Supplement of Atmos. Chem. Phys., 22, 11823–11843, 2022
https://doi.org/10.5194/acp-22-11823-2022-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

The effect of clouds and precipitation on the aerosol concentrations and composition in a boreal forest environment

Sini Isokääntä et al.

Correspondence to: Sini Isokääntä (sini.isokaanta@uef.fi) and Annele Virtanen (annele.virtanen@uef.fi)

S1 Data coverage and instrumentation

Table S1 shows the data coverage in % as available hours divided with total hours. Total hours include maximum of 24 data points per day (hourly data), from which points with wind direction 120°-140° and if the trajectory had crossed Kola peninsula were removed before calculating the data coverage. Table S2 shows summary statistics for the measured variables, calculated from the available data. Please see manuscript Section 2.1 for the explanations of the variable abbreviations.

Table S3 shows the instrumental details for each investigated variable.

Table S1 Data coverage (in %) for the studied observations between January 2005 and August 2019.

Variable	**Coverage**				**notes**
	DJF	MAM	JJA	SON	
Org, SO_4, NH_4, Chl and NO_3	20.0 (37.9)	33.4 (62.8)	40.1/36.0 (74.2/66.6)	28.5 (59.0)	Values in brackets () show the coverage for 2012-19, for JJA the latter value after "/" shows coverage for Chl and NH_4
T	100.0	100.0	98.7	98.9	
WS	100.0	100.0	99.7	98.0	
NOx	96.7	97.7	94.8	98.0	
CO	88.4	97.1	94.8	98.1	
SO_2	98.1	94.0	95.8	97.1	
RH	72.8	80.7	81.2	80.9	
SolR	97.4	99.9	99.4	100.0	
WD	100.0	100.0	100.0	100.0	
p	100.0	99.7	100.0	100.0	
eBC	64.8	79.1	66.4	72.9	Measurements start at January 2006
$rain_{local}$	94.0	93.2	93.4	99.6	

Table S2 Median and lower and upper quartiles (Q1 and Q3, respectively) for the measured variables for each season after removing data rows with wind direction between 120°-140° and data rows where trajectory crosses Kola peninsula. Data corresponds to Table S1.

Variable **Statistics for each season. Med. = Median.**

	DJF			MAM			JJA			SON		
	Q1	Med.	Q3	Q1	Med.	Q3	Q1	Med.	Q3	Q1	Med.	Q3
Org ($\mu g\ m^{-3}$)	0.45	0.88	1.64	0.64	1.27	2.41	1.32	2.21	3.68	0.56	1.03	1.86
NH_4 ($\mu g\ m^{-3}$)	0.05	0.18	0.36	0.04	0.15	0.29	0.04	0.15	0.30	0.03	0.14	0.30
SO_4 ($\mu g\ m^{-3}$)	0.21	0.52	1.07	0.20	0.37	0.29	0.23	0.39	0.68	0.13	0.33	0.78
Chl ($\mu g\ m^{-3}$)	0.00	0.01	0.03	0.00	0.01	0.02	0.00	0.01	0.02	0.00	0.01	0.02
NO_3 ($\mu g\ m^{-3}$)	0.06	0.14	0.35	0.05	0.11	0.22	0.07	0.12	0.20	0.05	0.10	0.19
T (°C)	-7.60	-3.18	-0.27	-0.37	3.26	8.63	12.20	14.94	18.15	1.25	5.17	9.34
WS ($m\ s^{-1}$)	1.41	1.94	2.62	1.41	1.90	2.57	1.23	1.69	2.27	1.29	1.85	2.58
NOx (ppb)	0.72	1.41	2.54	0.28	0.59	1.21	0.19	0.35	0.67	0.40	0.80	1.50
CO (ppb)	150.44	168.75	204.92	143.74	157.51	171.90	106.94	121.02	136.61	120.98	137.26	155.96
SO_2 (ppb)	0.02	0.10	0.37	0.02	0.08	0.21	0.01	0.04	0.10	0.01	0.03	0.10
RH (%)	92.33	96.23	98.48	50.68	72.05	91.50	55.13	72.67	89.11	85.19	93.74	97.80
SolR ($W\ m^{-2}$)	-0.11	0.60	6.94	0.49	40.64	239.02	3.70	103.72	333.19	-0.03	1.35	37.91
WD (°)	149.09	200.49	255.73	157.21	217.97	272.50	157.87	212.23	263.30	160.68	206.98	253.40
p (hPa)	977.38	988.05	997.76	984.07	991.15	998.07	986.36	991.47	996.41	982.17	990.95	999.57
eBC ($\mu g\ m^{-3}$)	0.13	0.26	0.53	0.09	0.17	0.33	0.08	0.15	0.26	0.10	0.19	0.36
$rain_{local}$ ($mm\ h^{-1}$)	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table S3 List of the measured variables and the measurement devices.

Variable	**Measuring instrument/method**
Org, SO_4, NH_4, Chl and NO_3 ($\mu g\ m^{-3}$)	ACSM
T (°C)	Ventilated and shielded sensor (Pt-100)
WS ($m\ s^{-1}$)	Thies Ultrasonic Anemometer 2D
NOx (ppb)	TEI 42 CTL chemiluminescence analyzer
CO (ppb)	Horiba APMA 370 infrared light absorption analyzer
RH (%)	Calculated from dewpoint until 5/2012, Rotronic MP102H RH sensor since 6/2012
SolR ($W\ m^{-2}$)	Reemann TP 3 pyranometer, Middleton solar SK08 pyranometer, λ=0.3-4.8 µm, (at height of 18m/35m)
WD (°)	Thies Ultrasonic Anemometer 2D, used value is average of above canopy records
p (hPa)	Druck DPI 260 barometer, at ground level (180 m a.s.l.)
eBC ($\mu g\ m^{-3}$)	Magee Scientific Aethalometer with PM10 inlet (models AE-31 and AE-33, AE-31 data corrected by the algorithm suggested by Virkkula et al., 2007), measured at ground level, mass concentration estimated with λ=880nm
$rain_{local}$ ($mm\ h^{-1}$)	Vaisala FD12P weather sensor at 18 m height

S2 Supporting figures and tables

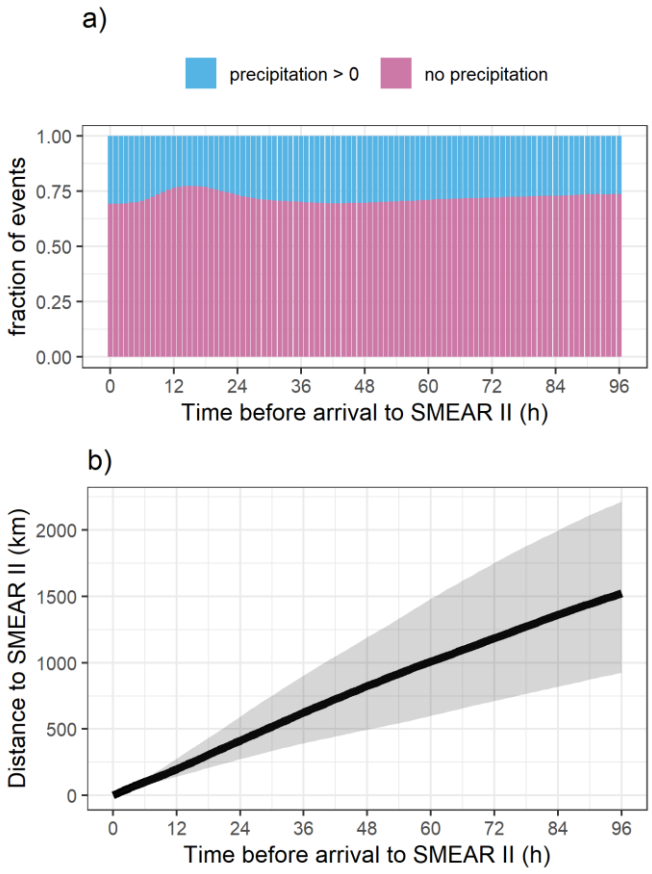

Figure S1 Fraction of precipitation events (based on ERA-Interim meteorology) along the trajectory is shown in a) and b) shows the median distance from SMEAR II with shaded area showing the 25th- 75th percentiles.

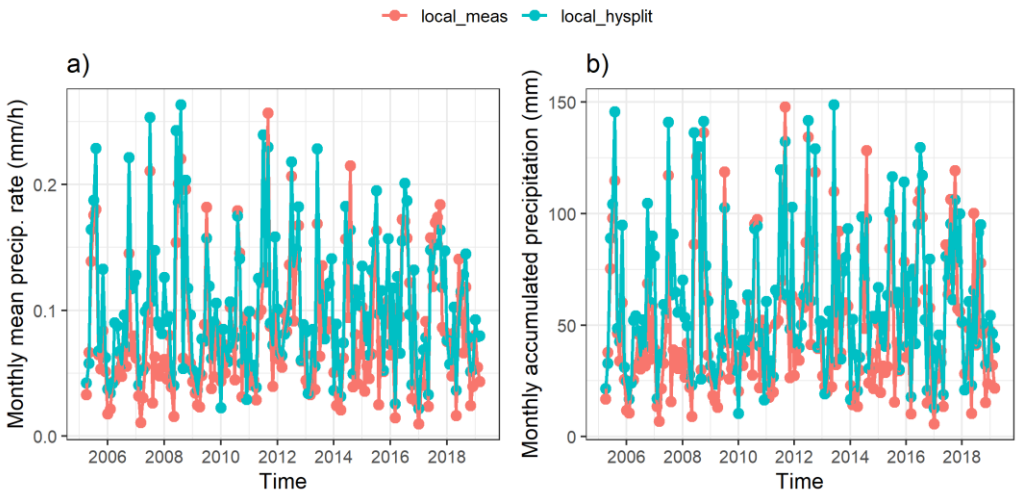

Figure S2 Precipitation time series of locally measured (red) and HYSPLIT estimate based on ERA-Interim meteorology (blue) as a a) monthly mean precipitation rate (mm/h) and b) monthly accumulated precipitation (mm) for SMEAR II. Spearman correlation coefficient between the two in a) is 0.87 and in b) 0.88.

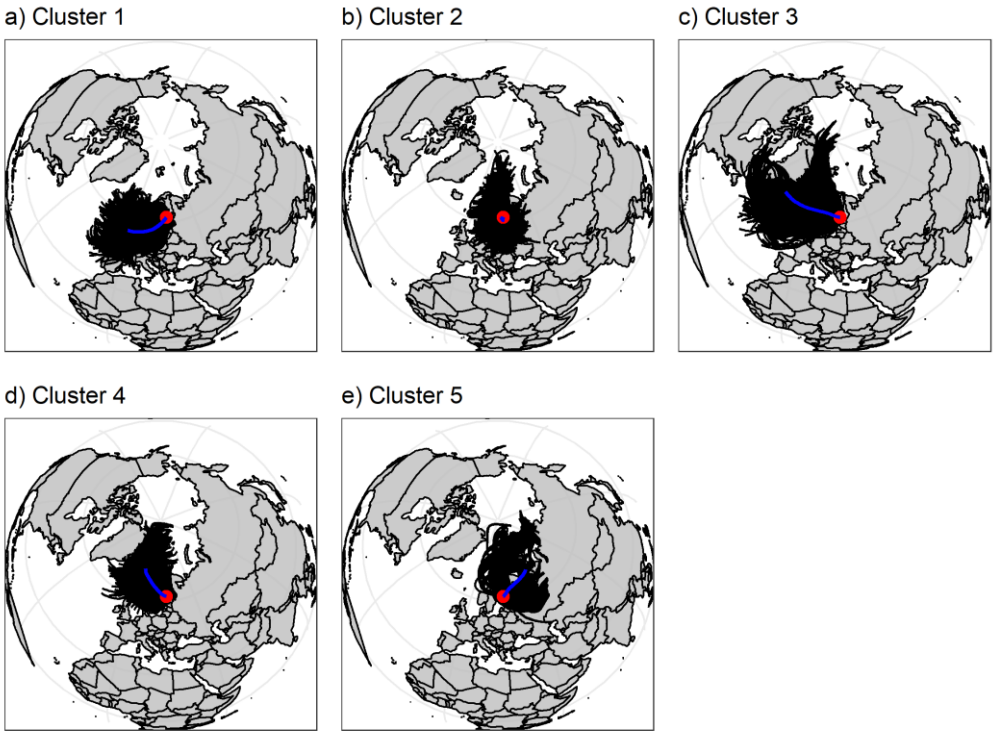

Figure S3 Geographical airmass source areas determined with the kmeans-clustering for the warm months. The red dot shows the location of SMEAR II and blue line is the cluster centre.

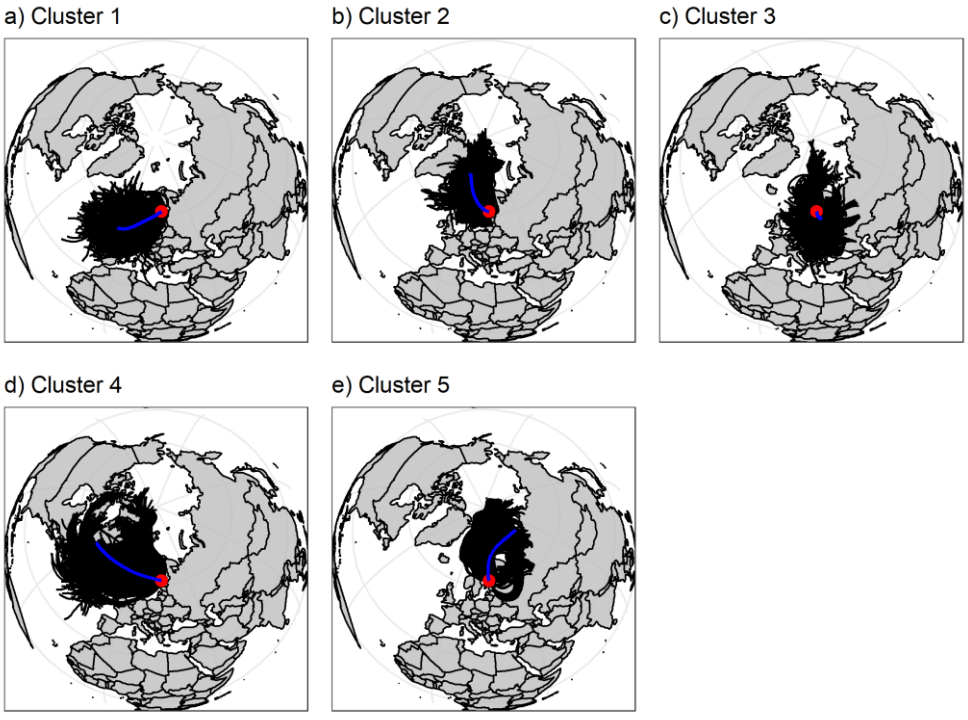

Figure S4 Geographical airmass source areas determined with the kmeans-clustering for the cold months. The red dot shows the location of SMEAR II and blue line is the cluster centre.

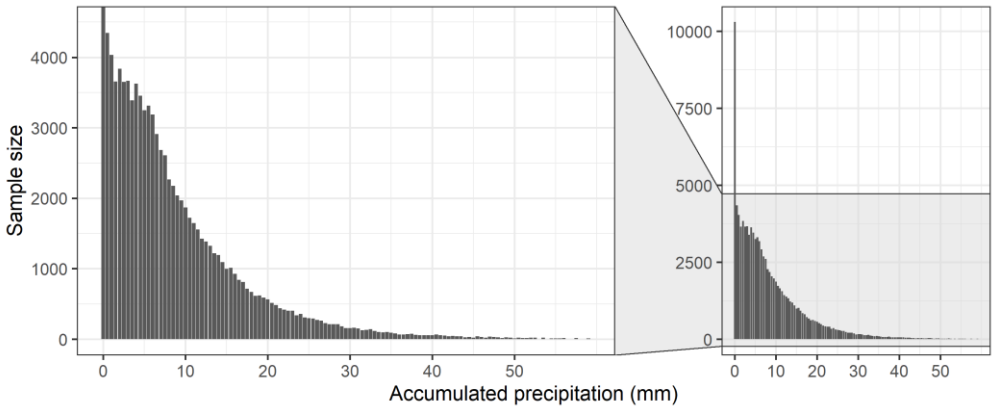

Figure S5 Sample size in each 0.5 mm bin of accumulated precipitation along the trajectory. Data based on DMPS observations corresponding to Fig. 1 in the main text.

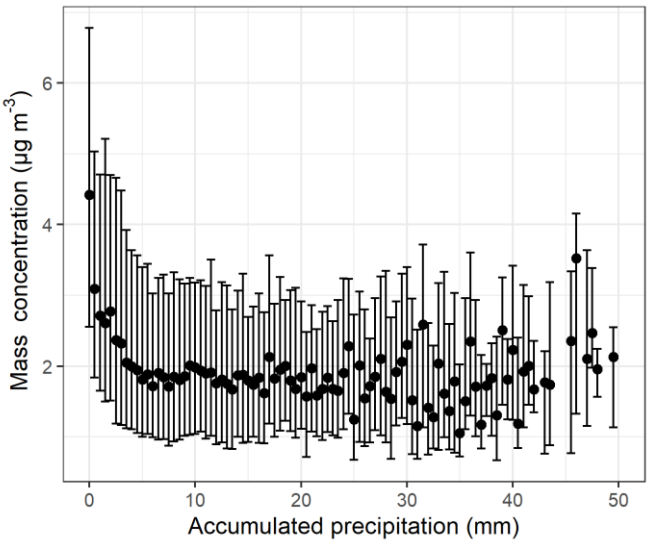

Figure S6 Particle mass (total sum of the constituents including Org, SO_4, NO_3, NH_4, Chl) evolution derived from the ACSM as a function of accumulated precipitation along the airmass trajectories. The black dots show the median values and bars highlight the 25th-75th percentiles for each 0.5 mm bin (with 10 or more data points) of accumulated precipitation.

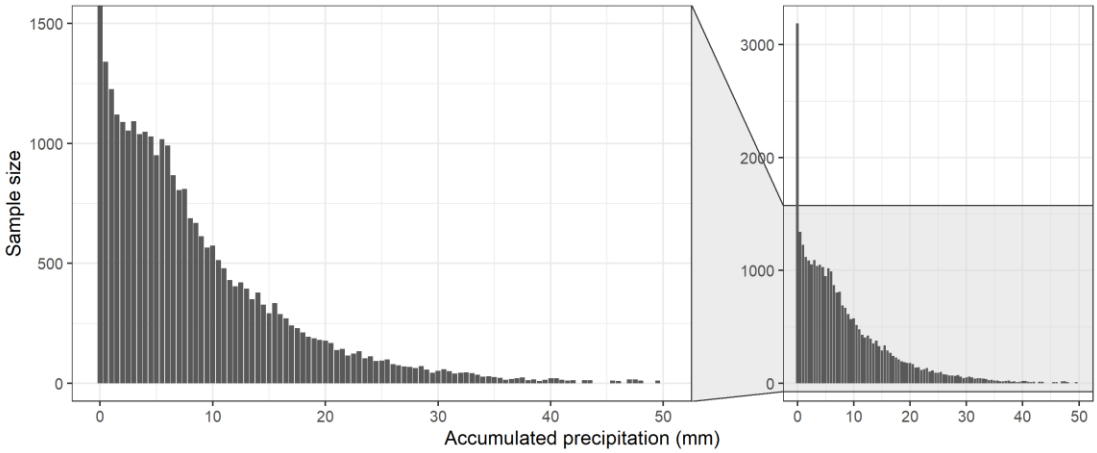

Figure S7 Sample size in each 0.5 mm bin of accumulated precipitation along the trajectory. Data based on ACSM observations, corresponding to Figure S6

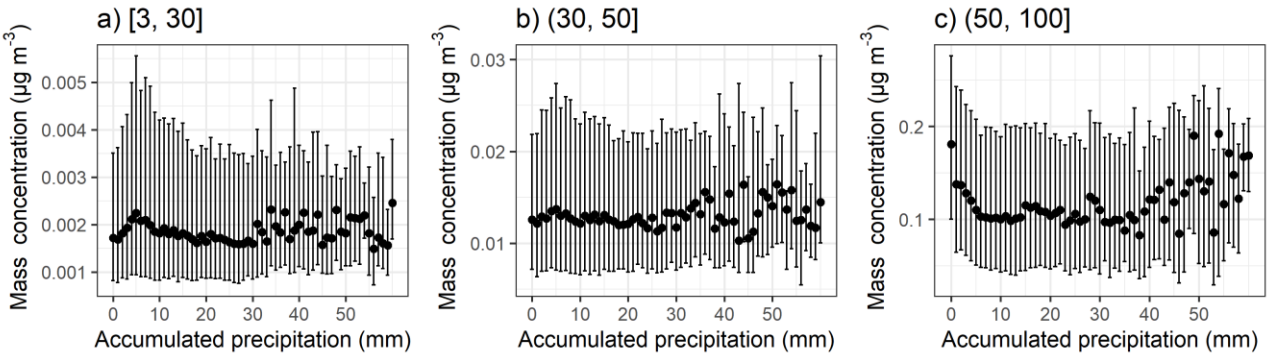

Figure S8 Particle number concentration as a function of 0-50 mm accumulated precipitation along the 96-hour HYSPLIT airmass trajectories for selected size ranges: (a) 3-50 nm, (b) 30-50 nm and (c) 50-100 nm. The black dots show the median values and bars highlight the 25th-75th percentiles for each 1 mm bin of accumulated precipitation. The figure includes DMPS data between January 2005 and August 2019.

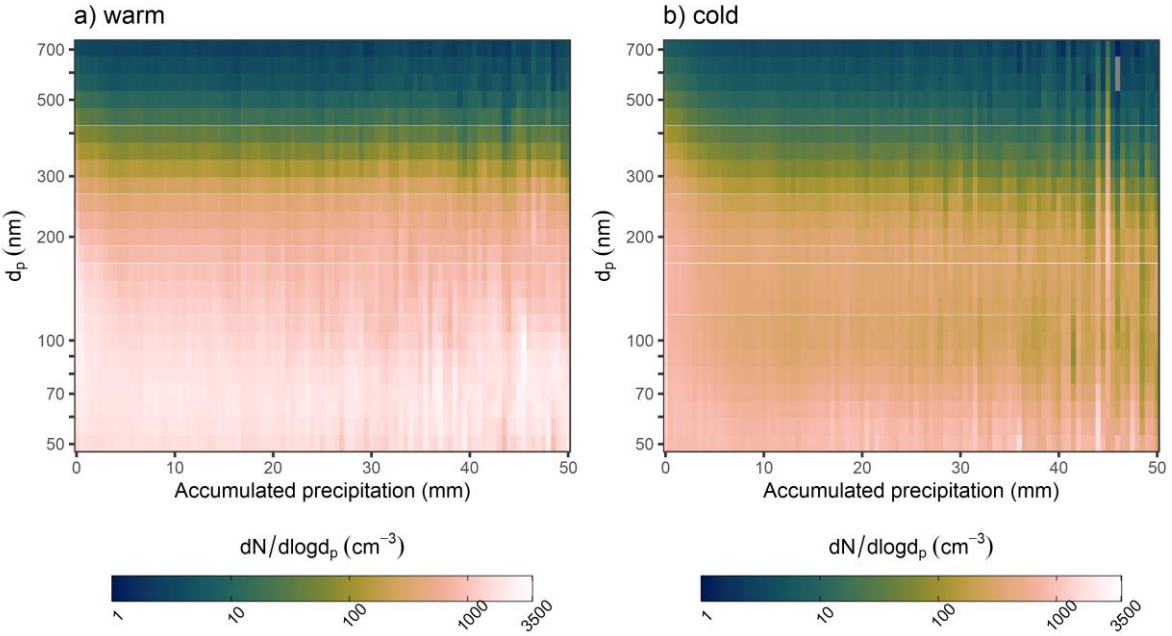

Figure S9 The aerosol size distribution evolution as a function of the 0-50 mm accumulated precipitation along the 96-hour airmass trajectories for a) warm and b) colder months for d_p = 50-700 nm (representing roughly the size range ACSM measures). Data is shown as medians for binned accumulated precipitation with bin size of 0.5 mm. Grey areas show missing data and/or values where dN/dlogd$_p$ < 1. The figure includes DMPS data between period of 2012-2019, which overlaps with the composition measurements.

Table S4 Statistical significance (p-values) of the differences between wet processing groups for particles in the accumulation mode (manuscript Fig. 5). P-values were obtained by comparing the data from the specified groups with Mann Whitney U test. Values smaller than 0.01 indicate the null hypothesis (distributions are identical) should be rejected.

Groups compared	Mass concentration	Number concentration
1 and 2	<0.001	<0.001
1 and 3	<0.001	<0.001

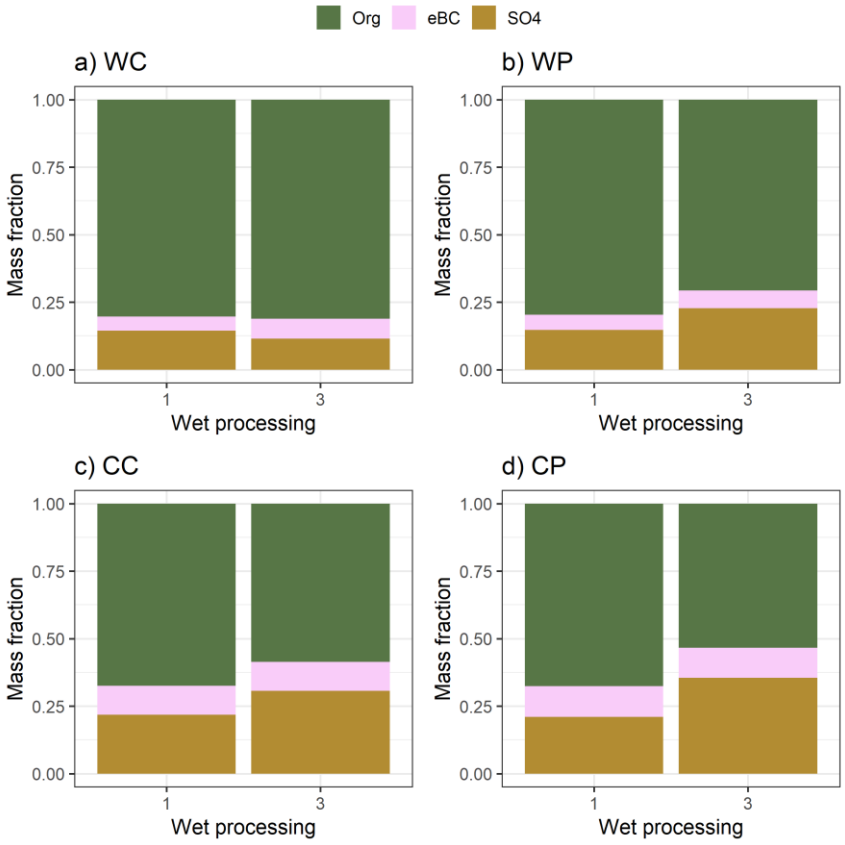

Figure S10 Chemical composition of the particles for each airmass history groups 1 and 3 as described in Table 2 (main text). Subplots show the airmass sectors (clean and polluted) with the seasonal (warm and cold) division following a) warm and clean, b) warm and polluted, c) cold and clean and b) cold and polluted. The figure includes simultaneous observations of these three species, spanning between March 2012 and August 2019.

Table S5 Statistical significance (p-values) of the differences between wet processing groups for Org, eBC and SO4 for each sector (manuscript Fig. 6). P-values were obtained by comparing the data from wet processing groups 1 and 3 with Mann Whitney U test. Values smaller than 0.01 indicate the null hypothesis (distributions are identical) should be rejected.

Species	**Sector**			
	WC	**WP**	**CC**	**CP**
Org	0.106	0.017	0.318	0.106
eBC	<0.001	<0.001	<0.001	<0.001
SO_4	<0.001	<0.001	<0.001	<0.001

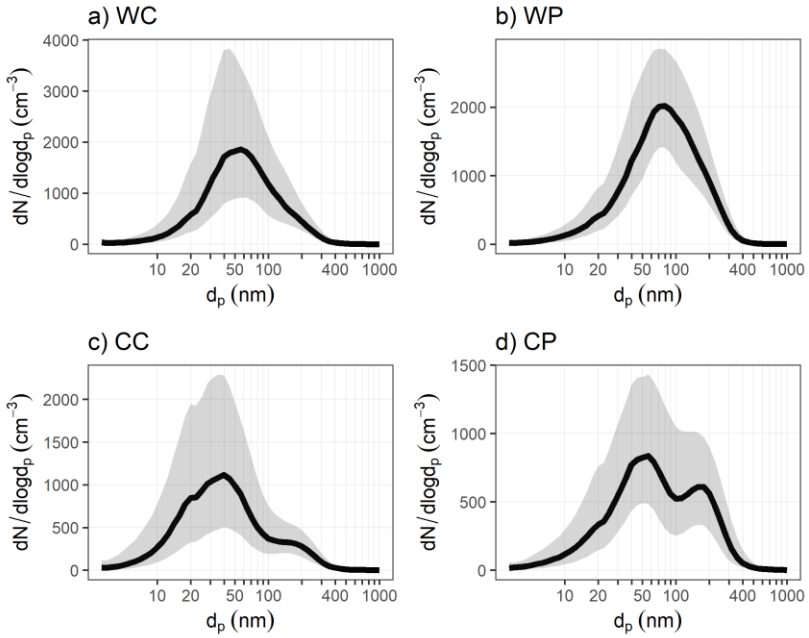

Figure S11 Median size distributions derived from DMPS for the period of 2012-2019, which overlaps with the composition measurements. Subplots show the airmass sectors (clean and polluted) with the seasonal (warm and cold) division: a) warm and clean, b) warm and polluted, c) cold and clean and b) cold and polluted. Please note that only particles larger than d_p = 50 nm are observed with the ACSM measuring the composition.

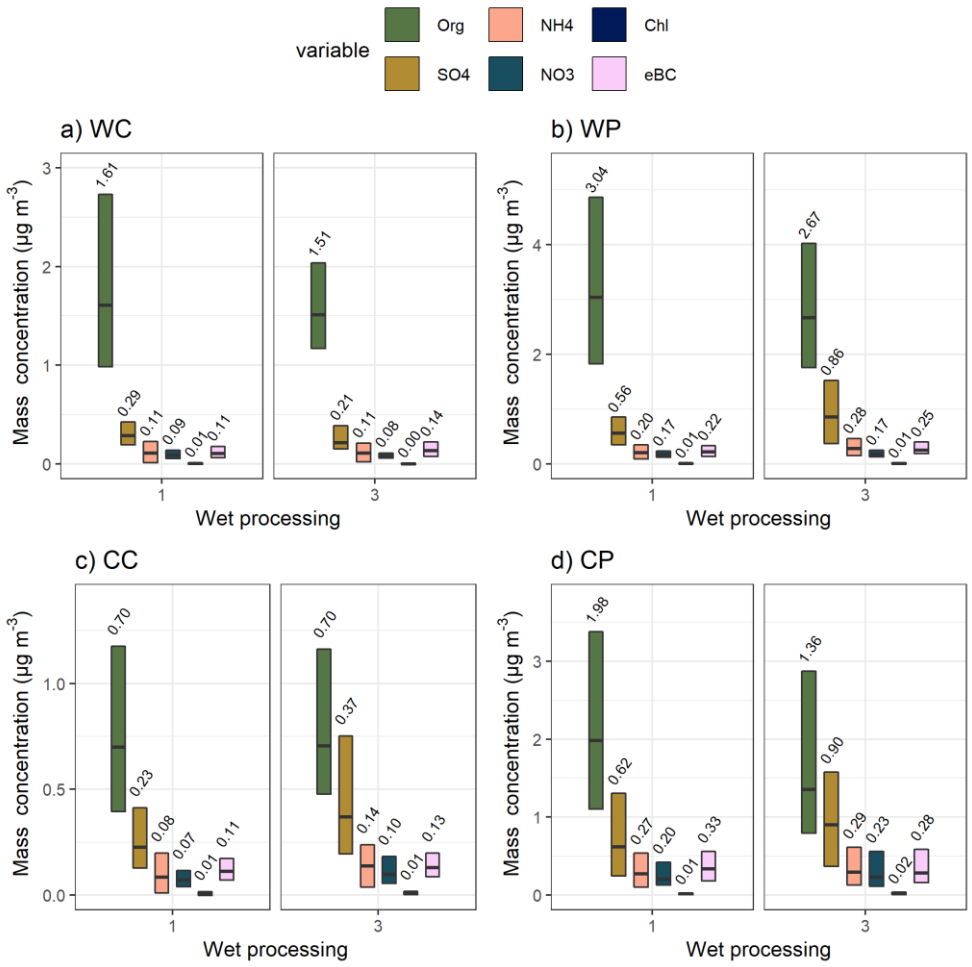

Figure S12 Median (black horizontal lines) particle mass concentration for Org, NH4, Chl, SO4, NO3 and eBC with 25th-75th percentiles (boxes) for airmass history groups 1 and 3 as described in Table 1 (main text). Subplots show the airmass sectors (clean and polluted) with the seasonal (warm and cold) division following a) warm and clean, b) warm and polluted, c) cold and clean and b) cold and polluted. The figure is based on simultaneous observations of these species between March 2012 and August 2019.

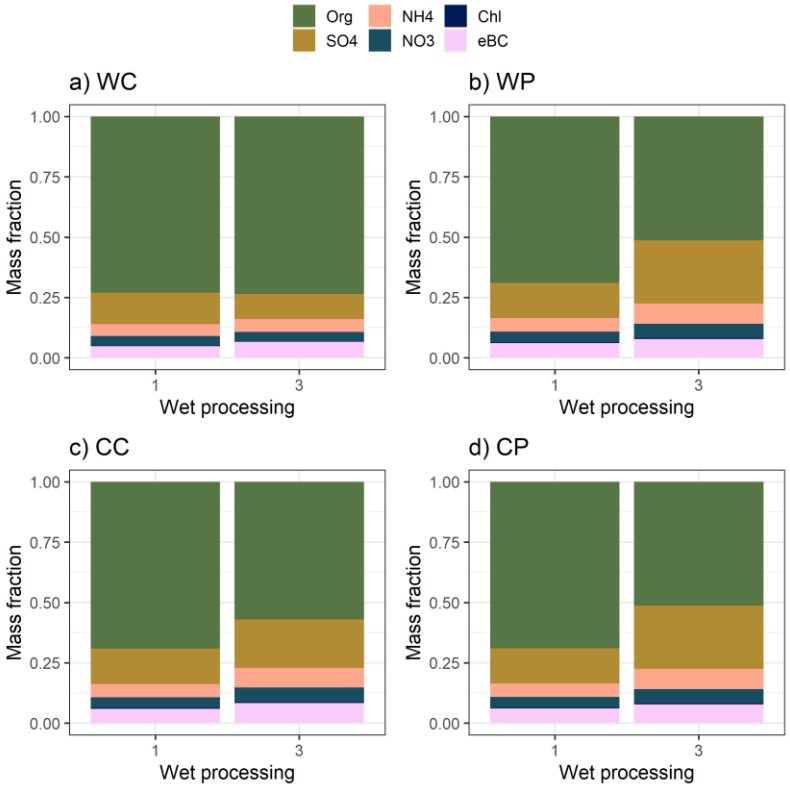

Figure S13 Chemical composition of the particles for each airmass history groups 1 and 3 as described in Table 2 (main text). Subplots show the airmass sectors (clean and polluted) with the seasonal (warm and cold) division: a) warm and clean, b) warm and polluted, c) cold and clean and b) cold and polluted. The figure includes simultaneous observations of these species, spanning between March 2012 and August 2019.

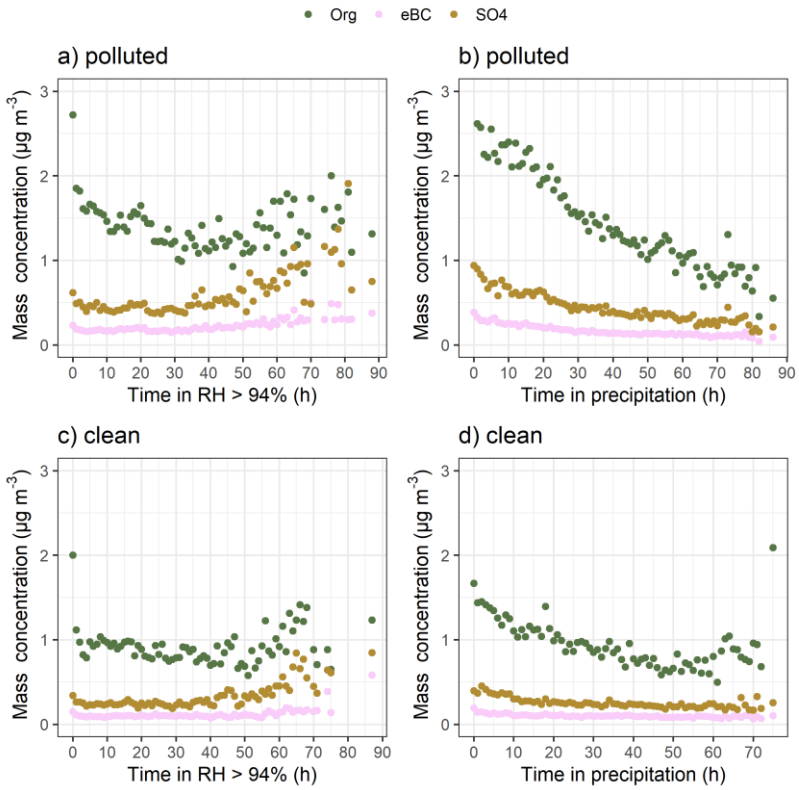

Figure S14 The mass concentrations of Org, SO_4 and eBC for clean and more polluted airmasses as a function of time spent in RH > 94 % (a and c) and in precipitation (b and d). Figure shows median values for each 1-hour bin, if 10 or more data points were available in the bin. The figure is based on observations between March 2012 and August 2019.

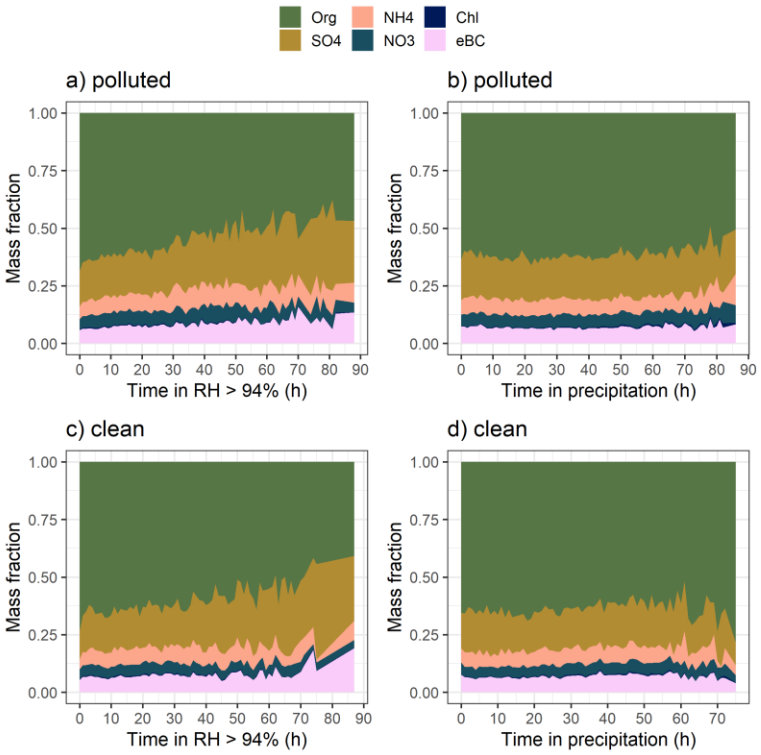

Figure S15 The mass fractions of Org, NH4, Chl, SO4, NO3 and eBC for clean and more polluted airmasses (calculated from median values if 10 or more data points were available in the bin) as a function of time spent in high humidity (RH > 94%) conditions (a and c) and in precipitation (b and d) with bin size being one hour for both cases. The figure is based on observations between March 2012 and August 2019.

Table S6 Statistical significance (p-values) of the differences between wet processing groups for the size classes based on DMPS data for each sector (manuscript Fig. 9). P-values were obtained by comparing the data from wet processing groups 1 and 3 with Mann Whitney U test. Values smaller than 0.01 indicate the null hypothesis (distributions are identical) should be rejected.

Size class	**Sector**			
	WC	**WP**	**CC**	**CP**
[3, 30]	<0.001	<0.001	<0.001	<0.001
(30, 50]	<0.001	<0.001	<0.001	<0.001
(50, 100]	<0.001	<0.001	<0.001	<0.001
(100, 200]	<0.001	<0.001	0.007	<0.001
(200, 350]	<0.001	<0.001	<0.001	0.771
(350, 600]	0.010	<0.001	<0.001	0.002
(600, 1000]	<0.001	<0.001	<0.001	0.205

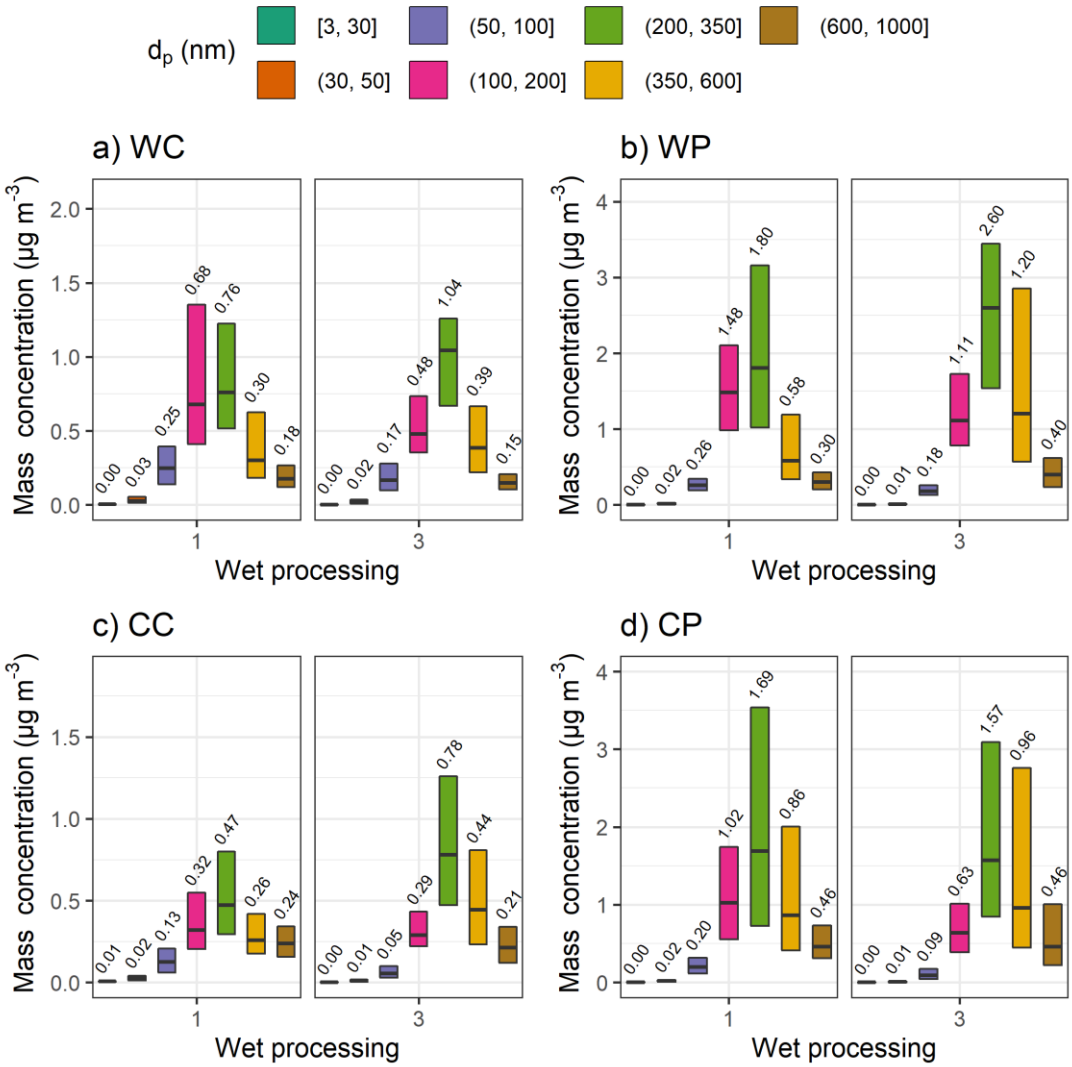

Figure S16 Median (black horizontal lines) particle mass concentration with 25th-75th percentiles (boxes) for size bins derived from the DMPS measurements for airmass history groups 1 and 3 as described in Table 2 in the main text. Subplots show the airmass sectors (clean and polluted) with the seasonal (warm and cold) division following a) warm and clean, b) warm and polluted, c) cold and clean and b) cold and polluted. The figure includes simultaneous observations from the DMPS and Org, BC and SO4 from the ACSM, thus spanning between March 2012 and August 2019.

S3 Variable selection and results from statistical mixed effects models

S3.1 Selection of relevant variables and determining the relative contribution of variable groups

In this section we justify our decision to leave out some of the variables, that could be considered relevant for the response variables (Org, eBC, SO4) we investigated. Variables that were investigated but were excluded from the final model were RH, SolR, WS, WD, $rain_{local}$ and p. RH has strong correlation coefficient (> 0.5) with MLH and having both in the model violates the assumption of the model on relatively independent predictors and causes collinearity issues on the computation (e.g. Dormann et al., 2013). When comparing models with either RH or MLH, the models with MLH had better predictive capability, thus MLH was selected. MLH also has high correlation coefficient SolR, but again the model with MLH had better predictive capability and thus SolR was discarded from further analysis. In addition, including both, RH and SolR, had negligible effect on goodness-of-fit indicators or did not change the other regression coefficients significantly indicating their presence in the model does not improve the overall fit. WS, WD, $rain_{local}$ and p were not significant predictors for the models based on a likelihood ratio test (Wilks, 1938) made between different model versions. Simultaneous observations of the selected variables were used in the final regression, including 22,778 observations in total from the period between March 2012 and August 2019. The division into two seasons based on monthly median temperature T_m (discussed in detail in Section 3.1 in the main manuscript) led to 14,501 observations for warm months (T_m > 10 °C) and 8,277 observations for the cold months (T_m < 10 °C).

To investigate how strong and/or significant an effect each predictor group shown in Table A1 (main text) has for the observed variable (Org, eBC, SO_4) relative to all other predictor groups, we applied Bayesian Information Criterion (BIC, Schwarz, 1978) derived for each of the fitted models. BIC is a criterion which can be used in model selection as models with lower BIC are preferred (Schwarz, 1978). It is based on likelihood-function and includes a penalty term for the number of variables in a model to avoid overfitting (Schwarz, 1978; Stoica and Selen, 2004). Each variable group was removed in turn from the full model (separate models for each species) presented in Eq. (A1), and the BIC for the reduced model was compared to the BIC of the full model. With this approach, we were able to determine the relative contribution of each variable group for the investigated species (response variable).

S3.2 Regression coefficients and relative contributions

Table S7-Table S9 show the relative changes in BIC for the fitted mixed effects models and Table S10-Table S12 show the obtained regression coefficients for the full models. Separate models were fitted for each species (Org, eBC and SO4) and temperature regimes (T_m > 10 °C and T_m < 10 °C) and the variable groups referred in the tables were presented in Table A1 in the main text. As group 4 (wet processing along trajectory) and its subgroups 4a and 4b are based on the processes along the airmass trajectories, their effects on the model can be considered highly collinear. On the other words, groups 4 and 5 (long-range transport) likely share some part of the model predictive potential.

Relative difference (Table S7-Table S9) close to zero indicates the variable group does not improve the overall fit much, though it does not necessary mean that the variable would not be important in the model considering other properties. The correlation coefficient describes how well the changes in the observed data are captured by the model, larger coefficient indicating from better fit. Each variable group was removed from the full model (marked as “none” in Table S7-Table S9) separately at a time, thus the change in BIC expresses how much each variable group contributes to the model in relative to all other used variable groups. Calculation of BIC is based on the log-likelihood function (Schwarz, 1978; Stoica and Selen, 2004) and is thus sensitive to changes in the data. It is commonly used to compare nested model structures where smaller model (here rows with removed variable group 1-5 in Table S7-Table S9) is compared to a model

with larger number of predictors also including those in the smaller model (here model marked as "none" in Table S7-Table S9). Both data and the number of available data rows for the temperature classes differ, and therefore, comparison of the relative contribution between the temperature classes should be done with caution. Larger differences can be interpreted as a qualitative indication (e.g. local meteorology, group 3, has stronger influence in the organic mass during warmer months than during colder months, see Table S7). Smaller differences, however, can be easily overinterpreted (see e.g. relative contributions for both temperature classes for group 4b in Table S7) and conclusions should not be drawn based those values only.

Table S7 Org as a response variable. Relative contribution as a difference in BIC as a percentage compared to full model (with all variable groups) when one group at a time is removed from the model. Correlation coefficient describes how well the changes in the observed data are captured by the model.

Removed variable group	Relative contribution (%)		Correlation coef.	
	warm	cold	warm	cold
none	NA	NA	0.81	0.77
1	6.03	2.36	0.76	0.74
2	4.79	10.40	0.77	0.66
3	10.62	1.76	0.71	0.75
4	0.60	2.26	0.80	0.75
4a	0.45	2.21	0.81	0.75
4b	0.22	0.11	0.81	0.76
5	2.13	1.20	0.79	0.75

Table S8 SO_4 as a response variable. Relative contribution as a difference in BIC as a percentage compared to full model (with all variable groups) when one group at a time is removed from the model. Correlation coefficient describes how well the changes in the observed data are captured by the model.

Removed variable group	Difference (%)		Correlation coef.	
	warm	cold	warm	cold
none	NA	NA	0.64	0.73
1	15.13	9.68	0.55	0.68
2	17.93	18.04	0.54	0.64
3	0.01	0.05	0.64	0.73
4	0.13	0.89	0.64	0.72
4a	0.09	0.65	0.64	0.73
4b	0.01	0.10	0.64	0.73
5	4.00	9.33	0.62	0.69

Table S9 eBC as a response variable. Relative contribution as a difference in BIC as a percentage compared to full model (with all variable groups) when one group at a time is removed from the model. Correlation coefficient describes how well the changes in the observed data are captured by the model.

Removed variable group	Difference (%)		Correlation coef.	
	warm	cold	warm	cold
none	NA	NA	0.74	0.79
1	6.08	27.04	0.70	0.74
2	15.95	78.68	0.64	0.63
3	5.80	10.21	0.71	0.77
4	0.62	3.54	0.74	0.78
4a	0.62	3.28	0.74	0.78
4b	0.05	0.77	0.74	0.79
5	1.44	14.10	0.73	0.76

Table S10 Regression coefficients and their standard errors from the full mixed effects model for Org. Model fixed intercepts (β_0) were -5.45/-4.42 (warm/cold, µg m^{-3}) and the standard deviations for random intercepts were 0.447/0.129 for hour, 0.631/0.493 for month, 0.565/0.243 for year and 0.284/0.048 for airmass source area for warm and cold months, respectively.

Model predictor variable	Regression coefficient				Unit
	warm		cold		
	Slope (β_n)	**Std. Error**	**Slope (β_n)**	**Std. Error**	
NOx	6.25E-01	3.83E-01	2.64E-01	5.62E-03	µg m^{-3}/ppb
SO_2	1.82E-01	2.92E-02	5.34E-01	5.29E-03	µg m^{-3}/ppb
O_3	-1.68E-02	2.82E-02	3.43E-02	5.60E-03	µg m^{-3}/ppb
CO	3.22E-02	1.87E-03	2.48E-02	1.75E-03	µg m^{-3}/ppb
T	2.57E-01	8.15E-04	4.80E-02	5.17E-04	µg m^{-3}/°C
MLH	-7.22E-01	3.44E-03	-6.85E-01	2.78E-03	µg m^{-3}/km
accum.precip*	-3.07E-02	3.26E-03	-6.47E-01	5.62E-03	mm^{-1}
time.in.cloud	-1.42E-02	1.31E-03	-6.25E-03	9.82E-04	µg m^{-3}/h
emission.col.time	-1.49E-04	5.65E-04	6.10E-03	5.57E-04	µg m^{-3}/h
time.in.land	8.62E-03	6.42E-04	7.25E-03	6.57E-04	µg m^{-3}/h

* Note that this coefficient is in the exponent, see Eq. (A1) from the main text.

Table S11 Regression coefficients and their standard errors from the full mixed effects model for SO_4. Model fixed intercepts (β_0) were -1.79/-2.36 (warm/cold, µg m^{-3}) and the standard deviations for random intercepts were 0.020/9.44E-11 for hour, 0.140/0.218 for month, 0.138/0.200 for year and 0.114/0.080 for airmass source area for warm and cold months, respectively.

Model predictor variable	Regression coefficient		Unit
	warm	cold	

	Slope (β_n)	**Std. Error**	**Slope (β_n)**	**Std. Error**	
NOx	1.07E-01	9.75E-02	7.03E-03	1.29E-01	µg m^{-3}/ppb
SO_2	3.01E-01	9.07E-03	7.62E-01	6.94E-03	µg m^{-3}/ppb
O_3	1.53E-02	8.82E-03	1.01E-02	2.56E-02	µg m^{-3}/ppb
CO	5.90E-03	5.83E-04	7.40E-03	9.87E-04	µg m^{-3}/ppb
T	4.34E-03	2.54E-04	7.24E-03	2.67E-04	µg m^{-3}/°C
MLH	-1.91E-02	1.05E-03	-1.94E-02	1.50E-03	µg m^{-3}/km
accum.precip*	-1.90E-03	4.24E-04	-7.96E-03	1.04E-03	mm^{-1}
time.in.cloud	1.37E-03	4.08E-04	2.00E-03	4.58E-04	µg m^{-3}/h
emission.col.time	7.38E-04	1.72E-04	2.49E-03	2.62E-04	µg m^{-3}/h
time.in.land	-9.09E-04	2.01E-04	6.00E-03	3.13E-04	µg m^{-3}/h

* Note that this coefficient is in the exponent, see Eq. (A1) from the main text.

Table S12 Regression coefficients and their standard errors from the full mixed effects model for eBC. Model fixed intercepts (β_0) were -1.38/-1.96 (warm/cold, µg m^{-3}) and the standard deviations for random intercepts were 0.016/0.005 for hour, 0.024/0.116 for month, 0.049/0.078 for year and 0.015/0.010 for airmass source area for warm and cold months, respectively.

Model predictor variable	**Regression coefficient**				**Unit**
	warm		cold		
	Slope (β_n)	**Std. Error**	**Slope (β_n)**	**Std. Error**	
NOx	9.08E-02	4.25E-03	4.79E-02	6.52E-03	µg m^{-3}/ppb
SO_2	1.66E-02	2.13E-03	1.10E-01	2.57E-03	µg m^{-3}/ppb
O_3	1.52E-03	2.09E-03	7.21E-03	5.54E-03	µg m^{-3}/ppb
CO	3.07E-03	1.55E-04	5.94E-03	3.66E-04	µg m^{-3}/ppb
T	9.38E-03	6.44E-05	8.81E-03	1.01E-04	µg m^{-3}/°C
MLH	-5.51E-02	2.87E-04	-1.28E-01	5.75E-04	µg m^{-3}/km
accum.precip*	-1.38E-03	1.12E-04	-4.33E-03	3.58E-04	mm^{-1}
time.in.cloud	-5.05E-04	1.10E-04	-1.25E-03	1.83E-04	µg m^{-3}/h
emission.col.time	2.38E-04	4.65E-05	2.52E-03	1.05E-04	µg m^{-3}/h
time.in.land	3.28E-04	5.40E-05	1.24E-03	1.20E-04	µg m^{-3}/h

* Note that this coefficient is in the exponent, see Eq. (A1) from the main text.

S4 Source contribution and vertical transport of the airmasses

Source contribution analysis was applied to further confirm that e.g. the seasonal patterns of trace gases like SO_2 and aerosols shown in Riuttanen et al. (2013) also hold for our study period. For each back trajectory arrival time, a temporally collocated measurement made at the SMEAR II station is assigned to all grid cells visited by the corresponding trajectory. Grid cells visited by fewer than 10 trajectories are discarded. The median of values for each grid cell is then calculated to produce a concentration field. Finally, the percentage difference from the total median is calculated. This approach is similar to Riuttanen et al. (2013) however does not apply a weighting according to travelled distance. Figure S17 and Figure S18 show examples of this analysis for accumulation mode particle number concentration and SO_2 concentration, respectively.

To investigate the vertical transport, and exclude possible influence caused by e.g., low level transport of sulfate from the oceans into our results (which could show up as high sulfate concentrations when the airmass has spent more time under high humidity conditions), we looked into the airmass height provided by the HYSPLIT airmass trajectories. For this, the airmass height data along the trajectories was clustered with kmeans-clustering using the Hartigan-Wong algorithm (Hartigan and Wong, 1979) and 3 clusters were selected for further analysis. Additional clusters did not show new patters, but only divided the cluster 3 into sub-clusters with different maximum heights. Figure S19 shows the median airmass height above ground along the trajectory with 25^{th}-75^{th} percentiles. Cluster 1 shows now the near-surface trajectories which could be transporting SO_4 from oceans along with high RH values, if the airmasses pass near sea surface. This cluster, however, has largest contribution of SO_4 from the continental Europe (Figure S20), indicating these airmasses are mostly affected by anthropogenic emissions of SO_2, as sulfur from anthropogenic sources is mainly emitted as SO_2 (e.g. Paulot et al., 2017).

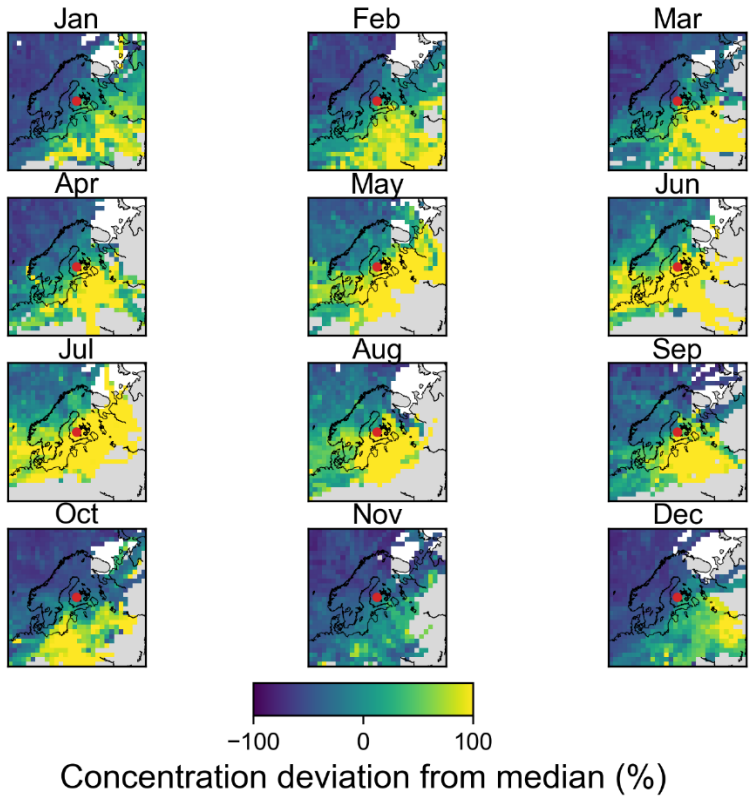

Figure S17 Potential source contribution of accumulation mode (dp = 100 nm – 1000 nm) particle number concentration in SMEAR II. Concentration fields are normalized to the median concentration over the whole measurement data. Data is shown for years between 2012-2019 which overlaps with the composition measurements. Pre-processed data is shown, thus trajectories crossing Kola Peninsula and trajectories arriving to SMEAR II with wind direction between 120°-140° are not included. Red dot shows the location of SMEAR II.

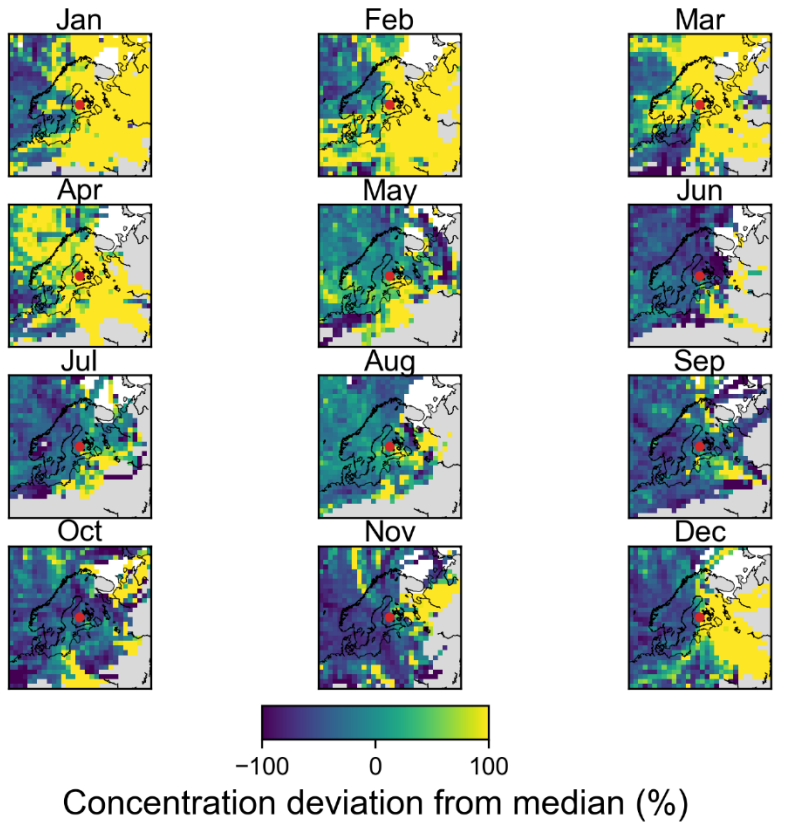

Figure S18 Potential source contribution of SO_2 concentration in SMEAR II. Concentration fields are normalized to the median concentration over the whole measurement data. Data is shown for years between 2012-2019 which overlaps with the composition measurements. Pre-processed data is shown, thus trajectories crossing Kola Peninsula and trajectories arriving to SMEAR II with wind direction between 120°-140° are not included. Red dot shows the location of SMEAR II.

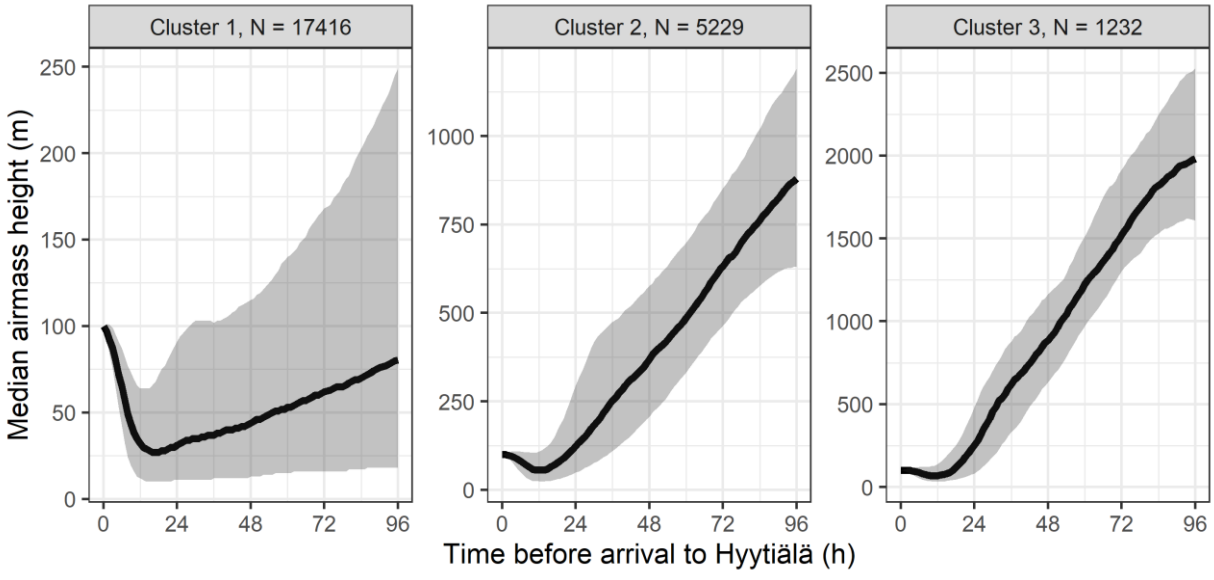

Figure S19 Median airmass height for each cluster (black lines) with 25th-75th percentiles (shaded area). The figure includes collocated observations of the trajectory data and measured SO_4/Org/BC, thus including time between March 2012 and August 2019. N shows the number of data rows in each cluster.

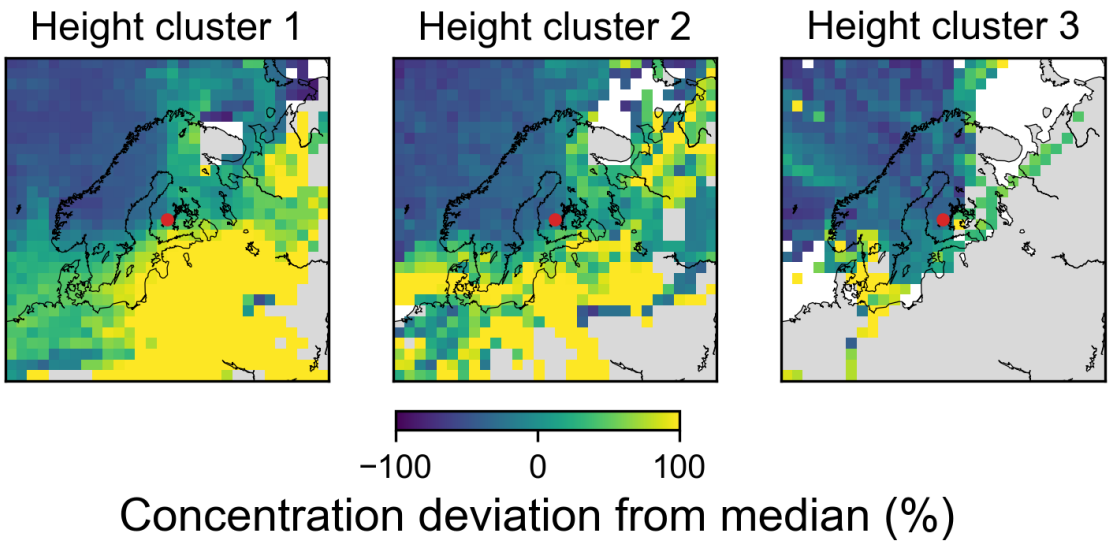

Figure S20 Potential source contribution of SO_4 for each height cluster shown in Figure S19. Red dot shows the location of SMEAR II.

S5 Main figures with HYSPLIT trajectories using GDAS reanalysis meteorology

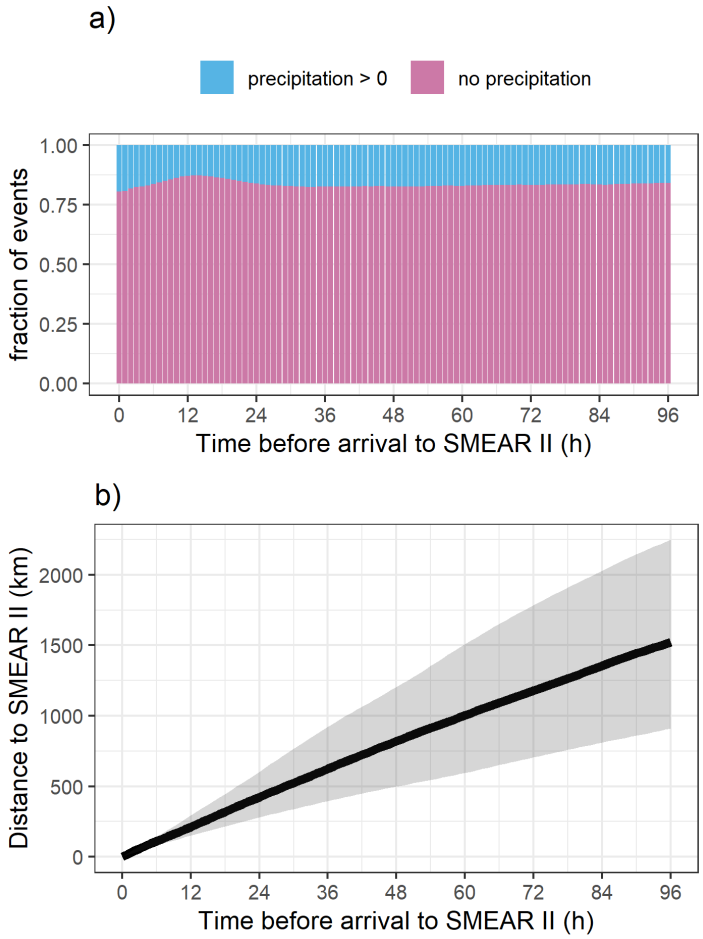

Figure S21 Fraction of precipitation events (based on GDAS meteorology) along the trajectory is shown in a) and b) shows the median distance from SMEAR II with shaded area showing the 25th- 75th percentiles.

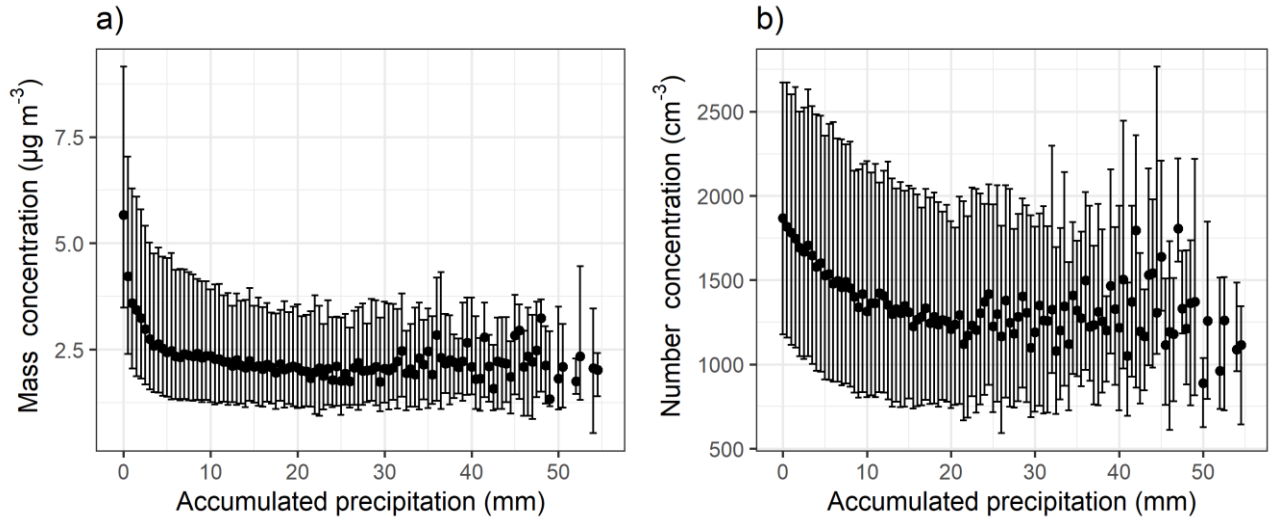

Figure S22 Total particle (d_p = 3-1000 nm) mass (a) and number (b) concentration as a function of 0-50 mm accumulated precipitation along the 96-hour HYSPLIT airmass trajectories based on GDAS meteorology. The black dots show the median values and bars highlight the 25th-75th percentiles for each 0.5 mm bin of accumulated precipitation. The figure includes DMPS data between January 2005 and August 2019.

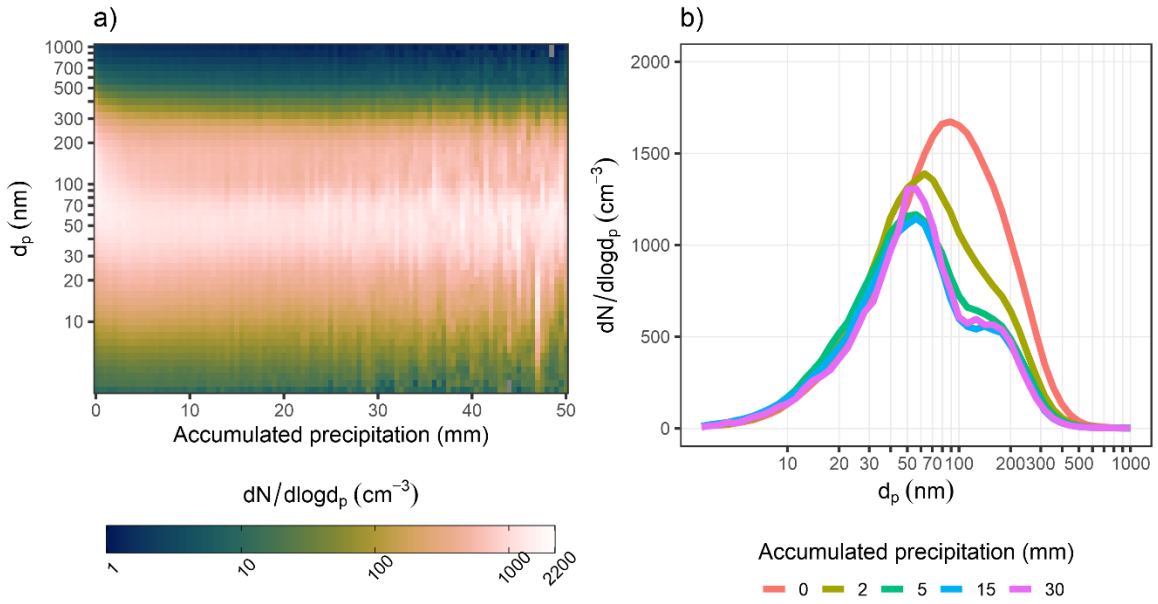

Figure S23 The aerosol size distribution (d_p = 3-1000 nm) as a function of the 0-50 mm of accumulated precipitation along the 96-hour airmass trajectories is shown in (a). Data is shown as medians for binned accumulated precipitation (bin size 0.5 mm) based on GDAS meteorology in the trajectory calculations. Grey areas show missing data and/or values where dN/dlogd$_p$ < 1. Median size distributions are presented for selected values of accumulated precipitation in b). The figure includes DMPS data between January 2005 and August 2019.

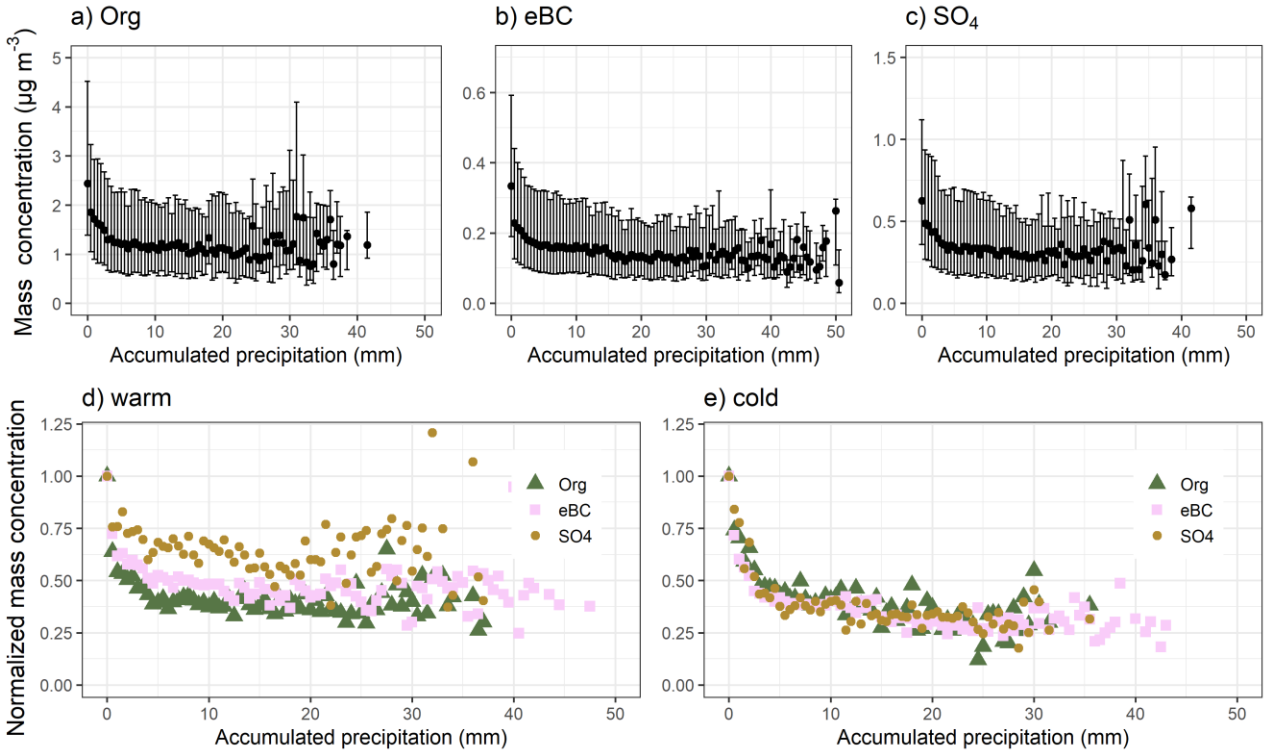

Figure S24 Particle mass concentration of a) Org, b) eBC and c) SO4 as a function of accumulated precipitation along the 96-hour airmass trajectories based on GDAS meteorology. The black dots in the top row show the median values and bars highlight the 25th-75th percentiles for each 0.5 mm bin of accumulated precipitation. Bottom row shows normalized particle masses (calculated from the medians) with temperature separation. Medians and normalized medians are shown for each bin having 10 or more data points. The figure includes data between 2006-2019 for eBC and 2012-2019 for Org and SO4.

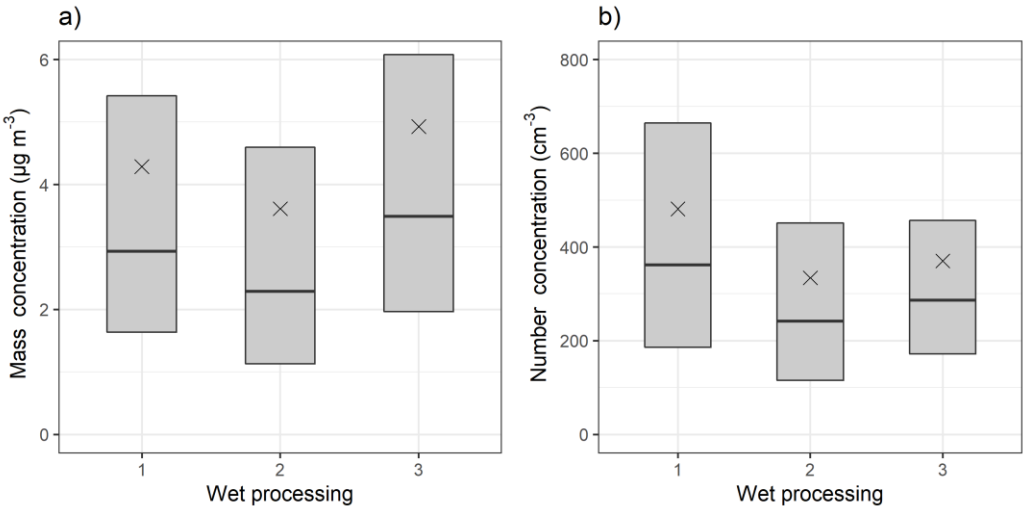

Figure S25 Median (black horizontal lines), mean (black crosses) and 25th -75th percentiles (boxes) for accumulation mode (d_p = 100-1000 nm) particle mass (a) and number (b) concentration for wet processing groups (trajectories based on GDAS meteorology) described in Table 1 in the main manuscript. The figure includes DMPS data between January 2005 and August 2019. The whiskers show the 99 % confidence intervals from 1000 bootstrap replicates.

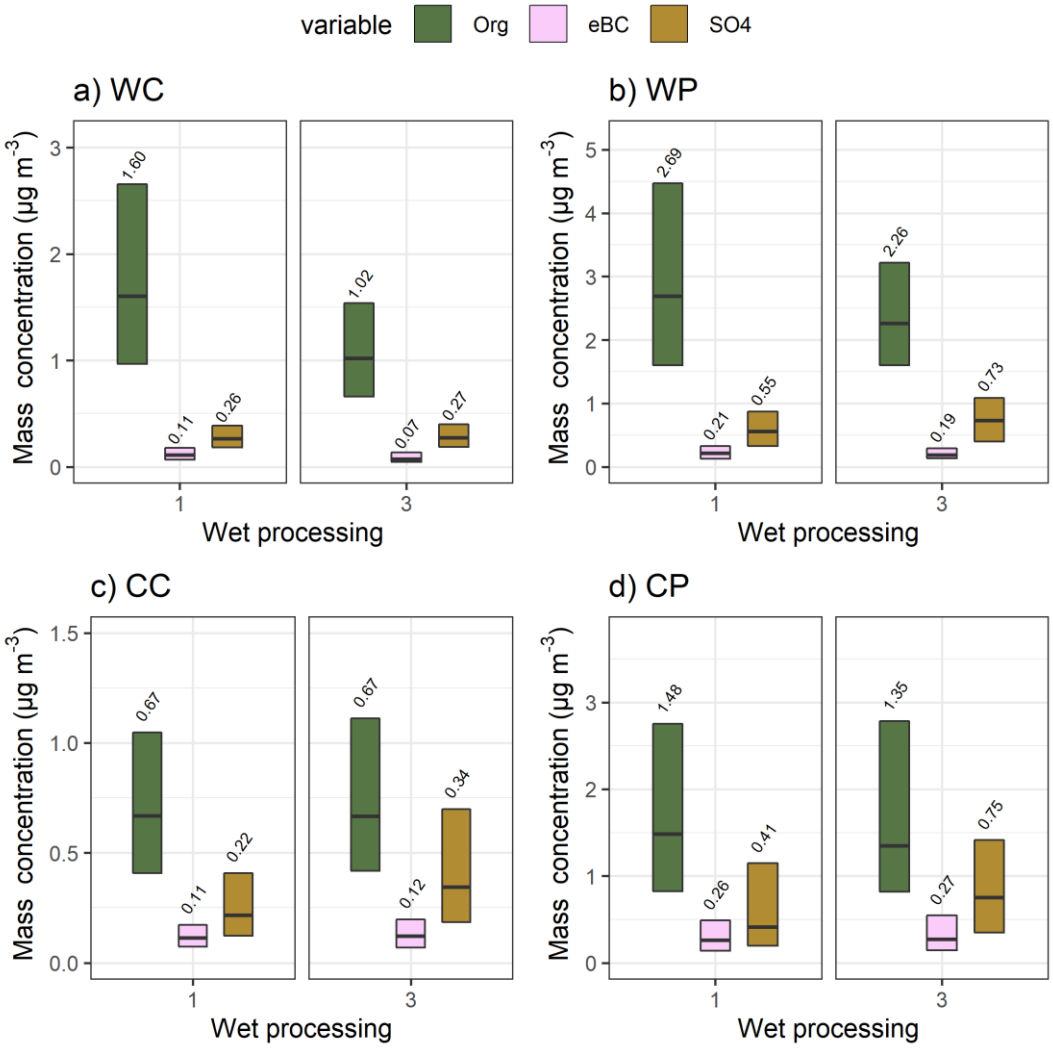

Figure S26 Median (black horizontal lines and numerical values) particle mass concentration with 25th-75th percentiles (boxes) for Org, eBC and SO_4 for wet processing groups 1 and 3 as described in Table 1 in the main manuscript. Subplots show the

airmass sectors (clean and polluted) with the seasonal (warm and cold) division: a) warm and clean, b) warm and polluted, c) cold and clean and b) cold and polluted. The figure is based on simultaneous observations of these three species between March 2012 and August 2019. Note the different y-axis limits in each subplot. Trajectories based on GDAS meteorology.

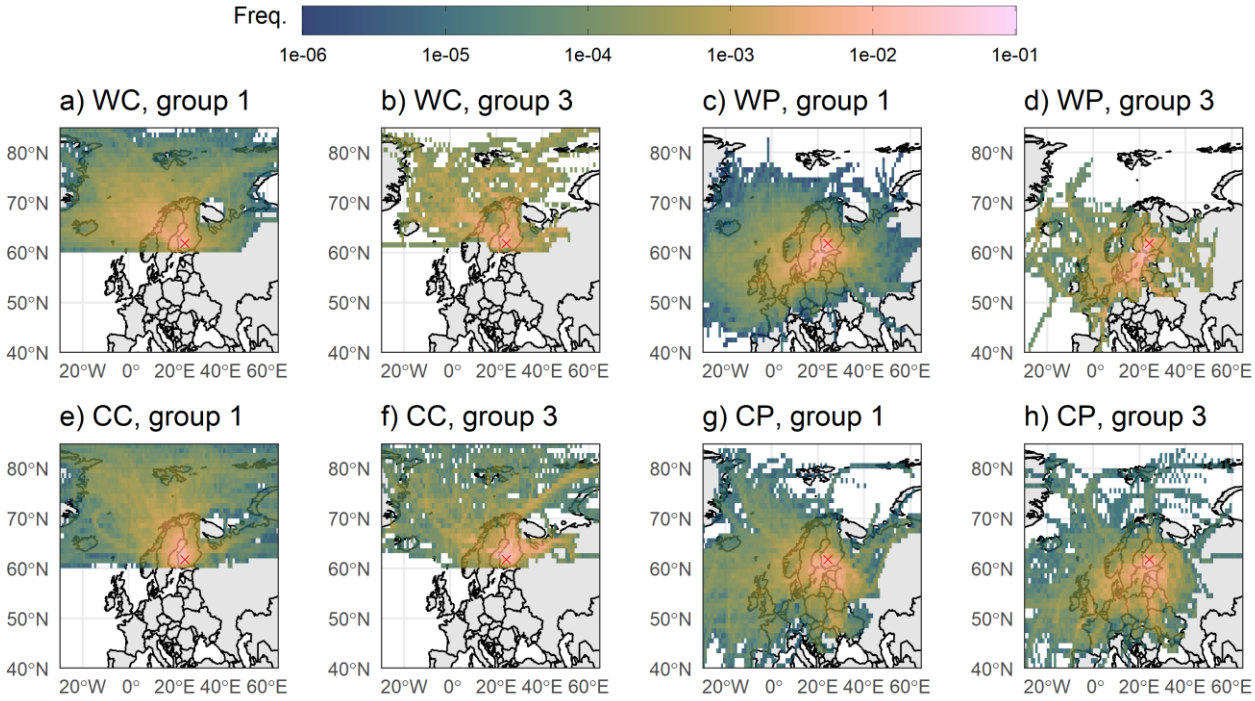

Figure S27 96-hour airmass history (trajectories based on GDAS meteorology) for the wet processing groups with the sector (clean and polluted) and temperature (warm and cold) division. Subplots show a)-b) warm and clean, c)-d) warm and polluted, e)-f) cold and clean and g)-h) cold and polluted. Colour scale shows the frequency (crossings in each 1° x 1° grid point divided by total number of crossings in each group) of trajectories crossing a grid point. The groups 1 and 3 correspond to the airmass history groups presented in Table 2 in the main text. The red cross shows the location for SMEAR II.

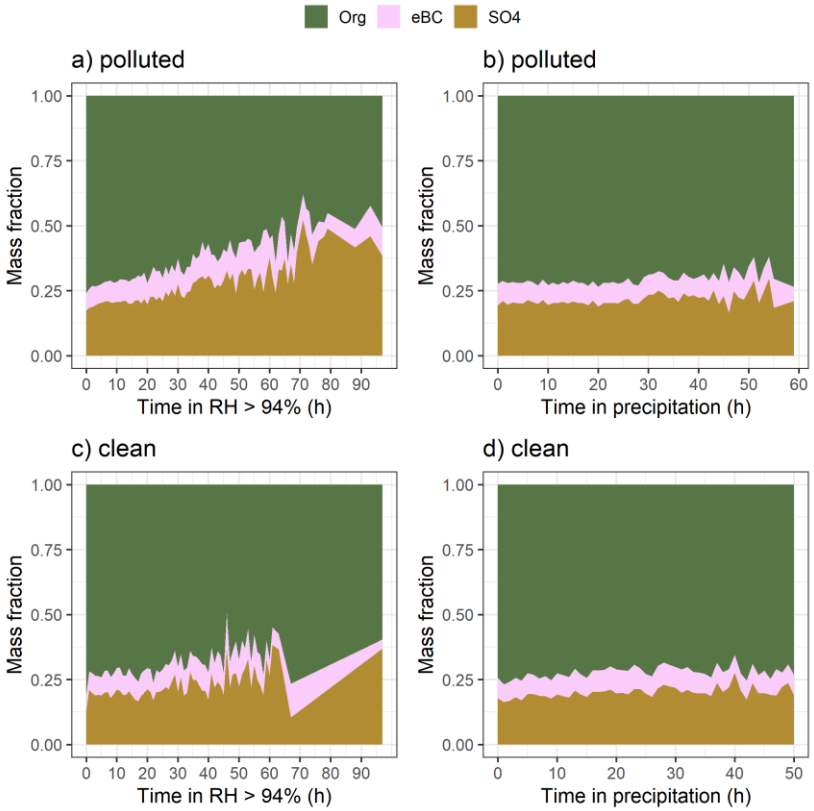

Figure S28 The mass fractions of Org, SO4 and eBC for clean and more polluted airmasses as a function of time spent in RH > 94 % (a and c) and in precipitation (b and d): Figure shows median values for each 1-hour bin, if 10 or more data points were available in the bin. The figure is based on observations between March 2012 and August 2019. Trajectories based on GDAS meteorology.

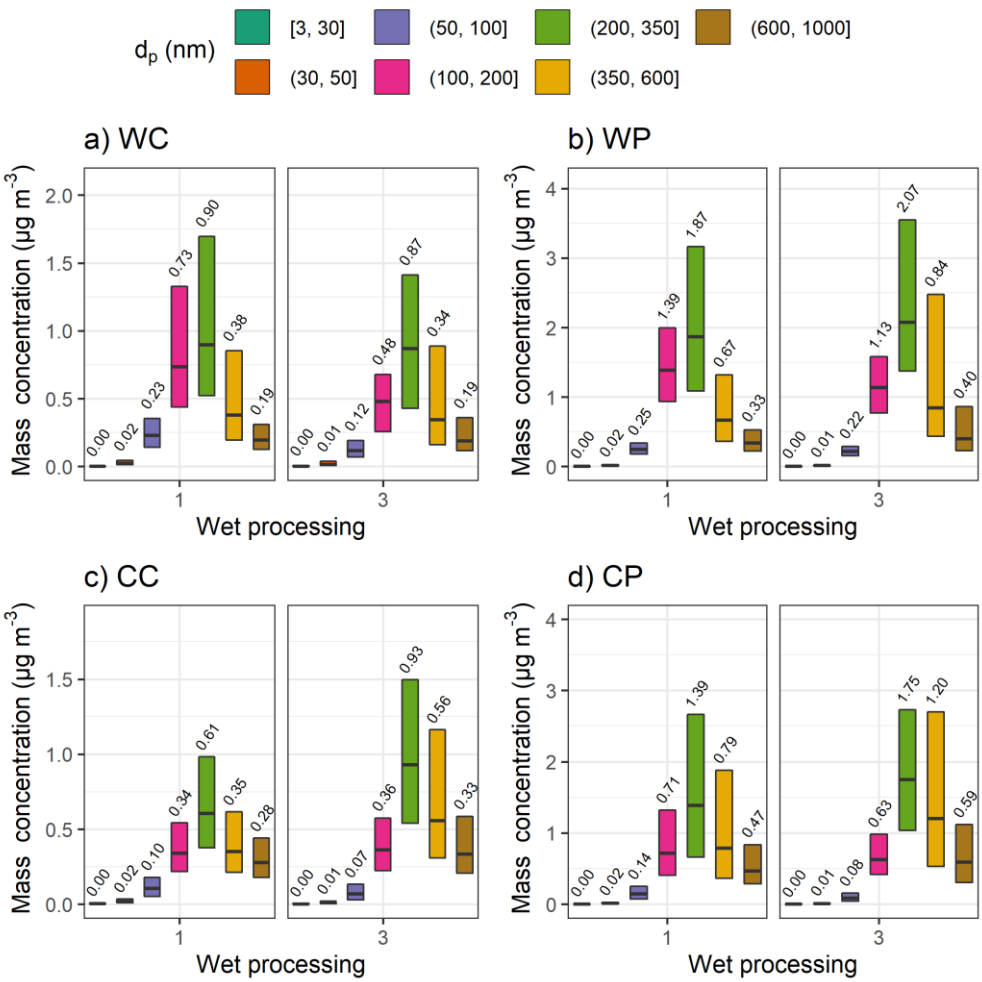

Figure 29 Median (black horizontal lines and numerical values) particle mass concentration with 25th-75th percentiles (boxes) for size bins derived from the DMPS measurements for wet processing groups 1 and 3 as described in Table 1 in the main manuscript. Subplots show the airmass sectors (clean and polluted) with the seasonal (warm and cold) division: a) warm and clean, b) warm and polluted, c) cold and clean and b) cold and polluted. The figure includes all available data between January 2005 and August 2019. Note the different y-axis scale in each subplot. The trajectories are based on GDAS meteorology.

References

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S.: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, Ecography, 36, 27-46, https://doi.org/10.1111/j.1600-0587.2012.07348.x, 2013.

Hartigan, J. A., and Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics), 28, 100-108, 10.2307/2346830, 1979.

Paulot, F., Fan, S., and Horowitz, L. W.: Contrasting seasonal responses of sulfate aerosols to declining SO2 emissions in the Eastern U.S.: Implications for the efficacy of SO2 emission controls, Geophysical Research Letters, 44, 455-464, https://doi.org/10.1002/2016GL070695, 2017.

Riuttanen, L., Hulkkonen, M., Dal Maso, M., Junninen, H., and Kulmala, M.: Trajectory analysis of atmospheric transport of fine particles, SO2, NOx and O3 to the SMEAR II station in Finland in 1996–2008, Atmos. Chem. Phys., 13, 2153-2164, 10.5194/acp-13-2153-2013, 2013.

Schwarz, G.: Estimating the Dimension of a Model, The Annals of Statistics, 6, 461-464, 464, 1978.

Stoica, P., and Selen, Y.: Model-order selection: a review of information criterion rules, IEEE Signal Processing Magazine, 21, 36-47, 10.1109/msp.2004.1311138, 2004.

Virkkula, A., Mäkelä, T., Hillamo, R., Yli-Tuomi, T., Hirsikko, A., Hämeri, K., and Koponen, I. K.: A Simple Procedure for Correcting Loading Effects of Aethalometer Data, Journal of the Air & Waste Management Association, 57, 1214-1222, 10.3155/1047-3289.57.10.1214, 2007.

Wilks, S. S.: The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, The Annals of Mathematical Statistics, 9, 60-62, 63, 1938.