Supplement: A process-evaluation of the impact of precipitation on aerosol particle number size distributions in three Earth System Models

Sara M. Blichner[1,2], Theodore Khadir[1,2], Sini Talvinen[1,2], Paulo Artaxo[3], Liine Heikkinen[1,2], Harri Kokkola[4,5], Radovan Krejci[1,2], Muhammed Irfan[4], Twan van Noije[6], Tuukka Petäjä[7], Christopher Pöhlker[8], Øyvind Seland[9], Carl Svenhag[10,11], Antti Vartiainen[4,12], and Ilona Riipinen[1,2]

[1]Department of Environmental Science, Stockholm University, Stockholm, 10691, Sweden
[2]Bolin Centre for Climate research, Stockholm University, Stockholm, 10691, Sweden
[3]Instituto de Física, Universidade de São Paulo, São Paulo, Brazil
[4]Department of Technical Physics, University of Eastern Finland, Kuopio, 70211, Finland
[5]Finnish Meteorological Institute, Kuopio FI-70211, Finland.
[6]Royal Netherlands Meteorological Institute, De Bilt, Netherlands
[7]University of Helsinki, Institute for Atmospheric and Earth System Research (INAR), Helsinki, Finland.
[8]Multiphase Chemistry Department, Max Planck Institute for Chemistry, 55128 Mainz, Germany
[9]Norwegian Meteorological Institute, Oslo, Norway
[10]Department of Physics, Lund University, Lund, Sweden
[11]Now at: Department of Environmental Science, Aarhus University, Roskilde, Denmark
[12]Advanced Computing Facility, CSC – IT Center for Science Ltd, Espoo, 02150, Finland

Correspondence: Sara M. Blichner (sara.blichner@aces.su.se)

Contents

S1 Model descriptions

S1.1 NorESM

Mode/Process	Below-cloud Scavenging Coefficient
Mode 00 (BC emitted externally mixed as fractal accumulation mode)	0.01
Mode 01 (NPF particles formed by SO_4 and SOA, usually Aitken mode)	0.02
Mode 02 (BC coated, Aitken mode)	0.02
Mode 04 (BC and OM (Organic matter) emitted internally, Aitken mode)	0.02
Mode 05 (SO_4 primary emitted, Accumulation mode)	0.01
Mode 06 (Dust, accumulation mode)	0.02
Mode 07 (Dust, Coarse mode)	0.2
Mode 08 (Sea salt, Aitken mode)	0.02
Mode 09 (Sea salt, accumulation mode)	0.02
Mode 10 (Sea salt, coarse mode)	0.5
Mode 12 (BC emitted externally mixed, nucleation mode)	0.08
Mode 14 (OM and BC emitted internally mixed, Aitken mode)	0.02
Process tracer SO_4_A1 condensate on extisting particles from gas phase	0.02
Process tracer SO_4_A2 formed from aquous phase chemistry	0.01
Process tracer SO_4_AC formed from coagulation with outer particles	0.02
Process tracer OM_AC OM and SOA formed from coagulation with outer particles	0.02
Process tracer BC_AC formed from coagulation with outer particles	0.02
Process tracer SOA_A1 condensate on existing particles from gas phase	0.02

Table S1. Below-cloud Scavenging Coefficients NorESM

S2 Basic evaluation

S2.1 Precipitation

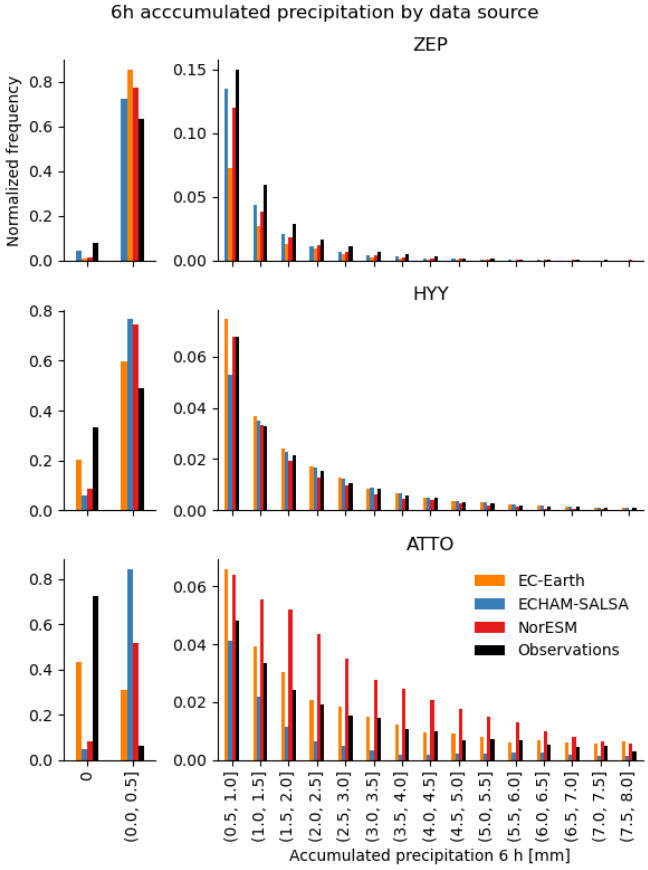


Figure S1. Frequency of 6 hours accumulated precipitation at each station [mm].

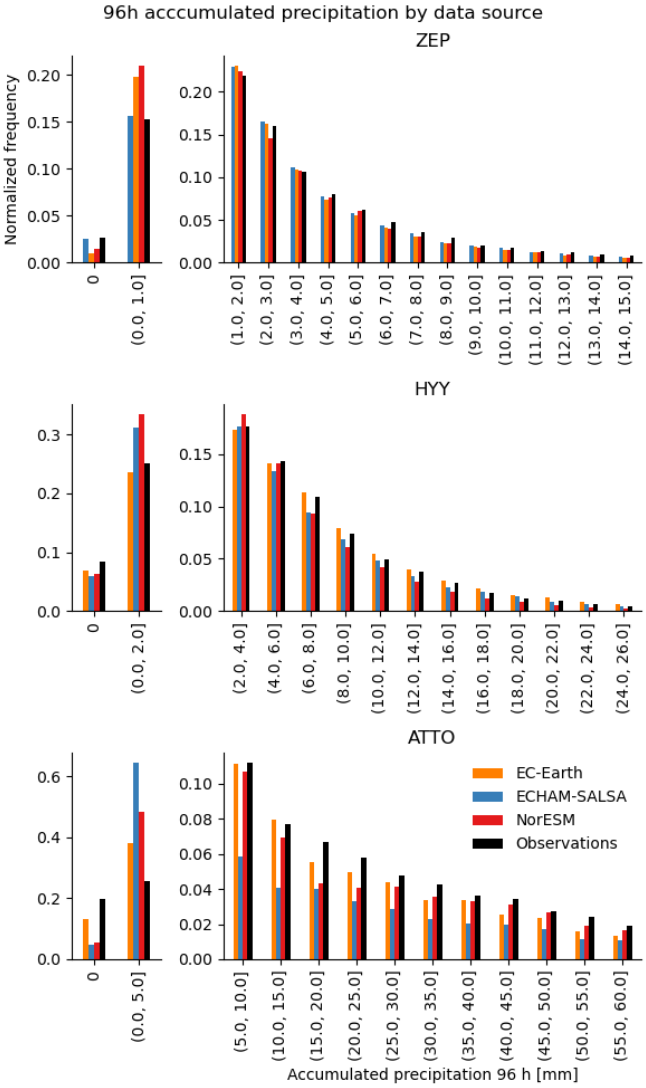


Figure S2. Frequency of 96 hours accumulated precipitation at each station [mm].

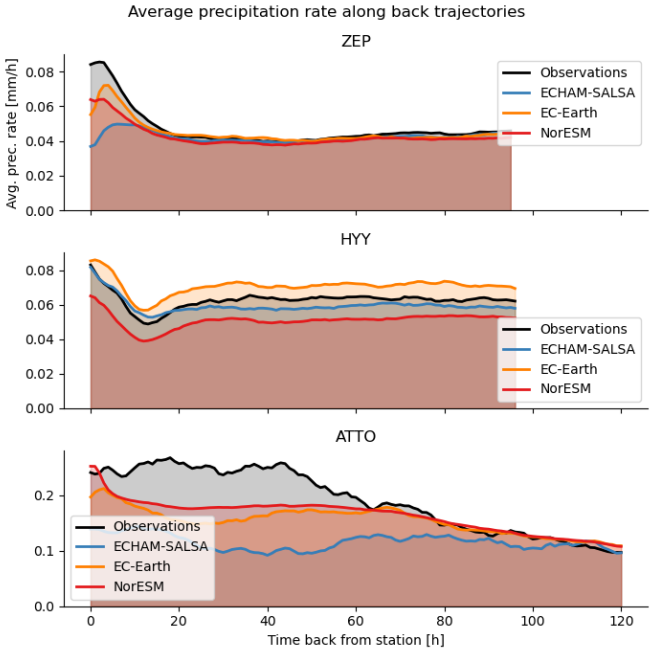


Figure S3. Average precipitation rate along back trajectories [mm/h].

S2.2 Diurnal variation

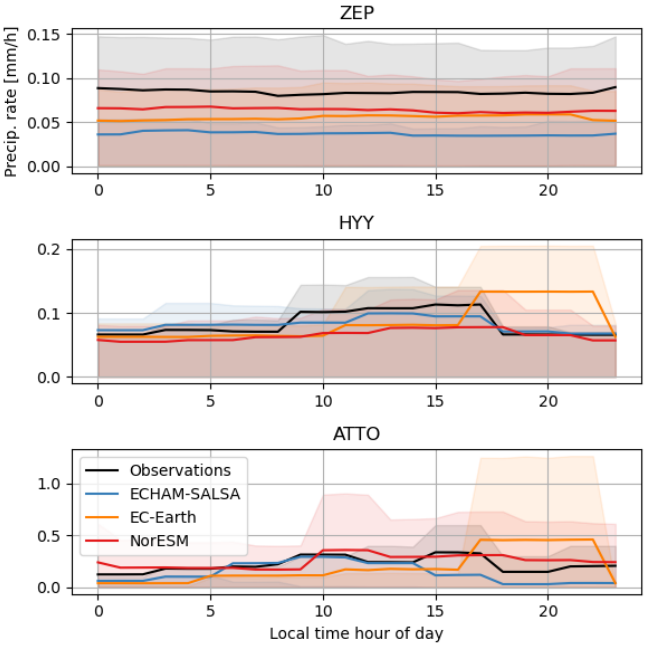


Figure S4. Mean diurnal variation in precipitation rate taken at the each station. The shaded area indicates the 16th-84th percentile.

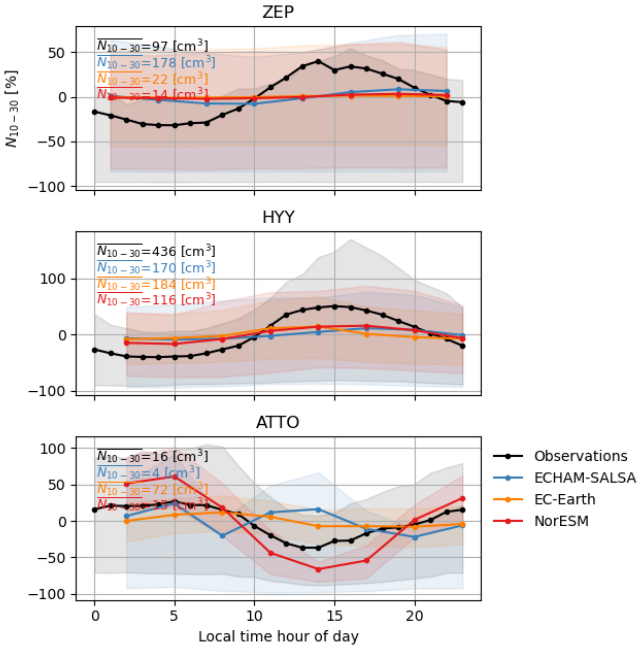


Figure S5. Mean diurnal variation in N_{10-30} at the each station in percentage deviation from the mean value of the total dataset (not each day). The shaded area indicates the 16th-84th percentile and the numbers indicate the mean at each station for each data source.

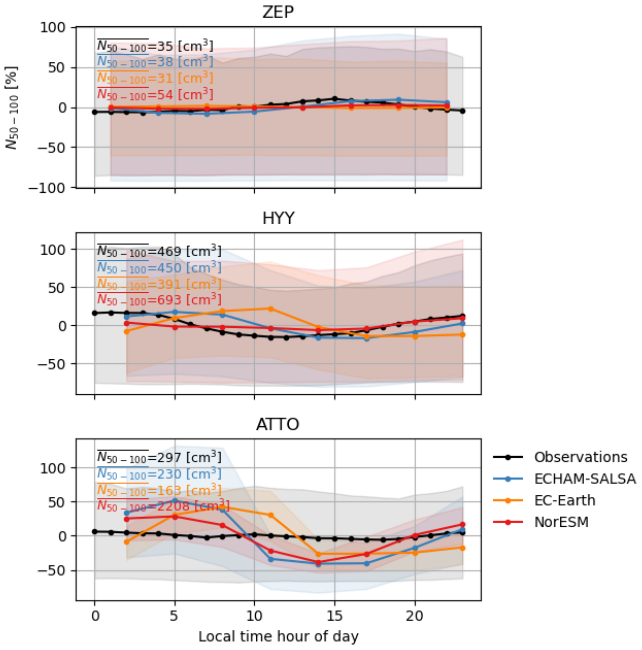


Figure S6. Mean diurnal variation in N_{50-100} at the each station in percentage deviation from the mean value of the total dataset (not each day). The shaded area indicates the 16th-84th percentile and the numbers indicate the mean at each station for each data source..

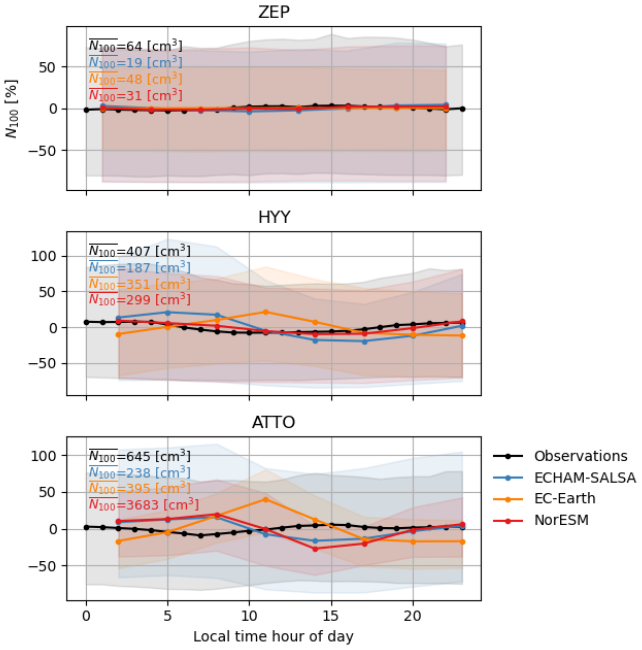


Figure S7. Mean diurnal variation in N_{100} at the each station in percentage deviation from the mean value of the total dataset (not each day). The shaded area indicates the 16th-84th percentile and the numbers indicate the mean at each station for each data source..

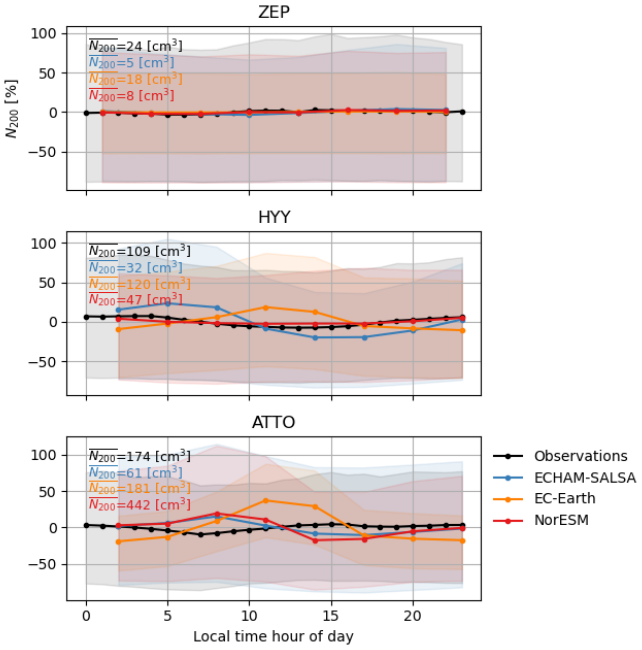
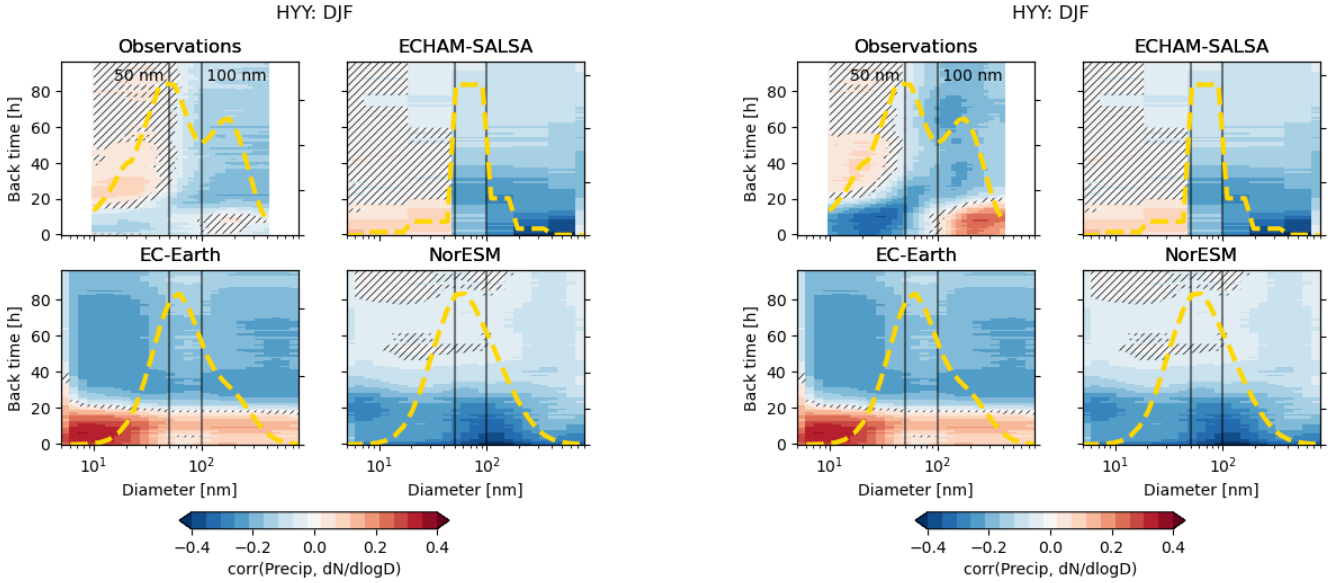


Figure S8. Mean diurnal variation in N_{200} at the each station in percentage deviation from the mean value of the total dataset (not each day). The shaded area indicates the 16th-84th percentile and the numbers indicate the mean at each station for each data source..

S3 Correlations with GDAS versus ERA

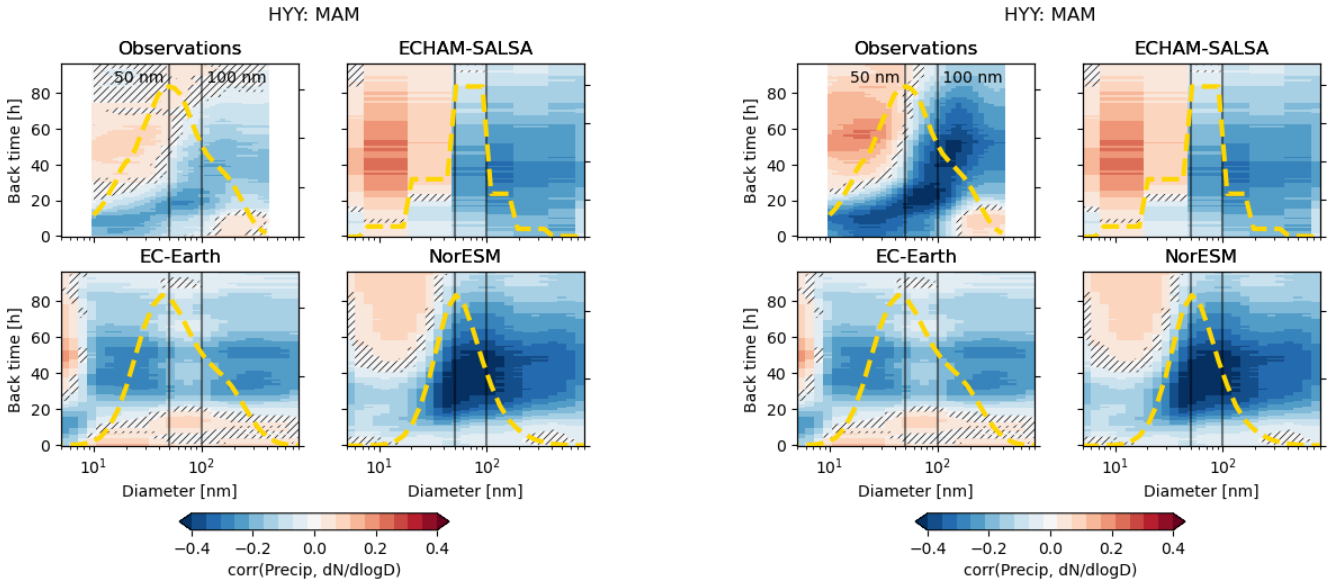
S3.1 Hyytiälä



(a) HYSPLIT, GDAS

(b) ERA5

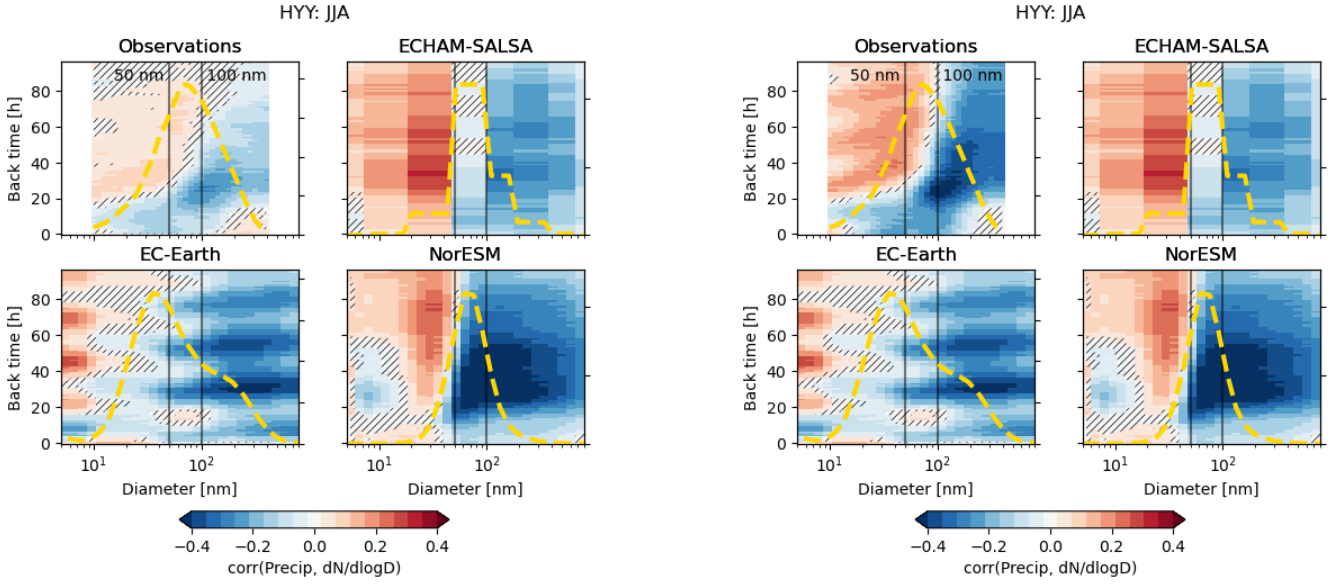
Figure S9. Comparison of GDAS and ERA5 correlations between precipitation and dNdlogD.



(a) Hysplit, GDAS

(b) ERA5

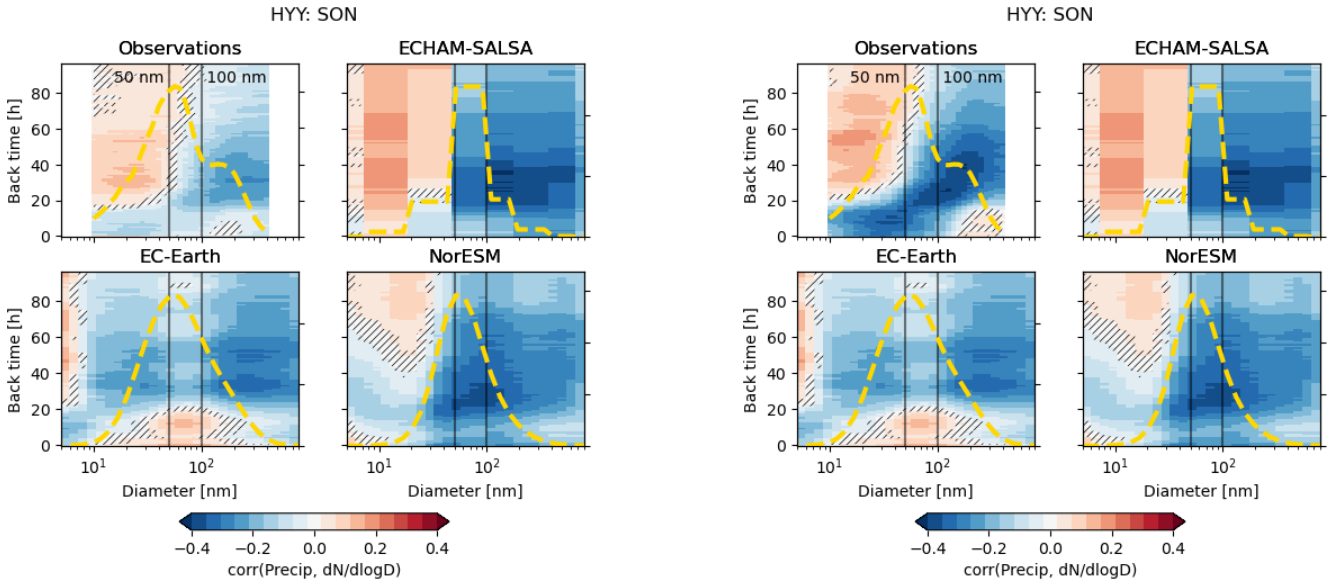
Figure S10. Comparison of GDAS and ERA5 correlations between precipitation and dNdlogD.



(a) HYSPLIT, GDAS

(b) ERA5

Figure S11. Comparison of GDAS and ERA5 correlations between precipitation and dNdlogD.

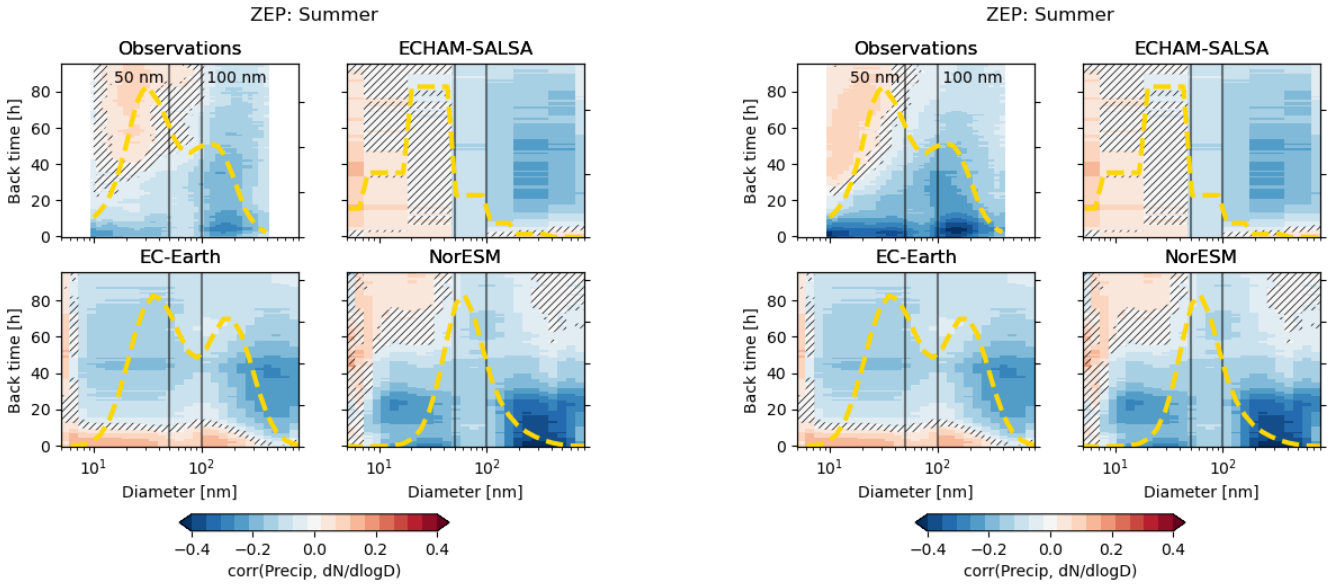


(a) HYSPLIT, GDAS

(b) ERA5

Figure S12. Comparison of GDAS and ERA5 correlations between precipitation and dNdlogD.

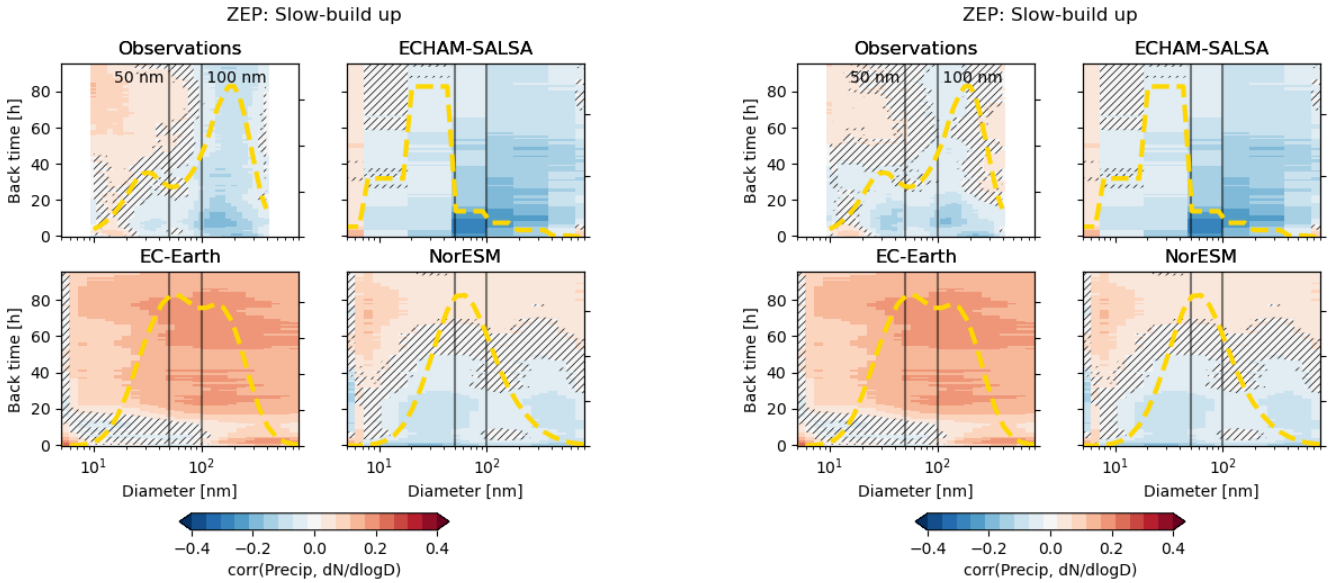
S3.2 Zeppelin



(a) HYSPLIT, GDAS

(b) ERA5

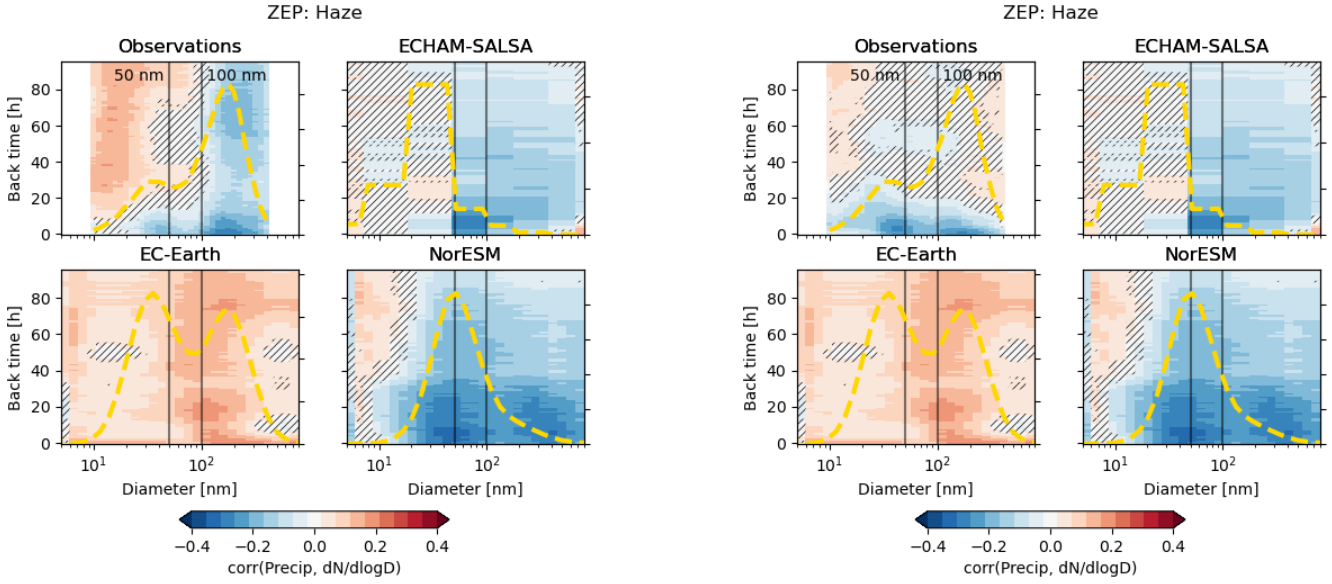
Figure S13. Comparison of GDAS and ERA5 correlations between precipitation and dNdlogD.



(a) HYSPLIT, GDAS

(b) ERA5

Figure S14. Comparison of GDAS and ERA5 correlations between precipitation and dNdlogD.

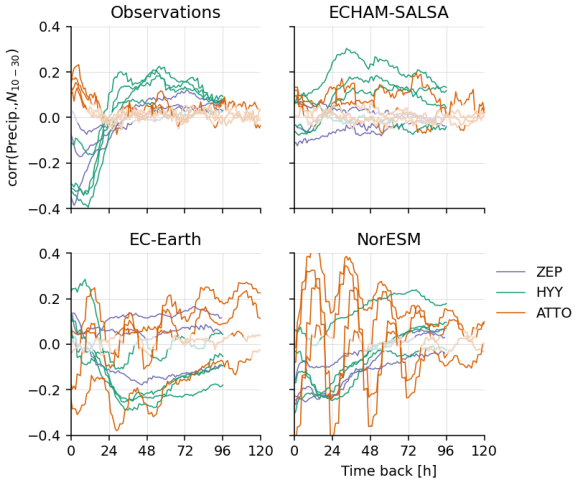


(a) HYSPLIT, GDAS

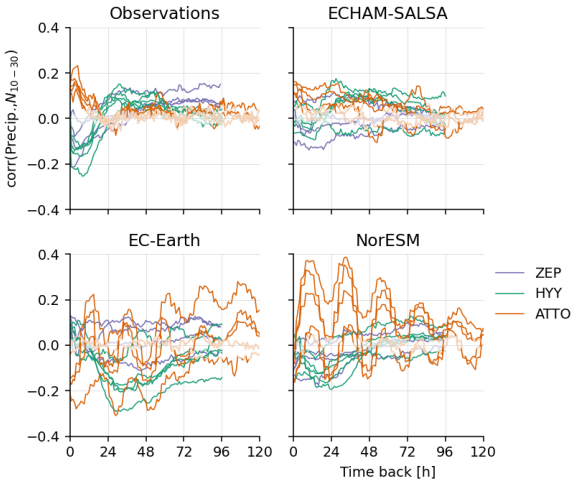
(b) ERA5

Figure S15. Comparison of GDAS and ERA5 correlations between precipitation and dNdlogD.

S4 Correlations with and without drizzle included in models

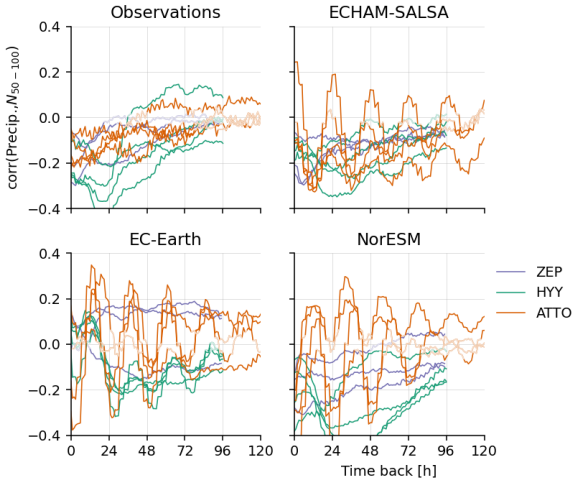


(a) No change to drizzle. ERA5 data for observations.

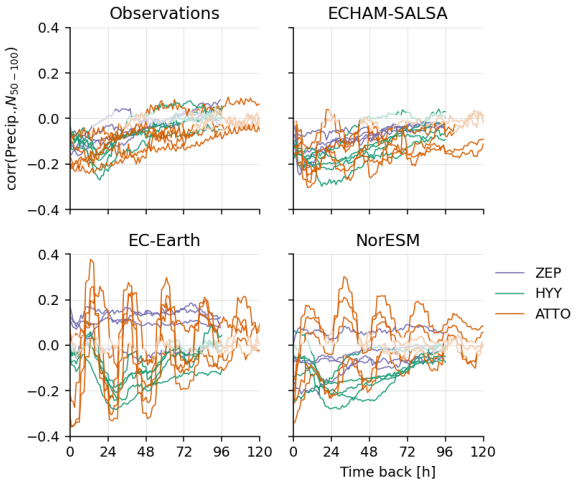


(b) Drizzle set to zero below 0.05 $\mathrm{mm/h}$. HYSPLIT data for observations

Figure S16. Comparison of drizzle treatments for correlations between precipitation and N_{10-30}.

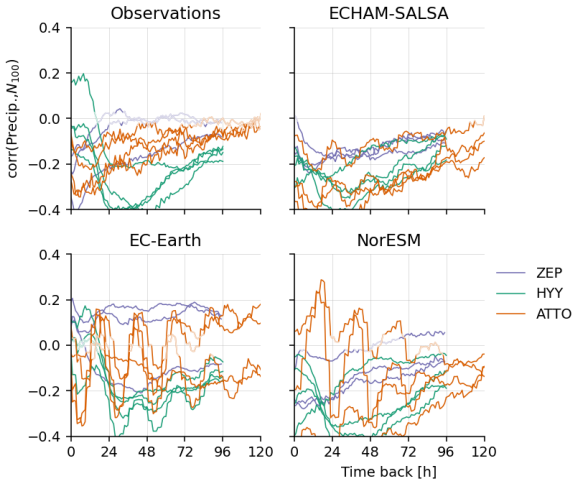


(a) No change to drizzle. ERA5 data for observations.

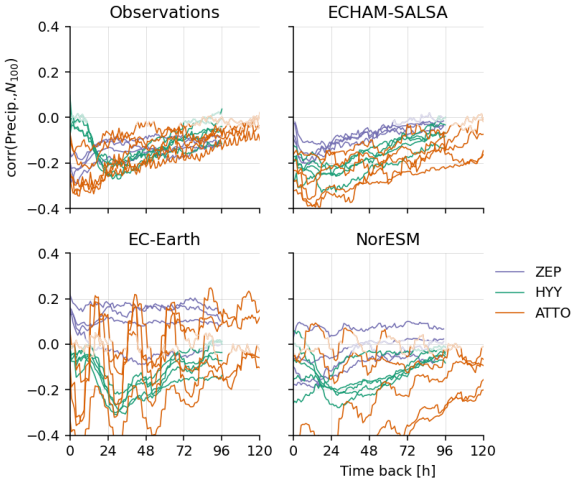


(b) Drizzle set to zero below 0.05 $\mathrm{mm/h}$. HYSPLIT data for observations.

Figure S17. Comparison of drizzle treatments for correlations between precipitation and N_{50-100}.

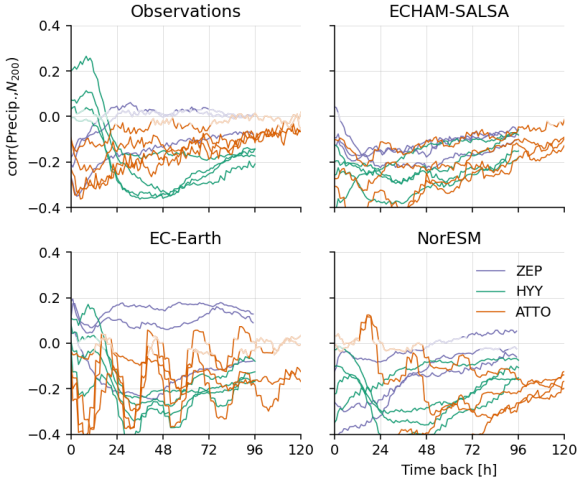


(a) No change to drizzle. ERA5 data for observations.

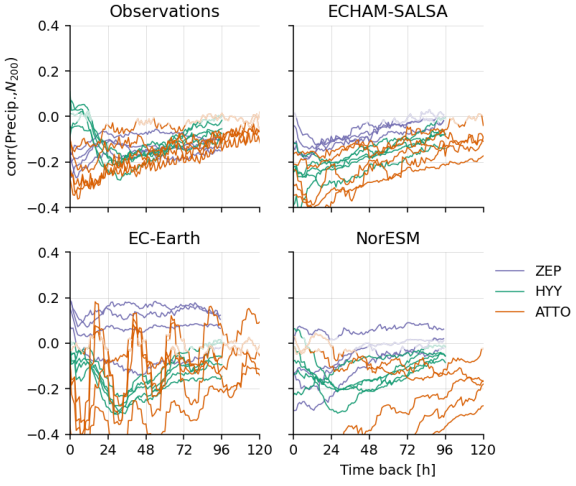


(b) Drizzle set to zero below 0.05 $\mathrm{mm/h}$. HYSPLIT data for observations.

Figure S18. Comparison of drizzle treatments for correlations between precipitation and N_{100}.



(a) No change to drizzle. ERA5 data for observations.



(b) Drizzle set to zero below 0.05 $\mathrm{mm/h}$. HYSPLIT data for observations.

Figure S19. Comparison of drizzle treatments for correlations between precipitation and N_{200}.

Compared to the HYSPLIT output which truncates the values close to zero to zero precipitation, the model have a lot more very low precipitation rates and this impacts the correlations. In Figs. S16–S19, we show how the correlations look if we use (right, a) for the full model dataset with small values of precip included and ERA5 precipitation data for the observations or

(left, b) the results when we artifically truncate the model data so that precipitation very close to zero is set to zero (below 0.05 $\mathrm{mm/h}$) and using the truncated HYSPLIT output for the observations. Mainly they display that the correlations become a little weaker in the models when drizzle is set to zero.

S5 Correlation between precipitation rate and particle number size distritbution

All except ATTO are displayed in the previous section.

S5.1 ATTO

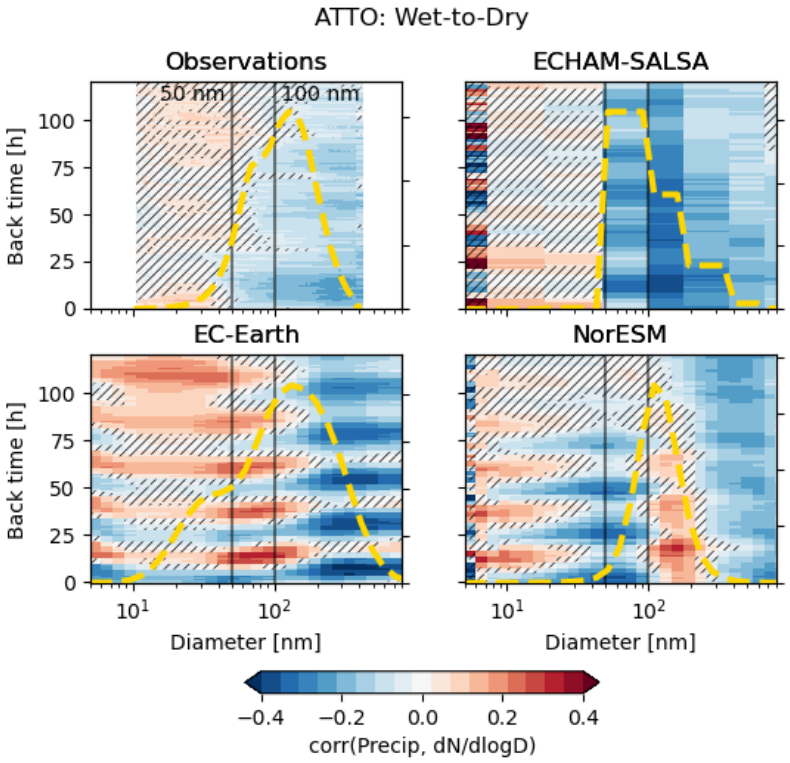


Figure S20. Correlation between rainfall and PNSD, Wet to dry season, ATTO.

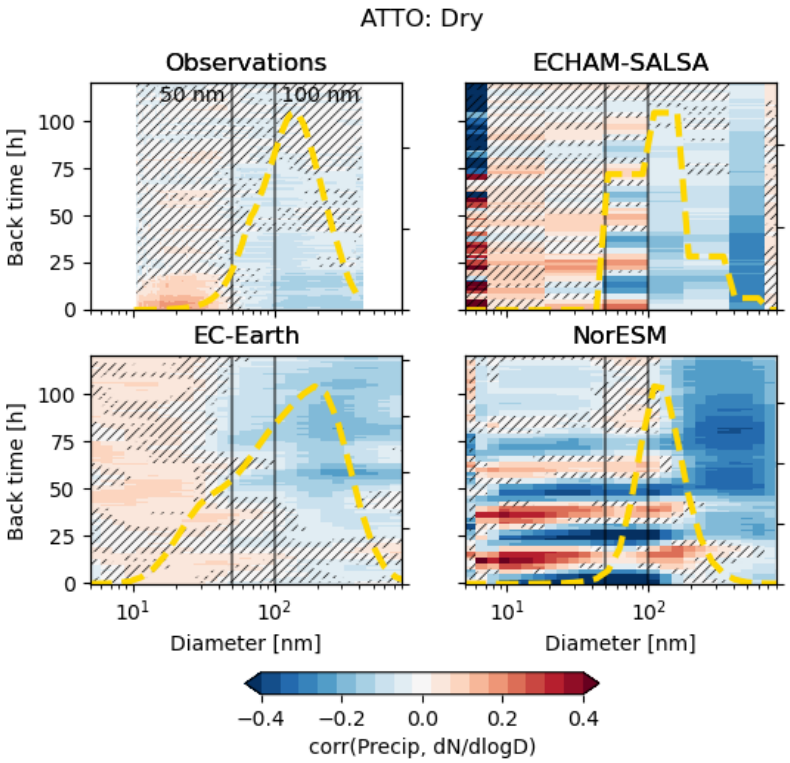


Figure S21. Correlation between rainfall and PNSD, Dry season, ATTO.

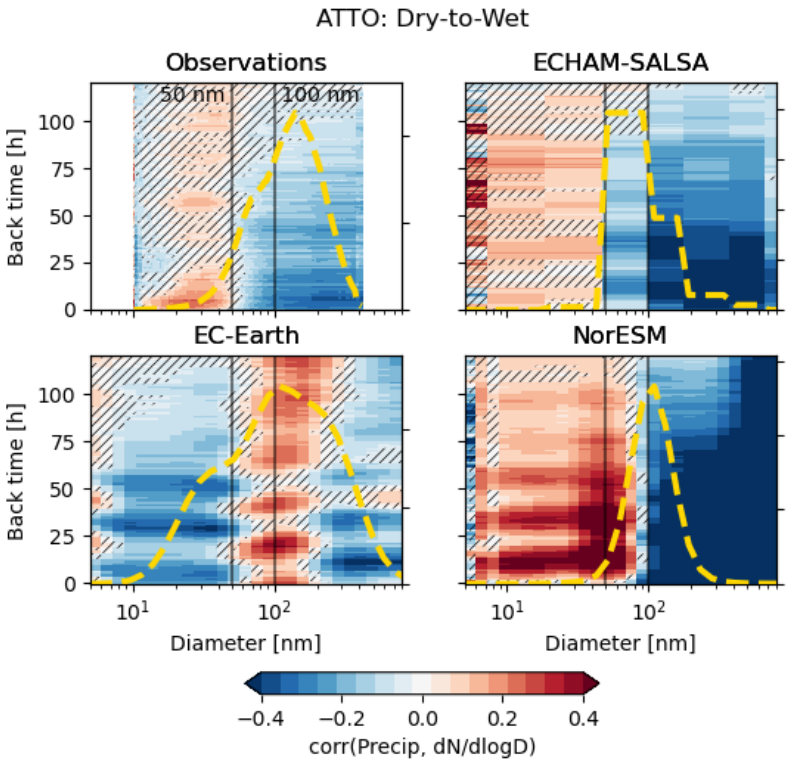


Figure S22. Correlation between rainfall and PNSD, Dry to Wet season, ATTO.

S6 Number concentration versus accumulated precipitation

S6.1 Recent precipitation: 6h accumulated

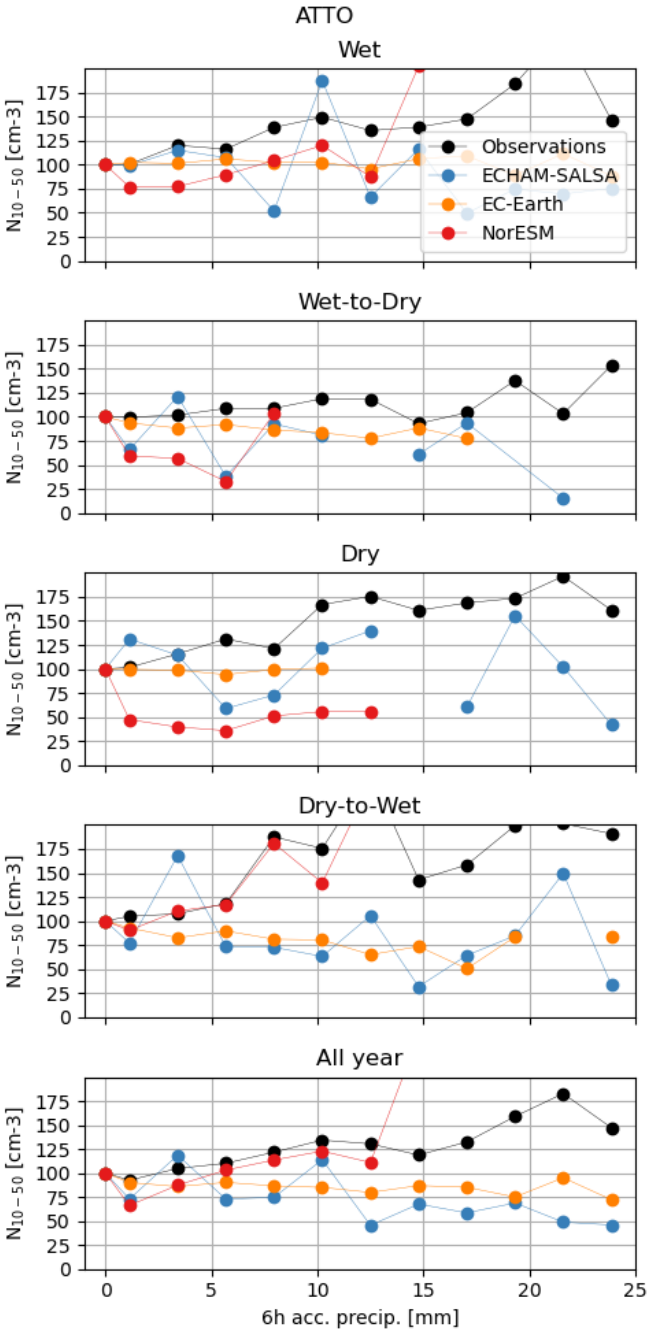


Figure S23. Median N_{10-50} (y-axis) binned by recent accumulated precipitation (6h) (x-axis) for ATTO station.

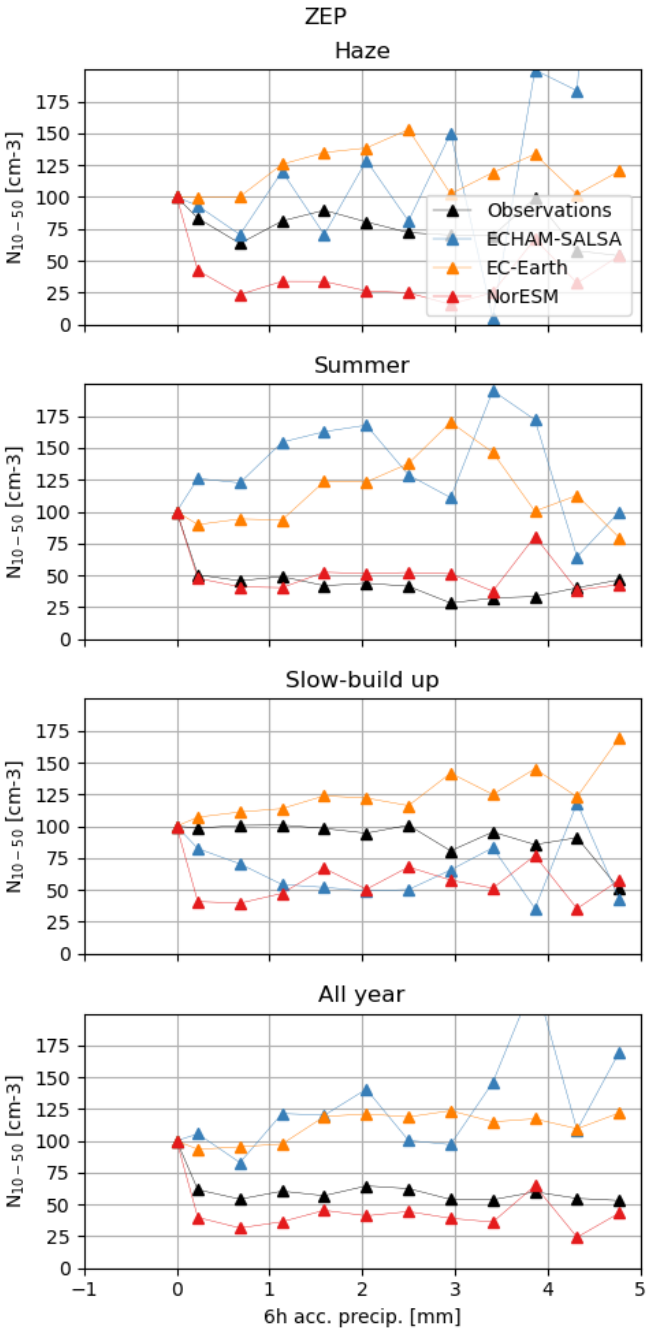


Figure S24. Median N_{10-50} (y-axis) binned by recent accumulated precipitation (6h) (x-axis) for Zeppelin station.

S7 New particle formation

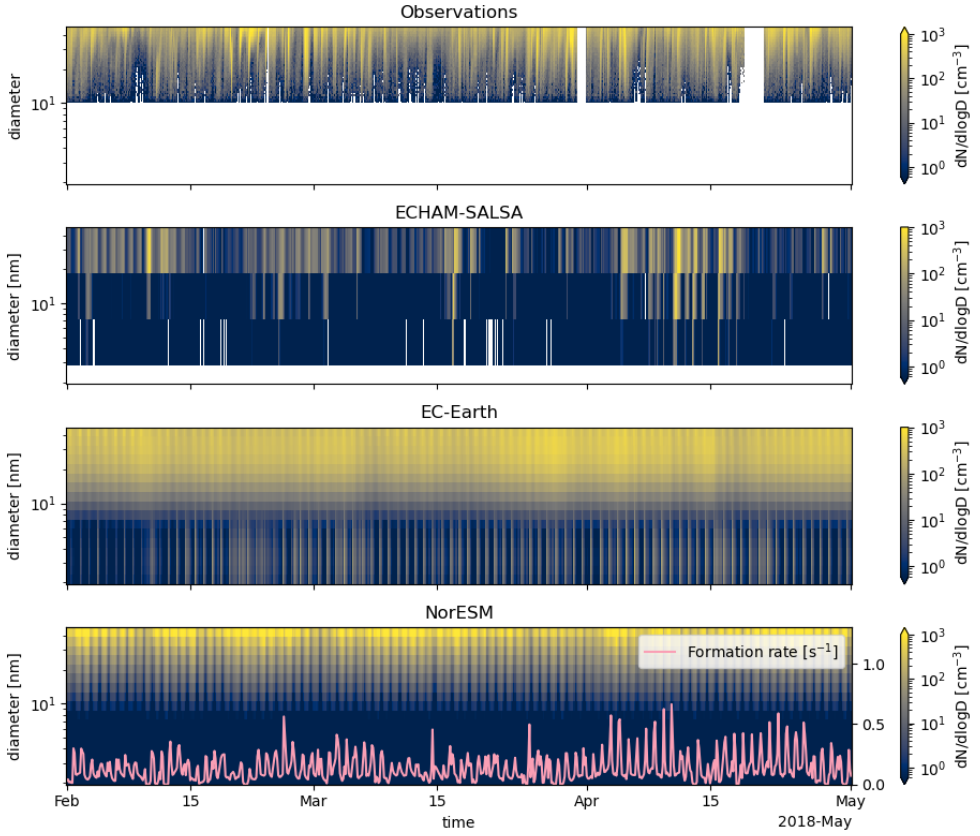


Figure S25. Size distributions in the range 1-50 nm during the wet season (Feb–May) during 2018. For NorESM the formation rate of particles is also shown.

S8 Trajectories

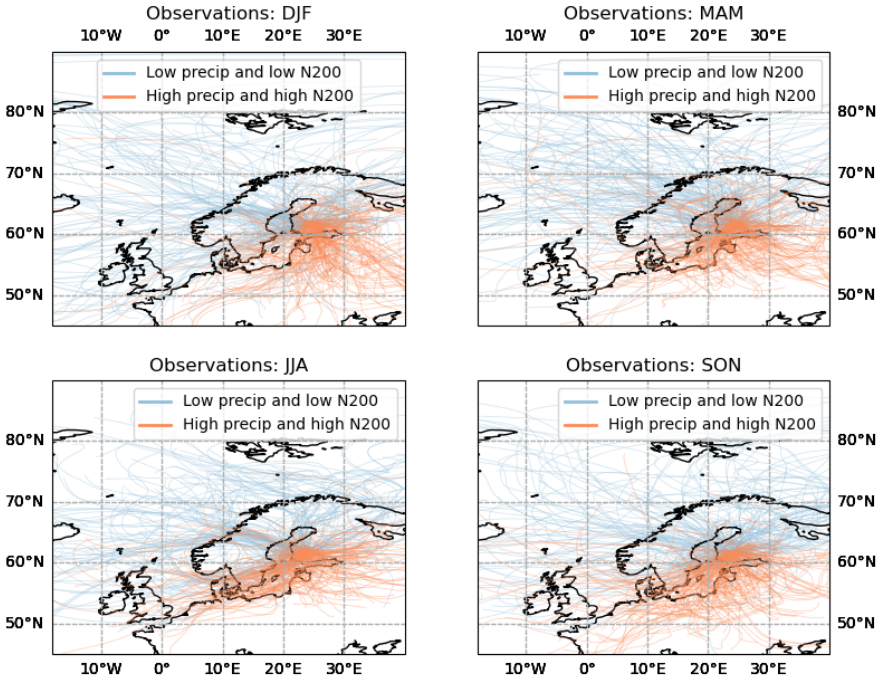


Figure S26. Trajectories for low recent precipitation and low N_{200} (blue, defined as below/above the mean) and high recent precipitation and high N_{200} (orange). Recent precipitation is calcualted as the average of the last 18 hours before the airmass arrival at the station. A random sample of 200 trajectories are plotted for each group.

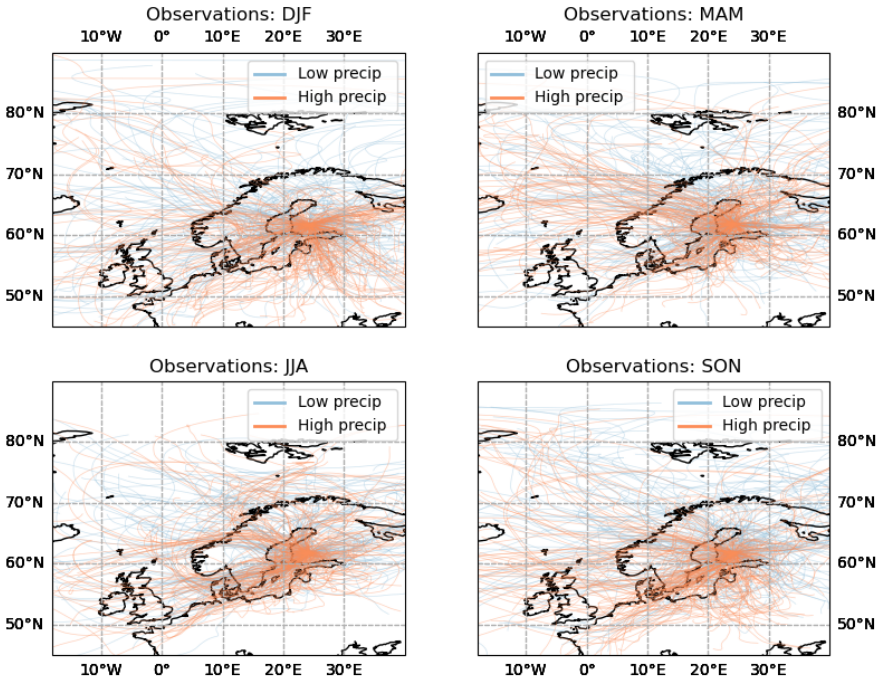


Figure S27. Trajectories for low recent precipitation (blue, defined as below/above the mean) and high recent precipitation (orange). Recent precipitation is calcualted as the average of the last 18 hours before the airmass arrival at the station. A random sample of 200 trajectories are plotted for each group.

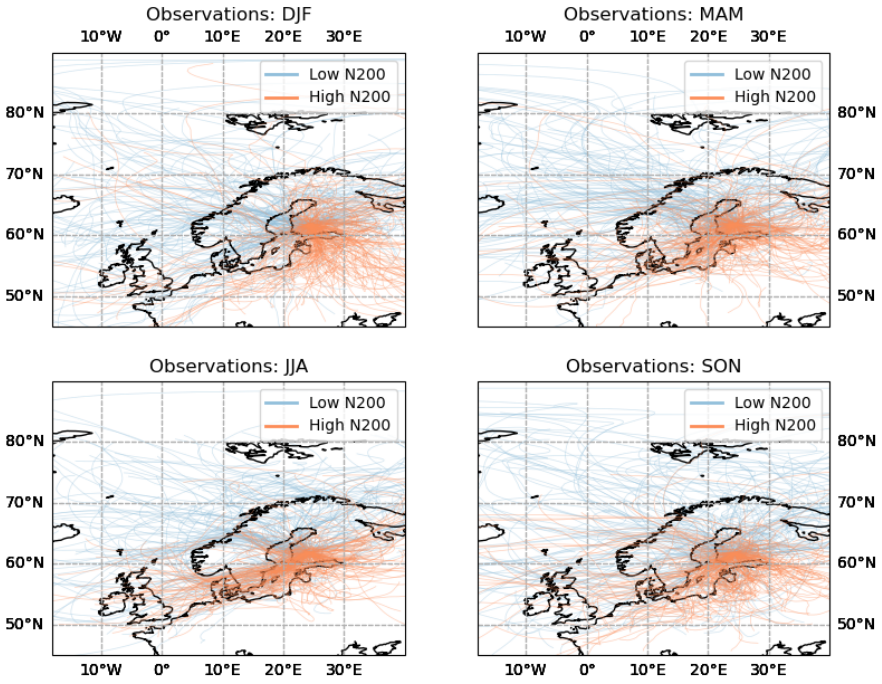


Figure S28. Trajectories for low N_{200} (blue, defined as below/above the mean) and high N_{200} (orange). Recent precipitation is calcualted as the average of the last 18 hours before the airmass arrival at the station. A random sample of 200 trajectories are plotted for each group.

S9 Correlation in different sectors for Hyytiälä

Note that these results use GDAS data for "observed" precipitation.

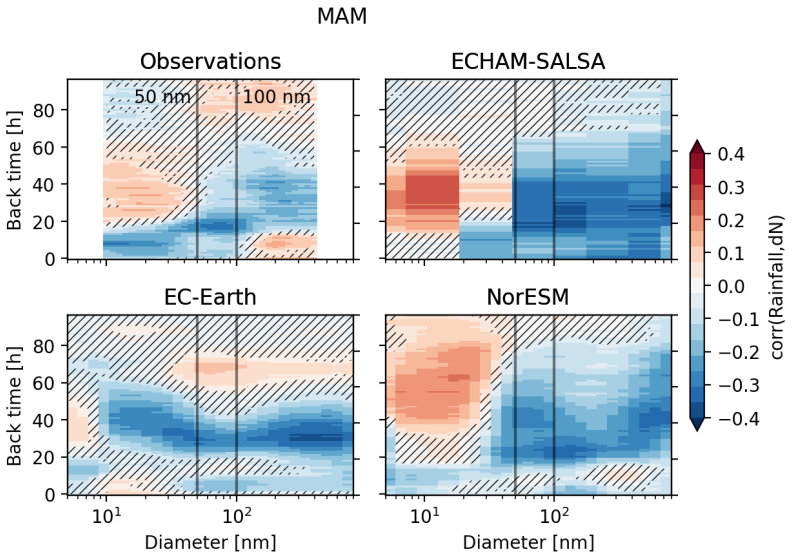


Figure S29. The Spearmann correlation between size distribution measured or modelled at the station and precipitation along the trajectory to the station when only trajectories in sector -175 to 5 degrees centered at the station are taken into account, assuming 0 degrees is straight east.

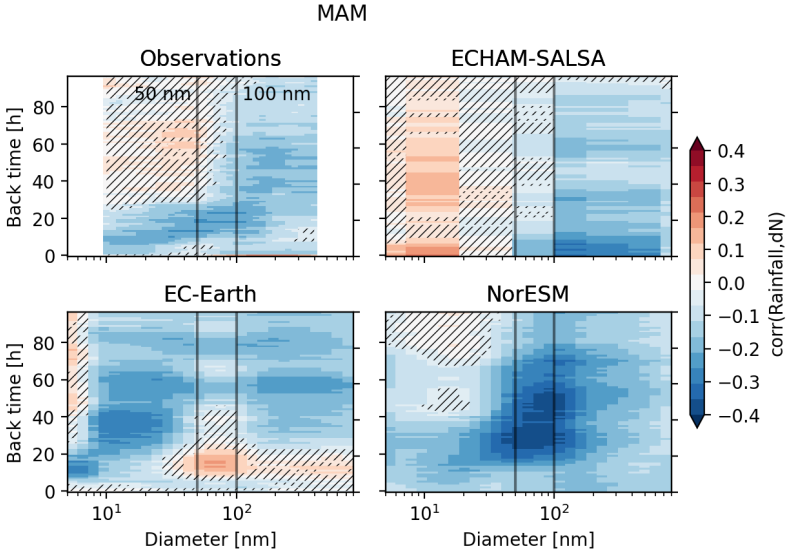


Figure S30. The Spearmann correlation between size distribution measured or modelled at the station and precipitation along the trajectory to the station when only trajectories in sector 5 to 185 degrees centered at the station are taken into account, assuming 0 degrees is straight east.

S10 Correlation between precipitation rate and particle number size distritbution calculated based on daily means

S10.1 Hyytiälä

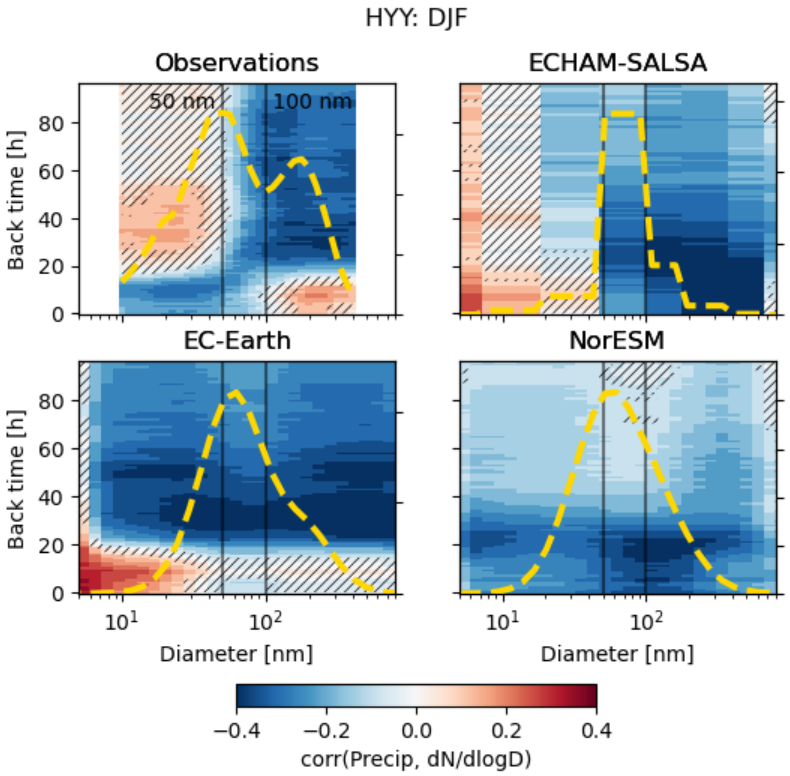


Figure S31. Correlation between rainfall and PNSD, DJF, Hyytiälä, calculated based on daily mean vales.

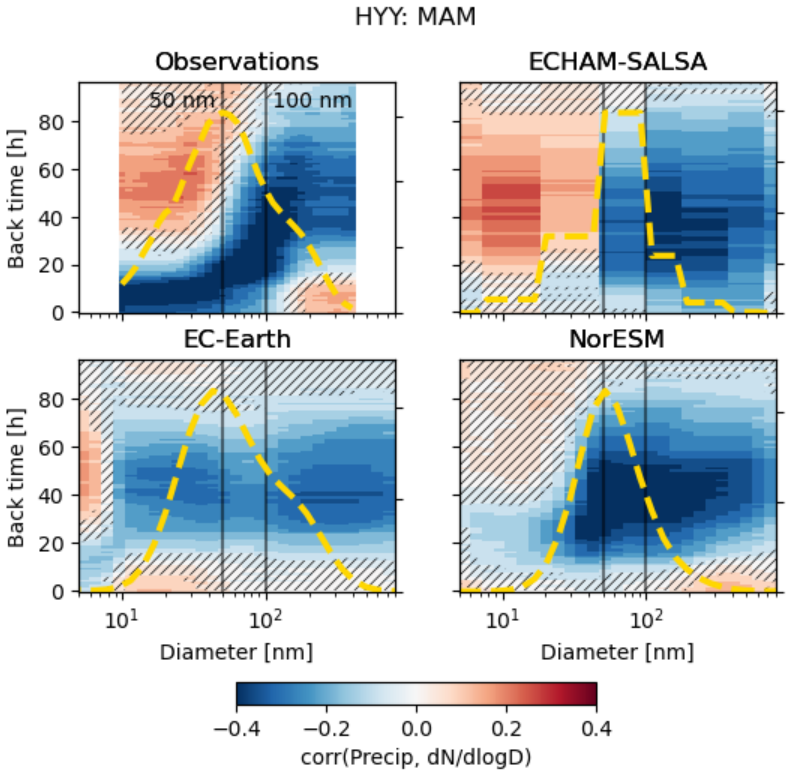


Figure S32. Correlation between rainfall and PNSD, MAM, Hyytiälä, calculated based on daily mean vales.

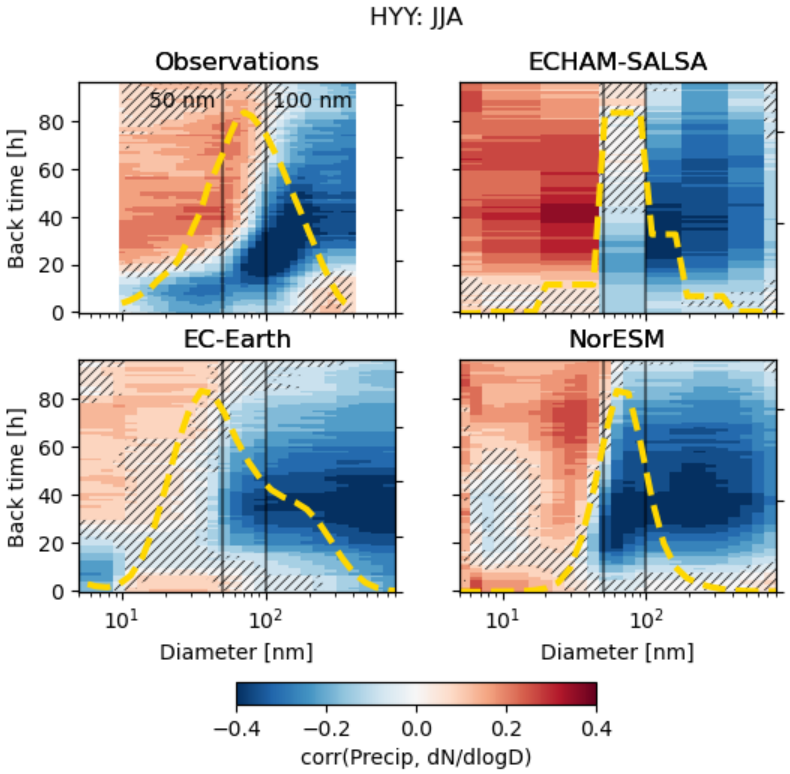


Figure S33. Correlation between rainfall and PNSD, JJA, Hyytiälä, calculated based on daily mean vales.

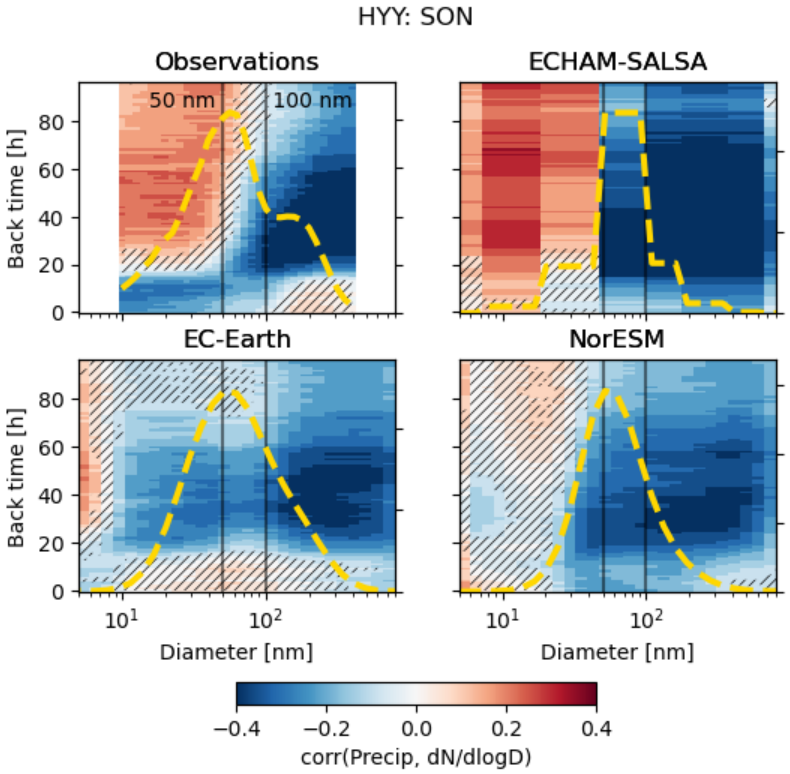


Figure S34. Correlation between rainfall and PNSD, SON, Hyytiälä, calculated based on daily mean vales.

S10.2 Zeppelin

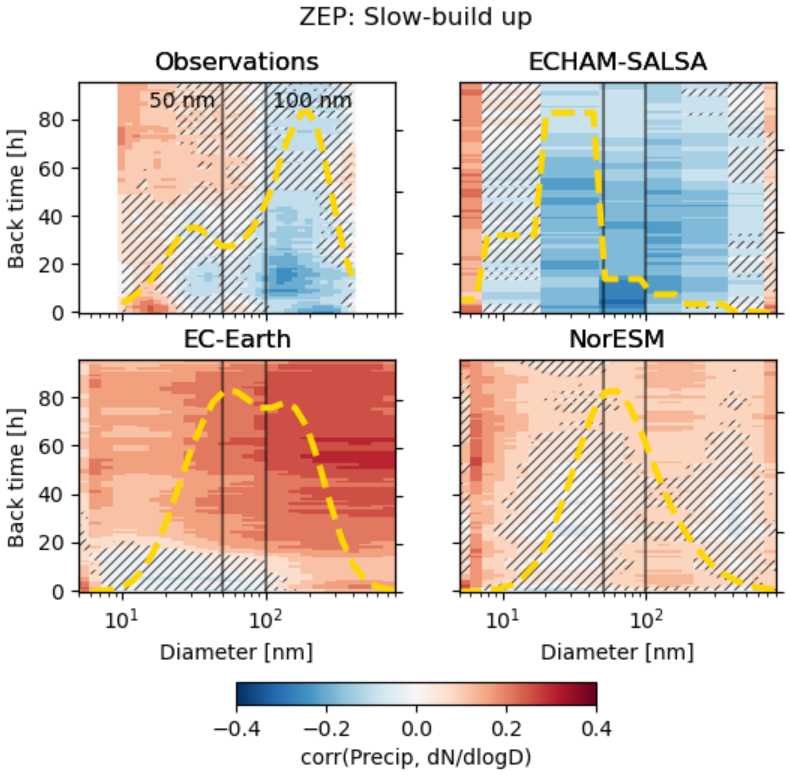


Figure S35. Correlation between rainfall and PNSD, Slow build up, Zeppelin, calculated based on daily mean vales.

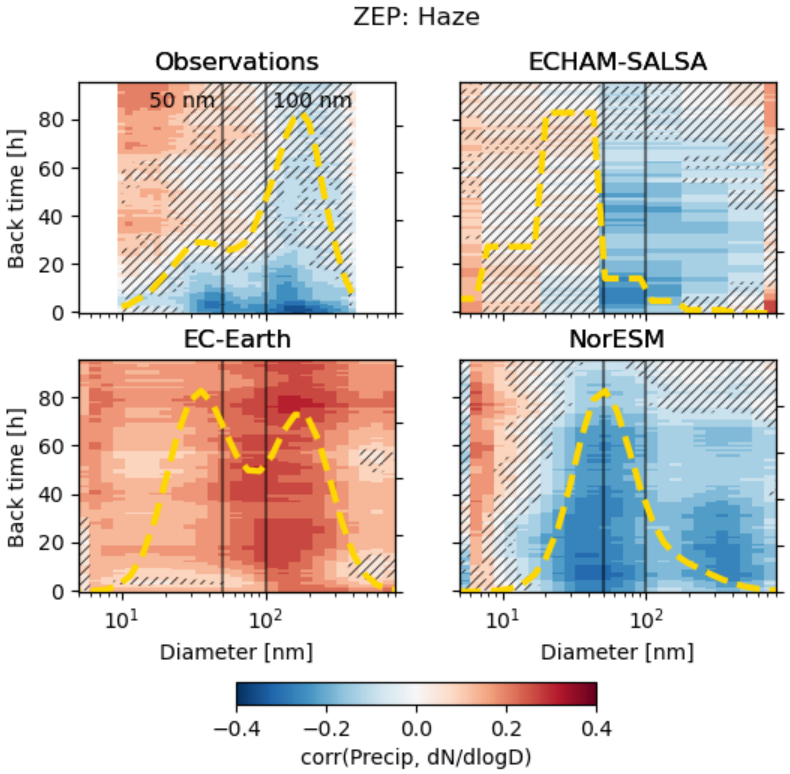


Figure S36. Correlation between rainfall and PNSD, Haze, Zeppelin, calculated based on daily mean vales.

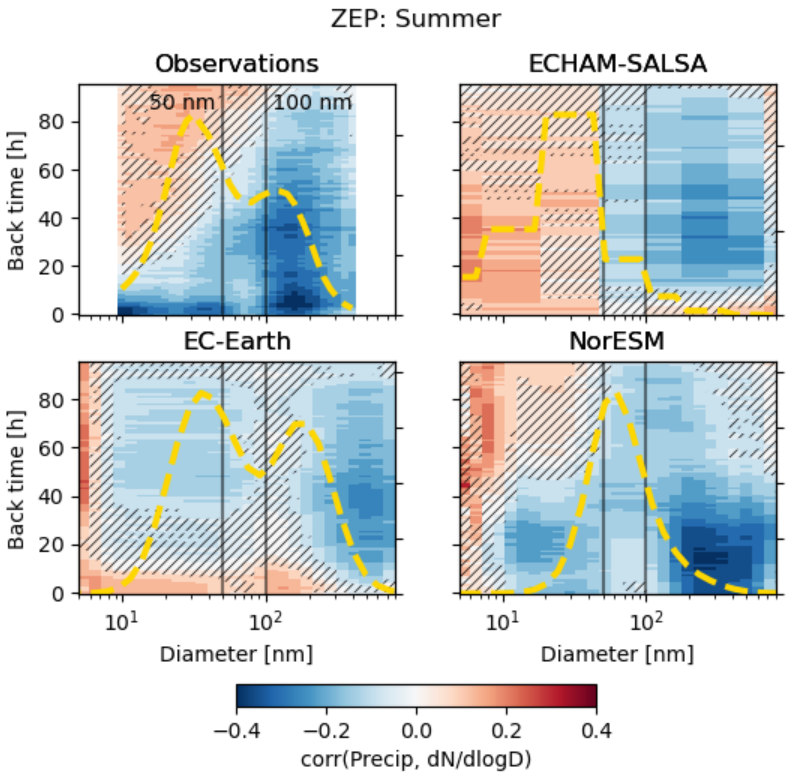


Figure S37. Correlation between rainfall and PNSD, Summer, Zeppelin, calculated based on daily mean vales.

S10.3 ATTO

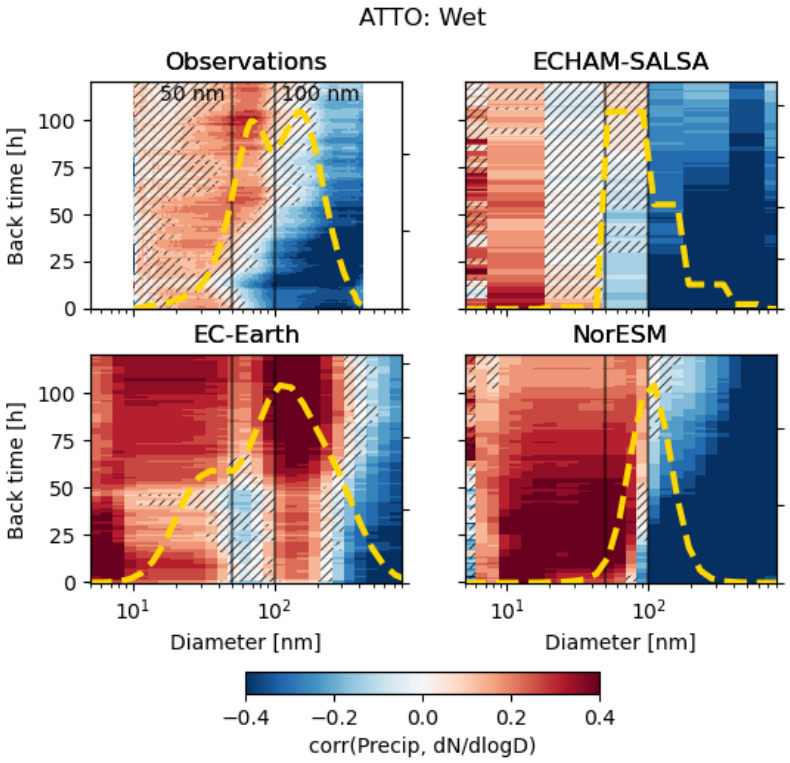


Figure S38. Correlation between rainfall and PNSD, Wet season, ATTO, calculated based on daily mean vales.

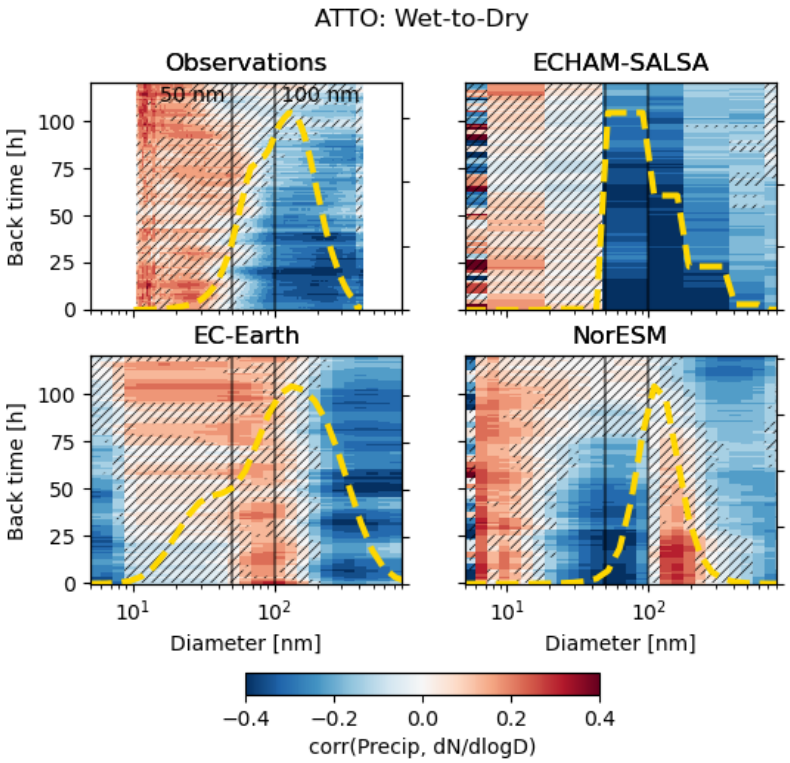


Figure S39. Correlation between rainfall and PNSD, Wet-to-Dry season, ATTO, calculated based on daily mean vales.

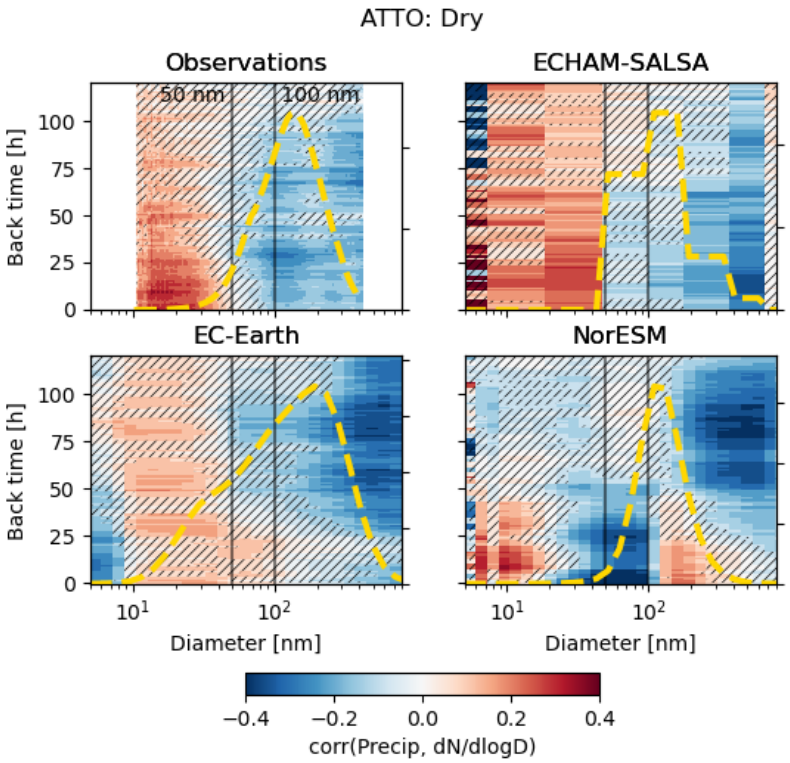


Figure S40. Correlation between rainfall and PNSD, Dry season, ATTO, calculated based on daily mean vales.

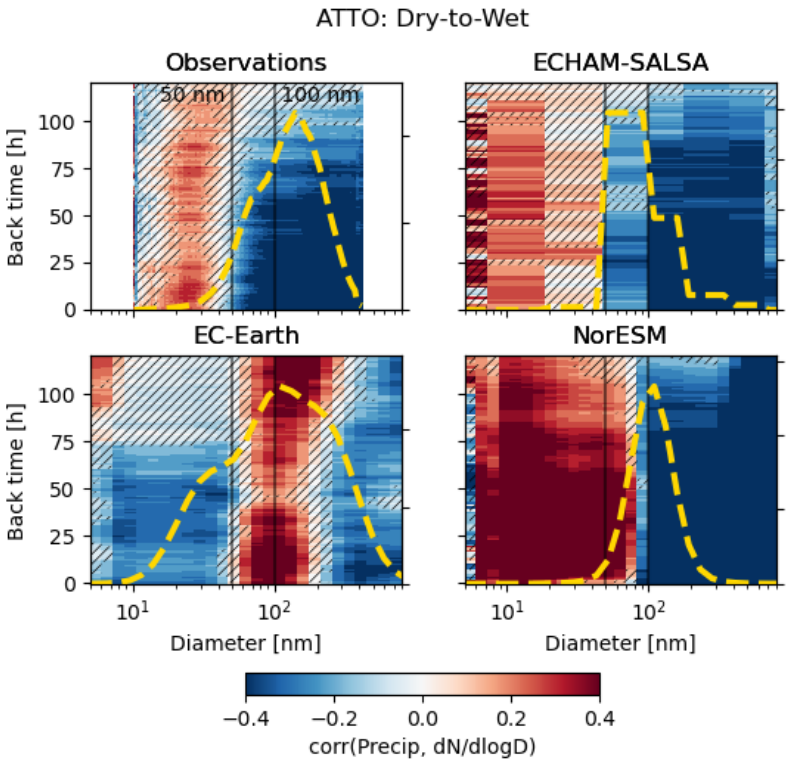


Figure S41. Correlation between rainfall and PNSD, Dry-to-Wet season, ATTO, calculated based on daily mean vales.

S11 Correlations for number concentrations in size ranges versus precipitation

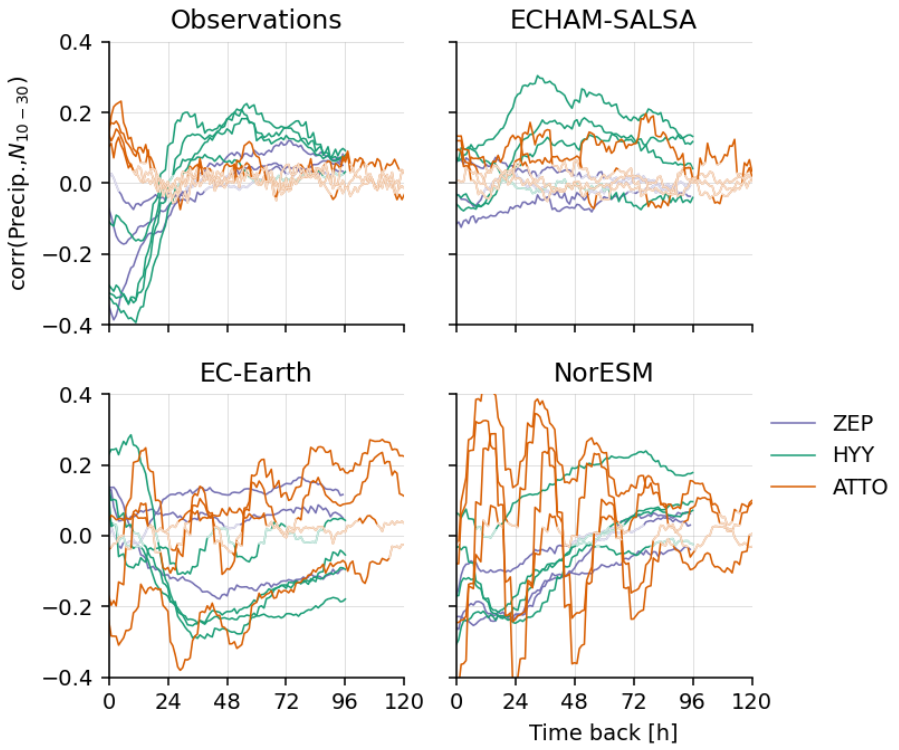


Figure S42. Spearman correlations between precipitation rate and N_{10-30} for observations (a) and models (b-d). Each line represents a season at a station.

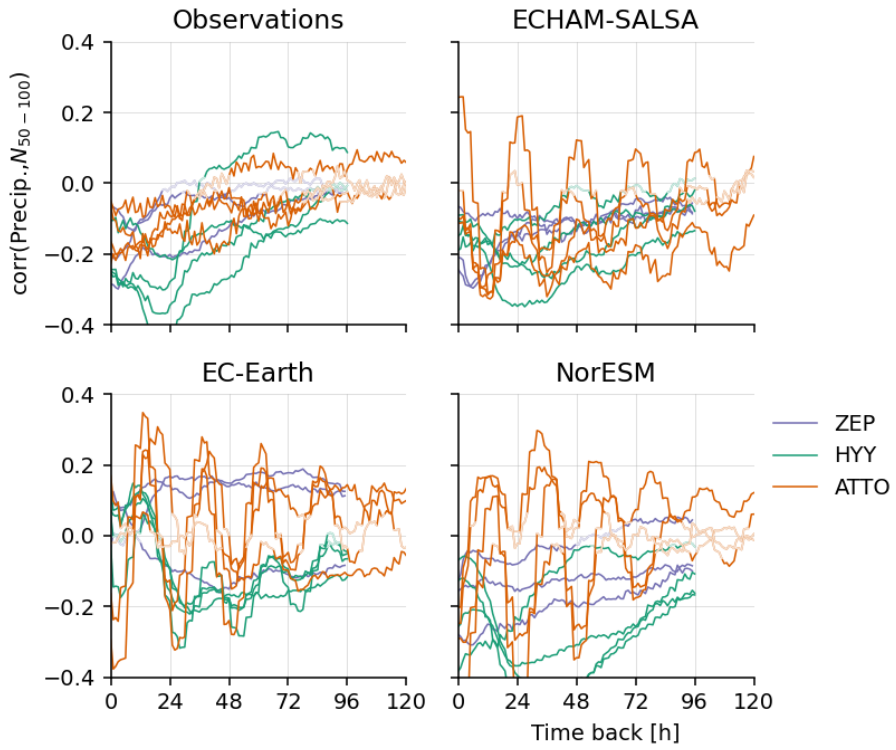


Figure S43. Spearman correlations between precipitation rate and N_{50-100} for observations (a) and models (b-d). Each line represents a season at a station.

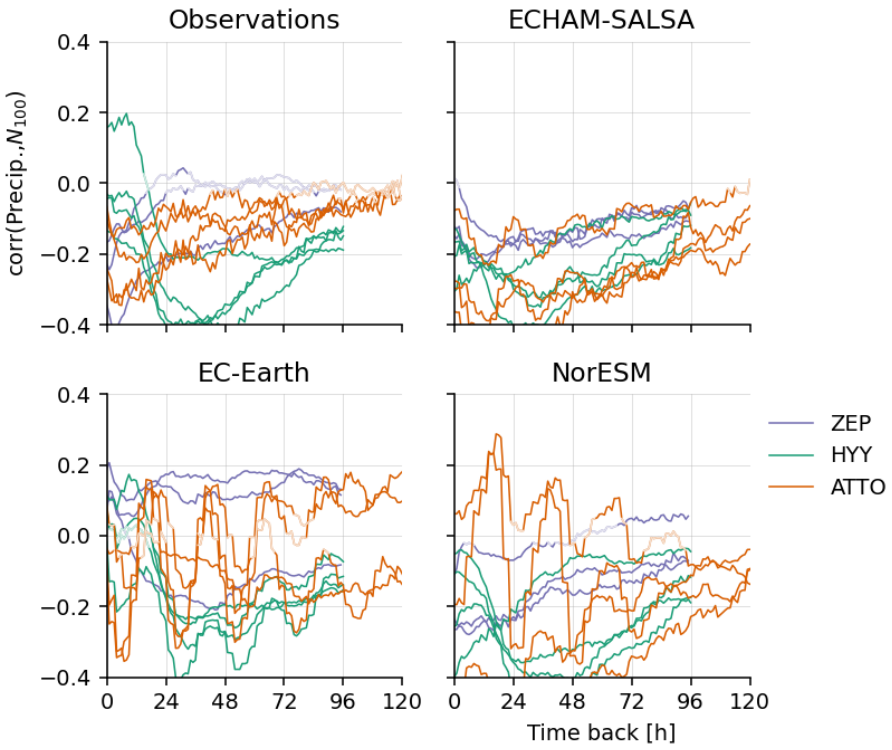


Figure S44. Spearman correlations between precipitation rate and N_{100} for observations (a) and models (b-d). Each line represents a season at a station.

S12 XGBoost SHAP values

SHAP (SHapley Additive exPlanations) values are a method from game theory used to explain the output of machine learning models. They quantify how much each input feature contributes to a specific prediction. In the below, we present SHAP value analysis for the XGBoost regressions created for the analysis in the manuscript.

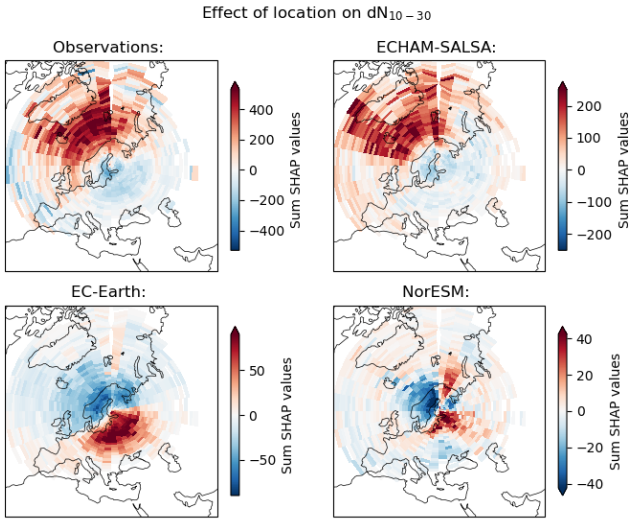


Figure S45. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{10-30} for Hyytiälä station. They are calculated by for each time, the sum of the SHAP values for all passings through that grid box is calculated (independent of when before arriving at the station), and finally the average of these sums is calculated (over the time dimension).

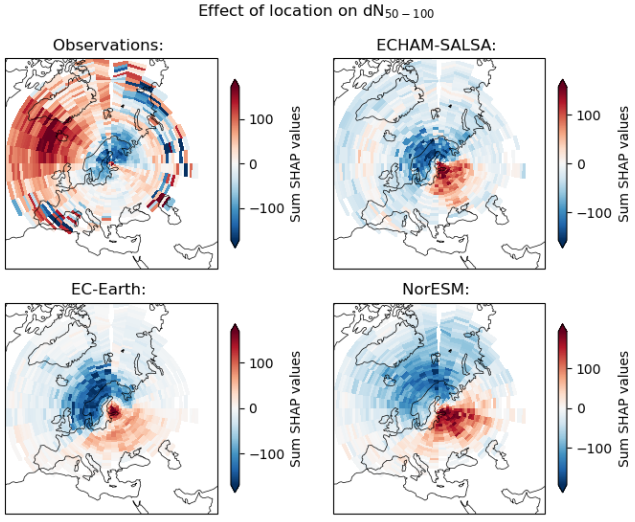


Figure S46. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{50-100} for Hyytiälä station. Please see caption of Fig. S45 for description of the calculation.

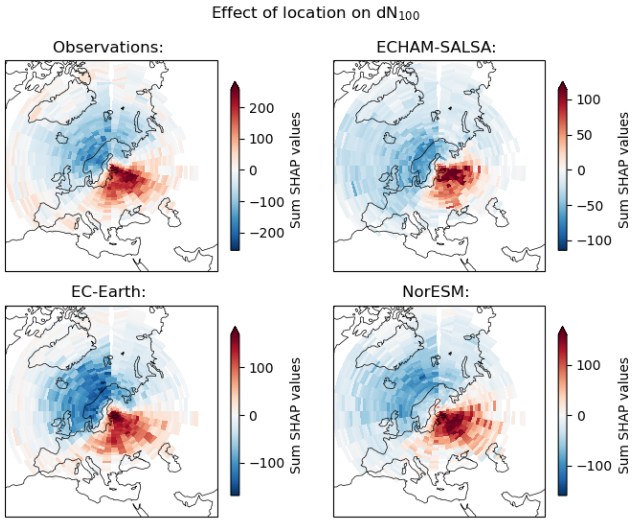


Figure S47. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{100} for Hyytiälä station. Please see caption of Fig. S45 for description of the calculation.

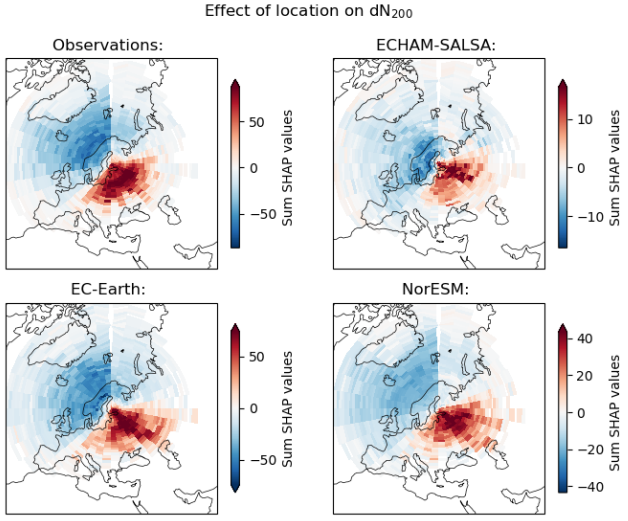


Figure S48. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{200} for Hyytiälä station. Please see caption of Fig. S45 for description of the calculation.

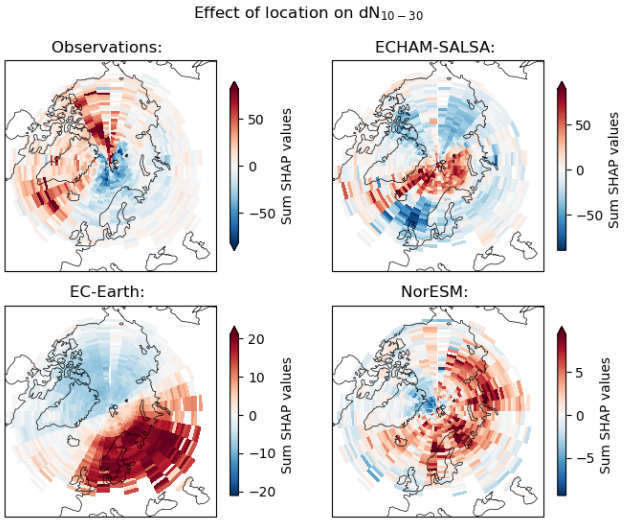


Figure S49. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{10-30} for Zeppelin station. Please see caption of Fig. S45 for description of the calculation.

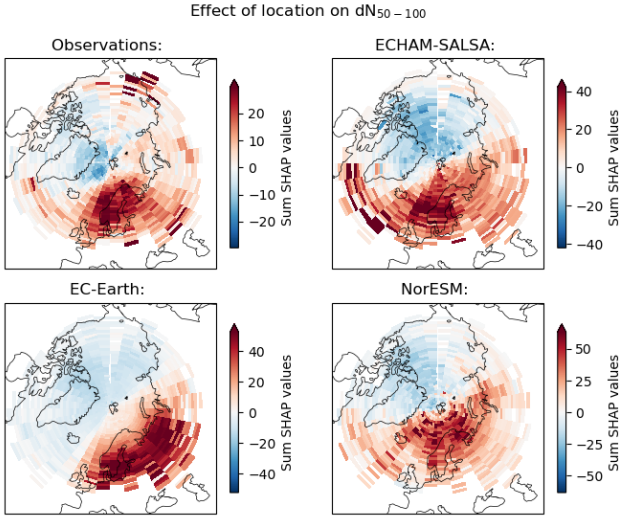


Figure S50. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{50-100} for Zeppelin station. Please see caption of Fig. S45 for description of the calculation.

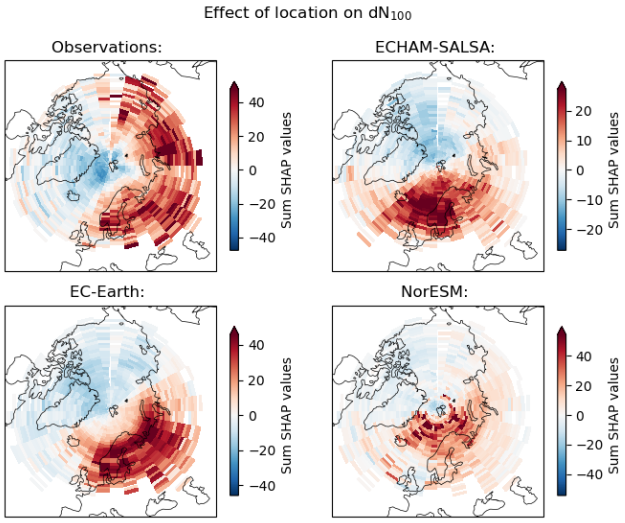


Figure S51. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{100} for Zeppelin station. Please see caption of Fig. S45 for description of the calculation.

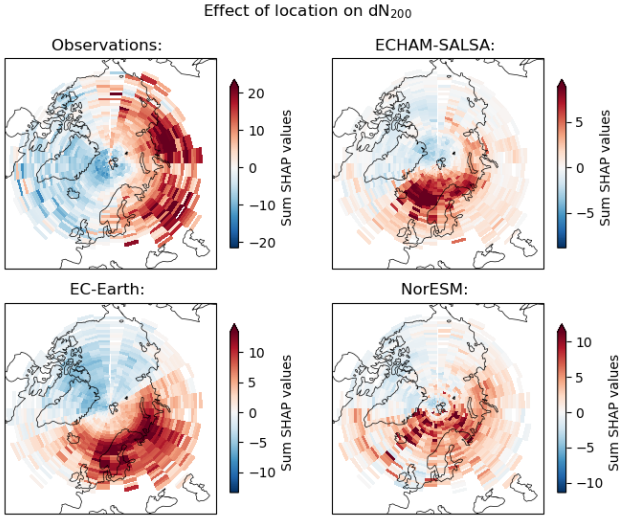


Figure S52. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{200} for Zeppelin station. Please see caption of Fig. S45 for description of the calculation.

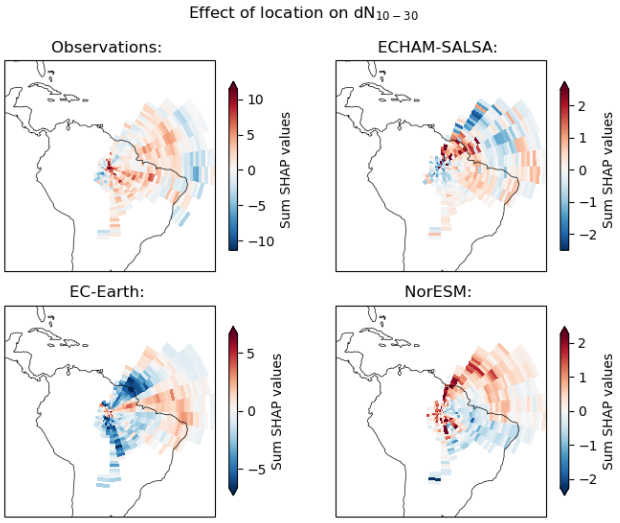


Figure S53. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{10-30} for ATTO station. Please see caption of Fig. S45 for description of the calculation.

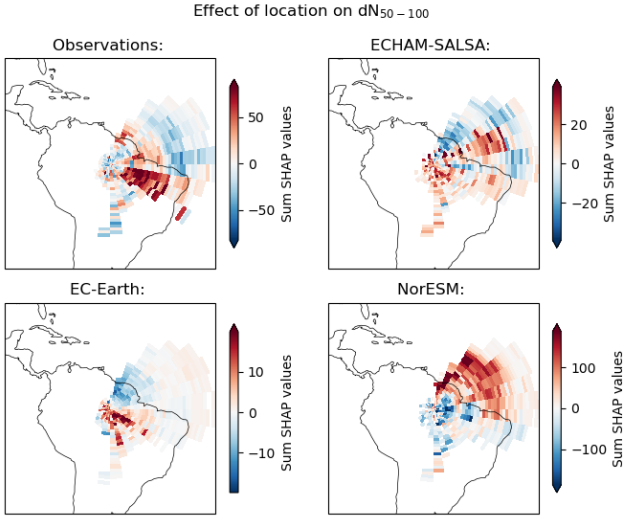


Figure S54. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{50-100} for ATTO station. Please see caption of Fig. S45 for description of the calculation.

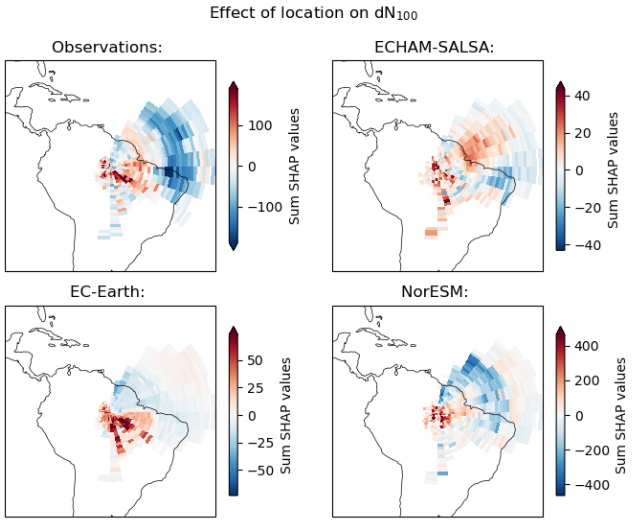


Figure S55. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{100} for ATTO station. Please see caption of Fig. S45 for description of the calculation.

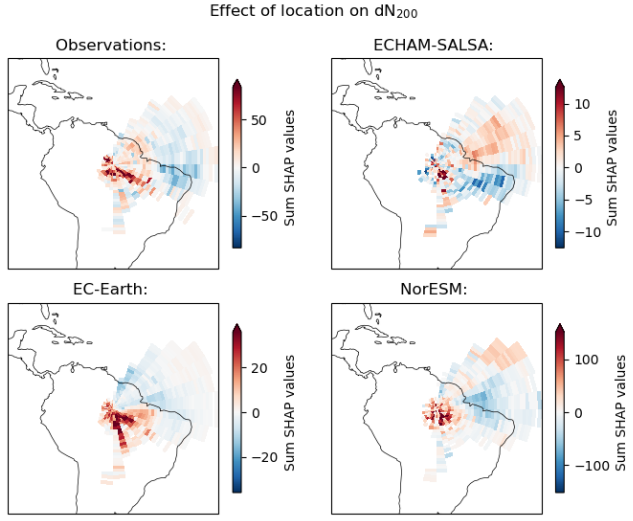


Figure S56. These maps show the "impact" in terms of SHAP values of a trajectory passing through each of grid cells on its way to the station on N_{200} for ATTO station. Please see caption of Fig. S45 for description of the calculation.

S12.1 SHAP values for recent (6h) precipitation

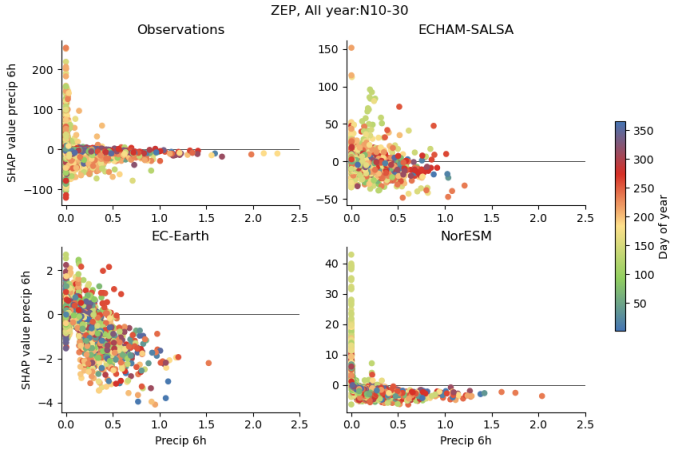


Figure S57. Zeppelin: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{10-30}. The color indicated the day of the year.

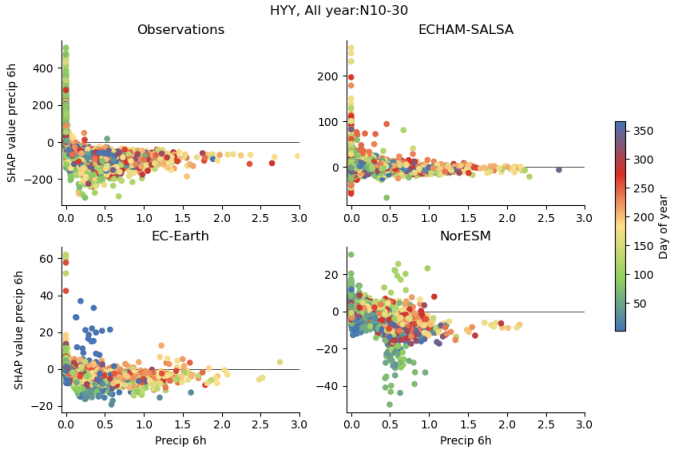


Figure S58. Hyytiälä: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{10-30}. The color indicated the day of the year.

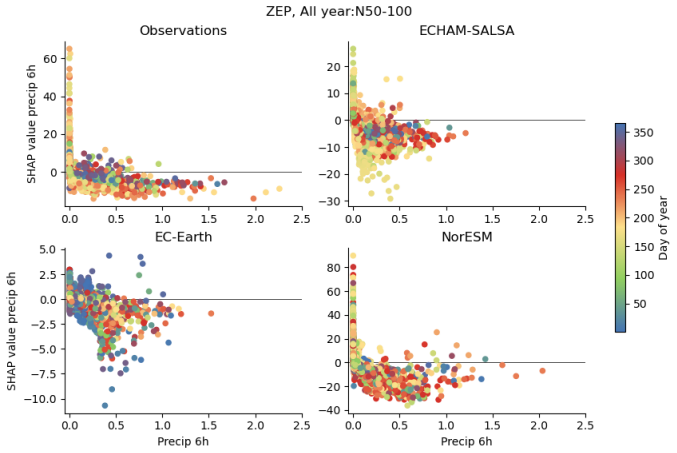


Figure S59. Zeppelin: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{50-100}. The color indicated the day of the year.

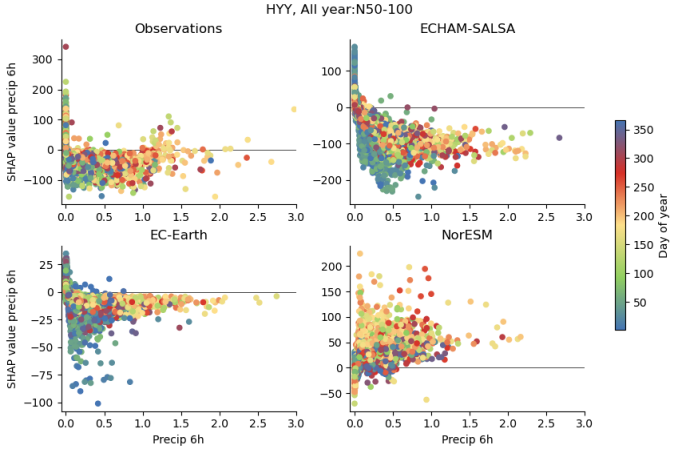


Figure S60. Hyytiälä: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{50-100}. The color indicated the day of the year.

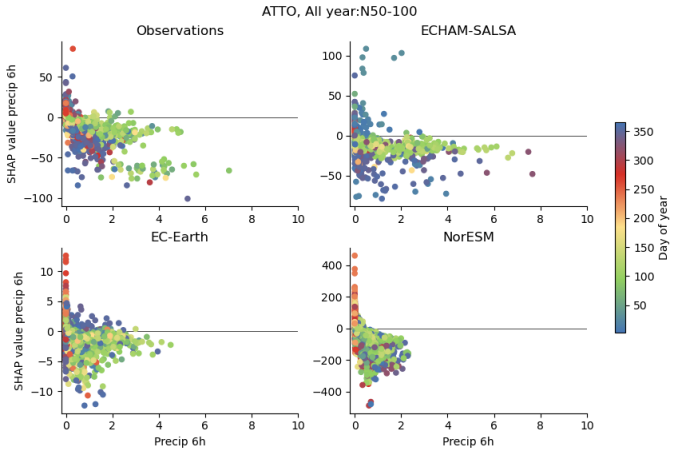


Figure S61. ATTO: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{50-100}. The color indicated the day of the year.

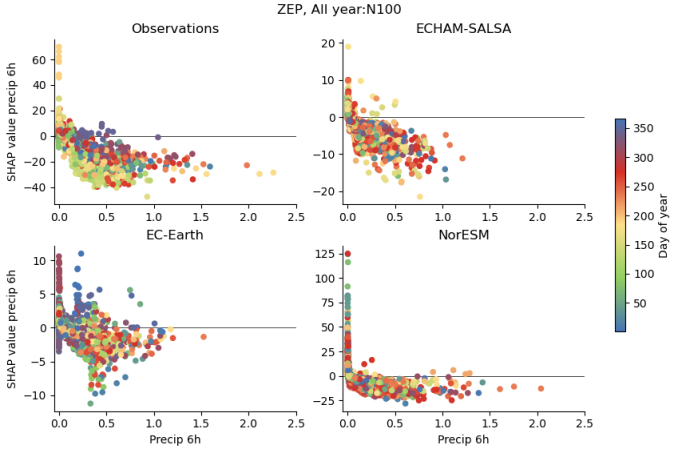


Figure S62. Zeppelin: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{100}. The color indicated the day of the year.

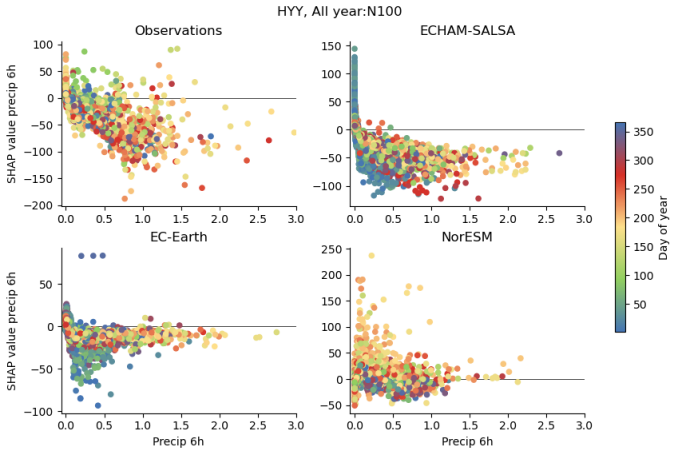


Figure S63. Hyytiälä: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{100}. The color indicated the day of the year.

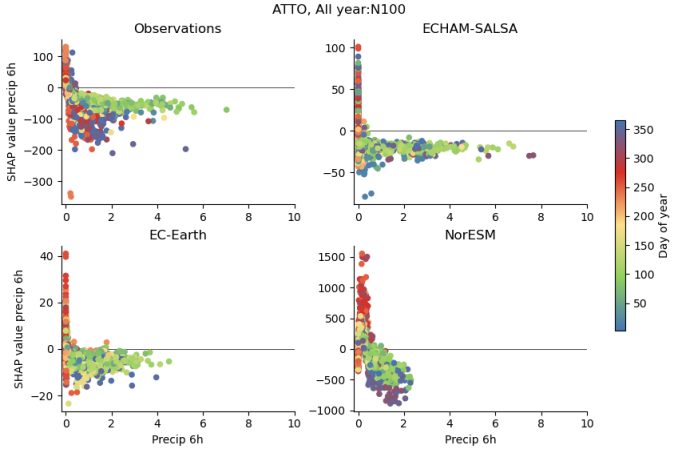


Figure S64. ATTO: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{100}. The color indicated the day of the year.

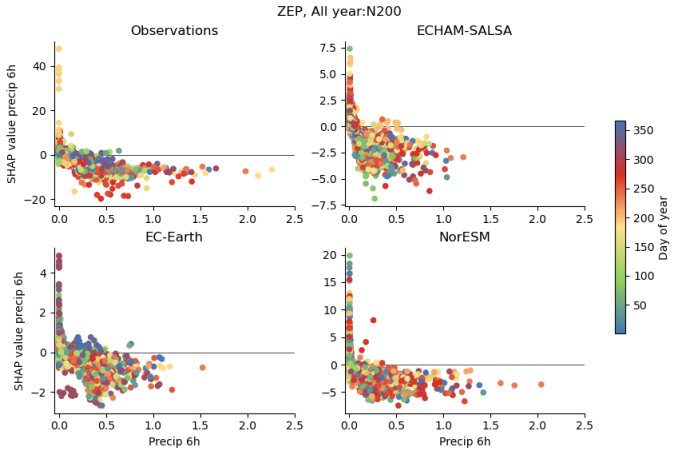


Figure S65. Zeppelin: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{200}. The color indicated the day of the year.

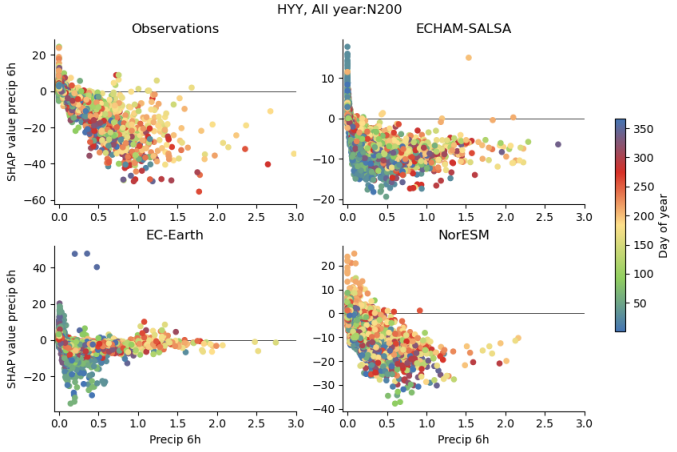


Figure S66. Hyytiälä: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{200}. The color indicated the day of the year.

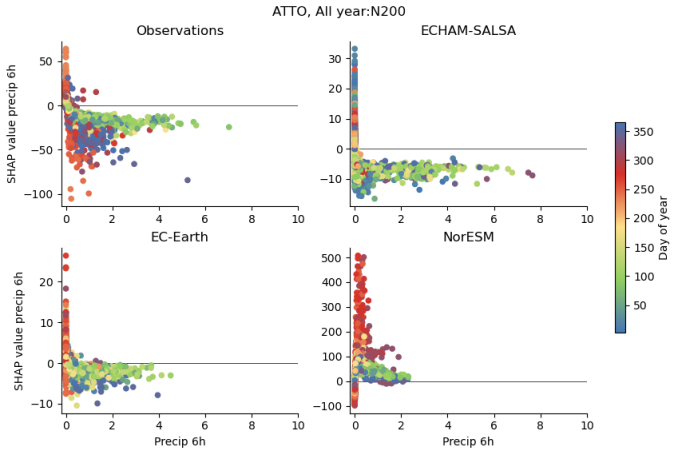


Figure S67. ATTO: Scatter plots showing the SHAP values for of the first 6h average precipitation rate prior to the airmass reaching the station for predicting N_{200}. The color indicated the day of the year.

S13 XGBoost evaluation

In the plots below, we have used all except the last year (2018) for training of the XGBoost, keeping the other parameters the same as in the main text (see the methods section). The results are thus for predictions for the testing year 2018.

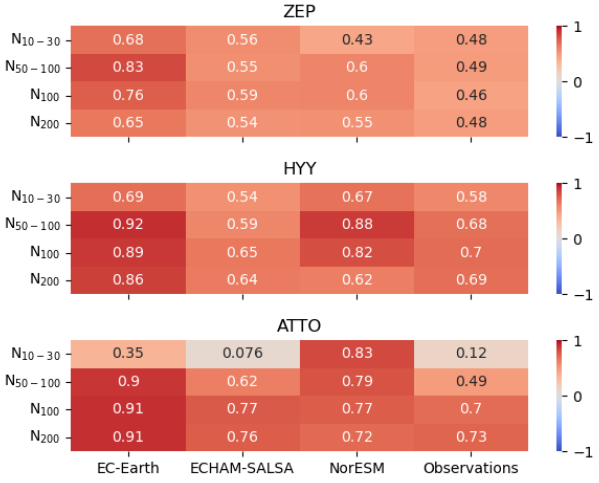


Figure S68. Pearson correlation coefficients for XGBoost modelled results versus observed results for 2018 for each station and size interval.

S13.1 Zeppelin

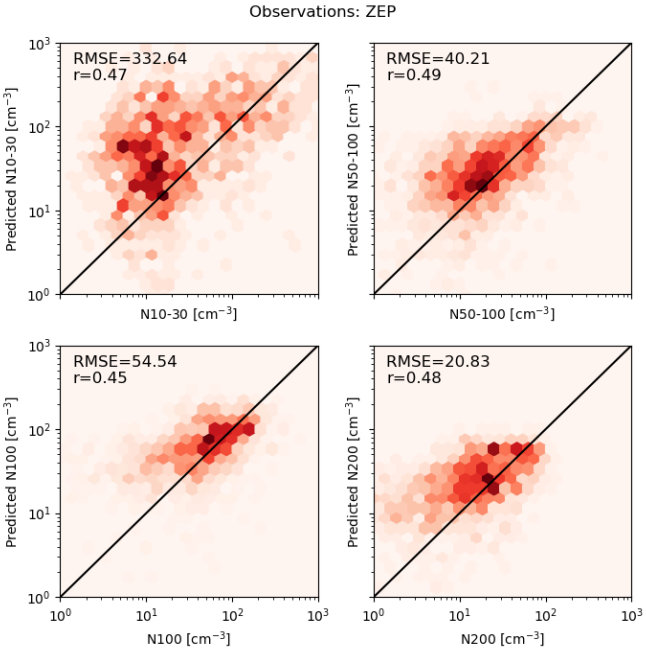


Figure S69. Evaluation of XGBoost model for Zeppelin station and observations/reanalysis.

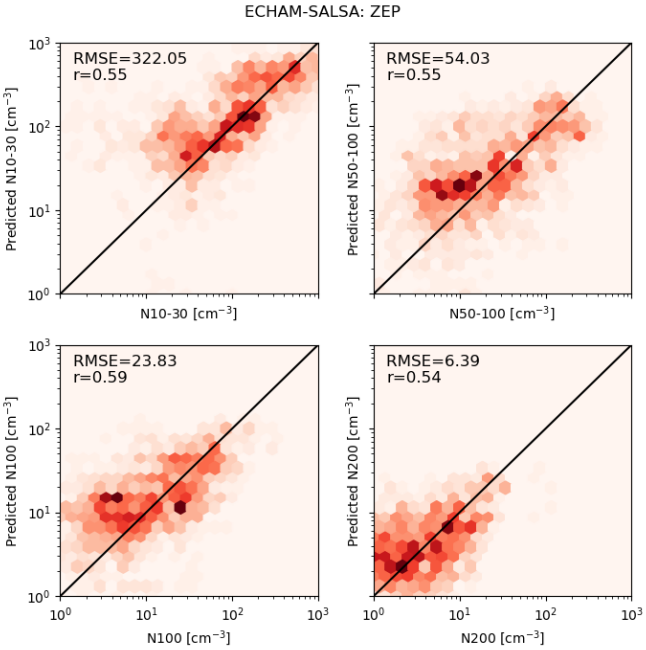


Figure S70. Evaluation of XGBoost model for Zeppelin station and ECHAM-SALSA.

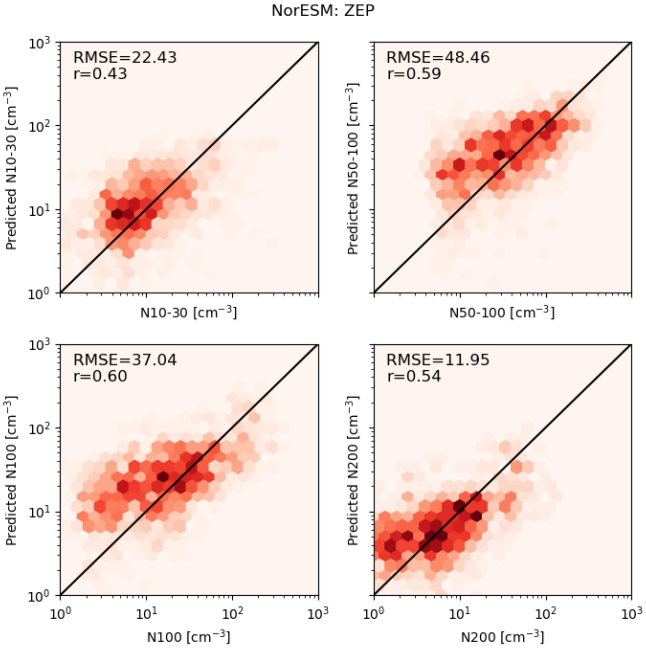


Figure S71. Evaluation of XGBoost model for Zeppelin station and NorESM.

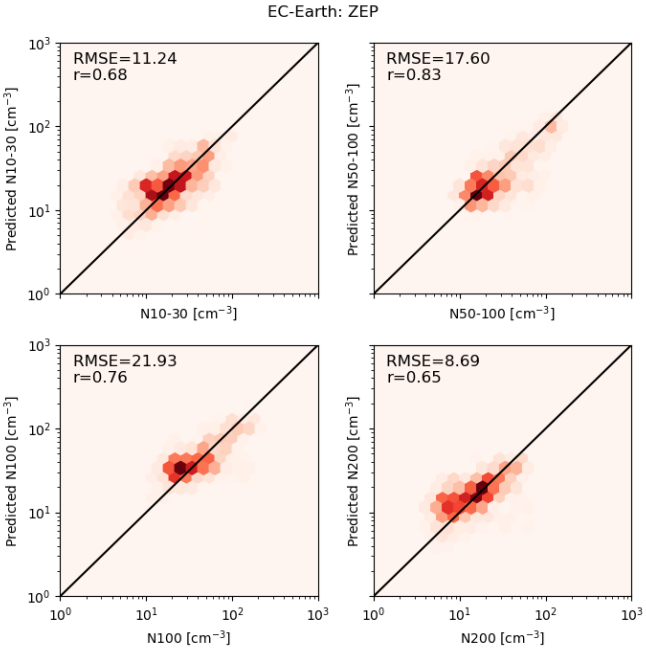


Figure S72. Evaluation of XGBoost model for Zeppelin station and EC-Earth.

S13.2 Hyytiälä

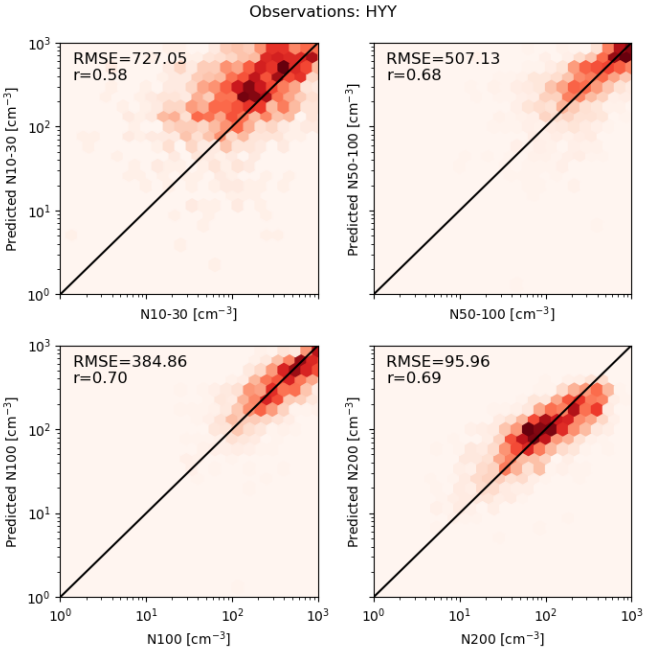


Figure S73. Evaluation of XGBoost model for Hyytiälä station and observations/reanalysis.

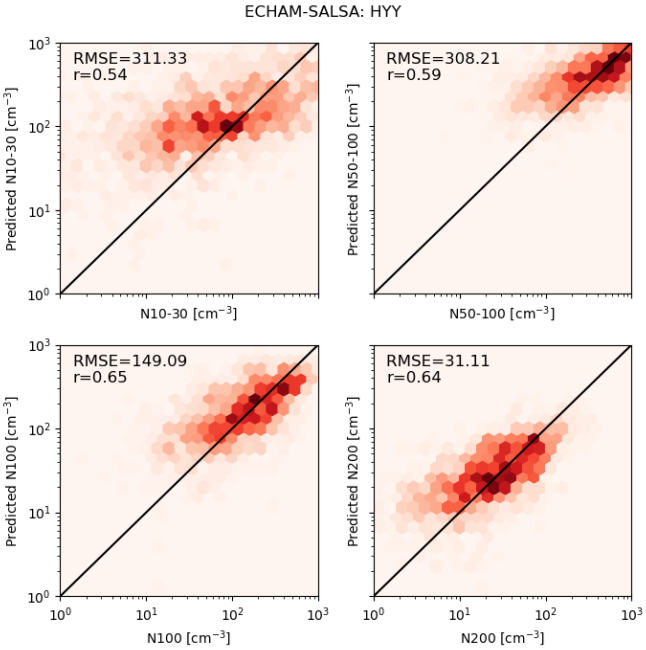


Figure S74. Evaluation of XGBoost model for Zeppelin station and ECHAM-SALSA.

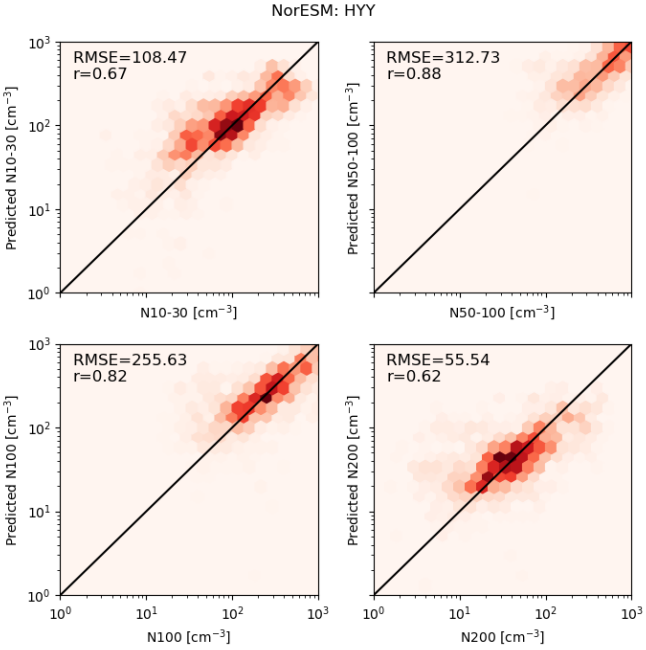


Figure S75. Evaluation of XGBoost model for Hyytiälä station and NorESM.

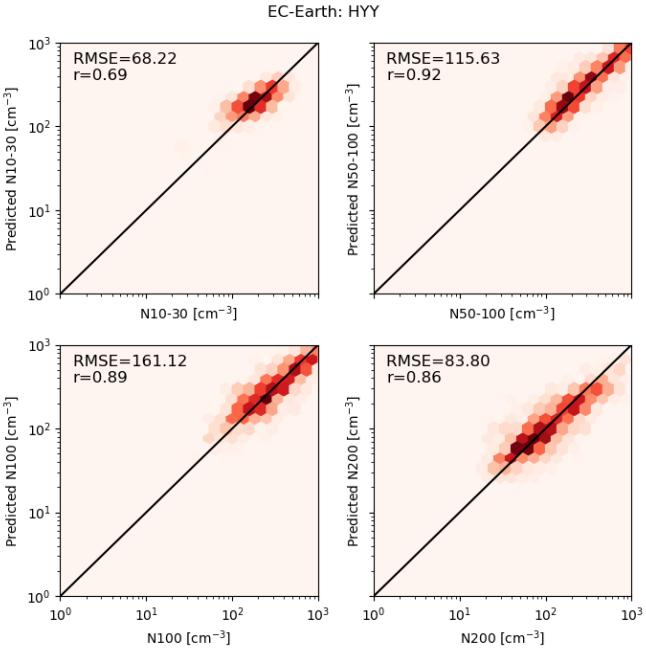


Figure S76. Evaluation of XGBoost model for Hyytiälä station and EC-Earth.

S13.3 ATTO

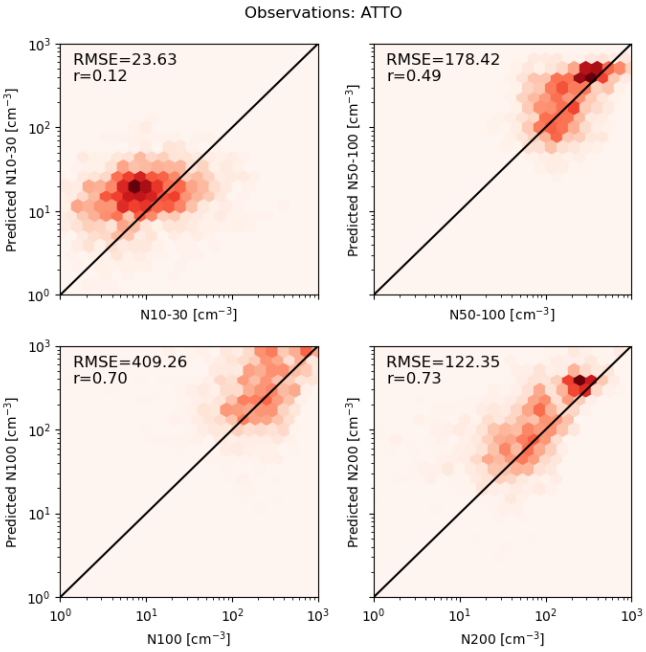


Figure S77. Evaluation of XGBoost model for ATTO station and observations/reanalysis.

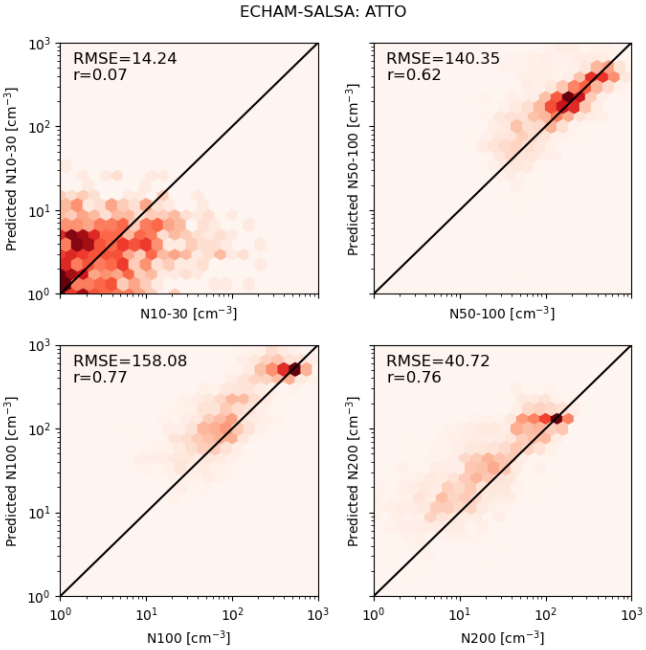


Figure S78. Evaluation of XGBoost model for ATTO station and ECHAM-SALSA.

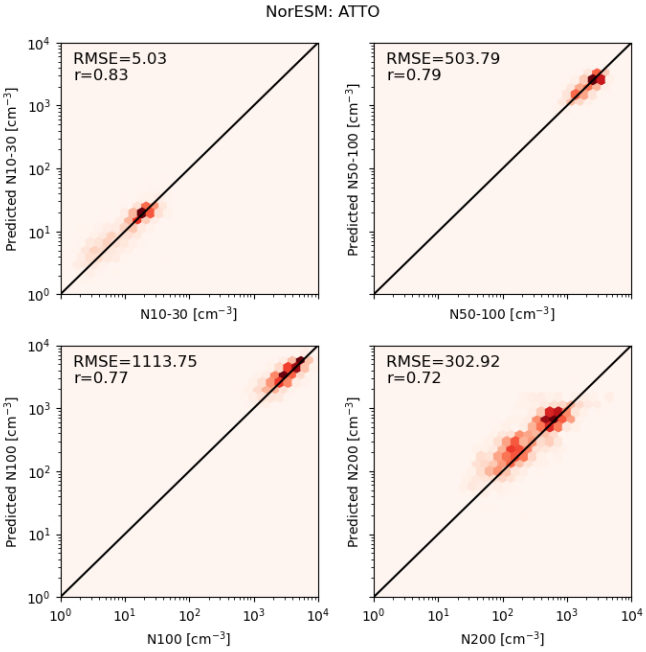


Figure S79. Evaluation of XGBoost model for ATTO station and NorESM.

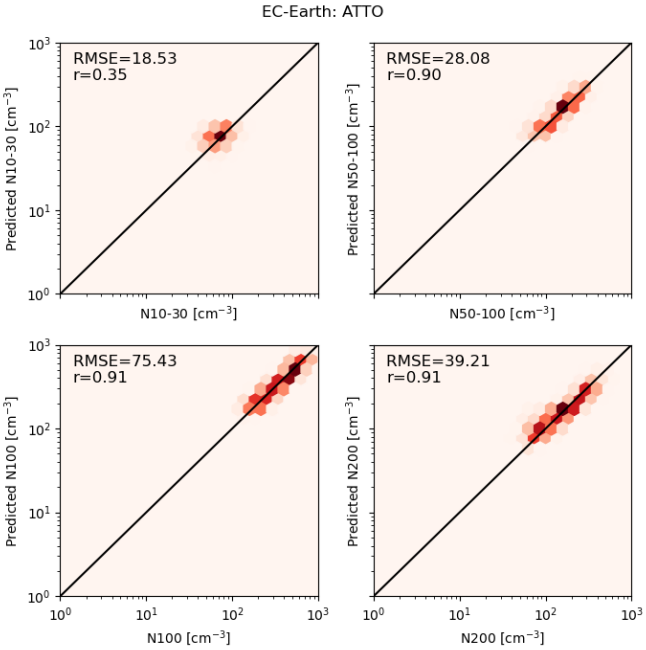


Figure S80. Evaluation of XGBoost model for ATTO station and EC-Earth.