# Long-term fluxes of carbonyl sulfide and their seasonality and interannual variability in a boreal forest

5 Timo Vesala<sup>1,2,3</sup>, Kukka-Maaria Kohonen<sup>1</sup>, Arnaud P. Praplan<sup>4</sup>, Linda M.J. Kooijmans<sup>5</sup>, Lenka Foltýnová<sup>6</sup>, Pasi Kolari<sup>1</sup>, Markku Kulmala<sup>1</sup>, Jaana Bäck<sup>2</sup>, David Nelson<sup>7</sup>, Dan Yakir<sup>8</sup>, Mark Zahniser<sup>7</sup>, Ivan Mammarella<sup>1</sup>

<sup>1</sup>Institute for Atmospheric and Earth System Research / Physics, University of Helsinki, Helsinki, Finland <sup>2</sup>Institute for Atmospheric and Earth System Research / Forest Sciences, University of Helsinki, Helsinki, Finland <sup>3</sup>Yugra State University, 628012, Khanty-Mansiysk, Russia

- <sup>3</sup>Yugra State University, 628012, Khanty-Mansiysk, Russia
  <sup>4</sup>Finnish Meteorological Institute, Helsinki, Finland
  <sup>5</sup>Meteorology and Air Quality, Wageningen University & Research, Wageningen, The Netherlands
  <sup>6</sup>Global Change Research Institute, Czech Academy of Sciences, Brno, Czech Republic
  <sup>7</sup>Aerodyne Research Inc., Billerica, MA, USA
- 15 <sup>8</sup>Weizmann Institute of Science, Rehovot, Israel

# Contents of this file

Texts S1 to S4

Figures S1 to S10

20 Tables S1 to S3

# Introduction

This Supplementary Information gives a description of the carbonyl sulfide (COS) eddy covariance (EC) flux data processing steps, quality criteria and gap-filling, and a description of the wavelet analysis used in the study. Figures provide additional information on environmental conditions as well as more details on COS flux (FCOS) variation and dependency

25 on environmental variables, and the difference between the Simple Biosphere Model Version 4 (SiB4) FCOS simulations and the upscaled FCOS parameterization. Tables provide information on the weekly, monthly, and seasonal variations of FCOS and start of the growing season by three different methods.

#### Eddy covariance data processing

EC data were processed using the EddyUH software (Mammarella et al., 2016) following the recommendations given in Kohonen et al. (2020). Raw data were despiked on the basis of the maximum difference allowed between two subsequent data points, two-dimensional coordinate rotation was used to rotate the coordinate frame, and turbulent fluctuations were

- determined from linear detrending. Lag times of COS, carbon monoxide (CO), and carbon dioxide (CO<sub>2</sub>) were determined from the maximum cross-covariance of CO<sub>2</sub> with vertical wind speed (*w*), while the lag time of water vapor (H<sub>2</sub>O) was determined from the maximum cross-covariance of H<sub>2</sub>O with *w*. High-frequency spectral corrections were calculated according to Mammarella et al. (2009), so that the response time of CO<sub>2</sub> was also used for COS and CO spectral corrections. Low-frequency correction was done according to Rannik & Vesala (1999).
- 40 Quality screening was applied to the data by tests for flux stationarity ( $\leq 0.3$ ) and limits for kurtosis (1 < Ku < 8) and skewness (-2 < Sk < 2). In addition, only 100 spikes were allowed for each 30 min period and the second wind rotation angle was not allowed to exceed 10 degrees. Friction velocity ( $u_*$ ) filtering with a threshold of 0.3 m/s was applied to exclude time periods with low turbulence. Finally, fluxes were gap-filled using the function

45 
$$FCOS = a*I/(I+b)+c*D+d$$

where *I* is photosynthetic photon flux density, *D* is vapor presssure deficit (VPD), and *a*, *b*, *c*, and *d* are fitting parameters.

The uncertainty of the cumulative flux was estimated using a bootstrap method. This was done because the uncertainty of single 30 min fluxes consists of both random and systematic uncertainties, as well as the uncertainty of the gap-filling function that cannot be mixed with flux measurement uncertainties. In the bootstrap method, we assumed a 20% total uncertainty that mostly consists of processing uncertainty (Kohonen et al., 2019), and a synthetic data set was randomly sampled from the data set (including the 20 % uncertainty) 10 000 times. The synthetic data sets consisted of an amount of data points equal to those in the original data set, but as the samples were drawn at random, some data points may be included several times while other points may not be included at all. The overall uncertainty was calculated as the difference

to the 95th percentile of the 10 000 bootstrap sample sums.

# Text S2

The VPD was calculated as the difference between the saturation water vapor pressure ( $e_s$ ) and the actual water vapor pressure ( $e_a$ ) as

$$VPD = e_s - e_a \tag{2}$$

60 where

$$e_{s} = 0.618 \exp\left(\frac{17.27 T_{a}}{T_{a} + 237.3}\right)$$
(3)

$$e_a = \frac{RH e_s}{100} \tag{4}$$

where  $T_a$  is air temperature and RH relative humidity of air, both calculated as an average of measurements done at 16.8 m and 33.6 m heights, representing the flux measurement height of 23 m.

65 Commencement of the growing season was determined, following the recommendations of Suni et al. (2003a), by a moving average  $T_a$  with a 5-day window (with a threshold of 3.3 °C indicating the start of photosynthetic uptake) and by a  $T_a$  dependent variable *S* that describes the stage of physiological development, introduced by Pelkonen & Hari (1980):

$$S_i = S_{i-1} + S_t \tag{5}$$

where

$$S_{t} = \frac{100}{1 + 100a^{-|T-S/c|}} - \frac{100}{1 + 100a^{|T-S/c|}}$$

In Eq. 6, a = 2 and c = 600 are fitted constants as in Suni et al. (2003a) and Pelkonen & Hari (1980), and Eqs 5 and 6 were solved with the Euler method with a half-hour time step dt. For simplicity, we scaled *S* by *c* as in Suni et al. (2003a). The threshold for the start of the growing season by this method was S/c > 1.81, as in Suni et al. (2003a), and *S* at i = 0 was defined as  $S_0 = 0$ .

(6)

75

70

## Text S3

#### Multivariate linear regression analysis

Multivariate linear regression analysis was used to find the dependencies of FCOS on various environmental factors: VPD,  $T_a$ , photosynthetically active radiation (PAR), RH, soil temperature ( $T_s$ ), soil water content (SWC), net radiation ( $R_n$ ), and atmospheric COS mixing ratio. All the computations were done in R (R Core Team, 2019) using base functions, and the

80 atmospheric COS mixing ratio. All the computations were done in R (R Core Team, 2019) using base functions, and th variance inflation factor (VIF) was computed using the vif function within package car (Fox & Weisberg, 2011).

We used VPD and  $T_a$  within the same linear regression model, despite knowing that VPD is obtained using  $T_a$ . However, VPD is computed as an exponential function of  $T_a$ , which technically does not break the assumption of linear independence of explanatory variables within the regression model. Nonetheless, there is naturally some multicollinearity among almost all environmental factors, which is necessary to treat. The VIF was used to see whether the multicollinearity among explanatory variables within each regression model is of an acceptable level to not brake the assumption of the linear regression analysis. The VIF quantifies the severity of multicollinearity in an ordinary least-squares regression analysis. It provides an index that measures how much the variance (the square of the estimate's standard deviation) of an estimated regression coefficient is increased because of collinearity. VIF < 10 is a commonly used cutoff for acceptable multicollinearity. All the models which included more than one variable were tested with VIF, which was in all cases less than 10. Two variables showing the same physical quantity were never used in the same model ( $T_a$  and  $T_s$ , PAR and  $R_n$ , or RH and VPD).

## Text S4

100

#### 95 Wavelet coherence analysis

Magnitude-squared wavelet coherence was computed using MATLAB's (MATLAB R2019a) wcoherence function, which uses the analytic Morlet wavelet (Grinsted et al., 2004; Lau & Weng, 1995). As input to the function, fluxes were first quality screened and gap-filled. For longer measurement gaps (e.g., winter), fluxes were forced to zero. Thus, for longer data gaps the wavelet coherence is also zero. Environmental variables (PAR,  $T_{\alpha}$ , and VPD) were averaged to 30 min values to match the time stamp of the flux data.



105 Figure S1. Measured (orange) and gap-filled (gray) 30 min carbonyl sulfide flux (FCOS) in 2013–2017. Black circles represent the monthly median FCOS and whiskers show the 25th and 75th percentiles.



110 Figure S2. Time series of photosynthetically active radiation (PAR) (a), air temperature (*T<sub>a</sub>*) (b), and vapor pressure deficit (VPD) (c). Dots represent 30 min values and black circles are the monthly medians. Whiskers show the 25th and 75th percentiles.



Figure S3. Soil temperature ( $T_s$ ) (a) and soil water content (SWC) (b) in Hyytiälä. Colored dots represent the 30 min average values, while black circles and whiskers are the monthly medians and their 25th and 75th percentiles, respectively.



Figure S4. Monthly medians of photosynthetically active radiation (PAR) (a), air temperature ( $T_a$ ) (b), vapor pressure deficit (VPD) (c), and soil water content (SWC) (d) for all measurement years. Error bars show the 25th and 75th percentiles. The monthly median of all years is presented in dark blue.



Figure S5. Cumulative carbonyl sulfide flux (FCOS) (left, in pmol m<sup>-2</sup> s<sup>-1</sup>) and heat sum (right, in °C) since the beginning of the year in different years. Note that the start date of FCOS measurements differs between years. FCOS measurements started on day 91 in 2013, day 87 in 2014, day 92 in 2016, and day 27 in 2017.



Figure S6. Daily median carbonyl sulfide flux (FCOS) against daily median of environmental variables: photosynthetically active radiation (PAR), net radiation, air temperature (*T<sub>a</sub>*), soil temperature (*T<sub>s</sub>*), vapor pressure deficit (VPD), soil water content (SWC), relative humidity (RH), and carbonyl sulfide (COS) mixing ratio. Black circles represent medians of data divided into 12 equal sized bins and black whiskers are the 25th and 75th percentiles.



Figure S7. Magnitude-squared wavelet coherence of carbonyl sulfide flux (FCOS) with photosynthetically active radiation (PAR), air temperature ( $T_a$ ), and vapor pressure deficit (VPD) for the gap-filled data (years 2015–2017). The color shows the coherence between the two variables at each timescale from 30 min to 512 days (y-axis). Arrows show significant correlation and the phase difference between the two time series. Arrows pointing to the right mean that the two time series are in phase, arrows pointing to the left mean antiphase, and arrows pointing upward mean a quarter cycle difference so that the first time series (FCOS) is leading the other.



135 Figure S8. Time series of measured (blue) and parameterized daily carbonyl sulfide flux (FCOS) (black). Subplots a-d show the different combinations of parameter functions, while (a) was chosen as the best option. FLAI = foliage and canopy light penetration, FPAR = stomatal response to PAR, FS = phenology of biochemical reactions, FVPD = stomatal regulation.



Figure S9. Scatter plots of the daily measured carbonyl sulfide flux (FCOS) against different combinations of parameter functions. Gray dots represent daily values, the red line the least-squares linear fit, and the dashed black line is the 1:1 line. FLAI = foliage and canopy light penetration, FPAR = stomatal response to PAR, FS = phenology of biochemical reactions, FVPD = stomatal regulation.



145 Figure S10. Difference between daily average carbonyl sulfide flux (FCOS) based on the Simple Biosphere Model Version 4 (SiB4) simulations and the upscaled parameterization (Param) for the period April–October (top) and November–March (bottom).

	S/c > 1.81	$T_5 > 3.3 \ ^{\circ}C$	$FCOS < 0.3*FCOS_{MIN}$
2013	105	104	105
2014	106	103	118
2015	98	97	NA
2016	88	89	<91
2017	120	120	118

150

Table S1. Commencement day of the growing season by different methods. Variable S/c describes the stage of physiological spring development (Pelkonen & Hari, 1980) and  $T_5$  is the moving average of air temperature using a 5-day window. The carbonyl sulfide flux (FCOS) criterion was filled when the uptake threshold (< 30 % of FCOS<sub>MIN</sub>) was exceeded for 5 consecutive days without falling back for more than 5 consecutive days.

	2013	2014	2015	2016	2017
January	NA	NA	NA	NA	-2.24 (0.10)
February	NA	NA	NA	NA	-2.35 (0.51)
March	NA	-2.52 (0.09)	NA	NA	-2.17 (0.82)
April	-9.17 (0.26)	-2.95 (0.81)	NA	-7.95 (0.38)	-4.98 (0.76)
May	-13.2 (0.58)	-8.61 (0.39)	NA	-12.8 (0.54)	-9.98 (0.62)
June	-18.5 (0.60)	-16.1 (0.51)	NA	-18.4 (0.54)	-18.1 (0.70)
July	-16.4 (0.62)	-9.96 (0.63)	-16.5 (0.59)	-17.6 (0.52)	-21.9 (0.63)
August	-7.84 (0.12)	-5.58 (0.32)	-11.8 (0.53)	-14.9 (0.49)	-17.8 (0.64)
September	-9.50 (0.29)	-1.21 (0.20)	-8.31 (0.51)	-10.9 (0.35)	NA
October	-3.18 (0.0)	NA	-6.25 (0.40)	-10.2 (0.61)	NA
November	NA	NA	NA	-1.48 (0.16)	NA
December	NA	NA	NA	NA	NA

Table S2. Monthly median carbonyl sulfide flux (pmol m<sup>-2</sup> s<sup>-1</sup>) in different years. Fraction of monthly measured 30 min fluxes 155 (non-gapfilled) are indicated in the brackets.

	2013	2014	2015	2016	2017
FCOS <sub>MAX</sub>	-23.2	-19.2	-19.8	-22.3	-23.9
Date	18.6	19.6	11.7	9.6	23.7
Week	25	25	28	23	29

Table S3. Maximum weekly median carbonyl sulfide flux (FCOS) (pmol m<sup>-2</sup> s<sup>-1</sup>) and its occurrence date and week.