Supplementary information to:
Polyols and glucose as tracers of primary biogenic organic aerosol: influence of environmental factors on ambient air concentrations and spatial distribution over France

Abdoulaye Samaké[1], Jean-Luc Jaffrezo[1], Olivier Favez[2], Samuël Weber[1], Véronique Jacob[1], Trishalee Canete[1], Alexandre Albinet[2], Aurélie Charron[1,16], Véronique Riffault[3], Esperanza Perdrix[3], Antoine Waked[1], Benjamin Golly[1], Dalia Salameh[1*], Florie Chevrier[1,4], Diogo Miguel Oliveira[2,3], Jean-Luc Besombes[4], Jean M.F. Martins[1], Nicolas Bonnaire[5], Sébastien Conil[6], Géraldine Guillaud[7], Boualem Mesbah[8], Benoit Rocq[9], Pierre-Yves Robic[10], Agnès Hulin[11], Sébastien Le Meur[12], Maxence Descheemaecker[13], Eve Chretien[14], Nicolas Marchand[15], and Gaëlle Uzu[1].

[1]University Grenoble Alpes, CNRS, IRD, INP-G, IGE (UMR 5001), 38000 Grenoble, France.

[2]INERIS, Parc Technologique Alata, BP 2, F-60550 Verneuil-en-Halatte, France

[3]IMT Lille Douai, University Lille, SAGE – Département Sciences de l'Atmosphère et Génie de l'Environnement, F-59000 Lille, France

[4]University Savoie Mont-Blanc, LCME, 73000 Chambéry, France

[5]LSCE, UMR CNRS-CEA-UVSQ, 91191 Gif-sur Yvette, France

[6]ANDRA DRD/GES Observatoire Pérenne de l'Environnement, F-55290 Bure, France

[7]Atmo Auvergne-Rhône-Alpes, 38400 Grenoble, France

[8]Air PACA, 03040, France

[9]Atmo Hauts de France, 59000, France

[10]Atmo Occitanie, 31330 Toulouse, France

[11]Atmo Nouvelle Aquitaine, 33000, France

[12]Atmo Normandie, 76000, France

[13]Lig'Air, 45590 Saint-Cyr-en-Val, France

[14]Atmo Grand Est, 16034 Strasbourg, France

[15]University Aix Marseille, LCE (UMR7376), Marseille, France

[16]IFSTTAR, F-69675 Bron, France

[*] Now at: Airport pollution control authority (ACNUSA), 75007 Paris, France

Corresponding author(s): A Samaké (abdoulaye.samake2@univ-grenoble-alpes.fr) and JL Jaffrezo (Jean-luc.Jaffrezo@univ-grenoble-alpes.fr)

Section 1: supplementary illustrations

Table S1: Characteristics of selected sites and data available

Sampling sites	Typology	Campaign periods	Long	Lat	Alt (m)	Data available
Grenoble_LF	Urban[a]	02/2012-03/2018	5.74	45.16	214	Polyols, glucose, cellulose, LAI
Grenoble_CB	Urban[a]	02/2017-03/2018	5.73	45.18	212	Polyols, glucose, LAI
Grenoble_VIF	Urban[a]	02/2017-03/2018	5.68	45.06	310	Polyols, glucose, cellulose, LAI
Passy	Urban[a]	11/2013-04/2015	6.71	45.92	588	Polyols, glucose, LAI, weather conditions
Marnaz	Urban[a]	07/2013-04/2015	6.53	46.06	504	Polyols, glucose, LAI, weather conditions
Chamonix	Urban[a]	11/2013-10/2014	7.05	45.92	1035	Polyols, glucose, LAI, weather conditions
Marseille	Urban	06/2014-12/2017	5.39	43.30	64	Polyols, glucose
Mallet	Urban	06/2014-06/2015	5.50	43.47	200	Polyols, glucose
Gardanne	Urban	07/2015-07/2016	5.47	43.45	214	Polyols, glucose
Meyreuil	Urban	01/2015-01/2016	5.50	43.47	235	Polyols, glucose
Port-de-Bouc	Urban	06/2014-12/2017	4.98	43.40	1	Polyols, glucose
Nice	Urban	06/2014-12/2016	7.28	43.70	9	Polyols, glucose
Rouen	Urban	01/2013-06/2014	1.08	49,44	6	Polyols, glucose
Roubaix	Traffic	01/2013-05/2014	3.18	50,71	10	Polyols, glucose
Nogent-sur-Oise	Sub-urban[b]	01/2013-12/2017	2.48	49.28	30	Polyols
OPE-ANDRA	Rural[b]	02/2012-12/2017	5.17	48.54	293	Polyols, glucose, LAI, weather conditions, agricultural activities records, cellulose

Symbols [(a)] stand for urban background sites located in the French Alp valley environment whereas symbols [(b)] outline background sites surrounded by crop field areas. Leaf Area Index (LAI) is a proxy of vegetation density evolution. Polyols are defined as the sum of mannitol and arabitol concentrations.

Table S2: Annual average values ± standard deviation of aerosol chemical data at each site (concentrations in ng m^{-3}).

Sampling sites	Polyols	Ratio mannitol-to-arabitol	Glucose	Ratio glucose-to-polyols
Grenoble_LF	41.2 ± 39.9	1.24 ± 0.36	26.8 ± 19.7	0.93 ± 0.63
Grenoble_CB	43.5 ± 41.9	1.07 ± 0.32	29.0 ± 22.6	0.94 ± 0.57
Grenoble_VIF	47.0 ± 48.8	1.11 ± 0.41	30.5 ± 26.2	0.92 ± 0.56
Passy	37.0 ± 23.2	0.94 ± 0.34	23.1 ± 13.3	0.70 ± 0.31
Marnaz	54.5 ± 42.6	1.03 ± 0.39	33.2 ± 23.3	0.72 ± 0.34
Chamonix	38.0 ± 28.0	1.08 ± 0.31	20.0 ± 11.9	0.73 ± 0.54
Marseille	26.1 ± 22.9	1.13 ± 0.34	21.2 ± 15.8	0.91 ± 0.45
Mallet	42.5 ± 31.5	0.99 ± 0.36	27.9 ± 17.4	0.66 ± 0.25
Gardanne	27.8 ± 20.8	0.98 ± 0.26	17.5 ± 10.6	0.72 ± 0.32
Meyreuil	27.8 ± 15.4	0.94 ± 0.32	17.6 ± 10.1	0.67 ± 0.28
Port-de-Bouc	21.1 ± 17.7	1.03 ± 0.37	14.4 ± 13.5	0.78 ± 0.49
Nice	37.6 ± 36.5	1.14 ± 0.40	24.5 ± 23.4	0.69 ± 0.35
Rouen	23.8 ± 34.2	1.27 ± 0.71	8.6 ± 11.1	0.52 ± 0.50
Roubaix	18.8 ± 22.2	1.67 ± 0.89	8.6 ± 8.6	0.72 ± 0.83
Nogent-sur-Oise	43.8 ± 42.9	1.53 ± 0.54	N/A	N/A
OPE-ANDRA	58.7 ± 90.4	1.01 ± 0.47	31.2 ± 32.6	0.86 ± 0.69

N/A: not available.

Table S3: Normalized cross-correlation coefficients (R) for sugar compounds and ratios between pairs of sites considering the sampling periods in common.

Pairs of sampling sites	Inter-site distance (Km)	Number of samples	Time periods	Arabitol vs Mannitol	Ratios Mannitol-to-Arabitol	Glucose vs Polyols	Ratios Glucose to Polyols
Grenoble_CB *vs* Grenoble_LF	2.5	125	02/2017-03/2018	0.967	0.988	0.973	0.968
Grenoble_LF *vs* Grenoble_VIF	12.4	125	02/2017-03/2018	0.962	0.963	0.957	0.985
Grenoble_CB *vs* Grenoble_VIF	14.4	127	02/2017-03/2018	0.930	0.982	0.930	0.968
Passy *vs* Chamonix	12.1	112	11/2013-10/2014	0.946	0.938	0.934	0.885
Marnaz *vs* Passy	20.4	159	11/2013-04/2015	0.951	0.918	0.938	0.935
Marnaz *vs* Chamonix	30.0	112	11/2013-10/2014	0.927	0.926	0.890	0.888
Marseille *vs* Gardanne	17.5	79	07/2015-07/2016	0.891	0.907	0.877	0.876
Marseille *vs* Mallet	20.1	79	06/2014-06/2015	0.829	0.717	0.862	0.843
Marseille *vs* Meyreuil	20.7	76	01/2015-01/2016	0.665	0.804	0.786	0.779
Marseille *vs* Port-de-Bouc	35.1	277	06/2014-12/2017	0.788	0.764	0.723	0.790
Grenoble_LF *vs* Marnaz	117.4	201	07/2013-04/2015	0.765	0.797	0.825	0.899
Nice vs Port-de-Bouc	188.7	218	06/2014-12/2017	0.723	0.678	0.713	0.530
Roubaix *vs* Rouen	205.4	154	01/2013-05/2014	0.635	0.659	0.474	0.440
Nogent-sur-Oise *vs* OPE-ANDRA	230.1	253	01/2013-12/2017	0.754	0.790	N/A	N/A
Grenoble_LF *vs* Marseille	270.1	240	06/2014-12/2017	0.438	0.309	0.344	0.419
Grenoble_LF *vs* OPE-ANDRA	374.9	297	02/2012-12/2017	0.390	0.234	0.275	0.394
Marseille *vs* OPE-ANDRA	581.0	194	06/2014-12/2017	0.307	0.180	0.124	0.236

N/A: not available.

The quality of the multiple linear regression model (linearity of the data, normality of residuals, homogeneity of residuals variance, independence of residuals error terms) was checked through several diagnostic plots:

- Figure S1A shows the residuals *vs* fitted values, which did not exhibit any significant pattern. Therefore a linearity relationship between log (polyols ± glucose) and the predictor variables can be assumed.
- In Figure S1B, the model residuals are correctly fitted with a straight line, indicative of a normal distribution.
- As evidenced in Figure S1C (Scale-Location plot), squared residuals are quite randomly distributed along the range of predictor variables. Thus the variance of the residuals is considered homogeneous.
- Finally, Figure S1D was used to examine the potential influential points (outliers or high-leverage points). Cook's distance (highlighted by the red dashed lines) measures the effect of deleting an extreme observation. Since numbered points are within Cook's distance scores (standardized residuals are also below 3), they are not considered as influencing the regression analysis.

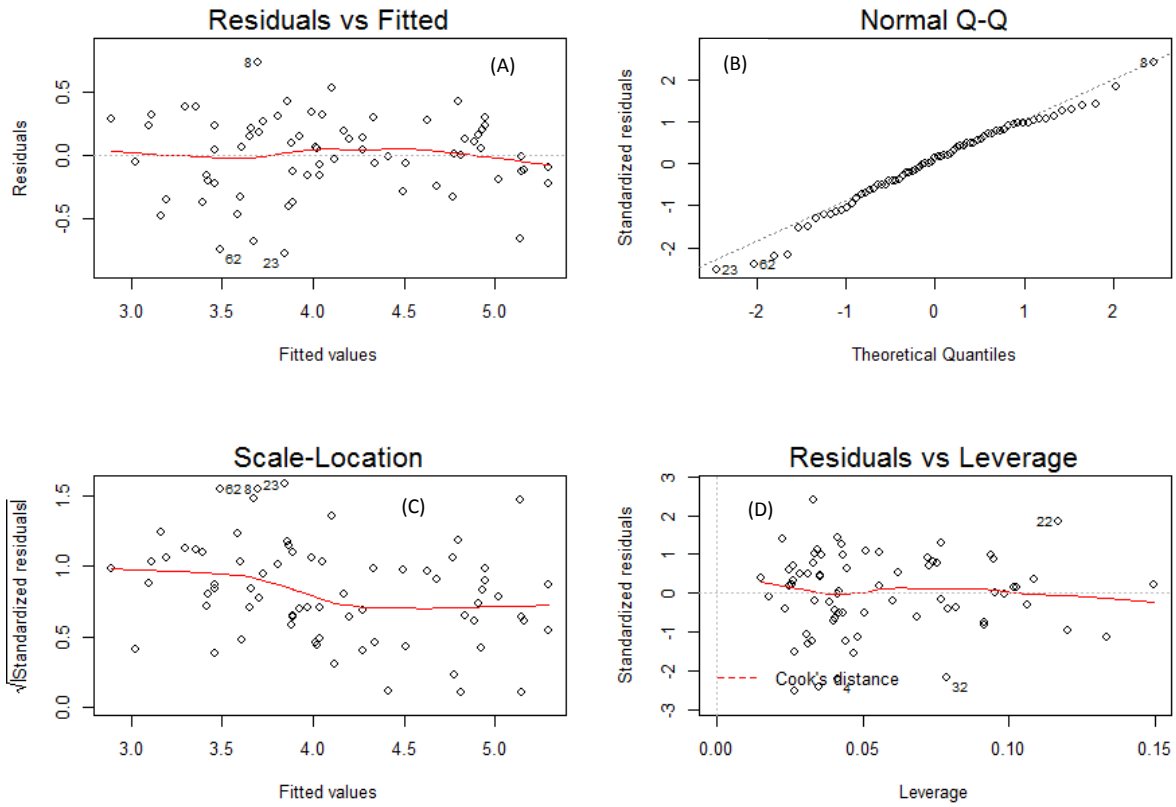

Figure S1: Diagnostic plots for the multiple linear regression analysis. The red solid lines are smoothed curves for detecting potential patterns.

Multicollinearity between the predictor variables was evaluated using variance inflation factors (VIF). These were performed using the *vif* function implemented in the open-access *"car package in R"* (Fox and Weisberg, 2018). Collinearity was not found to be a problem in our multiple regression analysis because all VIF values were less than ten for all predictor variables (Zuur et al., 2010).

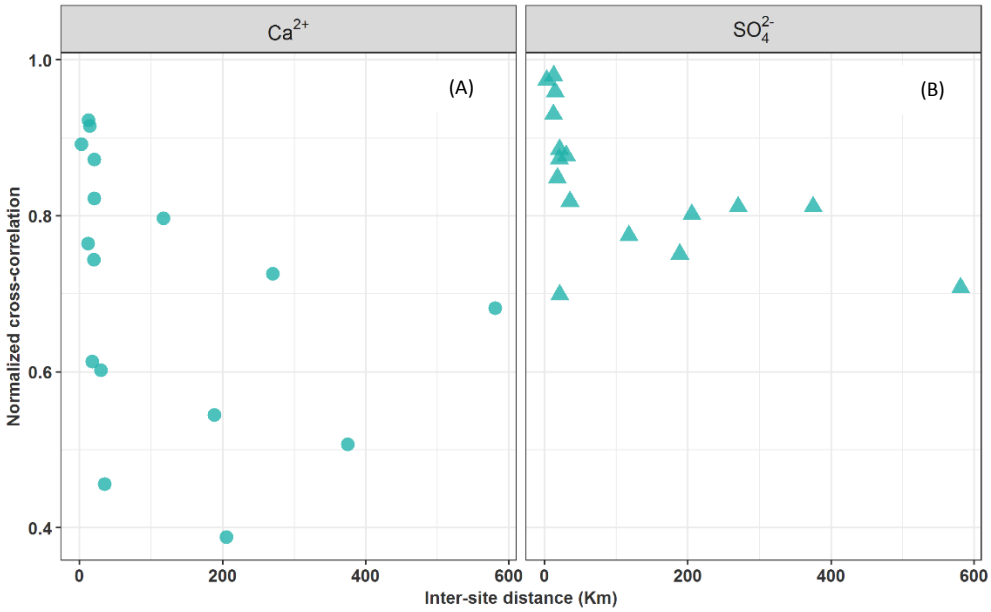

Figure S2: Normalized cross-correlation values for the daily evolution of particulate calcium (A) and sulfate (B) between pairs of sites located at increasing spatial scales across France.

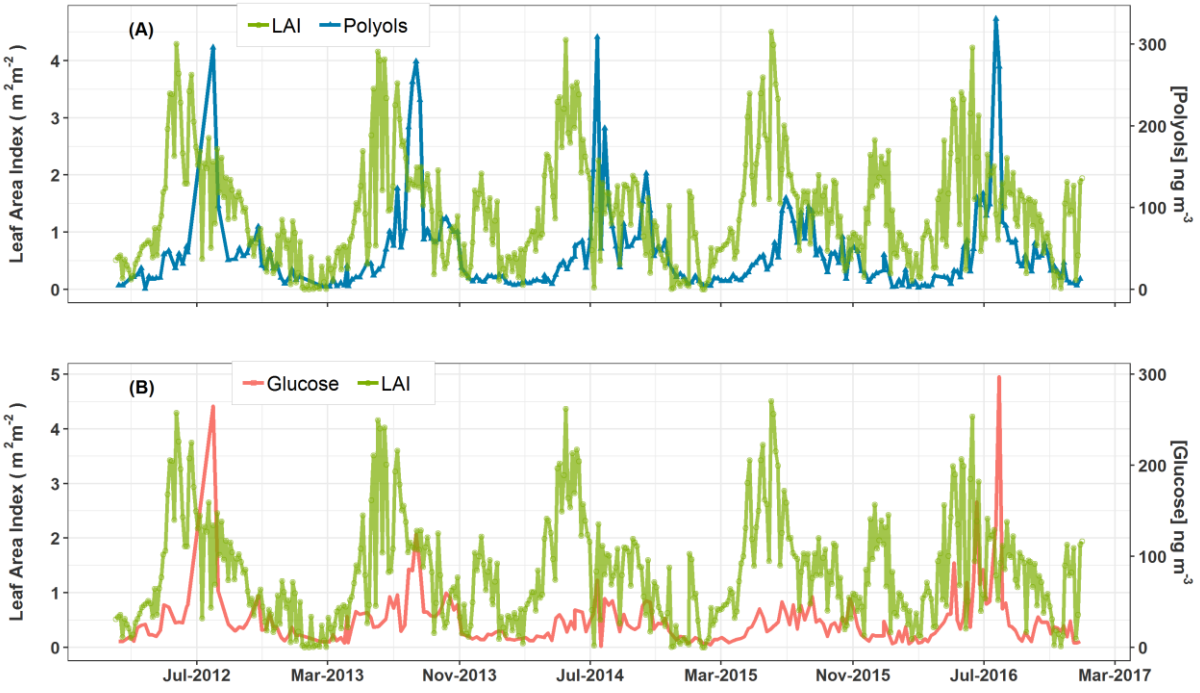

Figure S3: Covariation cycles of the daily concentrations of polyols (A) and glucose (B) and vegetation density (LAI) at OPE-ANDRA, from 2012 to 2016.

Section 2: Analysis of ambient particulate cellulose concentration

The analytical protocol is resumed below. A punch of 21 mm from a PM_{10} quartz filter sample is sonicated for 40 min in 3 mL of a 0.05 M acetate buffer (pH=4.8) in Milli-Q water containing 0.05% of thymol. 20 µL of the purified and diluted cellulase (70 u/g) and 60 µL of the diluted glucosidase (5 u/g) are added to the solution which is then incubated for 24 hours at 45 °C for the hydrolysis to occur. Enzymatic activity is stopped at the end of this step by heating the sample at 100°C for 45 minutes. After cooling down to room temperature, the sample is centrifuged for 10 minutes at 7,000 RPM and filtered through a 0.22 µm polyether sulfone membrane for the analysis of the glucose content using HPLC-PAD.

Since cellulase is the main contributor to the level of glucose in the blanks, this enzyme initial solution is purified by ultra-filtration at 15°C on a porous membrane of polyether sulfone (Hydrosart®, 2000 MWCO). Ten consecutive steps of ultra-filtration at 7,000 RPM in Vivaspin-15R tubes are needed to reduce the content of glucose in the blanks to an acceptable level (< 10 µg L^{-1} in analytical solutions). At the end of filtration, cellulase is diluted 10 times in Milli-Q water.

For each analytical batch, standard aqueous solutions of cellulose (microbeads of pure cellulose 20µm, Sigma Aldrich) are hydrolyzed in parallel under the same conditions in order to determine the conversion yield of cellulose. Although variable depending on the batch, it is generally in the range 65 – 80 %. Each analytical batch is then composed of glucose standards, hydrolyzed cellulose standards, hydrolyzed samples and hydrolyzed blanks filters. The final calculation of the atmospheric concentration of the free cellulose takes into account the conversion efficiency of cellulose, the cellulose on blank filters, and the initial concentrations of atmospheric glucose of each sample, determined in parallel using a similar HPLC-PAD technique (Waked et al., 2014).

The HPLC-PAD is composed of an AS50 autosampler, a LC30 oven, a GP40 pump and an ED50 detector (all from Dionex) working in the pulsed amperometric detection mode with a gold working electrode and an Ag/AgCl reference electrode. The analyses are performed on Dionex CarboPac PA1 columns (4 × 250 mm – analytical; 4 × 50 mm – guard), under gradient elution conditions (Table S4) at 30 °C. The mobile phase is made of sodium hydroxide and sodium acetate in Milli-Q water, at a flow rate of 1.1 mL min^{-1}. The waveform program applied to the detector is illustrated in Figure S5.

Table S4: Gradient operating conditions used for HPLC-PAD.

Time (min)	*Flow rate (mL min^{-1})*	*NaOH: 18 mM*	*NaOH: 200 mM*	*NaOH: 100 mM* *NaAc: 150 mM*
0	1.1	100 %	-	-
10	1.1	100 %	-	-
16	1.1	70 %	30 %	-
18	1.1	6 %	82 %	12 %
21	1.1	-	-	100 %
23.5	1.1	-	-	100 %
24	1.1	-	100 %	-
27	1.1	-	100 %	-
28	1.5	100 %	-	-
39.5	1.1	100 %	-	-

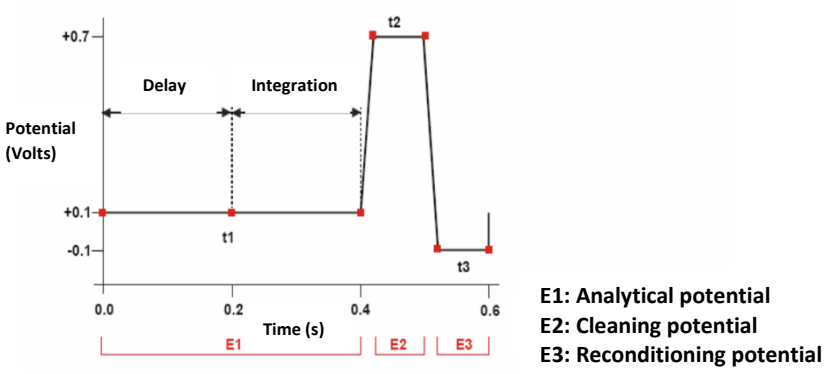

Figure S4 : Example of the applied waveform program.

References

Fox, J. and Weisberg, S.: Regression diagnostics for linear, generalized linear, and mixed-effects models, in An R Companion to Applied Regression, pp. 385–477, SAGE Publications. Available from: https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html, 2018.

Waked, A., Favez, O., Alleman, L. Y., Piot, C., Petit, J.-E., Delaunay, T., Verlinden, E., Golly, B., Besombes, J.-L., Jaffrezo, J.-L., and Leoz-Garziandia, E.: Source apportionment of PM_{10} in a north-western Europe regional urban background site (Lens, France) using positive matrix factorization and including primary biogenic emissions, Atmos. Chem. Phys., 14(7), 3325–3346, doi:10.5194/acp-14-3325-2014, 2014.

Zuur, A. F., Ieno, E. N., and Elphick, C. S.: A protocol for data exploration to avoid common statistical problems: data exploration, Methods Ecol. Evol., 1(1), 3–14, doi:10.1111/j.2041-210X.2009.00001.x, 2010.