Supplementary Material for "Quantifying the tropospheric ozone radiative effect and its temporal evolution in the satellite-era"

Richard J. Pope[1,2], Alex Rap[1], Matilda A. Pimlott[1], Brice Barret[3], Eric Le Flochmoen[3], Brian J. Kerridge[4,5], Richard Siddans[4,5], Barry G. Latter[4,5], Lucy J. Ventress[4,5], Anne Boynard[6,7], Christian Retscher[8], Wuhu Feng[1,9], Richard Rigby[1,10], Sandip S. Dhomse[1,2], Catherine Wespes[11] and Martyn P. Chipperfield[1,2]

Supplementary Material (SM) 1: Ozonesonde Bias Correction Factors

In this study, we used ozonesonde from 236 sites (**Figure S1**) from the WOUDC, SHADOZ and NOAA networks, which typically yielded several thousand sondes annually. Globally, the peak frequency of sondes annually was between approximately 6000 (substantial frequency drop in 2002) and 10,000, which were available between 1998 and 2013. The SHADOZ network starts in 1998 leading to the substantial increase in sonde frequency when compared with 1995 to 1997. After 2013, the sonde frequency decreases to between 1000 and 4000. On a hemispheric basis, there are more sonde sites in the northern hemisphere (NH). Between 1995 and 2019, the southern hemisphere (SH) sonde frequency peaks above 2000, while the NH frequency is more variable and ranges between 4000 and 7000 during 1998 to 2013. In other years of the record, the NH sonde frequency ranges between 1000 and 2000. An important point to note is that post-2013 there was a "drop-off" in absolute O_3 measurements from a third of ozonesondes, potentially linked to manufacturing issues or hardware changes, of 3-7% (Stauffer et al., 2020).

For comparisons with ozonesonde profiles, the satellite retrievals are spatiotemporally co-located within 500 km and 6 hours to allow for robust comparisons and reduce representation errors. Here, O_3 measurements were rejected if the O_3 or pressure values were unphysical (i.e. < 0.0), if the O_3 partial pressure > 2000.0 or the O_3 value was set to 99.9, and whole ozonesonde profiles were rejected if least 50% of the measurements did not meet these criteria. These criteria are similar, though not as extensive, to those applied by Keppins et al., (2018) and Hubert et al., (2016). To allow for direct like-for-like comparisons between the two quantities, accounting for the vertical sensitivity of the satellite, the instrument AKs are applied the ozonesonde profiles. Firstly, a co-located ozonesonde profile (in volume mixing ratio) were interpolated onto the satellite pressure grid in $\log(pressure)$. The sonde sub-columns were then derived using the hydrostatic balance approach:

$$mass\ density = mmr.\rho dz = mmr.\frac{-dp}{g} \qquad (S1)$$

where *mass density* is mass (kg) of O_3 per m^2 between two pressure levels, *mmr* is the O_3 mass mixing ratio from the sonde, ρ is the density (kg/m^3), *dz* is the distance between pressure levels (m), *dp* is the pressure difference between levels (Pa) and *g* is the acceleration due to gravity (-9.81 m/s^2). The application of the AKs from the three IASI products onto the ozonesondes is:

$$sonde_{AK} = AK.(sonde_{int} - apr) + apr \qquad (S2)$$

where ***sonde_AK*** is the modified ozonesonde sub-column profile, ***AK*** is the averaging kernel matrix, ***sonde_int*** is the sonde sub-column profile on the satellite pressure grid and ***apr*** is the apriori. For the application of the AKs to the ozonesonde profiles, the full ozone profile is required which is not available from the ozonesondes. Therefore, the ozonesonde profile above its minimum pressure level is extended using the apriori profile from the corresponding satellite product. The profile is smoothed vertically across the joining pressure level to avoid a profile discontinuity.

To derive tropospheric column ozone (TCO$_3$), we used the World Meteorological Organisation (WMO) lapse-rate tropopause definition of the lowest level where the temperature gradient is less than 2°C/km (WMO, 1957). Here, all the satellite and ozonesonde sub-columns are integrated from the surface up to the tropopause. Where the tropopause sits within the barrier pressure levels of a sub-column, the sub-column is sliced at the tropopause and the lower segment added to the integrated column.

Once the ozonesondes had been co-located with the satellite data and the AKs applied, the two datasets were compared across the full 2008-2017 period with the aim to generate annual-latitude bias correction factors (BCF) to harmonise the satellite records. Despite the all three products using the satellite IASI level-1 data, there differing retrieval schemes yielded sizable differences in the absolute TCO$_3$ values. Though, it is worth noting the spatial distributions are in good agreement. **Figure S2** shows the BCF for each product which have been duplicated longitudinally to provide a 2D map to impose on the gridded satellite TCO$_3$ data. The IASI-SOFRID has the best agreement with the ozonesondes (**Figure S2b**) where the mean bias (satellite-sonde$_{ak}$) ranges between -2.0 and -0.5 DU in the mid-latitudes and is near-zero in the tropics and southern high-latitudes. In the northern high-latitudes, the bias is small at less than 1.0 DU. For IASI-FOLRI (**Figure S2a**), the mean bias generally ranges between -3.0 and -1.0 DU with the peak underestimation of -5.0 DU in the southern mid-latitudes. In the southern high-latitudes, the mean bias tends towards zero. IASI-IMS overestimates the ozonesondes with positive biases of 1.0-3.0 DU in the tropics and mid-latitudes (**Figure S2c**). In the higher latitudes, then bias increases to 5.0-7.0 DU.

By subtracting these BCF from the respective satellite products, the TCO$_3$ records are harmonised in absolute terms to provide a more robust estimate of global TCO$_3$ values. While ozonesondes are used to ground true the satellite absolute values, the limited spatial coverage of the ozonesondes makes the bias corrected satellite products valuable resources to understand the spatial pattern of TCO$_3$ and its temporal evolution. As the BCFs are derived from the long-term (2008-2017) average, applying them to the individual years of the satellite data sets updates the absolute values but not the inter-annual variability of the records which remain independent of the ozonesondes.

While ozonesondes can be used to derive TCO$_3$ BCF, they cannot be used to derive profile BCFs. The IASI O$_3$ products tend to have approximately 1.0 degree of freedom of signal (DOFS) in the troposphere (i.e. one piece of information). For infrared instruments like IASI, this information peaks in the upper troposphere which is the most import region for the tropospheric ozone radiative effect (TO$_3$RE). Therefore, we can justify using the satellite data to derive the global TO$_3$RE (i.e. integrated effect over the upper troposphere) but the BCFs are applied at a tropospheric column level (i.e. the same percentage correction is applied at all levels within the troposphere). As a result, once the satellite TCO$_3$ and TO$_3$RE is derived, the BCFs are applied to the satellite data and then the factor difference between the original and corrected TCO$_3$ values is applied to the TO$_3$RE value (e.g. if the BCFs reduce the absolute TCO$_3$ value by 10% then the TO$_3$RE value is also reduced by 10%).

SM 2: TOMCAT Evaluation

We evaluate TOMCAT TO$_3$ using the IASI products and the ozonesondes. In **Figure S3**, TOMCAT TCO$_3$, averaged between 2008 and 2017, has had the satellite AKs applied using **Equation S2** and is directly comparable with the satellite TCO$_3$. Generally, TOMCAT+AKs is consistent with the IASI TCO$_3$ spatial distribution. In the NH, satellite TCO$_3$ typically ranges between 35.0 and 45.0 DU over the continental regions and decreases to 30.0-40.0 DU over the oceans. Here, TOMCAT tends to be

larger ranging from 40.0-50.0 and 35.0-45.0 DU over the continents and oceans, respectively. In the SH, IASI-SOFRID and IASI-IMS retrieve TCO_3 between 15.0 and 25.0 DU in the mid- and high-latitudes. IASI-FORLI/TOMCAT retrieves/simulates larger values between 20.0 and 30.0 DU in the same region. Over the central South Atlantic, the largest model-satellite discrepancy occurs where TOMCAT does not simulate the substantial TCO_3 plume from central/southern Africa. This is likely linked to biomass burning emissions used in the model and or the lighting parameterisation for NO_2 production (i.e. an ozone precursor gas).

The mean bias between TOMCAT and the satellite TCO_3 is similar for all the instruments. In the mid- and high-latitudes, TOMCAT overestimates TCO_3 by 2.0-7.0 DU. This is more prominent for the IASI-IMS comparisons where the overestimation peaks between 10.0 and 20.0 DU. However, as the IASI-IMS data is only sampled for 1 in 10 days and 1 in 4 pixels, the satellite sample sizes are lower potentially yielding a less robust observational constraint to compare against. In the high-latitudes for IASI-FORLI, the TOMCAT biases are moderately negative between -2.0 and -1.0 DU. The largest model-IASI-FORLI different is over India/central Asia where the biases peak at 10.0-15.0 DU. In the tropics, TOMCAT consistently underestimates TCO_3 with negative biases of -7.0 to -2.0 DU. This is most substantial over the central South Atlantic where TOMCAT understates the TCO_3 plume from Africa. For IASI-IMS, TOMCAT underestimates by more than 10.0 DU while in IASI-SOFRID and IASI-FORLI this underestimation is smaller at 3.0-8.0 DU. This underestimation is also similar to the surrounding tropical mean biases and not as pronounced as in the IASI-IMS comparisons.

When compared against the ozonesondes (**Figure S4**), the spatial mean biases are similar, and in absolute terms. In the mid- and high-latitudes, the TOMCAT-ozonesonde mean bias ranges between approximately 2.0 and 8.0 DU peaking at 10.0 DU in the SH between 60-90°S. In the tropics, the model underestimates with biases of -5.0 to 0.0 DU. Overall, TOMCAT-observational TCO_3 biases are consistent and generally between ±10 and 20%. The global mean bias ranges between -0.50 and 9.99%, while the root-mean-square-error is approximately 18-20% (approximately 30% for the IASI-IMS comparisons). There is reasonable to good spatial agreement in the TOMCAT and IASI TCO_3 spatial distributions with R^2 (i.e. correlation squared) ranging between 0.42 and 0.57. Overall, given the model simulates the satellite observed spatial distribution of TCO_3 with a small-moderate absolute mean bias in comparison to the observations, it provides confidence in our use of TOMCAT TO_3 to investigate the TO_3RE over the recent satellite-era (2008-2017).

References

Hubert D, et al. 2016. Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records. *Atmospheric Measurement Techniques*, **9**, 2497-2534, doi: 10.5194/amt-9-2497-2016.

Keppins A, et al. 2018. Quality assessment of the Ozone_cci Climate Research Data Package (release 2017) – Part 2: Ground-based validation of nadir ozone profile data products. *Atmospheric Measurement Techniques*, **11**, 3769-3800, doi: 10.5194/amt-11-3769-2018.

Stauffer RM, et al. 2020. A post-2013 Dropoff in Total Ozone at a Third of Global Ozonesonde Stations: Electrochemical Concentration Cell Instrument Artefacts?, *Geophysical Research Letters*, **47 (11)**, e2019GL086791, doi: 10.1029/2019GL086791.

WMO. World Meteorological Organization (WMO): Geneva, Switzerland, 1957.

Figures:
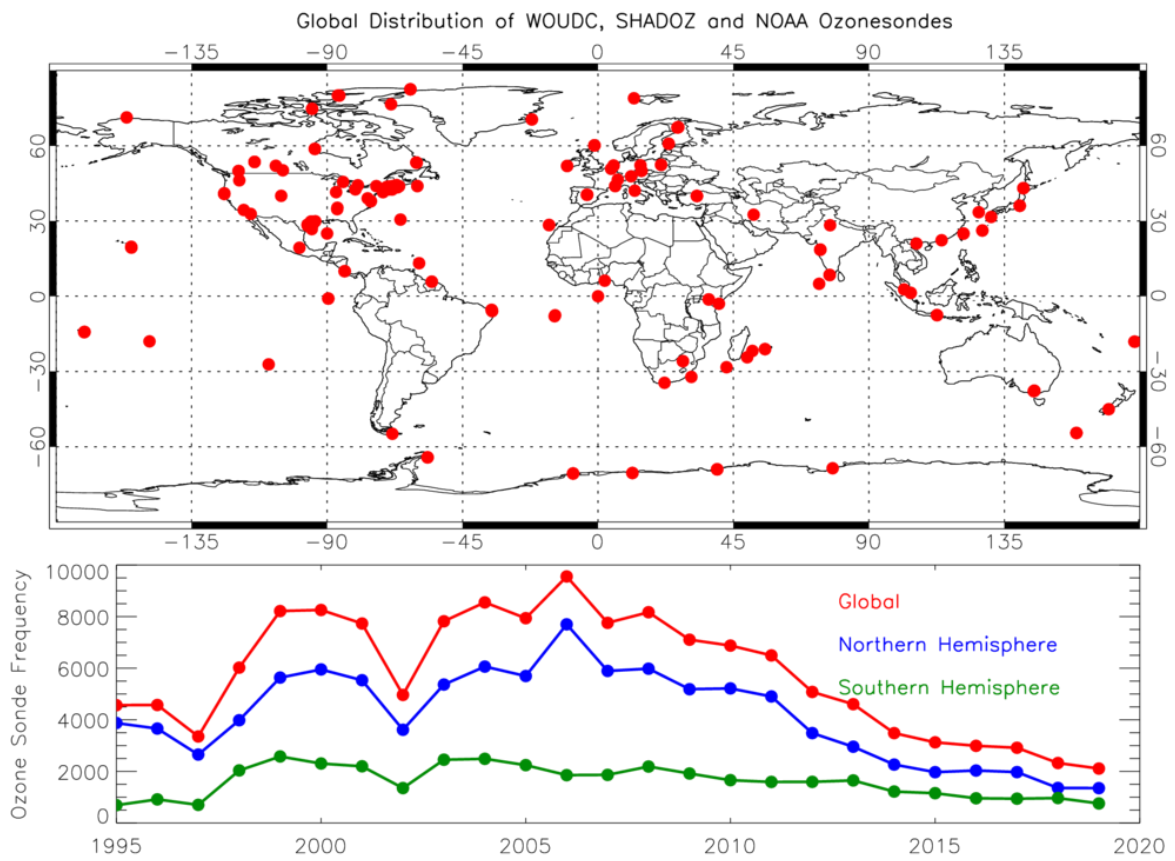
Figure S1: *Global distribution of WOUDC, SHADOZ and NOAA ozonesondes used in this study and time series of annual ozonesonde frequency (i.e. all sites and times) globally, in the northern hemisphere and in the southern hemisphere.*
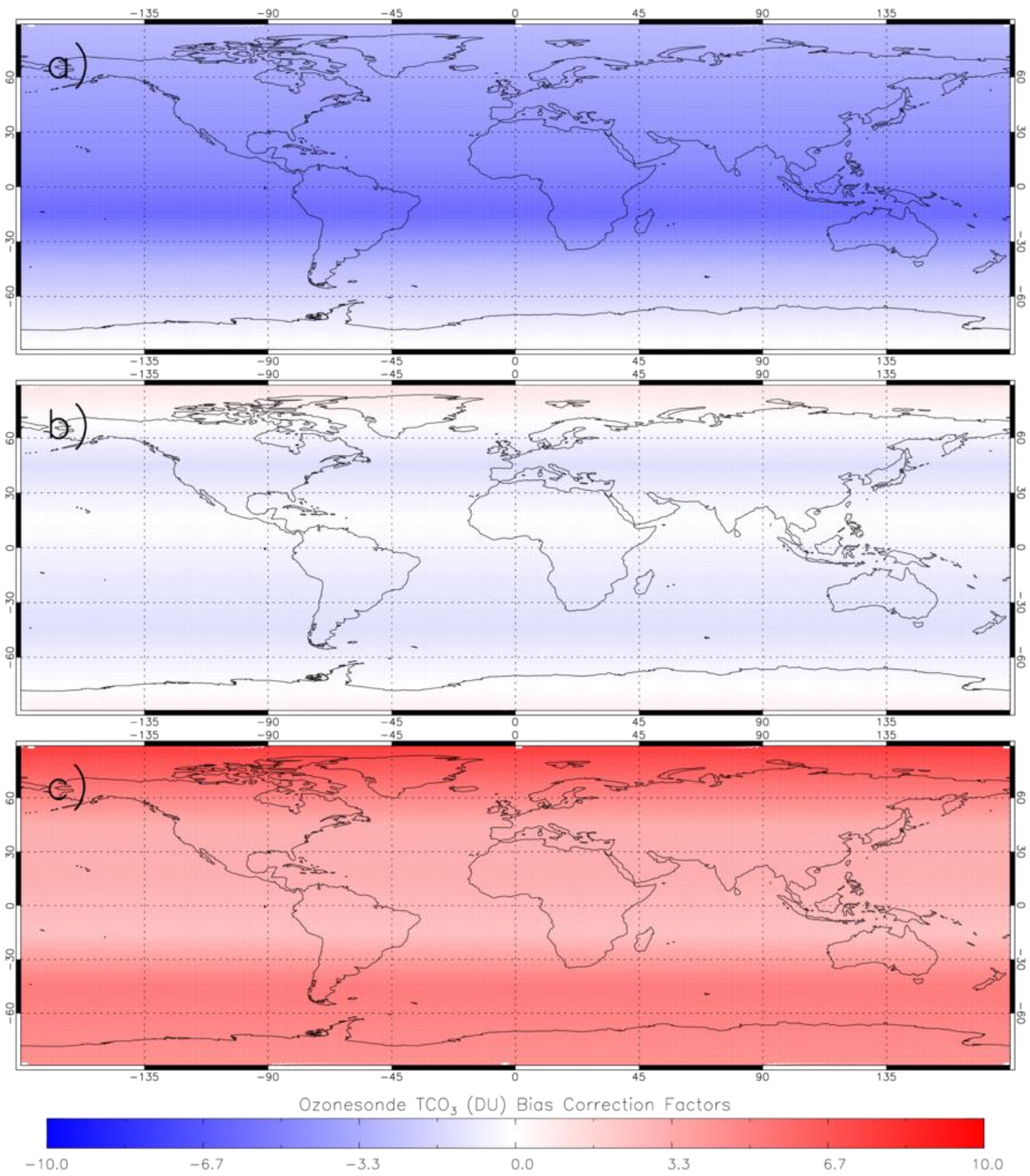
Figure S2: *Annual-latitude ozonesonde TCO$_3$ (DU) bias correction factors, duplicated zonally, for a) IASI-FORLI, b) IASI-SOFRID and c) IASI-IMS.*
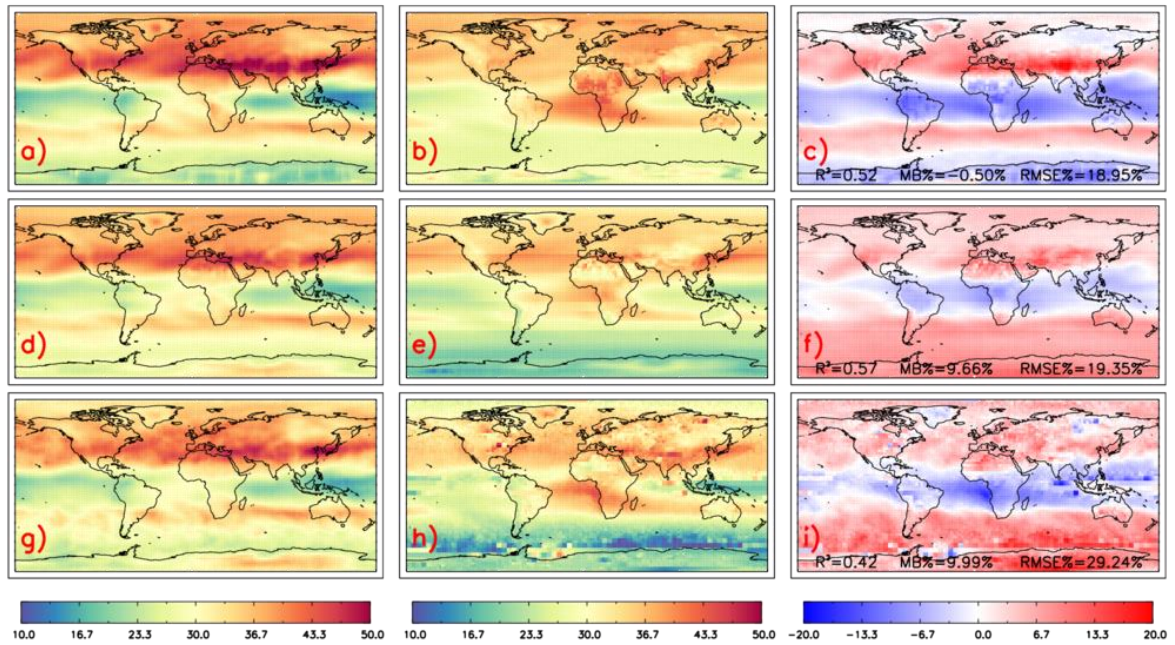
Figure S3: TCO$_3$ (DU), averaged between 2008 and 2017, for a) TOMCAT+IASI-FORLI AKs, b) IASI-FORLI, c) a)-b), d) TOMCAT+IASI-SOFRID AKs, e) IASI-SOFRID and f) d)-e), g) TOMCAT+IASI-IMS AKs, h) IASI-IMS, i) g)-h). The spatial mean bias (MB), root-mean-square-error (RMSE) and correlation squared (R^2) between TOMCAT and IASI are also shown.
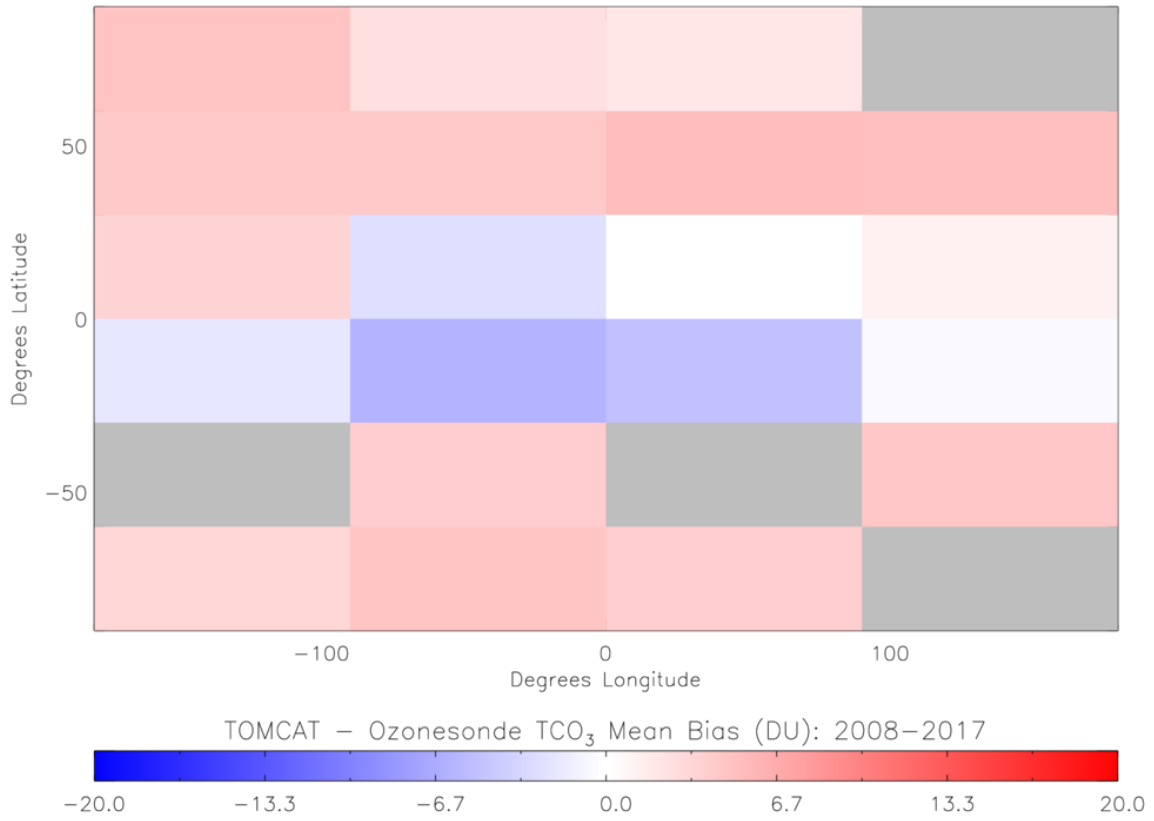
Figure S4: *Long-term (2008-2017) TCO$_3$ (DU) mean bias between TOMCAT and ozonesondes in 30° latitude and 90° longitude bins. Here, TOMCAT profiles have been spatiotemporally co-located (i.e. nearest model time-step and grid box) with individual ozonesondes to derive the TCO$_3$. Grey shading represents no ozonesonde data (and thus no model sub-sampled data).*