Geophysical Research Letters

Supporting Information for

Trends and interannual variability of the hydroxyl radical in the remote tropics during boreal autumn inferred from satellite proxy data

Daniel C. Anderson[1,2], Bryan N. Duncan[2], Junhua Liu[2,3], Julie M. Nicely[2,4], Sarah A. Strode[2,3], Melanie B. Follette-Cook[5], Amir H. Souri[2,3], Jerry R. Ziemke[2,3], Gonzalo González-Abad[6], Zolal Ayazapour[6,7]

[1]GESTAR II, University of Maryland Baltimore County, Baltimore, MD, USA, [2]Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [3]GESTAR II, Morgan State University, Baltimore, MD, USA, [4]Earth System Science Interdisciplinary Center, University of Maryland College Park, College Park, MD, USA, [5]Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, [6]Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA, [7]Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY, USA.

Contents of this file

Introduction

This supporting information contains a brief description of changes to the TCOH dataset due to incorporation of the most recent OMI formaldehyde retrieval. This section also contains four supporting figures for the main text as well as a supporting table, which outlines the satellite retrievals used in the TCOH product.

Text S1.

As discussed in the main text, we have updated the TCOH product described in Anderson et al., 2023 to use the v4 OMI HCHO retrieval. Major differences from the v3 HCHO retrieval include the use of updated Level 1B radiances (Kleipool et al., 2022), improved air mass factor calculations similar to the method applied to OMPS (Ozone Mapping and Profiler Suite) (Nowlan et al., 2023), and a new reference sector correction methodology. In agreement with the OMPS HCHO retrieval, the v4 SAO OMI retrieval shows little trend in the remote atmosphere over the 2005 – 2019 period, whereas the v3 SAO OMI retrieval has a distinct, increasing trend that is likely unphysical starting around 2015. This trend in HCHO propagates through to TCOH, necessitating the need to use the updated retrieval. Because the v4 HCHO is systematically lower than the v3 product and the MERRA2 GMI output is biased high with respect to OMI data, a higher frequency of HCHO data from the v4 retrieval are outside the bounds of the GBRT model training dataset. We exclude these data from the TCOH calculation to prevent extrapolation. The spatial extent of these low values, however, is primarily limited to one region off the western coast of South America and only reduces the number of valid TCOH datapoints by less than 5% per year.

Text S2.

We use the MERRA2-GMI simulation to create the training dataset for the GBRT model. We create a separate training dataset and GBRT model for each month of the year, following the results of Anderson et al, 2022. For a given month, we use instantaneous output from MERRA2-GMI at 14:00 LST (Local Solar Time) for all GBRT input variables, except CO, to coincide with the overpass time of the corresponding satellite retrieval. For CO, whose corresponding satellite has a morning overpass, we use output at 10:00 LST. For each day, we filter the data to include only those grid boxes with cloud fraction less than 30%, to align the data more closely with satellite retrievals. This results in approximately 43,000 valid grid boxes per day. For each year, we then average the data to the monthly resolution, resulting in approximately 600,000 total training values for each GBRT model input. For solar zenith angle, we use the noontime value for the 15th day of a given month, and sea surface temperatures are monthly averages of the 24-hour averaged means.

While other GBRT studies employing satellite data (e.g., Zhu et al 2022a) do convolve the training dataset with satellite averaging kernels and shape factors, as described in Anderson et al, 2023, we do not find that necessary here. To test the impact of omitting the averaging kernels, we trained a separate model for one month, where we convolved the MERRA2-GMI output with the corresponding averaging kernels from the satellite retrievals, finding that the resulting tropospheric column OH varied less than 1%, on average, from the GBRT model version omitting the averaging kernels. This difference is significantly smaller than the uncertainty in OH resulting from the uncertainties from the individual satellite retrievals. For regions where satellite retrievals have much smaller

random errors (e.g. over pollution source regions), it may become necessary to incorporate the averaging kernel information.

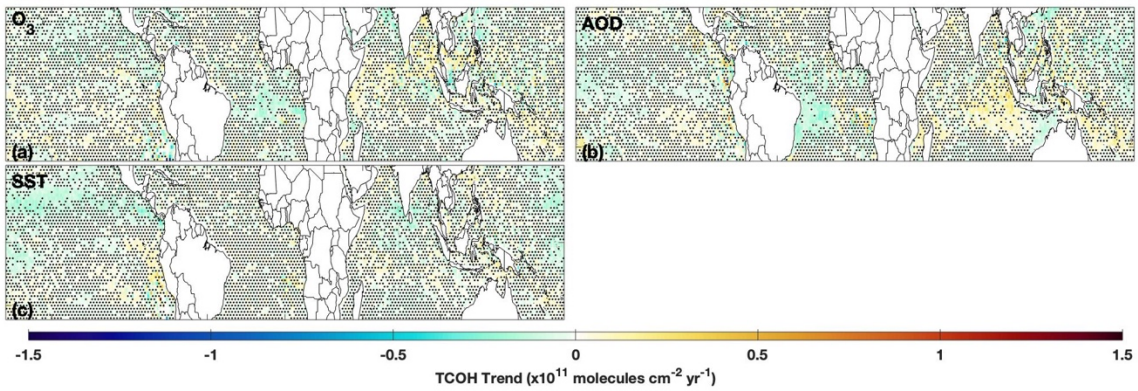

Figure S1. We show the inferred contribution of O_3 (**a**), AOD (**b**), and SSTs (**c**) to the total OH trend over 2005 – 2019. Stippling indicates regions where the inferred contribution to the trend is less than 25%.

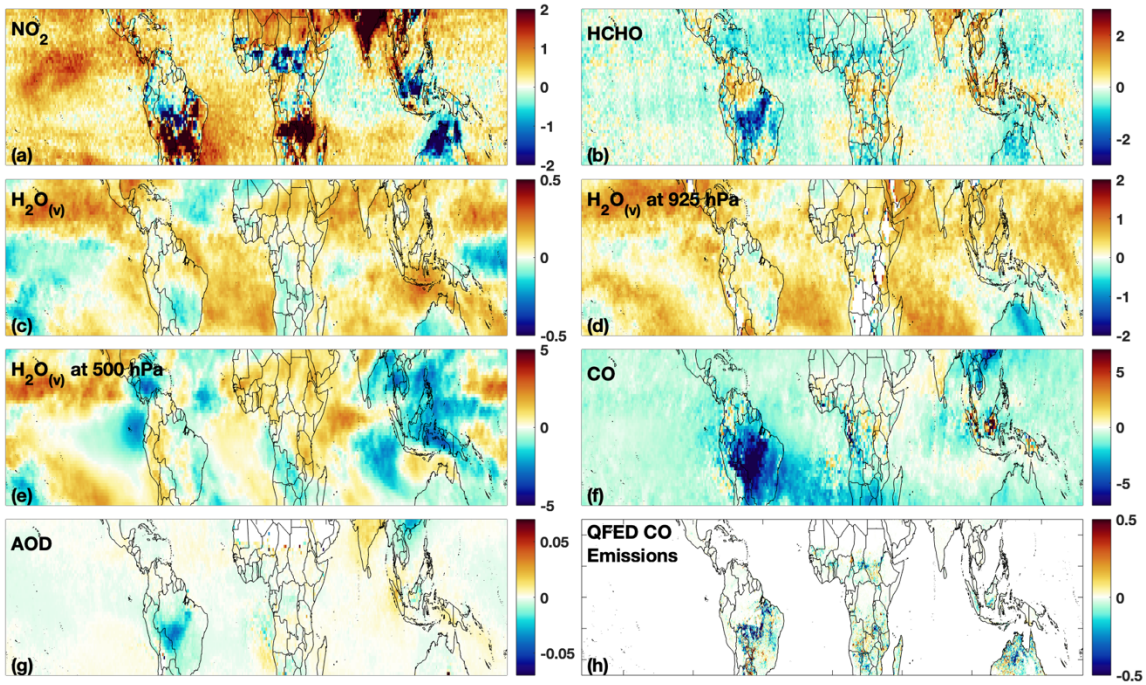

Figure S2. The trend over 2005 – 2019 for OMI NO_2 ($x10^{13}$ molecules cm^{-2} yr^{-1}) (**a**); OMI HCHO ($x10^{14}$ molecules cm^{-2} yr^{-1}) (**b**); AIRS $H_2O_{(v)}$ for the column (kg m^{-2}) (**c**), 925 hPa layer (g kg^{-1}) (**d**), and 500 hPa layer (g kg^{-1}) (**e**); MOPITT CO ($x10^{16}$ molecules cm^{-2} yr^{-1}) (**f**); MODIS AOD (**g**), and QFED CO emissions ($x10^{-8}$ kg m^{-2} s^{-1} yr^{-1}) (**h**). Note that for OMI NO_2 many of the trends over land extend beyond the range of the colorbar. See Figure 3 for a clearer representation of these values.

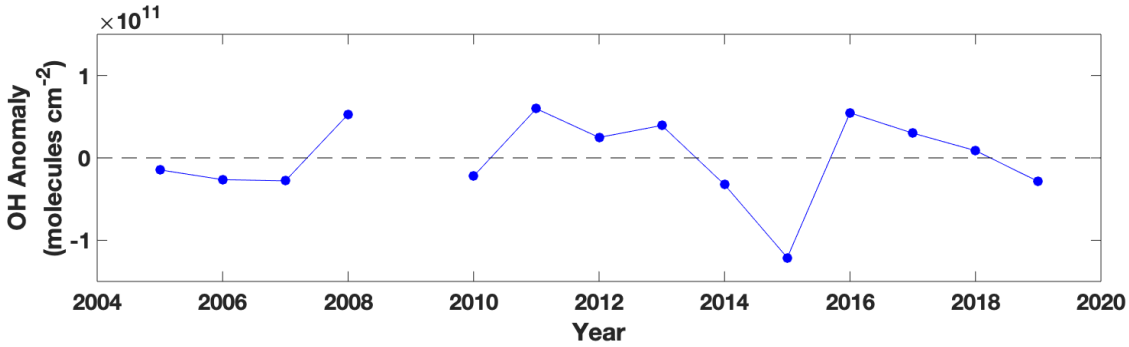

Figure S3. The domain-wide, detrended OH anomaly averaged as a function of time.

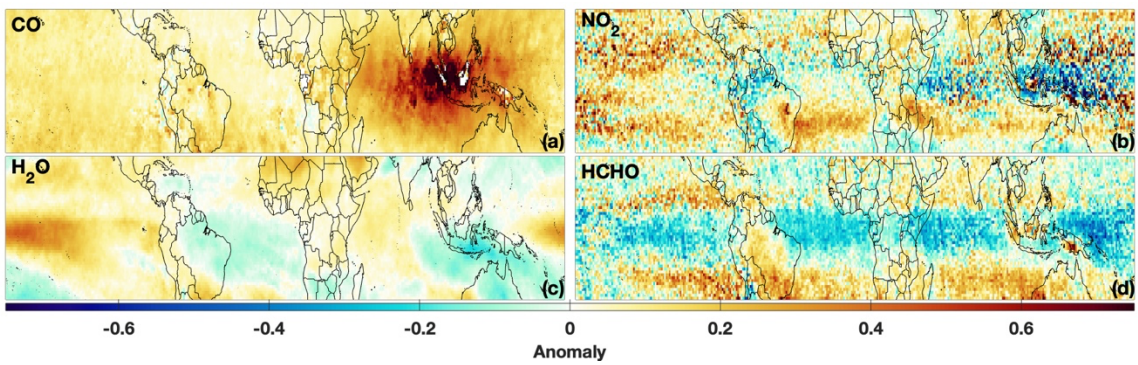

Figure S4. The detrended anomaly for MOPITT CO (**a**), OMI NO_2 (**b**), AIRS $H_2O_{(v)}$ column (**c**), and OMI HCHO (**d**) for 2015. The anomaly has been normalized by the climatological mean of the indicated species.

Variable	Satellite retrieval	Original horizontal and temporal resolution
Total O_3 column	OMI TOMS-Like L3 version 3	0.25° × 0.25°, daily
Tropospheric NO_2 column	OMI GSFC L3 version 4	0.25° × 0.25°, daily
CO column	MOPITT L3 version 8	1.0° × 1.0°, monthly
HCHO column	OMI SAO L3 version 4	0.1° × 0.1°, daily
$H_2O_{(v)}$ column	AIRS L3 version 6	1.0° × 1.0°, monthly
Sea surface temperature	MUR L4 version 4.2	0.25° × 0.25°, daily
Aerosol optical depth at 550 nm	MODIS Aqua L3 collection 6	0.5° × 0.5°, daily
$H_2O_{(v)}$ layers: 925 – 850 hPa, 850 – 700 hPa, 700 – 600 hPa, 600 – 500 hPa, 500 – 400 hPa, 400 – 300 hPa, and 300 - 250 hPa	AIRS L3 version 6	1.0° × 1.0°, monthly

Table S1. The satellite inputs to the GBRT model, the satellite retrieval used, and the original horizontal and temporal resolution of the retrieval. Latitude and solar zenith angle are also inputs to the GBRT model.