Tropical tropospheric aerosol sources and chemical composition observed at high-altitude in the Bolivian Andes

Supplementary material

--- Sampling site ---

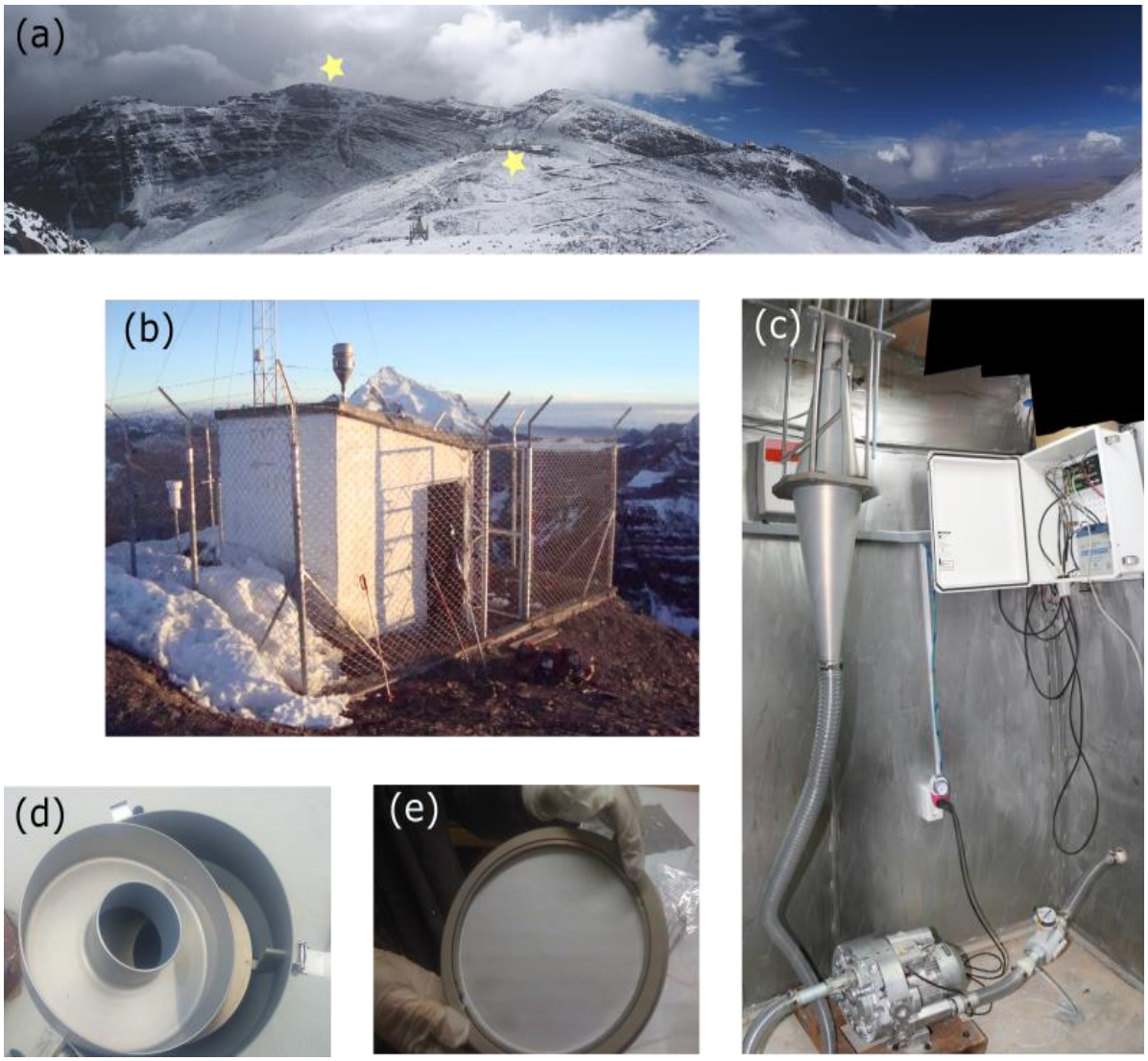

Figure S 1. (a) Measurement site: left star = high-volume sampling site, right star = main GAW-Cosmic Ray observatory managed by the University Mayor de San Andrés (the sampling site for aerosol in the 1970-1980 was located just above the main observatory, not in the summit). (b) Sampling hut with the Digitel head visible. Also meteorological variables are collected at the site (the 10-meter tower partially visible has an anemometer on top) (c) High-Volume sampler system (d) Water-collection system inside the sampling head (missing until 2016) (e) Filter holder with clean filter as set for sampling. Pictures taken by I. Moreno

--- Meteorological information ---

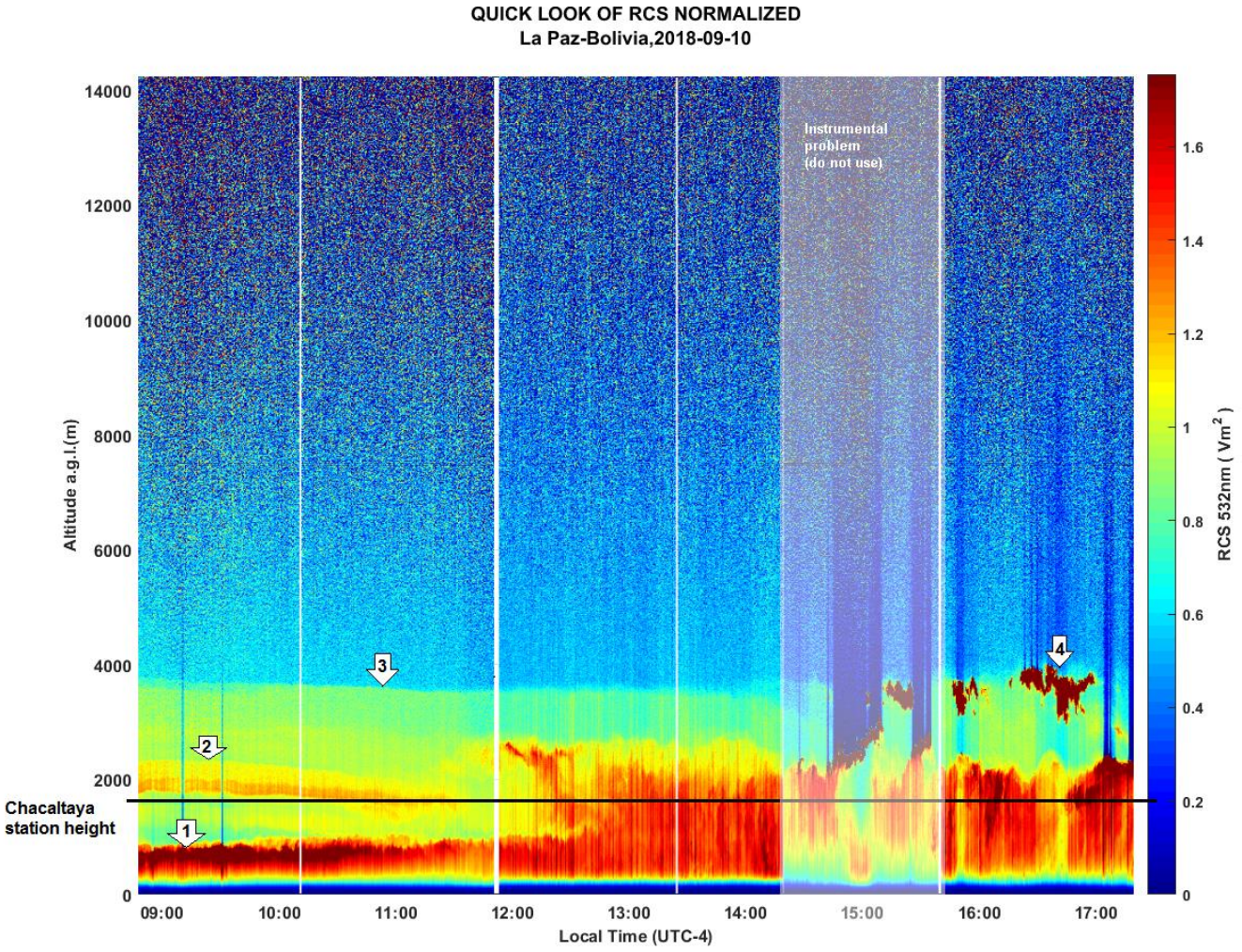

Figure S 2. Range-corrected signal (RCS) of lidar measurements at the Cota Cota location (LP-EA in Figure 1) showing an example of a multi-layered boundary/mixing layer. Ground level is 3640 m a.s.l. The horizontal line shows the Chacaltaya station height. The cleared area between 14:30 and 15:30 corresponds to an instrumental problem for which the absolute RCS is not correct. The numbered arrows stand for:
1= Top of the valley mixing layer
2= Top of the nocturnal residual layer
3= Top of the regional residual layer
4 = Clouds/ condensation level.
The image was taken during the core of the biomass burning season, when smoke and other by-products are emitted in large quantities in the Amazonian basin (with a comparatively small influence from the Altiplano). Prepared by R Forno, MF Sánchez, the use is authorized by the authors.

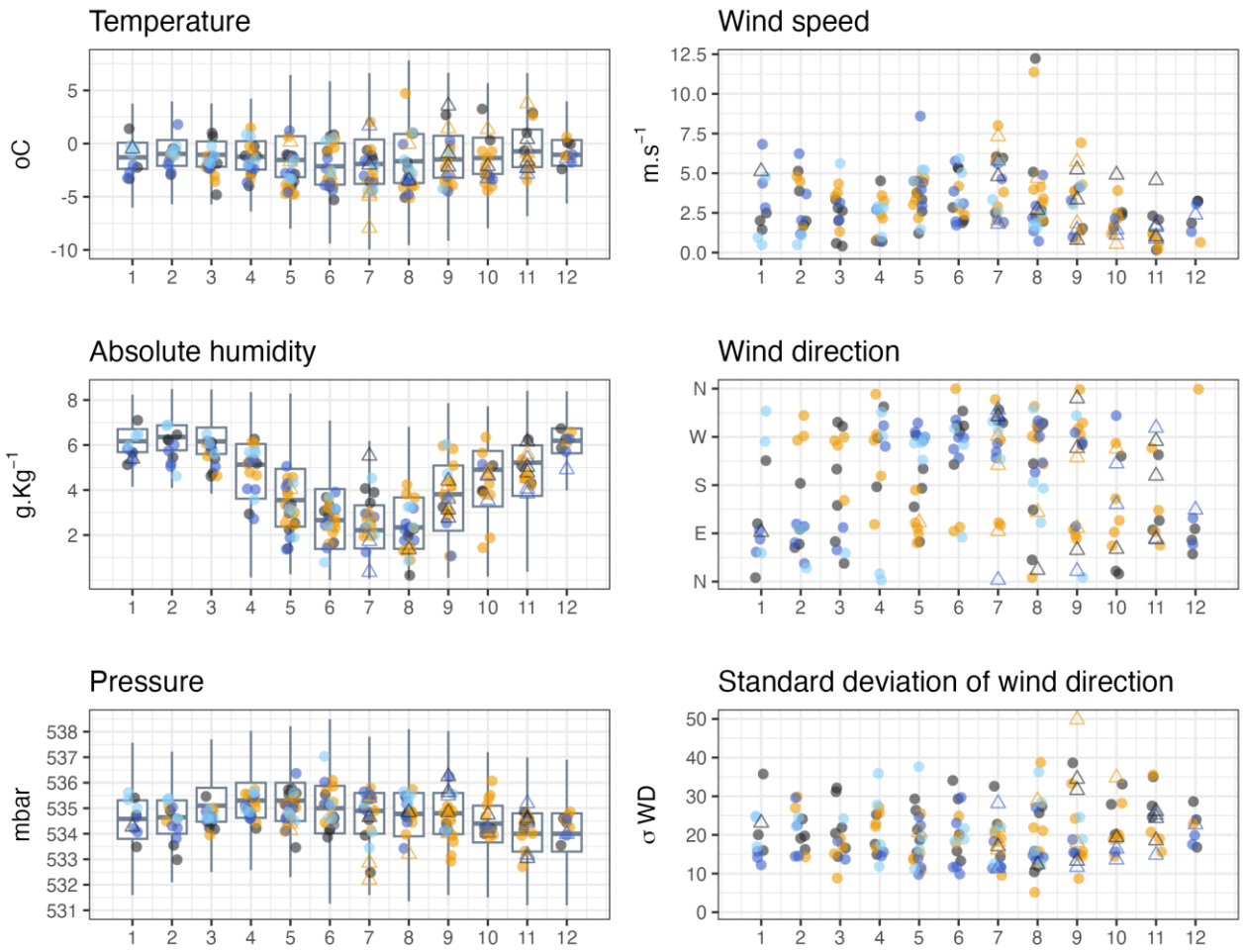

Figure S 3. Average meteorological conditions for each sample (for wind, vector average was used). Boxplots represent the climatology (data from 2011 to 2020), not presented for winds for clarity. Brown markers = PM_{10}-A, yellow markers = PM2.5, dark green markers = PM_{10}-B, light green markers = PM_{10}-C. Open triangles represent the outliers that were removed from Figures 5,6,7 and Figure 5S.

--- Standard temperature and pressure (STP) correction ---

The volume sampled in CHC was obtained in ambient conditions, 534 hPa, which is nearly half of the pressure at sea level. This implies that the volume sampled at CHC is equivalent to twice the volume sampled at sea-level. For comparison with other sites, the concentration was transformed to standard temperature and pressure (STP) conditions (P=1013.25 hPa, T=273ºK) taking into account the average meteorological conditions (pressure, P_{avg} and temperature, T_{avg}) for each sample (equations S1 and S2). When no meteorological parameters were available, the annual mean pressure and/or temperature were used to correct the volume to standard cubic meters.

$$C_{STP} = C_{CHC} \times F \text{ (equation S1)}$$

In which *F* mean value was 1.88 (σ = 0.01), ranging from 1.85 to 1.93 and was calculated for each sample following equation 2:

$$F = \frac{1013.25[hPa] \times (T_{avg}[℃] + 273)}{P_{avg}[hPa] \times 273[K]} \text{ (equation S2)}$$

Concentrations are then expressed in standard cubic meters for comparison to other sites (and other sites' ambient data were also converted to STP conditions). Note, however, that this extensively used simple correction does not take into account the effect of temperature and pressure changes in volatility of the measured species, and therefore the true reference values remain those measured at the site.

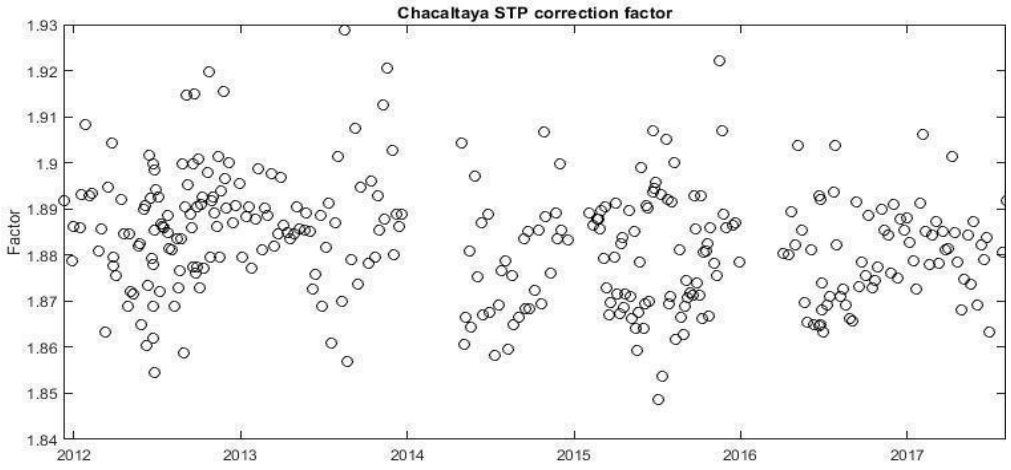

Figure S 4. Variability of STP correction factor for Chacaltaya samples

--- Detection limit per analyzed batch ---

Table S 1. Quantification limit for the analyzed species at Chacaltaya. Data from IGE, Grenoble.

Year (sample number)	**2011-2014 (1-171)**	**2015 (173-248)**	**2016 (250-270)**	**2017 (273-289)**	**2016 (B42-B43)**	**2017 (B44-B48)**	**2016 (INHALE)**	**2019 (291-323)**	**2020 (324-330)**	
Species	**QL**	**QL**	**QL**	**QL**	**QL**	**QL**	**QL**	**QL**	**QL**	**Units**
OC	0.03	0.09	0.08	0.03	0.20	0.20	0.05	0.08	0.02	$\mu g.m^{-3}$
EC	0.002	0.0001	0.003	0.001	0.02	0.02	0	$3x10^{-6}$	0	$\mu g.m^{-3}$
F^-	0.20	0.01	0.15	0.17	NA	NA	0.16	0.02	0.001	$ng.m^{-3}$
HCO_2^-	0.63	3.99	7.60	0.91	NA	NA	NA	0.99	0.03	$ng.m^{-3}$
$MeSO_3^-$	0.04	0.26	0	0.03	0.73	0.06	0.15	0.01	0.004	$ng.m^{-3}$
Cl^-	0.41	1.56	0.80	0.34	13.59	2.33	1.17	0.25	0.15	$ng.m^{-3}$
Br^-	0.10	0.12	0.08	0.09	NA	NA	0.09	0.02	0.03	$ng.m^{-3}$
NO_3^-	0.22	3.43	5.10	0.30	14.60	3.46	41.45	1.17	0.85	$ng.m^{-3}$
SO_4^{2-}	2.39	8.60	4.26	4.18	3.83	3.48	14.91	1.95	0.26	$ng.m^{-3}$
$C_2O_4^{-22-}$	0.21	0.44	0.15	0.16	1.67	1.05	0.84	0.39	0.06	$ng.m^{-3}$
Li^+		0.006			NA	NA	NA	0.03	0.001	$ng.m^{-3}$
Na^+	0.66	3.30	1.47	2.31	14.69	4.99	4.90	0.49	0.38	$ng.m^{-3}$
NH_4^+	0.96	2.88	3.95	1.61	9.81	5.40	6.63	1.06	0.60	$ng.m^{-3}$
K^+	0.39	0.81	0.60	0.34	7.96	1.63	17.76	0.47	0.39	$ng.m^{-3}$
Mg^{2+}	0.07	0.15	0.30	0.17	0.51	0.39	0.42	0.01	0.03	$ng.m^{-3}$
Ca^{2+}	1.49	1.83	1.40	2.21	10.17	2.53	2.29	0.12	0.07	$ng.m^{-3}$
arabitol	0.25	0.53	0.33	0.33	0.77	0.33	0.22	0.22	0.09	$ng.m^{-3}$
sorbitol	0.22	1.94	0.13	0.13	1.27	0.13	NA	1.08	0.04	$ng.m^{-3}$
mannitol	0.19	0.78	0.33	0.33	0.77	0.33	0.22	0.19	0.09	$ng.m^{-3}$
levoglucosan	0.27	0.26	1.65	1.65	0.62	1.65	0.17	0.95	0.46	$ng.m^{-3}$
mannosan	0.05	0.33	0.33	0.33	0.77	0.33	0.22	0.19	0.09	$ng.m^{-3}$
galactosan	0.29	0.13	0.13	0.13	0.31	0.13	0.09	0.08	0.04	$ng.m^{-3}$
glucose	0.51	0.33	0.33	0.33	0.77	0.33	0.22	0.19	0.09	$ng.m^{-3}$

---- Outlier values from figures 5, 6, 7 and S5 ----

Table S 2. Outliers of the dataseries

Ys	Ms	Ds	SN	ST	HeadN	ng/m3 amb	Species defining the outlier
2012	1	16	4	2	10	6,59	glucose
						2,97	mannitol
2012	7	27	47	2	10	5,38	arabitol
2012	8	9	49	-2	10	104	isolevo
2012	9	4	57	2	10	192	EC
2012	9	4	57	2	10	5,86	F
						6,67	F
2012	9	14	60	2	10	223	EC
						123	oxalate
2012	9	17	61	-2	10	0,114	Li
2012	10	26	74	-2	10	14,4	arabitol
2012	11	5	77	-2	10	66,3	Na
2012	11	16	80	-2	10	11,1	arabitol
						0,244	Li
2012	11	26	83	-2	10	3,33	mannitol
2013	9	9	124	4	2.5	0,123	Li
2014	5	29	145	5	2.5	2177	SO4
						15,5	Mg
						13,7	H
						240	Ca
2014	7	11	151	-5	2.5	176	Ca
2014	10	27	166	5	2.5	14,1	MSA
2015	5	12	197	6	2.5	3010	OC
2015	7	3	212	-6	2.5	6,14	F
						195	Ca
2015	7	31	218	-6	2.5	9,34	glucose
						17,3	H
						2007	SO4
2015	8	7	220	6	2.5	25,9	Cl
2015	9	4	226	6	2.5	13,1	H
2015	11	16	243	6	2.5	4,03	mannitol
2015	11	23	245	-6	2.5	10,9	H
2016	7	25	260	6	10	134	isolevo
						151	formato
2016	11	18	262	-6	10	182	EC
2016	12	2	263	99	10	189	EC
2016	12	16	265	-6	10	3,03	mannitol
2016	7	29	260.01	-6	10	104	isolevo
2016	9	16	260.08	-6	10	4,02	Br
						12,7	H
2016	10	21	260.13	-6	10	49,8	Na
2016	10	28	260.14	-6	10	31,5	Cl
						10,8	H
						81,3	Na
						17	Mg
2016	11	4	260.16	99	10	14,3	Mg

--- Complete dataseries ---

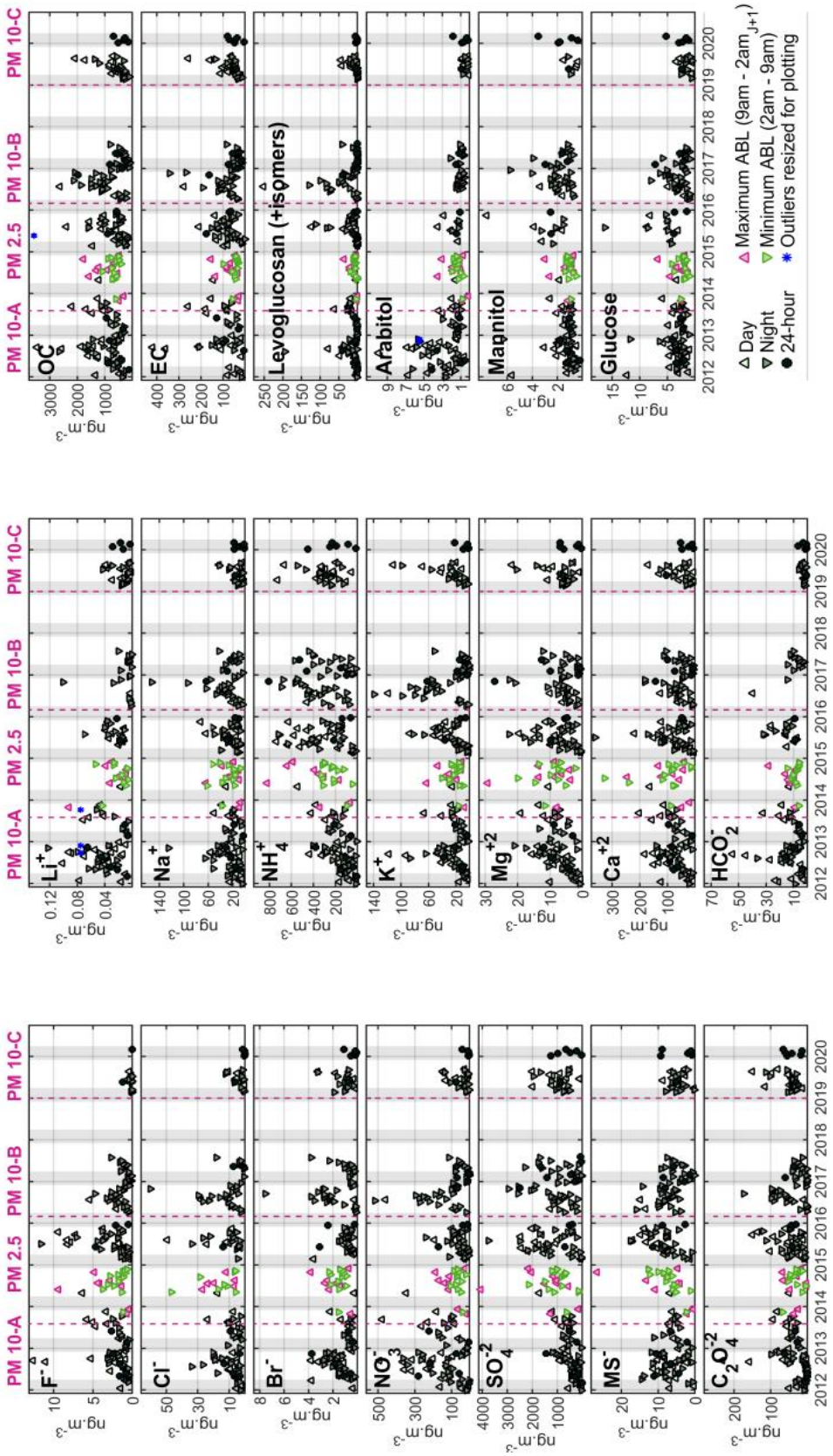

Figure S 5 Concentration of ions, OC, EC and anhydrosugars measured in Chacaltaya at STP conditions. The dashed lines show the difference between PM_{10} and PM 2.5 sampling periods. Upward pointing triangles are for daytime sampling, downward pointing triangles for nighttime and circles for 24-hour sampling. The color markers identify sampling conditions more likely to represent the difference between maximum (pink) and minimum (green) atmospheric boundary layer influence at Chacaltaya. Blue asterisks are extreme outliers whose values are found in the supplementary material. Gray shaded area highlights the wet season (DJFM). Note that F^-, Cl^-, NO_3^- and formiate concentrations for PM_{10}-C samples are suspected of presenting losses, and were excluded from the analysis but presented here to show the concentrations' drop.

In figure S5, the higher nitrate concentration of PM_{10}-A (149 ng m^{-3}) compared to PM_{10}-B (110 ng m^{-3}) remains unexplained.

--- Source apportionment using positive matrix factorization PMF EPA v 5.0 ---

- PM_{10} and $PM_{2.5}$ run together as most species seem to be in the $PM_{2.5}$ fraction
- Information about species used:

Table S 3. Species used and discarded for the source apportionment

N	Species	Uncertainty (%)	Total missing data (%)	Not measured (%)	Signal to noise ratio	Category
1	OC*	15%	4	0	5.7	Strong
2	EC	20%	14	0	3.4	Strong
3	$C_2O_4^{2-}$	20%	6	0	3.8	Strong
4	MSA	20%	6	0	3.8	Strong
5	SO_4^{2-}	15%	0.9	0	5.6	Strong
6	NH_4^+	15%	3	0	5.5	Strong
7	Br^-	20%	17	15	3.9	Strong
8	Na^+	20%	12	0	3.5	Strong
9	K^+	20%	6	0	3.7	Strong
10	Mg^{2+}	20%	5	0	3.8	Strong
11	Ca^{2+}	20%	4	0	3.8	Strong
12	Levoglucosan	20%	20	0	3.2	Strong
13	Arabitol	20%	30	0	2.8	Weak
14	Mannitol	20%	35	0	2.6	Weak
15	Glucose	20%	24	0	3.0	Weak
	For	20%	49	0.4	--	Bad
	Mannosan	20%	52	0	--	Bad
	NO_3^- PM_{10}-C lost	15%	20	0	4.5	Bad
	F^- PM_{10}-C lost	20%	25	0.4	3.0	Bad
	Cl^- PM_{10}-C lost	20%	43	0.4	2.3	Bad

- Species with no more than 50% of missing values should be used (Belis et al. 2019). In this regard, mannosan (52% of missing data) and formate (49% of missing data) were discarded
- NO_3^-, Cl^-, F^- were removed because they pull the solution to a factor only due to their absence in $PM_{10}C$.
- Outliers: 8 samples excluded (215 samples remaining): 2012-10-26, 2012-10-16 Highest arabitol concentration; 2015-5-12 high OC*; 2014-10-27 high MSA; 2014-5-29, 2015-7-31 high SO_4^{2-} and NH_4^+; 2012-9-14 highest EC; 2016-9-16 highest Br^-
- Missing values due to non-analyzed samples (very few, only missing for F^-, Cl^-, Br^-) were replaced by the median concentration of each species in the whole dataset including values of ½ QL
- As concentrations are very low compared to other sites, uncertainty for abundant species (OC*, SO_4^{2-}, NH_4^+ which have annual values of 0.8, 1.0, 0.3 µg/m^3 respectively) was set to 15%, and 20% for the rest of the species. Uncertainty for missing values was set to 5/6 QL
- Category assignment: "Weak" if > 20% of missing values, "Strong" if ≤ 20% of data missing.
- OC includes adjustment for the carbon contribution of organic species and it is represented by OC*:

$$OC^* = OC - \frac{oxalate}{3.67} - \frac{methanesulfonate}{7.92} - \frac{levoglucosan}{2.25} - \frac{arabitol}{2.53} - \frac{mannitol}{2.53} - \frac{glucose}{2.50}$$

- Factors obtained and constraints applied

Table S 4. Characteristics of the factors obtained by PMF

Factors	Main species	Constraints	Relationship with section 3 "seasonality"
1 Biomass burning	Levoglucosan, K^+		½ of the "biomass burning group", classical biomass burning source
2 Soil dust and bromide	Na^+, Mg^{2+}, Ca^{2+}, K^+, Br^-		½ of the "transport group" mixed with Br- from "biomass burning group". We gain insights of some K+ belonging to this factor
3 Combustion	EC, oxalate OC	Pull down maximally MSA (% dQ =0.50)	½ of the "biomass burning group" -- Coincides with deduction of EC+Ox having common source. May represent mostly the urban contribution.
4 Westward	$SO4^{2-}$, NH_4^+, MSA	Pull down maximally EC (% dQ =0.50)	½ of the "transport group", but represents the two sources (volcanism and marine) identified in the main text of the paper. It is always the best resolved factor. It points to low free tropospheric transport conditions.

- Bootstraps: Number of iterations set to 100, correlation coefficients >0.8.
- Solution evaluation criteria:
  - $Q_{true}=Q_{robust} < 1.5$
  - Residuals per species were mostly centered around 0, falling within the range of -3 and 5, with some exceptions for outliers. It was not possible to make the species fit in the -3 to +3 recommended range.
  - Displacement analysis did not show rotational ambiguity

Table S 5. Fpeak -0.5 run bootstrap

	Base Factor 1	Base Factor 2	Base Factor 3	Base Factor 4	Unmapped
Boot Factor 1	99	1	0	0	0
Boot Factor 2	0	99	1	0	0
Boot Factor 3	0	0	100	0	0
Boot Factor 4	0	0	0	100	0

Table S 6. Constrained run bootstrap

	Base Factor 1	Base Factor 2	Base Factor 3	Base Factor 4	Unmapped
Boot Factor 1	96	1	0	0	3
Boot Factor 2	0	99	0	0	1
Boot Factor 3	0	0	100	0	0
Boot Factor 4	0	0	0	100	0

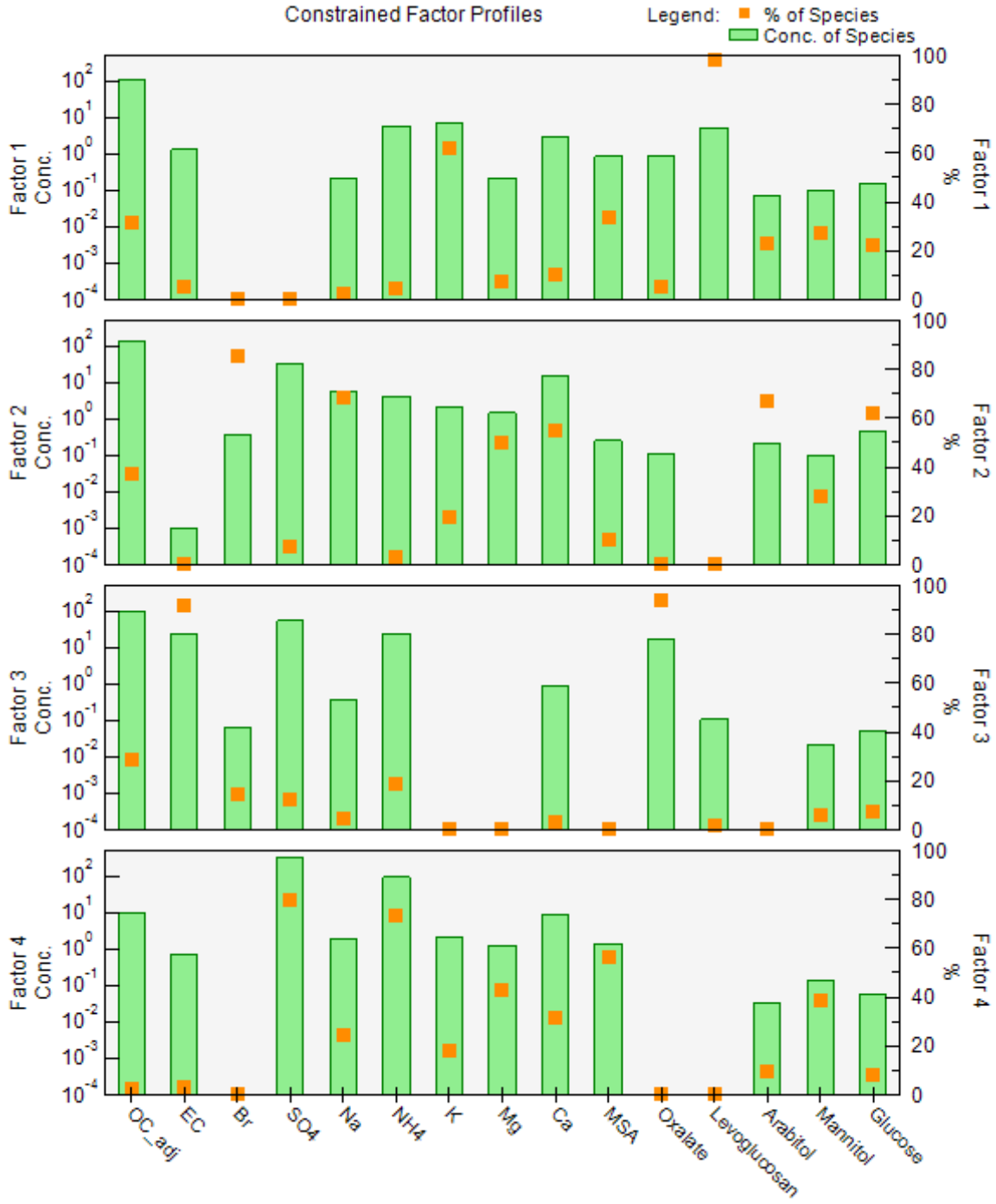

Figure S 6. Constrained factors obtained by PMF. Factor 1= Biomass burning, F

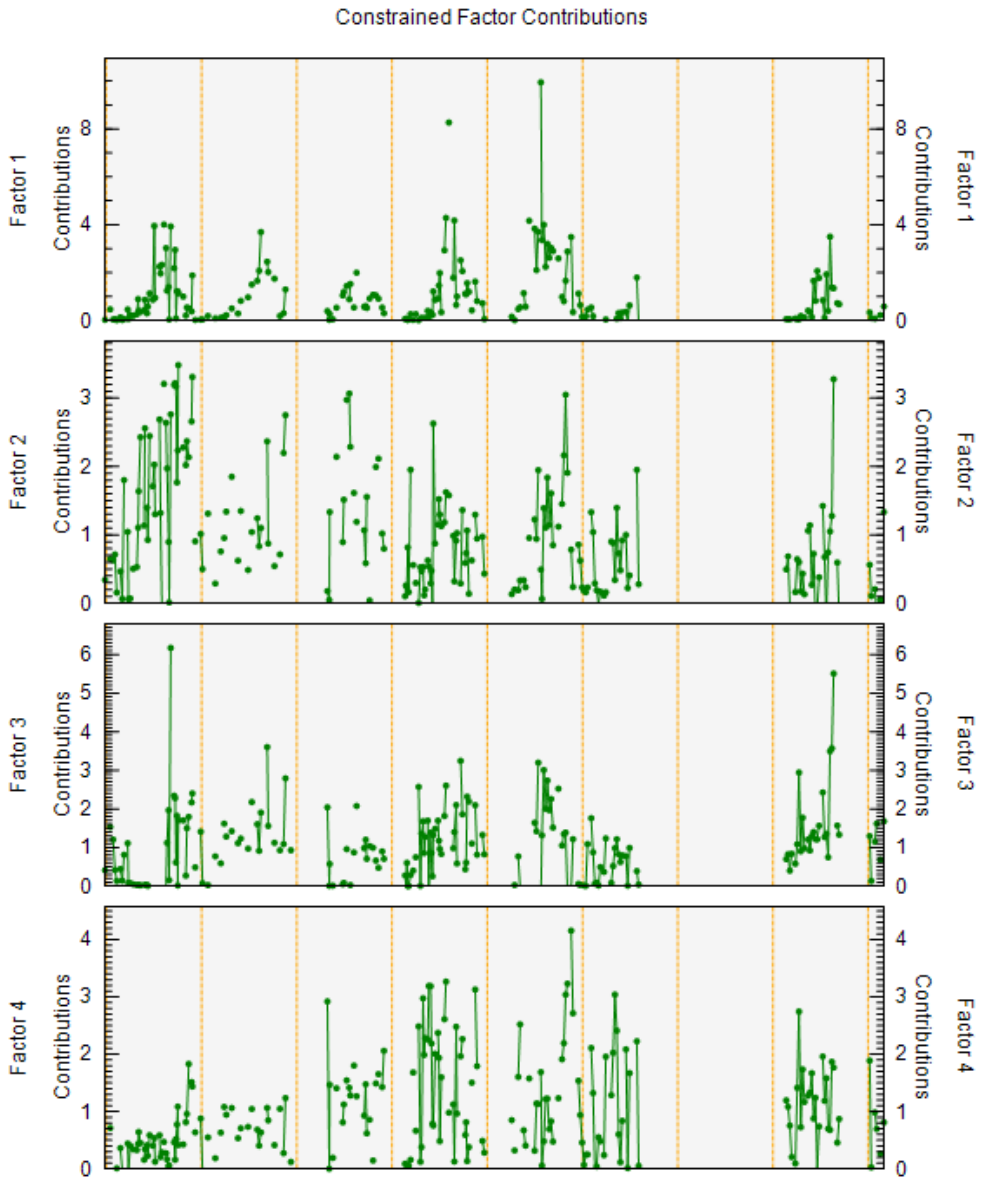

Figure S 7. Contributions of constrained factors obtained by PMF

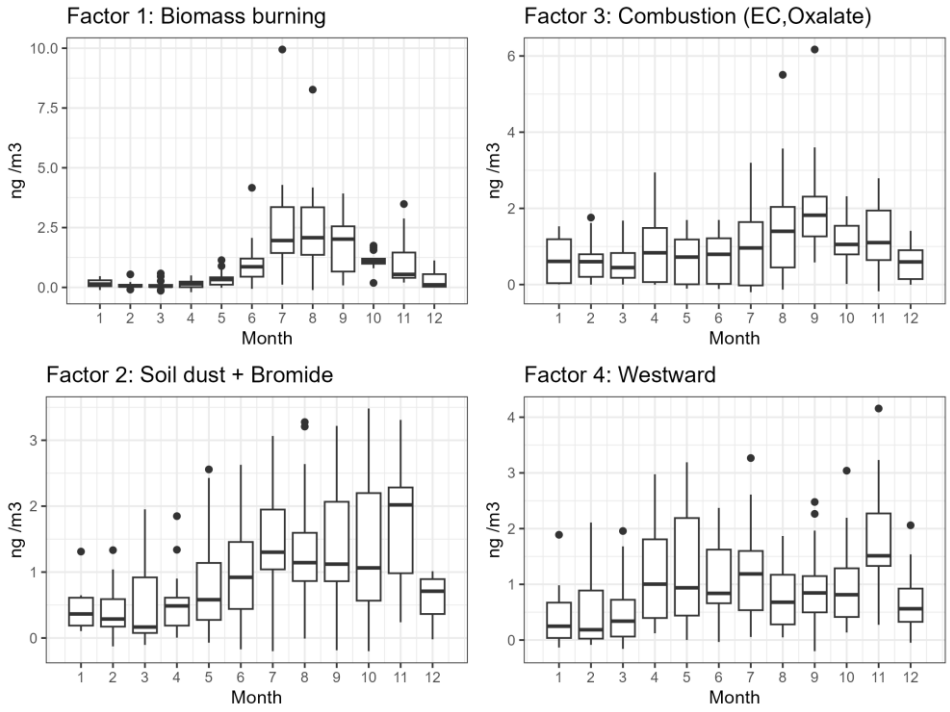

Figure S 8. Seasonality of contributions of constrained factors obtained by PMF

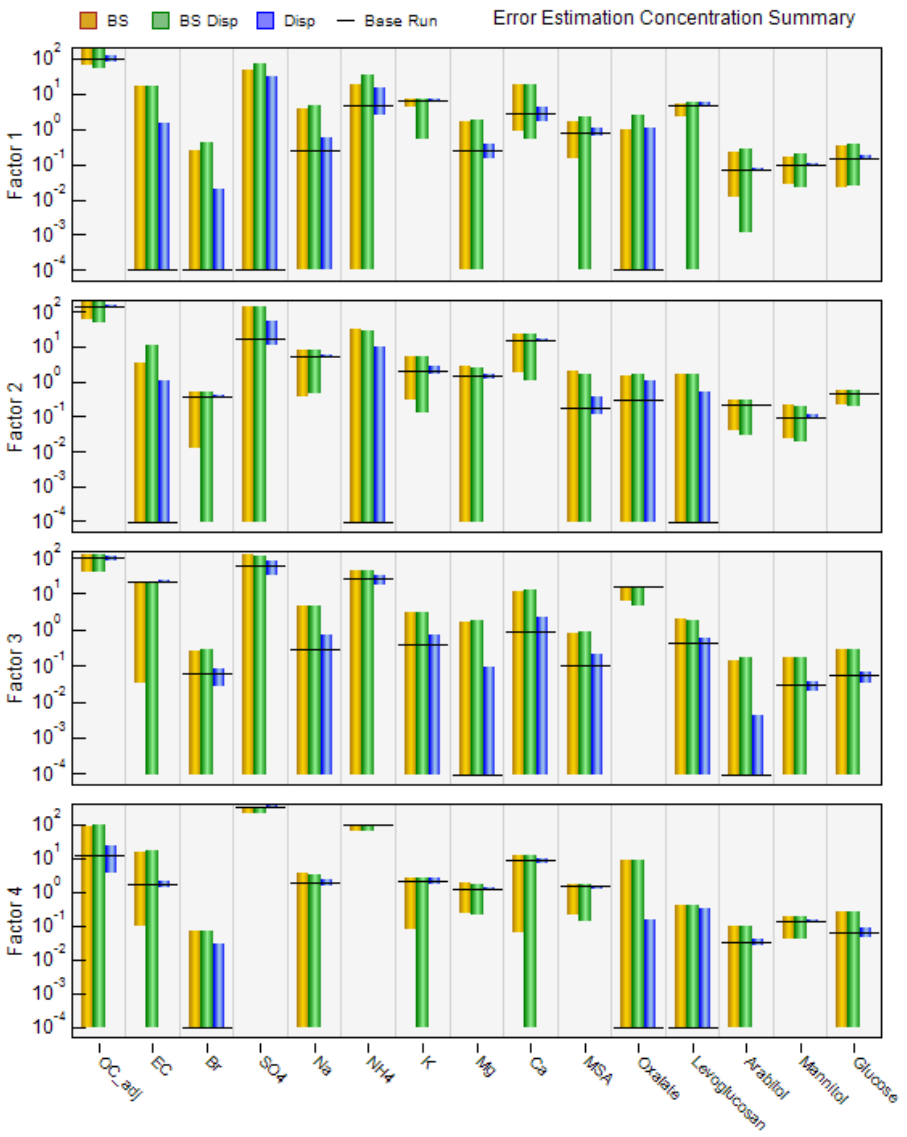

Figure S 9. Error estimation summary

- G-space plot inspection: Factors 1 and 3 are not completely independent in the G-space plot, likely indicating the influence of biomass burning in both factors. This is also the reason why both factors are resolved as a single group with clustering tools.

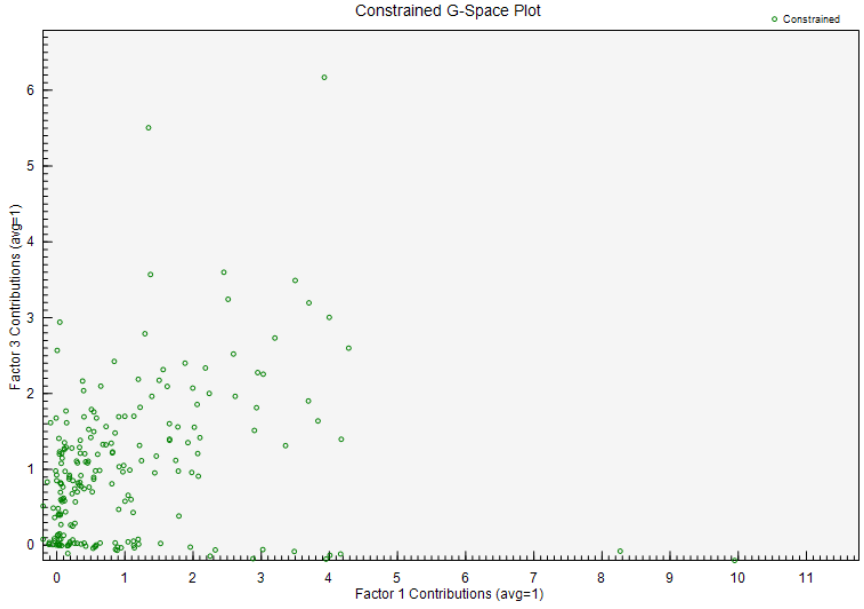

Figure S 10. G-space plot for factors 1 (Biomass burning) and factor 3 (EC+oxalate)

--- Source apportionment for biomass burning based on levoglucosan stereoisomers ---

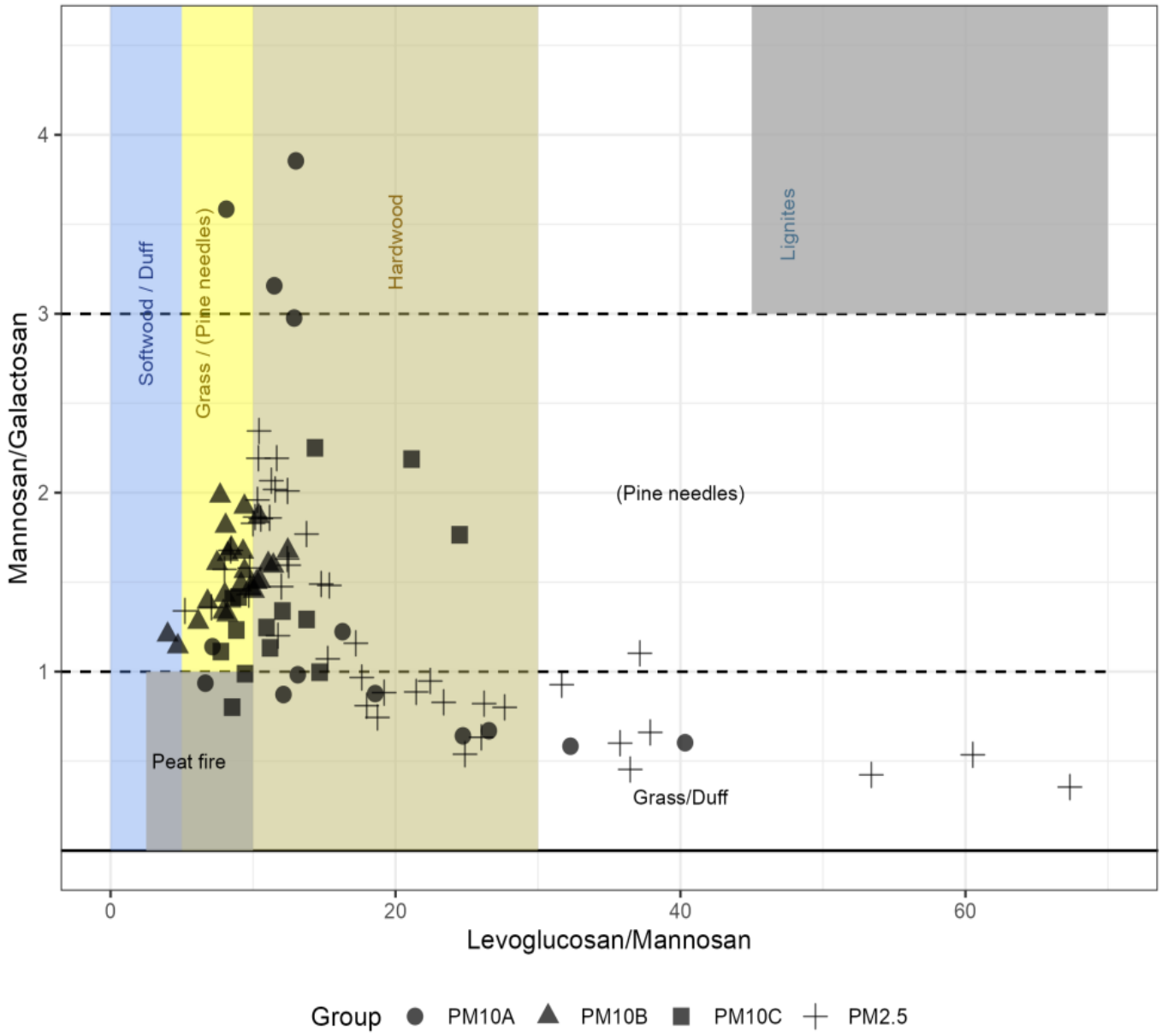

Figure S 11 Diagnostic diagram based on mass ratios of levoglucosan and its stereoisomers. Simplified plot adapted from (Marynowski and Simoneit, 2022) and (Xu et al., 2019) to cover the range of Chacaltaya data. Note that conifer forests are absent in the region and therefore the substrate "pine needles" is negligible for our case.

The highest levoglucosan+mannosan+galactosan concentration is 180 ng m^{-3} and it was found in 2016 (25 - 29/Jul/2016), early in the biomass burning season, as part of a regional event that spanned until mid-November. In our record, 2016 and 2017 were the years with the highest levoglucosan concentrations (Figure S5).

--- Species transported under maximum/minimum influence of the convective planetary boundary layer to the station---

A subset of $PM_{2.5}$ samples was studied to identify if there are differences between daytime and nighttime concentrations of the measured species at Chacaltaya. These 20 samples span from November 2013 to December 2014 (gap from January to March 2014), and they were obtained under maximum and minimum influence of the convective planetary boundary layer at the station. Samples obtained under maximum PBL influence (*maxPBL*) correspond to 17 hours continuously sampled (09:00-02:00 BOT) than include the diurnal development of the convective boundary layer and a nocturnal stable boundary layer that may be capped by a residual layer. Samples obtained under minimum PBL influence (*minPBL*) correspond to seven consecutive hours that include nighttime and early morning, before the arrival of the convective boundary layer to Chacaltaya (02:00-09:00 BOT). This is a period when there can be an influence of a reminiscent residual layer but it also corresponds to the time when we expect to capture more low-free-troposphere events.

Concentration and variability of concentration (represented by σ) for all species is higher during *maxPBL* than during *minPBL* influence are presented in Table S7. Only EC, NH_4^+, NO_3^- and glucose show a significance level (α) <10%. In practical terms, these four species are the only ones for which the transport under *maxPBL* influence (mostly daytime and residual layer) dominates over the transport under *minPBL* influence (nighttime and early morning). The rest of the species present statistically similar concentrations for day and nighttime sampling, and this is the justification to have pooled all types of samples in this work.

Table S 7. Chemical composition for $PM_{2.5}$ samples obtained during maximum (Mixing/Boundary Layer + Residual Layer) and minimum (Residual Layer + Low Free Troposphere) influence of the atmospheric boundary layer at the station Mean ($\overline{x}$), median ($\tilde{x}$), standard deviation (σ) and number of samples above quantification limit (N) are presented. Levoglucosan * includes its stereoisomers mannosan and galactosan. Two-tailed Wilcoxon Rank sum test results as α (significance level in %).

Ambient concentrations ng m^{-3}	Maximum influence (ML+RL) 0900 to 0200 of next day				Minimum influence (RL+LFT) 0200 to 0900				$\frac{\sigma_{maxPBL}}{\sigma_{minPBL}}$	$\frac{\tilde{x}_{maxPBL}}{\tilde{x}_{minPBL}}$	Significance level α (%)
	$\underline{x}$	$\tilde{x}$	σ	N	$\underline{x}$	$\tilde{x}$	σ	N			
NH_4^+	215	185	(122)	8	120	122	(49.7)	12	2.5	1.5	1.5
NO_3^-	61.5	62.0	(28.4)	8	32.8	35.4	(14.0)	12	2.0	1.8	2.3
Glucose	1.78	1.36	(0.94)	7	1.12	1.03	(0.41)	11	2.3	1.3	5.6
EC	42.6	39.0	(24.3)	8	23.0	22.4	(4.68)	11	5.2	1.7	7.5
SO_4^{-2}	824	655	(615)	8	520	490	(262)	12	2.3	1.3	14
OC	549	569	(277)	8	357	313	(95.5)	11	2.9	1.8	16
Mannitol	0.84	0.72	(0.46)	7	0.55	0.59	(0.14)	11	3.2	1.2	18
Br^-	1.12	1.24	(0.57)	7	0.82	0.78	(0.28)	10	2.0	1.6	23
F^-	1.86	1.86	(1.55)	8	0.97	0.71	(0.54)	11	2.8	2.6	24
K^+	17.9	19.4	(10.2)	7	12.03	10.1	(4.61)	11	2.2	1.9	29
Cl^-	8.10	8.36	(5.05)	6	4.84	3.51	(2.46)	7	2.1	2.4	30
$C_2O_4^{-2}$	21.1	20.0	(12.4)	7	15.1	13.5	(9.77)	10	1.3	1.5	32
Ca^{+2}	77.0	55.0	(76.3)	8	52.6	42.5	(51.4)	12	1.5	1.3	34
Levoglucosan *	6.79	4.37	(6.51)	8	4.1	3.81	(2.73)	12	2.4	1.1	47
Mg^{+2}	6.31	4.67	(4.46)	7	4.74	3.96	(2.92)	11	1.5	1.2	48
Arabitol	0.92	0.81	(0.61)	8	0.73	0.78	(0.22)	11	2.8	1.0	60
Li^+	0.02	0.01	(0.02)	7	0.01	0.01	(0.01)	10	1.8	1.1	-
Na^+	13.1	10.1	(10.3)	8	13.4	12.4	(8.16)	12	1.3	0.8	-
$MeSO_3^-$	4.65	3.30	(4.45)	8	3.64	3.27	(1.87)	12	2.4	1.0	-

--- Backtrajectory complementary information ---

The next figure highlights how different the topography of the Chacaltaya station can be, depending on the horizontal resolution of the models. This is the reason why WRF-nested backtrajectories represent the "control" case when compared to ERA-5 backtrajectories.

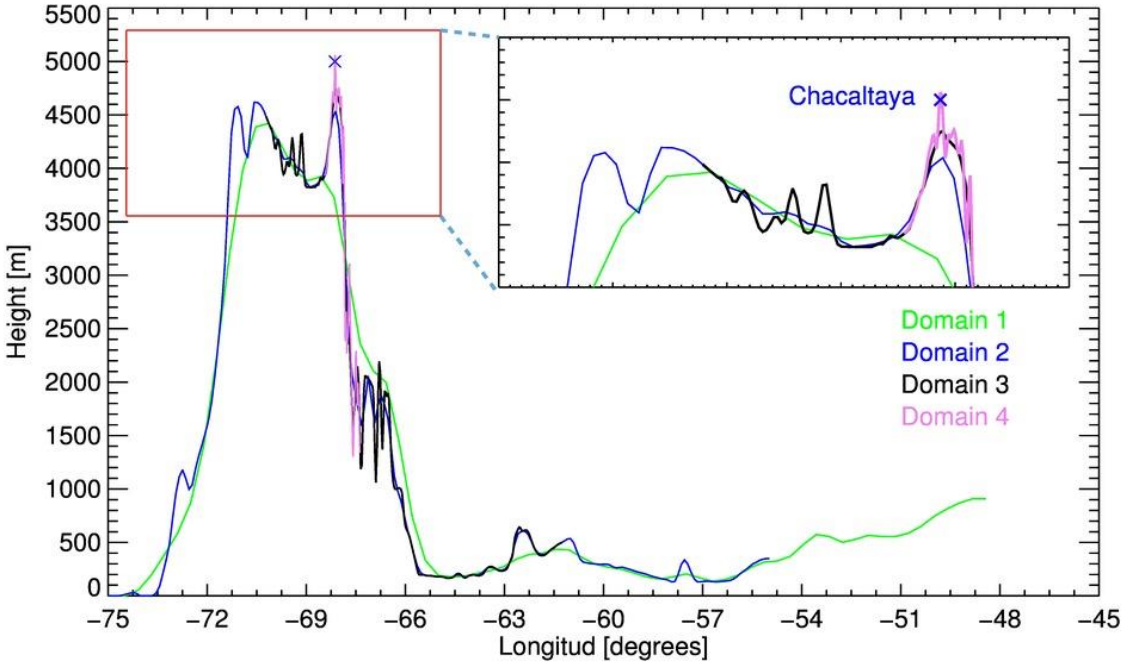

Figure S 12. Topography for the WRF domains used to obtain Hysplit backtrajectories. Latitudinal section at 16.351ºS for all domains. Note that domain 4 represents better the real altitude of Chacaltaya and the complex terrain around. Prepared by F. Velarde

Table S 8. WRF nested domains for generating meteorological data to be used in Hysplit Desktop v. 4.8. In all domains 28 pressure levels were used. Prepared by F. Velarde

Domain	Spatial resolution [km]	Height of Chacaltaya [masl]
WRFd01	38.00x38.00	3728.14
WRFd02	9.50x9.50	4535.99
WRFd03	3.17x3.17	4732.34
WRFd04	1.06x1.06	5058.44

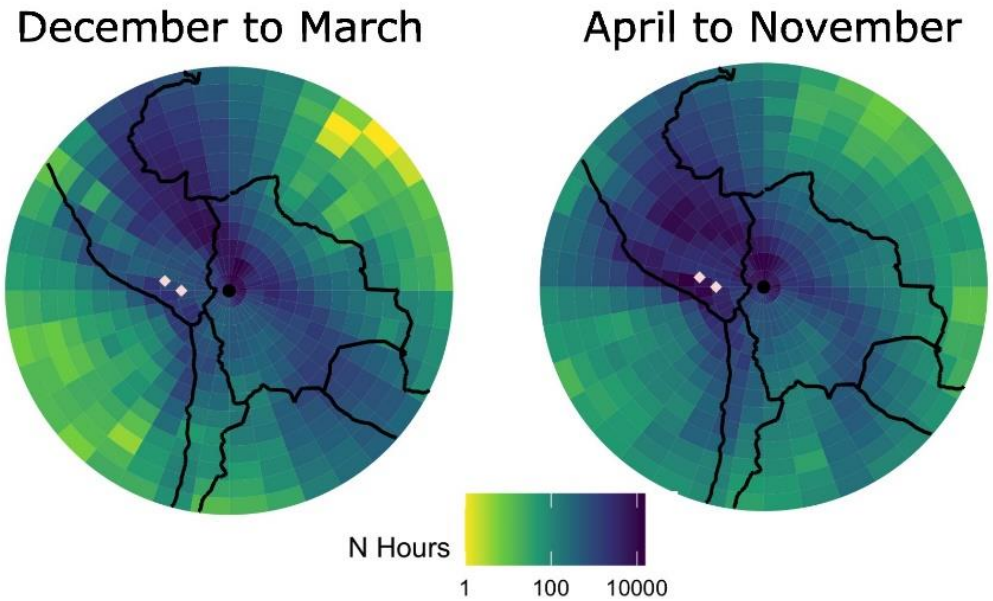

Figure S 13. Seasonality. WRF-based backtrajectories for all the samples taken for the wet season (December to March) and the rest of the year. The pink rhombuses show Sabancaya and Ubinas locations.

Figure S 14. Comparison of nested WRF (only central point) and ERA-5 backtrajectories (the only one point available) based on samples that present North (N) and/or Northwest (NW) pathways. The classification was made for each dataset (as there are not necessarily coincident) and grouped for the entire year. The number of 96h-bactrajectories used for each plot is in italics. The color scale corresponds to the number of hours of accumulated air passages over a pixel.

--- Clustering to identify seasonality patterns ---

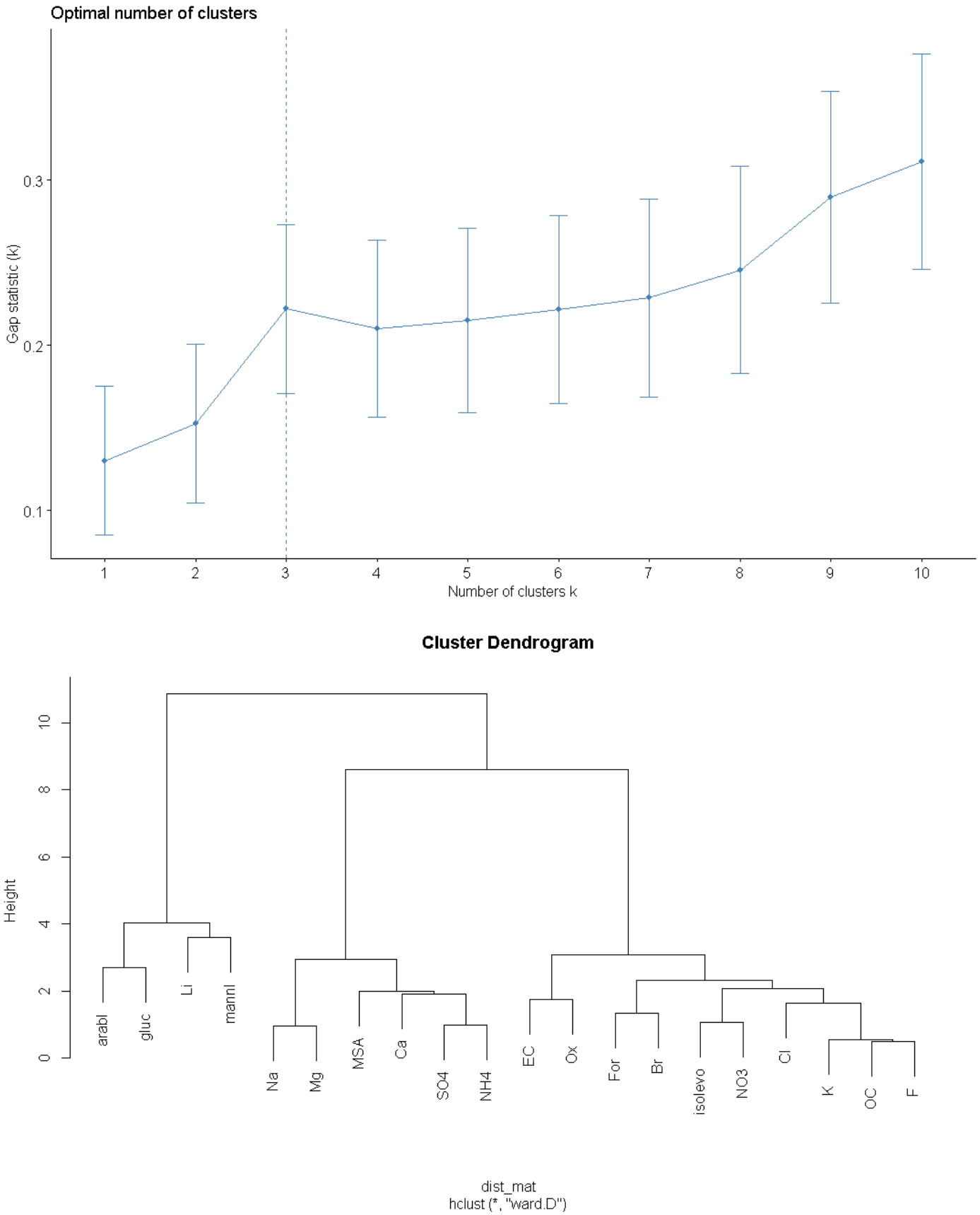

Figure S 15. Clustering by k-means of median concentration for ions, EC, OC, saccharides. Upper panel: k-means method optimal number of clustering. Lower panel: hierarchical clustering based on the Euclidean distance of concentrations using Ward method, prepared with R.

--- Control case for volcanism: methanesulfonate transported under W, SW ---

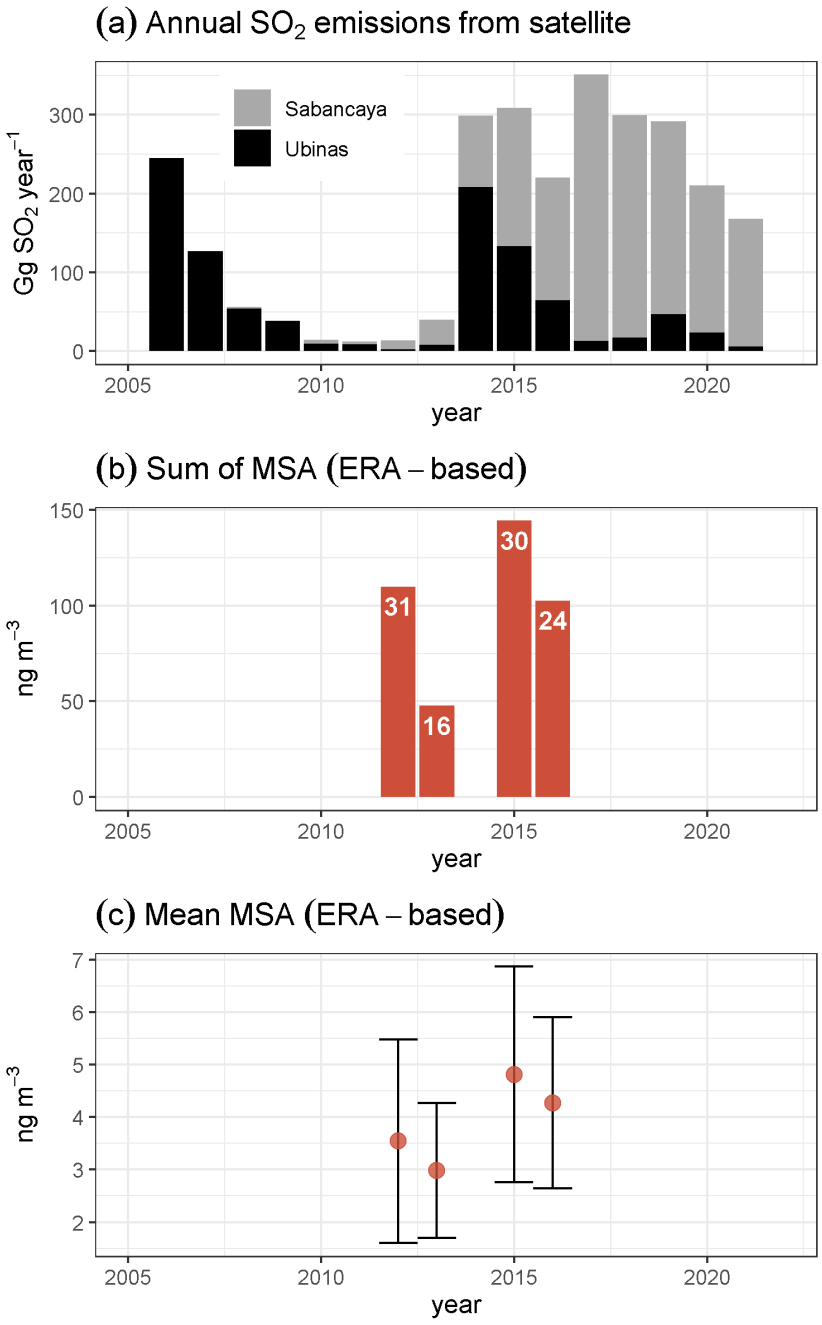

Figure S 16. Control case for marine transport using ERA-5 backtrajectories in complement to volcanic transport. Only years with at least 9 months of filter data were used. (a) Bulk SO_2 emissions for Sabancaya (grey) and Ubinas (black) from https://so2.gsfc.nasa.gov/measures.html (b) Accumulated SO_4^{-2} measured in the filters taken under W and/or SW influence grouped per year based on ERA-5 backtrajectories. The number of samples used is in white letters inside the red (c) Mean concentration and standard deviation for the selected samples. Note that there is no statistically significant difference between 2015-2016 and 2012.

---- References ----

Belis C.A., Favez O., Mircea M., Diapouli E., Manousakas M-I., Vratolis S., Gilardoni S., Paglione M., Decesari S., Mocnik G., Mooibroek D., Salvador P., Takahama S., Vecchi R., Paatero P., European guide on air pollution source apportionment with receptor models - Revised version 2019, EUR 29816 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-09001-4, doi:10.2760/439106, JRC117306

Marynowski, L. and Simoneit, B. R. T.: Saccharides in atmospheric particulate and sedimentary organic matter: Status overview and future perspectives, Chemosphere, 288, 132376, https://doi.org/10.1016/j.chemosphere.2021.132376, 2022.

Xu, J., Jia, C., He, J., Xu, H., Tang, Y.-T., Ji, D., Yu, H., Xiao, H., and Wang, C.: Biomass burning and fungal spores as sources of fine aerosols in Yangtze River Delta, China – Using multiple organic tracers to understand variability, correlations and origins, Environmental Pollution, 251, 155–165, https://doi.org/10.1016/j.envpol.2019.04.090, 2019.