United Kingdom Earth System Model – Ozone Monitoring Instrument Total Column Sulphur Dioxide Comparisons

Met Office Technical Document

Richard J Pope[1,2], Catherine Hardarce[3], Can Li[4,5] and Martyn P. Chipperfield[1,2]

1: School Earth and Environment, University of Leeds, Leeds, UK

2: National Centre for Earth Observation, University of Leeds, Leeds, UK

3: Met Office, Exeter, UK

4: Earth System Science Interdisciplinary Center, University of Maryland, MD, USA

5: Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, USA

Date: 14th May 2021

1. Background

A key component of the National Centre for Earth Observation (NCEO) – Long-term Science Multi-Centre (LTSM) project is to evaluate the United Kingdom Earth System Model (UKESM, Sellar et al., 2019), developed by the UK Met Office and the National Centre for Atmospheric Science (NCAS), using Earth observation (i.e. satellite) data sets.

Within the UKESM-LTSM programme, this work evaluates the atmospheric composition simulated within the model using the United Kingdom Chemistry and Aerosol (UKCA) sub-model; Archibald et al., (2020)). Here, we have evaluated UKESM atmospheric sulphur dioxide (SO_2) using total column SO_2 ($TCSO_2$) retrievals from the Ozone Monitoring Instrument (OMI) on-board NASA's Aura satellite.

2. Model Setup and Satellite Data

2.1. Modelling Framework

For this study we used the UK's Earth System model (UKESM1), the latest generation Earth system model to be developed in the UK. UKESM1 takes HadGEM3-GC3.1 as it is physical-dynamical core and couples it to additional Earth system components including the marine and terrestrial biogeochemical cycles and fully interactive stratospheric-tropospheric trace gas chemistry. UKESM1, its component models and the coupling between them are described fully in Sellar et al., (2019). Relevant to this study, the aerosol scheme, including SO_2 emissions and sulphur chemistry, is described in Mulcahy et al., (2020).

The version of UKESM1 used here has a regular latitude-longitude grid with a resolution of $1.25^o \times 1.875^o$ (~135 km at the mid latitudes). It has 85 vertical levels on a terrain following hybrid height coordinate with a model top at 85 km above sea level and 50 of these levels are below 18 km. The ocean has a horizontal resolution of 1^o and 75 vertical levels. Note that

UKESM1 has an atmospheric time step for the model physics of 20 minutes, but due to the inherent computational cost of the chemistry and aerosol components, both components are called once per hour. We made use of several different model simulations that utilised different model configurations in this study **(see Table 1)**.

The atmosphere only (AMIP) configuration used for u-bu004, u-bp880, u-bw217 were run from 1979 to the end of 2014 and allow additional simulations to be carried out at a much reduced computational cost. The AMIP configuration was driven by observed sea surface temperature (SST) and sea ice fields as it does not include the additional dynamic ocean and land surface components (Eyring et al., 2016). Instead the required vegetation (vegetation fractions, leaf area index, canopy height) and surface ocean biology fields (DMS and chlorophyll) are taken from a single UKESM1 historical member and are prescribed as ancillary data, thereby maintaining traceability to the fully coupled model. For u-bu004, the model meteorology (e.g. winds and temperature) was nudged to reanalysis meteorology (i.e. ECMWF ERA-Interim; Dee et al., (2011); Telford et al., (2008)).

The fully-coupled historical configuration of UKESM1 used for u-bc292 and u-bk575 was forced by transient external forcings of solar variability, well-mixed greenhouse gases and other trace gas emissions and aerosols. The volcanic forcing due to the stratospheric injection of SO_2 from volcanic eruptions is prescribed as a zonal mean climatology of the stratospheric aerosol optical properties over the historical period. All forcings and how they are implemented is described fully in (Sellar et al., 2020). The historical simulations cover the period from 1850 to the end of 2014 and therefore model the evolution of climate since the pre-industrial era.

As standard all configurations of UKESM have anthropogenic SO_2 emissions injected at the surface. However, to test the impact of vertically distributing the emissions, as per HadGEM3-GC3.1, we use job u-bw217. In HadGEM3-GC3.1, 50% of industrial emissions are emitted at 500 m, 50% are emitted at the surface; 100% of energy emissions are emitted at 500 m. The UKESM control run (i.e. emissions injected at the surface) is job u-bp880. We also investigate the impact of developments made to the SO_2 dry deposition parameterization as documented in Hardacre et al. (2021) and being implemented in the latest release of UKESM, UKESM1.1. The experiment (u-bk575) used a historical simulation (u-bc292, i.e. control run) with modifications of Hardacre et al. (2021) to the dry deposition parameterization. The first modification was to the surface resistance term for SO_2 dry deposition to water which was previously set to an unrealistically high value of 148 s cm^{-1}. Following an examination of the literature, this term has been set to 1 s cm^{-1}, SO_2 being very soluble in water. This change will increase SO_2 dry deposition, mostly over the ocean. The second change is a modification to better represent SO_2 dry deposition to land surfaces after rainfall. This change allows a surface to remain wet after rainfall for a period of three hours, where previously it would have been 'dry' immediately after the rainfall event. Because SO_2 is highly soluble this change will also increase SO_2 dry deposition, but only over land surfaces.

2.2 Satellite Instrument and Data Processing

Aura is a polar-orbiting satellite launched in 2004 with a local overpass time of ~13.45. OMI has a nadir footprint of 13 km × 24 km and a spectral viewing range of 270 to 500 nm (Levelt et al., 2018). The OMI $TCSO_2$ product was downloaded from https://aura.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level2/OMSO2.003 (Li et al., 2020). The data are quality controlled for cloud radiance fraction > 0.0 and <0.5, solar zenith angle < 65°, the South Atlantic Anomaly flag = 0, ice cover flag = 0, the air mass factor (AMF) > 0.3 and $TCSO_2$ > -1.0 DU. Generally, background $TCSO_2$ can have relatively large uncertainty ranges, where the mean background values are just above 0.0, but individual retrievals can be slightly negative. Such negative values such be included when calculating averages, else the mean background state can be positively skewed. Volcanic regions and high activity days where $TCSO_2$ is extremely large can positively skew satellite averaged data. To that end, OMI data (and corresponding model profiles) was excluded based on the spatio-temporal mask provided by Can Li (NASA), using the approach of Carn et al., (2017).

For the most accurate model-satellite comparisons, model data must be spatio-temporally co-located to the satellite swath data (i.e. reduce sampling biases) and the vertical sensitivity of the satellite accounted for. Often, this is done by mapping the vertical sensitivity of the satellite onto the model profile using a weighting function known as an averaging kernel (AK). This allows for a direct like-for-like comparison between both data sets. However, the OMI $TCSO_2$ product does not include AKs, so the satellite AMFs, used to derive the satellite total vertical column, are modified using the model vertical information.

The satellite total vertical column (*VCD*) is calculated by:

$$VCD = \frac{SCD}{AMF} \quad (1)$$

where *SCD* is the slant column density. The AMF is calculated by:

$$AMF = \int_0^{TOA} m(z)\, n_{SO2}(z) dz \quad (2)$$

where m is the layer Jacobians (sometimes also referred to as scattering weights or box AMFs), n_{SO2} is the normalised SO_2 profile, z is the altitude co-ordinate and *TOA* is the top of atmosphere. Here n_{SO2} is based on a model. Therefore, to allow for accurate model-satellite comparisons, the model and satellite vertical distributions need to be comparable. The UKESM model normalised SO_2 profile needs to be incorporated into **Equation 2** to derive a new AMF (incorporating the model vertical distribution) and then the satellite SCD observation can be converted to a VCD using **Equation 1**. All of which is discussed in the OMI SO_2 readme file (Krotkov et al., 2020; Li et al., 2020).

To achieve this, once a model profile has been co-located with the satellite retrieval, the model profile (in mass mixing ratio) is interpolated onto the pressure grid for which the layer Jacobians are specified. The sub-columns (e.g. kg/m^2 or molecules/m^2) between pressure levels are determined (using the hydrostatic balance approximation) and the model $TCSO_2$

used to normalise the profile sub-columns (n_{SO2}).The dot product of n_{SO2} and $m(z)$ then yields the AMF, from which the VCD (i.e. $TCSO_2$) can be calculated. As model pixels and satellite retrievals have to be co-located, both can be mapped back onto the model spatial grid.

3. Sensitivity of Comparison to Model Temporal Output

As above, robust model-satellite comparisons ideally require spatio-temporal co-location. However, this becomes problematic when standard UKESM1 (and other Earth system models) simulations typically output monthly means (MM) due to their long-term climate focus. For most comparisons, 3-hourly or 6-hourly (6H) model output of 3D-tracer and pressure fields are required over the analysis period (Pope et al., 2016; Monks et al., 2017). In the case of u-bu004 SO_2, this results in output volumes of 100s GBs/several TBs requiring substantial computing processing power and time, especially over long-term satellite time periods (e.g. 2005-2014, which is the focus in this study). The corresponding MM output is relatively small with data volumes of 10s GBs. For tropospheric column nitrogen dioxide ($TCNO_2$) between UKESM and OMI, we have found that the model fields can differ by up to 50-100% in these types of model-satellite comparisons. The MM output averages over all time steps (including peak NO_2 values), while high resolution output is co-located to the satellite overpass time (e.g. for OMI it is 13.45 local time at the NO_2 diurnal minima). Therefore, species with a strong diurnal cycle require the higher temporal resolution outputs.

For SO_2 we suspected that this would not be such an issue, and if we wanted to use OMI $TCSO_2$ data to evaluate several UKESM1 sensitivity runs, it was only realistically possible with the MM output. Therefore, **Figures 1** and **2** show the – OMI $TCSO_2$ comparisons for December-January-February (DJF) and June-July-August (JJA) between 2005 and 2014 using 6 hourly and monthly output data respectively. As satellite data (i.e. individual retrievals) can be subject to large errors and uncertainties it is important to generate satellite composite averages with substantial data samples to reduce the retrieval noise and achieve more accurate results.

In **Figure 1**, using the u-bu004 6H output, the model $TCSO_2$ values in DJF (**Figure 1a**) and JJA (**Figure 1b**) are larger than that of OMI over most source regions and outflows. During the hemispheric winter, the UKESM and OMI $TCSO_2$ values are larger over the source regions than in summer. In the median biases (**Figure e & f**), the model typically overestimates by 0.5-1.0 Dobson units (DU) over source regions and 0.1-0.3 DU in outflow regions. In **Figure 2**, using u-bu004 MM model data, the comparison gives consistent results to when 6 hourly data is used. The seasonal cycle is comparable in the model and satellite, and the model overestimates to a similar degree. Therefore, this suggests that 1) the model overestimates SO_2 and 2) the temporal sampling of the model is not overly critical for SO_2 over anthropogenic regions (i.e. modelled SO_2 has a sufficiently long lifetime to dampen the influence of diurnal sampling of the model).

Figure 3 supports this where the u-bu004 MM and u-bu004 6H $TCSO_2$ fields have small absolute average differences. In **Figure 3a** (u-bu004 MM - u-bu004 6H), there are negative differences of up to -0.25 DU over China in DJF. In the background regions, there are positive differences of 0.01-0.03 DU. In JJA there are mixed differences over China with absolute values of up to ~0.1 DU. Over other regions (e.g. India and eastern Europe), positive biases exist between approximately 0.02 and 0.08 DU. In percentage terms (**Figure 3c & d**), anthropogenic regions (e.g. China and India) have absolute difference peak between 10-30% and when compared to the model-satellite biases, are of secondary importance. This is shown by the green contouring, where these pixels have inter-model output absolute differences larger than that of the u-bu004 6H - OMI $TCSO_2$ differences. As the green contouring is limited to a few spatially random pixels with limited structure, it suggests that we can use UKESM1 MM data for model-satellite comparisons as the temporal sampling issues have limited influence on the model-satellite discrepancies. However, it is worth noting that there are large percentage biases over the remote regions, despite the absolute differences between the model simulations being very small. In the Arctic and Antarctic summer time, positive differences peak at 30-50%. In the Southern Ocean, background $TCSO_2$ differences reach -40%. These regions are also where OMI will have less sensitivity to retrieving $TCSO_2$, due to the low abundance of atmospheric SO_2, and critical interpretation of model-satellite comparisons in these remote regions is required.

The statistics in **Figure 3** show that the u-bu004 MM - u-bu004 6H correlation (linear fit) are 0.78 (0.97) and 0.77 (0.93) for DJF and JJA, respectively. The inter-quartile range (IQR) for u-bu004 MM for DJF and JJA are 0.02 DU and 0.03 DU, similar to that of u-bu004 6H (i.e. DJF and JJA IQRs are 0.03 DU and 0.03 DU). Overall, this will make evaluation of UKESM1 simulations (i.e. MM output) using OMI much quicker, yet still robust for anthropogenic regions, but with careful interpretation over remote regions.

4. Model Emission Representativeness

A possible cause of the model-satellite positive biases are the input emissions. The model SO_2 emissions come from CMIP6 (Coupled Model Intercomparison Project Phase 6, Erying et al., 2016) and here we compare the CMIP6 emissions with the equivalent emissions from OMI-HTAP (Liu et al., 2018) for all sectors (https://avdc.gsfc.nasa.gov/pub/data/project/OMI_HTAP_emis/). Here, OMI top-down SO_2 emissions have been merged with the bottom-up inventory from HTAP. Note that the elevated and surface emissions from both data sets have been merged into one level. **Figure 4** shows the difference between CMIP6 and OMI/HTAP over the seasonal cycle between 2005 and 2014. In general, CMIP6 emissions are larger than that of OMI/HTAP between 0.0 and $>1.0\times10^{-10}$ $kg/m^2/s$, especially over the source regions. There are examples where OMI/HTAP is larger at a similar magnitude, but these are more sporadic point sources. Most likely, the point sources giving negative differences are from OMI, while the more diffuse emission differences are dominated by HTAP.

As the emission differences are sporadic and noisy in **Figure 4**, the regional annual emission totals have been derived in **Table 2**. Green grid cells show where CMIP6 > OMI-HTAP and red is where OMI-HTAP > CMIP6. In all but three regions, the CMIP6 emissions are larger (i.e. well over 100% in many regions). In the 3 regions where OMI-HTAP is larger, I suspect these are dominated by large emission point sources (e.g. power stations, smelters and oil and gas; McLinden et al., 2016) detected by OMI. Overall, the annual global total for CMIP6 is 115 Tg/yr and 100 Tg/yr for OMI-HTAP.

A further comparison (**Figure 5**) highlighted a similar CMIP6 emission overestimation using EDGAR (Emission Database for Global Atmospheric Research, vn.4.3.2, https://edgar.jrc.ec.europa.eu/overview.php?v=432_AP) all sector data emissions (Crippa et al., 2018) for 2012. Overall, the annual global total for CMIP6 (111 Tg/yr) is moderately higher than EDGAR (102 Tg/yr), but there are widespread positive differences between 0.0 and $>1.0\times10^{-10}$ kg/m^2/s across most source regions globally. Similarly to OMI-HTAP, the large negative differences ($< -1.0\times10^{-10}$ kg/m^2/s) are more sporadic, isolated to individual grid cells. The EDGAR shipping emissions are larger by between 0.1×10^{-10} kg/m^2/s and 0.2×10^{-10} kg/m^2/s in the main shipping channels.

5. UKESM Experiment Comparisons

OMI $TCSO_2$ is further utilised to assess model skill when modifications of key processes are incorporated into the model. As stated in **Section 2.1**, u-bp880 (control) and u-bw217 (experiment) are free running simulations and investigate the impact of the injection height of SO_2 emissions into the model. As shown in **Figure 6a & b,** job u-bp880 simulates larger $TCSO_2$ compared to OMI as background positive biases exist in both DJF and JJA peaking at 0.6-0.9 DU. A similar pattern exists for u-bw217 (**Figure 6c & d**), but with slightly larger positive background biases peaking at > 1.0 DU in both seasons. **Figure 7a & b** inter-compares u-bp880 and u-bw217, where u-bw217 over source and outflow regions is larger by approximately 0.0-0.2 DU. The percentage difference (**Figure 7c & d**) supports this where background TSO_2 in u-bw217 is typically larger by 20-50%, peaking at 70-90% over Eurasia and North America in DJF. Therefore, vertically distributing the emissions degrades the comparisons with OMI, but in comparison to the absolute model (control, u-bp880) - satellite differences (i.e. green contouring), this perturbation to the model TSO_2 field is relatively small.

UKESM1 jobs u-bc292 (control) and u-bk575 (experiment) examine the impact of modifying the SO_2 dry deposition parameterization. These modifications act to increase SO_2 flux from the atmosphere to the Earth's surface, improving the positive SO_2 bias reported here and in Hardacre et al (2021). This would subsequently reduce atmospheric sulphate and associated mid-20th century aerosol forcing (Hardacre et al. 2021) which is believed to be a key contributor to a large mid-20th century cold temperature bias in UKESM1 (Sellar et al, 2019).

The increased dry deposition of SO_2 to the surface appears to decrease $TCSO_2$ and provide moderately better comparisons with OMI. In **Figures 8a & b**, the u-bc292-OMI $TCSO_2$ biases

peak at 0.6 - > 1.0 DU, with background differences of 0.2-0.5 DU, in both seasons (largest in DJF). However, in u-bk575, these differences decrease substantially with source region biases of 0.3-0.5 DU and background biases between 0.1-0.3. In **Figures 9a & b**, the modifications to the SO_2 dry deposition parameterization have decreased $TCSO_2$ by > 0.25 DU over China and India and by 0.05 to 0.1 DU over Eurasia and the USA. $TCSO_2$ in outflow regions has also decreased, but by < 0.05 DU. In the percentage differences (**Figures 9c & d**), the background $TSCO_2$ has dropped globally by 20-30%. Over the outflow regions (e.g. off the USA eastern seaboard), $TSCO_2$ has reduced by 30-50%. Over the source regions, this varies by 30-50% over South East Asia, 20-30% over Europe and 10-30% over the US. Therefore, modifications to the SO_2 dry deposition parameterization, have improved the representation of SO_2, but the inter-model differences are still smaller compared to the existing model-satellite systematic difference (i.e. green contouring remains sporadic with limited spatial coverage).

These model discrepancies with OMI are further highlighted in **Figure 10** and **Table 3**, where the regional temporal evolution of $TCSO_2$ over the USA, Europe and South East (SE) Asia are investigated. Over the USA (**Figure 10c**), the OMI regional average time series has a seasonal cycle amplitude ranging between 0.02 and 0.05 DU. Though model simulations have typically larger values ranging between 0.035 and 1.2 DU, the amplitude of the seasonal cycle is similar. While there is no obvious trend in the OMI time series, the model simulations have a decreasing tendency between 2005 and 2011/2012. Beyond 2012, the model and satellite time series have similar tendencies and absolute values. Over Europe (**Figure 10b**), there is a similar temporal pattern where OMI $TCSO_2$ is lower (0.025-0.09 DU), between 2005 and 2009, than the model simulations (0.07-2.28 DU). The model decreasing trends plateau in approximately 2009/2010 where the u-bk575 MM and u-bu004 6H time series have consistent absolute $TCSO_2$ values with OMI (0.025-0.09 DU). The seasonal amplitude of u-bw217 MM typically ranges between 0.08 and 0.3 DU representing the most extreme model simulation. In all cases, the phase of the satellite and model seasonal cycles are in reasonable agreement. Over SE Asia (**Figure 10a**), there is a systematic overestimation by the model simulations between 0.075 and 0.125 DU. OMI $TCSO_2$ ranges between 0.02 and 0.04 DU with a low seasonal amplitude (~0.02 DU). u-bk575 MM and u-bw217 MM have seasonal amplitudes of approximately 0.03 and 0.07 DU, respectively, representing the lower and upper extremes of the multiple model $TCSO_2$ time series.

Looking at **Table 3** (statistics of the time series in **Figure 10**), u-bk575 MM has the lowest mean bias (MB) ranging between 0.014 (56.1%) and 0.068 (227.9%) DU. UKESM job u-bw217 has the largest mean bias of 0.044 (112.8%)-0.152 (510.2%) DU. In most cases, the USA has the lowest regional MBs and root mean square errors (RMSE). However, except for u-bk575, Europe tends to have the lowest percentage mean biases (MB%). In all cases, SE Asia has the largest MBs, RMSEs and MB%s indicating that model has the great difficulties to accurately simulate $TCSO_2$ absolute values in this region. Interestingly though, the model simulations most accurately capture the observational seasonality in SE Asia with correlations (R)

between 0.44 and 0.67. The USA has the worst seasonal agreement with R between 0.19 and 0.28. European model-satellite regional time series correlations range between 0.43 and 0.58.

6. Conclusions

Overall, the model substantially overestimates OMI $TCSO_2$ for all simulations across most of the globe. The best comparisons are in the nudged version of UKESM1 (u-bu004) and the model experiment (u-bk575) which modified the SO_2 dry deposition parameterization to increase SO_2 removal from the atmosphere in UKESM1. Though typically, the nudged simulations compare more robustly with OMI then the free-running simulations. The 6H model output (from u-bu004) also provides marginally better comparisons with OMI than that of the MM output (from u-bu004). Though it is worth noting that these differences are small compared to the absolute model (u-bu004) - satellite biases and using SO_2 MM output from UKESM for these types of comparisons appears to be robust. As discussed above, the free running simulation (u-bp880) degrades the comparisons with OMI further, which is not improved by the vertical (0-500 m) emission of SO_2 from anthropogenic sources in the model. Overall though, we suspect that the dominate cause for the substantial model-satellite biases are the input emissions from CMIP6, which are larger than the OMI-HTAP and EDGAR (2012) inventories.

Acknowledgements:

RJP was supported by the UK Natural Environment Research Council (NERC) by providing funding for the National Centre for Earth Observation (NCEO) through grant number NE/R016518/1. OMI total column SO_2 data was obtained from NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC, https://disc.gsfc.nasa.gov/).
CH was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra (GA01101). We additionally acknowledge the EU Horizon 2020 CRESCENDO 30 project, grant number 641816. This work used JASMIN, the UK's collaborative data analysis environment (http://jasmin.ac.uk).

References:

Archibald et al. 2020. Description and evaluation of the UKCA stratosphere-troposphere chemistry scheme (StratTrop vn 1.0) implemented in UKESM, *Geophysical Model Development*, 13, 1223-1266, https://doi.org/10.5194/gmd-13-1223-2020.

Carn SA, et al. 2017. A decade of global volcanic SO_2 emissions measured from space, *Nature Scientific Reports*, 7:44095, https://doi:10.1038/srep44095.

Crippa M, et al. 2018. Gridded emissions of air pollutants for the period 1970-2012 within EDGAR v4.3.2, *Earth System Science Data*, 10, 1987-2013, https://doi.org/10.5194/essd-10-1987-2018.

Dee DP, et al. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553-597, https://doi.org/10.1002/qj.828.

Eyring V, et al. 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Hardarce C, et al. 2021. Evaluation of SO_2, SO_4^{2-} and an updated SO_2 dry deposition parameterization in UKESM1, *Atmospheric Chemistry and Physics Discussions*, https://doi.org/10.5194/acp-2021-238.
Krotkov NA, Li C and Leonard PJT. 2020. README Document for OMSO2: Aura/OMI Sulfur Dioxide Level 2 Product, *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, pp 1-22.

Levelt PF, et al. 2018. The Ozone Monitoring Instrument: overview of 14 years in space, *Atmospheric Chemistry and Physics*, 18, 5699-5745, https://doi.org/10.5194/acp-18-5699-2018.

Liu F, et al. 2018. A new global anthropogenic SO2 emission inventory for the last decade: a mosaic of satellite-derived and bottom-up emissions, *Atmospheric Chemistry and Physics*, 18, 16571–16586, https://doi.org/10.5194/acp-18-16571-2018.

McLinden CA, et al. 2016. Space-based detection of missing sulphur dioxide sources of global air pollution, *Nature Geoscience*, 8, 496-500, https://doi.org/10.1038/ngeo2724.

Monks SA, et al. 2017. The TOMCAT global chemistry transport model v1.6: description of chemical mechanism and model evaluation, *Geoscientific Model Development,* 10, 3025-3057, https://doi.org/10.5194/gmd-10-3025-2017.

Mulcahy JP, et al. 2020. Description and evaluation of aerosol in UKESM1 and HadGEM3-GC3.1 CMIP6 historical simulations, *Geoscientific Model Development*, 13, 6383-6423, https://doi.org/10.5194/gmd-13-6383-2020.

Pope RJ, et al. 2016. Intercomparison and evaluation of satellite peroxyacetyl nitrate observations in the upper troposphere – lower stratosphere, *Atmospheric Chemistry and Physics*, 16, 13541-13559, https://doi.org/10.5194/acp-16-13541-2016.

Sellar AA et al. 2019. UKESM1: Description and Evaluation of the U.K. Earth System Model, *Journal of Advances in Modeling Earth Systems*, 11, 4513-4558, https://doi.org/10.1029/2019MS001739.

Sellar AA, et al. 2020. Implementation of U.K. Earth System Models for CMIP6, *Journal of Advances in Modeling Earth Systems*,12, e2019MS001946, https://doi.org/10.1029/2019MS001946.

Telford PJ, et al. 2008. Technical Note: Description and assessment of a nudged version of the new dynamics Unified Model, *Atmospheric Chemistry and Physics*, 8, 1701-1712, https://doi:10.5194/acp-8-1701-2008.

Figures & Tables:

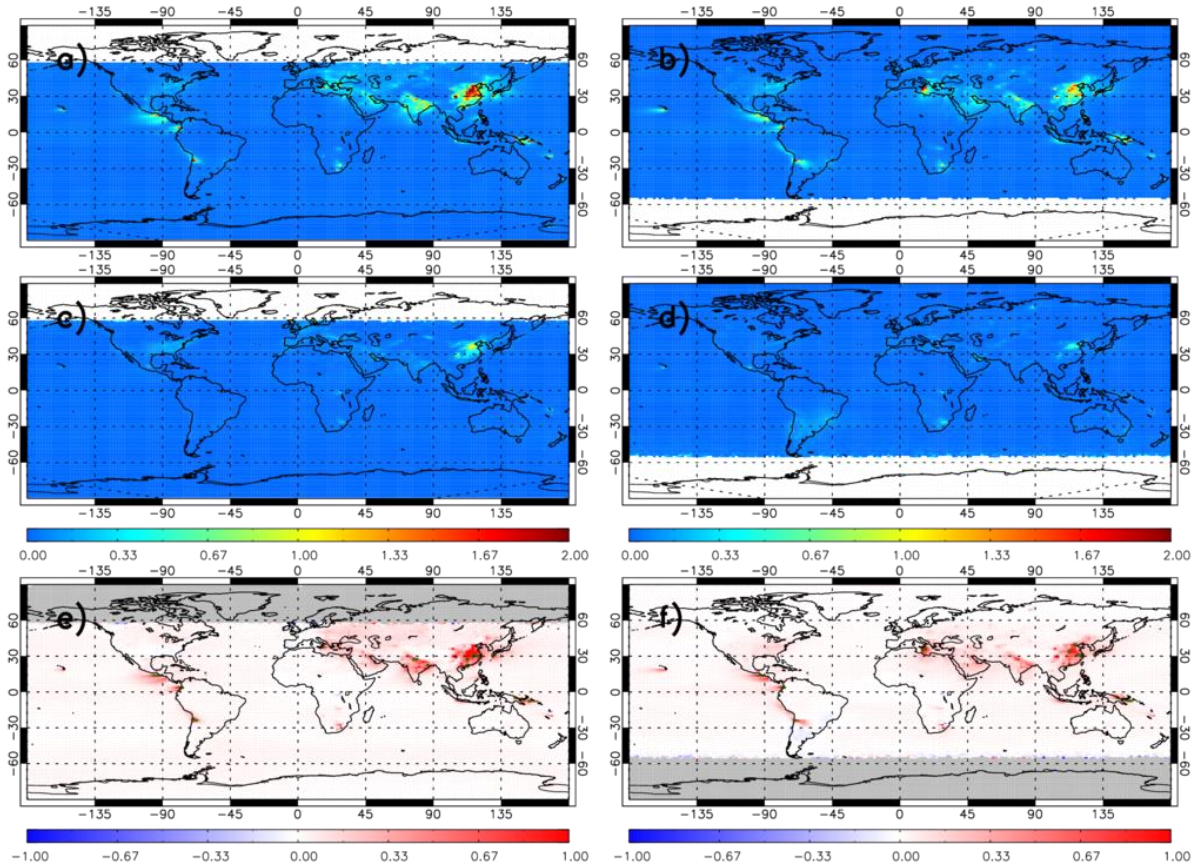

Figure 1: Median total column sulphur dioxide ($TCSO_2$; Dobson units, DU) for 2005 to 2014 for a) UKESM job u-bu004 6-hourly output (6H) in December-January-February (DJF), b) u-bu004 6H in June-July-August (JJA), c) OMI in DJF, d) OMI in JJA, e) u-bu004 6H - OMI in DJF and f) u-bu004 6H – OMI in JJA. Model fields have been spatio-temporally co-located with the OMI overpasses and the satellite AMFs updated to account for the model vertical structure. OMI data has been filtered for volcanic events, with associated model samples excluded as well. This is the same for all following figures.

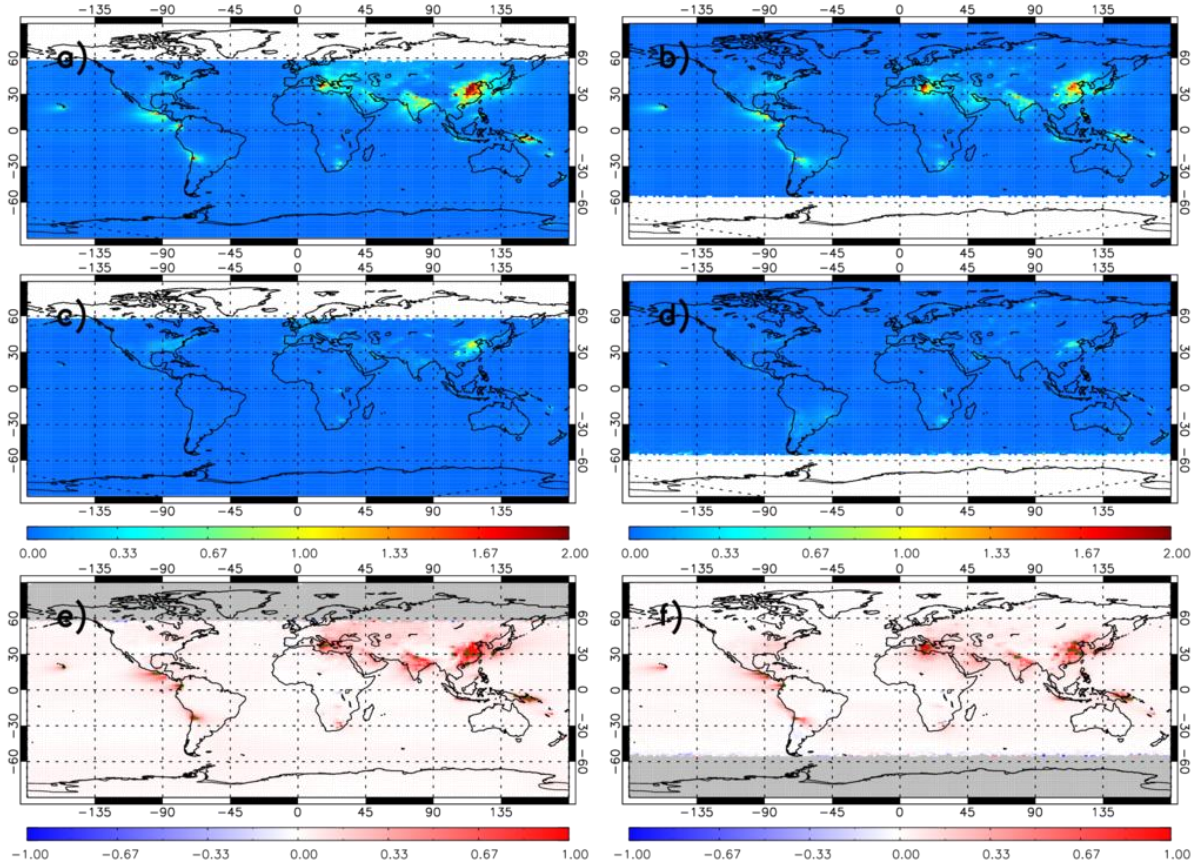

Figure 2: Same as Figure 1, but using monthly mean output (MM) for UKESM job u-bu004.

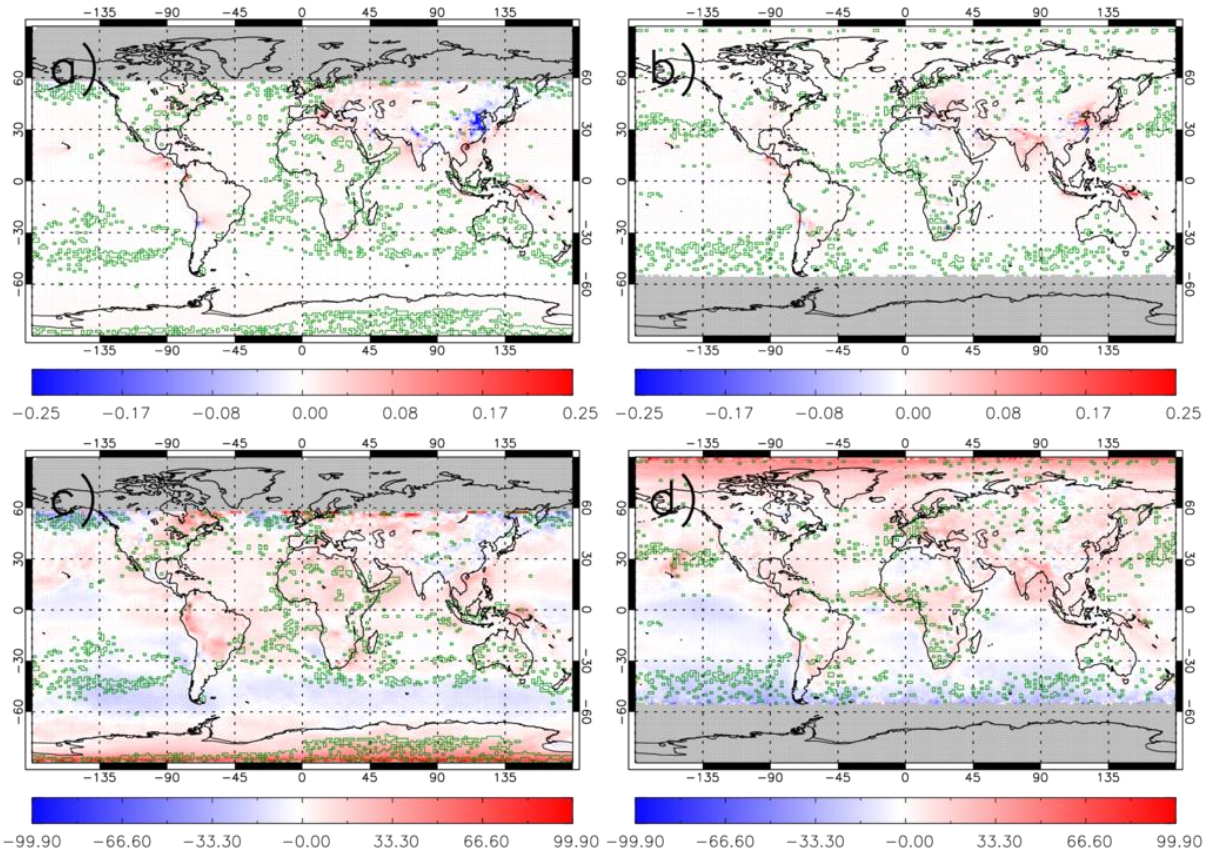

Figure 3: Mean $TCSO_2$ (DU) differences for 2005 to 2014 for a) u-bu004 MM – u-bu004 6H in DJF and b) u-bu004 MM – u-bu004 6H in JJA. c) and d) represent the percentage differences for a) and b), respectively. Green polygon-outlined regions represent where the absolute inter-model output difference is larger than the absolute u-bu004 6H – OMI differences (i.e. where the temporal output of the model will have a substantial impact on the model-satellite comparisons). The u-bu004 MM - u-bu004 6H correlation (linear fit) are 0.78 (0.97) and 0.77 (0.93) for DJF and JJA, respectively. The inter-quartile range (IQR) for u-bu004 MM for DJF and JJA are 0.02 DU and 0.03 DU. For u-bu004 6H, DJF and JJA IQRs are 0.03 DU and 0.03 DU.

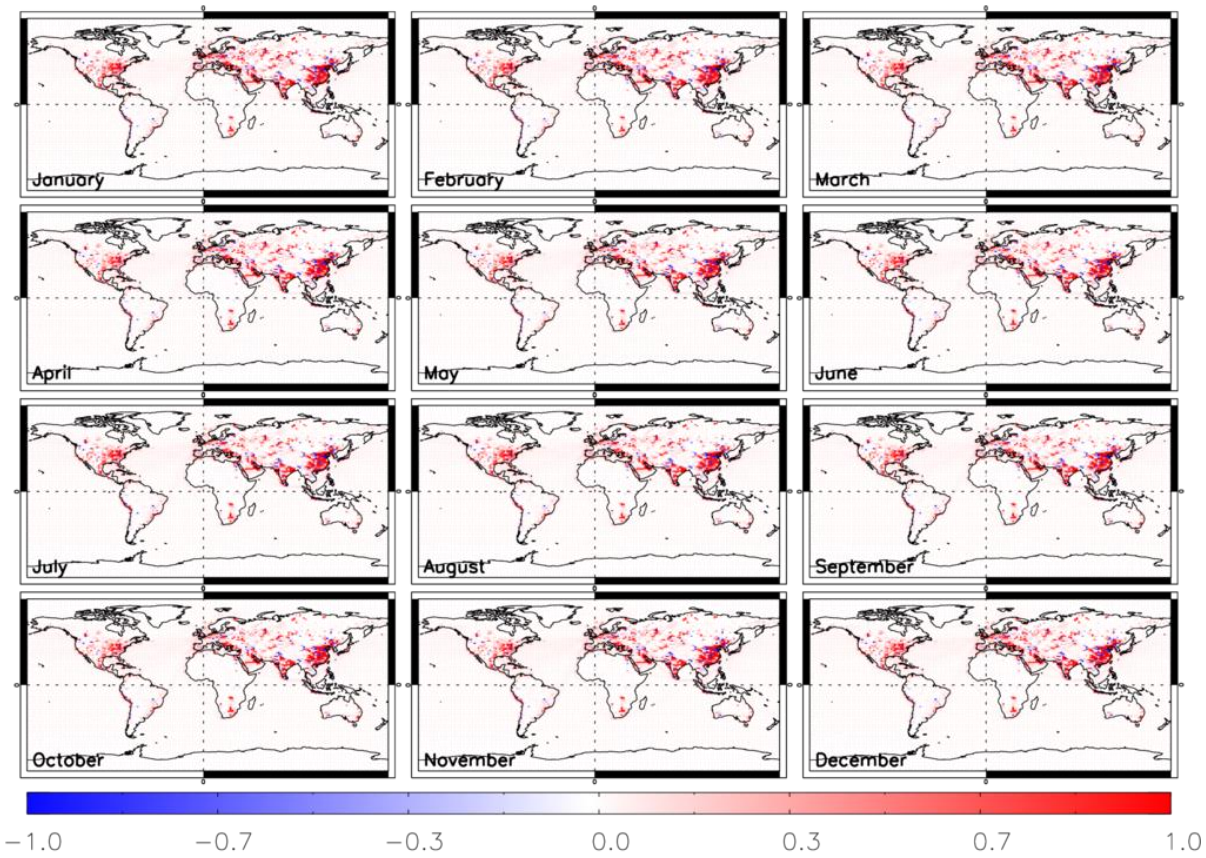

Figure 4: Sulphur dioxide (SO_2) emission ($\times 10^{-10}$ kg/m^2/s) monthly average differences between CMIP6 and OMI-HTAP for the 2005-2014 period.

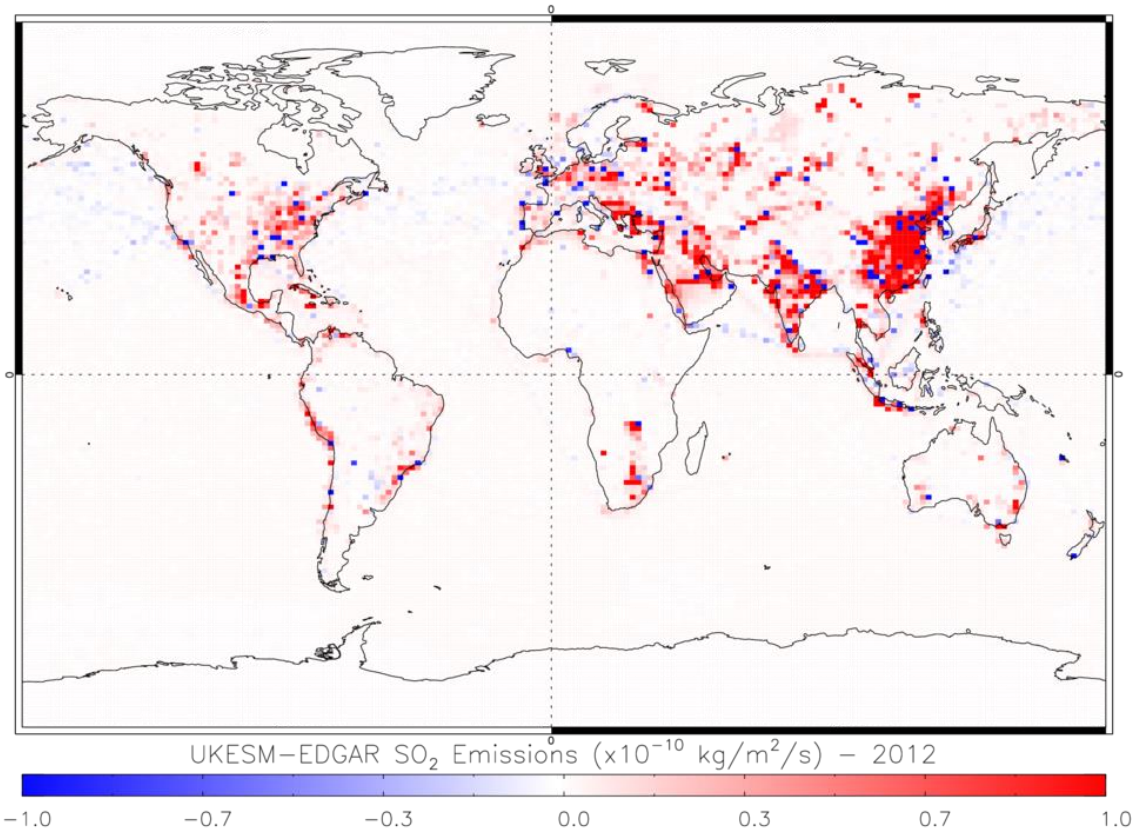

Figure 5: SO_2 emission ($\times 10^{-10}$ kg/m^2/s) annual average differences between CMIP6 and EDGAR for 2012.

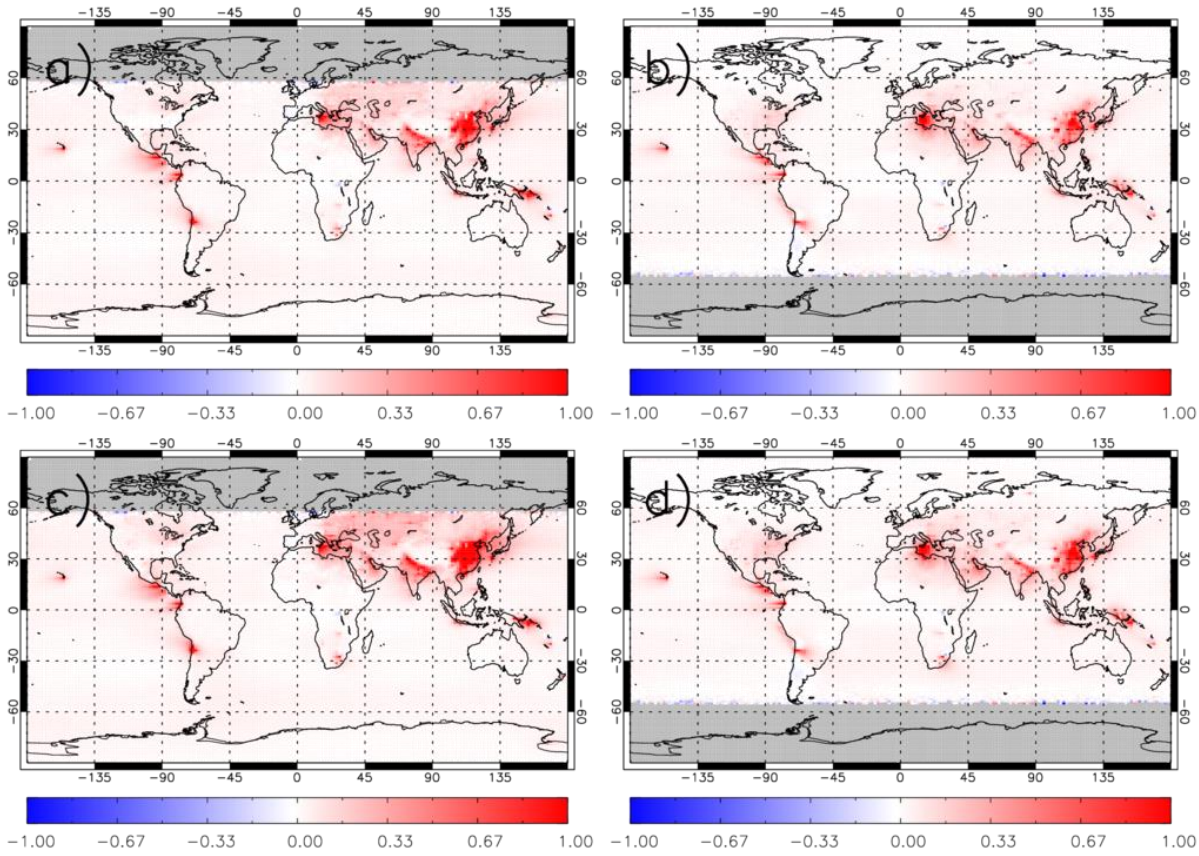

Figure 6: Median $TCSO_2$ (DU) differences for 2005 to 2014 for a) u-bp880 MM – OMI in DJF, b) u-bp880 MM – OMI in JJA, c) u-bw217 MM – OMI in DJF and d) u-bw217 MM – OMI in JJA.

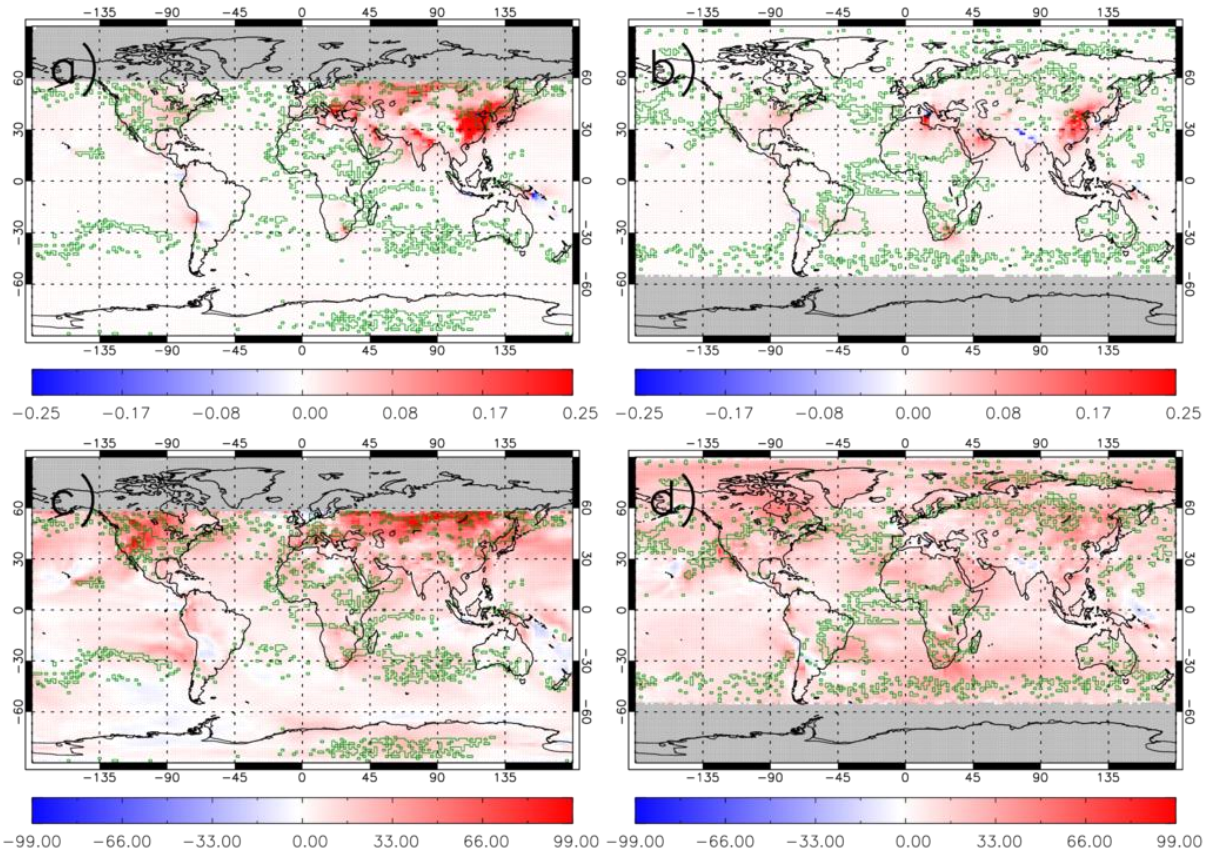

Figure 7: Mean $TCSO_2$ differences (DU) for 2005 to 2014 for a) u-bw217 MM– u-bp880 MM in DJF and b) u-bw217 MM– u-bp880 MM in JJA. c) and d) represent the percentage differences for a) and b), respectively. Green polygon-outlined regions represent where the absolute inter-model output difference is larger than the absolute u-bp880 MM – OMI differences.

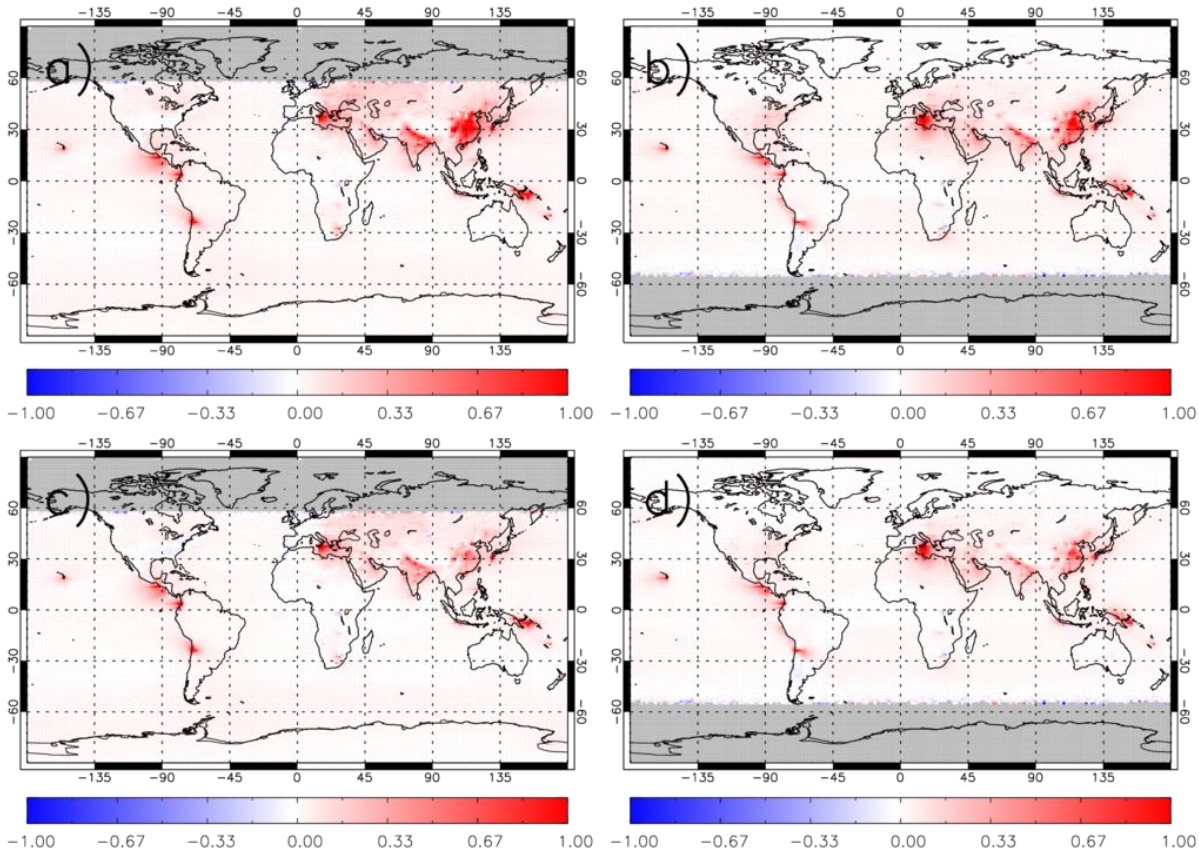

Figure 8: Median $TCSO_2$ (DU) differences for 2005 to 2014 for a) u-bC292 MM – OMI in DJF, b) u- bC292 MM – OMI in JJA, c) u-bK575 MM – OMI in DJF and d) u- bK575 MM – OMI in JJA.

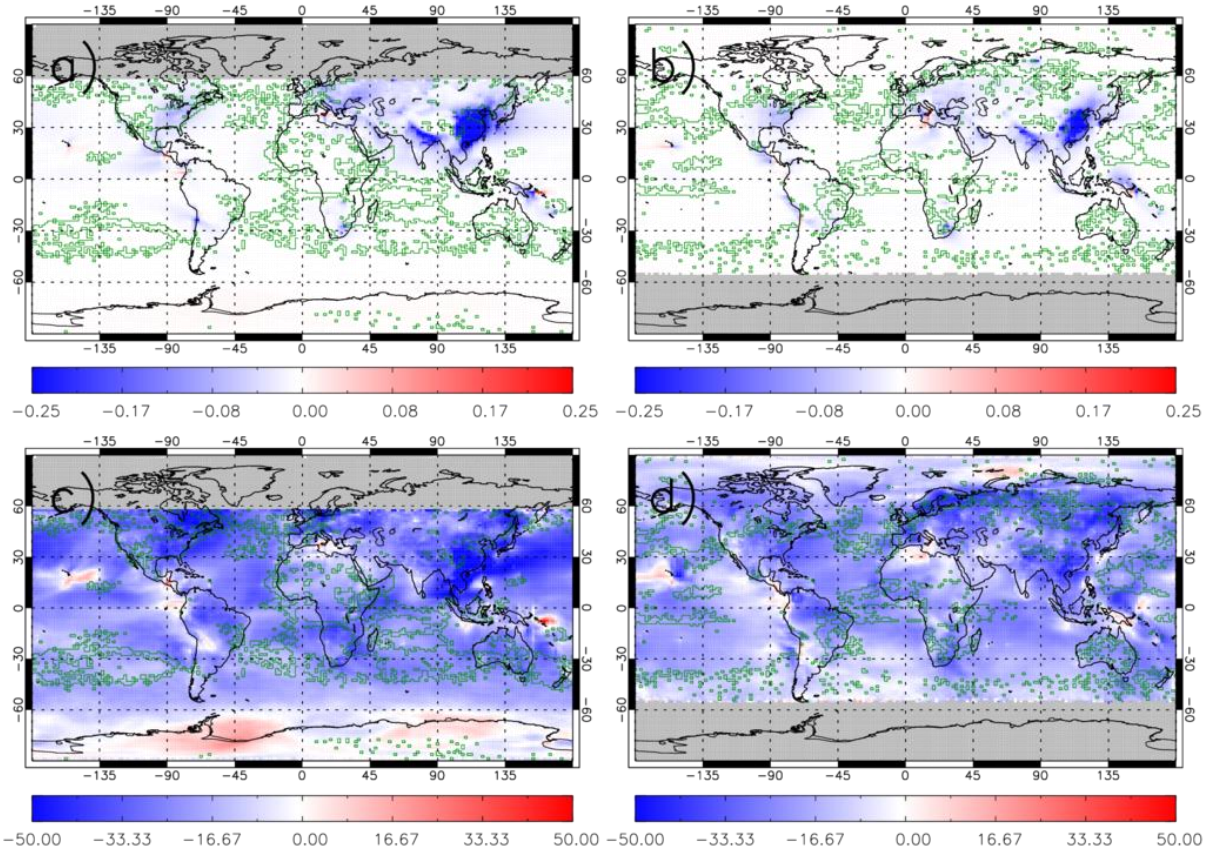

Figure 9: Mean $TCSO_2$ differences (DU) for 2005 to 2014 for a) u-bk575 MM– u-bc292 MM in DJF and b) u-bk575 MM – u-bc292 MM in JJA. c) and d) represent the percentage differences for a) and b), respectively. Green polygon-outlined regions represent where the absolute inter-model output difference is larger than the absolute u-bc292 MM – OMI differences.

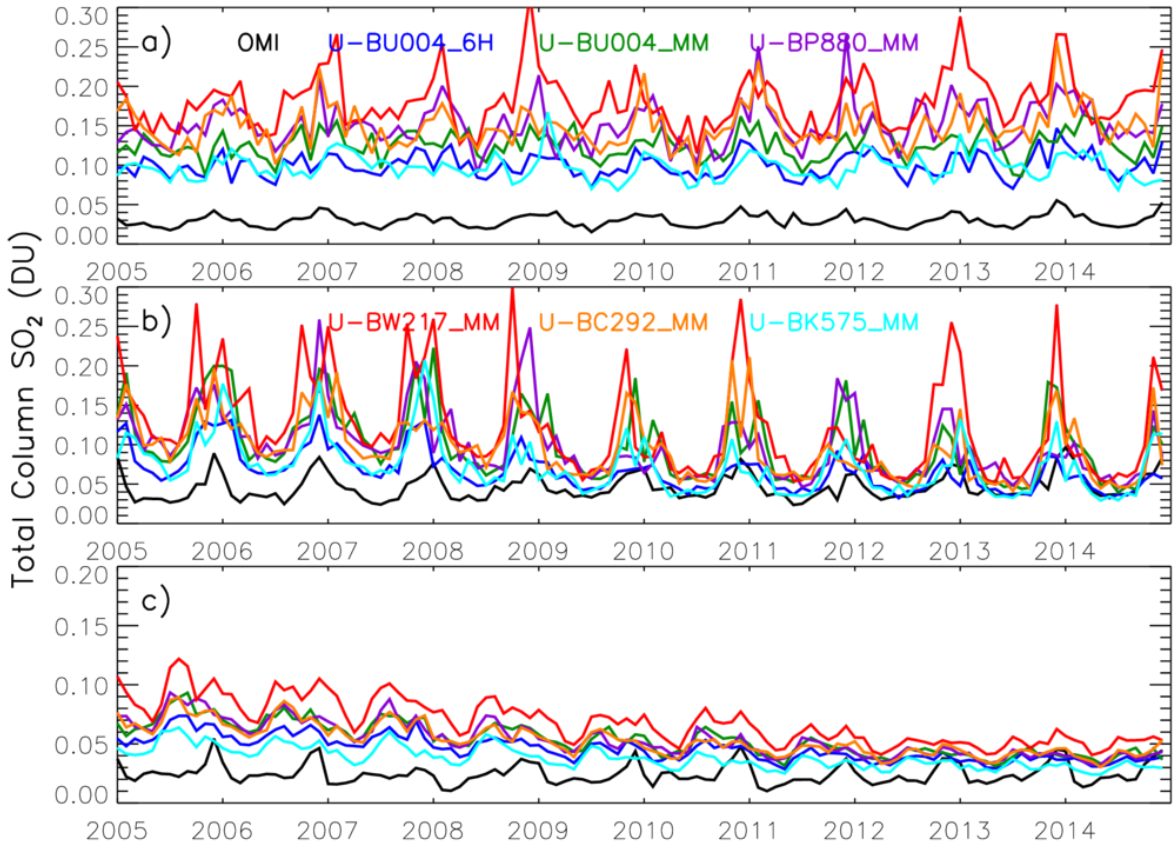

Figure 10: Monthly averaged (median) $TCSO_2$ time-series for OMI (black) and UKESM jobs u-bu004 6H (blue), u-bu004 MM (green), u-bp880 MM (purple), u-bw217 MM (red), u-bc292 MM (orange) and u-bk575 MM (cyan). Panels a), b) and c) represent South East Asia (75-125°E, 20-45°N), Europe (15°W-40°E, 35-65°N) and the USA (60-130°W, 25-50°N), respectively.

Table 1: Summary of the model simulations used in this study.

Experiment	**Job ID**	**Model configuration**
Assessment of temporal model sampling importance for comparisons with satellite data	u-bu004	Nudged UKESM1 AMIP N96 UM11.0 simulation run for the CRESCENDO project – included 6 hourly model output in addition to monthly outputs for SO_2.
Assessment of the impact of SO_2 emission injection height	u-bp880	Free-running UKESM1 AMIP N96 UM11.1 simulation
	u-bw217	As for u-bp880 but with SO_2 emissions injections heights as per GC3.1
Assessment of the impact of modifications to the SO_2 dry deposition parameterization	u-bc292	Historical UKESM1 run using UM10.9
	u-bk575	Historical UKESM1.0-beta run using UM11.2 with modifications to the SO_2 dry deposition parameterization

Table 2: OMI-HTAP (CMIP6) SO_2 annual average (2005-2014) emissions (Tg) totalled into multiple longitude-latitude bands. Green shading shows where CMIP6 > OMI-HTAP and red shading shows where OMI-HTAP > CMIP6.

OMI-HTAP (CMIP6) Climatological (2005-2014) Average SO2 Emissions (Tg)			
Longitude - Latitude Bins	90°-30°S	30°S-30°N	30°-90°N
0°-60°E	0.01 (0.65)	6.59 (9.02)	34.48 (16.5)
60°-120°E	0 (0.09)	10.10 (26.87)	35.29 (26.12)
120°-180°E	2.16 (1.02)	1.76 (2.71)	3.83 (8.47)
180°-120°W	0.00 (0.03)	0.00 (0.31)	0.08 (1.29)
120°-160°W	0.14 (0.32)	1.37 (6.78)	3.61 (8.68)
60°-0°W	0.03 (0.24)	0.50 (2.62)	0.48 (3.11)
Annual Global Total	**CMIP6 = 115 Tg**	**OMI-HTAP = 100 Tg**	

Table 3: Statistics between model runs and OMI $TCSO_2$ monthly averaged (median) time series (2005-2014) for the USA (60-130°W, 25-50°N), Europe (15°W-40°E, 35-65°N) and South East Asia (75-125°E, 20-45°N) - see Figure 10. The metrics include the mean bias (MB, DU), root mean square error (RMSE, DU), percentage mean bias (MB%) and correlation (R).

UKESM Job	Regions	MB	RMSE	MB%	R
u-bu004 6h	USA	0.021	0.024	86.4	0.28
	Europe	0.021	0.031	45.0	0.43
	SE Asia	0.070	0.071	235.2	0.67
u-bu004 mm	USA	0.029	0.033	119.5	0.20
	Europe	0.053	0.065	115.8	0.52
	SE Asia	0.095	0.097	322.8	0.44
u-bp880 mm	USA	0.029	0.033	119.4	0.19
	Europe	0.053	0.065	114.6	0.43
	SE Asia	0.121	0.124	40.7.5	0.57
u-bw217 mm	USA	0.044	0.048	179.5	0.24
	Europe	0.076	0.093	166.7	0.58
	SE Asia	0.152	0.155	510.2	0.67
u-bc292 mm	USA	0.029	0.032	116.5	0.26
	Europe	0.050	0.061	109.6	0.54
	SE Asia	0.120	0.122	402.4	0.64
u-bk575 mm	USA	0.014	0.018	56.1	0.19
	Europe	0.027	0.040	58.7	0.53
	SE Asia	0.068	0.070	227.9	0.48