The GERB Obs4MIPs Radiative Flux Dataset: A new tool for climate model evaluation

Jacqueline E. Russell[1], Richard J. Bantges[2], Helen E. Brindley[2], Alejandro Bodas-Salcedo[3]

[1]Physics Department, Imperial College, London, SW7 2BX, UK
5 [2]NERC National Centre for Earth Observation, Imperial College, London, SW7 2BX, UK
[3]Met Office Hadley Centre, Exeter, EX1 3PB, UK

Correspondence to: Jacqueline E. Russell (j.e.russell@imperial.ac.uk)

Abstract. A new radiative flux dataset, specifically designed to enable the evaluation of the diurnal cycle in top of the atmosphere fluxes, as captured by climate and Earth-system models is presented. Observations over the period 2007-2012
10 made by the Geostationary Earth Radiation Budget (GERB) instrument are used to derive monthly hourly mean reflected shortwave (RSW) and outgoing longwave fluxes (OLR) on a regular 1°x1° latitude-longitude grid covering 60° N-60° S and 60° E-60° W. The impact of missing data is evaluated in detail, and a data-filling solution is implemented using estimates of the broadband fluxes from the Spinning Enhanced Visible and Infrared Imager, flying on the same Meteosat platform, scaled to the GERB observations. This relatively simple approach is shown to deliver an approximate factor of ten improvement in
15 both the bias caused by missing data and the associated variability in the error. To demonstrate the utility of this GERB 'obs4MIPs' dataset, comparisons are made to radiative fluxes from two climate configurations of the Hadley Centre Global Environmental model: HadGEM3-GC3.1 and HadGEM3-GC5.0. Focusing on marine stratocumulus and deep convective cloud regimes, diurnally resolved comparisons between the model and observations highlight discrepancies between the model configurations in terms of their ability to capture the diurnal amplitude and phase of the top of atmosphere fluxes: details that
20 cannot be diagnosed by comparisons at lower temporal resolution. For these cloud regimes the GC5.0 configuration shows improved fidelity with the observations although notable differences remain. The GERB Obs4MIPs monthly hourly TOA fluxes are available from the Centre for Environmental Data Analysis with the OLR fluxes accessible at https://doi.org/10.5285/90148d9b1f1c40f1ac40152957e25467 (Bantges et al. 2023a) and the RSW fluxes at https://doi.org/10.5285/57821b58804945deaf4cdde278563ec2 (Bantges et al. 2023b).

25 # 1 Introduction

The Geostationary Earth Radiation Budget (GERB) experiment is designed to measure the top of the atmosphere broadband emitted thermal and reflected solar fluxes at high time resolution (Harries et al., 2005). Four GERB instruments have been deployed sequentially on the four Meteosat Second Generation satellites (Meteosat-8, 9 10 and 11). Since May 2004 they have provided top of atmosphere (TOA) outgoing longwave (OLR) and reflected solar (RSW) flux products broadly covering the
30 geographical region 60º E to 60º W and 60º N to 60º S at a 15-minute temporal resolution. The frequency and longevity of the

observations enables the diurnal cycle to be resolved and facilitates the study of fast climate processes, such as cloud and aerosol, by quantifying their changing effect on the radiation balance over a range of timescales from minutes to years. To date the instantaneous GERB products have been used to study and characterise diurnal variability (e.g. Comer et al. 2007, Gristey et al. 2018), the effects of cloud and aerosol on the radiation budget (e.g. Futyan et al. 2005, Slingo et al. 2006, Brindley and

35 Russell 2009, Pearson et al. 2010, Ansell et al. 2014, Banks et al. 2014) and to evaluate the representation of these processes in selected numerical weather prediction and climate models (e.g. Allan et al. 2007, Greuell et al. 2011, Haywood et al. 2011, Mackie et al. 2017).

While the instantaneous data have been extensively exploited, they are not currently provided in a format that facilitates easy comparison with climate or Earth-system model output. In particular, they suffer from irregular spatial sampling, have a

40 temporal resolution that is higher than that at which model radiation outputs are typically retained, and have a non-standard data format. This paper describes the production of a new monthly hourly mean data product, derived from the instantaneous GERB data, to circumvent these issues. This GERB 'obs4MIPs' dataset consists of monthly hourly mean TOA RSW and OLR fluxes provided at a one degree spatial resolution for the GERB observation region. It aims to provide a climatological record covering several years that resolves the diurnal variation in the TOA RSW and OLR and is compatible with climate model

45 output such as that produced for the recent Coupled Model Intercomparison Project 6 (CMIP6) (Eyring et al. 2016). The data are provided in CF compliant netCDF format meeting the Observations for Climate Model Intercomparison Projects (Obs4MIPs) submission requirements (Waliser et al. 2020). In the following sections we outline the basic methodology and provide a detailed analysis of the impact of missing data. We propose and evaluate a relatively simple approach to fill data gaps before providing an illustration of how the new dataset may be employed to assess climate model performance.

50 ## 2 Production of the GERB Obs4MIPs monthly hourly average products

2.1 Baseline Methodology

The GERB Obs4MIPs RSW and OLR fluxes discussed here are based on the observational record from the GERB-1 instrument on Meteosat-9, which runs from May 2007 to January 2013. As noted above, the goal is to create monthly mean, diurnally resolved RSW and OLR fluxes at hourly resolution on a regular 1°x1° latitude-longitude grid.

55 The starting point for creating the averages are the GERB level 2 High Resolution (HR) flux products (Brindley and Russell, 2017) which are produced to facilitate averaging and re-gridding of the GERB instantaneous fluxes. GERB HR fluxes are presented on a regular viewing angle grid which has a spatial resolution of 9 km at the sub-satellite point. They give a 'snapshot' of the fluxes at a 15-minute temporal resolution, aligned to the observation times of each Spinning Enhanced Visible and Infrared Imager (SEVIRI) flying on the Meteosat 2nd Generation series (Schmetz et al 2002).

60 The GERB instrument operates with the use of a rotating mirror which effectively steps the linear detector array, aligned approximately north-south with respect to the Earth, from east to west and then west-east across the Earth's disc. Early in the mission the mirror briefly became stuck in a position which allowed direct solar illumination of a portion of the detector array

resulting in several pixels being lost. To circumvent the possibility of this reoccurring, subsequent operations are restricted such that diurnally resolved observations are not collected for around 5 weeks either side of the equinoxes. As a result, the

65   production of unfilled GERB obs4MIPs monthly hourly fluxes was initially restricted to the months of November, December, January and May, June, July avoiding the months impacted by these operating restrictions. As will be demonstrated in Sect. 2.5, implementing a relatively simple data filling approach additionally allows the construction of February and August monthly hourly averages within tolerable uncertainties.

Figure 1 summarizes the steps used to produce an unfilled obs4MIPs product from the GERB HR 15-minute fluxes for both

70   the OLR and RSW. The initial step involves averaging of the GERB HR data to a 1°x1° hourly scale. To achieve this, an area weighted average of all the available points whose centres fall within each 1°x1° grid-box is performed across the region from 60° N-60° S and 60°E-60° W for points with a viewing zenith angle of less than 70°. This is followed by a straight average over all the available 15-minute products for each UTC hour, centred on the half hour. Both these steps proceed without prejudice if there is at least one GERB observation within the 1°x1° grid-box and at one time step in the hourly bin. For the

75   OLR this process is performed directly on the fluxes; for the RSW, the fluxes are converted to albedo before averaging and converted back to flux at the 1°x1° degree hourly scale, using the incoming solar flux representative of the centre of each 1°x1° grid-box and hourly bin (i.e. at 00:30, 01:30, UTC etc.). This treatment mitigates for any bias that might result from only some of the 15-minute observations within the hour being available and enables hourly fluxes to be derived in the presence of missing data, as long as at least one observation is available. We note that the GERB HR RSW products use a twilight model based on

80   the model derived from CERES observations (Kato and Loeb, 2003) for solar zenith angles between 85° to 100° and set RSW to zero for solar zenith angles greater than 100°. For consistency this treatment is also applied to the GERB obs4MIPs products at the daily hourly 1°x1° scale using the solar zenith angle of the centre of the grid-box and hourly bin. Hence, twilight and night-time RSW HR fluxes are not included in the averaging to the daily hourly scale if the central solar zenith angle is less than 85° but are used to replace grid-box values when the central solar zenith angle is equal to or exceeds 85°. For both OLR

85   and RSW, in the initial 'unfilled' product version, the resulting 1°x1° hourly fluxes are then averaged over all available days of the month to give the final 1°x1° unfilled monthly hourly products.

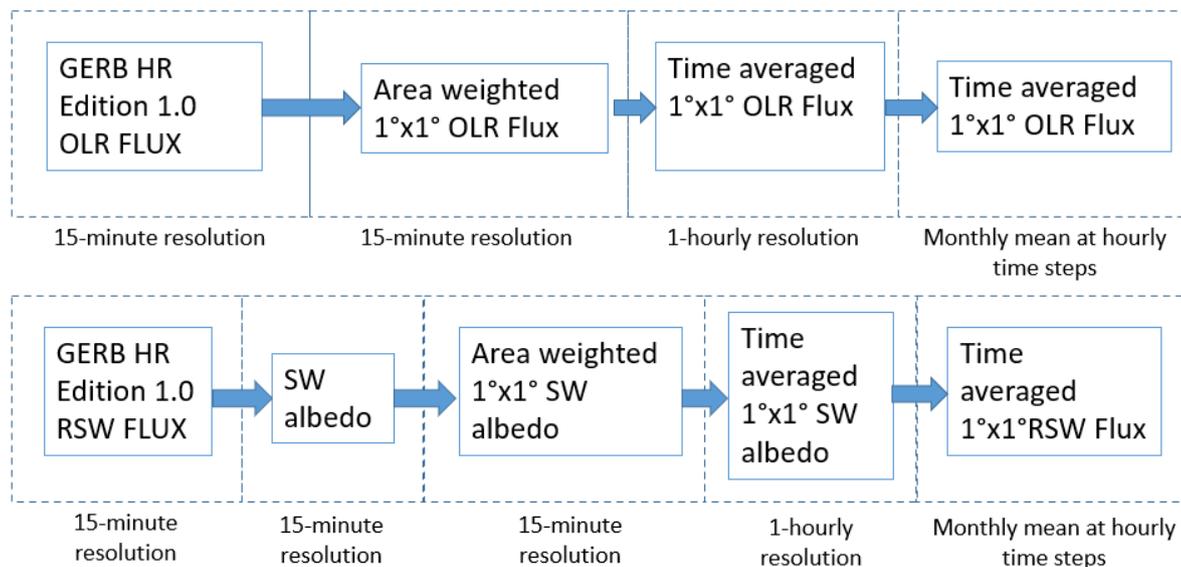



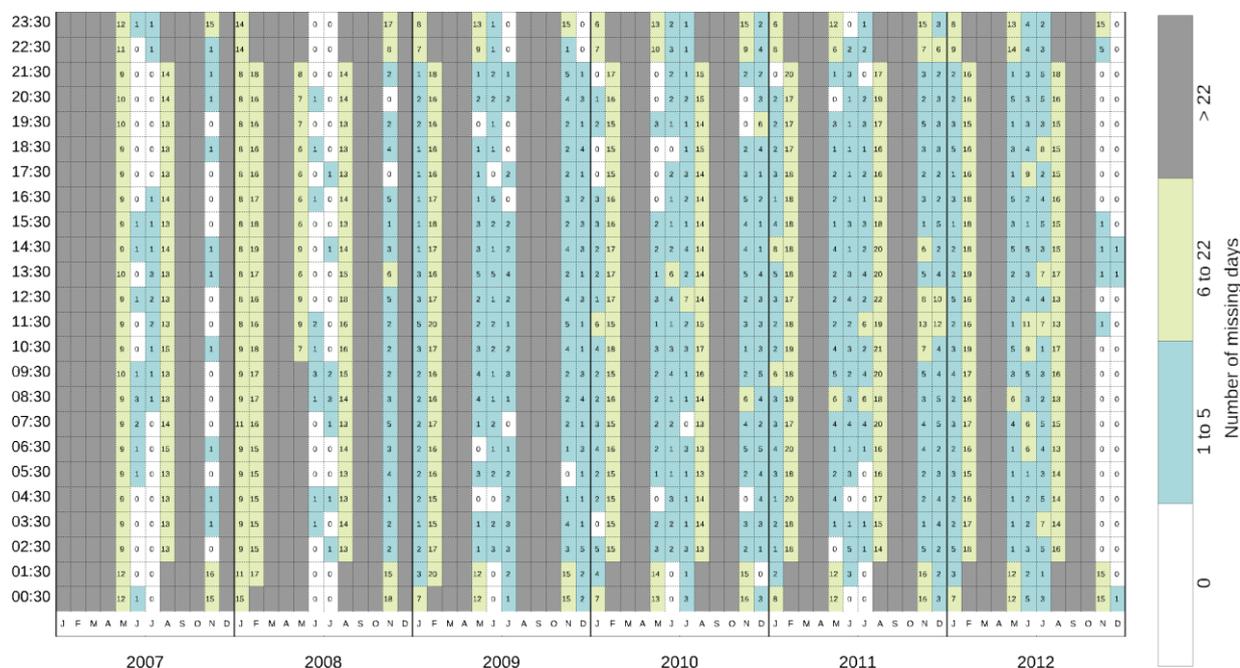
Figure 1: Schematic of the steps employed in the production of the OLR (top row) and RSW (bottom row) monthly hourly average Obs4MIPs products from the GERB HR Edition 1.0 fluxes without the implementation of data filling.

2.2 Diagnosing missing data

Whilst the instantaneous GERB data, including the HR products have been validated (Clerbaux et al. 2009; Parfit et al. 2016), the effect of missing data on the fidelity of the obs4MIPs averages needs additional consideration. The months for which the unfilled obs4MIPs averages are calculated are not affected by systematic outages but do still experience a significant quantity of missing observations. Calibration operations and other planned and unplanned operational issues result in observational gaps over the whole of the GERB region for one or more hours or, more occasionally, days at a time, and manifest as missing timeslots in the HR record. This leads to a significant number of cases where there are no observations available for a given hour on a particular day, which without further data processing, appear as gaps at the daily hourly scale that result in errors in the obs4MIPs monthly hourly averages.

A summary of the number of missing days of GERB data for the whole GERB-1 record is shown in Figure 2 as a function of hour and month. Hours with complete data are shown in white and those with more than 22 days missing are shaded grey. Hours where there are between 1 and 5 missing days in the month are shaded turquoise and cases with between 5 and 22 missing days are shaded pale green. It can be seen that there is an uneven distribution of missing data through the record, with a few months (e.g., December 2012) showing almost complete data coverage and others showing varying degrees of incomplete coverage at all hours. As previously discussed, operating restrictions in the months around the equinoxes are responsible for an almost complete absence of observations during March, April, September and October resulting in these months being greyed out. These restrictions are also responsible for the pattern of missing data in February and August, where the latter part of these months is always missing. Persistently higher amounts of missing data in the hours around midnight for

November and May is a result of data excluded due to stray light contamination at the start of each of these months. The other cases with more than 5 missing days across all hours (e.g., May and December 2007 and 2008) are at least in part associated

110 with extended instrument outages and in some cases satellite outages, leading to the loss of multiple days of data. Apart from these cases, missing data are generally randomly distributed through the month and the specific days that are missing generally change from hour to hour. Hence, the effect of missing data on the monthly hourly averages may also affect the fidelity of the diurnal cycle in unexpected ways.

115 **Figure 2: Number of missing days of data per month as a function of hour and year. Cells are coloured according to number of missing days for that hour and month, with turquoise indicating 5 or fewer missing days and pale green between 6 and 22 days missing and grey more than 22 missing days. Where there are 22 or fewer missing days the actual number of days missing is indicated in the box.**

2.3 Evaluation of the impact of missing data

120 The error in the monthly average due to missing data can be estimated by considering the effect of removing days from a month of data with complete, or nearly complete, coverage. Every UTC hour of the GERB-1 record with no more than one missing day during the month was used as a starting point for this analysis. This represents just over a third of the data for the months not affected by the systematic outages around the equinoxes. It also provides good coverage of the diurnal cycle for each of these months.

125 In this analysis, we consider each of the 'complete' or 'nearly complete' monthly hourly averages as the 'true' value. Differences between these true values and the averages calculated after the removal of a selected number of days provide an estimate of the error due to missing data. The effect of removing between 1 and 12 days randomly distributed through the month was calculated for eight different realisations of the days chosen. The effect of removing between 2 and 22 consecutive

days was also determined for three different patterns: all days missing at the start of the month, at the end of the month and
centred around the middle of the month.

Figure 3 displays example results for the removal of three randomly chosen days of data from the December 2012 11:30 UTC
monthly hourly average. Four different realisations of the missing days are shown. The variation in the spatial distribution of
the error (top 4 rows) highlights the effect of the altered sampling. The largest differences in averages are seen for the RSW in
the more strongly illuminated summer hemisphere and are for the most part associated with the averaging of synoptic
variability at higher latitudes. Notable errors are also present in other regions which exhibit significant day-to-day variability
in cloud coverage and/or properties, such as deep convective regimes over southern Africa. For both the OLR and RSW the
detail of the spatially resolved errors varies for each of the realisations depending on the meteorology on the individual days
removed. However, the overall distribution of errors shown in the bottom row for each case is relatively stable. For both the
OLR and RSW the distributions are relatively symmetrical about the mean, which is close to zero. As might be anticipated
from the spatial error patterns, the spread in the error is significantly larger for the RSW than the OLR with the associated
standard deviations between 3.5 to 4 times higher for the former. The mean and standard deviation of the error distribution are
considered sufficient summary statistics for interpreting the change in the errors as a function of number of days missing, time
of day and month.

Considering the results for all months and times of days used in this analysis we find that for the OLR, systematic variations
in the standard deviation and mean of the resulting error distribution, both seasonally and diurnally, are small and difficult to
distinguish from the variability resulting from the choice of days. Seasonal variation in the error distribution is also negligible
for the RSW, aside from a small reduction in variability in the standard deviation and a very slight reduction in its value for
July. This is associated with an increasingly dominant contribution from the Sahara which has low day to day variability.
However, there is a noticeable diurnal signal in the standard deviation of the RSW error distribution. Even when only calculated
over the locations which are not in twilight or night-time at any point in the month at that hour, the standard deviation, which
is relatively stable between 10:30 and 15:30 UTC when there is a high level of solar illumination, drops steadily for earlier
and later times of day, due to the overall reduction in the incoming solar flux. Results for hours earlier than 04:30 and later
than 19:30 UTC are more unpredictable and generally noisy as there are typically less than 20 % of the full number of points
represented in the statistics due to the limited portion of the disc illuminated at these times. Thus, for the RSW, combining
results for all months and for the hours 10:30 to 15:30 UTC, gives an indication of errors at the height of the disk illumination.
Errors at 04:30 and 19:30 UTC represent the error distribution for the low illumination case, when there are still a sufficient
number of points illuminated to obtain reasonable statistics.

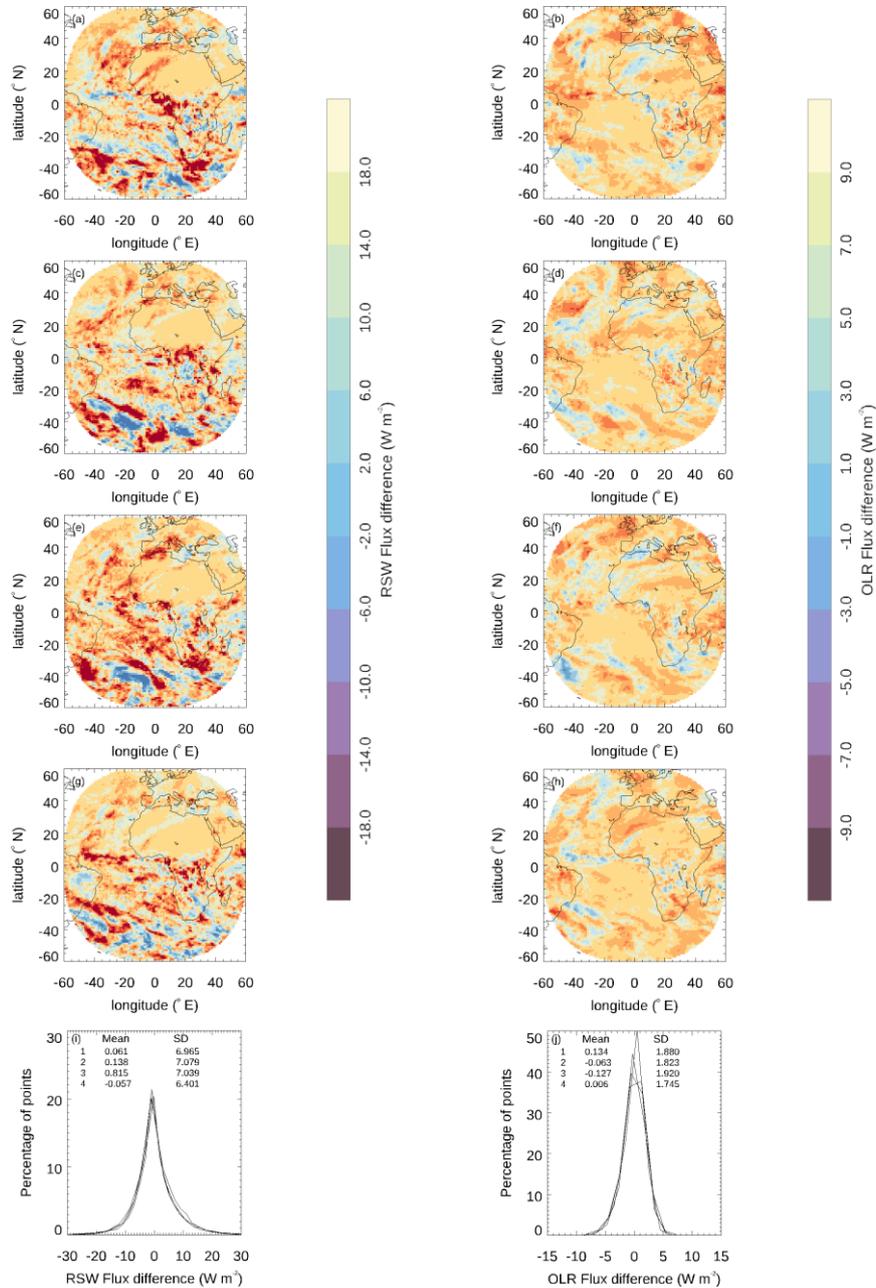



Figure 3: Impact on the 1°x1° monthly hourly mean fluxes of removing 3 randomly chosen days of data from December 2012 11:30 UTC. The results for four different random realisations of the days removed are shown for RSW in the left-hand panels (a, c, e, g, i) and OLR in the right-hand panels (b, d, f, h, j). The top four rows (panels a to h) show the spatially resolved difference compared to the full dataset and the bottom panels (i and j), show the distribution of these 1°x1° differences for the four realisations shown. The mean and standard deviation in each of the four cases is also shown.

Figure 4 summarizes the expected monthly hourly mean error due to missing data at the one degree scale, in terms of the

standard deviation and mean of the error distribution for both randomly and systematically removed data. The results show

that on average the mean and standard deviation increase roughly linearly as the number of missing days increases. The variability in the standard deviation and mean also increases as the number of missing days increases but in a less regular manner. For the 10:30 to 15:30 UTC time range the standard deviation of the RSW error distribution increases rapidly as the number of missing days increases, exceeding 10 W m⁻² for some cases with four or more consecutive missing days or five or

170 more missing days randomly distributed through the month. The corresponding standard deviation which is exceeded for the OLR in these cases is 3 W m⁻². For the mean of the error distribution, which is the overall image bias due to the missing data, individual realisations can see increasingly large biases as the number of missing days increases. When consecutive days are removed the bias may exceed 2 Wm⁻² for the RSW and 1 W m⁻² for the OLR for as few as 3 or 4 missing days for some of the cases.

175

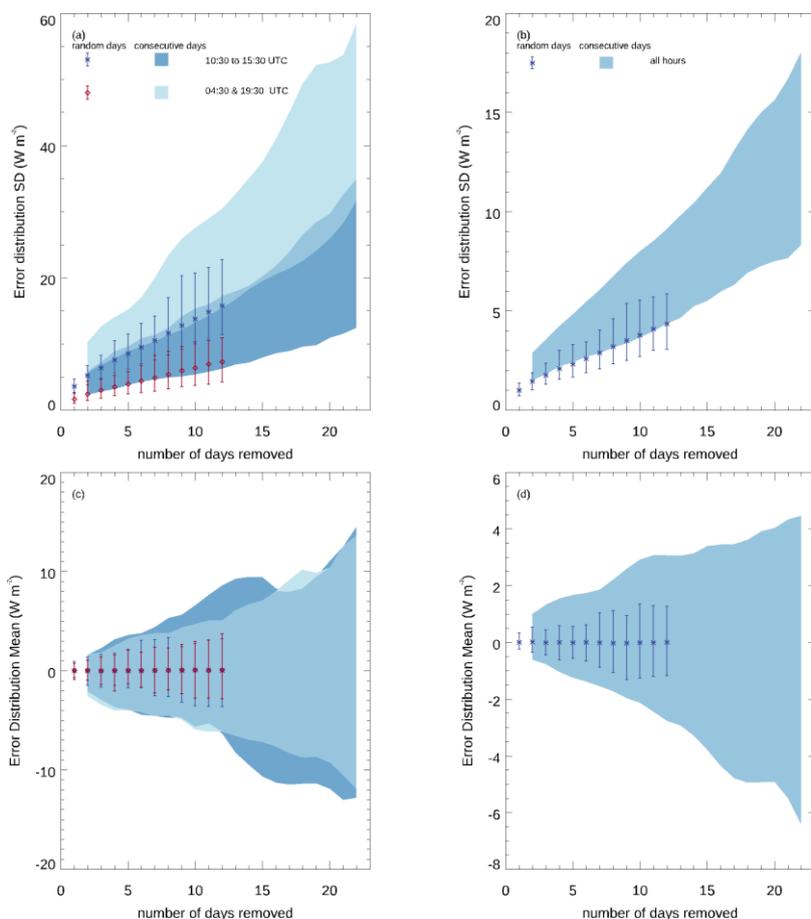



Figure 4: Summary statistics for the error distribution in the monthly hourly mean 1°x1° fluxes due to missing days. Standard deviation (panels a and b) and mean (panels c and d) of the error distributions are shown as a function of number of days removed for the RSW (a and c) and OLR (b and d) fluxes. The average and range over the realisations and months are shown for days removed chosen at random as points with bars. The corresponding range for points systematically removed from various points in

180 **the month is shown as shaded regions. For the RSW results shown separately for the UTC hours 10:30-15:30, representing high solar illumination of the GERB region and for 04:30 and 19:30 combined, representing low solar illumination. The OLR results are shown for all times together.**

2.4 Strategy for filling missing GERB data

Considering the amount of missing data in the GERB data set and the associated errors due to missing data summarised in
185 Figure 4, it is clearly desirable to investigate methods to fill some of the missing information, with the goal to reduce bias and spread in the resulting error distribution. Given the pattern of missing data, with multiple occurrences of several hours and indeed several days missing in some cases, filling the gaps by interpolating the exiting GERB observations is not viable. Ideally an alternative source of information responsive to the meteorology present during the periods of missing data that can be used to fill in the gaps in the record is required.

190 The prime instrument on the Metosat second generation satellites is the SEVIRI imager (Schmetz et al 2002). This instrument provides radiances in 11 narrow band channels from 0.635 to 13.4 μm every 15 minutes with a resolution of 3 km at the sub-satellite point. The GERB HR products, on which the obs4MIPs dataset is based, are provided as a snapshot at the time of the corresponding SEVIRI observation, at a resolution of 3x3 SEVIRI pixels, on a grid aligned with the SEVIRI grid. As part of the GERB processing an empirical narrowband to broadband conversion is applied to the SEVIRI radiances to derive estimates
195 of the broadband radiances. These so called 'GERB-like' radiances are converted to flux with the same conversion factor used to determine the GERB fluxes from the GERB radiances (Dewitte et al., 2008).

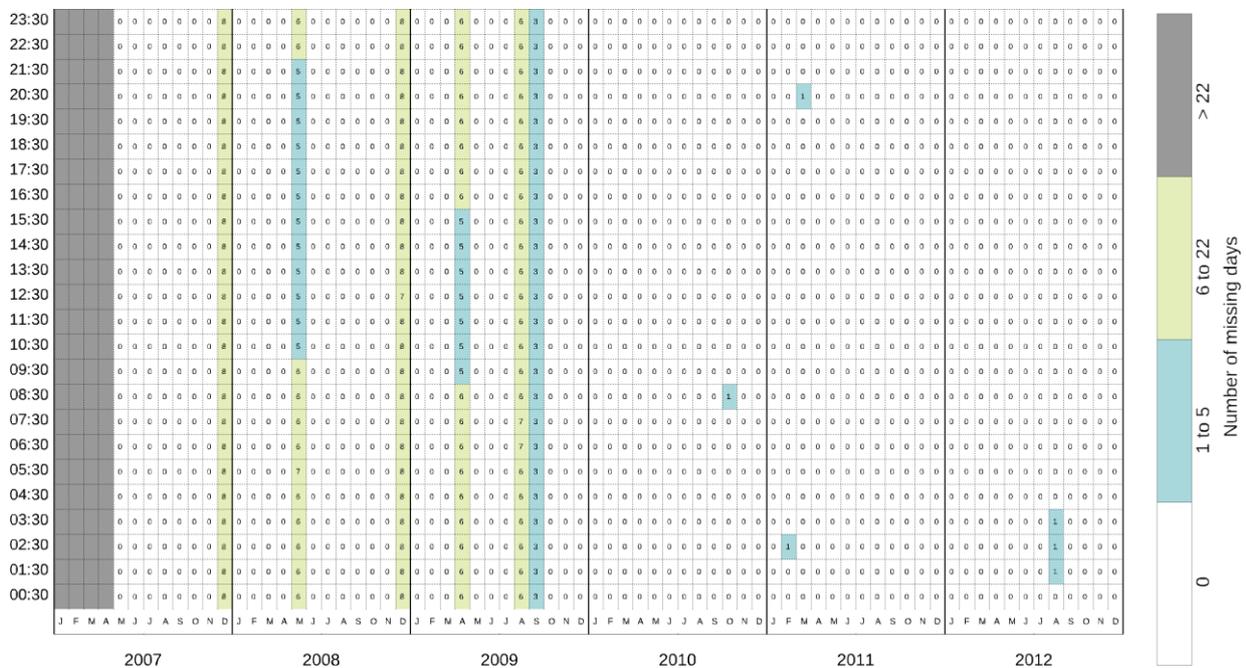



Figure 5: As **Figure 2** for GERB-like observations.

The SEVIRI based GERB-like fluxes suffer from significantly less missing data than the original GERB record. This is clearly
200 illustrated by Figure 5, which indicates the number of days of missing GERB-like data for each monthly hourly average for the whole GERB-1 period. As can be seen by comparing this to Figure 2, except for few extended outages in the first few years

which are a result of satellite level anomalies, nearly all the data missing in the GERB record are present in the GERB-like. Thus, the latter record may be useful for filling much of the missing GERB data.

The primary use of the GERB-like fluxes in the GERB processing is to provide spatial detail within the GERB footprint to
205 enable non-uniformities in the GERB pixel point spread function to be corrected. This use places no absolute accuracy requirements on GERB-like data, and their relative accuracy only needs to be maintained on a temporal and spatial scale of a little more than a single GERB pixel, as in the processing the GERB-like data are always used as a supplement to the GERB pixel level radiance. Therefore, the original regressions on which the narrowband to broadband conversion is based are not optimized to avoid larger scale angular biases that are negligible at the GERB pixel scale and neither the GERB-like data nor
210 the imager radiances used to produce them are calibrated against the GERB observations. Therefore, we expect a number of deficiencies with the GERB-like data will need to be addressed before they are suitable to be used to directly fill missing GERB data.

Our expectation is that the GERB and GERB-like fluxes will differ due to deficiencies in the narrowband to broadband conversion and due to the calibration of the original narrowband observations. Narrowband to broadband conversion errors
215 will likely have scene and angular dependencies that are not expected to vary a great deal over time, except in relation to these variables. Conversely, calibration related errors would be expected, at first order, to manifest across different scenes in a similar reproduceable way but may vary in time. There may also be cross terms, where calibration changes manifest across the scenes differently due to variation in the weighting of the channels between scenes. These differences need to be characterized and reduced as much as possible to optimize the usefulness of the GERB-like fluxes as a proxy for the missing GERB data.

220 Figure 6 and 7 show the spatially resolved monthly hourly mean GERB to GERB-like ratio, for the RSW and OLR, for a selection of different UTC hours and months. The ratios shown in these figures are determined from the 1x1° monthly hourly averages constructed from the GERB and GERB-like fluxes, where the available data used to construct these averages has been matched in both data sets. GERB-like data are always present when the GERB fluxes are available as they are a required part of the GERB processing, so matching the data availability simply involves removing GERB-like observations from the
225 average where the corresponding GERB data are missing.

The ratios shown in Figure 6 and Fig. 7 illustrate both a global bias between the two sets of fluxes and angularly dependent effects that manifest differently according to scene type. For the RSW, the ratio between the GERB and GERB-like fluxes generally varies between 0.95 and 1.2. Variations occur with viewing and solar angle and thus with both location and time of day and, more subtly, time of year associated with the variation of the solar zenith angle. The lowest RSW ratios tend to occur
230 at larger solar zenith angles over land. The highest RSW ratios occur over ocean and are mostly at larger solar zenith angles, especially when combined with large viewing zenith angles. For the OLR the ratios are generally less extreme than the RSW, with the lowest values of around 0.97 observed towards the edge of the GERB region at the largest viewing zenith angles for the coldest scenes. The fixed viewing geometry of the geostationary platform means that viewing zenith angle effects correspond to fixed locations. The diurnal variation in the GERB to GERB-like OLR ratio is small and is associated with
235 marked changes in scene, for example the daily heating of the land, seen most significantly over desert regions such as the

Sahara. Similarly, seasonal variations in the OLR ratio are associated with scene variations such as the seasonal variation in the positioning of the ITCZ and changes to solar induced land heating.

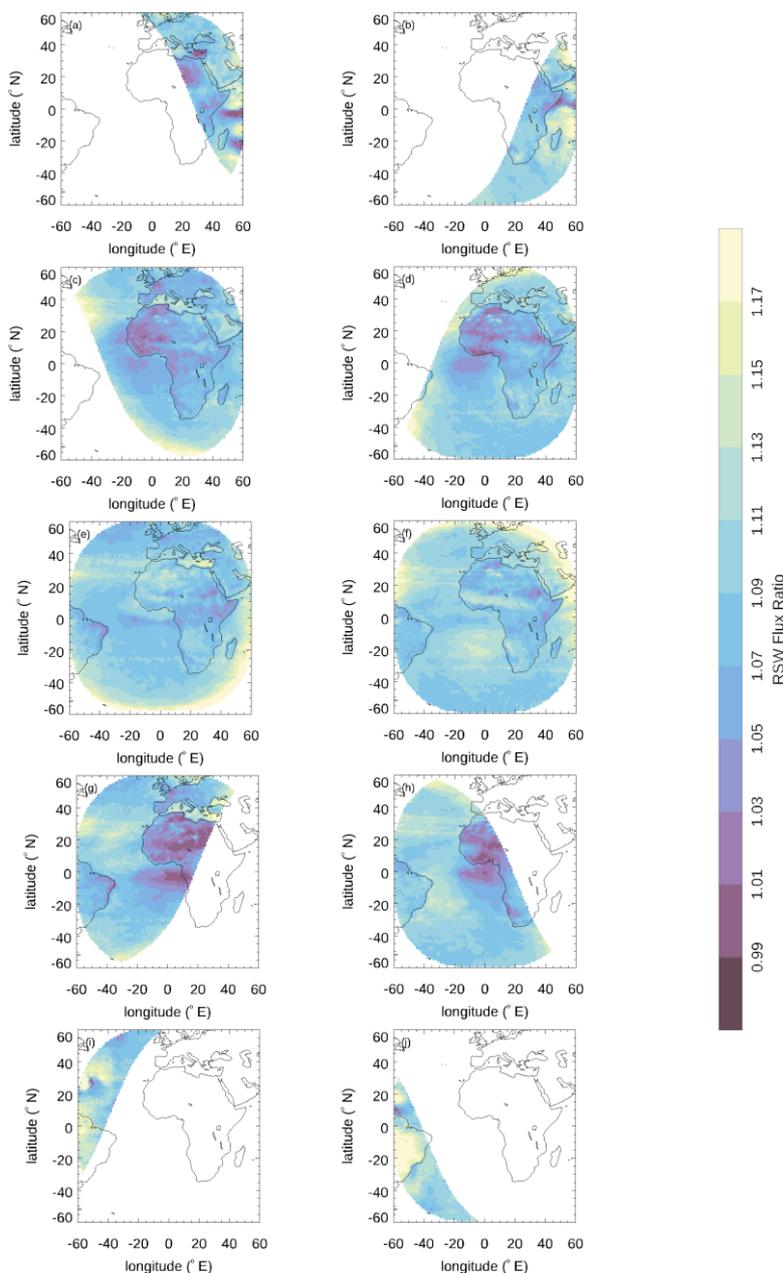



Figure 6: GERB/GERB-like RSW ratio of the monthly hourly mean at 1 x 1 degrees, for June 2009 in the left-hand panels (a, c, e, g, i) and December 2009 in the right-hand panels (b, d, f, h, j) for, from top to bottom, 04:30 (a and b), 08:30 (c and d), 12:30 (e and f), 16:30 (g and h) and 20:30 (i and j) UTC.

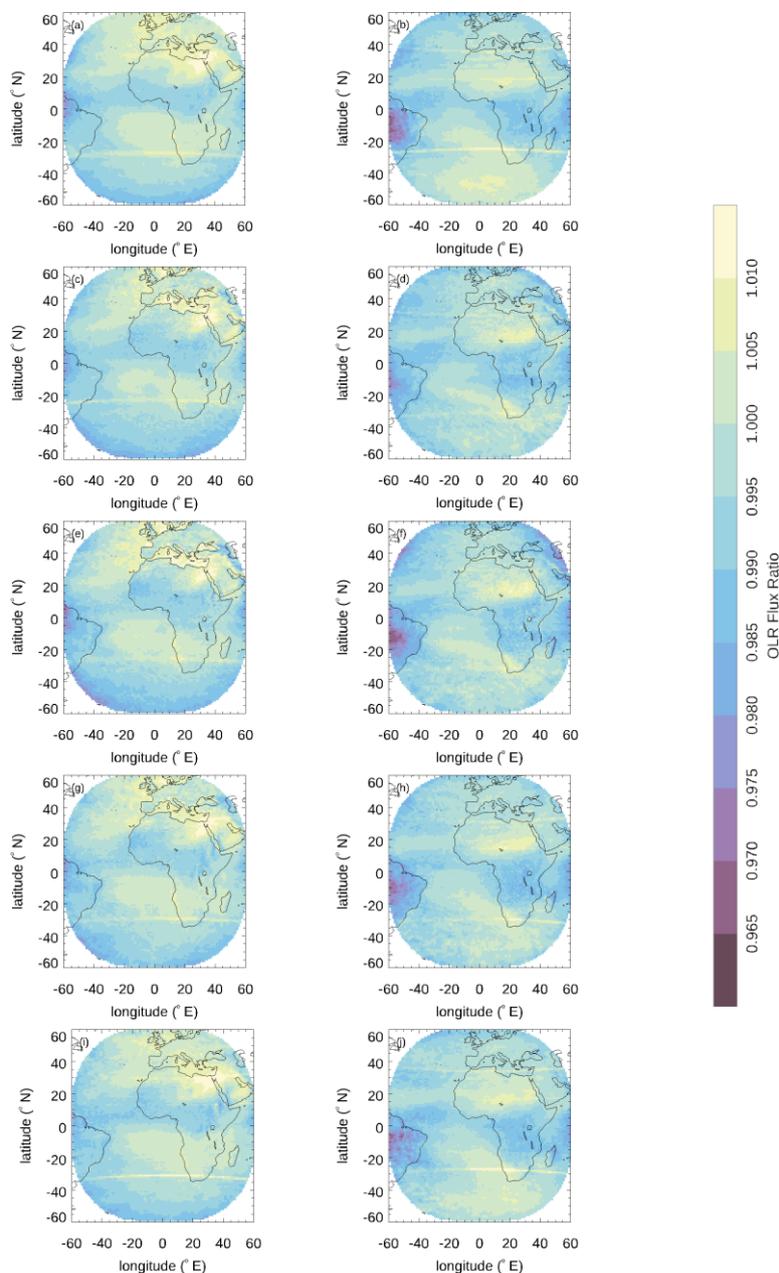



Figure 7: As Figure 6 but for the GERB/GERB-like OLR ratio for June 2009 in the left-hand panels (a, c, e, g, i) and December 2009 in the right-hand panels (b, d, f, h, j) for, from top to bottom, 04:30 (a and b), 08:30 (c and d), 12:30 (e and f), 16:30 (g and h) and 20:30 (i and j) UTC.

Figure 6 and 7 show that the ratio between the GERB and GERB-like fluxes does indeed show a variety of the expected variations between the two datasets, with strong angular and scene dependent patterns in the ratio of the fluxes dominating. However, we find the day-to-day variation in the overall bias between the two datasets (not shown) manifests at a much lower level in both the OLR and RSW and is difficult to distinguish from the combined effect of scene dependent bias and day-to-

250   day variation in scene make up. Hence, we hypothesize that it may be feasible to fill missing GERB fluxes with the
      corresponding GERB-like fluxes, adjusted by the GERB/GERB-like ratio calculated at the monthly hourly mean temporal and
      1º spatial scale. As a first test of this hypothesis, Figure 8 summarizes the difference between the GERB-like and GERB fluxes
      at the daily hourly 1º scale before and after adjustment of the GERB-like values by these ratios calculated from the monthly
      hourly means. Distributions of the 1º differences for each hour of every day where GERB and GERB-like data are available,

255   as long as there are no more than 22 missing days in the month, are determined. Figure 8 displays the average and range of the
      mean and standard deviation of these difference distributions as a function of UTC hour before and after adjustment of the
      GERB-like for both the OLR (left hand panels) and the RSW (right hand panels). The results show that both bias and spread
      of the error at the daily hourly 1º scale is noticeably reduced by correcting the daily data by the monthly hourly ratios. The
      daily hourly 1º biases are reduced to within a few W m^{-2} of zero and the standard deviations from an average of 10 to 4.6 W m^{-2}

260   ^{2} in the RSW and from 2.2 to 1.7 W m^{-2} in the OLR. The results imply that a single monthly hourly correction applied at the
      1º scale improves the fidelity between the GERB-like and GERB fluxes at the daily scale.

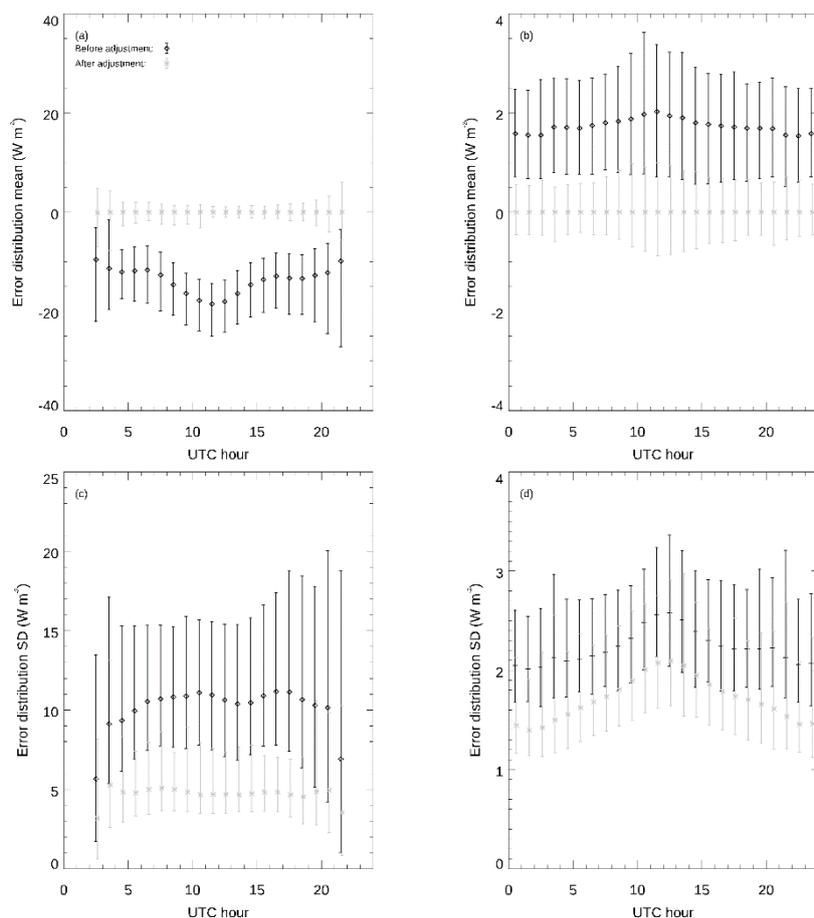



**Figure 8: Summary statistics for the GERB-like – GERB difference before (black) and after (grey) adjustment of the GERB-like
by the monthly hourly ratio. Points indicate the average and the bars the range of these statistics over all the days at each hour.**

265   **Results are shown as a function of UTC hour for the RSW (panels a and c) and the OLR (panels b and d) for the mean of the**

distribution (a and b) and the standard deviation (c and d). Times are on the half hour in all cases, but the plotting for the adjusted case is slightly offset on the x-axis for clarity.

2.5 Evaluation of filled data record

Whilst the improvement in correspondence between the GERB and GERB-like daily hourly fluxes after adjustment with the
270 monthly hourly ratio is encouraging, the results discussed thus far are not quite representative of the situation in the case of missing GERB data. In this case the monthly hourly ratio derived from incomplete GERB and corresponding GERB-like fluxes will need to be used to correct GERB-like fluxes that are not included in that average. Thus, it needs to be shown that rescaling by a monthly hourly average ratio derived from incomplete data can sufficiently improve the GERB-like fluxes at the daily-hourly 1º scale for the missing periods for the adjusted fluxes to be used to fill the missing GERB fluxes.

275 Analogous to the approach used in Sect. 2.3, starting with all the hours of the record with no more than one missing day of GERB data in the month, we determine the effect of removing increasing amounts of GERB data and replacing it with GERB-like data scaled by the monthly hourly ratio. In each case we match the data-coverage for both GERB and GERB-like: i.e. corresponding points are removed from both data records before calculating the monthly hourly means and the associated ratio. As was done for the unfilled average comparison described in Sect. 2.3, the resulting filled average can be compared to the
280 average calculated from the GERB data alone before any data were removed to assess error.

Figure 9 summarises statistics of the residual error at the monthly hourly average 1°x1° scale for the filled data. It can be directly compared to Figure 4 which shows the equivalent results for the unfilled averages. Comparing the two figures shows that filling the missing GERB fluxes with their scaled GERB-like equivalents reduces both the mean error and the spread in the error by more than a factor of 10 in all cases. Given these improved statistics we implement this filling approach to produce
285 our 'filled' GERB obs4MIPs product and use it in the next section to perform an initial evaluation of climate model performance. We note that the level of error reduction is retained for systematic data removal of up to 22 days such that we are also able to reinstate the months of February and August in the filled record.

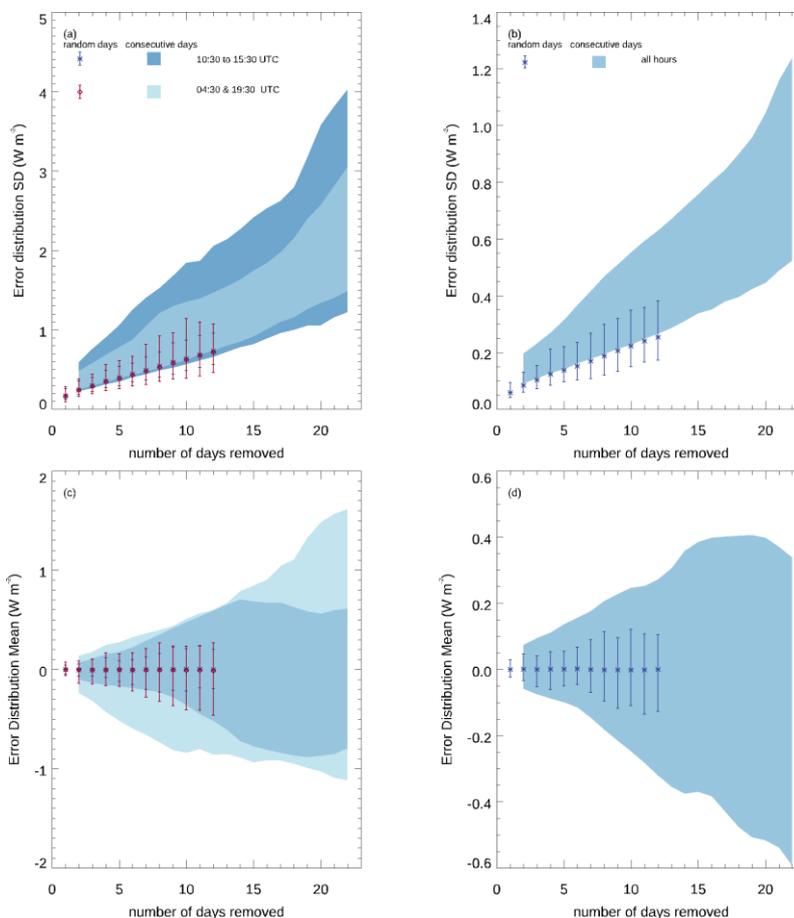



Figure 9: As Figure 4 but for GERB data that has been filled using scaled GERB-like data as described in the main text. Summary
290 **statistics for the error distribution in the monthly hourly mean 1°x1° fluxes due to missing days. Standard deviation (panels a and b) and mean (panels c and d) of the error distributions are shown as a function of number of days removed for the RSW (a and c) and OLR (b and d) fluxes. The average and range over the realisations and months are shown for days removed chosen at random as points with bars. The corresponding range for points systematically removed from various points in the month is shown as shaded regions. For the RSW results are shown separately for the UTC hours 10:30-15:30, representing high solar illumination of the GERB**
295 **region and for 04:30 and 19:30 combined, representing low solar illumination. For the OLR results are shown for all times together.**

Therefore, with the exceptions of December 2007, May and December 2008 and August 2009, where satellite outages result in a loss of both GERB and GERB-like data meaning missing GERB observations cannot be filled, a filled GERB monthly hourly average is produced for all the non-greyed out months shown in Figure 2 with the error associated with filling bounded by the values shown in Fig. 9.

300 **3. Application of the GERB Obs4MIPs filled product to climate model evaluation**

Top of the atmosphere radiative fluxes are routinely used as an evaluation metric for climate model performance, with model parameters often tuned to produce a realistic radiation budget. This is typically performed at a relatively coarse temporal and

spatial scale (monthly or global annual means) which has the potential to mask compensating errors. A more stringent test, at least at the process level, is to compare temporally resolved fluxes. This type of comparison has also been recognised as

305 potentially insightful for assessing cloud feedback (Webb et al., 2015) and has led to a limited number of modelling centres starting to produce and archive monthly hourly mean top of atmosphere radiative fluxes from AMIP type runs. Here we compare such fluxes, as simulated by two versions of the climate configuration of the Hadley Centre Global Environmental model HadGEM3, with the GERB obs4MIPs product.

3.1 HadGEM3 configurations and simulation description

310 The Global Coupled (GC) configurations of HadGEM3 contain the following subcomponents: Global Atmosphere (GA), Global Land (GL), Global Ocean (GO), and Global Sea Ice. We analyse historical Atmospheric Model Intercomparison Project (*amip*) simulations of two different GC configurations, 3.1 and 5.0. The *amip* simulations are forced with observations of sea-surface temperatures (SSTs), sea-ice cover and historical forcings (Eyring et al., 2016). Both configurations have 85 vertical layers and are run at N96 (1.875° longitude by 1.25° latitude) horizontal resolution.

315 GC3.1 is the configuration that underpinned the United Kingdom's contribution to CMIP6 (Williams et al., 2018; Mulcahy et al., 2018; Walters et al., 2019). The most recent configuration (GC5.0) has not been documented yet, but we briefly describe below three changes that are most relevant to the results presented here: prognostic-based convective entrainment, a new bimodal cloud scheme, and a reformulation of the 'cloud erosion' term.

The prognostic-based convective entrainment relates the entrainment rate to a 3-dimensional advected prognostic variable

320 based on surface precipitation. This links the prognostic rate to the amount of recent convective activity, introducing memory in the convection scheme. This change allows convection to behave in a more realistic manner, improving the diurnal cycle of convection over land. The new bimodal cloud scheme identifies entrainment zones associated with strong temperature inversions (Van Weverberg et al., 2021a & 2021b). Within these entrainment zones, two modes of variability are used to calculate the cloud liquid water content and the cloud fraction, conserving the grid-box mean value of saturation departure.

325 The large-scale cloud scheme (Wilson et al., 2008a & 2008b) contains a 'cloud erosion' term that represents the effect of entrainment of dry air into the cloudy part of the grid-box (Morcrette, 2012). GC5 uses a new implicit numerical method to calculate this term that avoids numerical overshooting and spurious removal of all the liquid cloud within the timestep. Both changes increase the amount and optical thickness of low-level cloud, particularly in the subtropics and lower midlatitudes.

Monthly mean diurnal cycles of TOA radiative fluxes (all-sky and clear-sky) are produced for the entire length of the amip

330 experiment. The radiative fluxes are hourly means, centred, as in the observations, on the half-hour, and the monthly mean diurnal cycle is constructed by averaging each UTC hourly mean over the entire month. These diagnostics were requested for the amip experiment of phase 3 of the Cloud Feedback Intercomparison Project (Webb et al., 2017). The HadGEM3 OLR diagnostics used in this study differ from those submitted to CFMIP3. The OLR diagnostics submitted to CFMIP3 contain a correction that accounts for the surface temperature adjustment by the boundary layer scheme in model time steps between

335 radiation time steps. This OLR diagnostic adjustment is introduced to conserve energy, but it significantly distorts the diurnal

cycle of OLR (its impact on daily and longer time averages is very small). Given that this OLR correction is purely diagnostic (i.e. it doesn't affect the model evolution) and it was not designed to work on sub-daily timescales, here we have used the OLR without this correction.

3.2 Model evaluation

340 For the purposes of highlighting the utility of the GERB obs4MIPs product we focus on two cloud regimes, marine stratocumulus and deep convection. Improving the representation of sub-tropical stratocumulus has been a focus for climate modellers for some time due to its importance in determining global cloud feedback (e.g. Bony and Dufresne, 2005). In general, models have tended to simulate too little marine stratocumulus, with what is present being too bright (e.g. Nam et al., 2012). In the multi-annual mean, Williams and Bodas-Salcedo (2017) report good agreement between GC3.1 and CALIPSO height-

345 frequency statistics over stratocumulus, but with a distribution that shows too few moderately optically thick clouds, compensated by too many optically thick clouds. Comparisons with CERES-EBAF top-of-atmosphere RSW fluxes imply that this translates into stratocumulus decks that are too reflective.

Deep convective regions continue to present a challenge because of the scale at which convection is typically parameterised in global climate models (e.g. Guichard et al., 2004, Hohenegger and Stevens, 2013, Christopoulos and Schneider, 2021).

350 Although improvements have been made (e.g. Stratton and Stirling, 2012), a persistent issue over land is that convective clouds tend to rain out too early, leading to too little cloud in the late afternoon to evening, when deep convection (and precipitation) typically peaks in observations (e.g. Yang and Slingo, 2001, Tan et al., 2019). Given the temporal resolution of the GERB obs4MIPs product it is ideally suited to investigate whether adjustments to the parameterisations that affect convective invigoration and lifecycle in GC5.0 are having a beneficial impact in terms of the top-of-atmosphere energy budget.

355 Figure 10 shows decadal average monthly mean January RSW fluxes as simulated by GC3.1 (a) and GC5.0 (b) over the region 60°S – 60°N and 60°E – 60°W. GERB obs4MIPs RSW fluxes are shown in panel (c), in this case averaged over the five years of GERB-1 January observations. The corresponding information for June is shown in panels (d)-(f), with, in this case six years of observations available for averaging. Broadly speaking the simulations capture the patterns seen in the observations, including the seasonal shift in the positioning and strength of features such as the ITCZ and stratocumulus decks off Angola and Namibia. There are differences: during the summer hemisphere GERB shows significantly higher RSW fluxes over the highest latitudes. It is noticeable that GC5.0 also tends to be brighter than GC3.1 in those regions. GC5.0 also appears to show more extensive, brighter marine stratocumulus off the west African coast in both seasons compared to GC3.1.

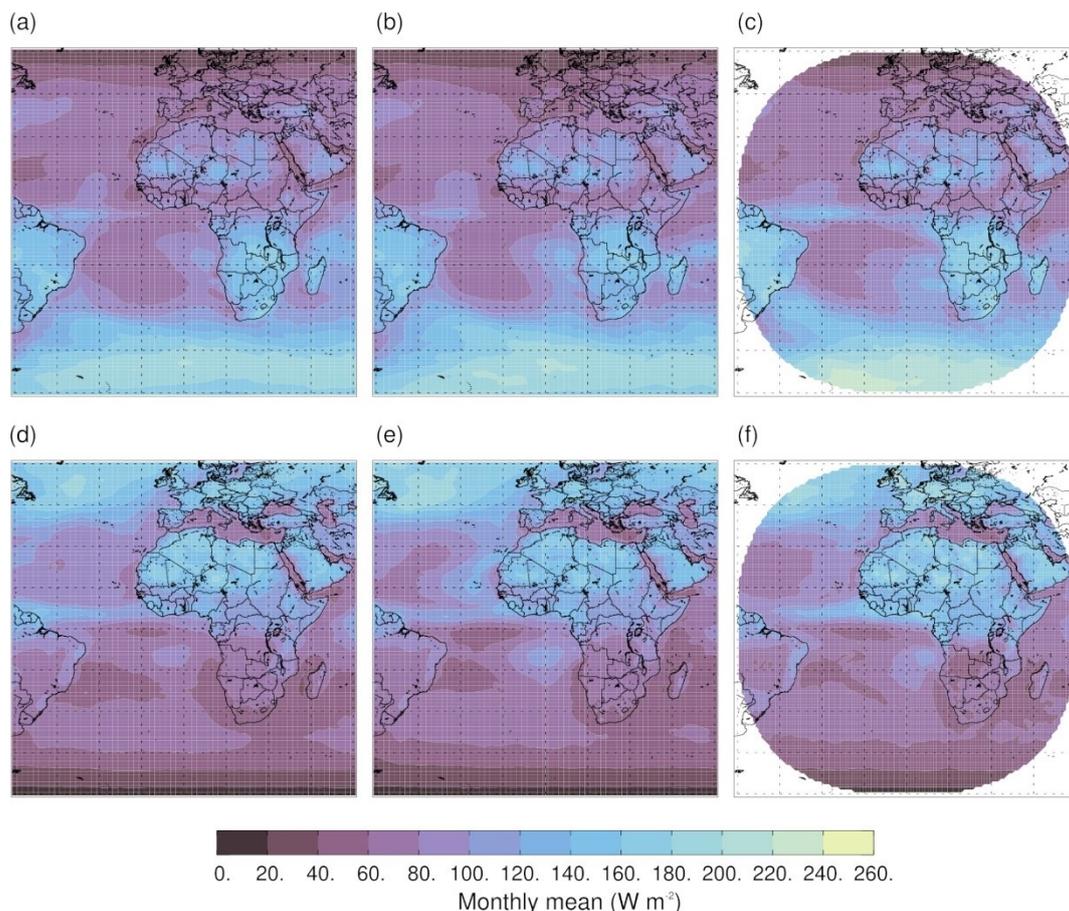



Figure 10: Monthly average top of atmosphere RSW fluxes for January in the top row (a, b and c) and June in the bottom row (d, e, and f) from GC3.1 (left panels a and d), GC5 (middle panels b and e) and GERB (right panels c and f). Simulated fluxes are a decadal mean (2000-2009). GERB fluxes are averaged over the duration of the GERB-1 observations.

Equivalent information to Figure 10 is shown in Figure 11 for OLR fluxes. In this case the most obvious differences between the two HadGEM3 simulations are located in regions of tropical deep convection. In the northern summer GC5.0 appears to shift the peak of convection within the ITCZ further east. In January the centres of deep convection over Brazil and central southern Africa are both strengthened in GC5.0 relative to GC3.1. Visually, both changes appear more in line with the GERB observations, although the intensity of land convection still appears greater in the observations.

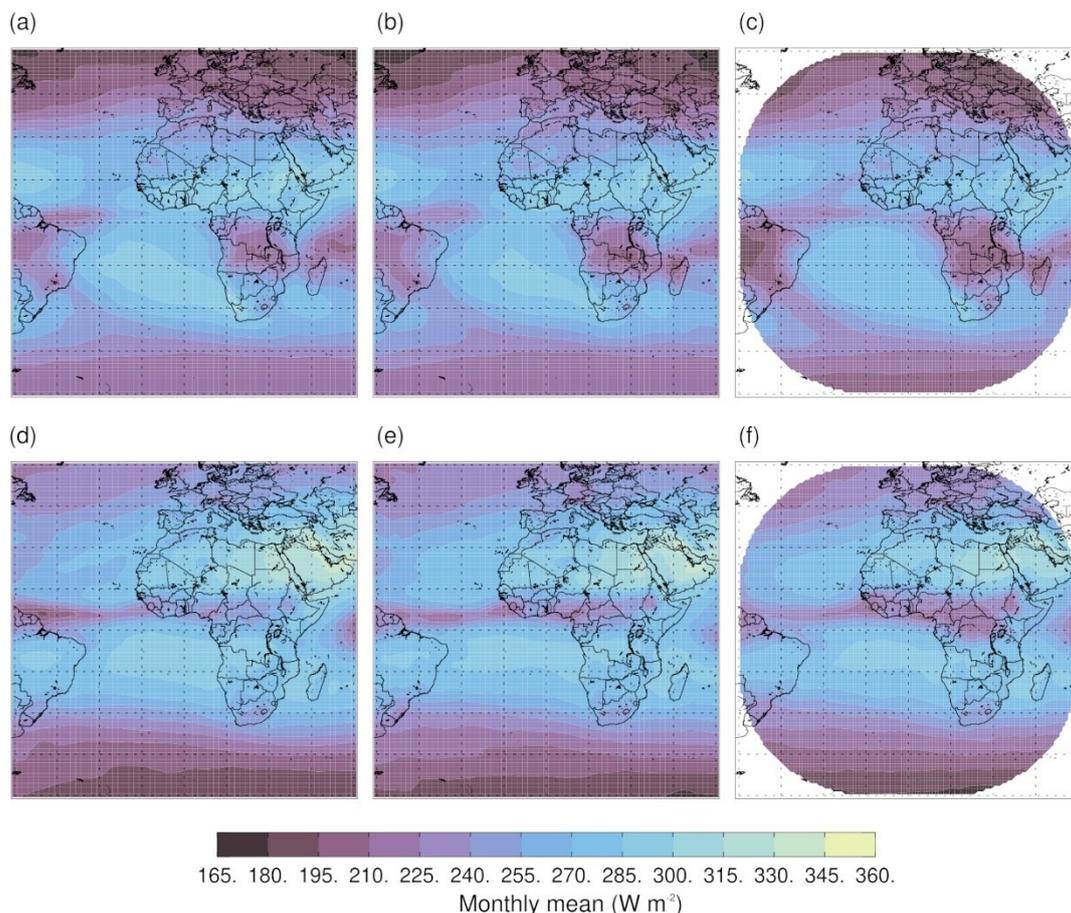



Figure 11: As Figure 10 for OLR fluxes.

To provide a more quantitative analysis we define two seasonally dependent latitude-longitude boxes encompassing the south-
east Atlantic stratocumulus deck and African deep convection. Table 1 shows the multi-year June and January monthly mean
fluxes obtained from both sets of simulations and from GERB in these regions. We note that shortening the period of averaging
in the simulated datasets to be commensurate with the length of the GERB record makes a difference of, at most, 3 W m⁻² in
the mean fluxes.

	South Atlantic Marine Stratocumulus				African Deep Convection			
	June		January		June		January	
	(-16-10° E, 3-22°S)		(-16-10° E, 3-28° S)		(14-37° E, -2-12° N)		(15-31° E,0-17° S)	
	RSW	OLR	RSW	OLR	RSW	OLR	RSW	OLR
GERB	76.8	283.9	94.4	275.1	129.5	228.6	161.6	208.3
GC3.1	67.6	287.4	82.1	284.2	105.3	260.2	139.7	227.3
GC5.0	82.1	284.8	92.5	281.2	106.5	253.4	141.6	221.6

Table 1: Multi-year June and January monthly mean RSW and OLR fluxes over regions characterised by marine stratocumulus
380   **and deep convective cloud as observed by GERB and simulated by the two configurations of HadGEM3 outlined in the main text.**

Over the stratocumulus region Table 1 reinforces the qualitative impression from Figure 10 and Figure 11, with the change in HadGEM3 configuration resulting in a distinct brightening in both June and January. In June, the degree of brightening means that the mean RSW flux exceeds that measured by GERB, whereas in January, the increment is still insufficient to reach the level of the observed fluxes. As might be anticipated given typical stratocumulus altitudes, the impact on the OLR fluxes is

385   less marked but consistent between the months, reducing by of the order 3 W m^{-2}. In concert, these two results imply an enhanced cloud fraction, optical depth or both in the GC5.0 configuration.

The largest differences between the two sets of simulated fluxes over deep convection are realised in the OLR. Moving from GC3.1 to GC5.0 results in a reduction in OLR of order 7 W m^{-2} in both months, while a small increase of less than 2 W m^{-2} is seen in the corresponding RSW fluxes (Table 1). These changes move the GC5.0 fluxes towards the observations but there is

390   still a notable overestimate in OLR and corresponding underestimate in RSW flux, particularly in June, consistent with the visual impression of 'missing' land convection in the simulations during this month (Figure 11).

To understand the reasons behind the changes in the model fluxes in both regions we use diagnostics produced by version 1.4 of the CFMIP (Cloud Feedback Model Intercomparison Project) Observational Simulator Package (COSP; Bodas-Salcedo et al., 2011). In particular, we use vertical profiles of cloud fraction of the Cloud–Aerosol Lidar and Infrared Pathfinder Satellite

395   Observation simulator (CALIPSO), and International Satellite Cloud Climatology Project (ISCCP) histograms of cloud fraction in intervals of cloud-top pressure (CTP) and cloud optical thickness (τ). The CALIPSO and ISCCP simulators are documented in Chepfer et al. (2008) and Klein and Jakob (1999), respectively.

Figure 12 illustrates these diagnostics for January. Results for June are qualitatively similar. GC5.0 shows a significant increase in cloud fraction in the stratocumulus region (Fig. 12a), with clouds also being optically thicker (Figs. 12 c and e). These two

400   changes contribute to the increase in RSW described above. In the deep convection region, GC5.0 shows an enhanced cloud fraction at high altitudes, coupled with lower cloud top height (Fig. 12b). The impact of these two changes on the OLR will partially cancel out. However, GC5.0 also shows optically thicker clouds (Figs. 12d and f). The combined increase in cloud fraction and optical thickness leads to a reduction in OLR in GC5.0 compared to GC3.1 (Table 1), despite the reduction in cloud top height.

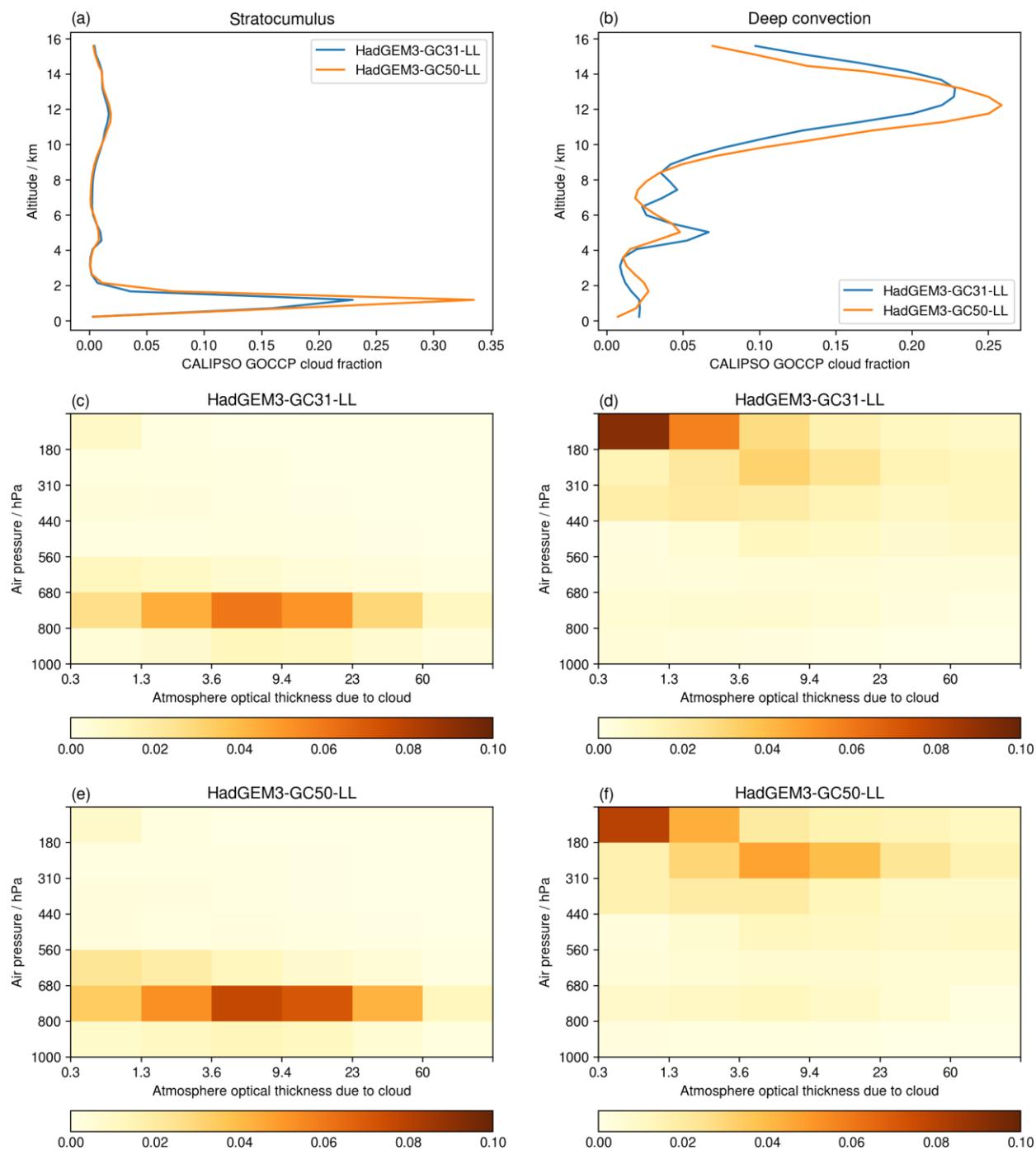



Figure 12. Multi-annual monthly average cloud fraction for January. Vertical profiles of COSP/CALIPSO cloud fraction (a and b), and COSP/ISCCP CTP-τ histograms of cloud fraction (c to f). The left columns show plots for the stratocumulus region and the right column for the deep convection region.

Utilizing the diurnally resolved GERB obs4MIPs fluxes we analyse these results further by decomposing them as a function

410   of time of day. Figure 13 shows the regional hourly monthly mean RSW fluxes from each HadGEM3 configuration for each

individual year of the simulation over the stratocumulus regions. Superposed in colour are the GERB obs4MIPs fluxes for

2007-2012. Figure 14 shows equivalent information for OLR fluxes over the regions of deep convection.

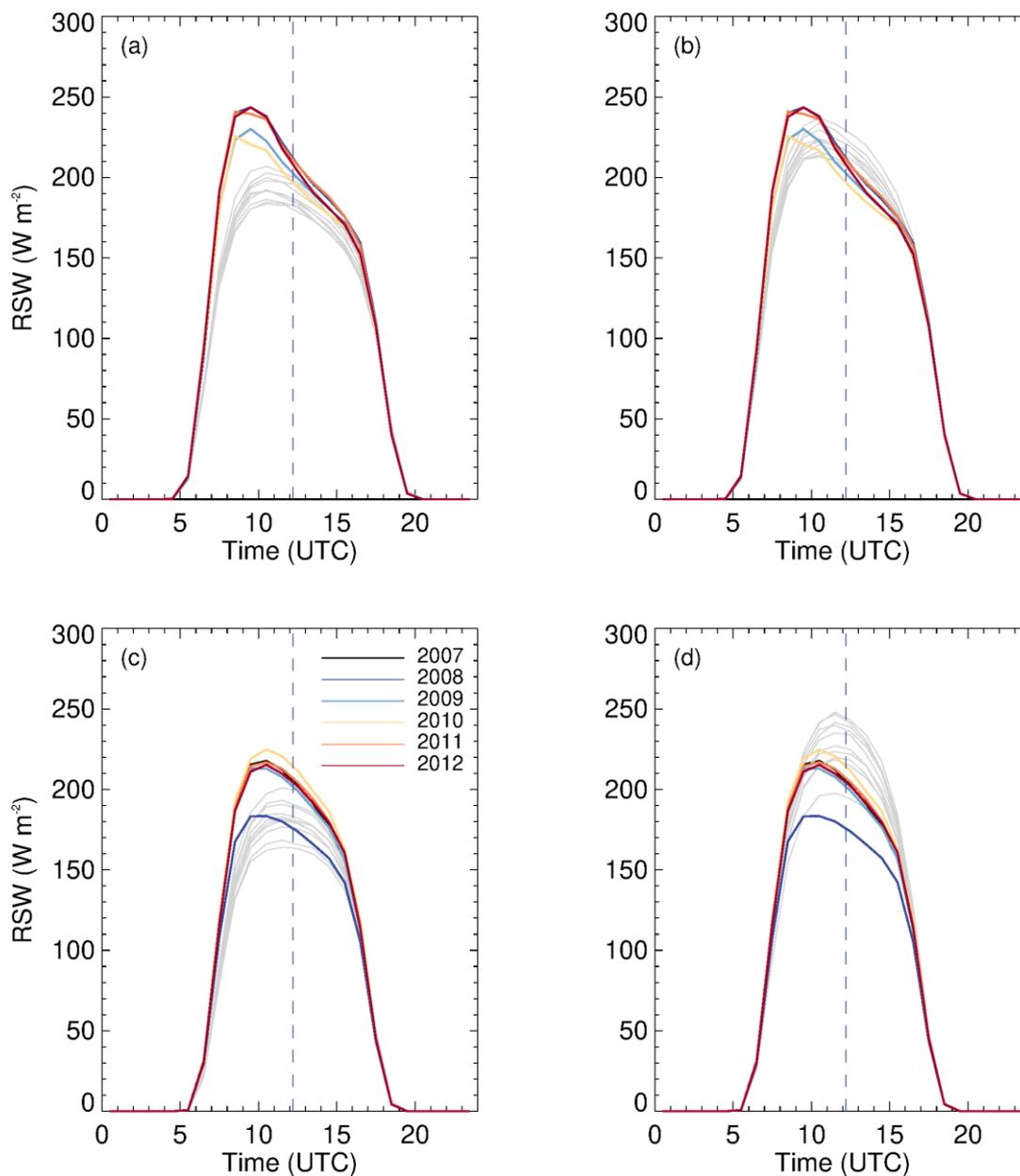



Figure 13: Monthly hourly mean RSW fluxes over the marine stratocumulus regions identified in Table 1 for January (a, b) and
415   **June (c, d). Coloured lines show GERB obs4MIPS fluxes for each year of the GERB observations. Grey lines show simulated fluxes**
for each simulation year for the HadGEM3-GC3.1 (a, c) and HadGEM3-GC5.0 (b, d) configurations. Dashed vertical lines show the
approximate timing of local noon.

Focusing on Figure 12 first, the observations show the classic signature of stratocumulus development and thickening in the morning prior to decay through the afternoon, manifested as a clear asymmetry in the RSW fluxes around local noon (e.g.

420 Gristey et al., 2019). This asymmetry is more pronounced in January than June. There is significant year-to-year variability in the magnitude of the observed fluxes (peak values can vary by up to 40 W m^{-2}) but they all have this characteristic phasing. The degree of observed inter-annual variability is smaller in January than June – behaviour that is also captured in the model simulations. While the simulations do exhibit a diurnal asymmetry they are unable to fully capture its observed magnitude. Similarly, although they show a constant diurnal phase from year-to-year, peak values for both model configurations are

425 typically delayed by an hour compared to the observations (Table 2). However, comparison of the GC3.1 and GC5.0 configurations does reinforce the impression that within these limitations, the latter is able to better capture the observed behaviour even if the improvement to the phasing between the configurations is slight.

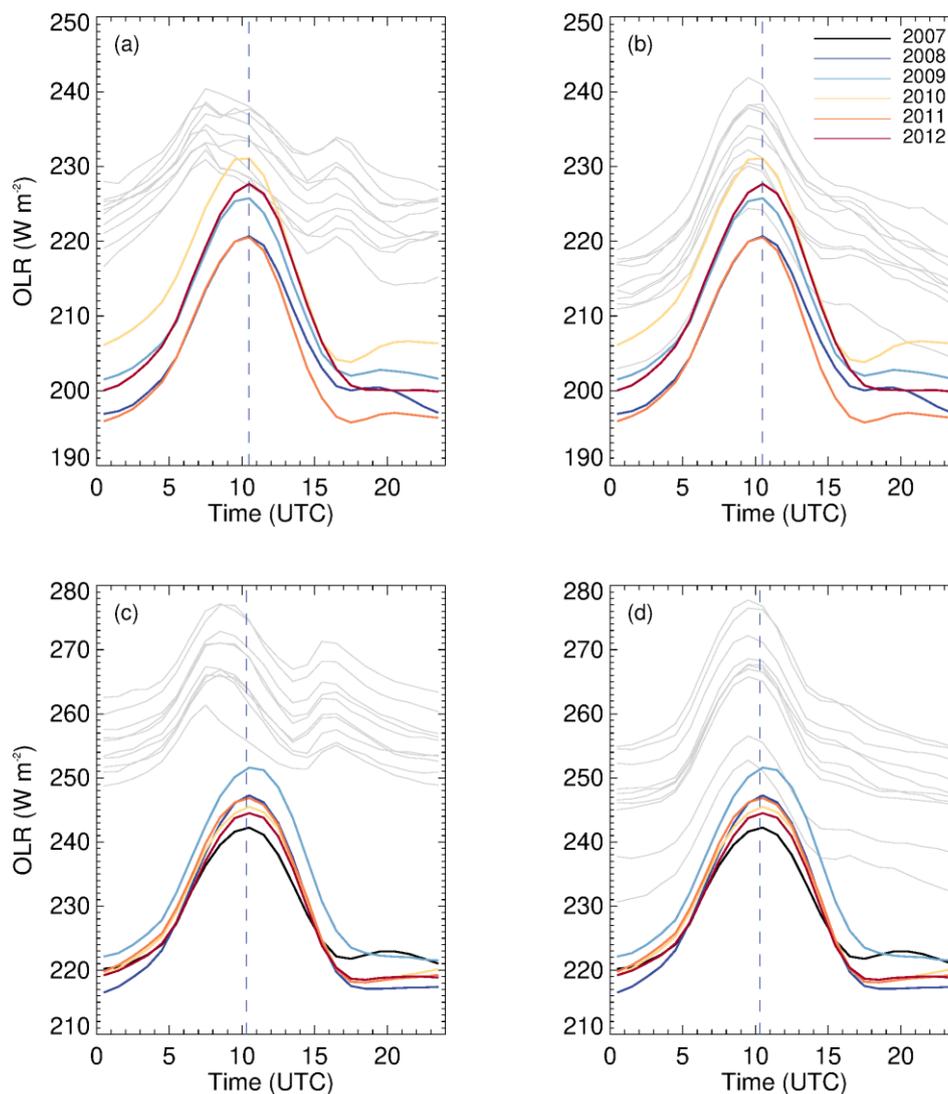



Figure 14: As Figure 13 for monthly hourly mean OLR fluxes over the deep convective regions identified in Table 1 for January in the top row (a and b) and June in the bottom row (c and d). Simulations for GC3.1 are shown in panels (a) and (c) and for GC5.0 in panels (b) and (d).

Turning to the deep convective regions (Figure 14), the observed OLR fluxes show a spread over the years considered which reaches of the order 10-15 W m⁻². The phasing of the cycle changes between the two months, with OLR fluxes reaching their maximum just after local noon in June and just before or at local noon in January. For both months the timing of the maximum is consistent from year to year although there is marked interannual variation in the shape of the cycle towards late afternoon and evening, particularly in January. The corresponding simulated values from GC5.0 highlight an improved ability to capture the general shape of the diurnal cycle, with the removal of what appears to be a spurious secondary peak in the OLR fluxes in late afternoon in GC3.1. The timing of the OLR maximum is shifted later in GC5.0 by between 1-2 hours and is more consistent with the observations, albeit still too early in the day. The amplitude of the cycle is also improved (Table 2). These improvements in the diurnal cycle are mainly driven by the introduction of the prognostic entrainment rate. Clearly other issues remain: in June the fluxes are consistently too high implying either missing convection or convection which is not vigorous enough. The interannual variability over the region is significantly higher than seen in the observations, which would be consistent with this interpretation. Both issues are present to a lesser extent in January. However, overall, the direction of travel from GC3.1 to GC5.0 is encouraging, particularly when viewed in a diurnally resolved comparison.

	South Atlantic Marine Stratocumulus (RSW)				African Deep Convection (OLR)			
	June (-16-10°E, 3-22°S)		January (-16-10°E, 3-28°S)		June (14-37°E, -2-12°N)		January (15-31°E,0-17°S)	
	Amp	Phase	Amp	Phase	Amp	Phase	Amp	Phase
GERB	135.0	10:30	141.0	09:30	17.7	10:30	16.9	10:30
GC3.1	114.8	11:30	111.7	10:30	9.2	08:30	8.1	07:30
GC5.0	147.6	11:30	130.8	10:30	13.8	09:30	12.6	09:30

Table 2: Amplitude and phase in multi-year June and January monthly mean RSW and OLR fluxes over marine stratocumulus and deep convective regions, as observed by GERB and simulated by the two configurations of HadGEM3. Amplitude, A, is defined as $A = Max(x_t - \overline{x_t})$ where x_t is the RSW or OLR flux as a function of hour through the day, and phase is the time (in UTC) at which the value of A is realised.

In climate models, the diurnal cycle of convection is typically evaluated using the diurnal cycle of precipitation (e.g. Stratton and Stirling, 2012). The remote sensing technology, spatio-temporal sampling and retrieval algorithms used in the precipitation retrievals introduce substantial uncertainty in the timing of the maximum of precipitation in the mean diurnal cycle (Dai et al., 2007; Minobe et al., 2020). The GERB dataset presented here provides a very accurate description of the monthly mean diurnal cycle- of the OLR and RSW fluxes, making it an excellent tool for the evaluation of the diurnal cycle of convection in models. It is worth noting that the minimum in OLR is delayed by around 3h with respect to the maximum in precipitation in convective regions (Dai et al., 2007), and therefore a combination of radiation and precipitation diagnostics can provide a more detailed picture of the evolution of precipitation and the anvil cloud associated with the development of deep convection.

4. Data availability

The GERB obs4MIPs OLR and RSW products presented in this paper are available from the Centre for Environmental Data Analysis (https://doi.org/10.5285/90148d9b1f1c40f1ac40152957e25467 (Bantges et al. 2023a) for the OLR and

460   https://doi.org/10.5285/57821b58804945deaf4cdde278563ec2 (Bantges et al, 2023b) for the RSW). The datasets are also available on the Earth System Grid Federation.

Model outputs used for the comparisons presented in Sect. 3 are available at
https://doi.org/10.5281/zenodo.10101394.

465   ## 5. Conclusions

The GERB obs4MIPs product is a new tool specifically designed to enable the evaluation of the diurnal cycle in top of the atmosphere fluxes, as captured by climate and Earth-system models. This paper has described in detail how it is derived from the baseline GERB measurements, providing monthly hourly mean RSW and OLR fluxes from 2007-2012 over the geographical region 60°N – 60°S, 60°E – 60°W for the months of November-February and May-August. Because of the

470   relative prevalence of missing observations, which occur both randomly throughout the record and systematically around the equinoxes, particular attention has been paid to the impact of missing data. Our results show how estimates of the instantaneous broadband 'GERB-like' fluxes from the SEVIRI narrowband imager can be used to fill missing GERB data. A scaling factor is calculated from the ratio of the monthly hourly 1° by 1° averages for the available GERB and matched GERB-like data and applied to the daily hourly GERB-like data. Using these scaled GERB-like fluxes to fill the missing GERB observations at the

475   daily hourly scale before averaging results in a reduction in uncertainty of more than a factor of 10 compared to the unfilled dataset. Even when a substantial amount of data is systematically removed, as is the case for GERB in the months of February and August, using the scaled GERB-like data to fill the missing periods leads to relatively small errors which are comparable to the error manifested in the unfilled dataset if just one day of data is missing.

We use the new filled product to perform a preliminary evaluation of two sets of *amip* type simulations for the HadGEM3

480   climate model. The two sets of simulations differ in their atmospheric components, with the newer configuration implementing a prognostic based entrainment rate scheme; a bimodal cloud scheme within entrainment zones associated with strong temperature inversions, and improvements to the influence of dry air entrainment on cloudy grid boxes. At the monthly mean level, there are noticeable differences in top-of-atmosphere fluxes, with an overall brightening in the newer GC5.0 configuration and an apparent strengthening of convection. Although such changes would be evident in comparisons with

485   existing radiative flux observations, further decomposing into the monthly hourly diurnal cycle allows insight into the amplitude and phasing of, in particular, different cloud regimes. Focusing on stratocumulus decks off south-western Africa and deep convection over Africa, the GERB obs4MIPs product indicates that the monthly mean changes are consistent with an improved diurnal amplitude and, in the case of the convective region, phase in these regions. Discrepancies still remain: for

example the simulated RSW asymmetry seen over the stratocumulus deck is not as pronounced as in the observations and tends to be delayed by around 1 hour compared to the observations, for both model configurations. Similarly, deep convection over Africa in boreal summer is too weak, and in both the winter and summer seasons it tends to occur slightly too early resulting in an earlier simulated peak in OLR than seen in the observations. Tying these initial results to the behaviour of the underlying driving fields will be one avenue for future investigation.

We have shown that the GERB obs4MIPs product is a very valuable complement to the traditional climatological averages of TOA radiation used for model evaluation. It provides a more direct connection with the model processes that control errors at both weather and climate timescales. Also, the fact that it is presented in CF-compliant netCDF format makes it extremely user-friendly, and ready to be incorporated into standard model evaluation tools like ESMValTool (Eyring et al., 2020).

Author contributions

The original draft manuscript was prepared by JER, HEB and RJB with substantial contributions to Sect. 3 from ABS. JER was responsible for developing the methodology of the GERB monthly hourly average product production and its filling. JER and RJB performed the error analysis related to missing data and data filling. RJB produced the software to generate the GERB obs4MIPs dataset and produced the datasets needed to perform the error analysis. ABS provided the HadGEM3 model output and COSP analysis and contributed expertise on the interpretation of the model-data differences. HEB carried out the comparisons between the HadGEM3 and GERB obs4MIPs data.

Competing interests

None of the authors have any competing interests.

References

Allan R., Slingo, A., Milton, S. and Brooks, M.,: Evaluation of the Met Office global forecast model using Geostationary Earth Radiation Budget (GERB) data. Q. J. Roy Meteo. Soc., 113, 1993-2010, https://doi.org/10.1002/qj.166, 2007.

Ansell C., Brindley, H., Pradhan, Y., and Saunders, R.: Mineral dust aerosol net direct radiative effect during GERBILS field campaign period derived from SEVIRI and GERB. J. Geophys. Res. – Atmos., 119, 4070-4086, https://doi.org/10.1002/2013JD020681, 2014.

Bantges, R.J.; Russell, J.E.; Brindley, H.E.: Obs4MIPs: Monthly-mean diurnal cycle of top of atmosphere outgoing longwave radiation from the GERB instrument (GERB-HR-ED01-1-1 rlut 1hrCM), v20231221. NERC EDS Centre for Environmental Data Analysis, 22 December 2023, https://dx.doi.org/10.5285/90148d9b1f1c40f1ac40152957e25467, 2023a.

Bantges, R.J.; Russell, J.E.; Brindley, H.E.: Obs4MIPs: Monthly-mean diurnal cycle of top of atmosphere outgoing shortwave radiation from the GERB instrument (GERB-HR-ED01-1-1 rsut 1hrCM), v20231221. NERC EDS Centre for Environmental Data Analysis, 22 December 2023, https://dx.doi.org/10.5285/57821b58804945deaf4cdde278563ec2, 2023b.

520 Bodas-Salcedo, A., Webb, M., Bony, S., Chepfer, H., Dufresne, J., Klein, S., Zhang, Y., Marchand, R., Haynes, J., Pincus, R., and John, V.: COSP: Satellite simulation software for model assessment, B. Am. Meteorol. Soc., 92, 1023-1043, https://doi.org/10.1175/2011BAMS2856.1, 2011

Bony, S. and Dufresne, J-L.,: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, Geophys. Res. Lett., 32, https://doi.org/10.1029/2005GL023851, 2005.

Brindley H. and Russell, J.; An assessment of Saharan dust loading and the corresponding cloud-free longwave direct radiative
525 effect from geostationary satellite observations. J. Geophys. Res. – Atmos., 114, https://doi.org/10.1029/2008JD011635, 2009

Brindley, H. E., and Russell, J. E.: Top of Atmosphere Broadband Radiative Fluxes from Geostationary Satellite Observations. In Comprehensive Remote Sensing: Vol. 5, Earth's Energy Budget, edited by S. Liang, 85–113. https://doi.org/10.1016/B978-0-12-409548-9.10368-9, 2017.

Chepfer, H., Bony, S., Winkler, D., Chiriaco, M., Dufresne, J-L., and Seze, G.: Use of CALIPSO lidar observations to evaluate
530 the cloudiness simulated by a climate model, Geophys. Res. Lett., 35, https://doi.org/10.1029/2008GL034207, 2008.

Clerbaux N., Russell, J., Dewitte, S., Bertrand, C., Caprion, D., De Paepe, B., Gonzalez Sotelino, L., Ipe, A., Bantges, R., and Brindley, H.: Comparison of GERB instantaneous radiance and flux products with CERES Edition-2 data, Remote Sens. Environ., 113, 102-114, https://doi.org/10.106/j.rse.2008.08.016, 2009.

Christopoulos, C. and Schneider, T.: Assessing biases and climate implications of the diurnal precipitaton cycle in climate
535 models, Geophys. Res. Lett., 48, https://doi.org/10.1029/2021GL093017, 2021.

Comer R., Slingo, A., and Allan, R.: Observations of the diurnal cycle of outgoing longwave radiation from the Geostationary Earth Radiation Budget instrument. Geophys Res Lett., 34, https://doi.org/10.1029/2006GL028229, 2007.

Dai, A., Lin X., and Hsu, K.: The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes. Clim. Dynam., 727–744. https://doi.org/10.1007/s00382-007-0260-y, 2007.

540 Dewitte S., Gonzalez, L., Clerbaux, N., Ipe, A., Bertrand C., and De Paepe, B.: The Geostationary Earth Radiation Budget Edition 1 data processing algorithms, Adv. Space Res., 41, 1906-1913, https://doi.org/10.1016/j.asr.2007.07.042, 2008.

Eyring V., Bony, S., Meehl, G., Senior, C., Stevens, B., Stouffer R., and Taylor, K.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev., 9, 1937-1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

545 Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., Arnone, E., Bellprat, O., Brötz, B., Caron, L.-P., Carvalhais, N., Cionni, I., Cortesi, N., Crezee, B., Davin, E. L., Davini, P., Debeire, K., de Mora, L., Deser, C., Docquier, D., Earnshaw, P., Ehbrecht, C., Gier, B. K., Gonzalez-Reviriego, N., Goodman, P., Hagemann, S., Hardiman, S., Hassler, B., Hunter, A., Kadow, C., Kindermann, S., Koirala, S., Koldunov, N., Lejeune, Q., Lembo, V., Lovato, T., Lucarini, V., Massonnet, F., Müller, B., Pandde, A., Pérez-Zanón, N., Phillips, A., Predoi, V., Russell, J., Sellar, A., Serva, F., Stacke, T.,

550 Swaminathan, R., Torralba, V., Vegas-Regidor, J., von Hardenberg, J., Weigel, K., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP, Geosci. Model Dev., 13, 3383–3438, https://doi.org/10.5194/gmd-13-3383-2020, 2020.

Futyan J., Russell, J., and Harries, J.: Determining cloud forcing by cloud type from geostationary satellite data. Geophys. Res.
555 Lett., 32, https://doi.org/10.1029/2004GL022275, 2005.

Gristey, J., Chiu, J. C., Gurney, R., Morcrette, C., Hill, P., Russell, J., and Brindley, H.: Insights into the diurnal cycle of global Earth outgoing radiation using a numerical weather prediction model. Atmos. Chem. Phys., 18, 5129-5145, http://doi.org/10.5194/acp-18-5129-2018, 2018.

Greuell, W., van Meijgaard, E., Clerbaux, N., and Meirink, J.: Evaluation of model-predicted top-of-atmosphere radiation and
560 cloud parameters over Africa with observations from GERB and SEVIRI, J. Climate, 24, 4015-4036, https://doi.org/10.1175/2011JCLI3856.1, 2011.

Guichard, F., Petch, J., Redelsperger, J.-L., Bechtold, P., Chaboureau, J-P., Cheinet, S., Grabowski, W., Grenier, H., Jones, C. G., Kohler, M., Piriou, J.-M., Tailleux, R., Tomasini, M.: Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single column models. Q. J. Roy. Meteor. Soc., 130, 3139-3172,
565 https://doi.org/10.1256/qj.03.145, 2004.

Harries J., Russell, J., Hanafin, J., Brindley, H., Futyan, J., Rufus, J., Kellock, S., Matthews, G., Wrigley, R., Last, A., Mueller, J., Mossavati, R., Ashmall, J., Sawyer, E., Parker, D., Caldwell, M., Allan, P., Smith, A., Bates, M., Coan, B., Stewart, B., Lepine, D., Cornwall, L., Corney, D., Ricketts, M., Drummond, D., Smart, D., Cutler, R., Dewitte, S., Clerbaux, N., Gonzalez, L., Ipe, A., Bertrand, C., Joukoff, A., Crommelynck, D., Nelms, N., Llewellyn-Jones, D., Butcher, G., Smith, G., Szewczyk,
570 Z., Mlynczak, P., Slingo, A., Allan, R. and Ringer, M.: The Geostationary Earth Radiation Budget (GERB) Project. B. Am. Meteorol. Soc., 86, 945-960, https://doi.org.10.1175/BAMS-86-7-945, 2005.

Haywood, J., Johnson, B., Osborne, S., Mulcahy, J., Brooks, M., Harrison, M., Milton, S., Brindley, H: Observations and modelling of the solar and terrestrial radiative effects of Saharan dust: a radiative-closure case study over oceans during the GERBILS campaign. Q. J. Roy. Meteor. Soc., 137, 1211-1226, https://doi.org/0.1002/qj.770, 2011.

575 Hohenegger, C. and Stevens, B.: Controls and impacts of the diurnal cycle of deep convective precipitation, J. Adv. Model. Earth Sy., 5, 801–815, https://doi.org/10.1002/2012MS000216, 2013.

Kato, S and Loeb, N.: Twilight Irradiance Refelected by the Earth Estimated from Clouds and the Earth's Radiant Energy System (CERES) Measurements. *J. Climate*, **16**, 2646-2650, https://doi.org/10.1175/1520-0442(2003)016<2646:TIRBTE>2.0.CO;2, 2003.

580 Klein, S. and Jakob, C.: Validation and sensitivies of frontal clouds simulated by the ECMWF model, Mon. Weather Rev., 127, 2514-2531 https://doi.org/10.1175/1520-0493(1999)127<2514:VASOFC>2.0.CO;2, 1999.

Mackie, A., Palmer P., and Brindley, H.: Characterising energy budget variability at a Sahelian site: a test of NWP model behaviour, Atmos. Chem. Phys., 17, 15095-15119, https://doi.org/10.5194/acp-17-15095-2017, 2017.

Milton, S., Greed, G., Brooks, M., Haywood, J., Johnson, B., Allan, R., Slingo A., and Grey, W.: Modeled and observed
585   atmospheric radiation balance during the West African dry season: Role of mineral dust, biomass burning aerosol, and surface
albedo, J. Geophys. Res. – Atmos., 113, https://doi.org/10.1029/2007JD009741, 2008.

Minobe, S., Park J., and Virts, K.: Diurnal cycles of precipitation and lightning in the tropics observed by TRMM3G68,
GSMaP, LIS and WWLLN. J. Climate, 33, 4293-4313, https://doi.org/10.1175/JCLI-D-19-0389.1, 2020.

Morcrette, C.J.: Improvements to a prognostic cloud scheme through changes to its cloud erosion parametrization. Atmos. Sci.
590   Lett., 13: 95-102. https://doi.org/10.1002/asl.374, 2012.

Mulcahy J., Jones, C., Sellar, A., Johnson, B., Boutle, I., Jones A., Andrews, T., Rumbold, S., Mollard, J., Bellouin, N.,
Johnson, C., Williams, K., Grosvenor, D., McCoy, D.: Improved aerosol processes and effective radiative forcing in HadGEM3
and UKESM1. Journal of Advances in Modeling Earth Systems, 10, 2786– 2805, doi:10.1029/2018MS001464

Nam, C., S. Bony, J-L Dufresne and H. Chepfer, 2012. The 'too few, too bright' tropical low cloud problem in CMIP5 models,
595   Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL053421, 2018.

Parfitt R., Russell, J., Bantges, R., Clerbaux N., and Brindley, H.: A study of the time evolution of GERB shortwave calibration
by comparison with CERES Edition-3A data. Remote. Sens. Environ., 186, 416-427,
https://doi.org/10.1016/j.rse.2016.09.005, 2016.

Pearson K., Hogan, R., Allan, R., Lister, G., and Holloway, C.: Evaluation of the model representation of the evolution of
600   convective systems using satellite observations of outgoing longwave radiation. J. Geophys. Res. – Atmos., 115,
https://doi.org/10.1029/2010JD014265, 2010.

Schmetz J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A.: An introduction to Metosat Second Generation
(MSG). B. Am. Meteorol. Soc., 83, 977-991, https://doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2, 2002.

Slingo A., Ackerman, T., Allan, R., Kassianoc, E., McFarlance, S., Robinson, G., Barnard, J., Miller, M., Harries, J., Russell,
605   J. and Dewitte, S.: Observations of the impact of a major Saharan dust storm on the atmospheric radiation balance. Geophys.
Res. Lett. 33, http://doi.org/10.1029/2006GL027869, 2006

Stratton R. and Stirling, A.: Improving the diurnal cycle of convection in GCMs. Q. J. Roy. Meteor. Soc., 138, 1121-1134,
https://doi.org/10.1002/qj.991, 2012

Tan J., Huffman, G., Bolvin, D. and Nelkin, E.: Diurnal cycle of IMERG V06 precipitation, Geophys. Res. Lett., 46,
610   https://doi.org/10.1029/2019GL085395, 2019

Waliser D., Gleckler, P., Ferraro, R., Taylor, K., Ames, S., Biard, J., Bosilovich, M., Brown, O., Chepfer, H., Cinquini, L.,
Durack, P., Eyring, V., Mathieu, P.-P, Lee, T., Pinnock, S., Potter, G., Rixen, M., Saunders, R., Schulz, J., Thépaut, J.-N. and
Tuma, M.: Observations for Model Intercomparison Project (Obs4MIPs): status for CMIP6. Geosci. Model Dev., 13, 2945-
2958, https://doi.org/10.5194/gmd-13-2945-2020, 2020.

615   Van Weverberg, K., Morcrette, C. J., Boutle, I., Furtado, K., and Field, P. R.: A Bimodal Diagnostic Cloud Fraction
Parameterization. Part I: Motivating Analysis and Scheme Description. Mon. Weather Rev., 149, 841–857,
https://doi.org/10.1175/MWR-D-20-0224.1, 2021a.

Van Weverberg, K., Morcrette, C. J., and Boutle, I.: A Bimodal Diagnostic Cloud Fraction Parameterization. Part II: Evaluation and Resolution Sensitivity. Mon. Weather Rev., 149, 859–878, https://doi.org/10.1175/MWR-D-20-0230.1, 2021b.

620 Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geosci. Model Dev.,

625 12, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019, 2019.

Webb, M., Lock, A. Bodas-Salcedo, A., Bony, S., Cole, J., Koshiro, T., Kawai, H., Lacagnina, C., Selten, F., Roehrig, R. and Stevens, B.: The diurnal cycle of marine cloud feedback in climate models. Clim. Dynam., 44, 1419-1436, https://doi.org/10.1007/s00382-014-2234-1, 2015.

Webb, M., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C., Chadwick, R., Chepfer, H., Douville, H., Good, P., Kay,

630 J., Klein, S., Marchand. R., Medeiros, B., Siebesma, A., Skinner, C., Stevens, B., Tselioudis, G., Tsushima, Y. and Watanabe, M.: The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6, Geosci. Model Dev., 10, 359–384, https://doi.org/10.5194/gmd-10-359-2017, 2017

Williams, K., Copsey, D., Blockley, E., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H., Hill, R., Hyder, P., Ineson, S., Johns, T., Keen, A., Lee, R., Megann, A., Milton, S., Rae, J., Roberts, M., Scaife, A., Schiemann, R.,

635 Storkey, D., Thorpe, L., Watterson, I., Walters, D., West, A., Wood, R., Woollings, T., and Xavier, P.: The Met Office Global Coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *J. Adv. Model. Earth Sys*., **10**, 357-380, https://doi.org/10.1002/2017MS001115, 2018.

Williams K. and Bodas-Salcedo, A.: A multi-diagnostic approach to cloud evaluation, Geosci. Model Dev., 10, 2547–2566, https://doi.org/10.5194/gmd-10-2547-2017, 2017.

640 Wilson, D.R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D. and Morcrette, C. J.: PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. Q. J. Roy. Meteor. Soc., 134: 2093-2107. https://doi.org/10.1002/qj.333, 2008a.

Wilson, D.R., Bushell, A.C., Kerr-Munslow, A.M., Price, J.D., Morcrette, C.J. and Bodas-Salcedo, A.. 2008b. PC2: A prognostic cloud fraction and condensation scheme. II: Climate model simulations. Q. J. Roy. Meteor. Soc., 134: 2109-2125. https://doi.org/10.1002/qj.332, 2008b

645 Yang, G. and Slingo, J.: The diurnal cycle in the tropics, Mon. Weather Rev., 129, 784-801, https://doi.org/10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2, 2001.