Supplement of Atmos. Chem. Phys., 20, 11747–11766, 2020
https://doi.org/10.5194/acp-20-11747-2020-supplement
© Author(s) 2020. This work is distributed under
the Creative Commons Attribution 4.0 License.

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Sources and sinks driving sulfuric acid concentrations in contrasting environments: implications on proxy calculations

Lubna Dada et al.

Correspondence to: Lubna Dada (lubna.dada@helsinki.fi) and Markku Kulmala (markku.kulmala@helsinki.fi)

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

1. Reaction rate constant from Mikkonen et al. (2011)

Derivation of the temperature dependent reaction rate constant (k) used in calculating the Mikkonen proxy from our data sets:

$$k\ (cm^3\ molec^{-1}\ s^{-1}\ ) = \frac{A\ .k_3}{(A+k_3)}\ x \exp\left\{ k_5 \left[1 + \log_{10}\left(\frac{A}{k_3}\right)^2\right]^{-1} \right\} \quad (S1)$$

$$A = k_1 . [M] . \left(\frac{300}{k}\right)^{k_2} \quad (S2)$$

$$[M] = 0.101\ .\ (1.381\ x\ 10^{-23}\ T)^{-1} \quad (S3)$$

M is the density of the air in molec cm^{-3}, $k_1 = 4\times10^{-31}$, $k_2 = 3.3$, $k_3 = 2\times10^{-12}$ and $k_5 = -0.8$.
k given in Eq.(S1) is scaled by multiplying it with 10^{12} as described in more detail in Mikkonen et al. (2011).

2. Bootstrap resampling and sensitivity analyses

When deriving the proxy equation for each site, 10 000 bootstrap resamples were drawn for each data set independently. Bootstrap resampling without disturbance generates extended data from the original data by randomly replacing an existing data point with another one from the same data set, resulting in different combinations of variables from the original data set. We accounted for the systematic uncertainty in H_2SO_4 and predictor variables arising e.g. from calibration uncertainties. For every bootstrap fit, we assumed both H_2SO_4 and all predictor variables to be affected by independent systematic errors between the upper and lower bound of their independent uncertainty ranges. Since the uncertainty related to the measurement accuracy was much larger than the precision of the measurement, we only accounted for the uncertainty arising from accuracy. In practice, we scaled the entire time series of each variable by a random set of numbers drawn from a uniform distribution of possible measurement biases.
Accordingly, a factor of 2 uncertainty was introduced in the sulfuric acid concentration, a 20% uncertainty in the condensation sink measurement, and a 10% in each trace gas concentration and global radiation. In the case of sulfuric acid concentrations, which have a factor of 2 uncertainty, the actual concentration of sulfuric acid at a certain point in time could be anywhere between a factor of 2 lower and a factor of 2 higher. Therefore, for each sulfuric acid measurement, we generated 10 000 concentrations by multiplying the original measured concentration by a uniform random array between the lower and upper bounds, which are 0.5 and 2 in the case of sulfuric acid. The same resampling method was applied for each other predictor variable independently, and the 10 000 possible combinations of the disturbed data sets were used to generate the fit and to derive the sulfuric acid proxy equation per site. A median of these 10 000 k value combinations which account for the error on the predictor variables was then used to form one equation per location after removing outliers outside the 1.5 interquartile ranges above the upper quartile (75th percent) or below the lower quartile (25th percent). The MATLAB code used to generate the boot resamples is shown in Code 1.

Code 1. MATLAB code used to generate the boot resamples and obtain the fitting coefficients *(k_1, k_2 and k_3)* using Eq. (3).

```
%% Derive k values for sulfuric acid proxy concentration using Dada et al. 2020 equation
% fitCoeff(1) = k1
% fitCoeff(2) = k2
% fitCoeff(3) = k3

data = [CS, SO2, O3, Alkene, GlobRad]; %CS in s-1, SO2,O3,Alkene in cm-3, GlobRad in W/m2
H2SO4; %measured sulfuric acid in cm-3

% Create the fitting function according to Equation 3

Y_fit = @(fitCoeff,data) (-1).*(data(:,1)./(2*fitCoeff(3))) + ...
    sqrt((data(:,1)./(2*fitCoeff(3))).^2 + data(:,2)./(fitCoeff(3)).*...
    (fitCoeff(1).*data(:,5) + fitCoeff(2).*data(:,3).*data(:,4)));


% Obtain the fitting coefficients were obtained by minimizing the sum of the squared logarithm
%of the ratio between the proxy values and measured sulphuric acid concentration

sum_squared_error = @(fit_coeff) sum((log10(H2SO4 ./ (Y_fit(fit_coeff,data)))).^2);

%introduce bootstrap resampling

fit_index = 10000; %number of bootstrap resampling

[~,bootsam] = bootstrp(fit_index,sum_squared_error,data); %bootstrap resampling

%introduce uncertainty estimates on the measured predictor varianbles
%create an array of random floating-point numbers that are drawn from a
%uniform distribution in the open interval between the lower and upper bound of accuracy

% 20% uncertainty on condensation sink
a = log10(1/1.2); %lower bound accuracy
b = log10(1.2); %upper bound accuracy
r_CS = 10.^((b-a).*rand(fit_index,1)+a);

% factor of 2 uncertainty on H2SO4 measurement
a = log10(0.5);
b = log10(2);
r_SA = 10.^((b-a).*rand(fit_index,1) + a);

% 10% uncertainty on trace gases and global radiation
a = log10(1/1.1);
b =log10(1.1);
r_SO2 = 10.^((b-a).*rand(fit_index,1) + a); %SO2

a = log10(1/1.1);
b =log10(1.1);
r_O3 = 10.^((b-a).*rand(fit_index,1) + a); %O3

a = log10(1/1.1);
b =log10(1.1);
r_MT = 10.^((b-a).*rand(fit_index,1) + a); %Alkenes

a = log10(1/1.1);
b =log10(1.1);
r_GR = 10.^((b-a).*rand(fit_index,1) + a); %GlobRadiation

%
k_all=[];
for i =1:fit_index

    %create bootstrapped data disturbed with uncertainty on predictor variables
data_boot = [data(bootsam(:,i),1)*r_CS(i),data(bootsam(:,i),2)*r_SO2(i),...
    data(bootsam(:,i),3)*r_O3(i), data(bootsam(:,i),4)*r_MT(i),...
    data(bootsam(:,i),5)*r_GR(i)];
H2SO4_boot=H2SO4(bootsam(:,i),:)*r_SA(i);

% Obtain the fitting coefficients for the bootstrap resamples

sum_squared_error = @(fit_coeff) sum((log10(H2SO4_boot ./ (Y_fit(fit_coeff,data_boot)))).^2);

% Assume initial values for the fitting parameters:
k0 = [1e-8, 1e-27,1e-9];

% Use built-in MATLAB function fminsearch to find the fitting parameters;
% the best fit parameters are in output into variable k:
[k, SSE] = fminsearch (sum_squared_error, k0, options);

k_all = [k_all;k(:,:)];
end
```

Table S 1. Summary of measurement locations and the instrumentation used for deriving the H_2SO_4 proxy (training data sets).

Location	Type	Measurement Period	Particle size distribution instrument	Trace Gases	Radiation
Hyytiälä, Finland	Boreal	August 18, 2016 to December 31, 2016 and March 8, 2018 to February 28, 2019	Twin – DMPS (Ground level)	SO_2 and O_3 are monitored using two Thermo Environmental Instruments (models 43i-TLE, 49i, respectively), at 16.8 m above ground level	[1]Global radiation was measured with Middleton solar SK08 pyranometer until August 24, 2017 and after that with Middleton solar EQ08-S pyranometer at 16.8 m.
Agia Marina, Cyprus[2]	Rural background	February 22 and March 3, 2018	2-20 nm using Airel NAIS and 20-800 nm using TSI SMPS	SO_2 is monitored using Ecotech instrument (9850).	Global radiation was measured by a Campbell Scientific weather station .
Budapest, Hungary[2]	Urban	March 21 and April 17, 2018	6-1000 nm using flow-switching type DMPS	SO_2 is measured using UV fluorescence (Ysselbach 43C).	Global radiation was measured by an SMP3 pyranometer (Kipp and Zonnen, The Netherlands).
Beijing, China	Megacity	March 15, 2019 – June 15, 2019	3 – 800 nm PSD system ~12 m above ground level.	SO_2 and O_3 are monitored using two Thermo Environmental Instruments (models 43i-TLE, 49i, respectively), ~12 m above ground.	[3]Global radiation was measured using CMP11 pyranometer (Kipp and Zonnen, Delft, Netherlands) at ~ 15 m above ground level.

[1] UVB radiation was measured with Solar SL 501A pyranometer.
[2] All variables are measured at the same height.
[3] UVB radiation was measured using a UVS-B-T radiometer (Kipp and Zonnen, Delft, Netherlands).

Table S 2. Summary of measurement locations and instrumentation used for verifying the predictive power of the derived proxies (testing data sets).

Location	Type	Measurement Period	Particle size distribution instrument	Trace Gases	Radiation
Hyytiälä, Finland	Boreal	January 1, 2017 – June 5, 2017	Twin – DMPS (Ground level).	SO_2 and O_3 are monitored using two Thermo Environmental Instruments (models 43i-TLE, 49i, respectively).	Global radiation was measured with Middleton solar EQ08-S pyranometer.
Helsinki, Finland	Semi-urban	July 1, 2019 – July 16, 2019	Twin DMPS at ground level	SO_2 was measured using UV-flurescence (Horiba APSA 360) at 31 m above ground.	Global radiation was monitored Kipp and Zonen CNR1 at 31 m above ground level.
Beijing, China	Megacity	September 8, 2019 – October 15, 2019	3 – 800 nm PSD system ~12 m above ground	SO_2 and O_3 are monitored using two Thermo Environmental Instruments (models 43i-TLE, 49i, respectively) ~ 12 m above ground.	Global radiation was measured using CMP11 pyranometer (Kipp and Zonnen, Delft, Netherlands) at ~ 15 m above ground level.
Kilpilahti, Finland	Industrial Area	June 07, 2012 – June 29, 2012	6 to 1000 nm DMPS.	SO_2 was monitored usingThermo Scientific ™ Model 43i SO_2 Analyser.	Acquired from SMEAR III station[4].

[4] Same as Helsinki site.

Table S 3. Summary of basic statistics of measurements of condensation sink, trace gases and global radiation at all locations and time periods included in this study. For Hyytiälä, Beijing and Kilpilahti we use the whole day time window (GlobRad > 0 W/m^2), for Agia Marina, Budapest and Helsinki we use daytime statistics (GlobRad > 50 W/m^2).

Location		**Hyytiälä, Finland**	**Hyytiälä, Finland**	**Agia Marina, Cyprus**	**Helsinki, Finland**	**Budapest Hungary**	**Beijing, China**	**Beijing, China**	**Kilpilahti, Finland**
Type		Boreal	Boreal	Rural	Semi-urban	Urban	MegaCity	MegaCity	Industrial Area
Measurement Period		August 18 - December 31, 2016 March 8 - February 28, 2019	January 1, 2017 – June 5, 2017	February 22 - March 3, 2018	July 1 – July 16, 2019	March 21 - April 17, 2018	March 15, 2019 – June 15, 2019	September 8, 2019 – October 15, 2019	June 07, 2012 – June 29, 2012
CS (10^{-3} s^{-1})	mean	4.48	2.88	4.43	3.38	11.74	24.20	23.22	5.25
	median	3.83	2.18	3.63	3.13	10.92	22.83	22.60	4.91
	5th percentile	0.85	0.74	1.37	1.25	5.03	7.60	5.14	2.61
	95th percentile	12.43	8.78	9.58	6.47	21.52	44.58	44.34	8.81
	sd	3.89	2.42	2.55	1.60	5.37	11.86	11.82	2.11
SO_2 (10^{10} cm^{-3})	mean	0.31	0.30	0.70	1.30	6.02	4.70	2.43	6.65
	median	0.12	0.16	0.46	0.87	5.45	3.49	1.35	2.98
	5th percentile	0.03	0.01	0.17	0.13	3.35	0.26	0.13	0.99
	95th percentile	1.24	1.01	1.96	2.19	12.42	13.71	8.47	26.00
	sd	0.54	0.47	0.65	3.19	2.54	4.59	3.56	11.46
O_3 (10^{10} cm^{-3})	mean	83.59	95.08				105.63	116.10	161.36
	median	80.27	97.10				95.66	102.53	178.15
	5th percentile	41.09	65.42				5.23	3.24	24.81
	95th percentile	134.85	118.42				238.26	260.97	234.37
	sd	28.52	16.80				72.22	80.99	62.92
Alkene (10^{10} cm^{-3})	mean	0.92	0.32				14.33	11.98	2.27
	median	0.39	0.15				12.29	11.91	0.72
	5th percentile	0.05	0.02				1.91	2.55	0.11
	95th percentile	3.54	0.85				34.40	19.51	10.20
	sd	2.03	0.98				9.68	4.96	3.38

Global Radiation ($W.m^{-2}$)	mean	149.25	93.06	283.71	353.67	322.90	243.72	221.27	307.86
	median	47.53	23.17	272.48	270.60	300.56	54.27	52.97	252.64
	5^{th} percentile	0.47	0.36	67.92	61.59	70.64	0.02	0.02	0.06
	95^{th} percentile	636.60	378.50	548.90	837.27	697.42	840.95	730.83	768.84
	sd	205.18	137.32	155.33	254.08	200.36	308.33	273.10	280.05
H_2SO_4 (10^6 cm^{-3})	mean	0.73	0.55	2.76	3.82	1.54	2.94	3.45	10.59
	median	0.28	0.18	1.81	2.55	1.02	1.61	2.00	3.19
	5^{th} percentile	0.02	0.02	0.17	0.41	0.23	0.37	0.37	0.19
	95^{th} percentile	2.55	2.01	8.22	11.71	4.76	8.63	10.98	37.08
	sd	1.40	1.06	3.06	4.57	1.77	3.00	3.74	28.25

Table S 4. Statistical parameters included in deriving the Aikake Information Criterion. Equation number refers to the number in the main text, N is the sample size (number of points), X is the number of coefficients (number of *k* values) and SSE is the sum of squared estimate of errors. AIC is calculated as AIC = 2X + N ln(SSE). The quantity $\exp((AIC_{min} - AIC_i)/2)$ describes the probability that the ith model minimizes the information loss. For example, Eq.(5) in Hyytiälä is 5.62E-8 times as probable as the Eq. (6) to minimize the information loss.

Hyytiälä Eq. (9)	Equation number	6	5	4	2
	number of coefficients	3	2	2	1
	N	1860	1860	1860	1860
	R	0.84	0.74	0.82	0.70
	Slope	0.80	0.78	0.96	1.84
	SSE	1.89E+02	3.00E+02	2.88E+02	1.17E+03
	AIC	4.24E+03	4.61E+03	4.58E+03	5.71E+03
	$\exp((AIC_{min} - AIC_i)/2)$	1	5.62E-81	5.09E-74	0
Agia Marina Eq. (10)	Equation number	6	5	4	2
	number of coefficients	3	2	2	1
	N		96		96
	R		0.88		0.80
	Slope		0.53		0.67
	SSE		2.02		5.22
	AIC		33.30		69.86
	$\exp((AIC_{min} - AIC_i)/2)$		1		1.15E-08
Budapest Eq. (11)	Equation number	6	5	4	2
	number of coefficients	3	2	2	1
	N		263		263
	R		0.59		0.49
	Slope		0.47		0.95
	SSE		10.73		30.10
	AIC		275.06		389.85
	$\exp((AIC_{min} - AIC_i)/2)$		1		1.19E-25
Beijing Eq. (12)	Equation number	6	5	4	2
	number of coefficients	3	2	2	1
	n	877	877	877	877
	R	0.72	0.89	0.70	0.90
	Slope	1.69	3.16	2.11	5.23
	SSE	189.72	318.04	275.05	769.09
	AIC	2003.90	2198.67	2143.37	2532.00
	$\exp((AIC_{min} - AIC_i)/2)$	1	2.57E-85	2.69E-61	4.4E-230

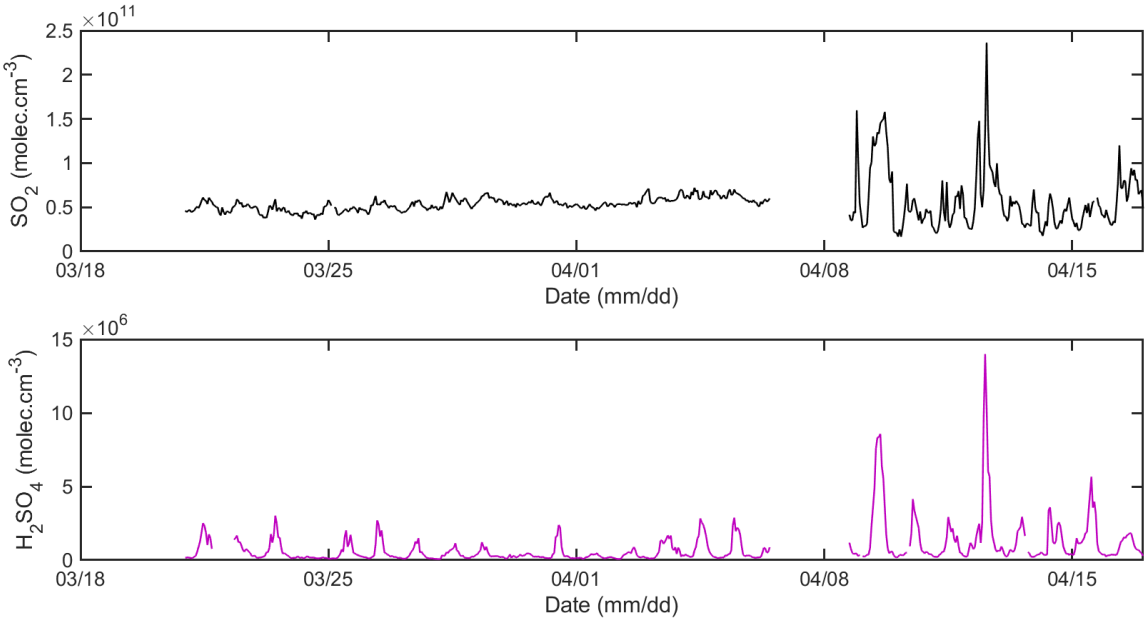


Figure S 1. SO_2 and measured H_2SO_4 concentrations in Budapest showing the change in concentration due to changes in meteorology mid-campaign.

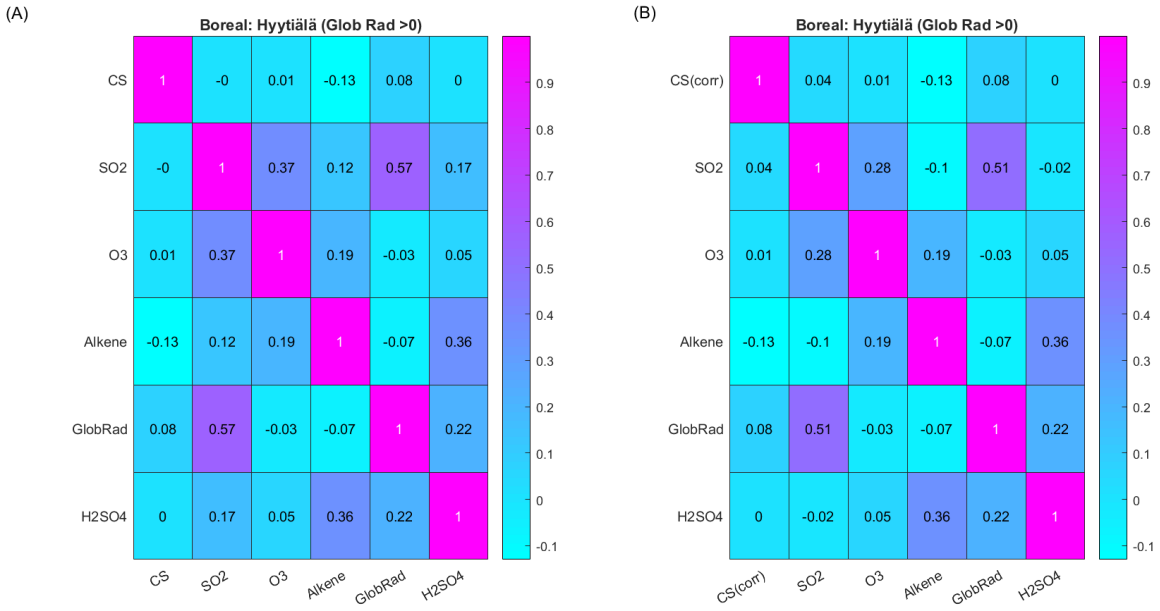


Figure S 2. Spearman's correlation coefficients matrix between variables involved in H_2SO_4 formation and loss at the Hyytiälä station (Global Radiation > 0 W/m^2). CS represents condensation sink in s^{-1}. SO_2, O_3 and MT (monoterpenes) in molecules/cm^{-3}. GlobRad is global radiation in W/m^2. H_2SO_4 is measured sulfuric acid in molecules/cm^{-3}. The color bar represents the Spearman's correlation coefficient. In (A) the condensation sink is not corrected for hygroscopic growth, while in (B) the condensation sink is corrected for hygroscopic growth using the parametrization given by Laakso et al. (2004).

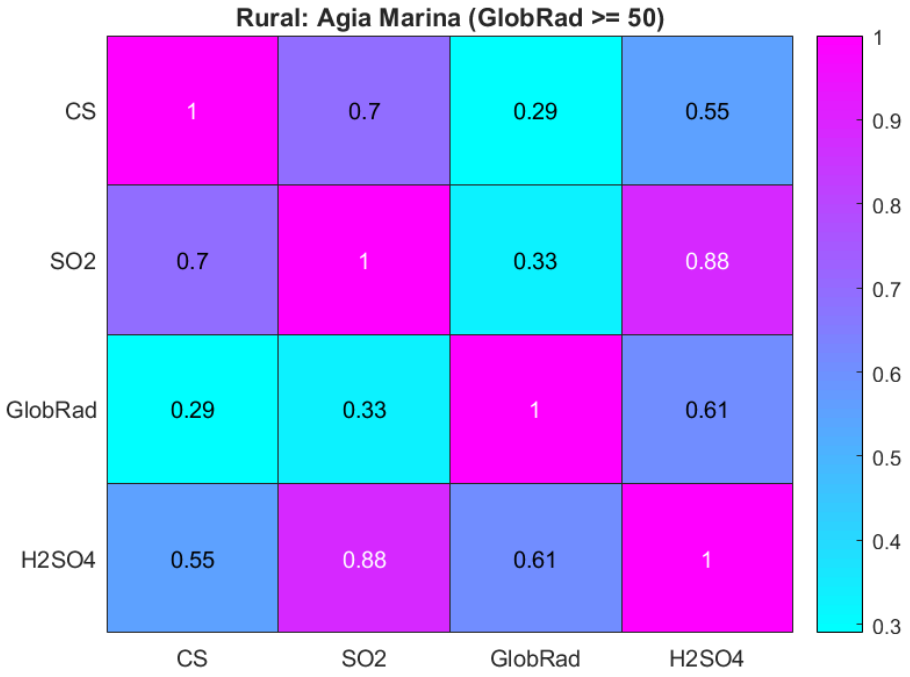


Figure S 3. Spearman's correlation coefficients matrix of variables involved in H_2SO_4 formation and loss at the Agia Marina station (Global Radiation > 50 W/m^2). CS represents condensation sink in s^{-1}. SO_2 is in molecules/cm^{-3}. GlobRad is global radiation in W/m^2. H_2SO_4 is measured sulfuric acid in molecules/cm^{-3}.The color bar represents the Spearman's correlation coefficient.

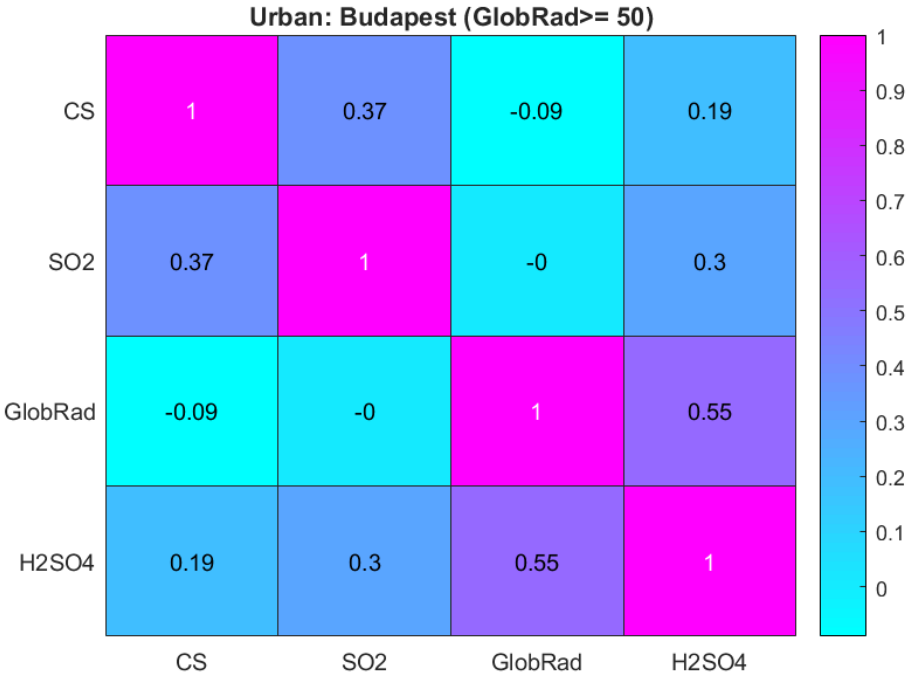


Figure S 4. Spearman's correlation coefficients matrix of variables involved in H_2SO_4 formation and loss at the Budapest station (Global Radiation > 50 W/m^2). CS represents condensation sink in s^{-1}. SO_2 in molecules/cm^{-3}. GlobRad is global radiation in W/m^2. H_2SO_4 is measured sulfuric acid in molecules/cm^{-3}. The color bar represents the Spearman's correlation coefficient.

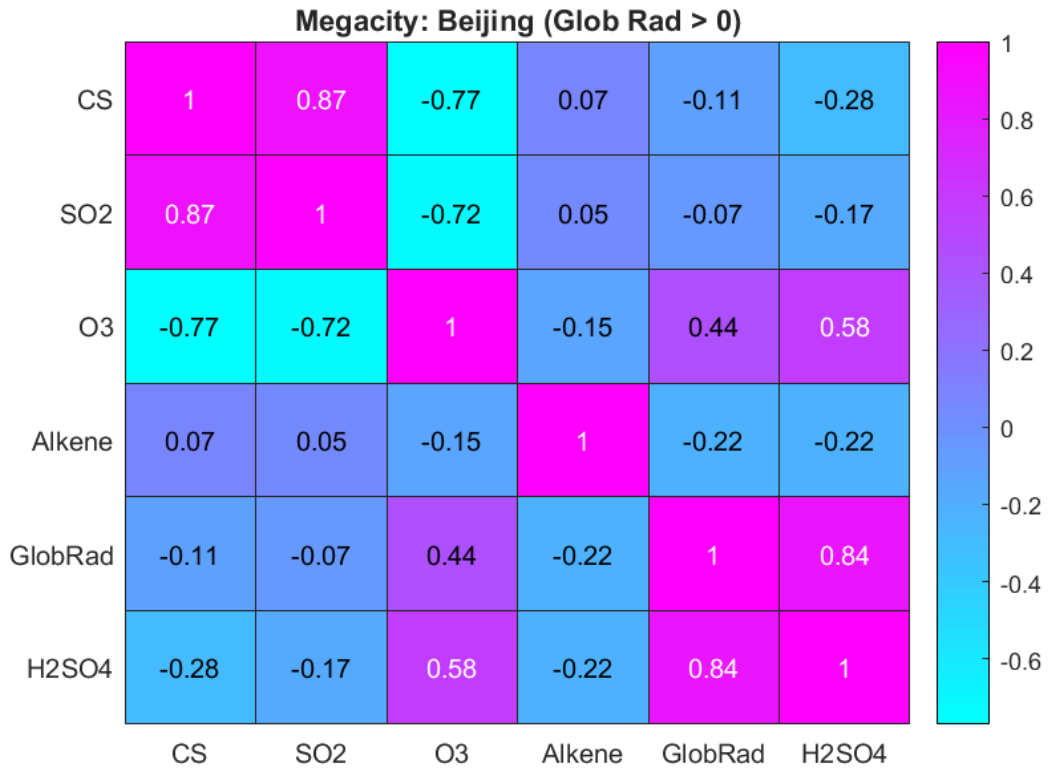


Figure S 5. Spearman's correlation coefficients matrix between variables involved in H_2SO_4 formation and loss at the Beijing station. CS represents condensation sink in s^{-1}. SO_2, O_3 and Alkenes in molecules/cm^{-3}. GlobRad is global radiation in W/m^2. H_2SO_4 is measured sulfuric acid in molecules/cm^{-3}. The color bar represents the Spearman's correlation coefficient.

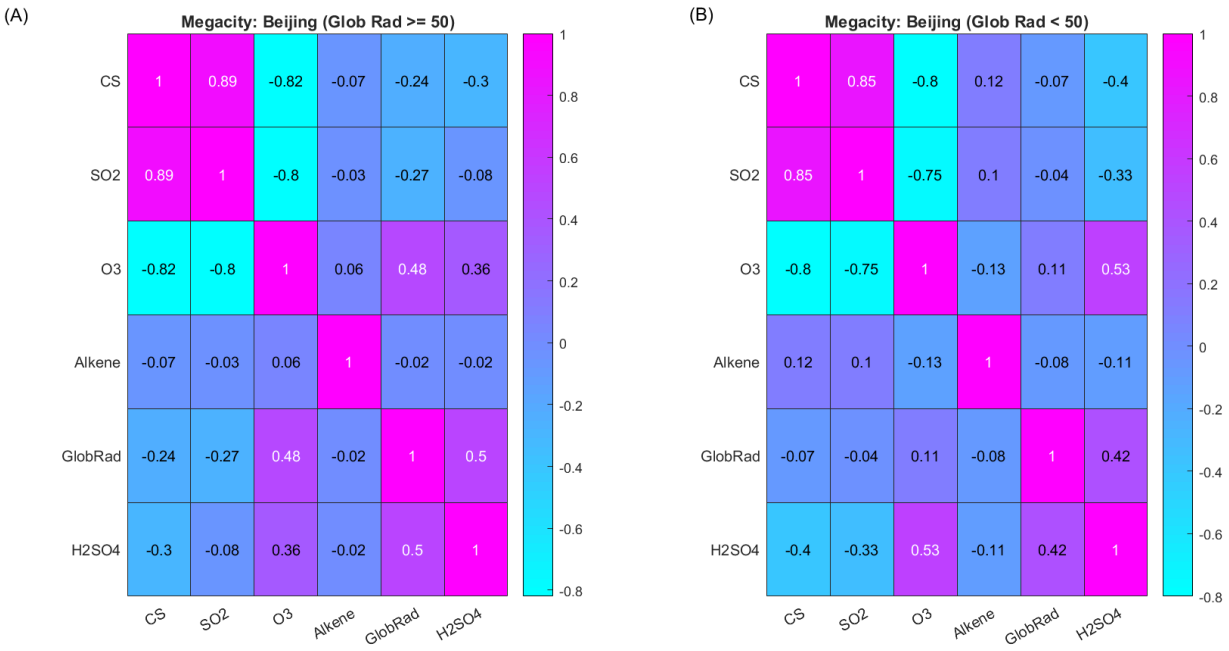


Figure S 6. Spearman's correlation coefficients matrix between variables involved in H_2SO_4 formation and loss at the Beijing station. CS represents condensation sink in s^{-1}. SO_2, O_3 and Alkenes in molecules/cm^{-3}. GlobRad is global radiation in W/m^2. H_2SO_4 is measured sulfuric acid in molecules/cm^{-3}. The color bar represents the Spearman's correlation coefficient. In (A) the daytime correlation coefficients are shown (Global radiation >= 50 W/m^2) and in (B) the nighttime correlation coefficients are shown (Global radiation < 50 W/m^2).

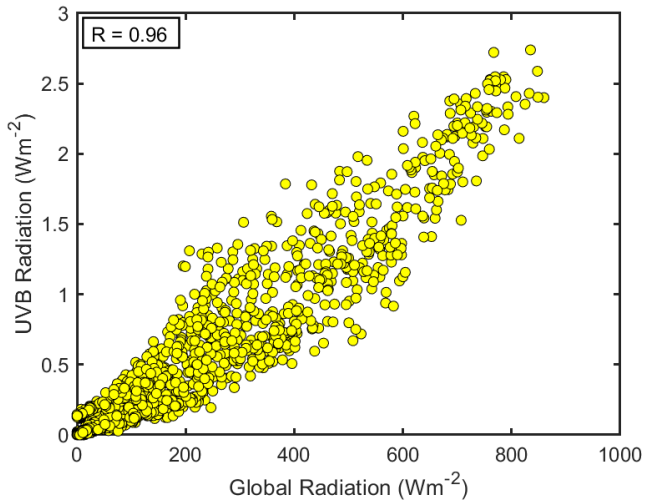


Figure S 7. Comparison between Global radiation and UVB in Hyytiälä. Hourly medians are shown. The total number of data points in the plot is 2306.

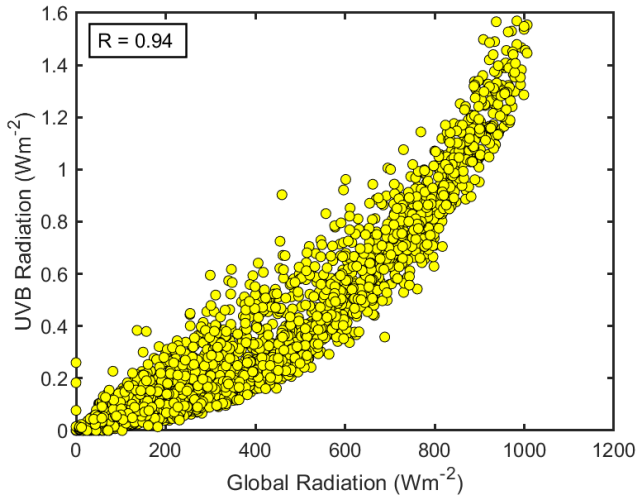


Figure S 8. Comparison between Global radiation and UVB in Beijing. Hourly medians are shown. The total number of data points in the plot is 7106.

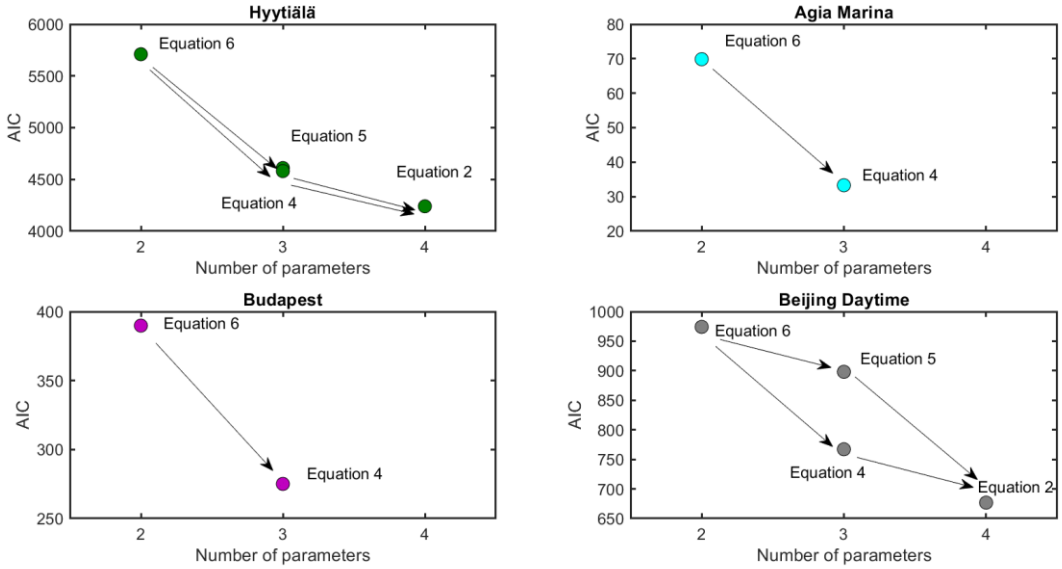


Figure S 9. Evaluation of the goodness of the fit using the Akaike information criterion (AIC) (McElreath, 2018). Number of parameters refers to the number of variables in each equation used. For example, Eq. (2) uses four parameters which are the two sources (Radiation and sCI) and the two sinks (CS and cluster formation).

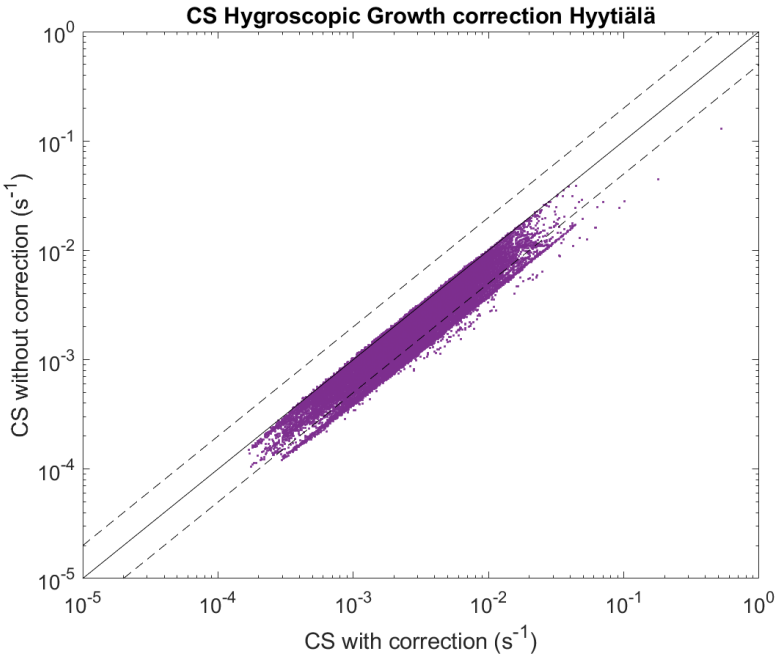


Figure S 10. Effect of hygroscopic growth correction on condensation sink calculation in the boreal forest. The solid line is the 1:1 line and the dashed lines are the 2:1 lines.

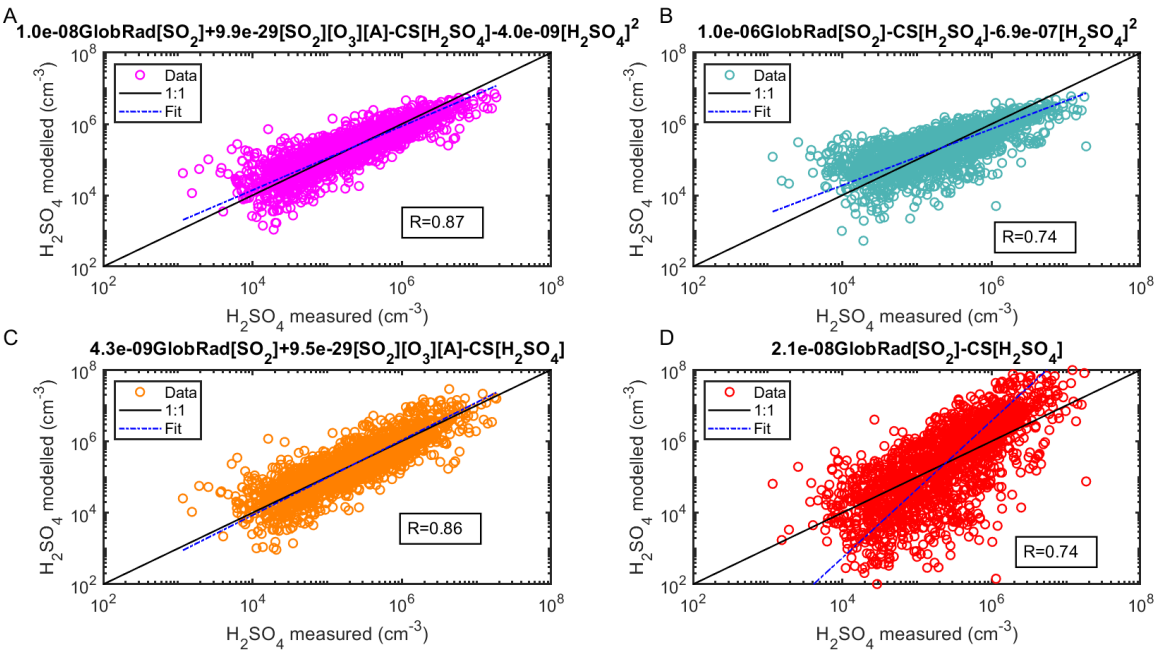


Figure S 11. Sulfuric acid proxy concentration as a function of measured sulfuric acid. Observation at SMEAR II station, Hyytiälä Finland **with CS corrected for hygroscopic growth**. The observed concentrations are measured 2016-2019 using CI-APi-ToF and are 3 h medians resulting in a total of 1594 data points. In (A), the full Eq. (2) is used, in (B) the equation without the stabilized Criegee intermediates source term (Eq. 4) is used, in (C) the equation without the cluster sink term (Eq. 5) is used and in (D) the equation without neither the stabilized Criegee intermediates source term nor the cluster sink term (Eq. 6) is used. The "fit" refers to the fitting between the measured and the proxy-calculated sulfuric acid concentration(log(y)=a.log(x)+b).

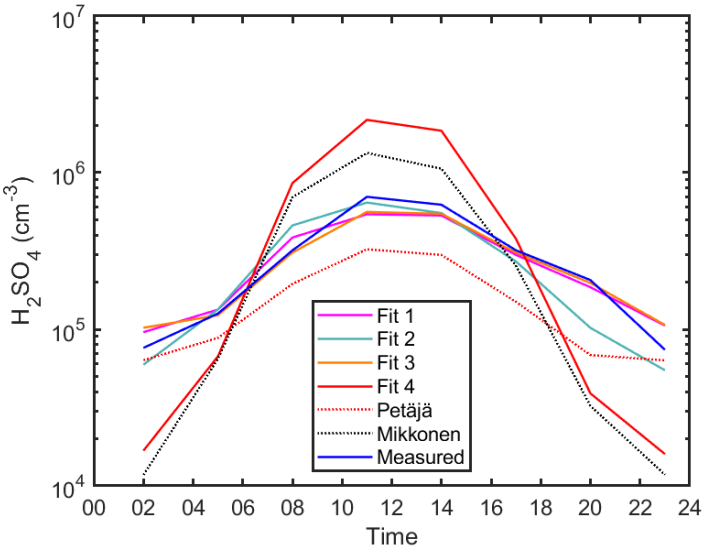
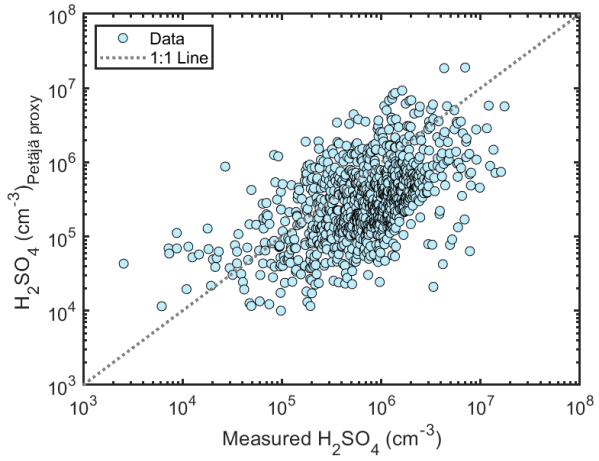
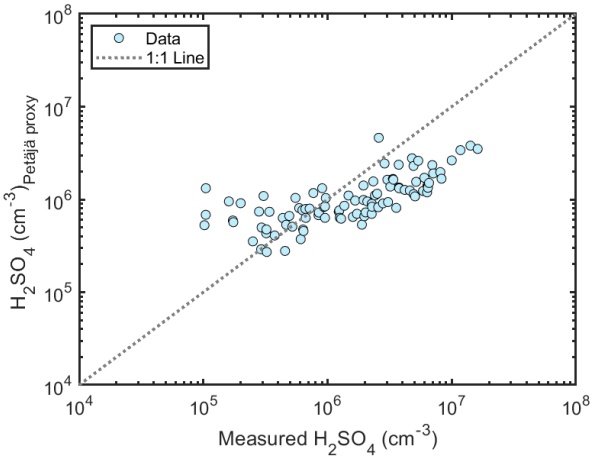


Figure S 12. The diurnal variation of sulfuric acid proxy concentrations using different fits and observed concentrations at SMEAR II in Hyytiälä, Finland **with CS corrected for hygroscopic growth**. Median values are shown. Fits 1, 2, 3 and 4 correspond to the Eqs. (2), (4), (5), and (6), respectively. The Petäjä fit shown is applied using the coefficients reported in Petäjä et al. (2009) (Eq. 7). The Mikkonen fit shown is applied using the coefficients reported in Mikkonen et al. (2011) (Eq. 8).

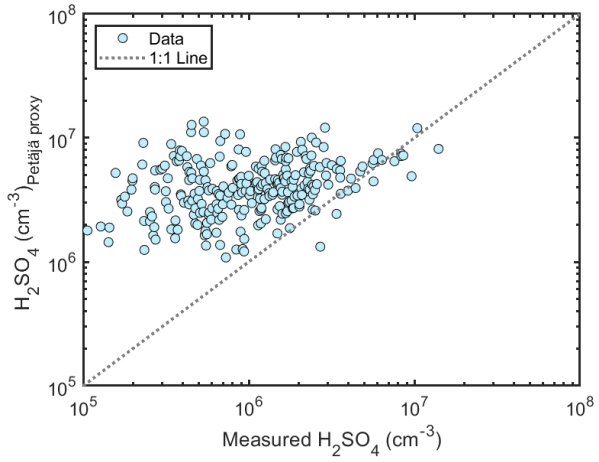
Hyytiälä



Agia Marina



Budapest



Beijing

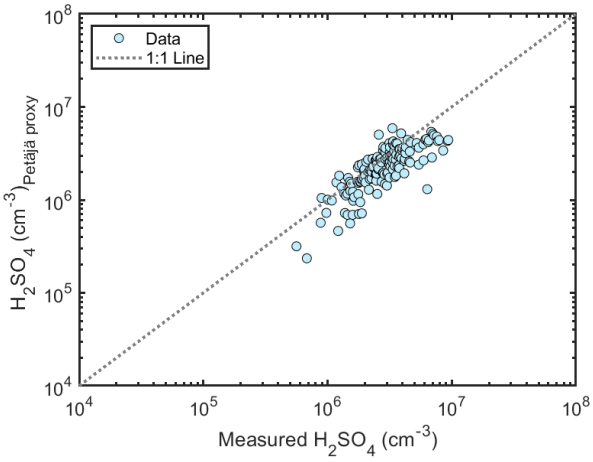
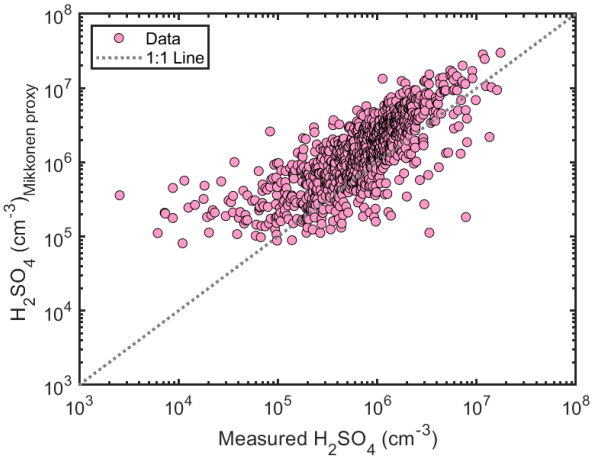
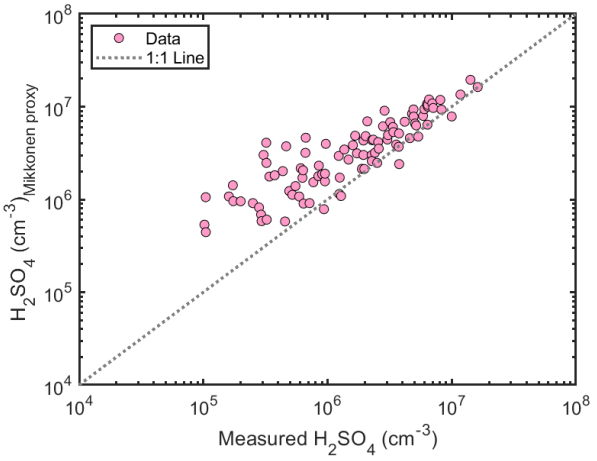


Figure S 13. Scatter plot showing the correlation between measured sulfuric acid and the sulfuric acid concentrations derived from the Petäjä et al. (2009) proxy at the 4 locations during daytime (GlobRad >= 50 W/m^2): Hyytiälä, Agia Marina, Budapest and Beijing.

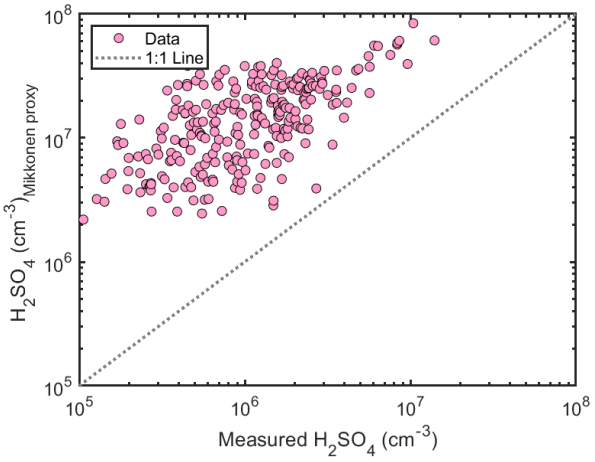
Hyytiälä



Agia Marina



Budapest



Beijing

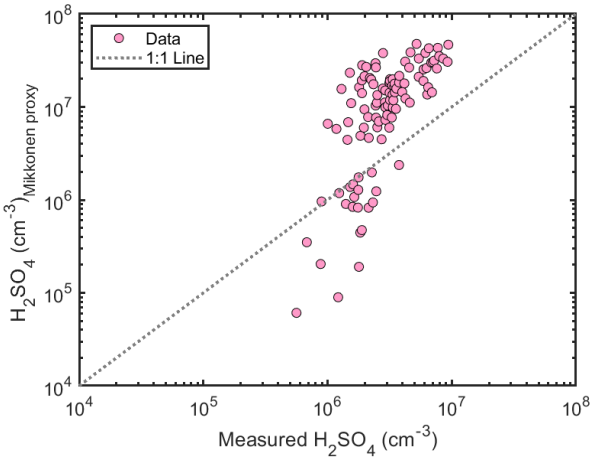


Figure S 14. Scatter plot showing the correlation between measured sulfuric acid and the sulfuric acid concentrations derived from the Mikkonen et al. (2011) proxy at the 4 locations during daytime (GlobRad >= 50 W/m^2): Hyytiälä, Agia Marina, Budapest and Beijing.

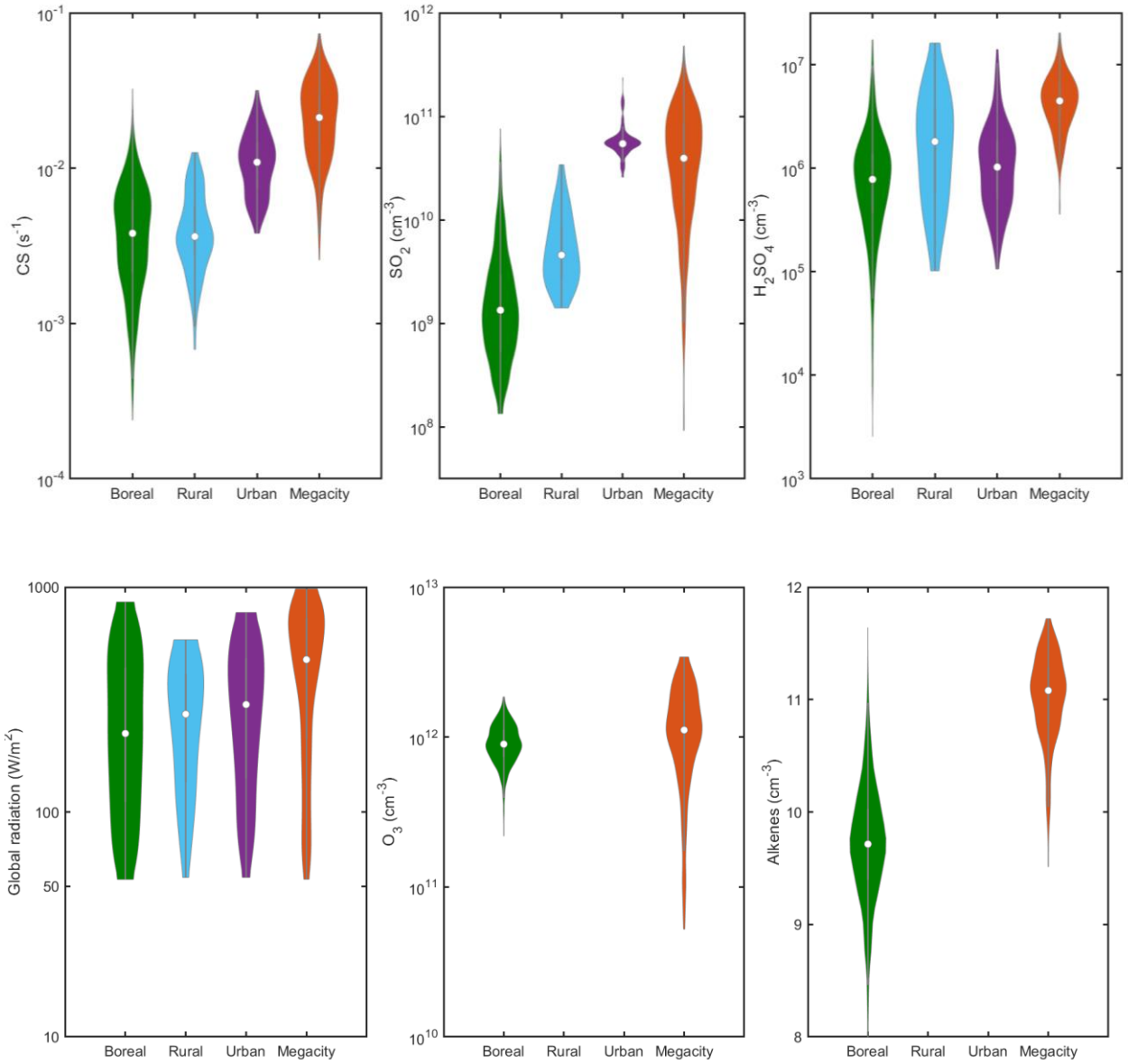


Figure S 15. Daytime data (GlobRad > 50 W/m^2) condensation sink, SO_2, GlobRad and measured H_2SO_4 concentrations in diffrerent environements. The concentrations are displayed as violin plots which are a combination of boxplot and a kernel distribution function on each side of the boxplots. The white circles define the median of the distribution and the edges on the inner grey boxes refer to the 25th and 75th percentiles, respectively.

(A) (B)

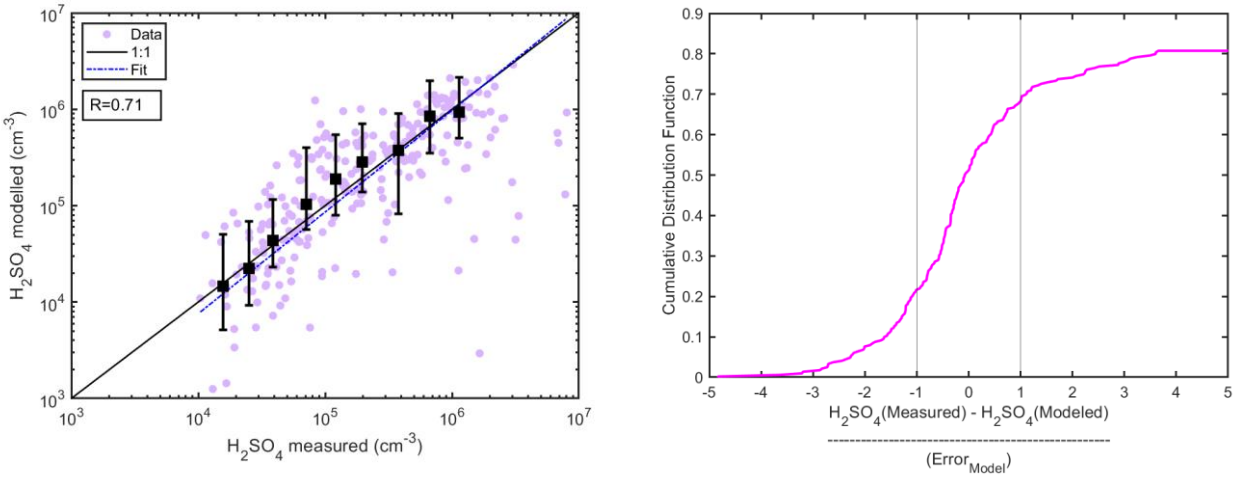


Figure S 16. (A) Sulfuric acid concentrations modeled as a function of measured sulfuric acid at Hyytiälä SMEAR II station. The concentrations shown are 3 h medians coinciding with the alkene measurements every 3 h resulting in a total of 257 data points. The modeled concentrations are the median derived using 10,000 *k* value combinations specific to the site. The colored data points refer to the modeled or predicted concentrations, and the dashed blue line refers to the fit (log(y) = a.log(x)+b) of the aforementioned data points. The black squares are the median modeled concentrations in logarithmically spaced measured sulfuric acid bins and their lower and upper whiskers correspond to 25th and 75th percentiles of the predicted concentrations. (B) Cumulative distribution function of the model error weighted difference between measured and modeled H_2SO_4 concentration (using 257 data points).

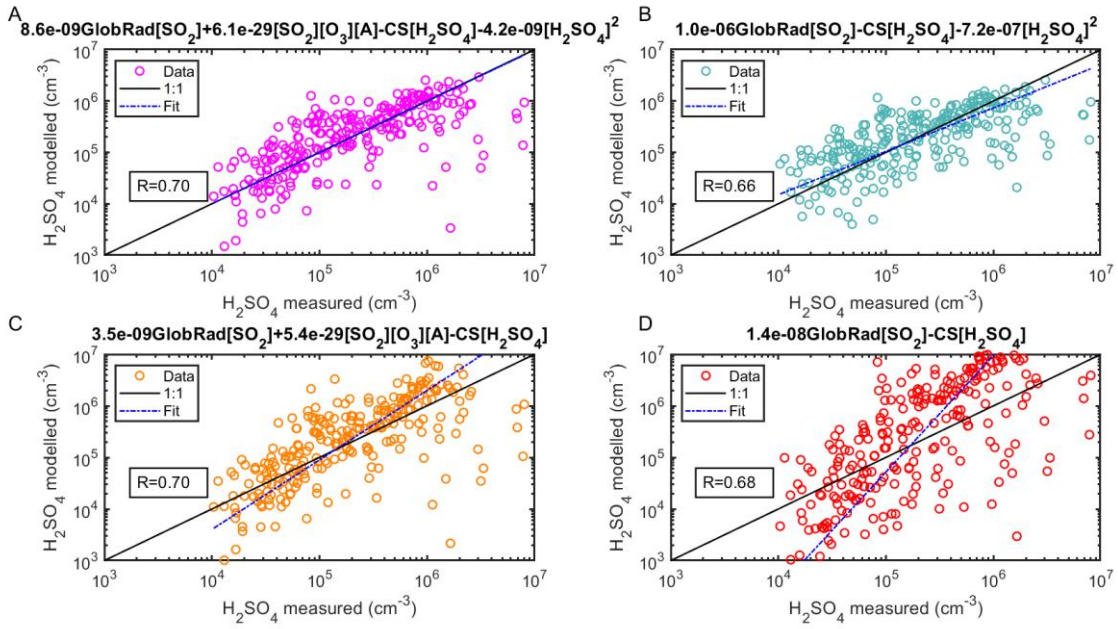


Figure S 17. Sulfuric acid proxy concentration as a function of measured sulfuric acid observed at SMEAR II station, Hyytiälä, Finland using the four different combinations of source and sink terms. The concentrations shown are 3 h medians coinciding with the alkene measurements every 3 h resulting in a total of 257 data points. In (A), the full Eq. (2) is used, in (B) the equation without the stabilized Criegee intermediates source term (Eq. 4) is used, in (C) the equation without the cluster sink term (Eq. 5) is used and in (D) the equation without neither the stabilized Criegee intermediates source term nor the cluster sink term (Eq. 6) is used. The "fit" refers to the fitting between the measured and the proxy-calculated sulfuric acid concentration (log(y)=a.log(x)+b).

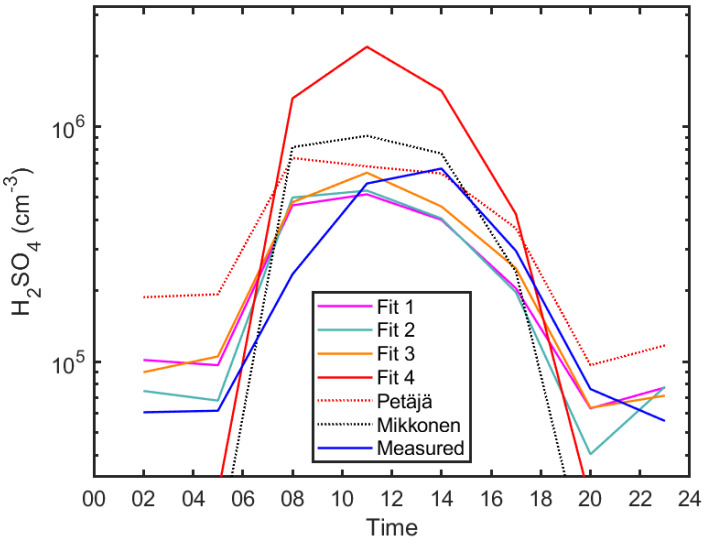


Figure S 18. The diurnal variation of sulfuric acid proxy concentrations using different fits and observed concentrations at SMEAR II in Hyytiälä, Finland. Median values are shown. Fits 1, 2, 3 and 4 correspond to the Eqs.(2), (4), (5), and (6), respectively. The Petäjä fit shown is applied using the coefficients reported in (Petäjä et al., 2009) (Eq.7). The Mikkonen fit shown is applied using the coefficients reported in Mikkonen et al. (2011) (Eq.8).

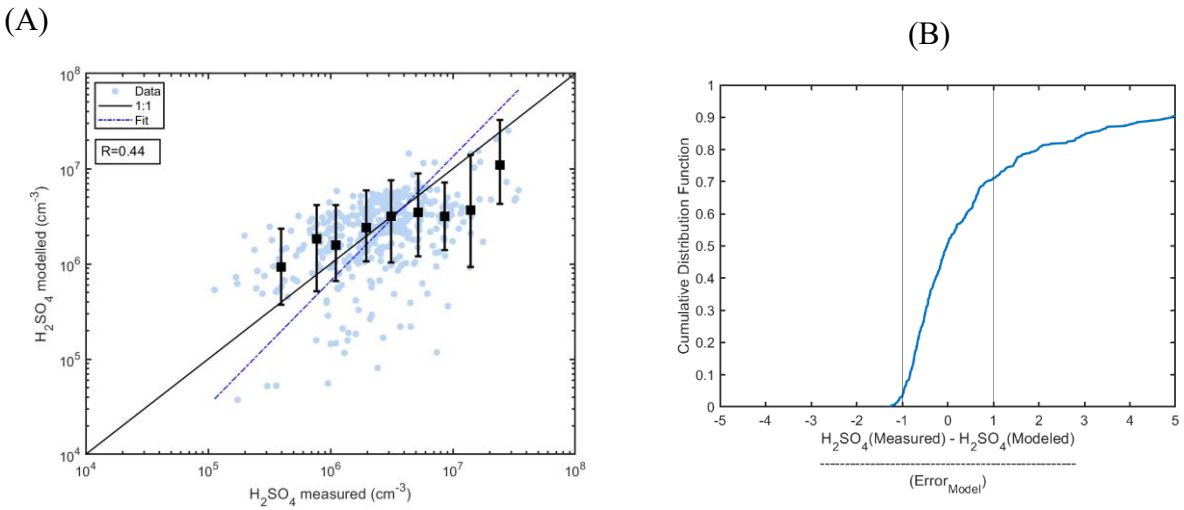


Figure S 19. Sulfuric acid concentrations modeled as a function of measured sulfuric acid at Helsinki SMEAR III station. The concentrations shown are 1 h medians resulting in a total of 416 data points. The modeled concentrations are the median derived using 10,000 *k* value combinations specific to the site. The colored data points refer to the modeled or predicted concentrations, and the dashed blue line refers to the fit (log(y) = a.log(x)+b) of the aforementioned data points. The black squares are the median modeled concentrations in logarithmically spaced measured sulfuric acid bins and their lower and upper whiskers correspond to 25th and 75th percentiles of the predicted concentrations. (B) Cumulative distribution function of the model error weighted difference between measured and modeled H_2SO_4 concentration (using 416 data points).

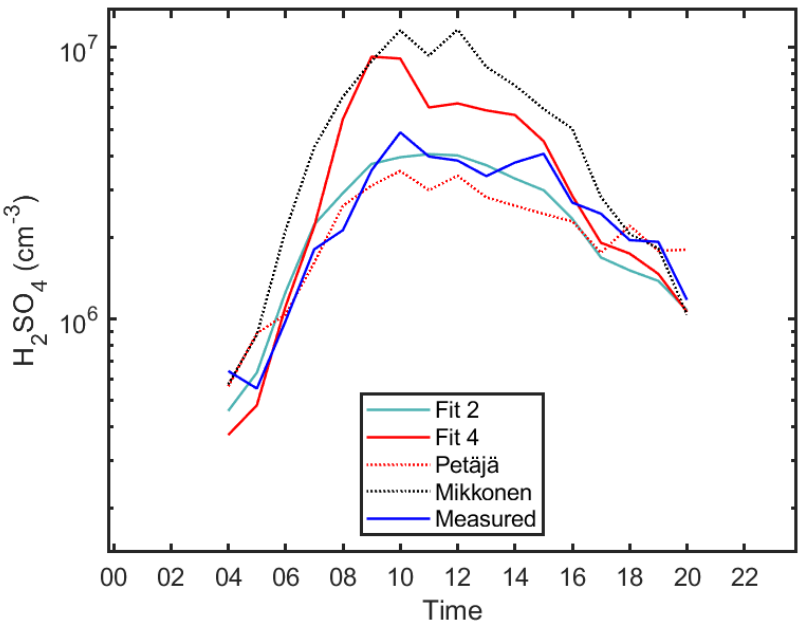


Figure S 20. The diurnal variation of sulfuric acid proxy concentrations using different fits and observed concentrations at SMEAR III in Helsinki, Finland. Median values are shown. Fits 1, 2, 3 and 4 correspond to the Eqs.(2), (4), (5), and (6), respectively. The Petäjä fit shown is applied using the coefficients reported in Petäjä et al. (2009) (Eq.7). The Mikkonen fit shown is applied using the coefficients reported in Mikkonen et al. (2011) (Eq.8).

(A) (B)

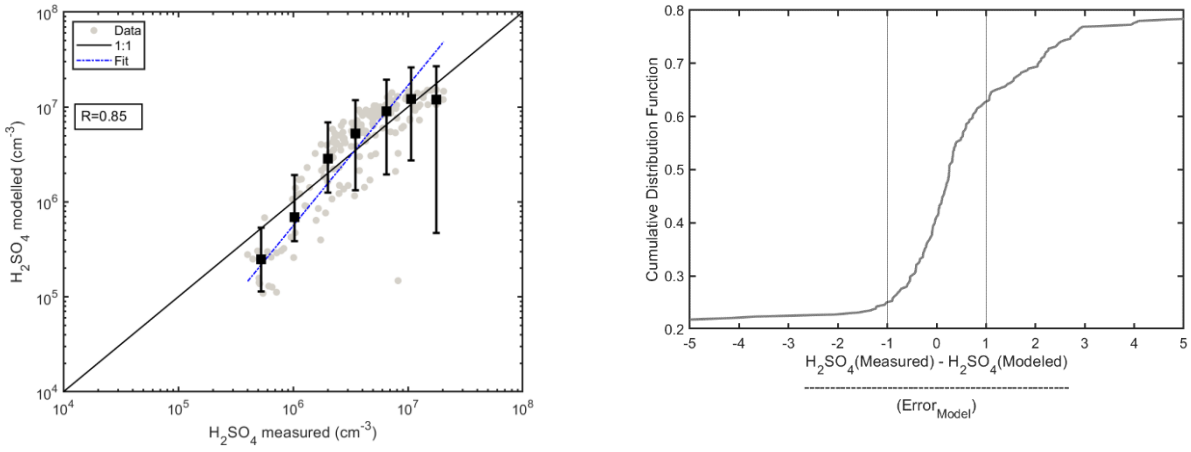


Figure S 21. Sulfuric acid concentrations modeled as a function of measured sulfuric acid in Beijing. The concentrations shown are 1 h medians resulting in a total of 263 data points. The modeled concentrations are the median derived using 10,000 *k* value combinations specific to the site. The gray data points refer to the modeled or predicted concentrations, and the dashed blue line refers to the fit (log(y) = a.log(x)+b) of the aforementioned data points. The black squares are the median modeled concentrations in logarithmically spaced measured sulfuric acid bins and their lower and upper whiskers correspond to 25th and 75th percentiles of the predicted concentrations. (B) Cumulative distribution function of the model error weighted difference between measured and modeled H_2SO_4 concentration (using 268 data points). H_2SO_4 concentration relative to the measured H_2SO_4 concentration (using 268 data points).

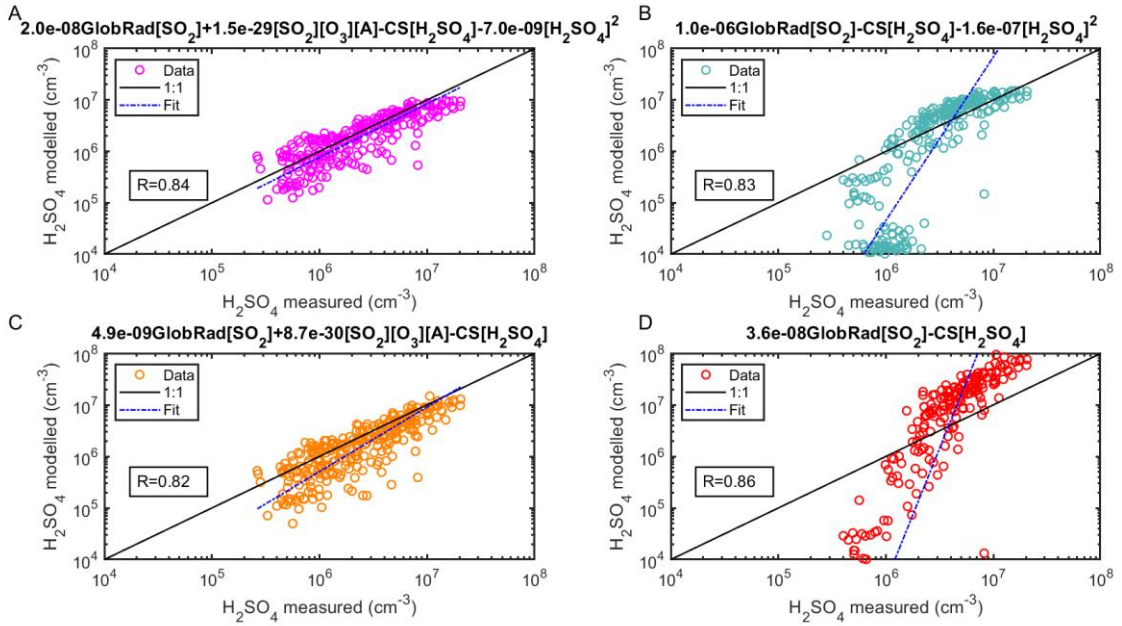


Figure S 22. Sulfuric acid proxy concentration as a function of measured sulfuric acid observed at Beijing, China for the testing data set using the four different combinations of source and sink terms. The concentrations shown are 1 h medians resulting in a total of 268 data points in each subplot. In (A), the full Eq. (2) is used, in (B) the equation without the stabilized Criegee intermediates source term (Eq. 4) is used, in (C) the equation without the cluster sink term (Eq. 5) is used and in (D) the equation without neither the stabilized Criegee intermediates source term nor the cluster sink term (Eq. 6) is used. The "fit" refers to the fitting between the measured and the proxy-calculated sulfuric acid concentration (log(y)=a.log(x)+b).

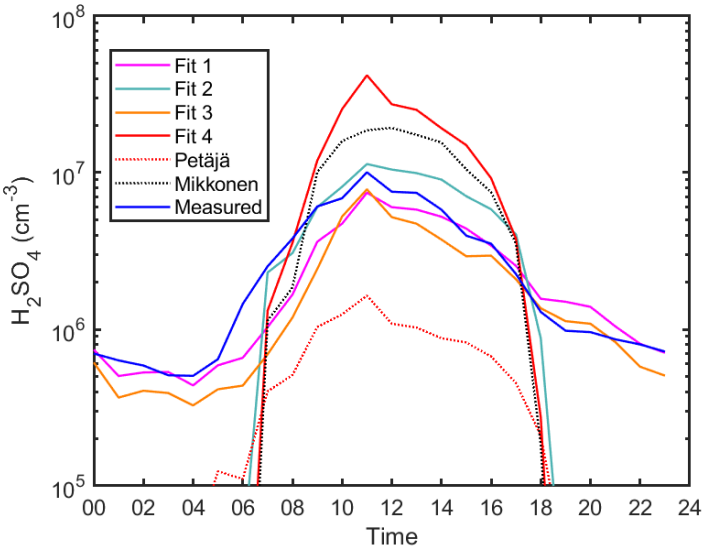


Figure S 23. The diurnal variation of sulfuric acid proxy concentrations using different fits and observed concentrations in Beijing, China for the testing data set. Median values are shown. Fits 1, 2, 3 and 4 correspond to the Eqs. (2), (4), (5), and (6), respectively. The Petäjä fit shown is applied using the coefficients reported in Petäjä et al. (2009) (Eq.7). The Mikkonen fit shown is applied using the coefficients reported in Mikkonen et al. (2011) (Eq.8).

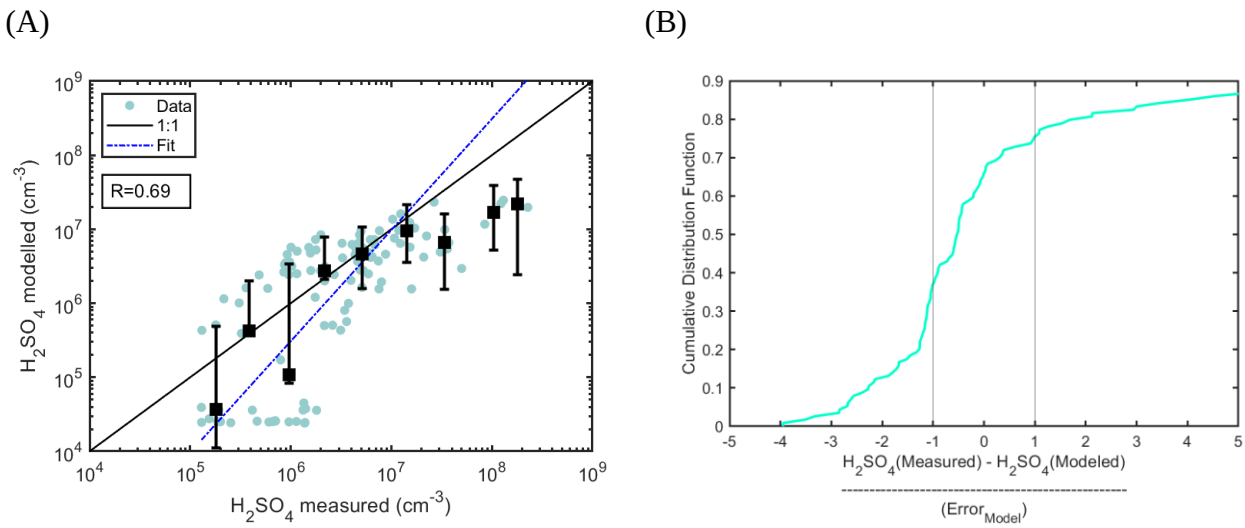


Figure S 24. Sulfuric acid concentrations modeled as a function of measured sulfuric acid at Kilpilahti, Finland. The concentrations shown are 1-hour medians resulting in a total of 114 data points. The modeled concentrations are the median derived using 10,000 *k* value combinations specific to the the boreal forest location. The colored data points refer to the modeled or predicted concentrations, and the dashed blue line refers to the fit (log(y) = a.log(x)+b) of the aforementioned data points. The black squares are the median modeled concentrations in logarithmically spaced measured sulfuric acid bins and their lower and upper whiskers correspond to 25th and 75th percentiles of the predicted concentrations. (B) Cumulative distribution function of the model error weighted difference between measured and modeled H_2SO_4 concentration (using 114 data points).

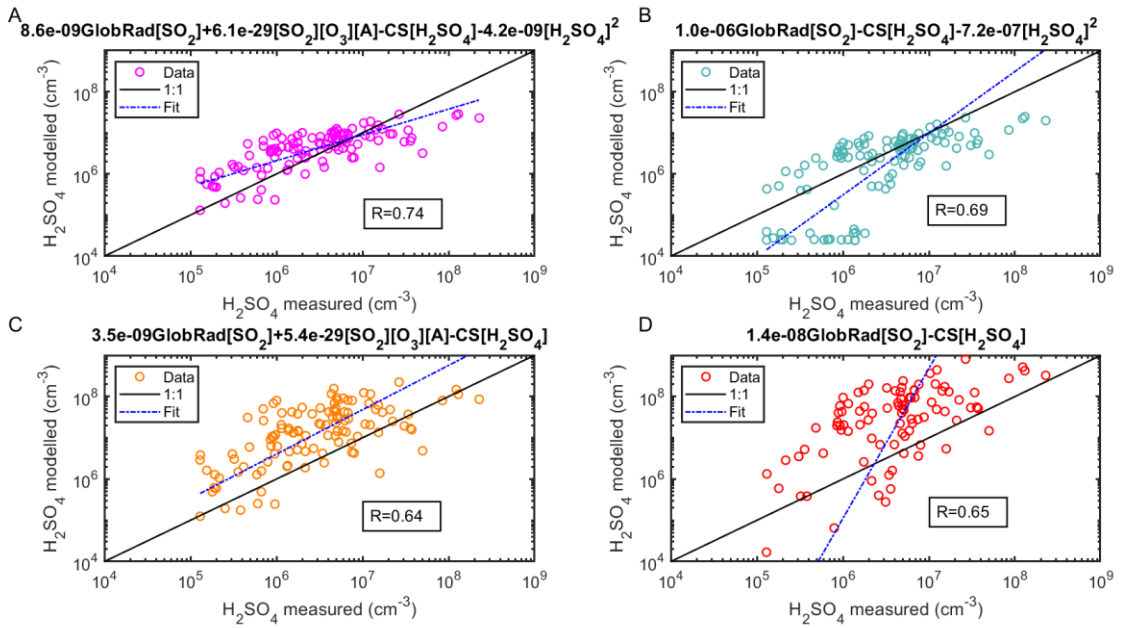


Figure S 25. Sulfuric acid proxy concentration as a function of measured sulfuric acid observed at Kilpilahti, industrial area using the four different combinations of source and sink terms derived from Hyytiälä. The concentrations shown are 1 h medians resulting in a total of 114 data points in each subplot. In (A), the full Eq. (2) is used, in (B) the equation without the stabilized Criegee intermediates source term (Eq. 4) is used, in (C) the equation without the cluster sink term (Eq. 5) is used and in (D) the equation without neither the stabilized Criegee intermediates source term nor the cluster sink

term (Eq. 6) is used. The “fit” refers to the fitting between the measured and the proxy-calculated sulfuric acid concentration (log(y)=a.log(x)+b).

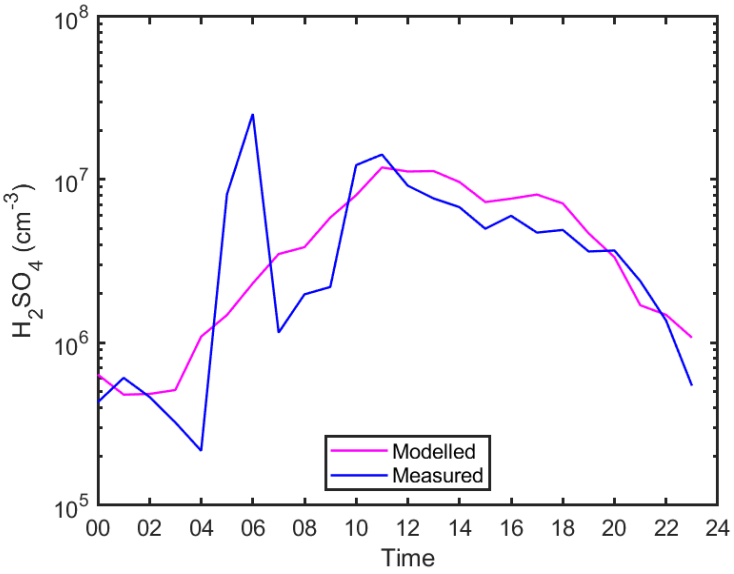


Figure S 26. The diurnal variation of sulfuric acid proxy concentrations observed concentrations at Kilpilahti, industrial area, Finland. Median values are shown. The modeled concentration is predicted using Eq. (9) using the *k* values derived from Hyytiälä SMEAR II station.

References

Laakso, L., Petaja, T., Lehtinen, K. E. J., Kulmala, M., Paatero, J., Horrak, U., Tammet, H., and Joutsensaari, J.: Ion production rate in a boreal forest based on ion, particle and radiation measurements, Atmos Chem Phys, 4, 1933-1943, DOI 10.5194/acp-4-1933-2004, 2004.

McElreath, R.: Statistical rethinking: A Bayesian course with examples in R and Stan, Chapman and Hall/CRC, 2018.

Mikkonen, S., Romakkaniemi, S., Smith, J. N., Korhonen, H., Petaja, T., Plass-Duelmer, C., Boy, M., McMurry, P. H., Lehtinen, K. E. J., Joutsensaari, J., Hamed, A., Mauldin, R. L., Birmili, W., Spindler, G., Arnold, F., Kulmala, M., and Laaksonen, A.: A statistical proxy for sulphuric acid concentration, Atmos Chem Phys, 11, 11319-11334, 10.5194/acp-11-11319-2011, 2011.

Petäjä, T., Mauldin Iii, R., Kosciuch, E., McGrath, J., Nieminen, T., Paasonen, P., Boy, M., Adamov, A., Kotiaho, T., and Kulmala, M.: Sulfuric acid and OH concentrations in a boreal forest site, Atmos. Chem. Phys., 9, 7435-7448, 10.5194/acp-9-7435-2009, 2009.