models underestimated the CCN number concentrations compared to the cases when the models overestimated N_{CCN}. As a result, the simulated uncertainty of cloud radiative forcing F_c could be significantly reduced from an overestimation of 1.89±0.78 W m^{-2} by the WRF-Chem model to 0.81±0.63 W m^{-2} by the RFRM model. 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.

4.2 Limitations and outlook

Although the machine learning-based model established in this study has been shown to significantly improve the prediction of CCN number concentration, the prediction results still exhibit considerable deviations, which may be attributed to several factors. First, it is related to the inherent limitation of the random forest method in describing time series with extreme values and large short-term fluctuations. In the future, other advanced machine learning algorithms (e.g., Long Short-Term Memory, Transformer) can be integrated to optimize and improve the results. Second, in this study, the observational data only from six campaigns at three sites are analyzed. 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. Overall, the RFRM framework presented here relies on available atmospheric state variables (e. g., chemical compositions, gas

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

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. This new modeling framework could also guide the development of machine learning based models to predict CCN vertical profiles, which are critical for the accurate evaluation of the ACI effect.

Code and Data availability

The data and code are publicly accessible at https://zenodo.org/records/18932004 (Ren et al., 2026). This includes the WRF-Chem model version 4.1.5 used in this study, the machine learning code, the corresponding training, testing datasets and the CCN observation datasets, the emissions inventory and scripts used in WRF-Chem and the scripts used for plotting, supporting the findings of this study. The release version of WRF-Chem is also open-access and can be publicly available at NCAR https://www2.mmm.ucar.edu/wrf/users/download/get_source.html (Skamarock et al., 2019, last access: 10 May 2025). The initial meteorological variables are from the National Center for Environmental Prediction's Final Operational Global (NCEP/FNL) and available at https://doi.org/10.5065/D6M043C6 (NCEP, 2000).

Supplement

The Supplement contains the information of additional descriptions of the WRF-Chem simulation (parameterization scheme, emission inventory and initial and boundary

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

conditions), the method to evaluate aerosol indirect effects and additional figures applied in this study.

Author contributions

JYR: Conceptualization, Investigation, Formal analysis, Writing – original draft. SJZ: Data curation. HHX and GQL: Formal analysis. ZW: Formal analysis. ARZ: Data curation. CFZ: Technical support. MH: Data curation. DJS: Data curation. LZT: Data curation. RJH: Funding acquisition. YLS: Technical support. FZ: Writing-Reviewing and Editing, Funding acquisition. All authors read and approved the final manuscript.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

This work was funded by the National Natural Science Foundation of China (NSFC) research project (Grant No. 42405118, 42475112, 41975174), the State Key Laboratory of Loess Science, the Institute of Earth Environment, Chinese Academy of Sciences (SKLLQG2429), Shenzhen Science and Technology Plan Project (Grant No. KCXST20221021111404011; GXWD20220818172959001; SYSPG20241211173609007), Guangdong Natural Science Foundation (Grant No. 2024A1515011005). We thank all participants in the field campaigns for their tireless work and cooperation. We also sincerely thank the three anonymous reviewers for their

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

valuable comments and suggestions, which have greatly improved the quality of this paper.

References

Albrecht, B.: Aerosols, cloud microphysics, and fractional cloudiness, Science, 245(4923), 1227–1230, https://doi.org/10.1126/science.245.4923.1227, 1989.

Charlson, R., Lovelock, J., Andreae, M., Warren, S.: Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate, Nature, 326(6114), 655–661, https://doi.org/10.1038/326655a0, 1987.

Chen, L., Zhang, F., Zhang, D. et al.: Measurement report: Hygroscopic growth of ambient fine particles measured at five sites in China, Atmos. Chem. Phys., 22(10), 6773–6786, https://doi.org/10.5194/acp-22-6773-2022, 2022.

Charlson, R. et al.: Climate Forcing by Anthropogenic Aerosols, Science, 255, 423–430, https://doi.org/ 10.1126/science.255.5043.423, 1992.

Dinar, E., Anttlia, T., Rudich, Y.: CCN activity and hygroscopic growth of organic aerosols following reactive uptake of ammonia, Environmental Science & Technology, 42(3), 912-918, https://doi.org/10.1021/es071874p, 2008.

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.

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.

Fan, T., Ren, J., Liu, C., Li, Z., Liu, J., Sun, Y., et al.: Evidence of surface‐tension lowering of atmospheric aerosols by organics from field observations in an urban atmosphere: Relation to particle size and chemical composition, Environmental

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

Science & Technology, 58(26), 11363 – 11375. https://doi.org/10.1021/acs.est.4c03141, 2024.

Grange, S., Carslaw, D., Lewis, A. et al.: Random Forest meteorological normalisation models for Swiss PM_{10} trend analysis, Atmos. Chem. Phys., 18(9), 6223–6239, https://doi.org/10.5194/acp-18-6223-2018, 2018.

Geng, G., Xiao, Q., Liu, S. et al.: Tracking air pollution in China: near real-time PM2.5 retrievals from multisource data fusion, Environmental Science & Technology, 55(17), 12106–12115, https://doi.org/10.1021/acs.est.1c01863, 2021.

Guo, J., Miao, Y., Zhang, Y. et al.: The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data, Atmos. Chem. Phys., 16(20), 13309–13319, https://doi.org/10.5194/acp-16-13309-2016, 2016.

IPCC. Summary for Policymakers. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA. (2021).

Liu, S., Geng, G., Xiao, Q., Zheng, Y., Liu, X., Cheng, J., & Zhang, Q.: Tracking daily concentrations of $PM_{2.5}$ chemical composition in China since 2000, Environ Sci Technol, 56, 16517–16527, https://doi.org/10.1021/acs.est.2c06510, 2022.

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.

Liu, C., Wang, T., Rosenfeld, D. et al.: Anthropogenic effects on cloud condensation nuclei distribution and rain initiation in East Asia, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL086184, 2020.

Liu, J., Li, Z.: Significant underestimation in the optically based estimation of the aerosol first indirect effect induced by the aerosol swelling effect, Geophysical Research Letters, 45(11), 5690–5699, https://doi.org/10.1029/2018GL077679, 2018.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

574 Liang, M., Tao, J., Ma, N. et al.: Prediction of CCN spectra parameters in the North
575 China Plain using a random forest model, Atmospheric Environment, 289, 119323,
576 https://doi.org/10.1016/j.atmosenv.2022.119323, 2022.
577 Li, S., Zhang, F., Jin, X. et al.: Characterizing the ratio of nitrate to sulfate in ambient
578 fine particles of urban Beijing during 2018-2019, Atmospheric Environment, 237,
579 https://doi.org/10.1016/j.atmosenv.2020.117662, 2020.
580 Liu, J., Zhang, F., Xu, W. et al.: Hygroscopicity of organic aerosols linked to formation
581 mechanisms, Geophysical Research Letters, 48,
582 https://doi.org/10.1029/2020GL091683, 2021.
583 Lin, Y., Farley, R., Orville, H.: Bulk parameterization of the snow field in a cloud model,
584 Journal of Applied Meteorology and Climatology, 22(6): 1065–1092,
585 https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2, 1983.
586 Menon & Rotstayn.: The radiative influence of aerosol effects on liquid-phase cumulus
587 and stratiform clouds based on sensitivity studies with two climate models,
588 Climate Dynamics, 27, 4, 345–356, https://doi.org/10.1007/s00382-006-0139-3,
589 2006.
590 Moore, R., Cerully, K., Bahreini, R. et al.: Hygroscopicity and composition of
591 California CCN during summer 2010, Journal of Geophysical Research:
592 Atmospheres, 117, https://doi.org/10.1029/2011JD017352, 2012.
593 Morrison, H., Thompson, G., Tatarskii, V.: Impact of cloud microphysics on the
594 development of trailing stratiform precipitation in a simulated squall line:
595 Comparison of one-and two-moment schemes, Monthly Weather Review, 137(3),
596 991–1007, https://doi.org/10.1175/2008MWR2556.1, 2009.
597 Nair, A. A., Yu, F.: Using machine learning to derive cloud condensation nuclei number
598 concentrations from commonly available measurements, Atmos. Chem. Phys.,
599 20(21), 12853–12869, https://doi.org/10.5194/acp-20-12853-2020, 2020.
600 Nair, A. A., Yu, F., Campuzano-Jost, P. et al.: Machine learning uncovers aerosol size
601 information from chemistry and meteorology to quantify potential cloud-forming

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

particles, Geophysical Research Letters, 48, https://doi.org/10.1029/2021GL094133, 2021.

NCEP: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999, National Centers for Environ mental Prediction [Data set], https://doi.org/10.5065/D6M043C6, 2000 (last access: 10 May 2025).

Petters, M. D., Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmos. Chem. Phys., 7(8), 1961–1971, https://doi.org/10.5194/acp-7-1961-2007, 2007.

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.

Ren, J., Chen, L., Liu, J., and Zhang, F.: The density of ambient black carbon retrieved by a new method: implications for cloud condensation nuclei prediction, Atmos. Chem. Phys., 23, 4327–4342, https://doi.org/10.5194/acp-23-4327-2023, 2023.

Rose, C., Collaud Coen, M. et al.: Seasonality of the particle number concentration and size distribution: a global analysis retrieved from the network of Global Atmosphere Watch (GAW) near-surface observatories, Atmos. Chem. Phys., 21, 17185–17223, https://doi.org/10.5194/acp-21-17185-2021, 2021.

Rosenfeld, D., Zheng, Y., Hashimshoni, E. et al.: Satellite retrieval of cloud condensation nuclei concentrations by using clouds as CCN chambers, Proceedings of the National Academy of Sciences, 113(21), 5828–5834, https://doi.org/10.1073/pnas.1514044113, 2016.

Ren, J., Zou, S., Xu, H., Liu, G., Wang, Z., Zhang, A., Zhao, C., Hu, M., Shang, D., Tang, L., Huang, R.-J., Sun, Y., & Zhang, F.: Machine learning significantly improves the simulation of hourly-to-yearly scale cloud nuclei concentration and radiative forcing in polluted atmosphere [Data set]. Zenodo. https://zenodo.org/records/18932004, 2026.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

630 Stockwell, W. R., Middleton, P., Chang, J. S., et al.: The second generation regional
631 acid deposition model chemical mechanism for regional air quality modeling,
632 Journal of Geophysical Research: Atmospheres, 95(D10): 16343–16367,
633 https://doi.org/10.1029/JD095iD10p16343, 1990.
634 Sotiropoulou, R. E. P., Nenes, A., Adams, P. J. et al.: Cloud condensation nuclei
635 prediction error from application of Köhler theory: Importance for the aerosol
636 indirect effect, Journal of Geophysical Research: Atmospheres, 112(D12)
637 https://doi.org/10.1029/2006JD007834, 2007.
638 Shen, Y., Virkkula, A., Ding, A. et al.: Estimating cloud condensation nuclei number
639 concentrations using aerosol optical properties: role of particle number size
640 distribution and parameterization, Atmos. Chem. Phys., 19(24), 15483–15502,
641 https://doi.org/10.5194/acp-19-15483-2019, 2019.
642 Sha, T., Ma, X.Y., Jia, H. L. et al.: Exploring the influence of two inventories on
643 simulated air pollutants during winter over the Yangtze River Delta, Atmospheric
644 Environment, 206, 170–182, https://doi.org/10.1016/j.atmosenv.2019.03.006,
645 2019.
646 Schmale, J., Henning, S., Decesari, S. et al.: Long-term cloud condensation nuclei
647 number concentration, particle number size distribution and chemical composition
648 measurements at regionally representative observatories, Atmos. Chem. Phys.,
649 18(4), 2853–2881, https://doi.org/10.5194/acp-18-2853-2018, 2018.
650 Song, C., Becagli, S., Beddows, D. C. S. et al.: Understanding Sources and Drivers of
651 Size-Resolved Aerosol in the High Arctic Islands of Svalbard Using a Receptor
652 Model Coupled with Machine Learning, Environmental Science & Technology,
653 56(16), 11189–11198, https://doi.org/10.1021/acs.est.1c07796, 2022.
654 Shang, D., Tang, L., Fang, X. et al.: Variations in source contributions of particle
655 number concentration under long-term emission control in winter of urban Beijing,
656 Environmental Pollution, 304, 119072,
657 https://doi.org/10.1016/j.envpol.2022.119072, 2022.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

658 Stier, P., van, den, Heever, S, C., Christensen, M. W. et al.: Multifaceted aerosol effects
659 on precipitation, Nature Geoscience, 17(8), 719–732,
660 https://doi.org/10.1038/s41561-024-01482-6, 2024.
661 Skamarock, W., Klemp, J., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers,
662 J. G., Duda, M. G., Barker, D., and Huang, X.-Y.: A Description of the Advanced
663 Research WRF Model Version 4.1, UCAR/NCAR, https://doi.org/10.5065/1dfh-
664 6p97, 2019 (code available at https://www2.mmm.ucar.edu/wrf/users/download/
665 get_source.html, last access: 10 May 2025).
666 Tao, J., Kuang, Y., Ma, N., Hong, J., Sun, Y., Xu, W., Zhang, Y., He, Y., Luo, Q., Xie,
667 L., Su, H., and Cheng, Y.: Secondary aerosol formation alters CCN activity in the
668 North China Plain, Atmos. Chem. Phys., 21, 7409–7427,
669 https://doi.org/10.5194/acp-21-7409-2021, 2021.
670 Twomey, S.: The influence of pollution on the shortwave albedo of clouds, Journal of
671 the atmospheric sciences, 34(7), 1149–1152, https://doi.org/10.1175/1520-
672 0469(1977)034<1149:TIOPOT>2.0.CO;2, 1977.
673 Wei, J., Li, Z., Wang, J. et al.: Ground-level gaseous pollutants (NO_2, SO_2, and CO) in
674 China: daily seamless mapping and spatiotemporal variations, Atmos. Chem.
675 Phys., 23, 1511–1532, https://doi.org/10.5194/acp-23-1511-2023, 2023.
676 Wang, Y., Niu, S., Lv, J. et al.: A new method for distinguishing unactivated particles in
677 cloud condensation nuclei measurements: implications for aerosol indirect effect
678 evaluation, Geophysical Research Letters, 46, 14,185–14,194,
679 https://doi.org/10.1029/2019GL085379, 2019.
680 Wang, J., Lee, Y.-N., Daum, P. H., Jayne, J., and Alexander, M. L.: Effects of aerosol
681 organics on cloud condensation nucleus (CCN) concentration and first indirect
682 aerosol effect, Atmos. Chem. Phys., 8, 6325–6339, https://doi.org/10.5194/acp-8-
683 6325-2008, 2008.
684 Xu, W., Fossum, K. N., Ovadnevaite, J. et al.: The impact of aerosol size-dependent
685 hygroscopicity and mixing state on the cloud condensation nuclei potential over

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

the north-east Atlantic, Atmos. Chem. Phys., 21, 8655–8675, https://doi.org/10.5194/acp-21-8655-2021, 2021.

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.

Yu, F., Luo, G., Nair, A. A., Tsigaridis, K., & Bauer, S. E.: Use of machine learning to reduce uncertainties in particle number concentration and aerosol indirect radiative forcing predicted by climate models, Geophysical Research Letters, 49, https://doi.org/10.1029/2022GL098551, 2022.

Yu, F., Luo, G., Nair, A. A. et al.: Wintertime new particle formation and its contribution to cloud condensation nuclei in the Northeastern United States, Atmos. Chem. Phys., 20(4), 2591–2601, https://doi.org/10.5194/acp-20-2591-2020, 2020.

Zhu, Y., Shen, Y., Li, K. et al.: Investigation of particle number concentrations and new particle formation with largely reduced air pollutant emissions at a coastal semi-urban site in northern China, Journal of Geophysical Research: Atmospheres, 126, e2021JD035419, https://doi.org/10.1029/2021JD035419, 2021.

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.

Zhang, F., Li, Y., Li, Z. et al.: Aerosol hygroscopicity and cloud condensation nuclei activity during the AC3Exp campaign: Implications for cloud condensation nuclei parameterization, Atmos. Chem. Phys., 14(24), 13423–13437, https://doi.org/10.5194/acp-14-13423-2014, 2014.

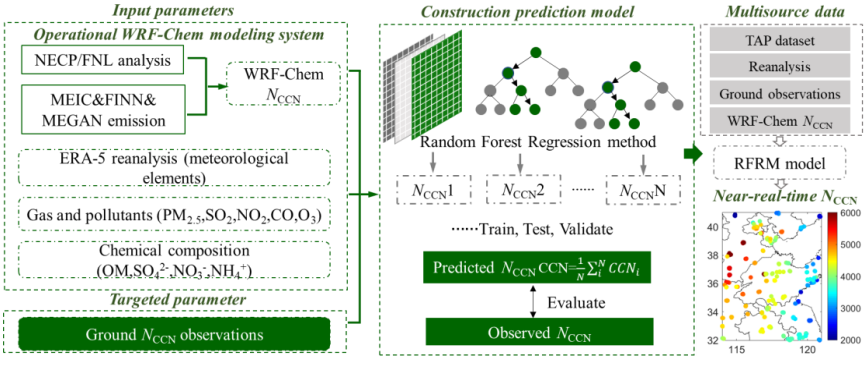
Zhang, F., Li, Z., Li, Y. et al.: Impacts of organic aerosols and its oxidation level on CCN activity from measurement at a suburban site in China, Atmos. Chem. Phys., 16(8), 5413–5425, https://doi.org/10.5194/acp-16-5413-2016, 2016.

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

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

716 Beijing, Journal of Geophysical Research: Atmospheres, 124(24), 14102–14113,
717 https://doi.org/10.1029/2019JD031457, 2019.
718 Zhang, F., Wang, Y., Peng, J. et al.: Uncertainty in predicting CCN activity of aged and
719 primary aerosols, Journal of Geophysical Research: Atmospheres, 122(21),
720 11723–11736, https://doi.org/10.1002/2017JD027058, 2017.
721 Zaveri, R. A., Peters, L. K.: A new lumped structure photochemical mechanism for
722 large-scale applications, Journal of Geophysical Research: Atmospheres, 104,
723 30387–30415, https://doi.org/10.1029/1999JD900876, 1999.
724 Zheng, B., Tong, D., Li, M., Liu, F. et al.: Trends in China's anthropogenic emissions
725 since 2010 as the consequence of clean air actions, Atmos. Chem. Phys., 18,
726 14095-14111, https://doi.org/10.5194/acp-18-14095-2018, 2018.

727 ## Figures

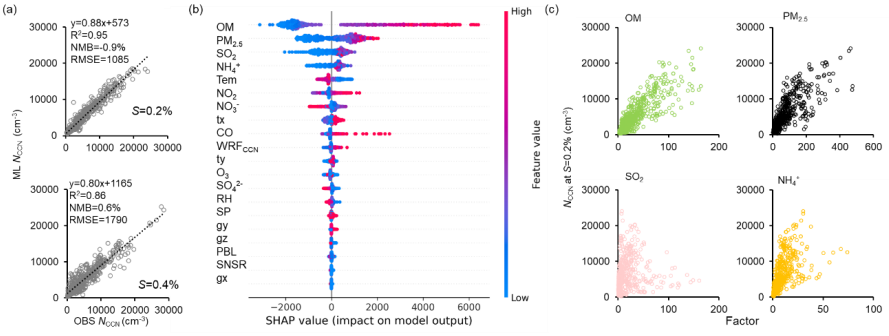


728

729 **Fig. 1** Methodological framework of CCN number concentration prediction.

730

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026



731

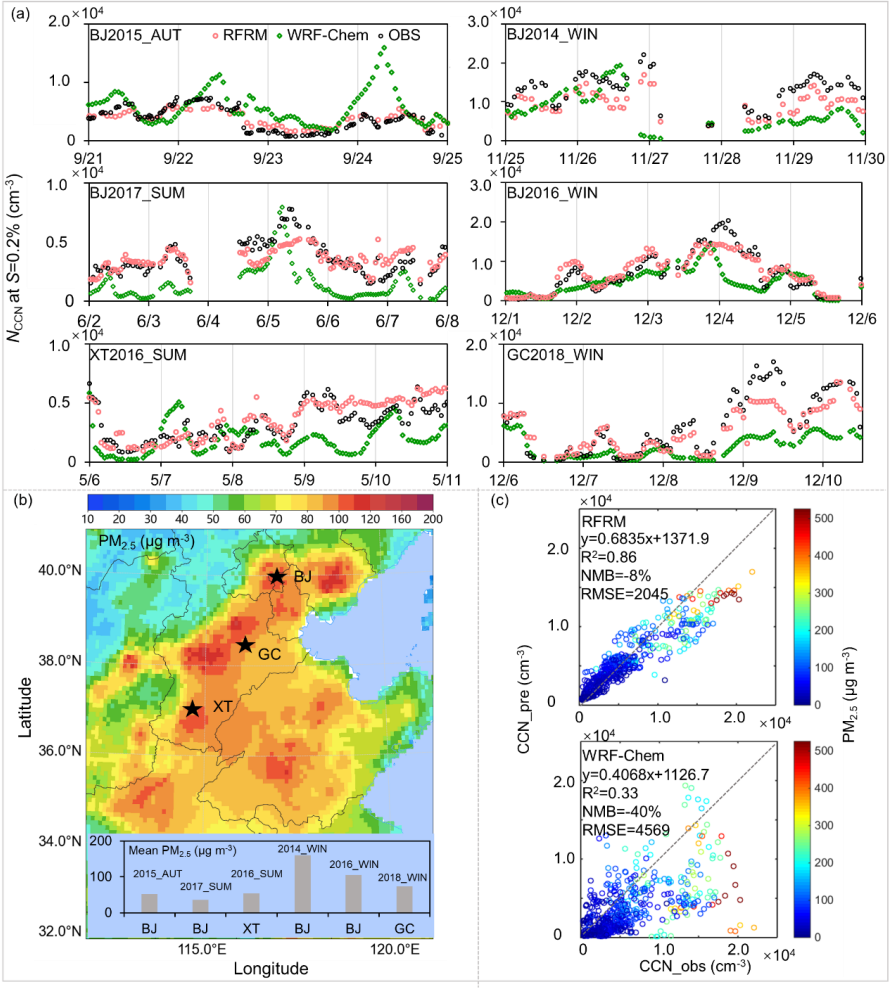
732 **Fig. 2** Comparison of RFRM model retrieval and observed N_{CCN} at S=0.2% and 0.4%

733 for the testing dataset (a), The SHAP value of the input parameters to the prediction of

734 N_{CCN} (b), Scatter plots of the input parameters (OM, $PM_{2.5}$, SO_2, NH_4^+) with CCN

735 number concentration at S=0.2% (c).

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

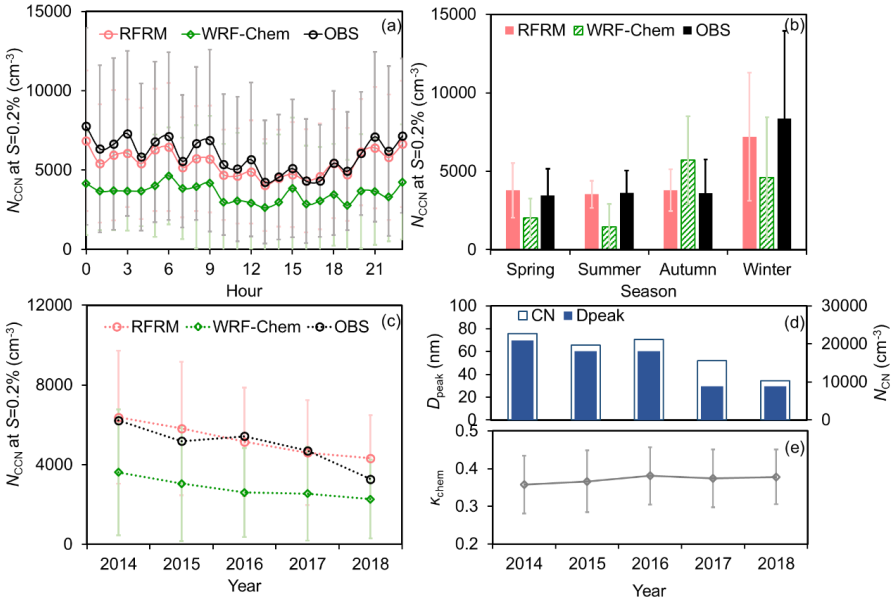


736

737 **Fig. 3** Performance of the RFRM model in predicting N_{CCN} at field sites in NCP. (a)
738 Time series of the observed and predicted CCN number concentrations at S=0.2% for
739 the six periods (BJ2015_AUT, BJ2017_SUM, XT2016_SUM, BJ2014_WIN,
740 BJ2016_WIN, GC2018_WIN) in the North China Plain; (b) Map for average mass
741 concentration of $PM_{2.5}$ of 2014 from TAP dataset in NCP (http://tapdata.org.cn/) and
742 field observed average mass concentration of $PM_{2.5}$ during the six field campaigns (see

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

743 embedded histogram); (c) Scatter plots of the observed N_{CCN} at S=0.2% with the RFRM
744 model predicted (top) and WRF-Chem simulated (bottom) respectively.



745

746 **Fig. 4** Performance of the RFRM model in predicting hourly-to-yearly scale N_{CCN}. (a)
747 Diurnal variations of N_{CCN} at S=0.2% derived from the RFRM model, the WRF-Chem,
748 and the observations from the field campaigns; (b) Seasonal variations, here the
749 comparison in spring, summer, autumn and winter were conducted using the campaign
750 averages of XT2016_SPR, BJ2017_SUM, BJ2015_AUT, and
751 BJ2014_WIN&BJ2016_WIN respectively with the RFRM model and WRF-Chem
752 predictions at corresponding periods; (c) Trends of annual mean N_{CCN} from 2014 to
753 2018; (d) Trends of annual mean particle number concentration and peak diameter; (e)
754 Trends of annual mean of the hygroscopic parameter κ_{chem} calculated from TAP dataset
755 in Beijing.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

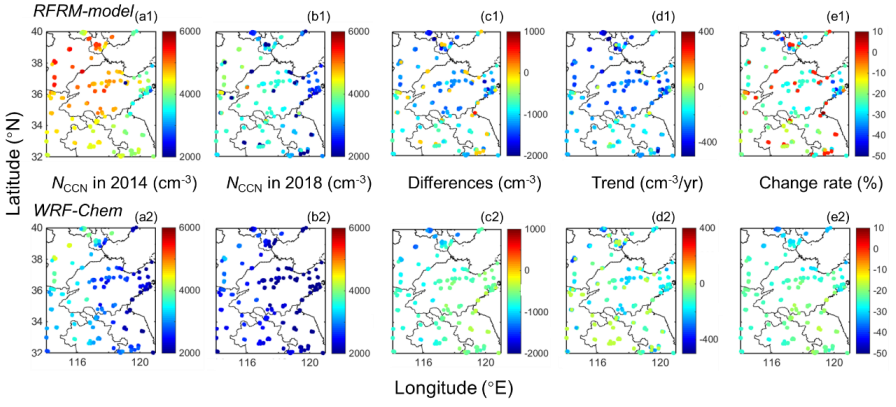


Fig. 5 Spatial variations of N_{CCN} derived by the RFRM model (top) and WRF-Chem (bottom) at the sites in the studied region. (a1) Average N_{CCN} at S=0.2% in 2014 predicted by the RFRM model; (a2) Average N_{CCN} at S=0.2% in 2014 by the WRF-Chem; (b1 and b2) Same as a1 and a2 but in 2018; (c1) Differences in N_{CCN} at S=0.2% between the year of 2014 and 2018 predicted by the RFRM model; (c2) Same as (c1) but by the WRF-Chem; (d1) Trends of N_{CCN} at S=0.2% from 2014 to 2018 predicted by the RFRM model ; (d2) Same as (d1) but by the WRF-Chem; (e1) Change rates of N_{CCN} at S=0.2% from 2014 to 2018 predicted by the RFRM model; (e2) Same as (e1) but by the WRF-Chem.

https://doi.org/10.5194/egusphere-2026-1347
Preprint. Discussion started: 21 April 2026

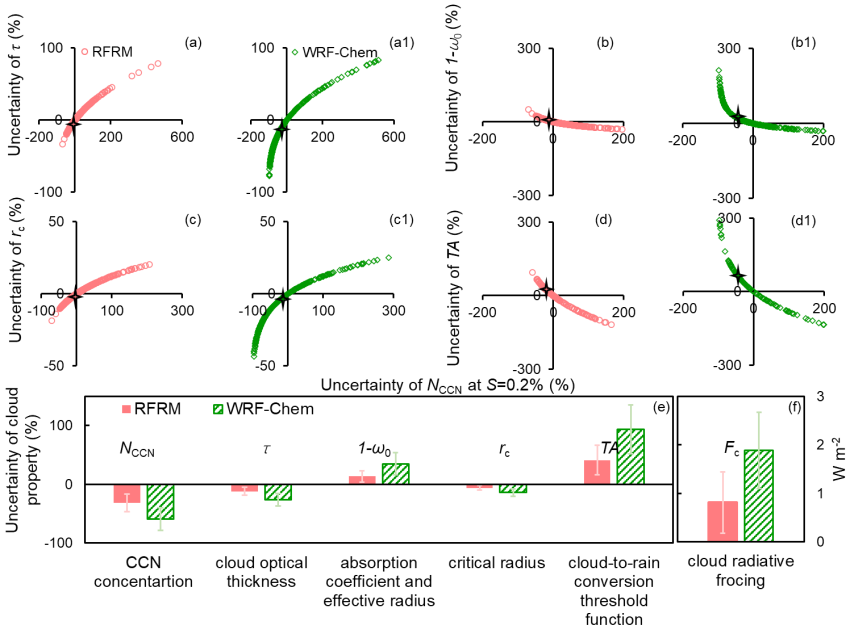


Fig. 6 Sensitivity of the cloud parameters and radiative forcing to CCN prediction biases. (a) Dependence of the uncertainty of the cloud optical thickness (τ) on the uncertainty of N_{CCN} at S=0.2% with the RFRM model; (a1) Same as (a) but by the WRF-Chem; (b) Dependence of the uncertainty of the absorption coefficient (1-ω_0) on the uncertainty of N_{CCN} at S=0.2% with the RFRM model; (b1) Same as (b) but by the by the WRF-Chem; (c) Dependence of the uncertainty of the critical radius (r_c) on the uncertainty of N_{CCN} at S=0.2% with the RFRM model; (c1) Same as (c) but by the WRF-Chem; (d) Dependence of the uncertainty of the cloud-to-rain conversion threshold function (TA) on the uncertainty of N_{CCN} at S=0.2% with the RFRM model; (d1) Same as (d) but by the WRF-Chem; (e) Mean uncertainty in simulating the cloud properties and (f) radiative forcing (F_c) by the RFRM model and the WRF-Chem; Black star shows the mean value for the observation.