Preliminary evaluation of the UKESM aerosol component

Adam C. Povey[1] and Roy G. Grainger[1]

[1] *National Centre for Earth Observation, University of Oxford, Oxford, UK*

March 29, 2018

Fields of aerosol optical depth (AOD) simulated by the UK Earth System Model (UKESM) are evaluated against an ensemble of ten satellite datasets. Both daily and monthly averages of the satellite data are considered.

1 Introduction

Aerosols are an important part of the Earth's radiation budget as they interact both directly and indirectly with radiation. Several reports of the IPCC have described them as the greatest source of uncertainty in predictions of the future climate. Numerous modelling centres work to address these uncertainties, brought together by the AeroCom project (Kinne et al., 2006).

Aerosols present a significant observational challenge as they are highly variable in both time and space due to their diverse range of sources and sinks. Passive satellite imagery provides a long-term, global record with which to monitor aerosols. Numerous such datasets have been generated, tabulating variables such as aerosol optical thickness in the visible or absorbing aerosol index in the infrared. However, a nadir image does not fully constrain the observed environment, such that significant assumptions must be made about the properties of the surface and/or composition of the aerosol. Observations at multiple angles improve matters somewhat, but the most direct and accurate measurements of aerosol are provided by the Aerosol Robotic Network (AERONET), a global network of ground-based, automated sun-photometers.

In theory, this plethora of observations provides a sound foundation from which to validate the representation of aerosol in models. However, because of the large spatio-temporal variability, the difference in scale between observations produces significant, unavoidable differences in their results (Povey and Grainger, 2015). AERONET observations are point-like while satellites footprints cover 1–10 km and models average over tens to hundreds of kilometres. The latter is comparable to the scales over which aerosol properties are correlated: 50 km (Anderson et al., 2003), though *in situ* studies (e.g. Shinozuka and Redemann, 2011) have found that can vary by an order of magnitude in either direction. Further, observations can only be made in cloud-free conditions while models provide all-sky averages.

A sequence of papers by Oxford's Climate Processes group have used high-resolution model data to estimate the errors caused by spatial (Schutgens et al., 2016) and temporal (Schutgens, Partridge, and Stier, 2016) sampling or subgrid variability (Weigum, Schutgens, and Stier, 2016). A summary of their work, Schutgens et al. (2017), recommends temporally collocating hourly model data to valid observations (thereby accounting for cloud cover) and averaging observations over 4 hr and 110 km before comparison.

Consideration of such errors is vital when thoroughly evaluating the performance of an aerosol model. However, it is a burden for routine work: the storage requirements for hourly model output are substantial and models must be nudged for their output to resemble real-world observations. Though an in-depth validation of the UKESM aerosol component will be pursued eventually, it is also important to consider how the model is evaluated during day-to-day development work using monthly averages of AOD for a few free-running years (by, for example, ESMValTool).

This report discusses our thoughts on aerosol model evaluation and presents some initial results. It is hoped to be part of an ongoing discussion.

2 Current work

When working with free-running models, there is no reason to expect the output to exactly replicate a particular set of observations. Collocating the model to data could represent a satellite's spatial sampling but won't change the disconnect between the underlying measurands. As such, we aim to approach model evaluation statistically — evaluating how likely a given model output field was, given the satellite observations. That requires assessing the probability density function (PDF) of AOD globally (a.k.a. generating a climatology).

Seasonal (rather than monthly) climatologies were explored as the three-fold increase in data was found to improve performance. For brevity, only winter (DJF) is shown here with summer versions of each plot gathered in the Appendix. Data from ATSR-2, AATSR, MODIS Terra, MISR, and SeaWiFS are presented. For the ATSR instruments, three different realisations are provided: ORAC from Oxford University, SU from Swansea University, and ADV from the Finnish Meteorological Institute. For MODIS, the Dark Target and Deep Blue algorithms are considered separately. (Note that SeaWiFS uses the Deep Blue algorithm.) All available years are considered for each instrument except daily MISR, for which only 2008 and 2015 were available due to storage constraints. Comparison against AERONET will be explored in the future.

2.1 Comparison of means

The most common method of aggregating data is the simple mean. Accompanied by a standard deviation, it represents normally distributed data. The monthly mean fields output by the model and those distributed as Level 3 data can easily be combined to produce seasonal averages. This is done by many of the current evaluations, comparing the model to the merged MODIS product or AERONET.

Figure 1 shows the difference, δ, between the model's 1995–2008 winter mean and the full-mission winter mean for the ten satellite records. The diversity of the satellite data is clear, though the spatial patterns share many features. Mean δ are 0.016–0.075. Modal δ are all positive but are close to zero for ORAC, ADV, and DT. The distribution of δ is positively skewed for all satellites except MODIS.

Figure 2 combines the ten satellite means to give a single map of δ. The model underestimates AOD over the Sahara and continental land while overestimating over the African biomass-burning region and east Asia. As the datasets differ substantially in their estimates of remote ocean AOD, the model's performance over sea is broadly fine except for the outflows from central America and India. Excluding mentioned regions, the distribution of δ is fairly normal with a half-width of ~0.03.
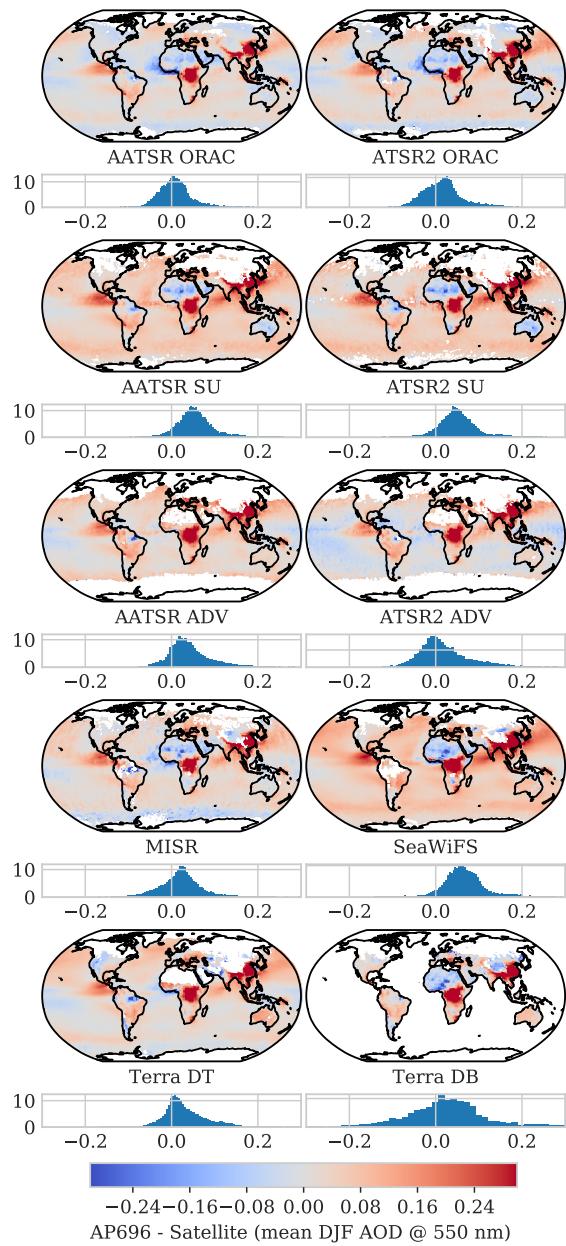

Figure 1: *Difference between the winter mean AOD from model configuration AP696 and ten satellite datasets. The histogram of differences is shown below each plot.*
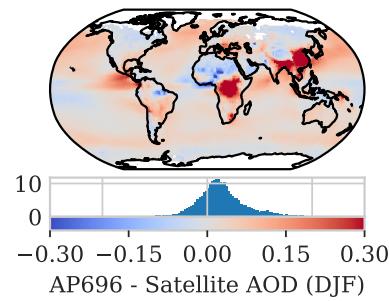
Figure 2: *As Fig. 1 but averaging the ten satellite datasets.*
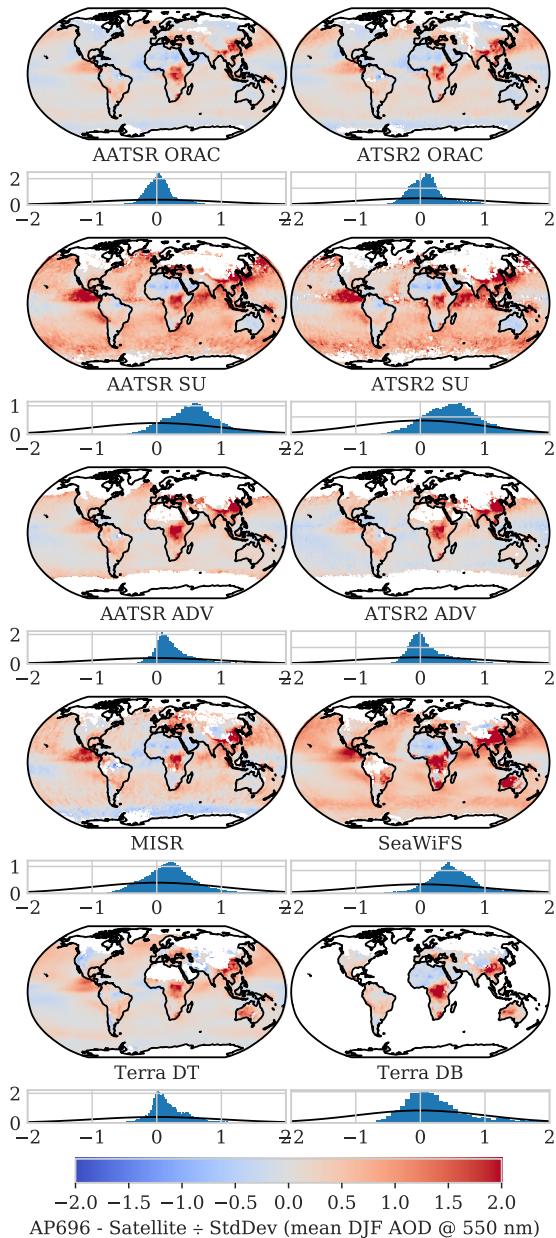
Figure 3: *As Fig. 1 but normalising the difference by the satellite data's standard deviation. The $\mathcal{N}(0,1)$ distribution shown on each histogram for reference.*

AP696 - Satellite AOD ÷ StdDev (DJF)
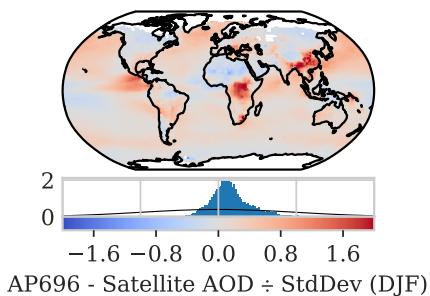
Figure 4: *As Fig. 3 but averaging the ten satellite datasets.*

To be clear, Fig. 2 should be considered qualitative rather than quantitative. A simple mean of all satellite observations does not account for differences in the volumes of data produced, the spatio-temporal sampling, nor the algorithms' systematic errors. There is colloquial evidence (M. Schultz, personal communication, 2017) that the median of satellite datasets compares quite favourably to AERONET, and this will be investigated in future.

Figures 3 and 4 show the same difference as the previous plots but normalised by the satellite's standard deviation (a difference we denote by Δ) to provide a globally consistent scaling. If differences between the model and satellites were purely due to natural variation, Δ would be a normal distribution with zero mean and unit width. The actual distributions are narrower, which is expected as the variability of instantaneous 1 km AOD will certainly exceed that of model grid cells.

Comparisons such as these have been used to argue that the UKESM model fields are broadly sensible. Though easy to perform, these are the manner of comparisons that Schutgens et al. wished to discourage as sampling differences make them difficult to interpret. For example, the difference over the biomass burning region could result from temporal sampling (e.g. the satellites overpass between 10.30am and 2.00pm; if fires tend to be lit in the evening, they won't be observed but will impact the model data). Significant variations can be lost within such multi-year means and offsets between the satellite datasets make it difficult to establish a model bias as, in any given location, the model tends to be close to at least one dataset.

2.2 Monthly distribution at $1° \times 1°$

Alternatively, we can compare the observed distribution of monthly mean AOD to that modelled. Two examples are shown in Figure 5. Each pane shows a grid of 7×5 plots over a map. Each sub-plot shows the distribution of daily mean AOD observed in that $1° \times 1°$ region by each of the ten datasets. (Each tick is 10 % frequency along the y-axis and 0.2 units of AOD along the x-axis.) The mean of each distribution (not shown) tends to be larger than the mode as the distributions are positively skewed.

The top image is off the southwestern coast of India, where all datasets agree the model overestimates AOD. The satellites observe broadly similar distributions near the land (somewhat surprisingly) but increasingly diverge over the sea. MODIS and MISR report narrow peaks centred around 0.1 while the other instruments see broad distributions, with AOD > 0.2 being relatively common. ORAC (blue) tends to report the largest values. The bottom image in Fig. 5 inspects a region near Cambodia found to have a close agreement between the satellite datasets. The spread of distributions is smaller than near India, but the difference between MISR/MODIS and the other datasets remains.
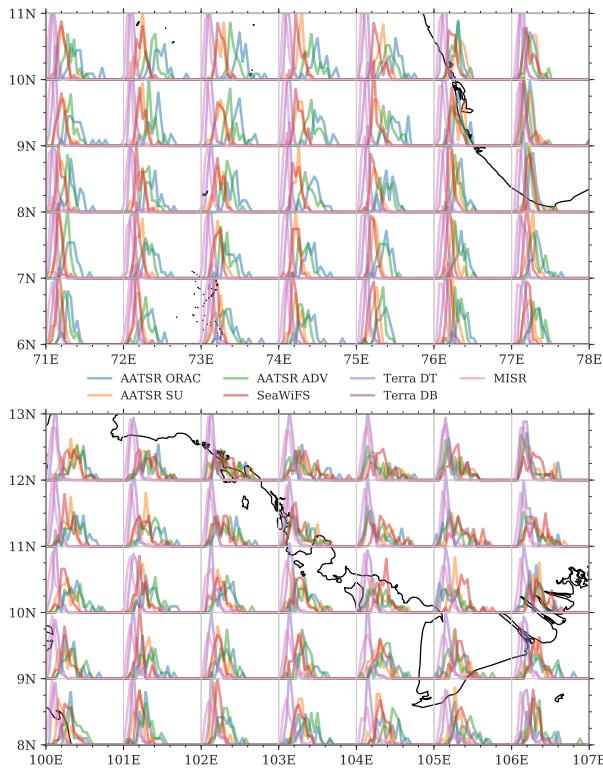
Figure 5: *Distribution of winter monthly mean AOD near southwestern India (top) and Cambodia (bottom). Axis labels give latitude and longitude. Each $1° \times 1°$ grid cell displays the frequency of AOD in 25 bins between 0 and 1 with $1°$ of latitude corresponding to a relative frequency of 40 %. Colour denotes the satellite as defined in the legend.*

Figure 6: *Comparison of model configuration AP696's monthly mean AOD to the ensemble of ten satellite distributions, on the axes of Fig. 5. The ensemble spread is plotted in light blue, with the ensemble mean in dark blue. The frequency of model monthly mean is in orange.*

This ensemble of frequency distributions is compared to model output in Figure 6. The mean of the distributions from the ten satellites is shown in blue, with the min-max range shaded. The distribution of model monthly mean is shown in orange. In the top frame, we see that the model's overestimation of AOD results from a systematic lack of low-AOD events, though the shape of the distribution is approximately correct. Over sea in the bottom frame, the model reproduces the observed distribution quite well, though it still lacks AOD < 0.1. Over land, the model produces AOD > 0.3 events that do not appear in the data. (The distributions from model run AP680, Fig. B.3, are very similar so this is unlikely due to the tuning of LAI_min_io.)

Most of the model distributions are rather noisy, which is a sign that more model years should be considered. 1995–2008 were used to correspond to the satellite data record. Adding 2009–2017 is an easy addition, while earlier years may sample different emission scenarios.

2.3 Daily distribution at $1° \times 1°$

As stated in the Introduction, the difference between the observed and modelled distributions of AOD could result from differences in sampling. For example, the prevalence of low AOD in the observations could be caused by thick aerosol events being mistaken for cloud. Most satellite products provide daily Level 3 data, which is closer to the averaging scales recommended by Schutgens et al. (2017). The distribution of daily mean AOD should give a better understanding of short-term variability and help to interpret the differences described in the previous section.

Figure 7 shows the distribution of daily mean AOD in the satellite datasets. It is strikingly different to Fig. 5 in that all of the datasets now report similar distributions, especially over Cambodia. ORAC (blue) still reports large AOD most frequently while SeaWiFS (red) reports small AOD most frequently. The relatively small number of MISR observations (pink) produces noisy distributions, which will be addressed when a more compact product is downloaded. The broadening of the distributions from MODIS and MISR is thought to result from their averaging technique: They calculate monthly means by averaging daily means in an effort to minimise clear-sky bias (i.e. a single clear day could provide hundreds of observations that would swamp a handfull of observations from the remainder of a cloudy month). The reduced spread of daily data provides a clearer comparison to the model in Fig. 8.
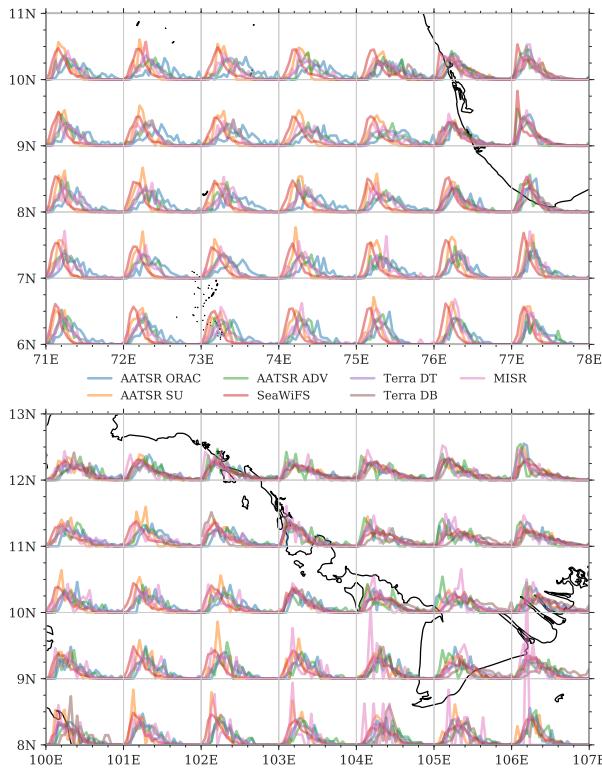
Figure 7: *As Fig. 5, but considering the distribution of daily means for the satellite. (The model data is unchanged.)*

Figure 8: *As Fig. 6, but considering the distribution of daily means for the satellite. (The model data is unchanged.)*

The figure also clearly shows that AOD conforms to a log-normal distribution (as described in, e.g. O'Neill et al. (2000)). In that case, it would be better to average $\ln$ AOD. No data provider currently does. In theory, we could produce our own Level 3 data but this would require significant time (e.g. to download the NASA products). Especially considering behaviour in summer (Fig. A.6), the modelled distribution of AOD appears to be too sharply peaked around 0.15. Without higher time resolution output, though, it is difficult to say why.

A simple metric to assess the agreement between the model and satellite distributions is the root-mean-square (RMS) difference. The top of Fig. 9 shows this, with the bottom normalising the difference by the spread of the ensemble (coarsely defined here as $\frac{1}{2}[\max - \min]$). Absolute differences are small in dusty or monsoon regions and are high over the western US and eastern Australia. Relative differences resemble Fig. 2.

However, these fields require some care to interpret. Consider the Asian peninsula (which was shown at the bottom of Fig. 7). The absolute differences are small but the relative differences are large. This follows from the model occasionally observing large AOD never seen in the observations. The absolute difference in distribution is small but, because the ensemble spread is small, the relative difference is large. Conversely, the southern Atlantic shows large absolute differences, but these aren't relatively important because the spread

of data is larger. In fact, the different spatial patterns between the two panes of Fig. 9 indicate that the model is accurately reproducing the variability of AOD across most of the planet. Comparing this model version to its predecessor (Fig. B.5), the newer version slightly reduces the difference between model and satellite.

3   Discussion

A preliminary evaluation of the UKESM aerosol component has been outlined. Based on this analysis, we would say that the model is behaving realistically. AOD is slightly overestimated, possibly due to a lack of clean air (i.e. low AOD events), and highly overestimated in a few regions. The tuning of LAI_min_io has improved the model, if only slightly.

Comparing monthly means, the modelled AOD is slightly high, with biases over Africa and southeastern Asia. Looking at the distribution of monthly means, it appears the model fails to include events with AOD < 0.1. However, there is disagreement in the satellite datasets considered as to the frequency of such events.

The datasets are in agreement if daily means are considered instead, highlighting the importance of using suitable spatio-temporal scales when comparing models to satellite observations. A simple metric, the root-mean-square difference, was used to quantify the model's discrepancy. That particular plot was produced

AP696) DJF
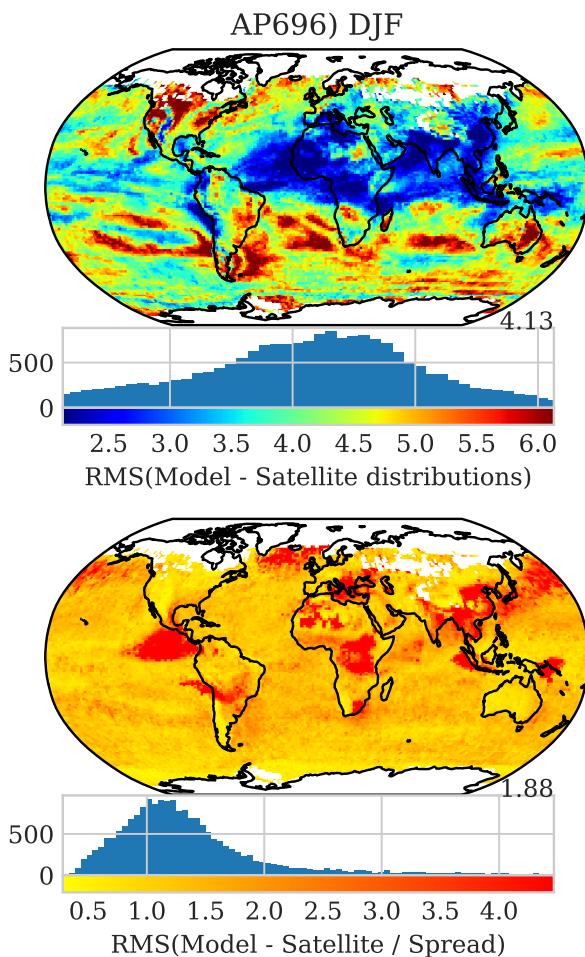
Figure 9: *RMS difference between the modelled distribution of AOD and the satellite ensemble mean. The absolute difference is shown on the top plot, which is normalised by the ensemble spread in the bottom plot. The global mean of each field is given to its right, with the histogram shown below.*

mere hours before this report was distributed, so feedback would be appreciated on how to best consolidate and communicate those statistics.

Looking forward, the goal is to produce a simple metric from these analyses that can be included within ESMValTool. This should, in theory, be straightforward as all of this work was done in Python — the format necessary to integrate with ESMValTool. A third view of the satellite data has been processed but not yet analysed — 10 km spatial resolution. The daily/monthly difference shown above explored the impact of temporal averaging. The hope is to use the 10 km statistics to explore the impact of spatial averaging.

Broadly, though we are clearly fond of histograms, they are awkward to communicate and use. We intend to simplify this using some parametric statistics: fit a log-normal curve to the distributions shown. This should streamline comparisons and present a better understanding of the data, but will likely require making our own Level 3 data, which will be time consuming.

There is also the possibility of using gappy EOF analysis to evaluate the modes of spatial variation in the data (as all of the analyses presented implicitly treat each pixel as independent, which is plainly untrue).

Otherwise, we are eager for feedback on the work presented and look forward to a high temporal resolution model run to come in the future.

References

Anderson, T L et al. (2003). "Mesoscale Variations of Tropospheric Aerosols". In: *Journal of the Atmospheric Sciences* 60.1, pp. 119–136. DOI: 10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2.

Kinne, S. et al. (2006). "An AeroCom initial assessment – optical properties in aerosol component modules of global models". In: *Atmospheric Chemistry and Physics* 6.7, pp. 1815–1834. DOI: 10.5194/acp-6-1815-2006.

O'Neill, N. T. et al. (2000). "The lognormal distribution as a reference for reporting aerosol optical depth statistics; Empirical tests using multi-year, multi-site AERONET Sunphotometer data". In: *Geophysical Research Letters* 27.20, pp. 3333–3336. DOI: 10.1029/2000GL011581.

Povey, A C and R G Grainger (2015). "Known and unknown unknowns: uncertainty estimation in satellite remote sensing". In: *Atmospheric Measurement Techniques* 8, pp. 4699–4718. DOI: 10.5194/amt-8-4699-2015.

Schutgens, N. A. J., D. G. Partridge, and P. Stier (2016). "The importance of temporal collocation for the evaluation of aerosol models with observations". In: *Atmospheric Chemistry and Physics* 16.2, pp. 1065–1079. DOI: 10.5194/acp-16-1065-2016.

Schutgens, N. A. J. et al. (2016). "Will a perfect model agree with perfect observations? The impact of spatial sampling". In: *Atmospheric Chemistry and Physics* 16.10, pp. 6335–6353. DOI: 10.5194/acp-2015-973.

Schutgens, Nick et al. (2017). "On the spatio-temporal representativeness of observations". In: *Atmospheric Chemistry and Physics* 17, pp. 9761–9780. DOI: 10.5194/acp-17-9761-2017.

Shinozuka, Y. and J. Redemann (2011). "Horizontal variability of aerosol optical depth observed during the ARCTAS airborne experiment". In: *Atmospheric Chemistry and Physics* 11.16, pp. 8489–8495. DOI: 10.5194/acp-11-8489-2011.

Weigum, Natalie, Nick Schutgens, and Philip Stier (2016). "Effect of aerosol subgrid variability on aerosol optical depth and cloud condensation nuclei: Implications for global aerosol modelling". In: *Atmospheric Chemistry and Physics* 16.21. DOI: 10.5194/acp-16-13619-2016.

A   AP696 in summer

AATSR ORAC   ATSR2 ORAC

AATSR SU   ATSR2 SU

AATSR ADV   ATSR2 ADV

MISR   SeaWiFS

Terra DT   Terra DB

AP696 - Satellite (mean JJA AOD @ 550 nm)
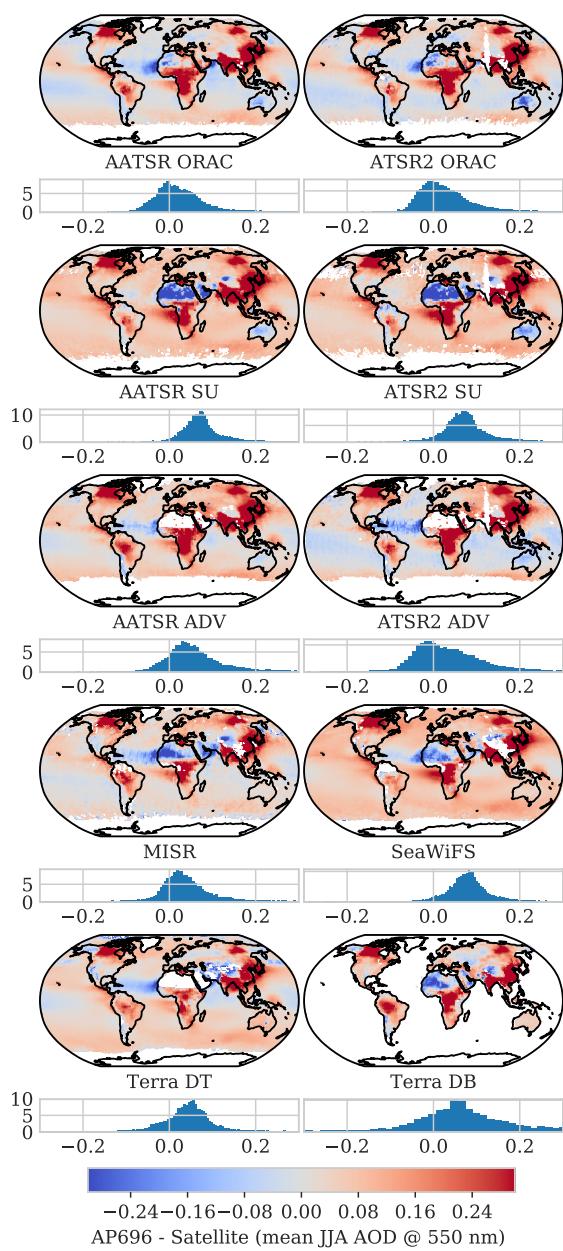
Figure A.1: *As Fig. 1 but for JJA.*

AP696 - Satellite AOD (JJA)
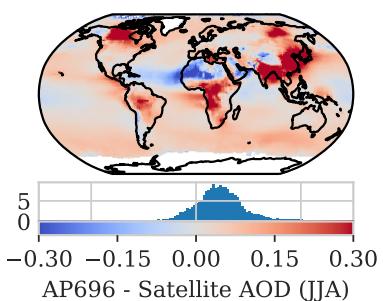
Figure A.2: *As Fig. 2 but for JJA.*

AATSR ORAC   AATSR ADV   Terra DT   MISR
AATSR SU   SeaWiFS   Terra DB
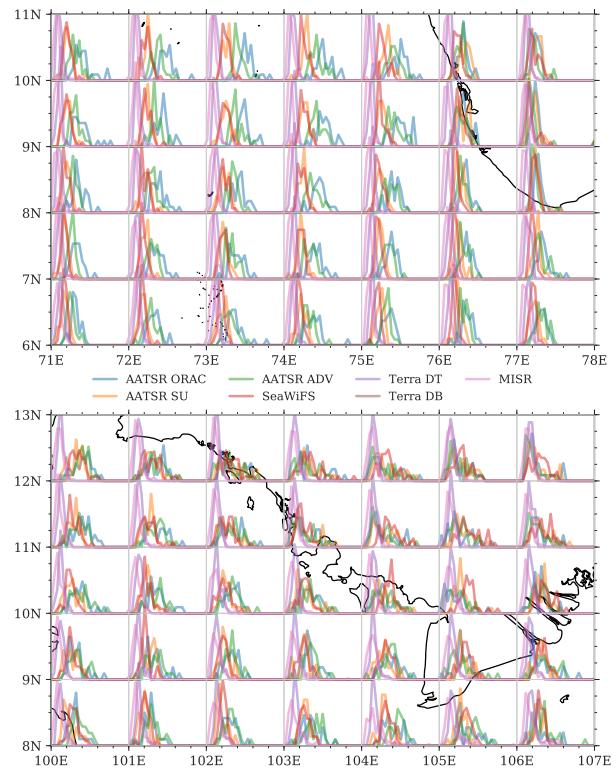
Figure A.3: *As Fig. 5 but for JJA.*

AP696   Monthly ensemble mean   Monthly ensemble spread
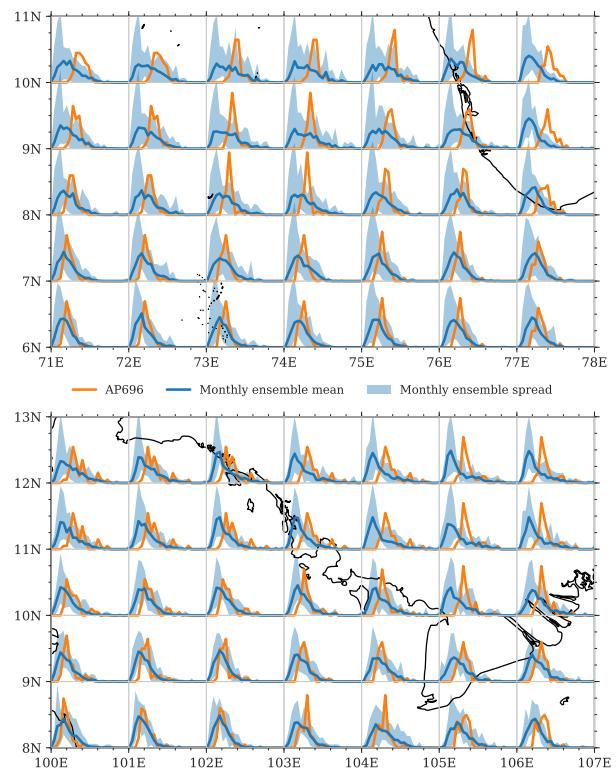
Figure A.4: *As Fig. 6 but for JJA.*
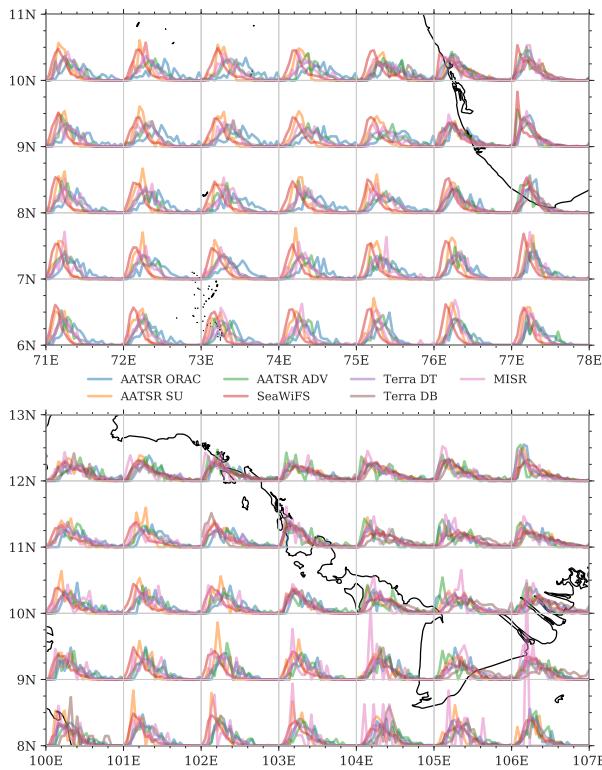
Figure A.5: *As Fig. 7 but for JJA.*
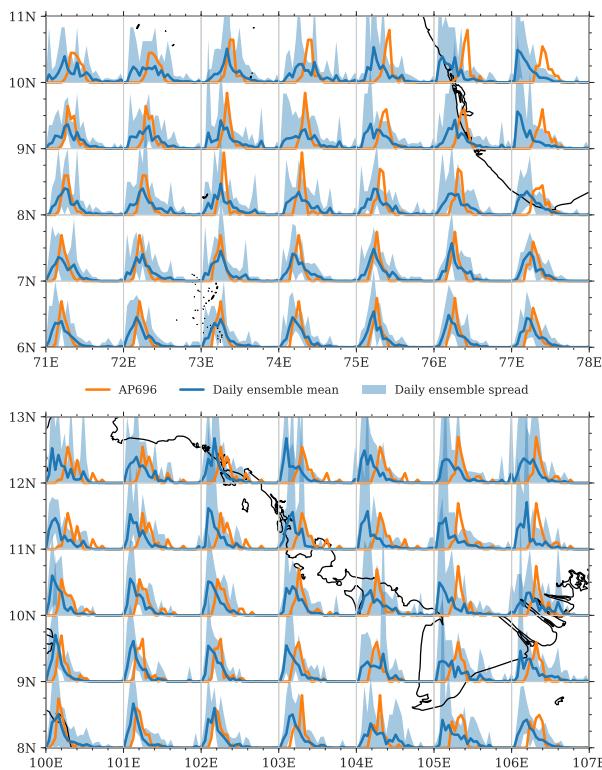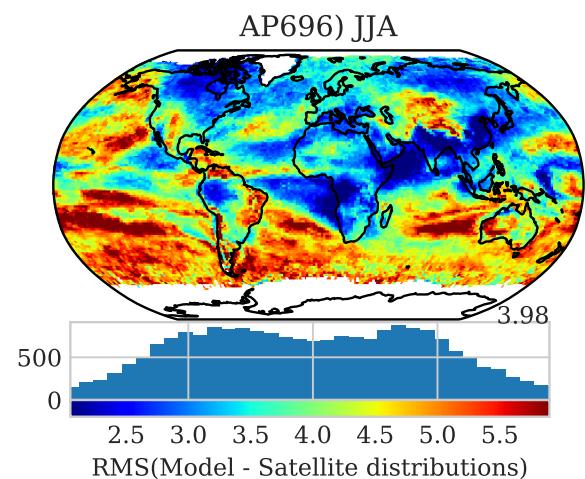
Figure A.6: *As Fig. 8 but for JJA.*

RMS(Model - Satellite distributions)

RMS(Model - Satellite / Spread)
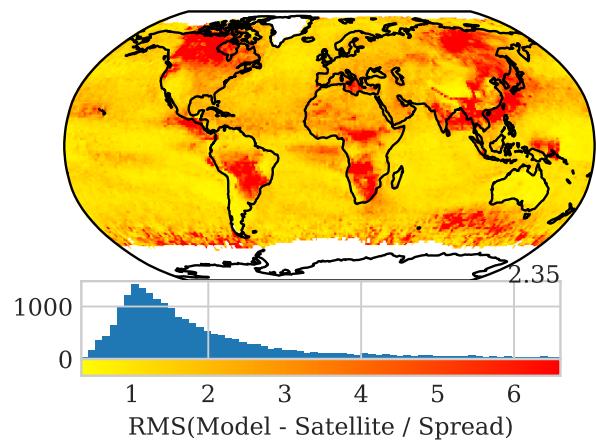
Figure A.7: *As Fig. 9 but for JJA.*

B   AP680 in winter
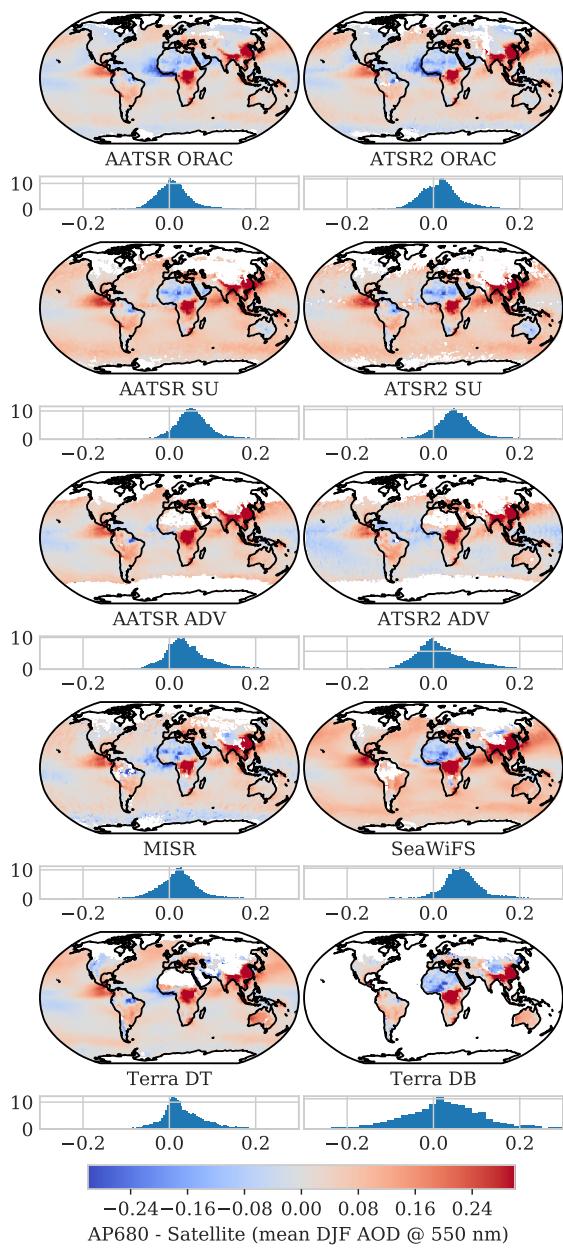
Figure B.1: *As Fig. 1 but for AP680.*
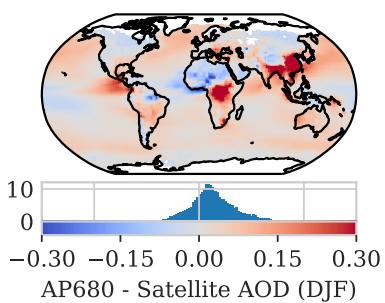
Figure B.2: *As Fig. 2 but for AP680.*
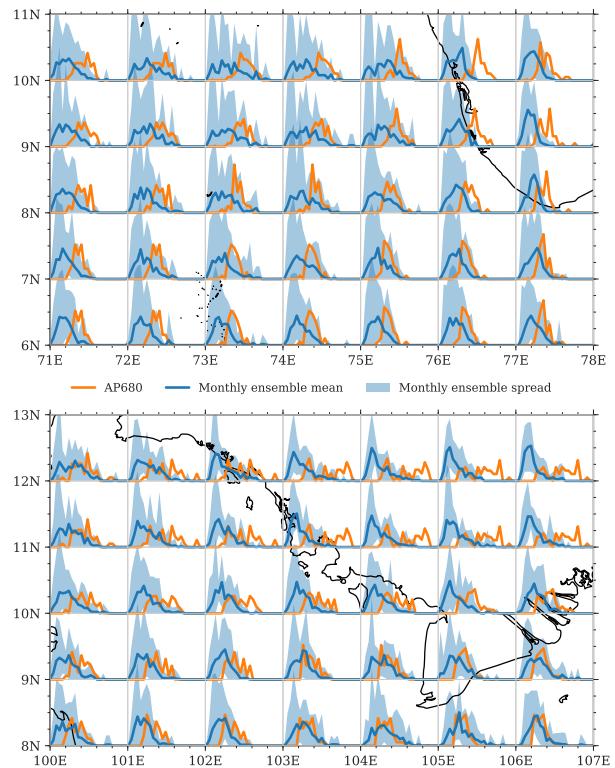

Figure B.3: *As Fig. 6 but for AP680.*
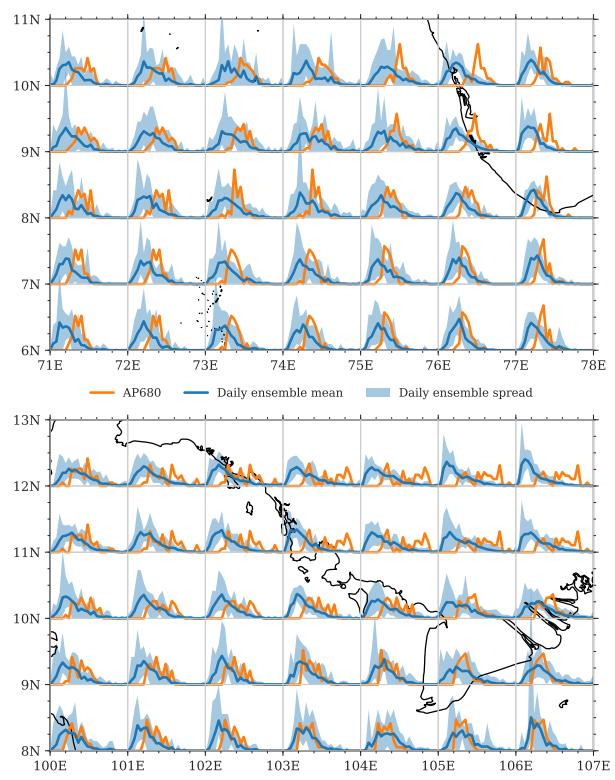
Figure B.4: *As Fig. 8 but for AP680.*
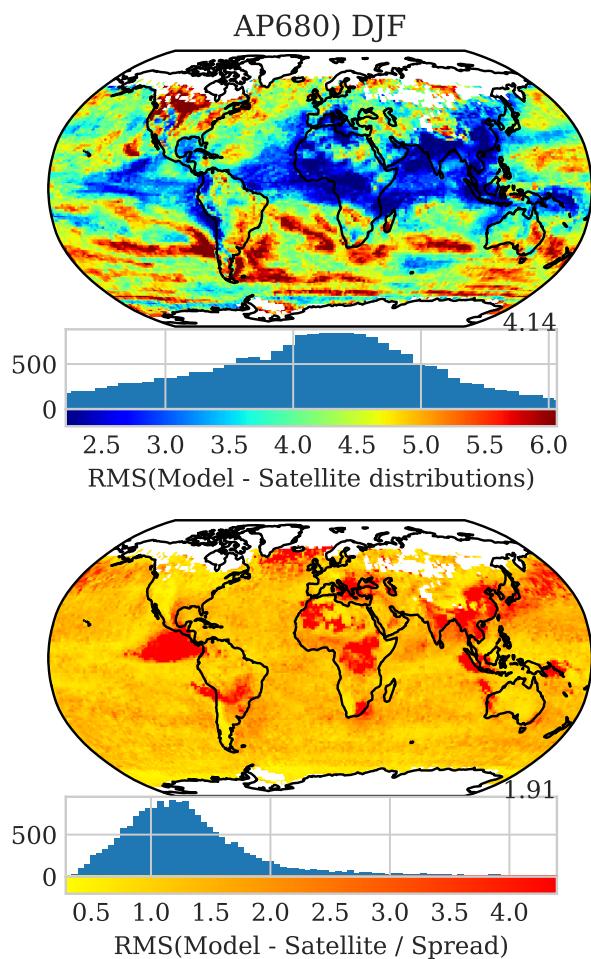
**Figure B.5:** *As Fig. 9 but for AP680.*

C AP680 in summer
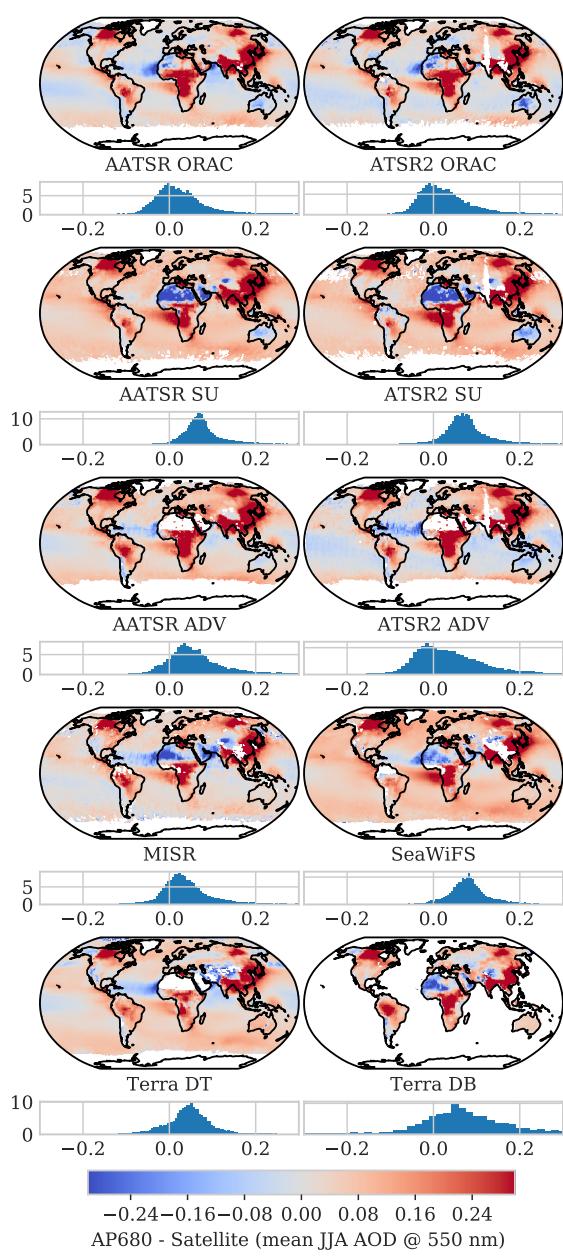
**Figure C.1:** *As Fig. 1 but for AP680 in JJA.*
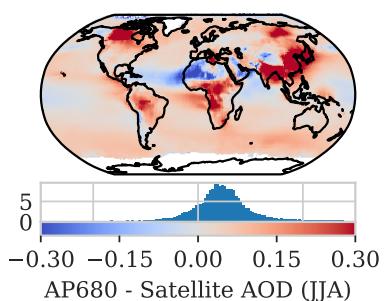

Figure C.2: *As Fig. 2 but for AP680 in JJA.*
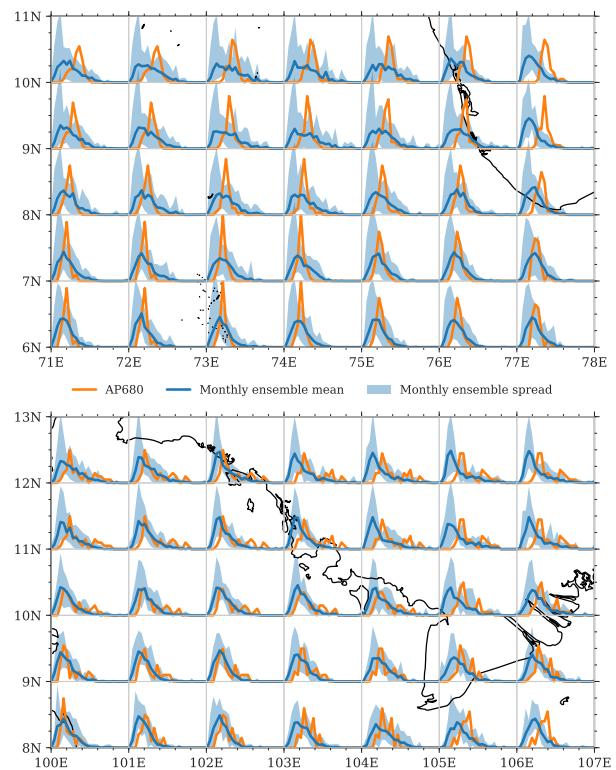
**Figure C.3:** *As Fig. 6 but for AP680 in JJA.*
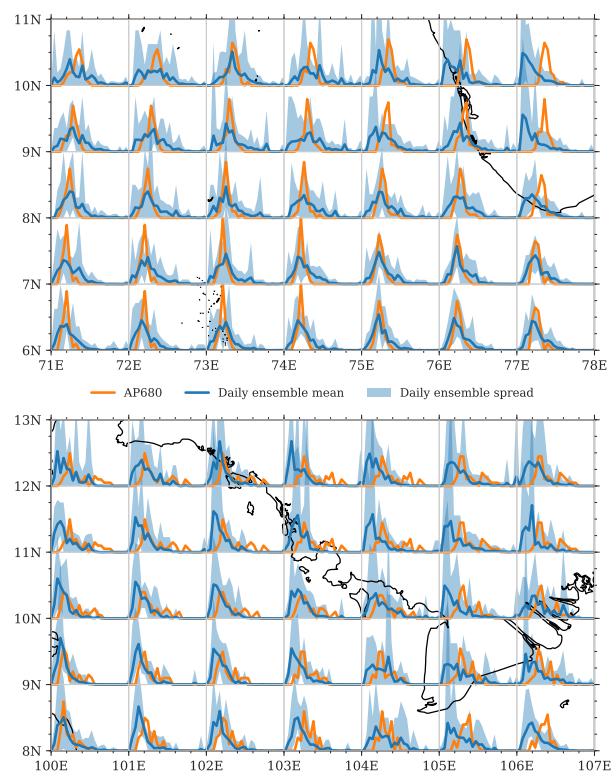
Figure C.4: *As Fig. 8 but for AP680 in JJA.*
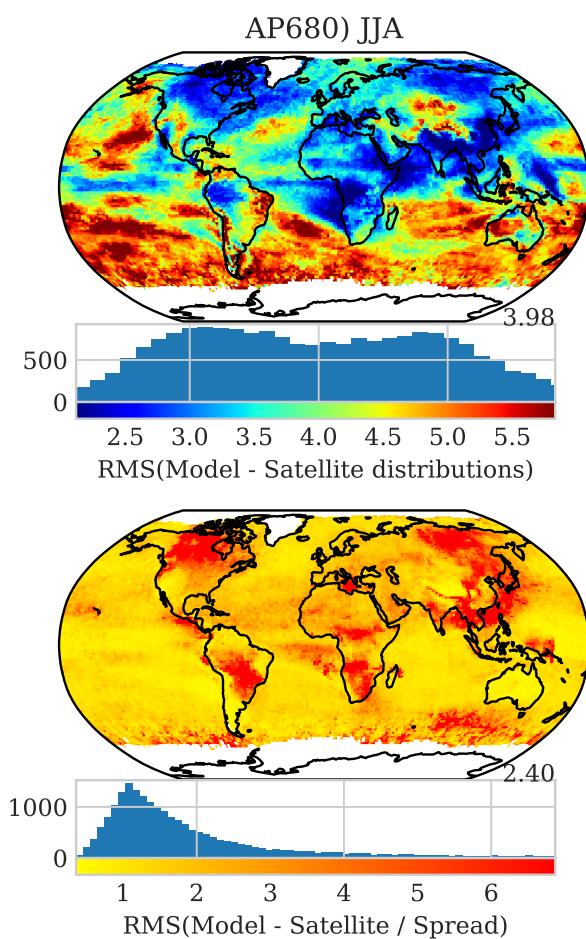
Figure C.5: *As Fig. 9 but for AP680 in JJA.*