Supporting Information for

Sources of Formaldehyde in U.S. Oil and Gas Production Regions

Barbara Dix[*,1], Meng Li[1,2], Esther Roosenbrand[1,3,+], Colby Francoeur[1,2,4], Steven S. Brown[2,5], Jessica B. Gilman[2], Thomas F. Hanisco[6], Frank Keutsch[7,#], Abigail Koss[8,##], Brian M. Lerner[9,##], Jeff Peischl[1,2], James M. Roberts[2], Thomas B. Ryerson[10,###], Jason M. St. Clair[6,11], Patrick R. Veres[12,##], Carsten Warneke[1,2], Robert J. Wild[13,##], Glenn M. Wolfe[6], Bin Yuan[14,15##], J. Pepijn Veefkind[3,16], Pieternel F. Levelt[17,16,3], Brian C. McDonald[2], and Joost de Gouw[1,5]*

1. Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, 80309, United States

2. NOAA Chemical Sciences Laboratory, Boulder, Colorado, 80305, United States

3. Department of Civil Engineering and Geosciences, Technical University of Delft, 2628 CN Delft, Netherlands

4. Department of Mechanical Engineering, University of Colorado, Boulder, Colorado, 80309, United States

5. Department of Chemistry, University of Colorado, Boulder, Colorado, 80309, United States

6. Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, United States

7. Paulson School of Engineering and Applied Science and Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, United States

8. TOFWERK USA, Boulder, CO 80301, United States

9. Aerodyne Research, Billerica, MA 01821, United States

10. ChampionX Emissions Technologies, The Woodlands, TX 77381, United States

11. Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, MD, 21228, United States

12. Research Aviation Facility, NCAR Earth Observing Laboratory, Boulder, Colorado, 80307, United States

13. Institute for Ion Physics and Applied Physics, University of Innsbruck, 6020 Innsbruck, Austria

14. Institute of Environmental and Climate Research, Jinan University, Guangzhou 511443, China

15. Guangdong-Hongkong-Macau Joint Laboratory of Collaborative Innovation for Environmental Quality, Guangzhou 511443, China

16. Royal Netherlands Meteorological Institute, 3731 GA De Bilt, Netherlands

17. NCAR Atmospheric Chemistry Observations & Modeling Lab, Boulder, Colorado, 80307, United States

[+] now at Faculty of Aerospace Engineering, Technical University of Delft, 2628 CN Delft, Netherlands

was at Department of Chemistry, University of Wisconsin, Madison, Wisconsin, 53706, United States, during SONGNEX

was at 1. and 2. during SONGNEX

was at 2. during SONGNEX

Summary of Supporting Information:

S1. Calculation of primary formaldehyde emissions from flaring

S2. Box model simulation of formaldehyde abundance from primary emissions

Figures S1-S7

S1 Calculation of primary formaldehyde emissions from flaring

In order to estimate primary formaldehyde emissions from flaring, we first calculate flare counts weighted by their detection frequency and number of days observed from VIIRS data, to get the annual total of continuously lit flares in 2018 for the Permian basin study area, which is 168.3. The actual number could be lower or higher, depending on whether flares were lit or unlit on days without valid satellite observations.

Formaldehyde emissions for flares were reported by Pikelnaya et al. (2013)[1] based on measurements on petrochemical flares in the Houston-Galveston area in Texas. Since flared volumes and emissions can vary strongly in size, we take the median of their exponentially distributed reported observations, which is 0.79 kg h^{-1} per flare, to define the upper limit of our emission estimate for the Permian basin.

The VIIRS satellite flaring data set also provides flared gas volumes. A comparison of flared gas volumes of the exclusively upstream flares in the Permian basin, with downstream flares in the Houston-Galveston area, shows that the median upstream volumes are smaller by a factor of 0.44 compared to the downstream flares. Downscaling the upper limit by 0.44 provides the lower limit of 0.35 kg h^{-1} per flare.

## S2 Box model simulation of formaldehyde abundance from primary emissions

In order to calculate formaldehyde columns from our scaled emissions within the Permian basin study area, we treat the study area as a 2-dimensional box and apply a simple box model calculation as follows:

$$VCD_{avg} = (EM_{avg} / k) * (1-e^{(-k*t)})$$

Where VCD_{avg} is the average formaldehyde vertical column density within the box, EM_{avg} are the averaged primary formaldehyde emissions within the box, and k is the loss rate, which consists of chemical removal (assumed chemical life time) and flow out of the box (constrained by ECMWF winds). Figure S7 shows the VCD evolution over time, t. Steady state conditions are typically reached after 6-12 hours and later for winter, due to the long chemical lifetime. For this study we report the values at hour 23 in Table 2 of the main text. Note that the box model does not resolve gradients within the box, but it is well suited to provide an order of magnitude information based on box averages.

References

(1) Pikelnaya, O.; Flynn, J. H.; Tsai, C.; Stutz, J. Imaging DOAS Detection of Primary Formaldehyde and Sulfur Dioxide Emissions from Petrochemical Flares. *Journal of Geophysical Research: Atmospheres* **2013**, *118* (15), 8716–8728. https://doi.org/10.1002/jgrd.50643.
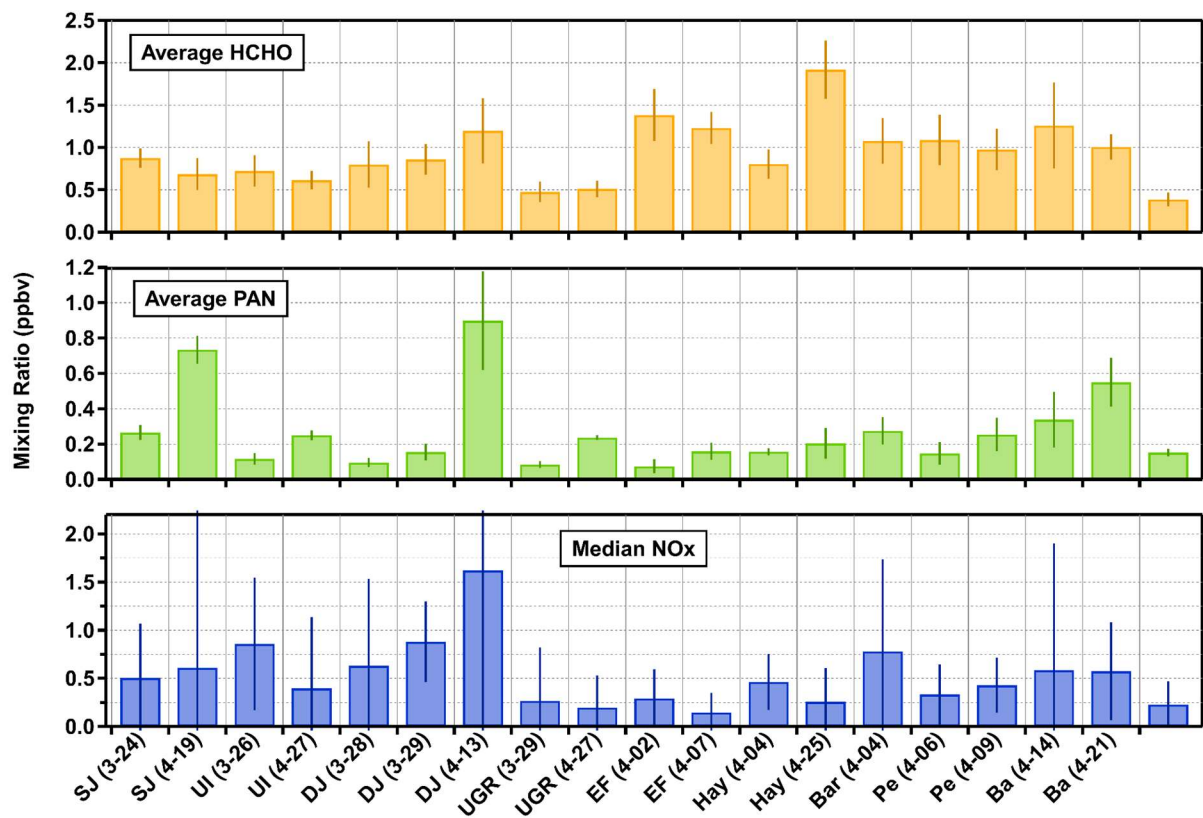
Figure S1: Average formaldehyde and PAN and median NO_x with standard deviations for each SONGNEX flight. Given the large variability in NO_x, the median is plotted here.
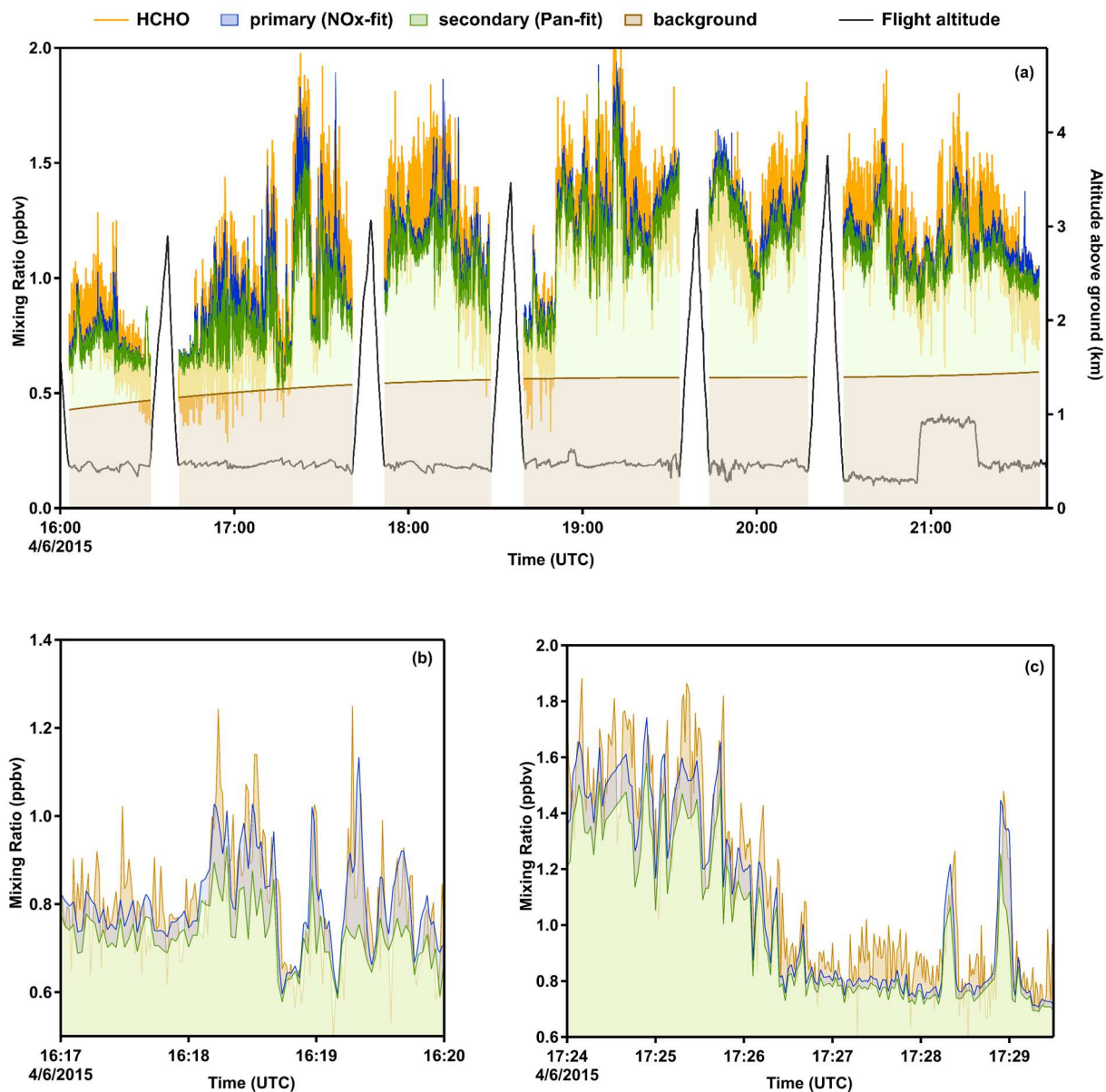
Figure S2: Example of a multivariate fit result, shown for the (a) Permian basin flight on 6 April, 2015 and (b-c) exemplary zooms to illustrate the varying correlation of formaldehyde with NO_x and PAN. The measured formaldehyde is simulated by a fitted background (5th order polynomial) and fitted fractions of the measured PAN and NO_x data.
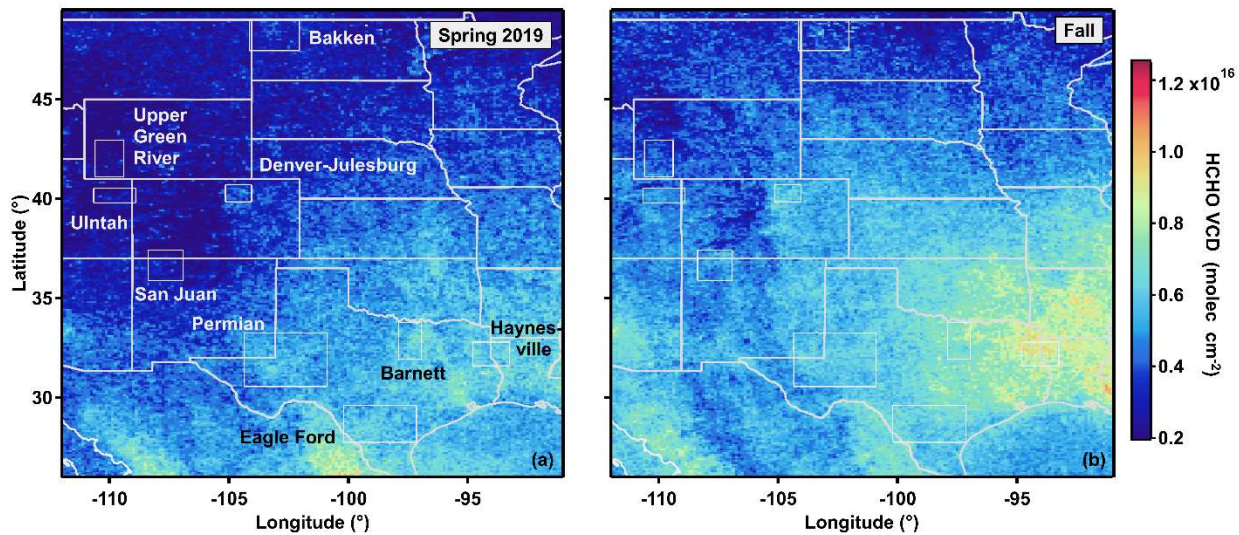
Figure S3: Average TROPOMI formaldehyde VCDs for (a) spring 2019 (March to May 2019) and (b) fall (September to November 2018 and 2019). Outlined boxes denote the oil and gas production areas probed by the SONGNEX campaign. Note that spring data for 2018 is not available from TROPOMI.
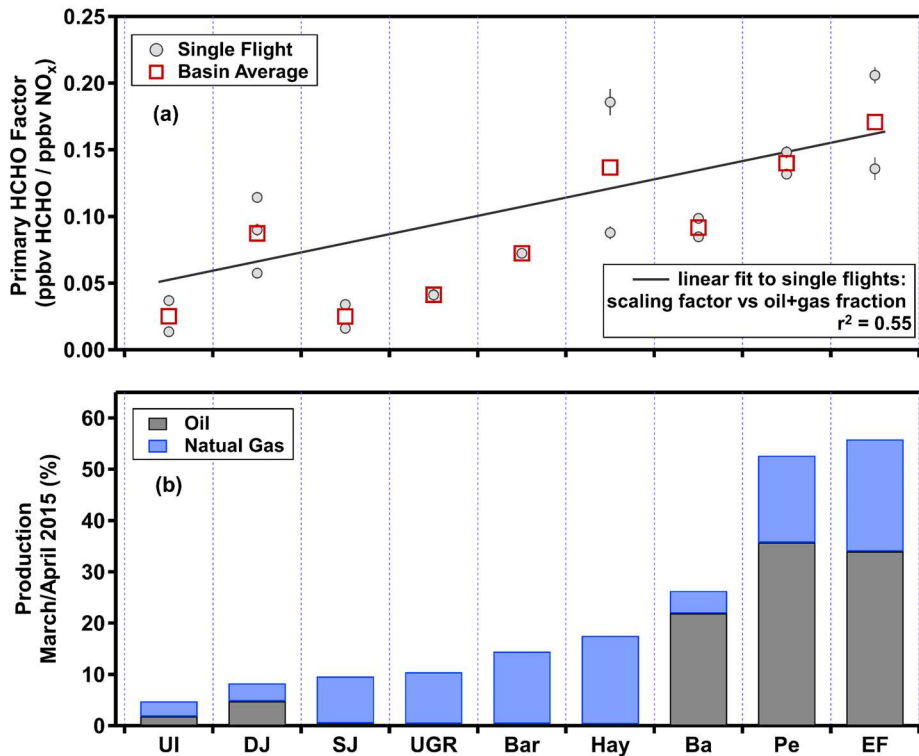
Figure S4: (a) Basin averages of the multivariate NO_x fit coefficients that are used to scale FOG NO_x emission to primary HCHO emissions. (b) Fractional oil and natural gas production volumes for each basin. These fractions are calculated so that the sum of production volumes over all basins equals 100% for oil and gas, respectively. The line fit in panel (a) denotes a linear fit to the single flight factors over the combined oil and gas fraction shown in panel (b). There is some correlation between primary emissions and the amount of oil and gas produced in each basin with an r^2 of 0.55. This correlation might reflect the balance between production increases that are accompanied by growing infrastructure and production increases that are leveraged by existing infrastructure.
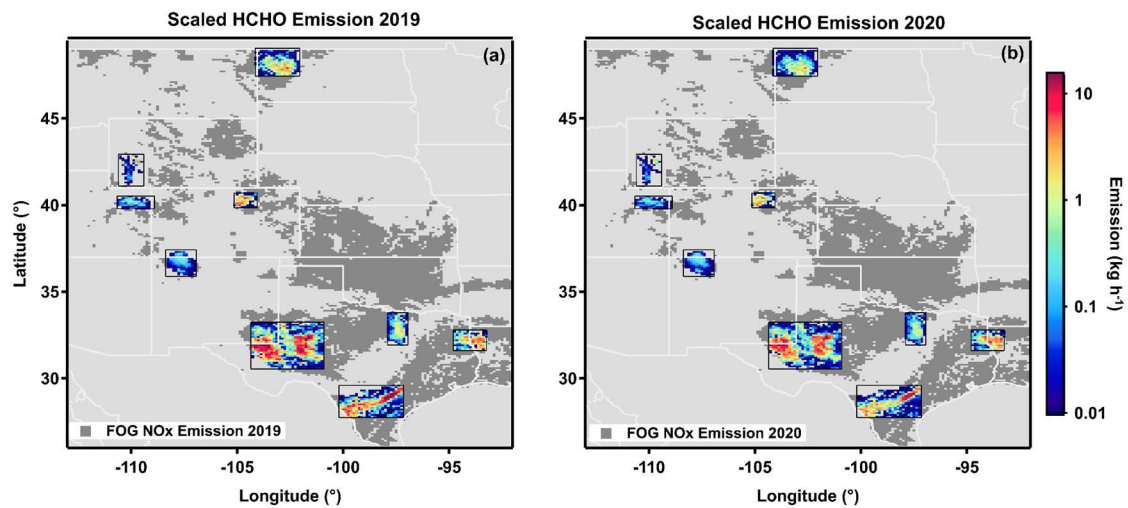
Figure S5: Primary formaldehyde emissions scaled from FOG NO$_x$ emissions for the years (a) 2019 and (b) 2020. Decreased emissions visible in some basins in 2020 compared to prior years are caused by the economic impact of the COVID-19 shutdowns. Note that the color scale is logarithmic.
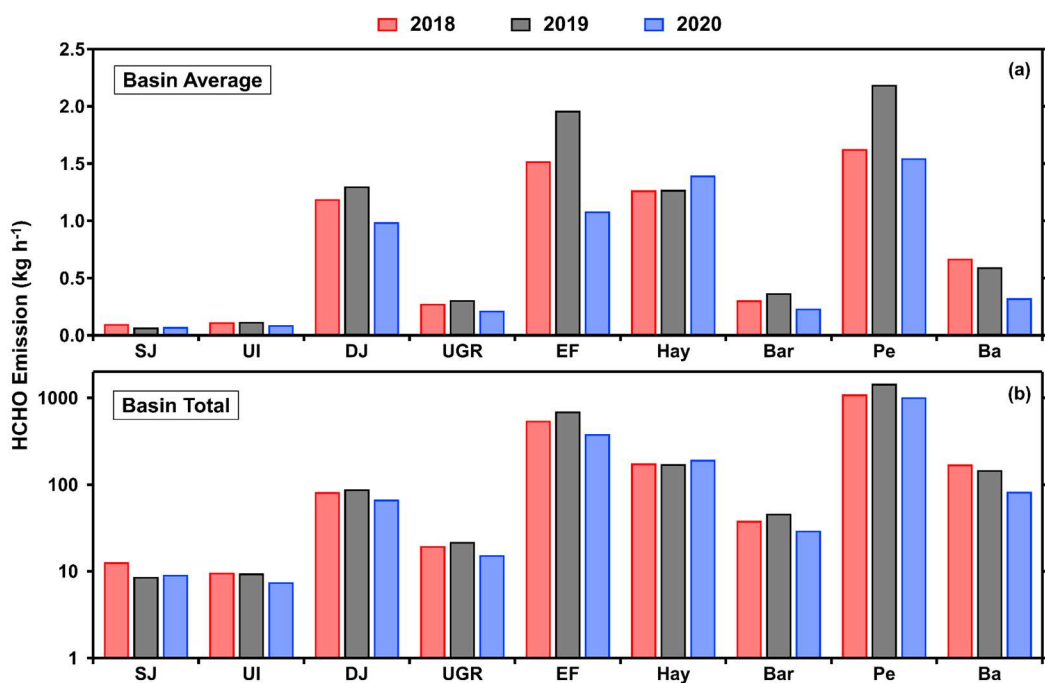
Figure S6: Primary formaldehyde emissions scaled from FOG NO$_x$ emissions for the years 2018, 2019 and 2020. Shown here are (a) basin averages and (b) totals for the SONGNEX study areas as outlined in Figs. 1 and S5. Decreased emissions in some basins in 2020 compared to prior years are caused by the economic impact of the COVID-19 shutdowns. Note the logarithmic scale in panel b.
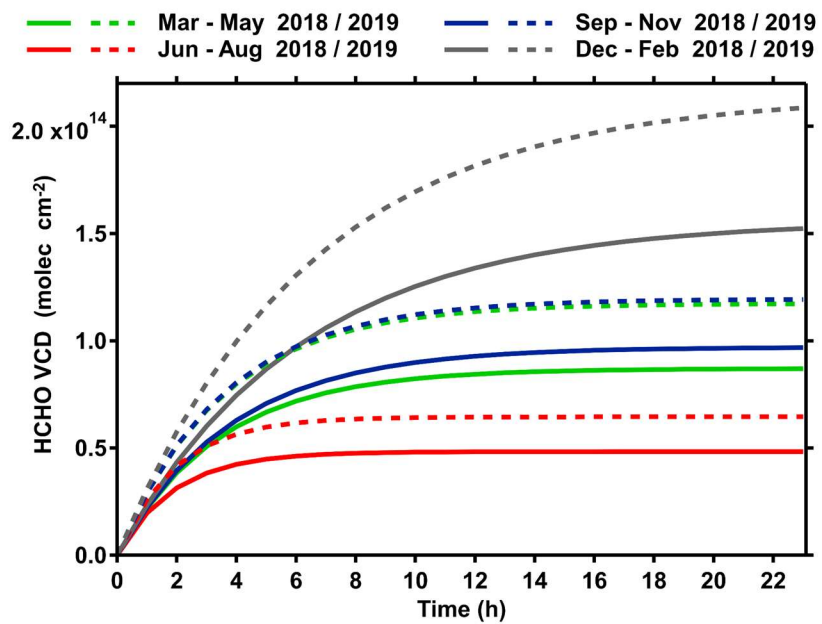
Figure S7: Box model results. The formaldehyde vertical columns are simulated as described in Section S2. Values reported in Table 2 in the main text are taken at hour 23.