Supplement of Atmos. Chem. Phys., 18, 12185–12206, 2018
https://doi.org/10.5194/acp-18-12185-2018-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Urban influence on the concentration and composition of submicron particulate matter in central Amazonia

Suzane S. de Sá et al.

Correspondence to: Scot T. Martin (scot_martin@harvard.edu)

S1. Positive-matrix factorization

S1.1 Diagnostics of the six-factor solution

Quantification of mass concentrations by the AMS was obtained from "V-mode" data, which corresponds to the shorter ion time-of-flight path and is therefore the more sensitive mode. The choice of ions to fit was aided by "W-mode" data, which correspond to the longer ion time-of-flight path and is therefore the mode with highest mass resolution. V-mode data were collected continuously, and W-mode data were collected for one of every five days. The time series of organic mass spectra measured by the AMS in V-mode was analyzed by positive-matrix factorization (PMF) using a standard analysis toolkit (Ulbrich et al., 2009) The PMF solution was based on minimization of the "Q-value" (i.e., the sum of the weighed squared residuals for a chosen number of factors) and the physical meaningfulness of factors, as evaluated by profile characteristics and correlations with gas and particle phase measurements by other instruments.

Technical diagnostics of the six-factor solution are presented in Figure S3 in complement to the diagnostics presented in de Sá et al. (2017). The analysis was run for a number of factors from 1 to 10, and the rotational ambiguity parameter f_{peak} was varied from -1 to 1 in intervals of 0.2. Panel a shows the statistics of residuals for solutions with different number of factors. There was a large improvement in the solution when a sixth factor was introduced, as shown by a significant decrease in residuals, and only a marginal improvement when a seventh factor was added. Panel b shows, on the ordinate, the correlation between the time series of loadings for each pair of factors and, on the abscissa, the correlation between the profiles of each pair of factors. For the six-factor solution, the correlations among factor profiles are overall lower, also suggesting a better separation of factors and an improvement in the solution. Figure S4 corroborates this analysis by showing the factor profiles and loading time series of the 5- and 7-

factor solutions. In the 5-factor solution, factors 4 and 5 seem to be a result of mixing of the three factors that are associated with secondary processing in the 6-factor solution (MO-OOA, LO-OOA, IEPOX-SOA). Conversely, in the 7-factor solution, some splitting seems to occur as factor 7 is physically meaningless, and a few pairs of factors have higher correlations between their loading time series (cf. Figure S3). An f_{peak} of zero was chosen for the final 6-factor solution, since it yielded the minimum quality of fit parameter $Q/Q_{expected}$ de Sá et al. (2017), and no significant improvements in the external validation of factors were observed by varying f_{peak}.

S1.2 Discussion of the ADOA PMF factor

ADOA is interpreted as a primary anthropogenic factor due to the correlation of its loadings with several tracers of anthropogenic activities (Figure 5), its spectral profile, and its diel behavior (Figure 4). Even though factors containing a characteristic *m*/*z* 91 have been reported in the literature as a biogenic factor (Robinson et al., 2011; Budisulistiorini et al., 2015; Chen et al., 2015; Riva et al., 2016), the ADOA of this study showed similarity with primary organic material from cooking activities. Figure S5 shows the high similarity of ADOA of this study to a factor representing cooking emissions at an urban background site in Barcelona, Spain (Mohr et al., 2012), and to a factor representing a cooking source tied to restaurants in an urban background site in Zurich, Switzerland (Lanz et al., 2007). By contrast, a lower similarity is found with the "91fac" factor found in the Borneo forest, a predominantly biogenic site. This result emphasizes that a characteristic marker ion $C_7H_7^+$ at *m*/*z* 91 does not directly imply either biogenic or anthropogenic origin, and the interpretation of a PMF factor with such marker should also strongly rely on the atmospheric context of the measurements, including the correlations of the factor loadings with external measurements and the diel behavior.

S2. Estimates of organic and inorganic nitrates based on AMS analysis

The typical AMS analysis reports total nitrate, meaning that nitrate fragments originating from both organic and inorganic nitrates are reported indistinctively as nitrate. In the absence of external measurements of inorganic nitrate, an estimation method using the ratio of NO_2^+ to NO^+ signal intensities measured by the AMS was employed (Figure S6; Fry et al., 2009; Farmer et al., 2010; Fry et al., 2013). Calculations were done on a 60-min time base to increase signal over noise. The obtained organic and inorganic nitrate time series were then interpolated into the original AMS timestamp for ambient measurements (i.e., one point every 8-min interval). The analysis excluded points that had total nitrate below the estimated detection limit, $DL_{Nitrate}$, which was estimated as three times the standard deviation for "closed AMS spectra", i.e., when chopper was in closed position and particles did not reach the vaporizer. Mathematically, $DL_{Nitrate} = 3\times\sqrt{E}$, where E is the "closed" error calculated by the standard *PIKA* software (Ulbrich et al., 2009). The dark blue dashed line in Figure S6c that defines NO_2^+/NO^+ for inorganic nitrate was determined by linear fit of ammonium nitrate calibrations performed regularly, as shown by the grey triangles. The small drift over time can be attributed to a gradual clean-up of the vaporizer. Worth noting, whether the linear fit or an average value was used for the calculations, the overall results did not change considerably, as all calibration ratios lied within ± 20% of the campaign-average ratio. The ratio NO_2^+/NO^+ for organic nitrates was assumed to be a factor of 2.25 lower than that of inorganic nitrate based on previous field studies (Farmer et al., 2010; Fry et al., 2013). The resulting IOP1-average for the fraction of organic nitrate in total nitrate (Figure S6b) was 87%.

S3. Fuzzy c-means clustering

Fuzzy c-means (FCM) clustering was applied to the dataset consisting of concentrations of particle number, NO_y, ozone, black carbon, and sulfate (Bezdek et al., 1984). The use of a

fuzzy clustering method stems from the understanding that any point in time may be affected by a combination of different sources and processes and could therefore be anywhere on the scale between pristine background and extreme polluted conditions, as opposed to a simpler binary classification. Given the scope of the analysis as non-overcast afternoon times, data points were restricted to (i) local 12:00-16:00 h, (ii) local solar radiation over the past 4 h not less than 200 W m^{-2} (i.e., excluding the lower 20 percentile), and (iii) insignificant precipitation (< 0.1 mm) over the previous 10 h along backward trajectory (a threshold was used as most rain radar grid cells had non-zero yet negligible values). The data were normalized prior to the FCM analysis using the z-score method, which transforms all variables into a common scale with an average of zero and standard deviation of one.

The FCM algorithm minimizes the objective function represented in Eq. S1, which is a weighted sum of squared errors where the error is the Euclidean distance between each data point and a cluster centroid.

$$J(U,\nu)= \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik}^{m} \, ||y_k-\nu_i||^2 \qquad \text{(Eq. S1)}$$

The input data is given by the matrix $Y = [y_1, y_2, \ldots, y_N]$, where y_k is a vector of length X at the *k-th* time point. X is the number of variables (i.e., measurements) used as input in the analysis. The number of time points is represented by N, and the associated running index is k. N in this case was 313. The number of clusters is represented by c, and the corresponding running index is i. The coordinates of the centroid of each cluster i are represented by ν_i, a vector of length X. The exponent of the Fuzzy partition matrix is represented by m. The algorithm returns (1) the Fuzzy partition matrix of Y, given by $U = [u_{ik}]$ where u_{ik} is the degree of membership of time point k to cluster i, (2) the vectors of coordinates of cluster centers, given by $\nu = [\nu_i]$, as well as (3) the value J of the objective function.

The analysis was performed in MATLAB® using the "fcm" function in the Fuzzy logic toolbox™. The stop criterion of the algorithm is that either the maximum number of iterations is reached or the improvement of the objective function between two consecutive iterations is less than the minimum amount of improvement specified. The default value of 1×10^{-5} was used for the minimum amount of improvement, and the maximum number of iterations was set to 1000 so that convergence always happened before this maximum was reached. A default value of 2 was used for the exponent *m* of the partition matrix. Fuzzy clustering algorithms are not sensitive to small fluctuations in *m* (Chatzis, 2011), and a value in the range of 1.5 to 3 is recommended (Bezdek et al., 1984; Hathaway and Bezdek, 2001).

The analysis was run for a number of clusters varying from two to eight, and the value of the objective function for each run is shown in Figure S7. The choice of number of clusters hinges on a balance between increased complexity and additional information provided by each extra cluster. The improvement in the objective function was larger in the range of two to four clusters, with marginal improvements above four clusters (Figure S7). The location of cluster centroids was also examined for evaluation of cluster overlap (Figure S8). The addition of a fifth cluster made two pairs of clusters very similar, as can be seen by the locations of cluster centroids in Figure S8. The solution of four clusters was therefore a reasonable choice to represent the studied system. The subsequent characterization of the PM chemical composition associated with each cluster further confirmed the meaningfulness of the solution. Although the three-cluster solution could also provide a reasonable representation of the system, the four-cluster solution provided further insight by differentiating two background and two polluted conditions.

Subsequently, the PM composition associated with each of the clusters was determined by calculating the corresponding coordinates of the centroids for AMS species concentrations and PMF factor loadings, which were not input to the FCM analysis. The calculation followed the mathematical definition of the centroid (Eq. S2). The resulting characterization of clusters is shown in Figure 8 and Table 2.

$$v_i = \frac{\sum_{k=1}^{N} (u_{ik})^m y_k}{\sum_{k=1}^{N} (u_{ik})^m} \qquad \text{(Eq. S2)}$$

References

Bezdek, J. C., Ehrlich, R., and Full, W.: FCM: The fuzzy c-means clustering algorithm, Computers & Geosciences, 10, 191-203, 1984.

Budisulistiorini, S. H., Li, X., Bairai, S. T., Renfro, J., Liu, Y., Liu, Y. J., McKinney, K. A., Martin, S. T., McNeill, V. F., Pye, H. O. T., Nenes, A., Neff, M. E., Stone, E. A., Mueller, S., Knote, C., Shaw, S. L., Zhang, Z., Gold, A., and Surratt, J. D.: Examining the effects of anthropogenic emissions on isoprene-derived secondary organic aerosol formation during the 2013 Southern Oxidant and Aerosol Study (SOAS) at the Look Rock, Tennessee ground site, Atmos. Chem. Phys., 15, 8871-8888, 2015, 10.5194/acp-15-8871-2015.

Chatzis, S. P.: A fuzzy c-means-type algorithm for clustering of data with mixed numeric and categorical attributes employing a probabilistic dissimilarity functional, Expert Systems with Applications, 38, 8684-8689, 2011, https://doi.org/10.1016/j.eswa.2011.01.074.

Chen, Q., Farmer, D. K., Schneider, J., Zorn, S. R., Heald, C. L., Karl, T. G., Guenther, A., Allan, J. D., Robinson, N., Coe, H., Kimmel, J. R., Pauliquevis, T., Borrmann, S., Pöschl, U., Andreae, M. O., Artaxo, P., Jimenez, J. L., and Martin, S. T.: Mass spectral characterization of submicron biogenic organic particles in the Amazon Basin, Geophys. Res. Lett., 36, L20806, 2009, 10.1029/2009GL039880.

Chen, Q., Farmer, D. K., Rizzo, L. V., Pauliquevis, T., Kuwata, M., Karl, T. G., Guenther, A., Allan, J. D., Coe, H., Andreae, M. O., Pöschl, U., Jimenez, J. L., Artaxo, P., and Martin, S. T.: Submicron particle mass concentrations and sources in the Amazonian wet season (AMAZE-08), Atmos. Chem. Phys., 15, 3687-3701, 2015, 10.5194/acp-15-3687-2015.

de Sá, S. S., Palm, B. B., Campuzano-Jost, P., Day, D. A., Newburn, M. K., Hu, W., Isaacman-VanWertz, G., Yee, L. D., Thalman, R., Brito, J., Carbone, S., Artaxo, P., Goldstein, A. H., Manzi, A. O., Souza, R. A. F., Mei, F., Shilling, J. E., Springston, S. R., Wang, J., Surratt, J. D., Alexander, M. L., Jimenez, J. L., and Martin, S. T.: Influence of urban pollution on the production of organic particulate matter from isoprene epoxydiols in central Amazonia, Atmos. Chem. Phys., 17, 6611-6629, 2017, 10.5194/acp-17-6611-2017.

Farmer, D. K., Matsunaga, A., Docherty, K. S., Surratt, J. D., Seinfeld, J. H., Ziemann, P. J., and Jimenez, J. L.: Response of an aerosol mass spectrometer to organonitrates and organosulfates and implications for atmospheric chemistry, Proc. Natl. Acad. Sci. USA, 107, 6670-6675, 2010, 10.1073/pnas.0912340107.

Fry, J. L., Kiendler-Scharr, A., Rollins, A. W., Wooldridge, P. J., Brown, S. S., Fuchs, H., Dubé, W., Mensah, A., dal Maso, M., Tillmann, R., Dorn, H. P., Brauers, T., and Cohen, R. C.: Organic nitrate and secondary organic aerosol yield from NO₃ oxidation of β-pinene evaluated using a gas-phase kinetics/aerosol partitioning model, Atmos. Chem. Phys., 9, 1431-1449, 2009, 10.5194/acp-9-1431-2009.

Fry, J. L., Draper, D. C., Zarzana, K. J., Campuzano-Jost, P., Day, D. A., Jimenez, J. L., Brown, S. S., Cohen, R. C., Kaser, L., Hansel, A., Cappellin, L., Karl, T., Hodzic Roux, A., Turnipseed, A., Cantrell, C., Lefer, B. L., and Grossberg, N.: Observations of gas- and aerosol-phase organic nitrates at BEACHON-RoMBAS 2011, Atmos. Chem. Phys., 13, 8585-8605, 2013, 10.5194/acp-13-8585-2013.

Hathaway, R. J. and Bezdek, J. C.: Fuzzy c-means clustering of incomplete data, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 31, 735-744, 2001, 10.1109/3477.956035.

IBGE: Malhas digitais, Setor Censitário, 2010, https://mapas.ibge.gov.br/bases-e-referenciais/bases-cartograficas/malhas-digitais.html, last access: January 2018.

Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prévôt, A. S. H.: Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra, Atmos. Chem. Phys., 7, 1503-1522, 2007, 10.5194/acp-7-1503-2007.

Medeiros, A., Souza, R. A. F., and Martin, S. T., in preparation.

Mohr, C., DeCarlo, P. F., Heringa, M. F., Chirico, R., Slowik, J. G., Richter, R., Reche, C., Alastuey, A., Querol, X., Seco, R., Peñuelas, J., Jiménez, J. L., Crippa, M., Zimmermann, R., Baltensperger, U., and Prévôt, A. S. H.: Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data, Atmos. Chem. Phys., 12, 1649-1665, 2012, 10.5194/acp-12-1649-2012.

Riva, M., Budisulistiorini, S. H., Chen, Y., Zhang, Z., D'Ambro, E. L., Zhang, X., Gold, A., Turpin, B. J., Thornton, J. A., Canagaratna, M. R., and Surratt, J. D.: Chemical characterization of secondary organic aerosol from oxidation of isoprene hydroxyhydroperoxides, Environ. Sci. Technol., 2016, 10.1021/acs.est.6b02511.

Robinson, N. H., Hamilton, J. F., Allan, J. D., Langford, B., Oram, D. E., Chen, Q., Docherty, K., Farmer, D. K., Jimenez, J. L., Ward, M. W., Hewitt, C. N., Barley, M. H., Jenkin, M. E., Rickard, A. R., Martin, S. T., McFiggans, G., and Coe, H.: Evidence for a significant proportion of Secondary Organic Aerosol from isoprene above a maritime tropical forest, Atmos. Chem. Phys., 11, 1039-1050, 2011, 10.5194/acp-11-1039-2011.

Schneider, J., Freutel, F., Zorn, S., Chen, Q., Farmer, D., Jimenez, J., Martin, S., Artaxo, P., Wiedensohler, A., and Borrmann, S.: Mass-spectrometric identification of primary biological particle markers and application to pristine submicron aerosol measurements in Amazonia, Atmos Chem Phys, 11, 11415-11429, 2011, doi.org/10.5194/acp-11-11415-2011.

Ulbrich, I., Canagaratna, M., Zhang, Q., Worsnop, D., and Jimenez, J.: Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data, Atmos. Chem. Phys., 9, 2891-2918, 2009, 10.5194/acp-9-2891-2009.

List of Supplementary Figures

Figure S1. Location of the GoAmazon2014/5 sites relevant for this study. Image data: Google earth.

Figure S2. Scatter plot of the AMS signal fraction at *m*/*z* 44 (f_{44}) against that at *m*/*z* 43 (f_{43}). Green and yellow markers correspond to measurements made by two different AMS instruments at T0t in the wet season of 2008 during the AMAZE-08 campaign (Chen et al., 2009; Schneider et al., 2011). Red markers correspond to measurements made at the T0a (ATTO) by an ACSM during the wet season of 2015. A correction factor of 0.75 was applied to the f_{44} values of the ACSM based on calibrations with standards. Solid squares represent median values, and whiskers represent 10 and 90 percentiles. The plot shows a significant variability between the observations of 2008 and 2015 for the two background sites. An explanation of the differences is not attempted herein and warrants further investigation through longer-term continuous measurements.

Figure S3. Diagnostics of the PMF analysis. (a) Statistics of the sum of for solutions with different number of factors. Box plots show the interquartile ranges, including the medians as a horizontal line. Red markers show the means. Whiskers show the 5 and 95 percentiles. (b) Correlations expressed as between each pair of factors within each PMF solution, with number of factors varying from 2 to 7. The Pearson *R* value between factor loadings is shown on the coordinate and between factor profiles is shown on the abscissa. Numbers in red indicate the identity of the pair of factors.

Figure S4. Results of the PMF analysis for 5 factors (a and b) and 7 factors (c and d). Panels on the left (a and c) show the time series of factor loadings and panels on the right (b and

d) show the profiles of factors. The signals shown in panels b and d were summed to unit mass resolution.

Figure S5. Comparison of the ADOA factor profile from the present study to factors found in three other field studies. "COA" are factors representative of cooking activities, and the "91fac" from Robinson et al. (2011) was tied to biogenic sources.

Figure S6. Summary of the analysis for estimating organic and inorganic nitrates from AMS bulk measurements. (a) Resulting time series of organic and inorganic nitrates are shown together with the original nitrate AMS times series. (b) Time series of the fraction of organic nitrate in total nitrate. (c) Time series of the measured NO_2^+/NO^+ ratio is shown in red and values of NO_2^+/NO^+ from ammonium nitrate calibrations are shown in gray triangles. A linear fit to those calibration ratios is shown by the dashed dark blue line and constitutes the reference ratio for inorganic nitrate over time. The dashed light blue line is the reference ratio for organic nitrates over time. Calculations were done for data binned to one hour (as plotted), and the resulting time series were interpolated to the native time stamp for evaluation of correlations in the PMF analysis.

Figure S7. Value of the objective function of the FCM analysis (Eq. 1) in the last iteration plotted against the number of clusters.

Figure S8. Locations of cluster centroids from the FCM analysis as visualized by a 2-D projection on the plane defined by each pair of input variables. Results for two to five clusters are shown in panels a to d. Red circles are observational data and black squares are cluster centroids.

Figure S9. Map of Manaus city depicting population density as well as main avenues and representative locations of industry, restaurants, and other businesses. Population density data are from the 2010 census by the Brazilian Institute of Geography and Statistics (IBGE, 2010).

Figure S10. Measurements showing the geographical heterogeneity of emissions from Manaus. On the top row, concentrations of sulfate (red) and particle number (white) measured onboard the G-1 aircraft on (a) March 19 and (b) Mar 21. Image data: Google earth. On the bottom row, rose plots of mean (c) sulfate mass concentrations and (d) particle number concentrations observed at T2 during IOP1. The angles represent wind direction, the radial scale (0 to 5 m s^{-1}) represents wind speed, and the color scale represents the concentrations. The interactions of emissions from Manaus with the daily river breeze is complex, and the detailed interpretation of the data sets is not fully attempted herein. Of importance, the river breeze terminates well below 500 m based on the G-1 flights so that the complexities of the river breeze largely do not affect the measurements at T3 because most pollution is lofted above the river breeze before reaching T3 (Medeiros et al., in preparation). These surface-level plots, although complicated by the river breeze, demonstrate the heterogeneity of Manaus emissions.

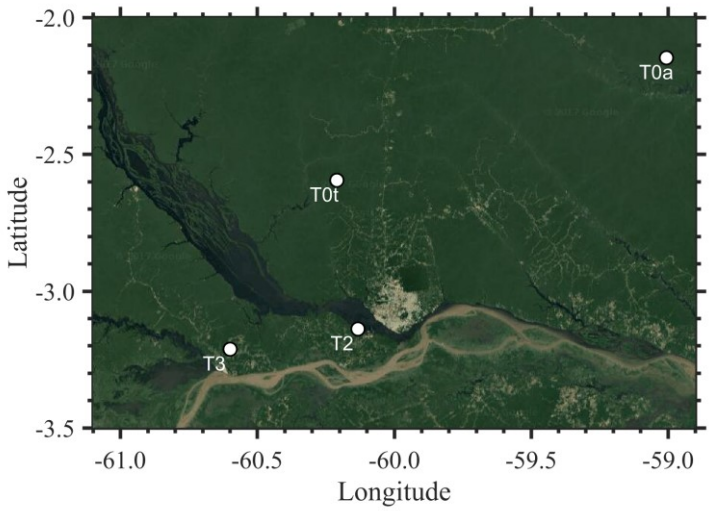
-2.0
-2.5
-3.0
-3.5
Latitude
-61.0
-60.5
-60.0
-59.5
-59.0
Longitude
T0a
T0t
T2
T3

Figure S1

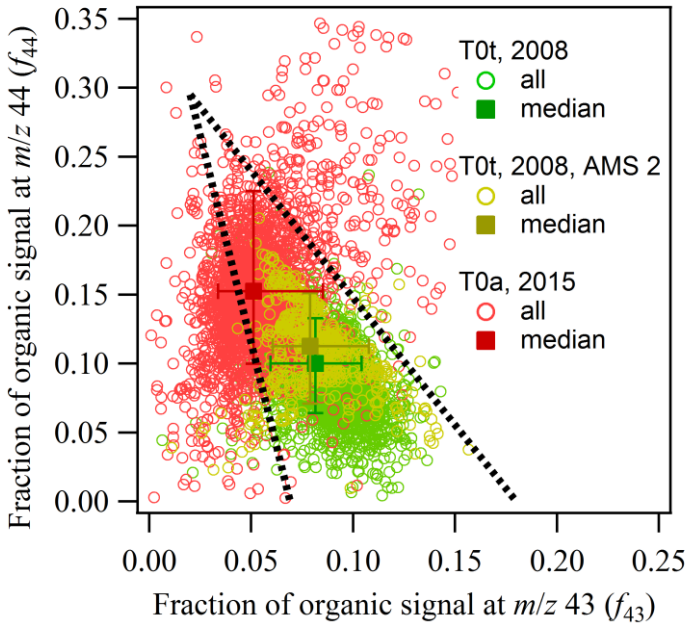
T0t, 2008
all
median
T0t, 2008, AMS 2
all
median
T0a, 2015
all
median
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.00
0.05
0.10
0.15
0.20
0.25
Fraction of organic signal at m/z 44 (f44)
Fraction of organic signal at m/z 43 (f43)

Figure S2

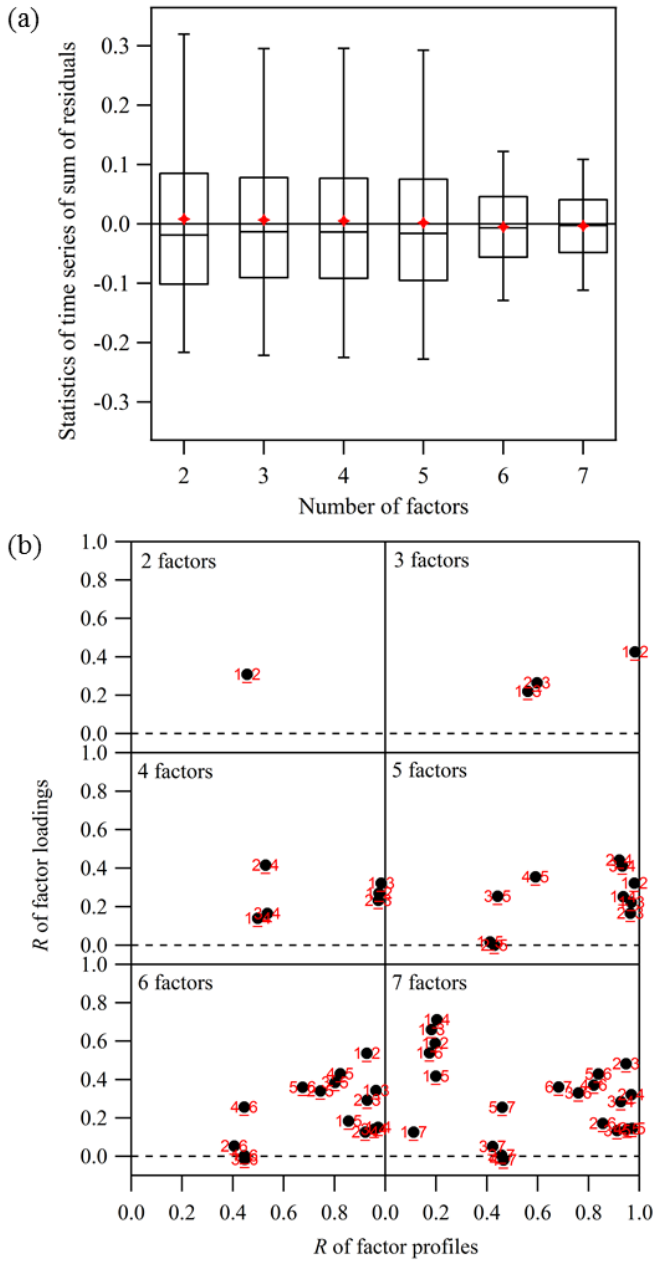
(a)
Statistics of time series of sum of residuals
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
2
3
4
5
6
7
Number of factors
(b)
2 factors
3 factors
4 factors
5 factors
6 factors
7 factors
R of factor loadings
R of factor profiles
0.0
0.2
0.4
0.6
0.8
1.0

Figure S3

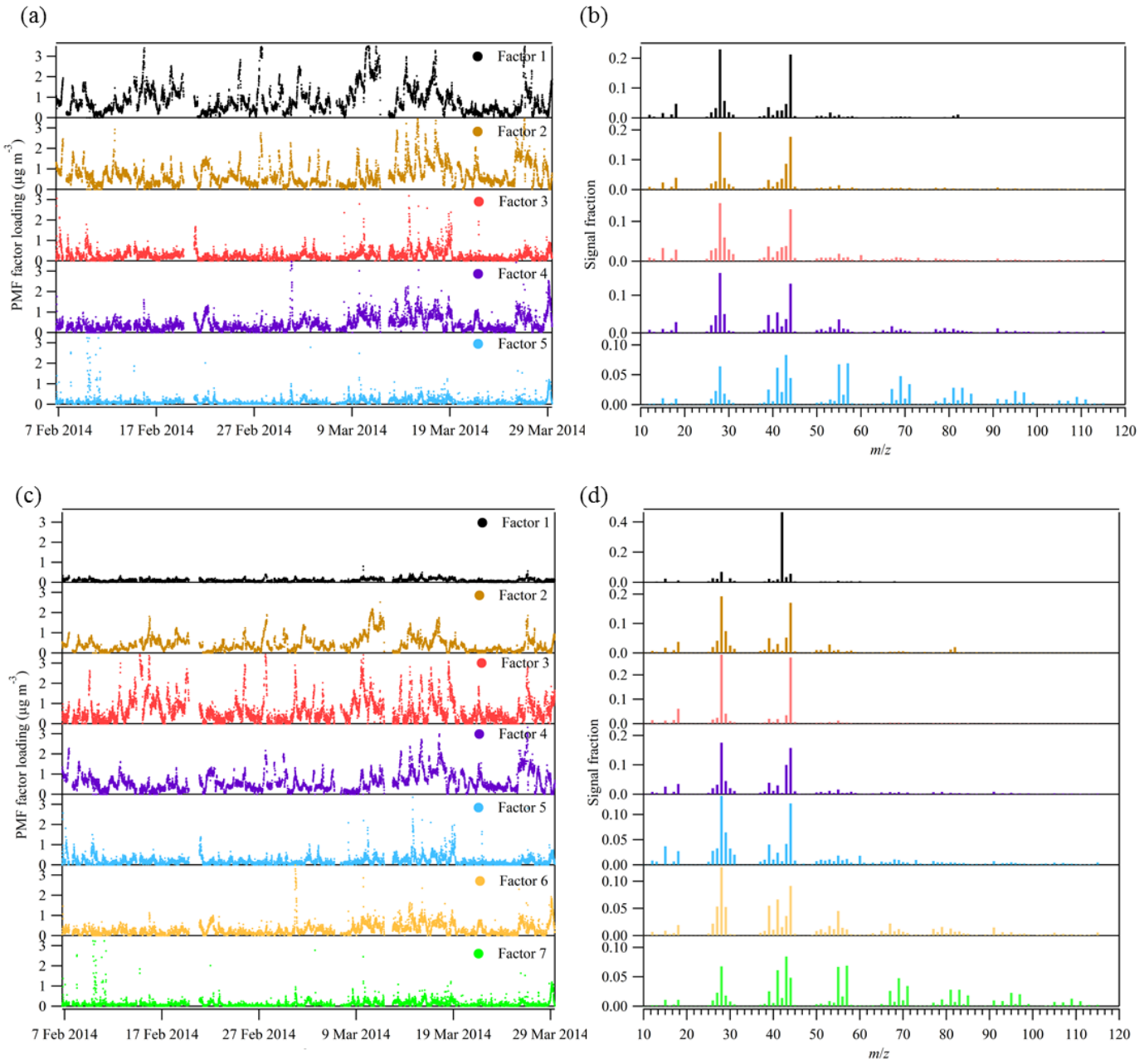
(a)
(b)
(c)
(d)
Factor 1
Factor 2
Factor 3
Factor 4
Factor 5
Factor 6
Factor 7
PMF factor loading (μg m-3)
Signal fraction
m/z
7 Feb 2014
17 Feb 2014
27 Feb 2014
9 Mar 2014
19 Mar 2014
29 Mar 2014

Figure S4

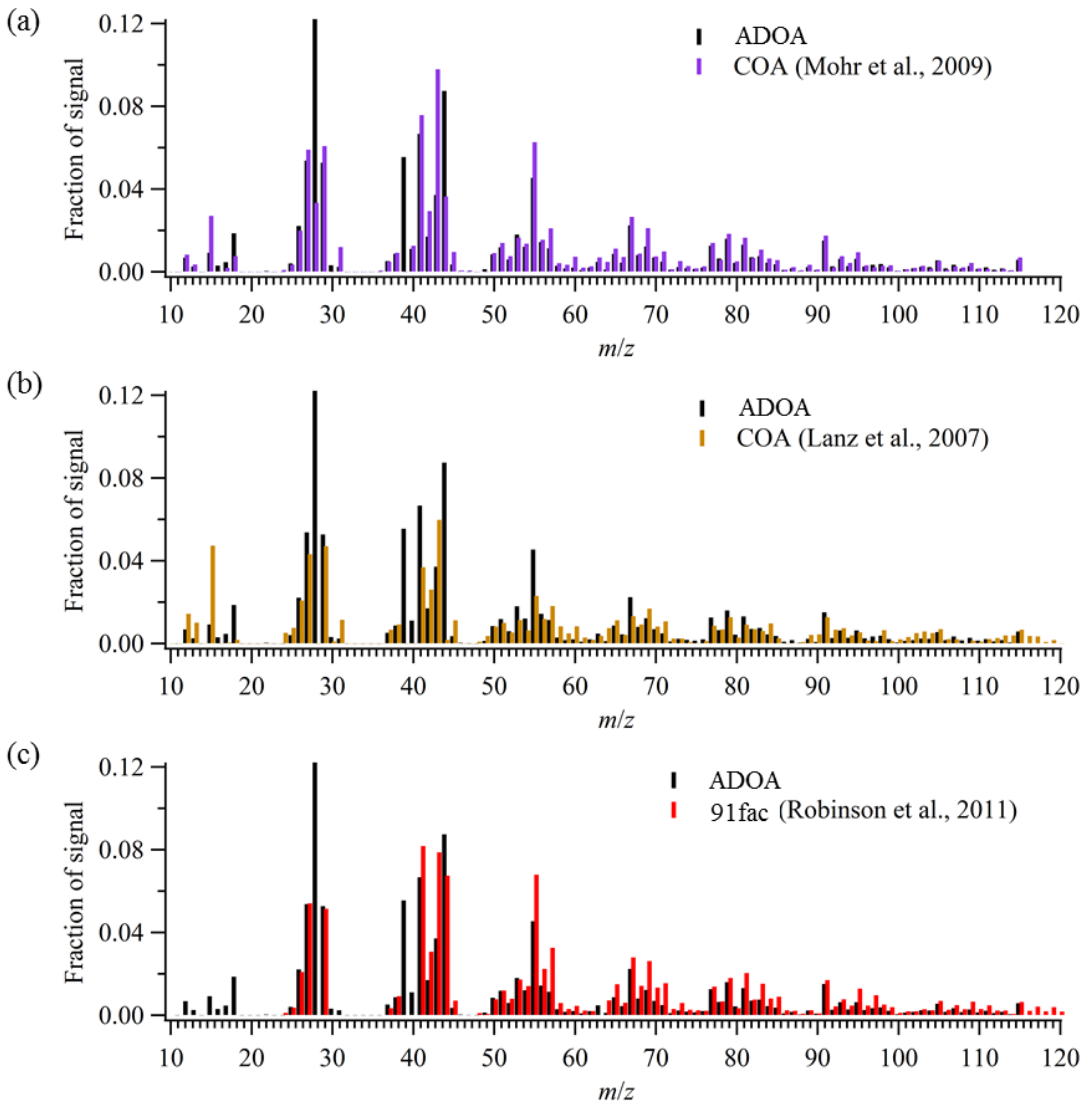
(a)
ADOA
COA (Mohr et al., 2009)
Fraction of signal
0.12
0.08
0.04
0.00
10 20 30 40 50 60 70 80 90 100 110 120
m/z
(b)
ADOA
COA (Lanz et al., 2007)
Fraction of signal
0.12
0.08
0.04
0.00
10 20 30 40 50 60 70 80 90 100 110 120
m/z
(c)
ADOA
91fac (Robinson et al., 2011)
Fraction of signal
0.12
0.08
0.04
0.00
10 20 30 40 50 60 70 80 90 100 110 120
m/z

Figure S5

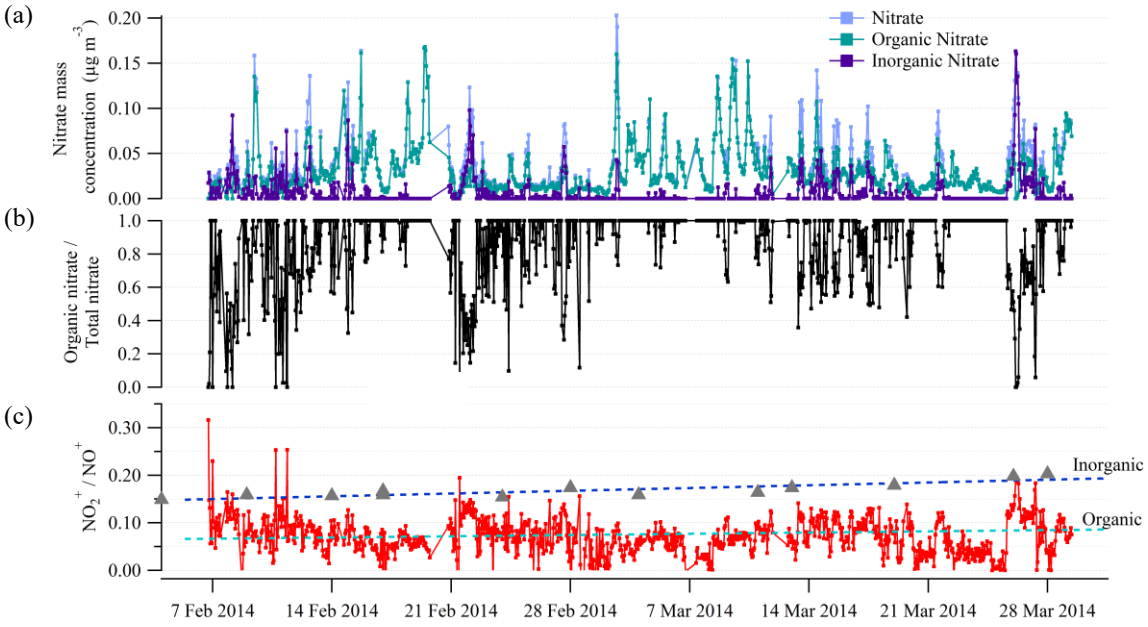
(a)
Nitrate mass concentration (μg m-3)
0.20
0.15
0.10
0.05
0.00
Nitrate
Organic Nitrate
Inorganic Nitrate
(b)
Organic nitrate / Total nitrate
1.0
0.8
0.6
0.4
0.2
0.0
(c)
NO2+ / NO+
0.30
0.20
0.10
0.00
Inorganic
Organic
7 Feb 2014
14 Feb 2014
21 Feb 2014
28 Feb 2014
7 Mar 2014
14 Mar 2014
21 Mar 2014
28 Mar 2014

Figure S6

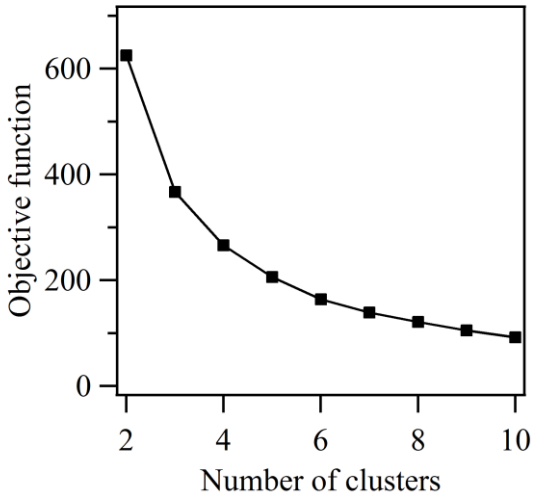

Figure S7

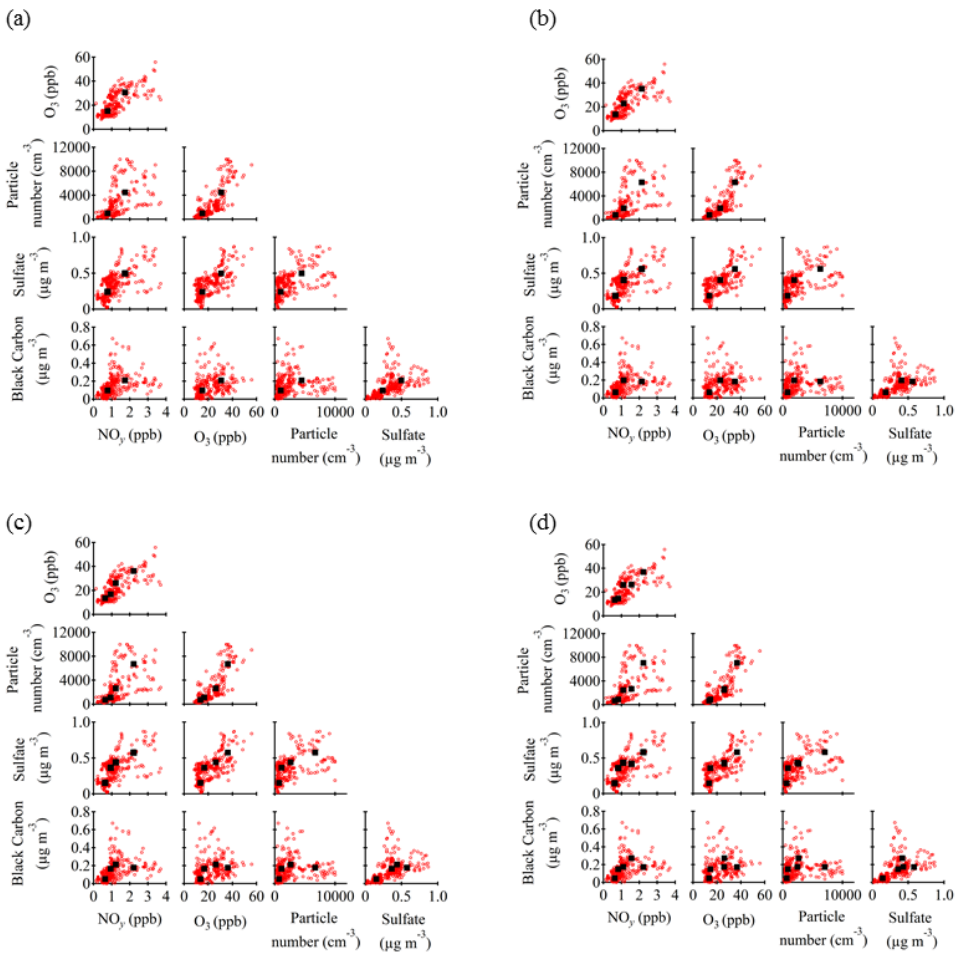
(a)
(b)
(c)
(d)
O3 (ppb)
Particle number (cm-3)
Sulfate (μg m-3)
Black Carbon (μg m-3)
NOy (ppb)
O3 (ppb)
Particle number (cm-3)
Sulfate (μg m-3)

Figure S8

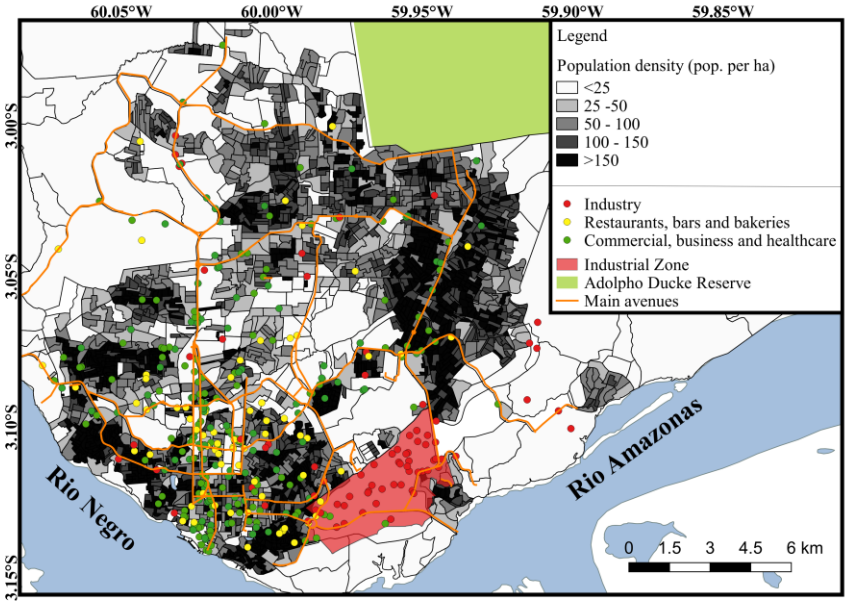
60.05°W
60.00°W
59.95°W
59.90°W
59.85°W
3.00°S
3.05°S
3.10°S
3.15°S
Legend
Population density (pop. per ha)
<25
25 -50
50 - 100
100 - 150
>150
Industry
Restaurants, bars and bakeries
Commercial, business and healthcare
Industrial Zone
Adolpho Ducke Reserve
Main avenues
Rio Negro
Rio Amazonas
0 1.5 3 4.5 6 km

Figure S9

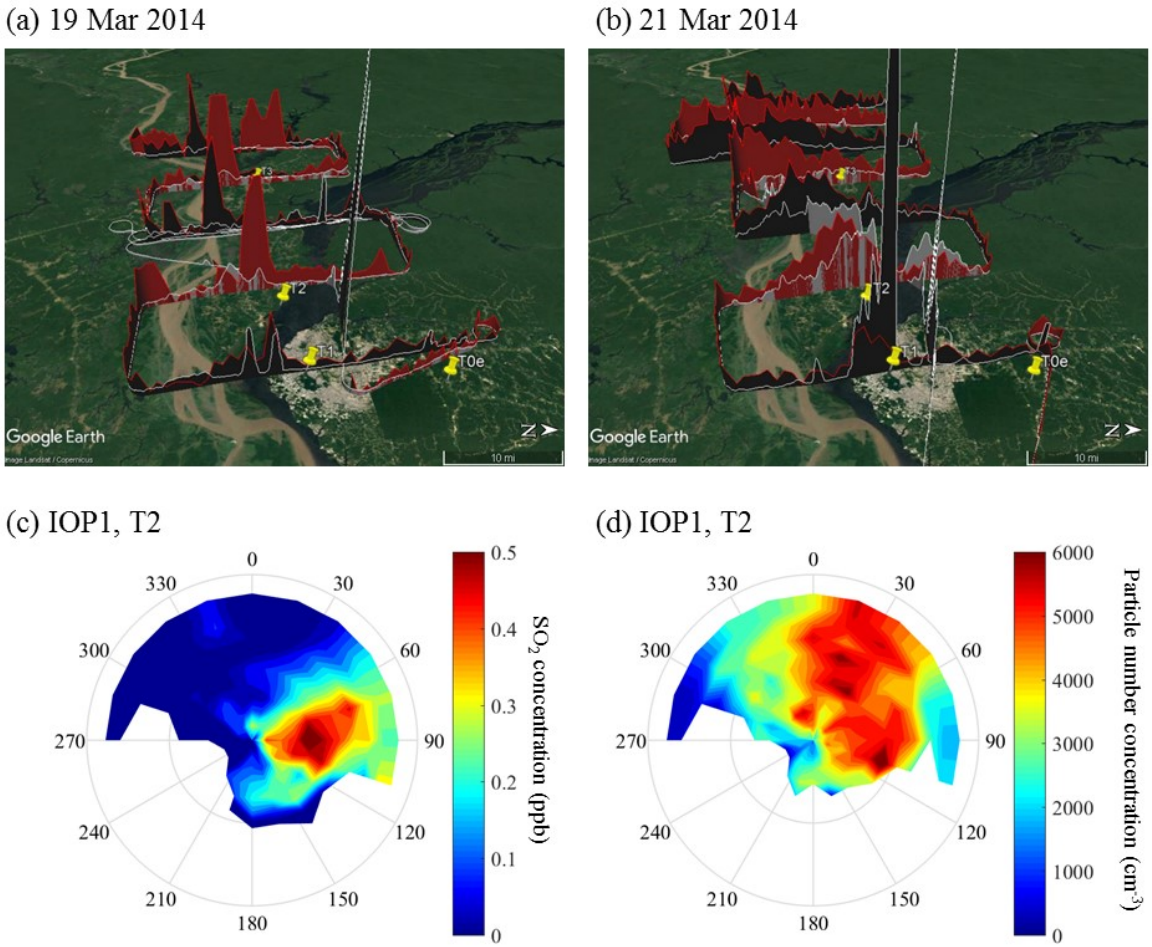
(a) 19 Mar 2014
(b) 21 Mar 2014
T3
T2
T1
T0e
Google Earth
10 mi
(c) IOP1, T2
(d) IOP1, T2
0
30
60
90
120
150
180
210
240
270
300
330
0.5
0.4
0.3
0.2
0.1
0
SO2 concentration (ppb)
6000
5000
4000
3000
2000
1000
0
Particle number concentration (cm-3)

Figure S10