## S1  Time series of ground-based and satellite HCHO columns

Fig. S1 displays the time series of monthly-averaged HCHO columns from the seven FTIR stations used in our study, along with the TROPOMI observations and the bias-corrected TROPOMI HCHO data. Similarly, in Fig. S2, we show the comparison with MAX-DOAS data. The station data are smoothed following the procedure described in Sect. 4. Furthermore, the data are
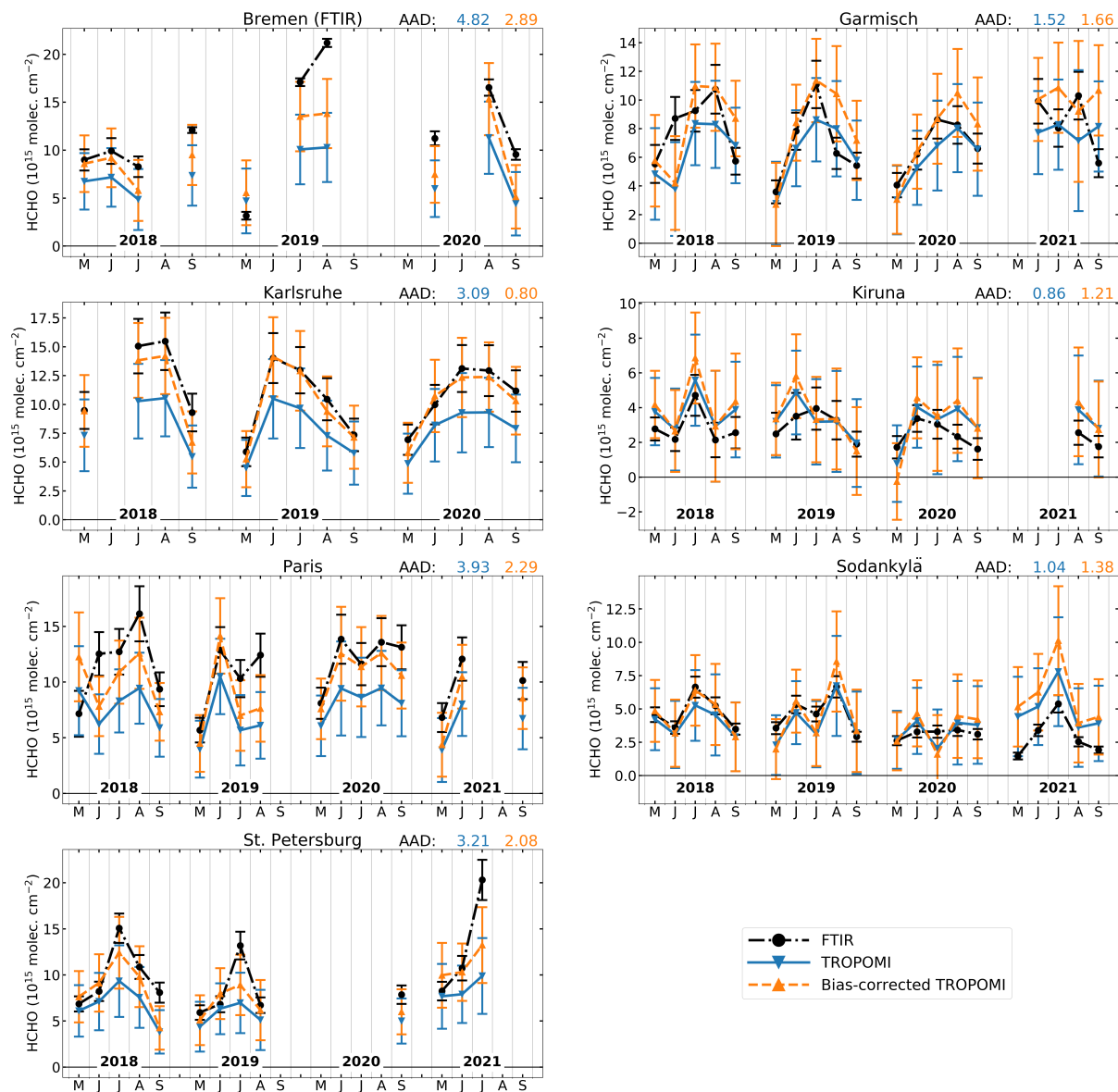
**Figure S1.** Time series of monthly averages of FTIR, TROPOMI, and bias-corrected TROPOMI HCHO vertical columns for all stations. The bias correction is based on the fit in Fig. 3 in the main text. The $x$ axis depicts the month and year, running from May to September. The average absolute deviation (AAD), calculated as the mean of the absolute differences between TROPOMI and FTIR in units of $10^{15}$ molec. cm$^{-2}$, is given at the top right of each panel.

collocated, taking only TROPOMI scenes within 20 km of the station, and taking only ground-based measurements within 3 h of the TROPOMI overpass. The bias correction generally decreases the discrepancy between ground-based and satellite data, except at few locations (Kiruna FTIR and Athens MAX-DOAS).
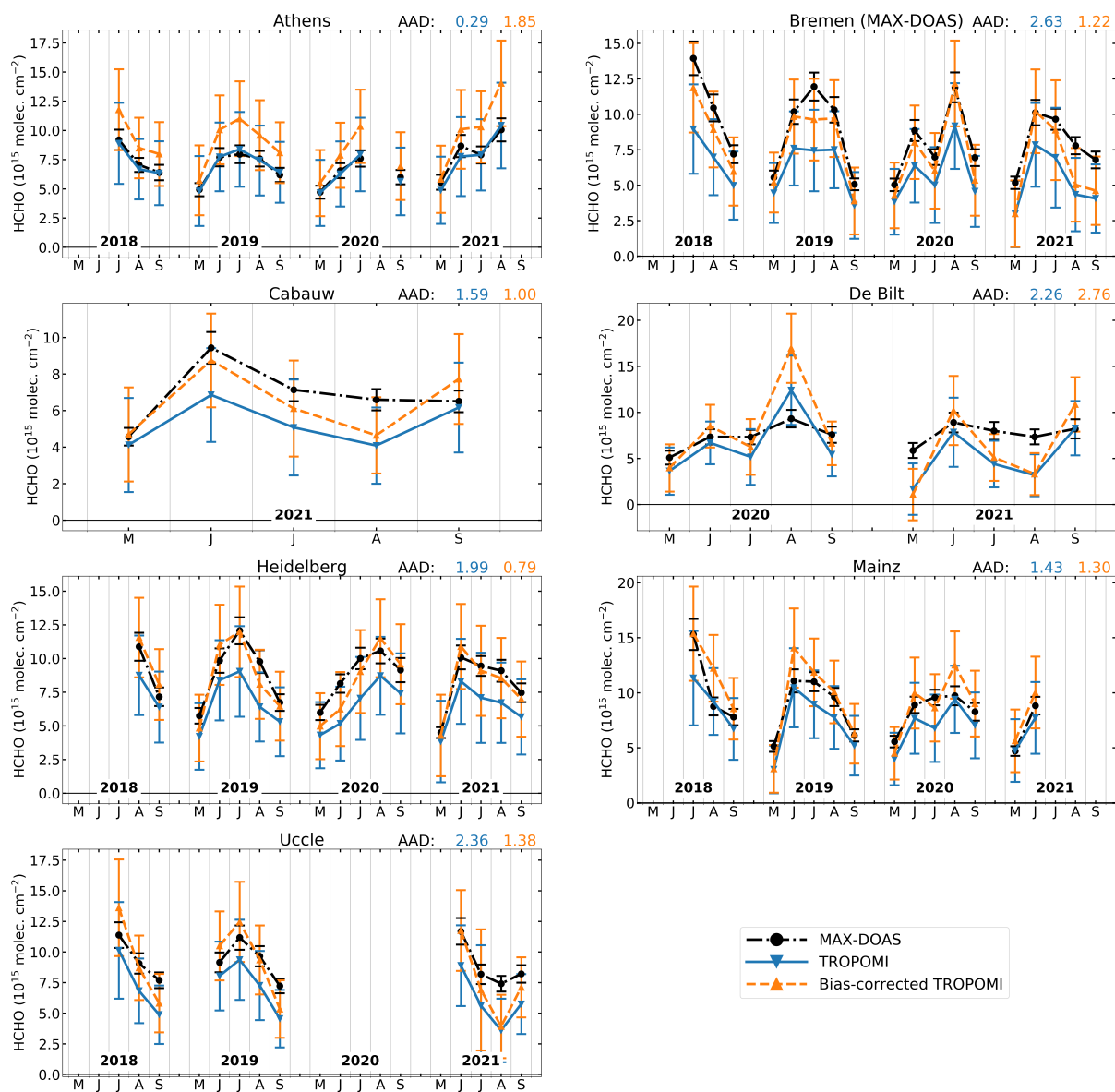
**Figure S2.** Time series of monthly averages of MAX-DOAS, TROPOMI, and bias-corrected TROPOMI HCHO vertical columns for all stations. The bias correction is based on FTIR data (Fig. 3 in the main text). The $x$ axis depicts the month and year, running from May to September. The average absolute deviation (AAD) between TROPOMI and MAX-DOAS is given at the top right of each panel.

## S2    Top-down pyrogenic and anthropogenic VOC emissions

The top-down inversion scheme optimizes three emission categories: biogenic emissions, pyrogenic emissions, and anthropogenic VOC emissions. The former are the main focus of this work and are discussed at length in the main text. This section presents the biomass burning VOC emissions (Fig. S3) and the anthropogenic VOC emissions (Fig. S4).
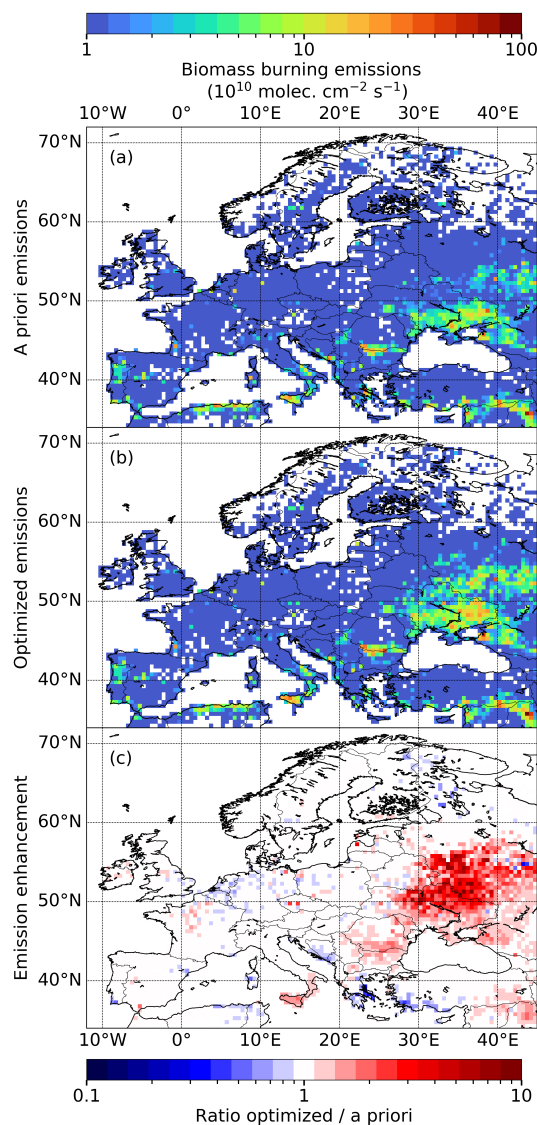
**Figure S3.** The top panel (**a**) shows a priori pyrogenic emissions from the QFEDv2.4 inventory (Darmenov and da Silva, 2015) using emission factors of Andreae (2019). The middle panel (**b**) shows the optimized fire emissions resulting from our top-down inversion. The bottom panel (**c**) shows the enhancement map derived as the ratio of optimized over a priori emissions (**b** divided by **a**). The data are averaged over the summer months (May to September) from 2018 to 2021.
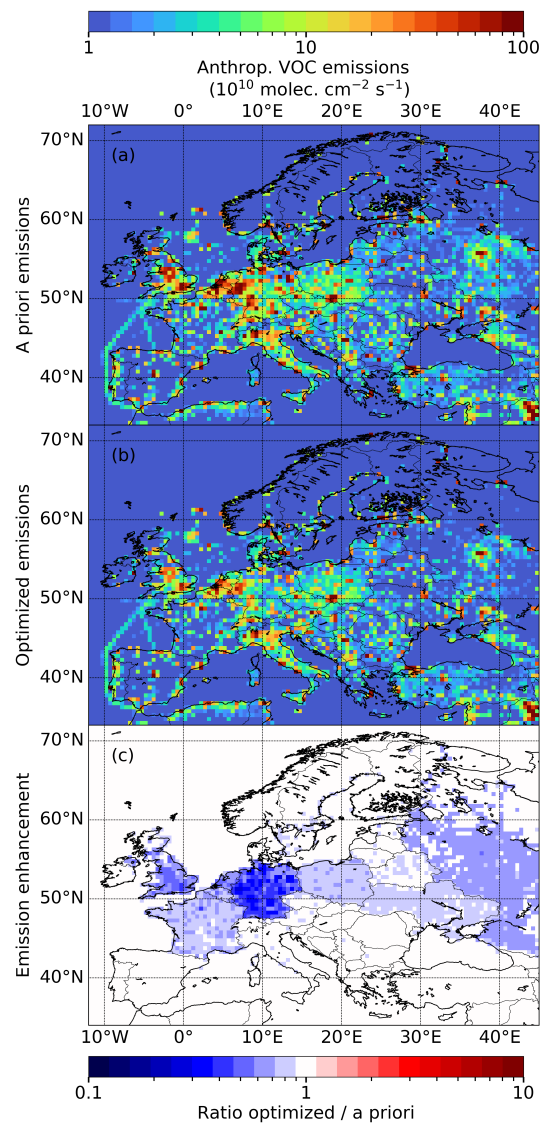
**Figure S4.** Same as Fig. S3, but for anthropogenic VOC emissions. Panel (**a**) shows a priori emissions from CAMS-GLOB-ANT version 4.3 (Granier et al., 2019).

## S3   Evaluation of model against in situ measurements

To evaluate the model, in situ measurements at different stations are compared to near-surface concentrations of the a priori and optimized model. The median ratios of model and observations, which are shown in Fig. 10 of the main text, are provided in Table S1. Ratios closer to a value of 1 imply improvement of the model. Of the four stations with temporally collocated data (LAH, PEY, PAB, and TAR), we find significant improvement of the model at the two French stations in La Tardière and Peyrusse Vieille. At San Pablo de los Montes in Spain, the HCHO concentrations of the optimized model are too high. At the other in situ stations, climatological averages are used for the comparison. The inversion shows mixed results at these sites.

**Table S1.** Ratios of the modeled near-surface HCHO concentrations and the measured in situ concentrations. The ratio is calculated as the median of model over measurement. The locations of the stations are shown in Figs. 2 and 11, and correspond to the in situ stations presented in Table 1 of the paper. The upper four stations (LAH, PEY, PAB, and TAR) have temporally collocated measurements, hence provide the most reliable validation of the model.

| Station | code | mean concentration ($\mu g\,m^{-3}$) | a priori ratio | optimized ratio |
|---|---|---|---|---|
| Lahemaa | LAH | 1.45 | 0.61 | 0.64 |
| Peyrusse Vieille | PEY | 2.04 | 0.55 | 0.89 |
| San Pablo de los Montes | PAB | 1.30 | 1.11 | 1.58 |
| La Tardière | TAR | 1.17 | 0.83 | 1.00 |
| Birkenes | BIR | 0.58 | 1.28 | 1.23 |
| Brotjacklriegel | BRO | 0.56 | 2.11 | 2.53 |
| Campisábalos | CAM | 0.62 | 2.00 | 2.86 |
| Donon | DON | 1.41 | 1.02 | 1.20 |
| Ispra | ISP | 2.75 | 0.74 | 1.10 |
| Košetice | KOS | 1.19 | 1.14 | 1.13 |
| Montelibretti | MON | 2.33 | 0.76 | 1.24 |
| Schmücke | SCH | 1.24 | 1.01 | 0.99 |
| Waldhof | WAL | 1.38 | 0.92 | 0.80 |
| Zingst | ZIN | 1.13 | 0.93 | 0.90 |

## S4    Sensitivity runs

The optimized isoprene emissions of the different sensitivity runs are shown in Fig. S5. Figure S6 shows the scaling factors of the biogenic emissions derived from the sensitivity inversions described in Sect. 6. Most distributions do not show large differences compared to the standard run (see Fig. 6). The most notable differences occur for the S1 sensitivity run, which ignores the bias correction of the TROPOMI columns. For this scenario, the lower HCHO columns result in much lower optimized isoprene emissions, as described in Table 3. For the S10 run, which uses monthly averages in the optimization, the emissions in northern Europe are much lower compared to the other sensitivity runs. This is due to the longer time averages leading to less filtering of low HCHO columns, and thus lower HCHO columns on average. The lower isoprene emissions in northern Europe only have a minor impact on the total emission over the domain for the S10 run, however, since biogenic emissions are mainly located in southern Europe.
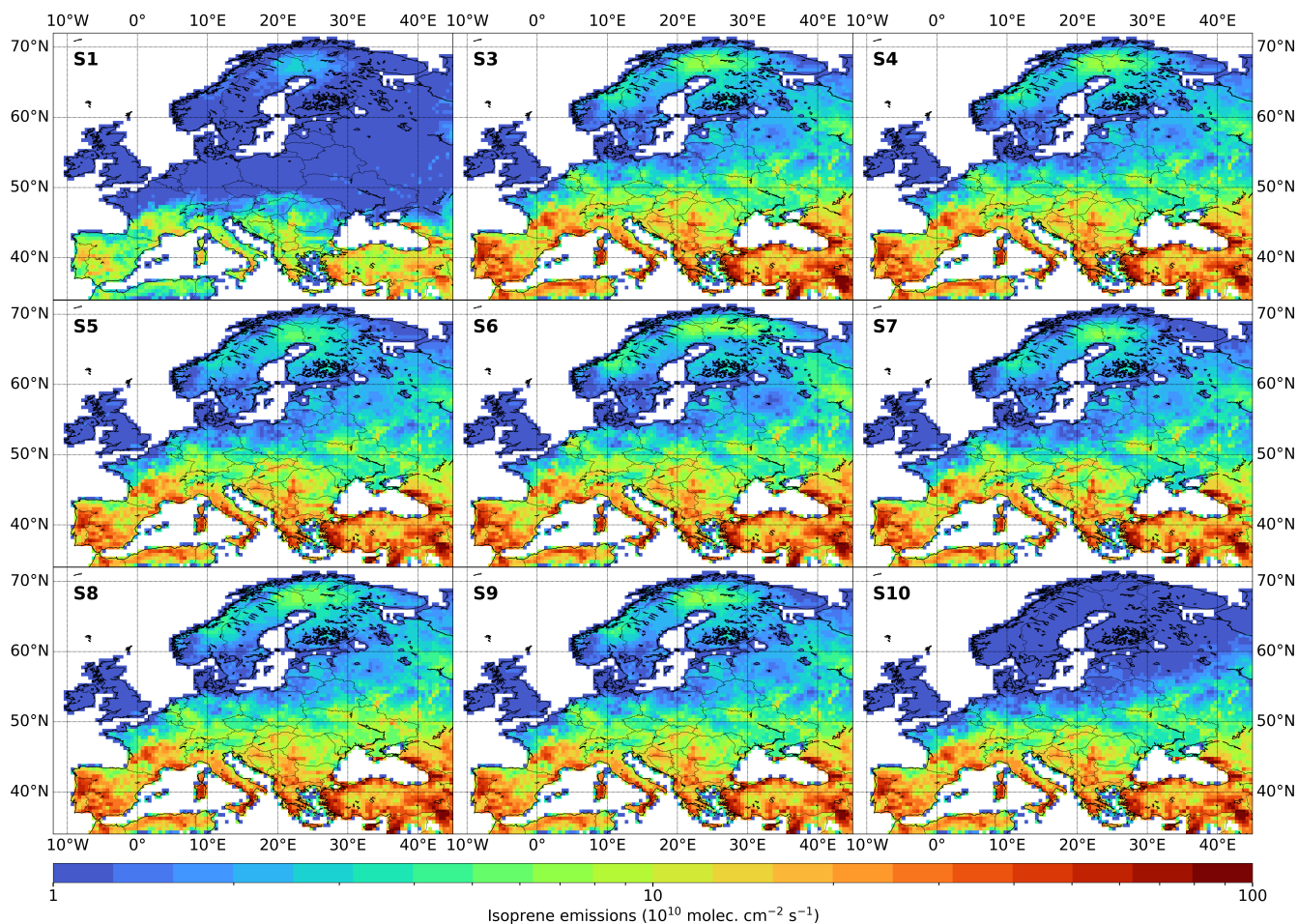
**Figure S5.** Optimized isoprene emissions of the sensitivity runs discussed in Sect. 6 and described in Table 3. The enhancement map for the S2 run is shown in Fig. 11c.
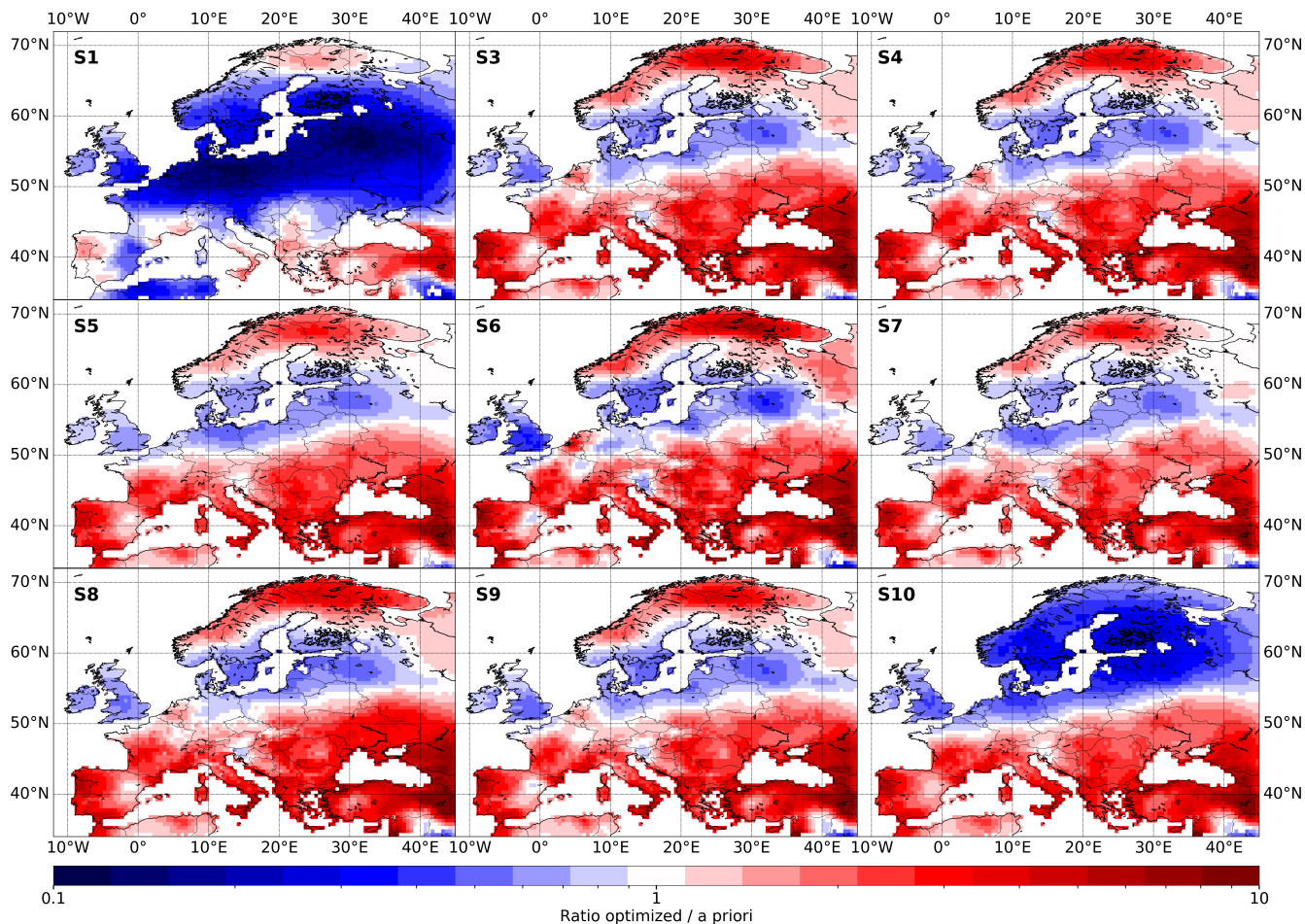
**Figure S6.** Emission enhancement maps for isoprene emissions in the sensitivity runs discussed in Sect. 6 and described in Table 3. The emission enhancements are calculated as the optimized emissions (Fig. S5) divided by the a priori biogenic emissions (from MEGAN-MOHYCAN, Stavrakou et al., 2018). The enhancement map for the S2 run is shown in Fig. 11c.

## References

Andreae, M. O.: Emission of trace gases and aerosols from biomass burning - an updated assessment, Atmos. Chem. Phys., 19, 8523–8546, https://doi.org/10.5194/acp-19-8523-2019, 2019.

Darmenov, A. and da Silva, A.: The Quick Fire Emissions Dataset (QFED): Documentation of versions 2.1, 2.2 and 2.4, NASA Technical Report Series on Global Modeling and Data Assimilation NASA TM-2015-104606, 38, 2015.

Granier, C., Darras, S., Denier Van Der Gon, H., Doubalova, J., Elguindi, N., Galle, B., Gauss, M., Guevara, M., Jalkanen, J. P., Kuenen, J. J. P., Liousse, C., Quack, B., Simpson, D., and Sindelarova, K.: The Copernicus Atmosphere Monitoring Service global and regional emissions (April 2019 version), Copernicus Atmosphere Monitoring Service (CAMS) report, https://doi.org/10.24380/d0bn-kx16, 2019.

Stavrakou, T., Müller, J. F., Bauwens, M., De Smedt, I., Van Roozendael, M., and Guenther, A.: Impact of Short-Term Climate Variability on Volatile Organic Compounds Emissions Assessed Using OMI Satellite Formaldehyde Observations, Geophys. Res. Lett., 45, 8681–8689, https://doi.org/10.1029/2018GL078676, 2018.