Supplement of Geosci. Model Dev., 19, 6403–6416, 2026
https://doi.org/10.5194/gmd-19-6403-2026-supplement

Geoscientific Model Development
Open Access

Supplement of

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

Jingye Ren et al.

Correspondence to: Fang Zhang (zhangfang2021@hit.edu.cn)

S1. Data and Methods

S1.1. WRF-Chem model

In this study, WRF-Chem v 4.1.5 is used to simulate CCN concentration for training the machine learning model. The physical parameterizations employed in this study include Morrison 2-monment scheme (Morrison et al., 2009), RRTMG short-wave and long-wave radiation scheme (Iacono et al., 2008), Grell 3-D cumulus scheme (Grell et al., 2002), Yonsei University boundary layer scheme and Noah land surface model (Hong et al., 2006). The Carbon Bond Mechanism Z photochemical mechanism and 4-bin sectional Model for Simulating Aerosol Interactions and Chemistry (MOSAIC) aerosol model with aqueous chemistry are chosen in this study (Zaveri et al., 1999, 2008). The initial meteorological variables are provided from the National Center for Environmental Prediction's Final Operational Global (NCEP/FNL) datasets. The horizontal resolution of the reanalysis data was 0.25°×0.25°, and the temporal resolution was 6 h. The initial and boundary chemical conditions are from the Community Atmosphere Model with Chemistry model (Buchholz et al., 2019). The anthropogenic emissions are from Multi resolution Emission Inventory for China with a horizontal resolution of 0.25°×0.25° and a temporal resolution of monthly data (Zheng et al., 2018), corresponding to the years 2014-2018. The biological and biomass emissions are taken from the Model of Emissions of Gases and Aerosols from Nature and the Fire Inventory from NCAR, respectively.

S1.2. Evaluation of the aerosol indirect effects

Atmospheric particles acting as cloud condensation nuclei could alter the cloud-rain interactions and further have impact on the climate change. The methods for evaluating the cloud and radiative properties caused by the deviations of CCN concentrations output from WRF-Chem simulation and random forest technique was refereed from Wang et al. 2019 and Wang et al. 2008.

As for the first indirect effect, the cloud optical thickness (τ) and single scatter albedo coefficient (ω_0) are the crucial estimators and can be calculated as:

$$\tau \approx \frac{3}{2} W {r_e}^{-1} \quad \text{(S1)}$$

$$1 - \omega_0 = 1.7 k_w r_e \quad \text{(S2)}$$

where W means the liquid water path, k_w is the complex part of the refractive index of water, r_e is the effective radius of cloud droplets and can be expressed by:

$$r_e = \beta (\frac{3LWC}{4\pi \rho_w N_C})^{1/3} \quad \text{(S3)}$$

where LWC is the liquid water content in cloud, β is the scaling factor. Here, we mainly focused on the response of cloud droplet effective radius to the change of cloud droplet number concentration N_c, so β is assumed to be constant (Quaas et al., 2004). Cloud number concentration can be calculated based on the N_{CCN}-SS parameterizations in many climate models (Fan et al., 2012). Note that the power relationship between the r_e and N_c referred from the previous observations is only an approximate relationship (Garrett et al., 2004).

For evaluating second aerosol indirect forcing, it is critical to estimate the relevant variables associated with the cloud-rain process. The parameterizations of the cloud-to-rain conversion process can be expressed by conversion threshold function (*TA*):

$$TA = \left[\frac{\int_{r_c}^{\infty} r^6 n(r)dr}{\int_0^{\infty} r^6 n(r)dr}\right]\left[\frac{\int_{r_c}^{\infty} r^3 n(r)dr}{\int_0^{\infty} r^3 n(r)dr}\right] \quad \text{(S4)}$$

where r is the cloud droplet radius, $n(r)$ is the size distribution of the cloud droplet (here refereed from Wang et al. 2019 to evaluate). According to the Liu et al. (2004), the critical radius of the auto-conversion process r_c can be expressed by

$$r_c \approx 4.09 \times 10^{-4} \beta_{con}^{1/6} \frac{{N_C}^{1/6}}{LWC^{1/3}} \quad \text{(S5)}$$

where β_{con} is a fixed parameter of 1.15×10^{23} s^{-1} (Liu et al., 2004). Here, the focus is mainly on the effect of the ΔN_{CCN} between WRF-Chem and RF model on the aerosol indirect effects in our research.

We also evaluate the perturbation of cloud radiative forcing due to the uncertainty in CCN concentration estimated from machine learning (RFRM) and numerical model (WRF-Chem). Here under the assumption of the fixed liquid water volume, an approximate analytical expression for estimating the cloud radiative of the change in N_{CCN} as follows (Charlson et al, 1992; Wang et al., 2008):

$$\Delta F_C \approx -\frac{F_T}{4} A_{mst} T_r^2 \Delta R_C \approx -2.2 \times \Delta\ln N_{CCN} \quad \text{(S6)}$$

$$\Delta R_C = \frac{1}{3}[R_C(1 - R_C)] \times \Delta ln N_d \approx 0.075 \times \Delta ln N_d \quad \text{(S7)}$$

$$\Delta ln N_d \approx 0.5 \times \Delta\ln N_{CCN} \quad \text{(S8)}$$

where ΔF_c is the uncertainty in cloud radiative forcing, F_T is the solar constant, A_{mst} is the cloud fractional coverage, T_r is the fractional transmission of shortwave radiation above the cloud layer, ΔR_c is the uncertainty in cloud albedo, N_d is the cloud droplet number concentration, N_{CCN} is the cloud condensation nuclei number concentration. In equation (S6), A_{mst} and T_r were set as 0.3 and 0.76, respectively (Charlson et al, 1987; Schwartz et al., 1996), which represented an average fraction of nonoverlapped marine stratus and stratocumulus. Here the uncertainty in ΔR_c in equation (S7) for the typical range of cloud albedo was calculated according to the Wang et al. 2008. The uncertainty N_d associated with change in N_{CCN} was referred from Sotiropoulou et al. (2006) with a value of 0.5 in equation (S8).

References

Buchholz, R. R., Emmons, L. K., Tilmes, S., & The CESM2 Development Team.: CESM2.1/CAM-chem Instantaneous Output for Boundary Conditions. UCAR/NCAR - Atmospheric Chemistry Observations and Modeling Laboratory. e.g. Lat: -10 to 10, Lon: 100 to 150, September 2015 - February 2016, 2019.

Charlson, R. J., Schwartz, S. E., Hales, J. M., Cess, R. D., Coakley Jr., J. A., Hansen, J. E., and Hofmann, D. J.: Climate forcing by anthropogenic aerosols, Science, 255, 423–430, https://doi.org/ 10.1126/science.255.5043.423, 1992.

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.

Fan, J., Leung, L. R., Li, Z., Morrison, H., Chen, H., Zhou, Y., Qian, Y, Wang, Y.: Aerosol impacts on clouds and precipitation in eastern China: Results from bin and bulk microphysics, Journal of Geophysical Research, 117, D00K36, https://doi.org/10.1029/2011JD016537, 2012.

Garrett, T., Zhao, C., Dong, X., Mace, G., Hobbs, P.: Effects of varying aerosol regimes on low-level arctic stratus, Geophysical Research Letters, 31, https://doi.org/10.1029/2004GL019928, 2004.

Grell, G. A., Dezső, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophysical Research Letters, 29, 1693, https://doi.org/10.1029/2002GL015311, 2002.

Hong, S., Noh, Y., Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev, 134(9), 2318, https://doi.org/10.1175/MWR3199.1, 2006.

Iacono, M., Delamere, J., Mlawer, E., Shephard, M., Clough, S., Collins, W.: Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres, 113 (D13), https://doi.org/ 10.1029/2008JD009944, 2008.

Liu, Y., Daum, P. H., & McGraw, R.: An analytical expression for predicting the critical radius in the autoconversion parameterization, Geophysical Research Letters, 31, L06121, https://doi.org/10.1029/2003GL019117, 2004.

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.

Quaas, J., Boucher, O., & Breon, F. M.: Aerosol indirect effects in POLDER satellite data and the Laboratoire de Meteorologie Dynamique Zoom (LMDZ) general circulation model, Journal of Geophysical Research, 109, D08205, https://doi.org/10.1029/2003JD004317, 2004.

Schwartz, S. E. and Slingo, A.: Enhanced shortwave cloud radiative forcing due to anthropogenic aerosols., in: Clouds, Chemistry, and Climate., edited by: Crutzen, P., and Ramanathan, V., Springer, Heidelberg. 191–236, 1996.

Sotiropoulou, R. E. P., Medina, J., and Nenes, A.: CCN predictions: Is theory sufficient for assessments of the indirect effect?, Geophysical Research Letters, 33, L05816, https://doi.org/10.1029/2005GL025148, 2006.

Wang, Y., Niu, S., Lv, J., Lu, C., Xu, X., Wang, Y., Ding, J., Zhang, H., Wang, T., Kang, B.: A new method for distinguishing unactivated particles in cloud condensation nuclei measurements: implications for aerosol indirect effect evaluation, Geophysical Research Letters, 46, 14,185–14,194, https://doi.org/10.1029/2019GL085379, 2019.

Wang, J., Lee, Y.-N., Daum, P. H., Jayne, J., and Alexander, M. L.: Effects of aerosol organics on cloud condensation nucleus (CCN) concentration and first indirect aerosol effect, Atmos. Chem. Phys., 8, 6325–6339, https://doi.org/10.5194/acp-8-6325-2008, 2008.

Zaveri, R. A., Easter, R. C., Fast, J. D., and Peters, L. K.: Model for simulating aerosol interactions and chemistry (MOSAIC), Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD008782, 2008.

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.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., and Zhang, Q.: 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.

Table S1. Summary of the datasets used in this study from multiple sources

Data category	Content (Abbreviation)	Unit	Spatial resolution	Temporal resolution	Data source
N_{CCN}	CCN concentration (N_{CCN}) at S=0.20% and S=0.40%	cm^{-3}	10 km	Hourly	WRF-Chem model
Chemical compositions	Organic matter (OM)	$\mu g\ m^{-3}$	10 km	Daily	TAP (http://tapdata.org.cn/?page_id=59&item=pm25)
	Sulfate (SO_4^{2-})	$\mu g\ m^{-3}$	-	-	
	Nitrate (NO_3^-)	$\mu g\ m^{-3}$	-	-	
	Ammonium ($NH4^+$)	$\mu g\ m^{-3}$	-	-	
Meteorological data	2m temperature (TEM)	K	0.25°×0.25°	Hourly	ERA-5 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form)
	Relative humidity (RH)	%	-	-	
	Boundary layer (BL)	M	-	-	
	Surface pressure (SP)	hPa	-	-	
	Surface net solar radiation (SNSR)	$W\ m^{-2}$	-	-	
Gas pollutants	$PM_{2.5}$	$\mu g\ m^{-3}$	-	Hourly	CEMC (https://quotsoft.net/air/)
	NO_2	-	-	-	
	SO_2	-	-	-	
	CO	-	-	-	
	O_3	-	-	-	

Figures

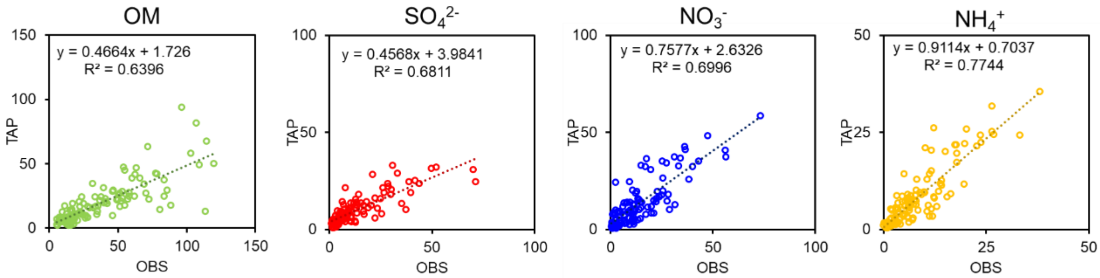

Figure S1. Comparison of the mass concentration of organic, sulfate, nitrate, ammonium between the TAP dataset and observed values at BJ site.

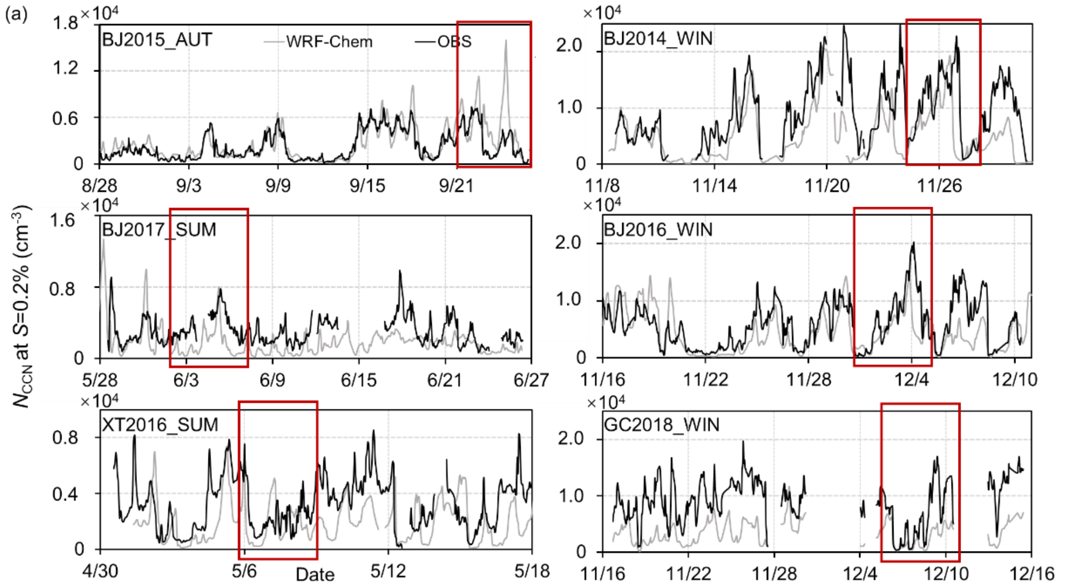

Figure S2. Time series of the WRF-Chem simulated and observed N_{CCN} at S=0.2% for the six campaigns, red box indicates the validation dataset.

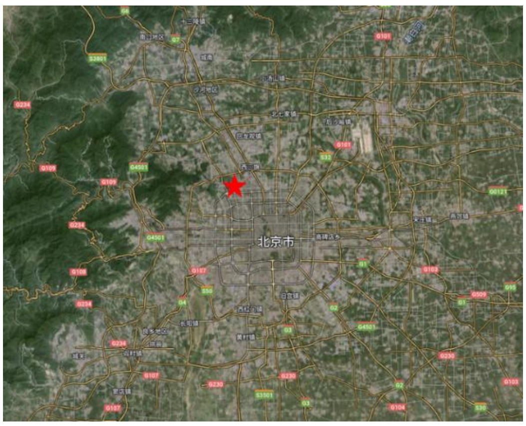

Figure S3. Map location of the sampling site (red star represents the Peking University). (Imagery © 2025 NASA, Map data © 2025 Google, https://maps.google.com/, last access: 8 September 2024).

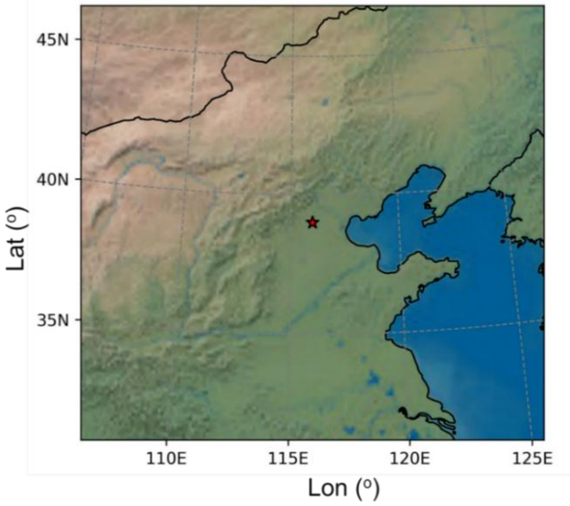

Figure S4. WRF-Chem model simulation domain diagram drawn by using the Cartopy library. Red star represents the center point of the simulation domain.

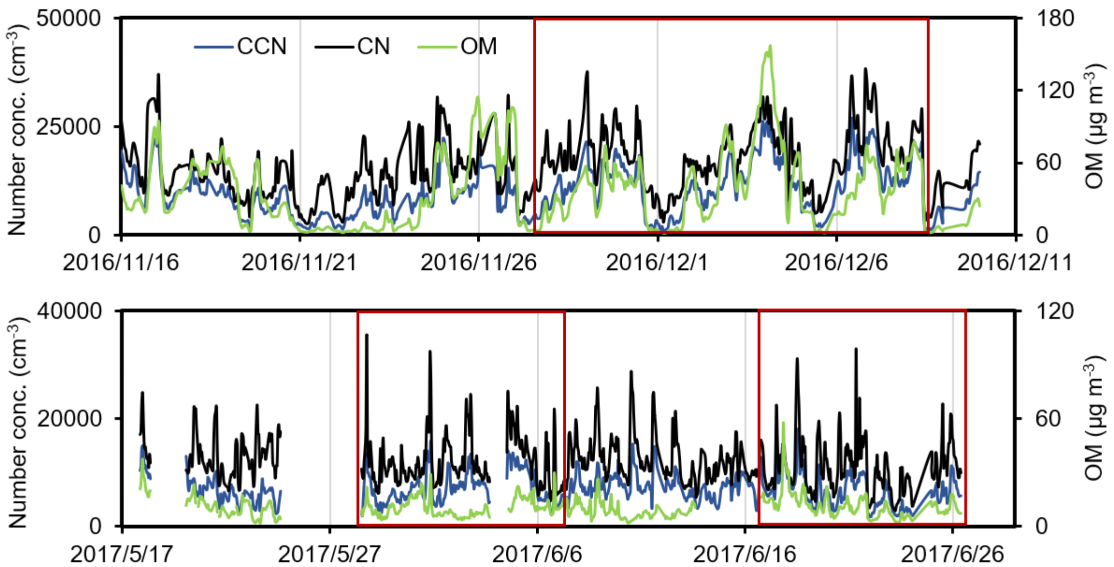

Figure S5. Time series of CN, CCN number concentration and organic mass concentration in the campaigns of 2016BJ and 2017BJ.

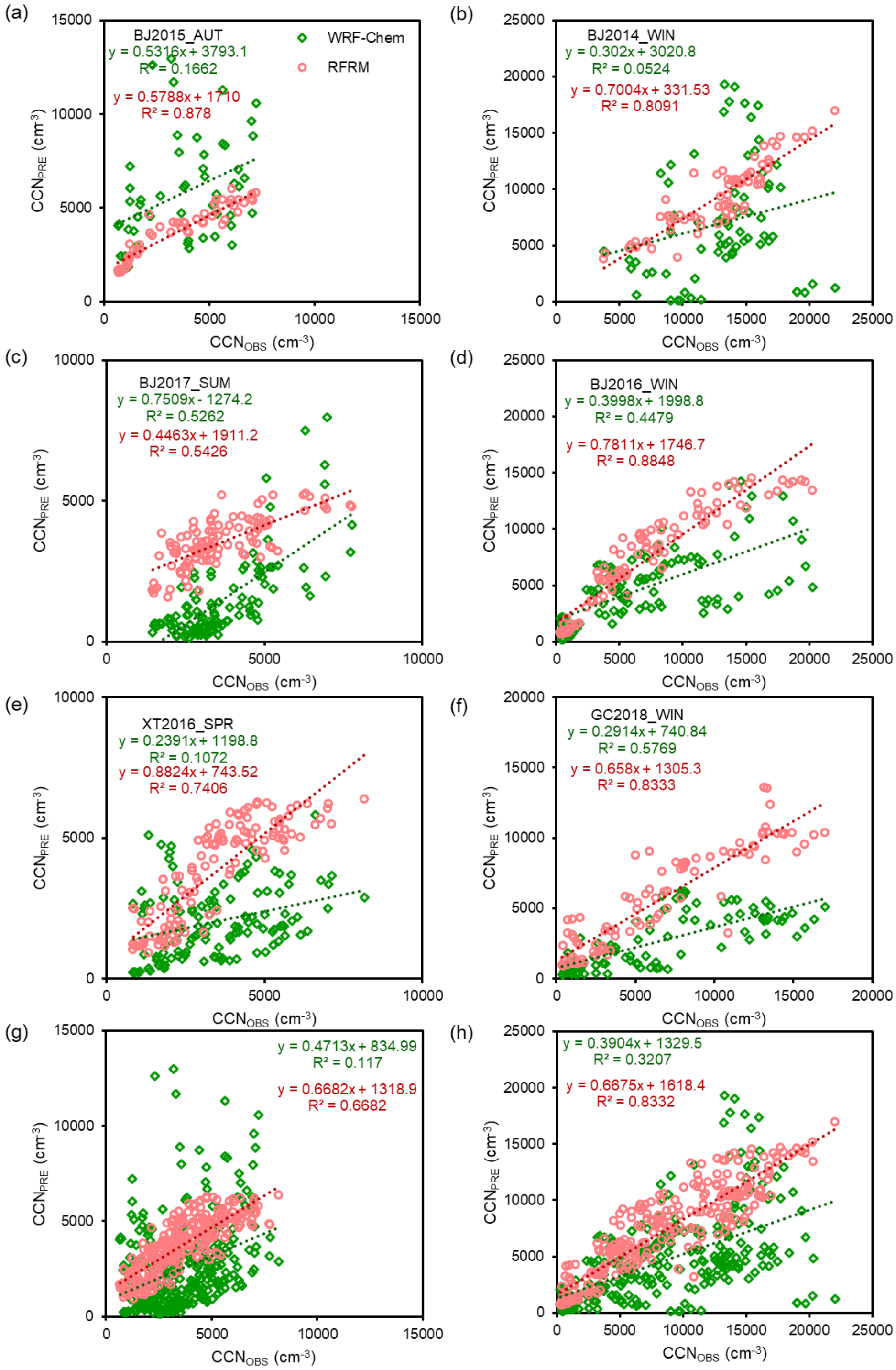

Figure S6. Scatter plots of the observed N_{CCN} with the RFRM-predicted and WRF-Chem-simulated for each campaign at S=0.2%.

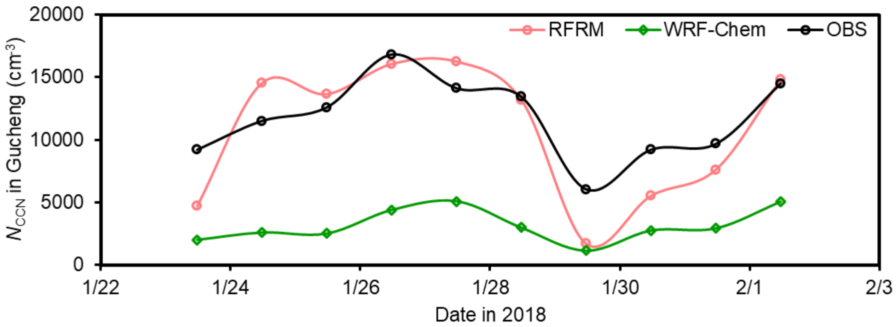

Figure S7. Time series of the RFRM-predicted, WRF-Chem simulated and observed N_{CCN} at S=0.2% at Gucheng.

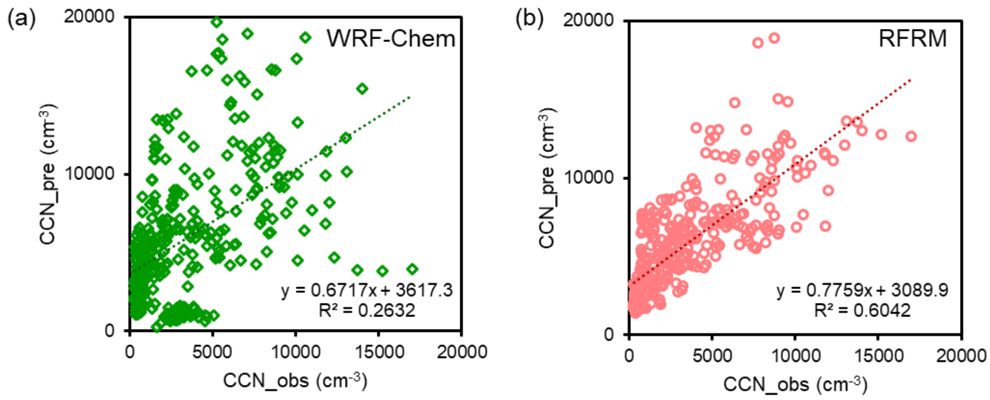

Figure S8. Scatter plots of the observed N_{CCN} at S=0.4% with the WRF-Chem simulated and RFRM model predicted, respectively. Here the training, testing, and validation datasets are independent of the period corresponding to the supersaturation of 0.2%.

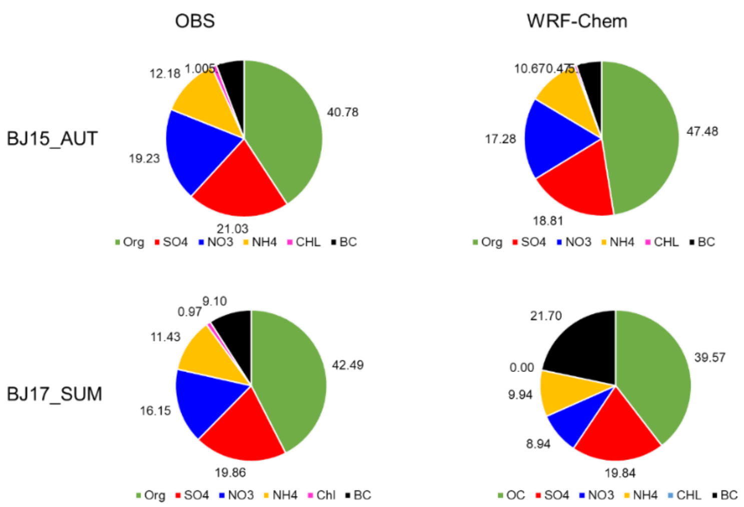

Figure S9. Pie charts of the mass fraction of the chemical composition between the observation and simulation from WRF-Chem, taken the BJ15_AUT and BJ17_SUM as the examples.

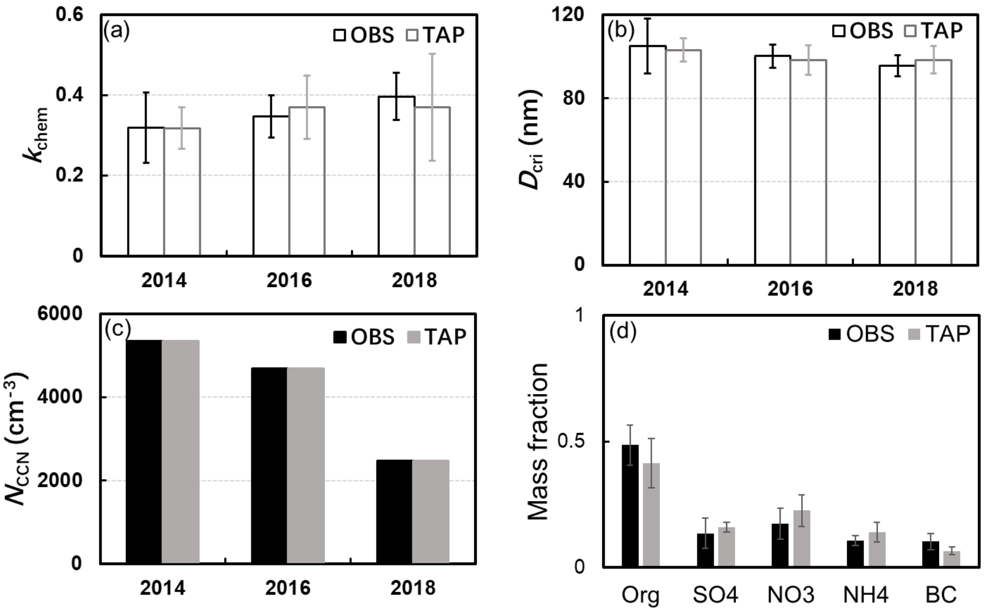

Figure S10. Average annual value of the k_{chem} calculated from chemical composition (a), the critical diameter at S=0.2% (b), N_{CCN} at S=0.2% (c), the mass fraction (d) between the observed and TAP dataset in the winter of 2014, 2016, 2018.

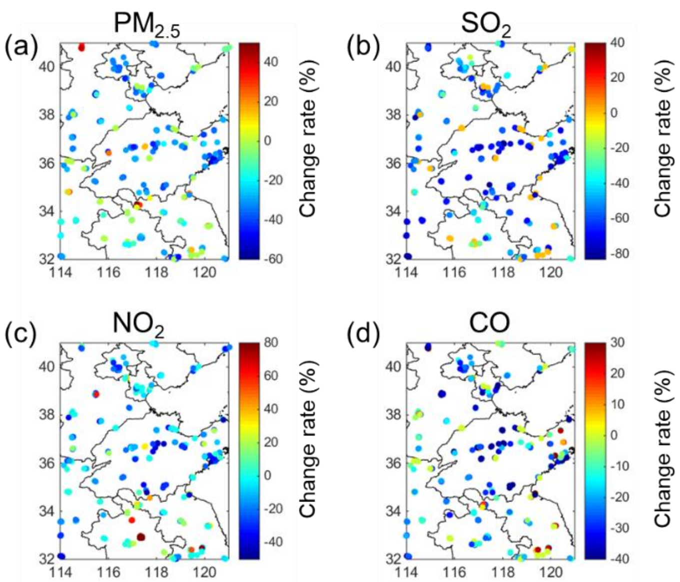

Figure S11. Spatial pattern of the change rates of the $PM_{2.5}$, SO_2, NO_2 and CO from 2014 to 2018 in NCP.

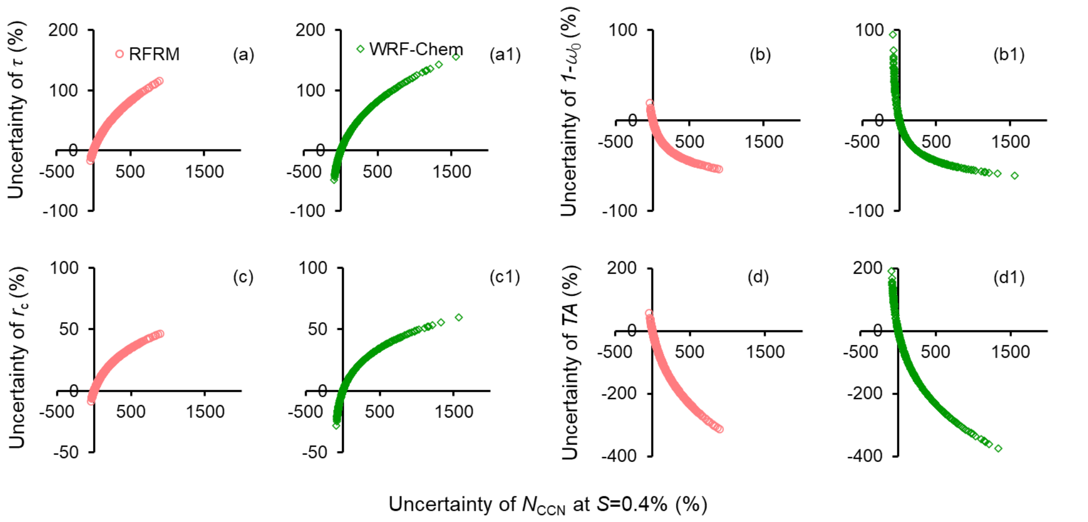

Figure S12. Sensitivity of the cloud properties on the CCN concentration at S=0.4%, uncertainty of the CCN number concentration on cloud optical thickness (τ) between the observed with the RFRM model (a) and with the WRF-Chem model (a1), on absorption coefficient (1-ω_0) and effective radius (r_e) (b and b1), on critical radius (r_c) (c and c1), the cloud-to-rain conversion threshold function (TA) (d and d1).

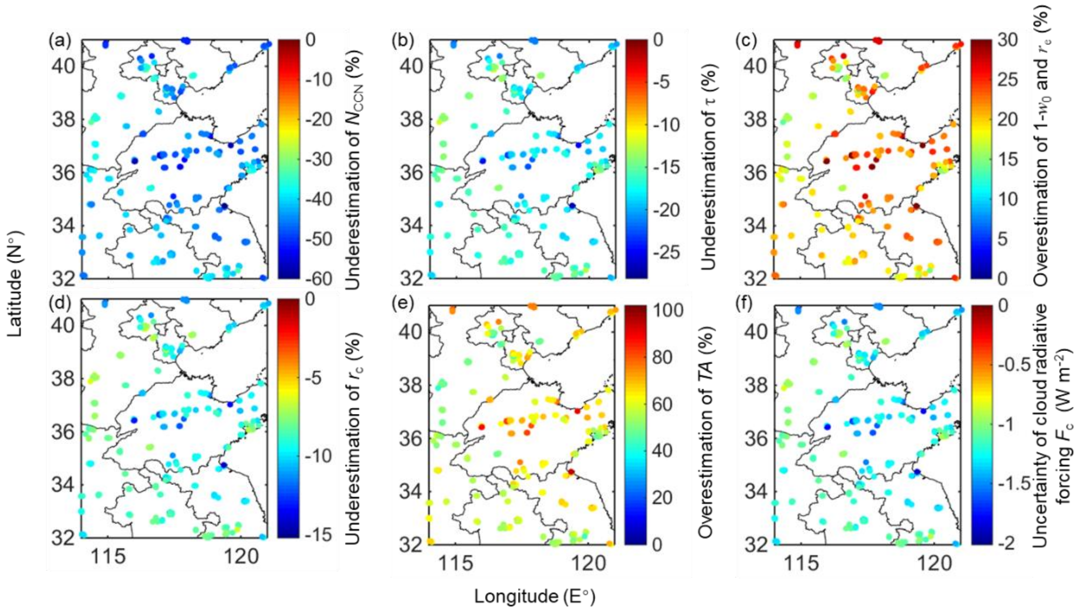

Figure S13. Spatial pattern of the differences of CCN number concentration between the RFRM model and WRF-Chem (a), effects of the discrepancy from CCN number concentration on cloud optical thickness (τ) (b), on absorption coefficient (1-ω_0) and effective radius (r_e) (c), on critical radius (r_c) (d), the cloud-to-rain conversion threshold function (TA) (e), cloud radiative forcing (F_c) (f).