Supplementary Material for "Investigation of satellite vertical sensitivity on long-term retrieved tropospheric ozone trends"

Richard J. Pope[1,2], Fiona M. O'Connor[3,4], Mohit Dalvi[3], Brian J. Kerridge[5,6], Richard Siddans[5,6], Barry G. Latter[5,6], Brice Barret[7], Eric Le Flochmoen[7], Anne Boynard[8,9], Martyn P. Chipperfield[1,2], Wuhu Feng[1,10], Matilda A. Pimlott[1], Sandip S. Dhomse[1,2], Christian Retscher[11], Catherine Wespes[12] and Richard Rigby[1,13]

Supplementary Material (SM)-1: Ozonesondes and Averaging Kernels

We have used ozonesonde data between 2008 and 2017 from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC, https://woudc.org/), the Southern Hemisphere ADditional Ozonesondes (SHADOZ) project (https://tropo.gsfc.nasa.gov/shadoz/) and from the National Oceanic and Atmospheric Administration (NOAA, https://gml.noaa.gov/ozwv/ozsondes/). Here, a month-latitude long-term set of bias correction factors (BCF) has been generated in 30° latitude bins for 12-months (i.e. a climatological of monthly averages) over the record for each instrument. As satellite records can have systematic biases in column ozone (e.g. Gaudel et al., 2018), we use these BCFs in an attempt to harmonise the records in absolute value terms. Thus, as the BCFs are generated from a long-term average, they should improve absolute column values, but not interfere with the long-term change in the record. This was done for lower tropospheric column ozone ($LTCO_3$), as discussed in the main manuscript.

To derive the BCFs, each ozonesonde profile was spatiotemporally co-located with the nearest satellite retrieval within 500 km and 6 hours to allow for robust comparisons and reduce sampling 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 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 to the ozonesonde profiles. Firstly, the co-located ozonesonde profile (in volume mixing ratio) is interpolated onto the satellite pressure grid in $\log(pressure)$. The sonde sub-columns are then derived using the hydrostatic balance approximation:

$$mass\ density = mmr \times \rho \times dz = mmr \times \frac{-dp}{g} \qquad (1)$$

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 (m) between pressure levels, *dp* is the pressure difference (Pa) between levels and *g* is the acceleration due to gravity (-9.81 m/s^2). The application of the AKs for the Ozone Monitoring Instrument (OMI, **Equation 2**) and the Infrared Atmospheric Sounding Interferometer products (IASI, **Equation 3**) are:

$$\boldsymbol{sonde_{AK} = AK.sonde_{int} + imak\_sc\_apr} \qquad (2)$$

$$\boldsymbol{sonde_{AK} = AK(sonde_{int} - apr) + apr} \qquad (3)$$

where $\boldsymbol{sonde_{AK}}$ is the modified ozonesonde sub-column profile (Dobson units, DU), ***AK*** is the averaging kernel matrix, $\boldsymbol{sonde_{int}}$ is the sonde sub-column profile (DU) on the satellite pressure grid and ***apr*** is the apriori (DU) and ***imak_sc_apr*** represents the term ***(I-AK).apr*** where ***I*** is the identity matrix. The application of the satellite AKs to the UKESM model profiles is the same. The only

differences are that for each satellite retrieval, the closest UKESM grid box is used and is within 3-hours.

For the RAL OMI products, the data is already represented as $LTCO_3$ in the lowest layer. For the IASI products, given its greater vertical resolutions, the sub-columns between levels were totalled up to the 450 hPa layer for the $LTCO_3$. The satellite and ozonesonde, with AKs applied, $LTCO_3$ quantities were then binned into the respective latitude and monthly bins and the median biases (satellite-ozonesonde) or offsets were determined. Therefore, whenever the satellite datasets listed in **Table 1** of the main manuscript are used (e.g. for trends or comparison with UKESM), the BCFs are subtracted from the satellite data for the relevant latitude and monthly bins. An example of the OMI $LTCO_3$ BCFs is shown in **Figure S1**.

SM-2: HTAP Mask

The Hemispheric Transport of Air Pollution (HTAP) Task Force, in 2012, launched a co-ordinated multi-model and analysis programme (HTAP Phase 2) to help inform the Convention on Long-range Transboundary Air Pollution (LRTAP), national governments and multi-lateral cooperative efforts on appropriate actions to decrease air pollutant and its associated impacts (European Commission, 2016). Within HTAP Phase 2, a useful land mask was developed (**Figure S2**) to focus analysis on sub-global regions. Each region is assigned a code, so regions of e.g. O_3 data can be extracted from datasets, on the same spatial resolution of the HTAP mask, and averaged together to derive regional quantity time-series for analysis. This is a more robust approach than using a square/rectangular longitudinal-latitudinal box to approximate an area of interest. The subsequent satellite-UKESM trend and seasonal cycle analyses in **Section 3** of the main manuscript uses this HTAP mask to derive regional information.

SM-3: UKESM Evaluation

For comparison with the ozonesondes (**Figure S3**), the model was co-located in time (within 6 hours) and space (nearest model grid box) with each of the ozonesondes. The analysis has been split up into three latitude ranges (90-30°S, 30°S-30°N & 30-90°N) and four seasons (December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON)). In the both hemispheres and seasons, UKESM $LTCO_3$ is in good agreement with that of the ozonesondes with a near-zero bias and similar overlap between the model and observational variability (i.e. the 25^{th}-75^{th} percentile range). Overall, UKESM successfully reproduces the majority of the latitudinal-seasonal absolute median values and variability seen in the ozonesondes for $LTCO_3$.

UKESM (with the corresponding AKs applied) was compared with OMI, IASI-FORLI and IASI-SOFIRD $LTCO_3$ for DJF and JJA between 2008 and 2017. Note, the ozonesonde BCFs have been applied to all the satellite products. In comparison to OMI (**Figure S4**), UKESM typically simulates the $LTCO_3$ spatial distribution and seasonality. In DJF, UKESM has both regional negative (-3.0 DU to 0.0 DU) and positive (0-3.0 DU) biases, while in JJA there are widespread biases of 0.0-4.0 DU over the tropics and sub-tropics. However, in general, the absolute UKESM-OMI biases sit within the satellite uncertainty ranges. When compared to IASI-FORLI (**Figure S5**), UKESM simulates similar spatial distributions and seasonality, but largely underestimates the retrieved $LTCO_3$ by 3.0-5.0 DU. These biases tend to be classed as "substantial" biases as the absolute bias is often larger than the retrieved $LTCO_3$ uncertainty range. These low biases are most prominent over the high-latitudes (-5.0 to -3.0 DU). Against IASI- SOFRID (**Figure S6**), the comparisons become more complex due to the $LTCO_3$ latitudinal banding caused by the dynamic apriori used in the retrieval scheme. This is potentially suggestive that IASI-SOFRID has less vertical sensitivity in the surface-450 hPa range in

comparison to the other products. In general, UKESM overestimates $LTCO_3$ by 2.0 to 4.0 DU, but some regions are more substantial (e.g. northern sub-tropics in DJF, >5.0 DU, and in the Middle East/southern Africa in JJA, > 5.0 DU). Overall, UKESM robustly simulates $LTCO_3$ spatially and seasonally in comparison to the ozonesondes and satellite instruments (i.e. typically within the ozonesonde variability and satellite uncertainty range).

Figures:

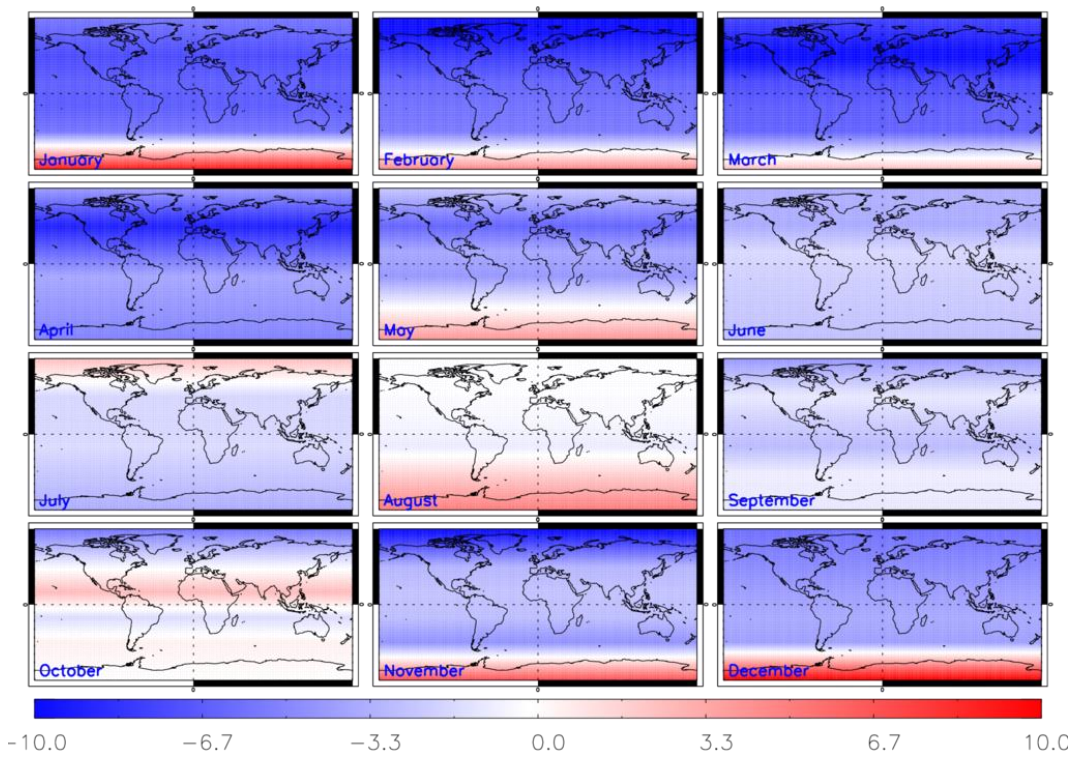

Figure S1*: OMI-ozonesonde (with AKs applied) bias correction factors (BCFs, Dobson Units (DU)) for OMI lower tropospheric column O_3 ($LTCO_3$) using the instrument record between 2008 and 2017.*

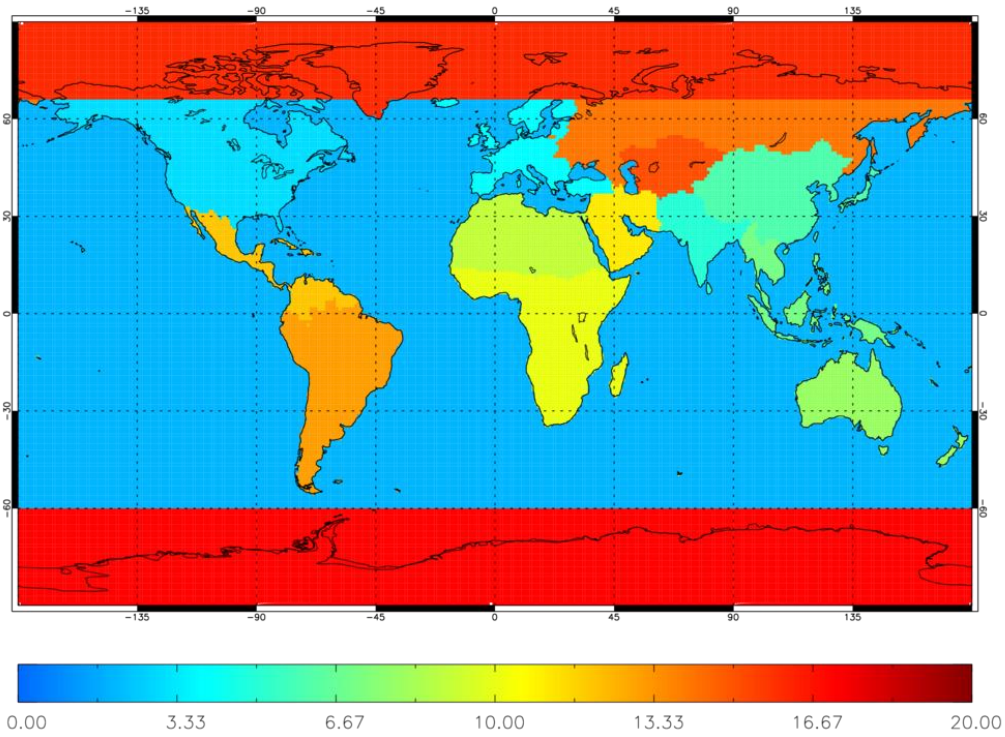

Figure S2: *Mask of different regions provided by the Task Force on Hemispheric Transport of* Air *Pollution (HTAP) on a 1°×1° horizontal resolution. For instance, mask code 4 represents Europe.*

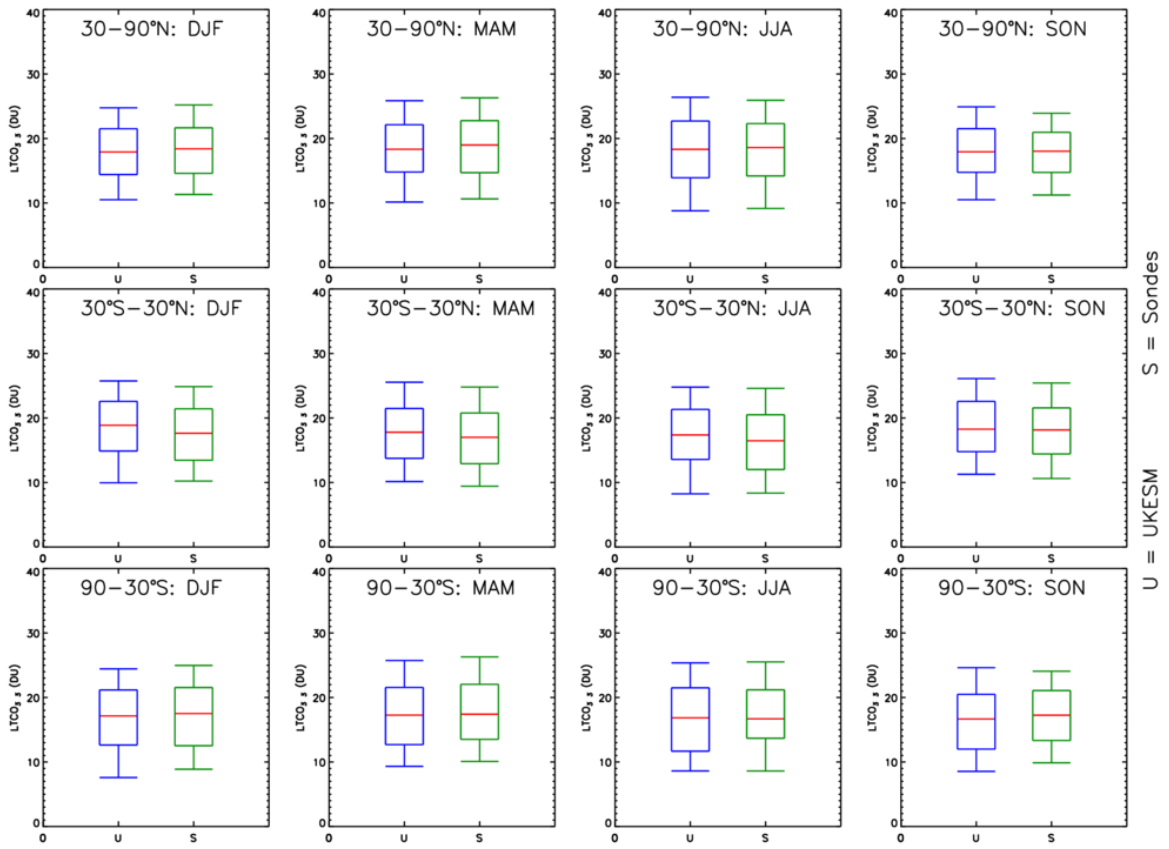

Figure S3: *Comparison of UKESM and ozonesonde $LTCO_3$ (DU) between 2008 and 2017 are shown in blue and green respectively for December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON) across the latitude bands: 90-30°S, 30°S-30°N and 30-90°N. The sample median (red), 25th, 75th, 10th and 90th percentiles are shown in the box and whisker plots.*

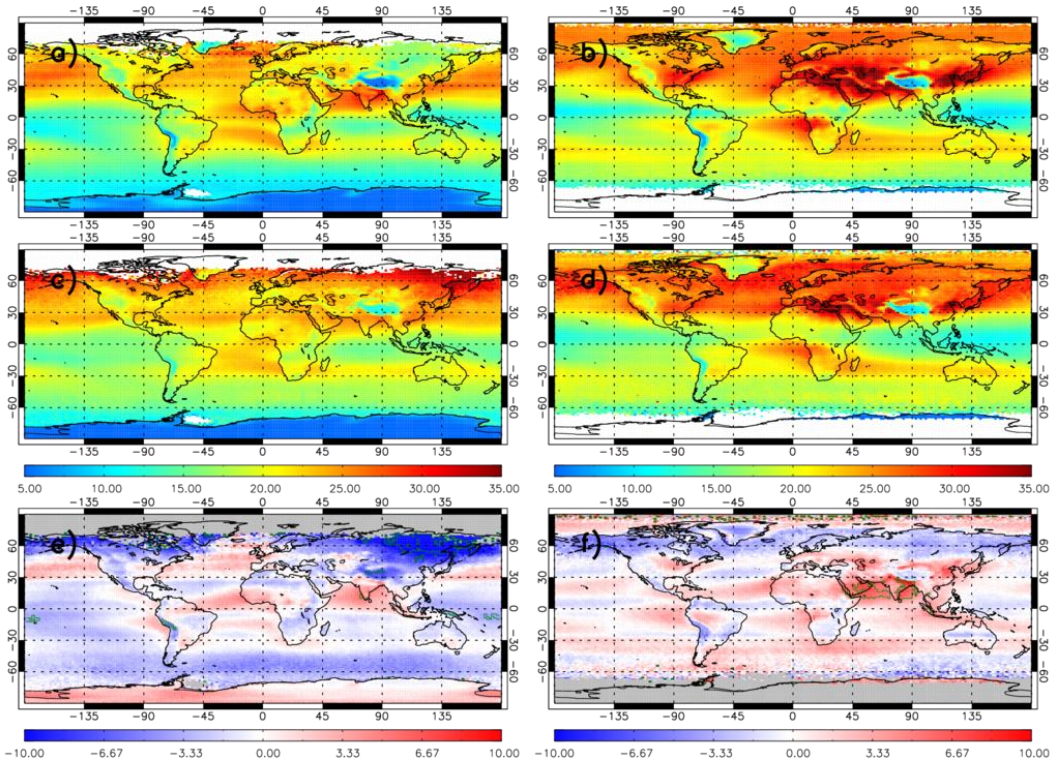

Figure S4: *$LTCO_3$ (DU) between 2008 and 2017 for a) UKESM with OMI averaging kernels (AKs) applied in DJF, b) UKESM with OMI AKs applied in JJA, c) OMI in DJF and d) OMI in JJA. Panels e) and f) show the UKESM-OMI mean bias for DJF and JJA, respectively. Green polygon-outlined regions show where the model-satellite biases are larger than the satellite error, but where the absolute bias is greater than 1.0 DU (i.e. focus on more substantial absolute biases).*

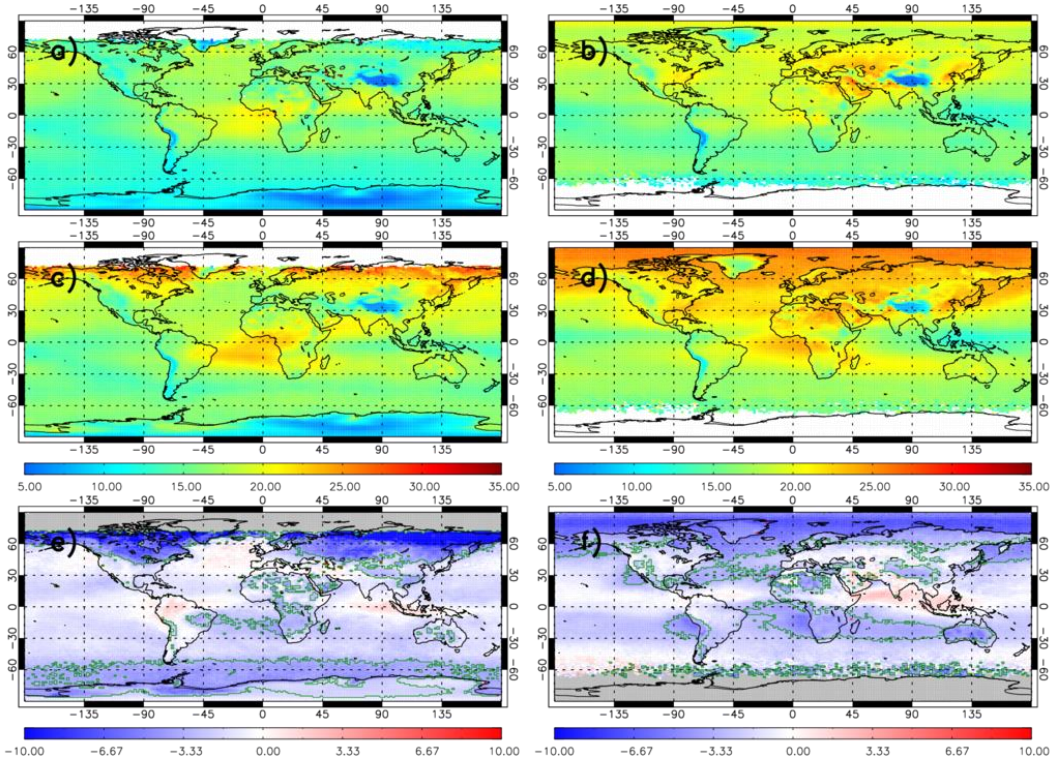

Figure S5: *Same as **Figure S4** but for IASI-FORLI $LTCO_3$ (DU).*

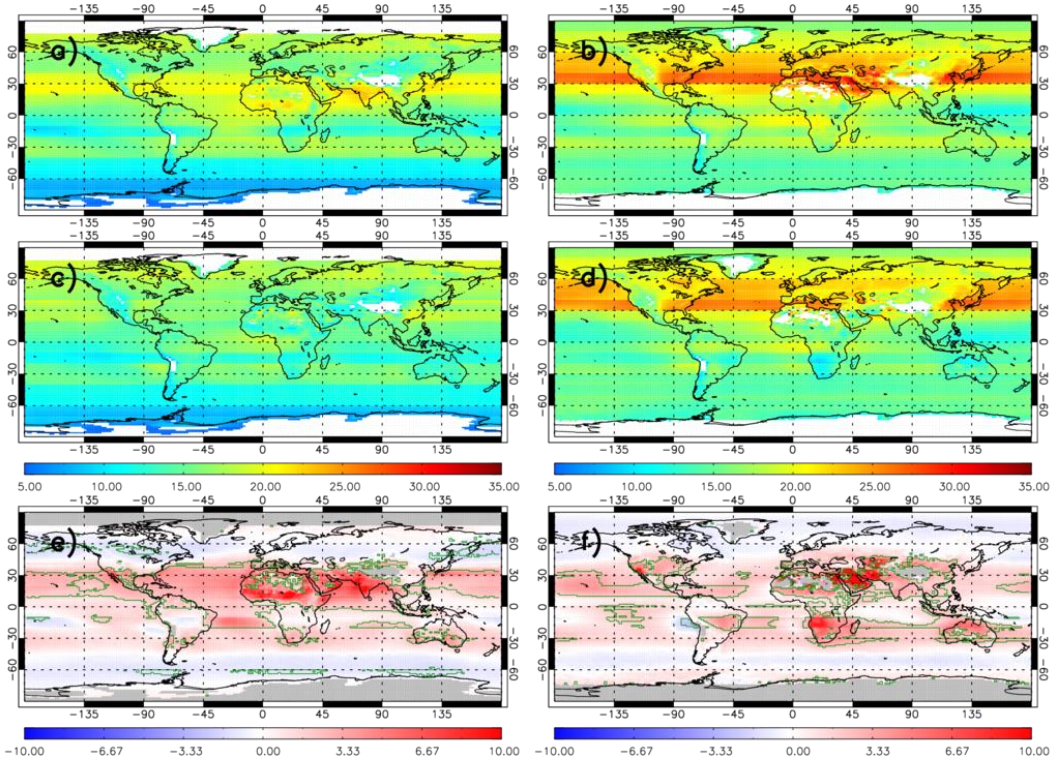

Figure S6: *Same as **Figure S4** but for IASI-SOFRID LTCO$_3$ (DU).*