# A comparison between HadCM3 integrations for COAPEC
using Beowulf (UM version 4.5) and Cray T3E (UM version 4.4)

Alan Iwi & Bryan Lawrence
Rutherford Laboratory
July 2004

## Executive Summary

This report documents a comparison between two long integrations of HadCM3 which have been performed for COAPEC: one on a Cray T3E using Unified Model version 4.4, the other on a Beowulf cluster using UM version 4.5.

A detailed comparison of difference plots between the two models shows a number of statistically significant differences in some fields, but these are small in magnitude, and both models are judged to be equally valid for scientific use for climate studies. Data are available from both runs at the NCAS/British Atmospheric Data Centre, and can be supplied to the research community.

## 1 Introduction

The atmospheric science community has expressed concern as to whether Beowulf clusters are suitable platforms for climate model integrations based on the Met Office Unified Model (UM). (The issue as to whether the Beowulf clusters are in general suitable for atmospheric modelling has long been resolved.) There are three particular issues:

- Whether 64-bit integrations of fully coupled models are stable and are comparable with integrations on more traditional architectures,
- whether 32-bit integrations of any version of the Unified Model remain stable and are comparable with integrations on more traditional architectures, and
- more specifically, in the latter case, even if the atmosphere integrations are reliable, will the ocean integrations be stable?

Of course these quesitons are actually not specific to Beowulf clusters and need to be answered for any new architecture (and model version).

This report addresses the first of these concerns for one instance; a fully coupled climate version of the UM - HadCM3, by comparison of two 64-bit integrations:

- A 100-year subset of a 1000-year integration, performed on a Cray T3E within the Met Office, using UM version 4.4
- A 500-year integration, performed on a Beowulf cluster at Rutherford Laboratory, using UM version 4.5.

Both integrations were performed for the COAPEC programme, and the output of both is available in the COAPEC section of the NCAS/British Atmosphere Data Centre. Obviously there were significant differences both between the hardware and software configurations of the models, so the comparison starts from the premise that the two models are different, so the questions are:

1. How different are the two climates?
2. Is the climate of the Beowulf simulation realistic and suitable for scientific use?

The remainder of this document consists of four sections: section 2 outlines a more detailed comparison of the model configurations and the methodology used; section 3 gives the main results from this detailed comparison; section 4 addresses the first question above: "how different are the two simulations?", based on a more detailed case-study of one of the quantities shown in section 3; section 5 summarises and draws some conclusions about the overall usability of the Beowulf simulations.

# 2 Description of models and comparison methods

## 2.1 Description of models

The model runs described in this report were both time-slice experiments of the fully coupled climate model, HadCM3, with greenhouse gas forcing held constant at pre-industrial levels, and without the use of air-sea heat flux correction (although a salinity flux correction was applied).

In both experiments, all calculations were performed using 64-bit data types.

### 2.1.1 Differences in model code

Although the two model runs can both be described as integrations of HadCM3, they were run using model code taken from different stages of development of the Unified Model. UM version 4.4 was used for the Cray T3E integration, whereas version 4.5 was used on the Beowulf cluster. In fact, these model version numbers oversimplify the situation somewhat. In reality, there are frequent minor changes to the model code: the integration at 4.4 contained a number of code modifications which later became a standard part of version 4.5; likewise the integration at 4.5 contained further code modifications developed after the release of 4.5. Most of these code changes consist either of bug fixes or of optimisations. In a few cases the changes introduce wholly new physics schemes, but often these are only enabled if the user specifically selects the new scheme as an alternative. In summary: there is a long list of code changes between the two models, although in terms of model evolution they are very close.

Amongst the list of changes are a few salient physics scheme differences between the model integrations: in the atmosphere, updated spectral coefficients for longwave radiation, a 3-dimensional $CO_2$ field, and an accurate treatment of precipitation phase change; in the ocean, the Griffies isopycnal diffusion scheme replacing the older Redi scheme, and a new parameterisation of Mediterranean and Hudson bay outflow.

### 2.1.2 Model initialisation

There are also differences in model initialisation: both integrations were performed from initial conditions supplied by the Met Office from previous integrations of HadCM3, but the version 4.5 run used a run dated 2789, and the version

4.4 a run dated 1849. While these dates are arbitrary (as external parameters such as greenhouse gases are constant), they do reflect a greater level of spin-up equilibration in the version 4.5 initialisation. This difference was in part due to the fact that it was not originally intended to run the V4.5 simulation for as long, and it was thought the better spun-up start would result in a better quality simulation. Given the other significant differences between the simulations it was thought this would not be a problem, but it did lead to a systematic offset between the two simulations as discussed below.

#### 2.1.3 Nomenclature

In order to reflect the fact that the integrations being compared in this report differed not only in computer hardware platform, but also in the model setup, the following labels will be used:

- **the run at version 4.5 on the Beowulf cluster will be denoted "B45"**
- **the run at version 4.4 on the Cray T3E will be denoted "C44"**

### 2.2 Description of comparison techniques

To show the effects of model spinup, timeseries of the following annually averaged quantities were calculated:

- surface temperature: global mean, area-weighted average over all gridboxes
- surface temperature and precipitation: unweighted average over a $3 \times 4$ rectangle of gridboxes covering the British Isles
- ocean root-mean-square horizontal velocity on various levels: this is calculated from monthly mean total ocean velocity (sum of barotropic and baroclinic components), then an area-weighted global horizontal average over all ocean gridboxes ($\overline{u^2 + v^2}$) is taken at each level, regardless of ocean basin, then annually averaged and the square-root taken.

This report mostly presents anomaly fields, calculated and presented as B45 – C44, to show detailed differences between the simulations.

For the atmosphere, the set of diagnostics was chosen (as far as available) based on those used by the "validation note" tests which are performed within the Met Office (although that specific test suite was not available for technical reasons[1]). Mostly these fields are seasonal averages for DJF and JJA, although a few fields for MAM/SON are also included. Additionally some representative fields are shown for the ocean; some ocean fields which have little interseasonal variability are shown as annual rather than seasonal means.

For all anomaly calculations, an associated significance level was calculated. This was performed separately at each grid point to be plotted (after any spatial and seasonal averaging), using the composite multi-annual timeseries [2] for the given season. Anomalies locally significant at the 95% confidence level using a two-tailed $t$-test are indicated on the plots. The statistical implications of the use of this technique are further discussed in section 4.

## 3 Detailed comparison of model runs

This section comprises a discussion of many of the differences seen between the runs. It will be seen that most of them are minor, and attributable to either the initial conditions or the physics changes between the models.

[1] It has not been ported from the Cray; neither does it support the NetCDF file format to which the output files had been converted.

[2] In all cases, the maximum number of available complete periods was used for any composite; for example, in C44 (100 years January–December) the JJA composites have 100 members, but the DJF composites, and the annual mean composites (which were constructed from seasonal means, hence run from December to November), each have 99 members.

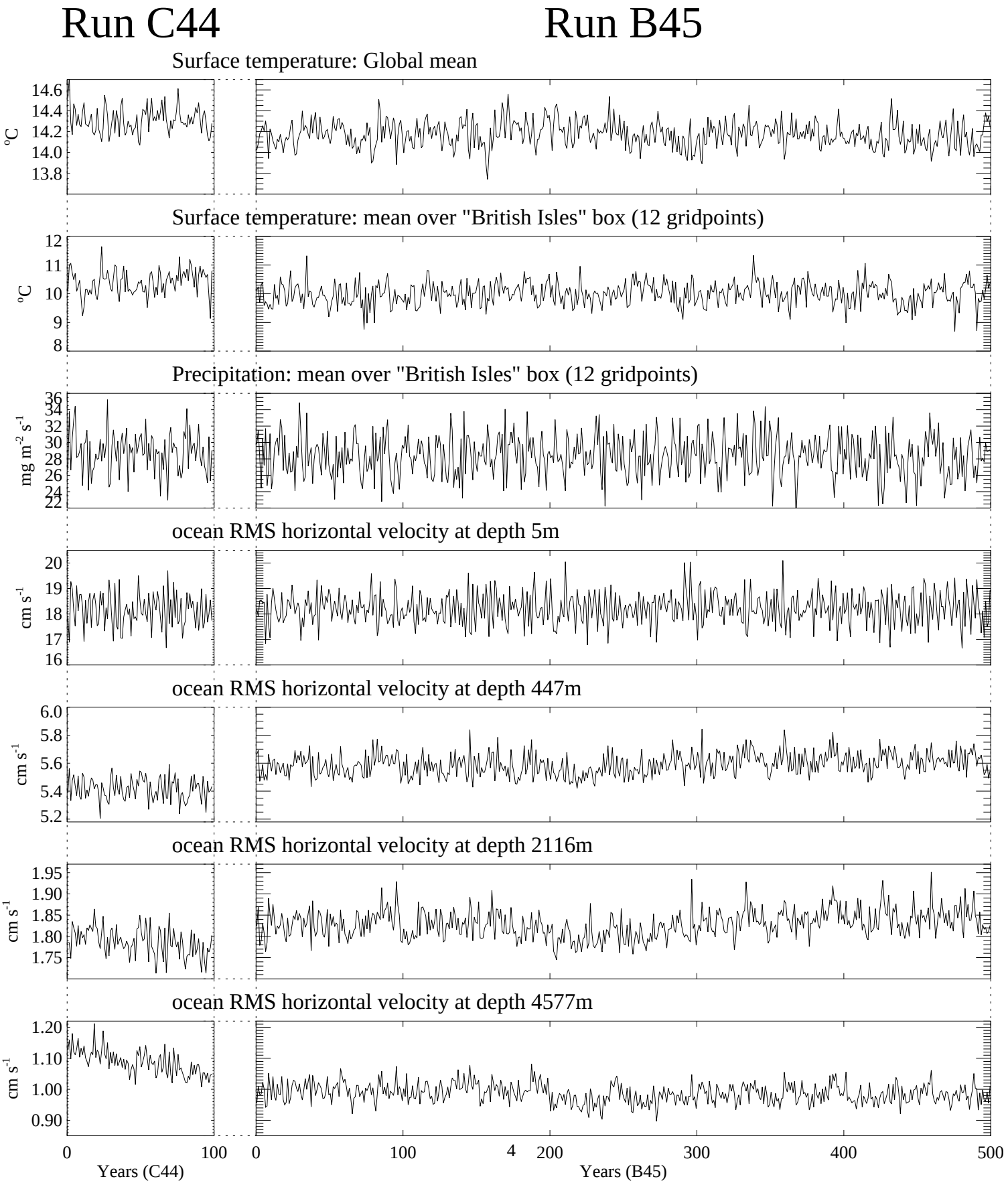


Figure 1: Timeseries of selected quantities from the model integrations (see text)

## 3.1 Timeseries

Figure 1 shows timeseries plots from the integrations for the quantities described in section 2.2 (using identical scales for C44 and B45 in both time and independent variable axes).

These timeseries suggest an overall broad similarity between the results from runs B45 and C44 in the selected diagnostics. Average values are similar, although global mean surface temperature is approximately 0.1 K warmer in C44 compared to B45. Interannual variability of the various diagnostics shown has similar magnitude in the two integrations, and also similar temporal structure (i.e. high- versus low-frequency variability).

The timeseries also suggest that, at least as regards the selected diagnostics, integration B45 is well equilibrated without obvious drifts. Integration C44 (based on less spun-up initial conditions, as discussed) is largely well equilibrated, although there is a drift evident in the deep ocean. This confirms the appropriateness of using the entire run duration as an averaging period in the anomaly plots shown in later figures (which do not extend down to ocean levels where the drift is seen in C44).

There is actually also a small bug in the C44 run which results in atmospheric mass loss, and a gradual drift downwards in the surface pressure, as can be seen in figure 2. The rate of decrease (initially $0.025$ hPA/decade, though slightly non-linear with time) results in a mean surface pressure anomaly of 1.7 hPA between the two simulations, attributable to the fact that B45 was initialised from a C44 run after approximately 1000 years of mass-loss.

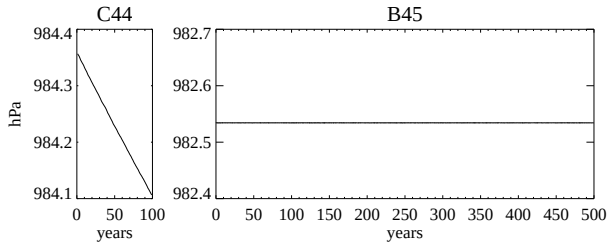


Figure 2: Annual-mean global-mean surface pressure from the two integrations

## 3.2 Atmospheric dynamical fields

In all the figures to follow, the quantity shown is the anomaly of the named field, with the sign B45 minus C44, which is also the meaning given in the discussion to the sign of the anomaly. Solid contours denote positive anomalies; dashed contours negative anomalies; faint dotted lines denote the zero contour. The labelling "[c=...]" in the plot titles denotes the contour interval. Shaded regions correspond to the 5% statistical significance as previously described. For brevity, where possible we have chosen to highlight differences by choosing examples from one (rather than both) of the solsticial seasons.

As explained above, there is a surface pressure difference between the simulations which is attributable to a (minor) bug in the C44 simulation. While the pressure difference (1.7 hPa) is small, it could have a slight effect on the location of the hybrid $\sigma$-$p$ levels which would probably be most obvious in comparisons near orography. In any case, because pressure *gradients* are more important dynamically than the absolute value, the pressure anomaly fields are shown with the global mean removed (see figure 3). While this is relatively featureless, a few anomalies (of order 0.5–1 hPa) exist, most notably a positive anomaly in the North Atlantic in spring/summer, a persistent positive anomaly over the Himalayas, and various anomalies in the southern ocean. In addition, the American mountain chains (Rockies, Andes) show significant seasonally consistent differences. We have not yet attempted to ascertain as to whether that can be attributed to the surface pressure difference or other processes.

The global mean geopotential height (not shown) shows a systematic negative anomaly of $15 - 20$m with height and latitude which is consistent with the pressure difference. When the global mean is removed (figure 4), a weak Rossby wave train in the mid-latitude is significant in DJF, but not so significant in JJA.

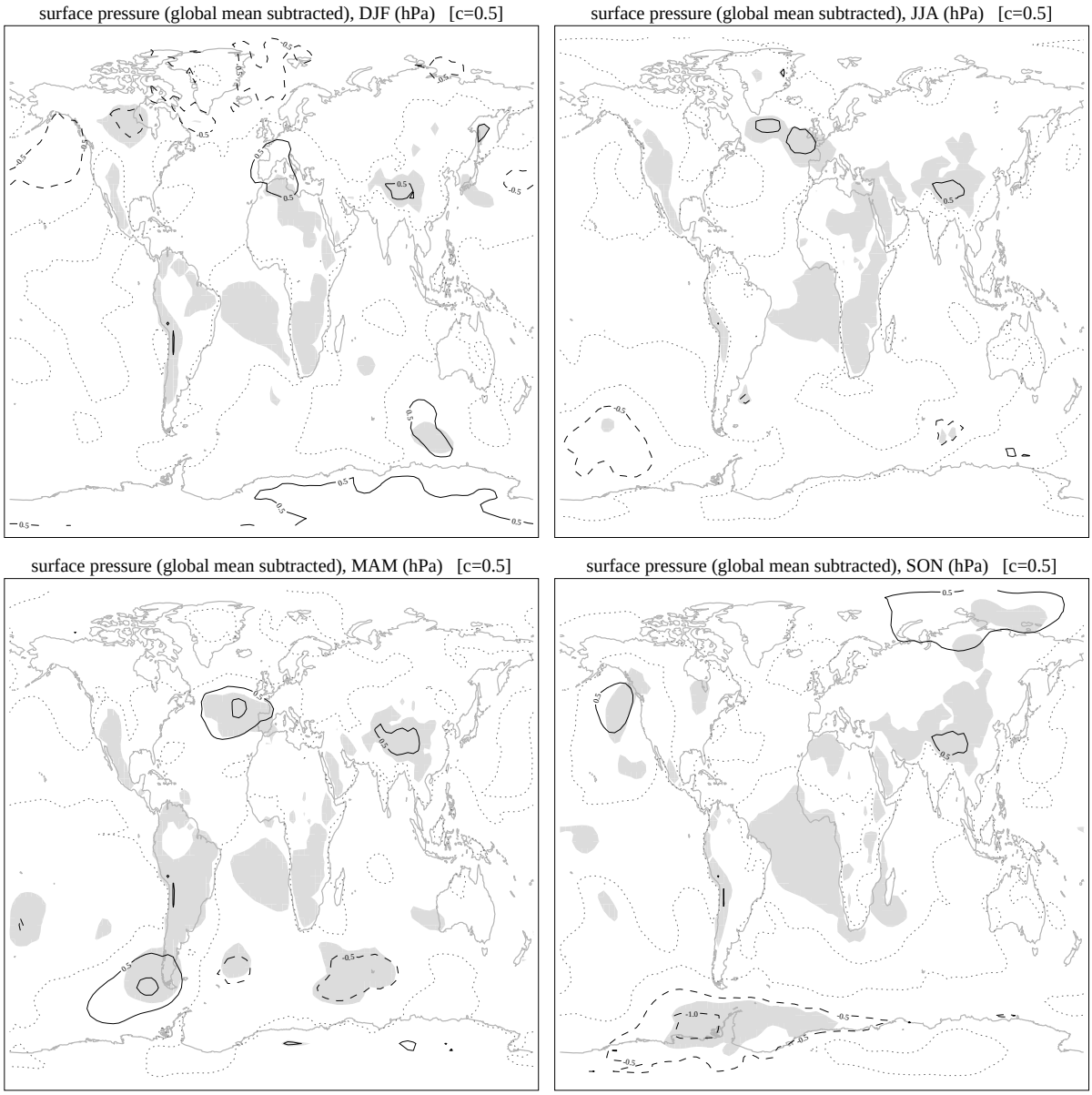


Figure 3: Surface pressure differences for all seasons (after subtraction of global mean).

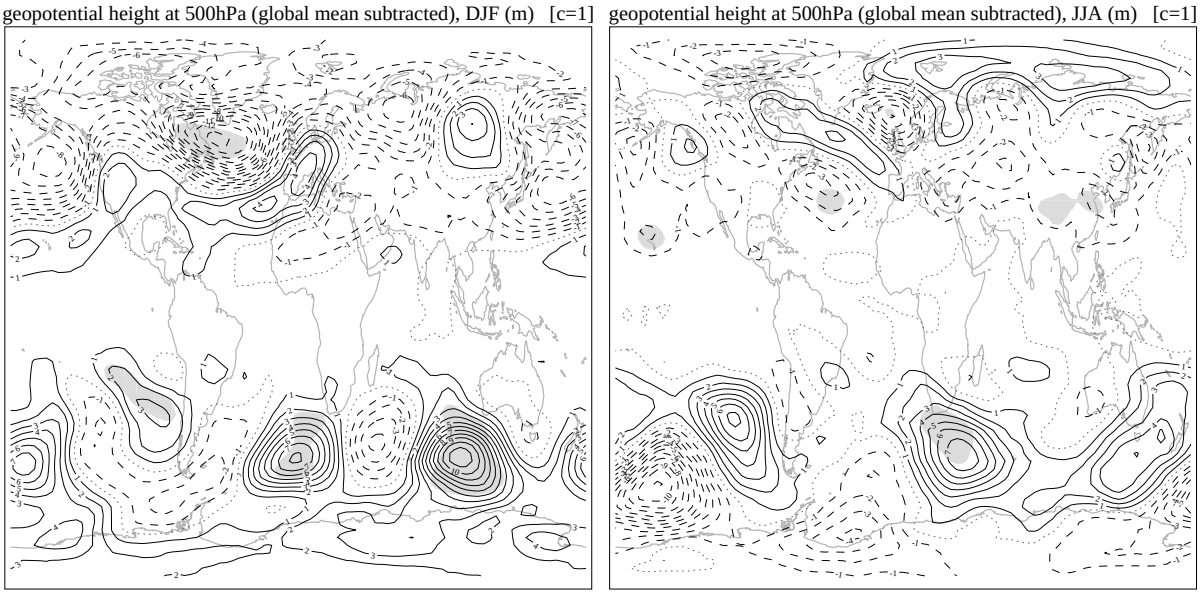


Figure 4: Geopotential anomalies at 500 hPa (after subtraction of global mean)

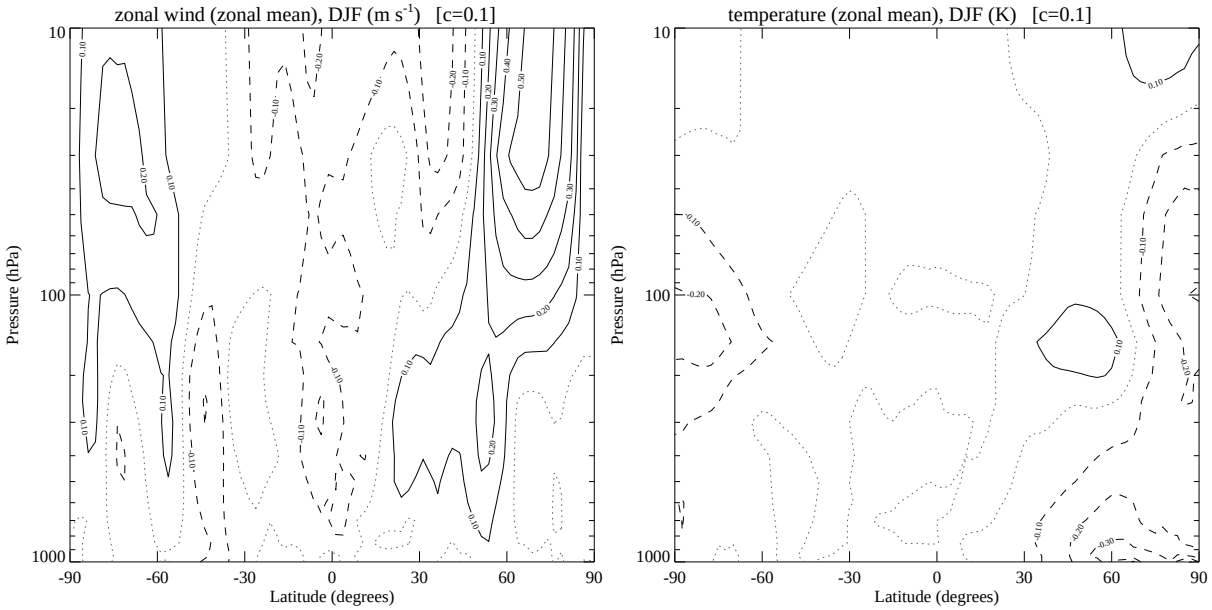


Figure 5: Zonal-mean zonal wind and temperature anomalies for DJF

Upper-air winds and temperatures show minimal anomalies, none significant in the zonal mean (e.g. figure 5). An examination of the longitudinal mean structures (not shown) shows weak significant features associated with the Rossby-wave train anomaly seen already and nothing else of signficance.

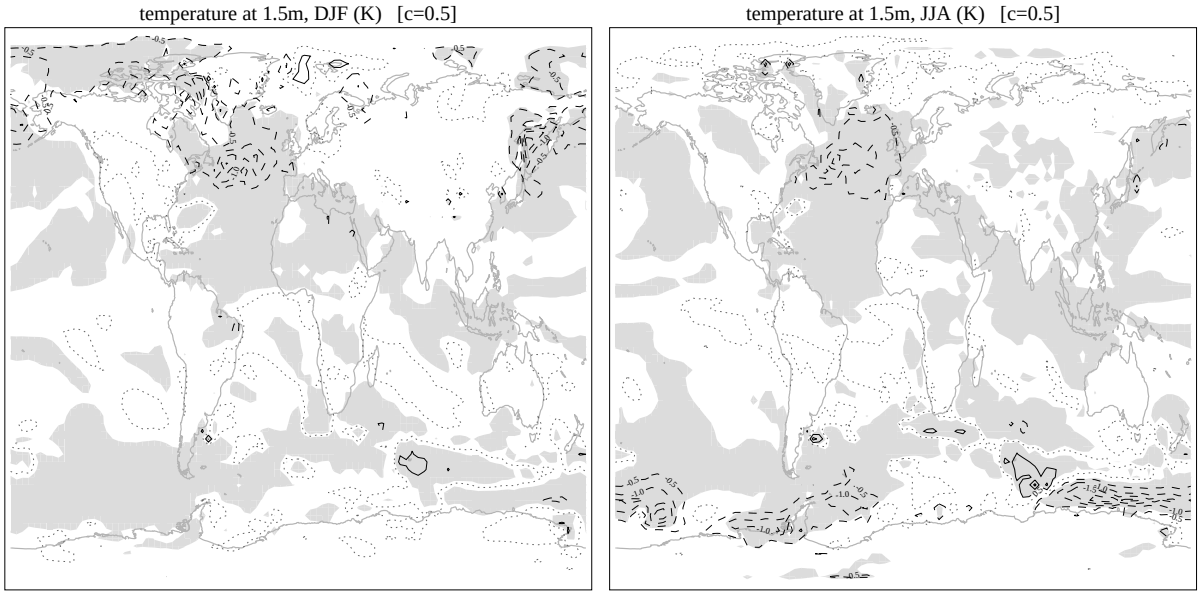


Figure 6: 1.5m temperature anomalies

The near-surface air temperature (figure 6) is very slightly different throughout the globe (but especially over the oceans), with a persistent anomaly in the North Atlantic, and a winter anomaly above the polar oceanic region. Again these are probably consistent with the global mean air-mass difference associated with the initialisation. Near-surface zonal-wind anomalies are rather uninteresting and are not shown. There is a region of the tropical Atlantic in which the meridional wind is persistently different (of order $\sim 0.2\mathrm{m\,s}^{-1}$ throughout the year, see figure 7).

## 3.3 Humidity, Cloud, and Precipitation

There are significant physics changes between V4.4 and V4.5 of the UM, and so it is important not to over-interpret differences in the water and cloud variables.

There are large differences in the relative humidity in the troposphere (e.g. see figure 8 for DJF), but these are not significant. Differences in the stratosphere are signficant, but small — of order 1–3% of relative humidity.

Similar results are found for the specific humidity (not shown), with the largest differences in the middle stratosphere where the anomalies are of order of 30% of the background value. However, because there is no methane oxidation in these models, the stratospheric water budget is rather unphysical anyway (the specific humidity decreases with height instead of increasing).

Zonally averaged cloud water and vertically resolved layer and convective cloud fraction do not show significant differences between the simulations (zonal mean convective cloud fraction anomalies, not shown, are less than 1% nearly everywhere). However, an examination of the geographical distribution of clouds does show signficant differences (e.g. JJA: figure 9, where the North Atlantic cloud anomaly may be associated with the surface temperature anomaly seen in figure 6).

The geographical anomalies in cloud distribution are associated with precipitation differences in the DJF season (figure 10), although not as obviously in the JJA season. A comparison of the convective precipitation with the large-scale

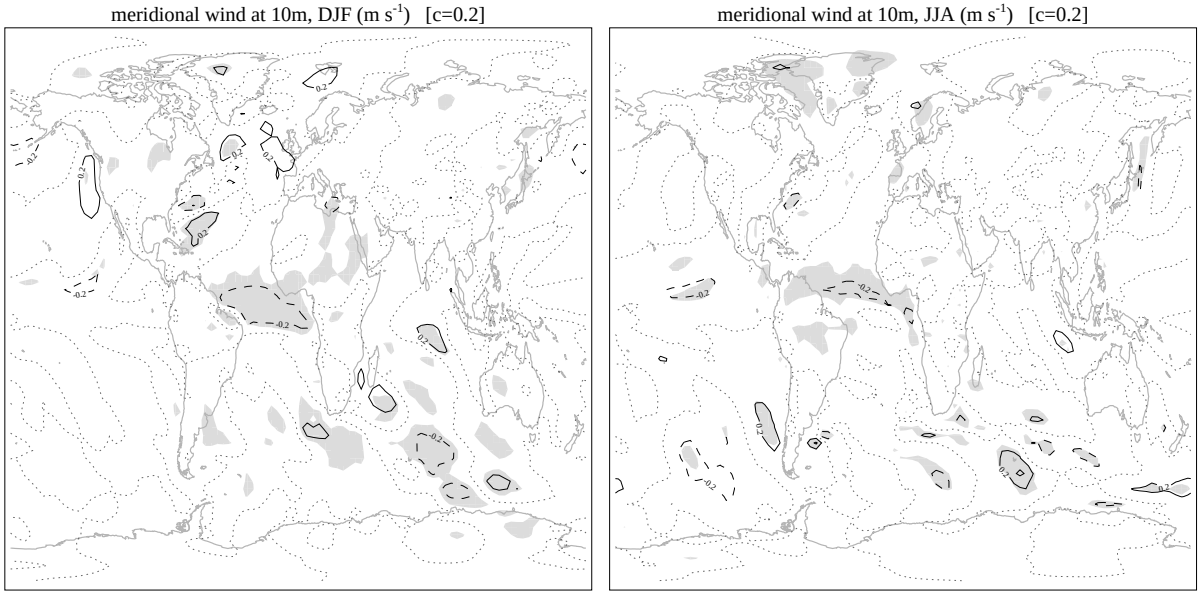


Figure 7: Meridional wind anomaly at 10m.

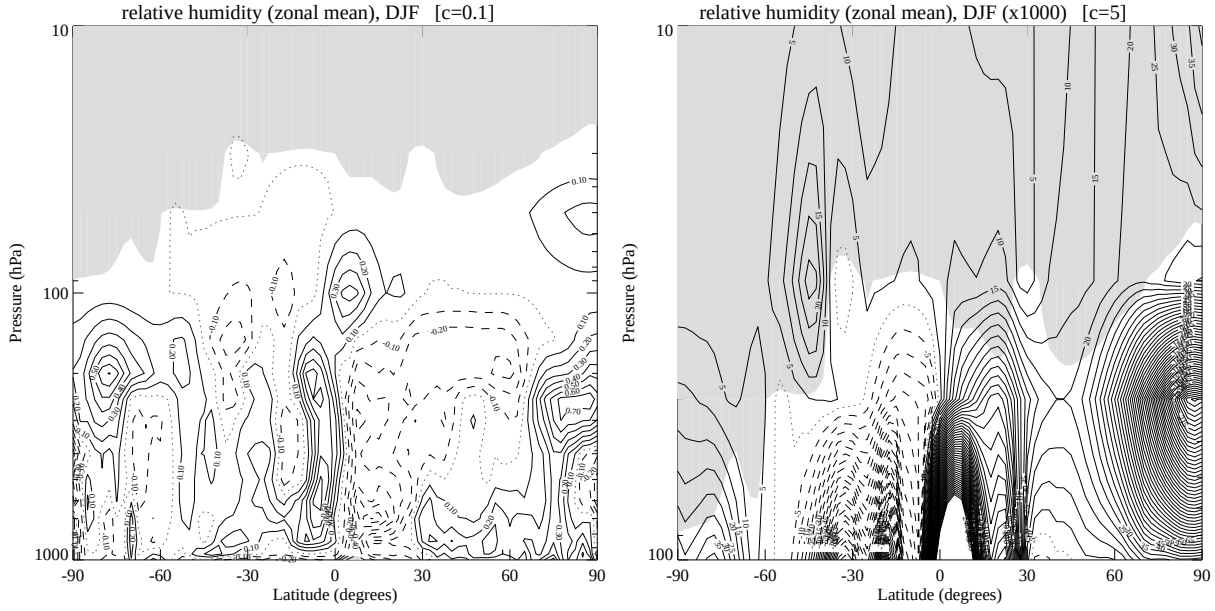


Figure 8: Relative humidity anomaly in DJF, with tropospheric contour intervals and stratospheric contour intervals (Left and right panels respectively).

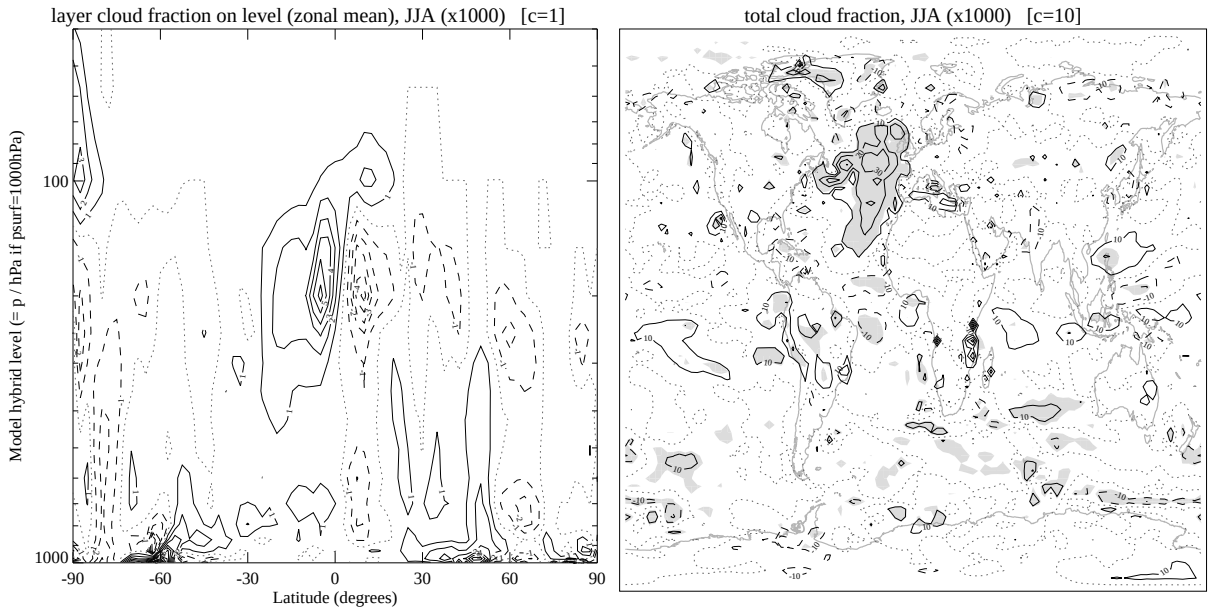


Figure 9: JJA Anomaly comparison between zonal mean layer cloud fraction and geographical distribution of total cloud fraction

precipitation (figure 11) shows that the differences in precipitation in the tropical Atlantic are due entirely to convective precipitation (which suggests that the cloud differences in this region are also convective, despite the small zonal mean differences in convective cloud). These differences in precipitation also lead to a significant difference in soil moisture in the Brazilian region of South America (not shown).

Figure 12 shows that the boundary layer height anomalies are of order a few tens of metres, and the main differences are consistent with the temperature and convection anomalies already noted.

### 3.3.1 Radiation

As one might expect, the radiation anomalies are dominated by the effects of the cloud differences (e.g. outgoing radiation — figure 13). Although the surface downwelling radiation is not shown here, the patterns are consistent with what is seen in the top of atmosphere (outgoing) fluxes. The differences can be summarised as:

- Shortwave: increased cloudiness increases outgoing shortwave but decreases surface downward shortwave (due to scattering).
- Longwave: increased cloudiness increases surface downward longwave, but reduces outgoing longwave (OLR) (because clouds emit longwave both upward and downward, but by an amount which is less than the surface-emitted OLR which they absorb, due to temperature differences).

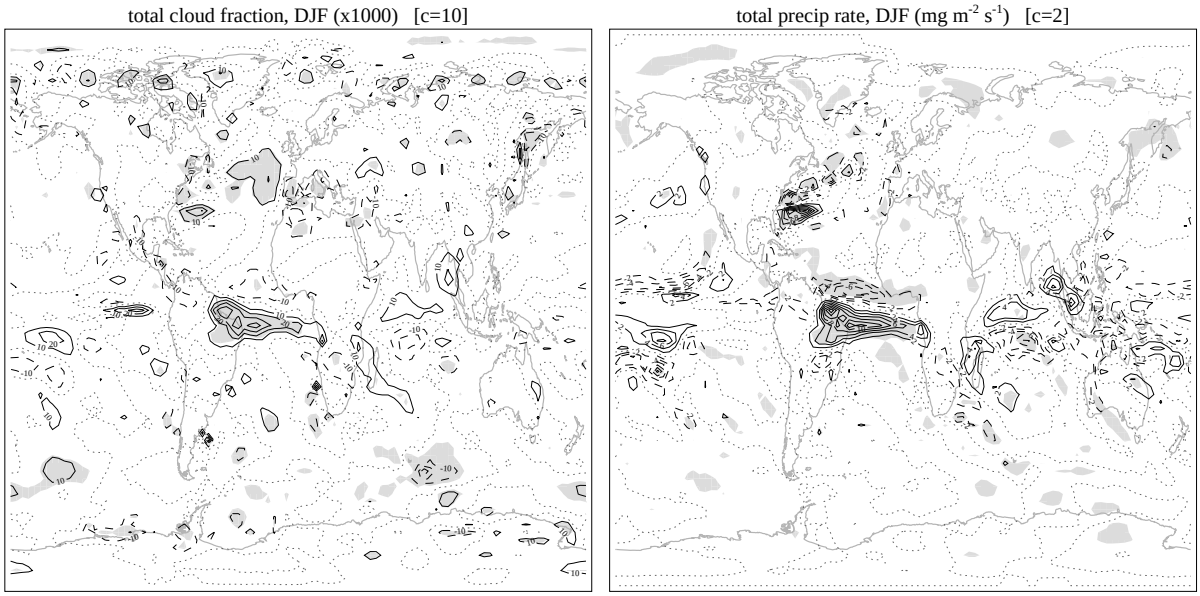


Figure 10: DJF anomalies in geographical distribution of total cloud and precipitation

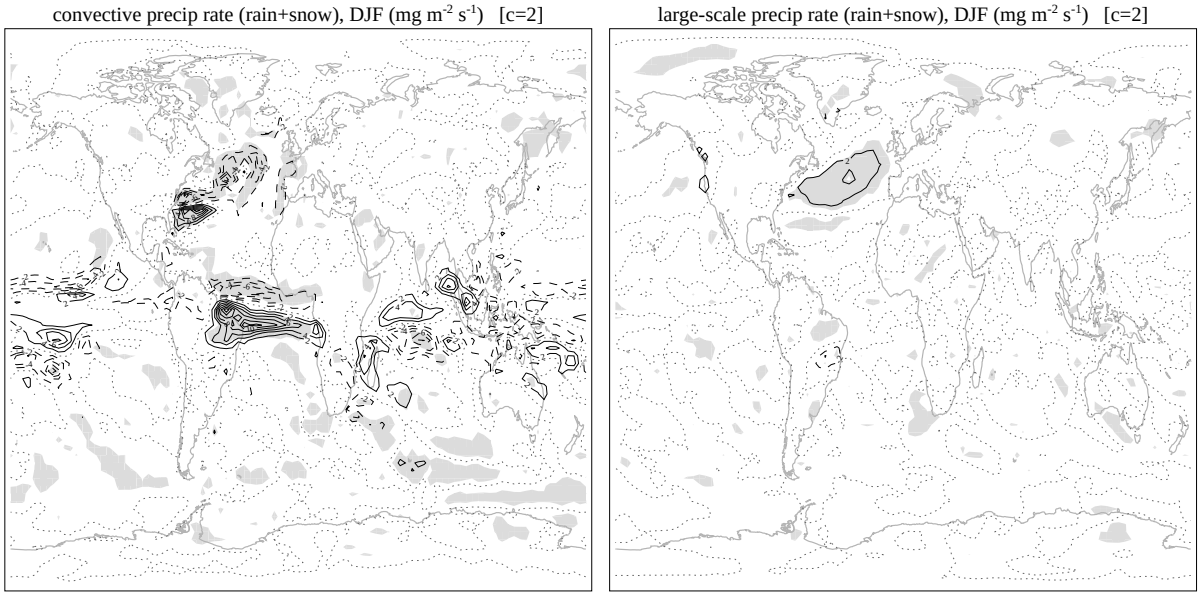


Figure 11: Partition of DJF anomalies in precipitation between convective and large-scale precipitation (rain plus snow).

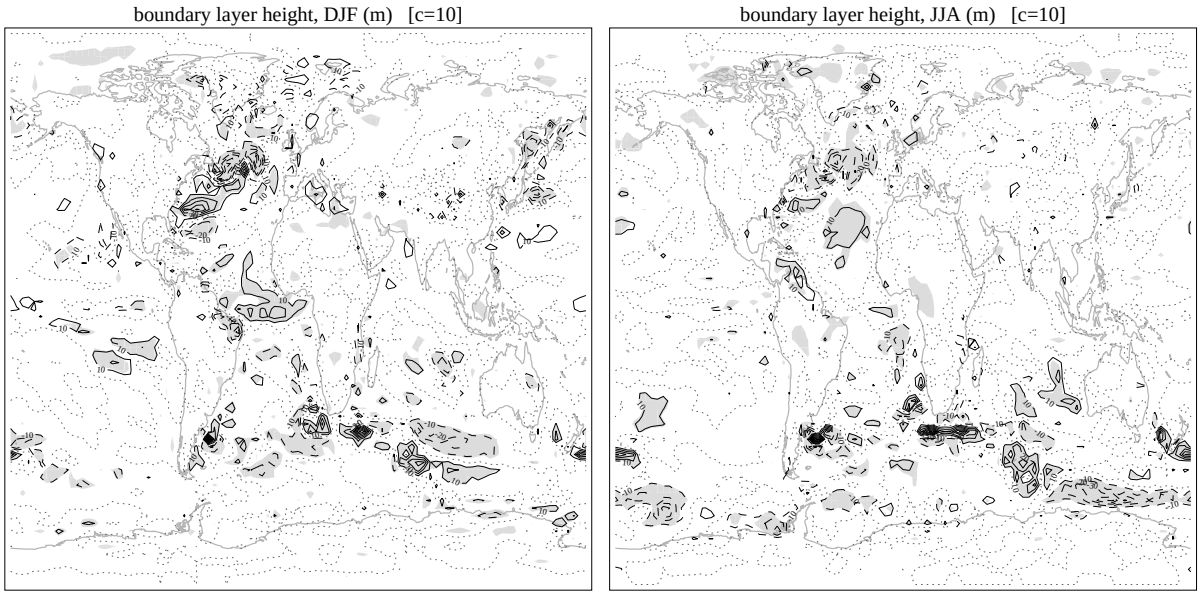


Figure 12: Boundary layer height

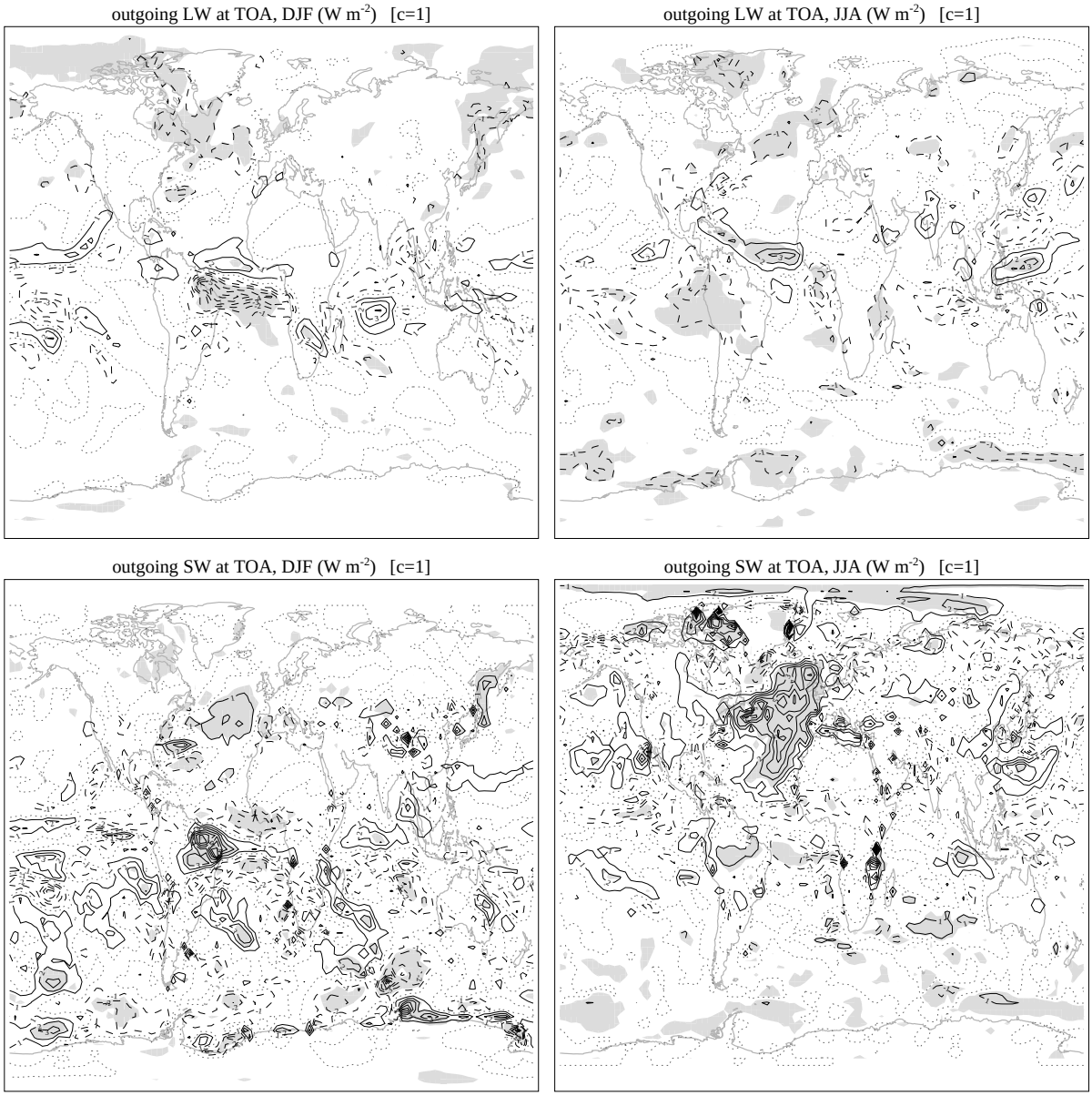


Figure 13: Outgoing longwave and shortwave radiation (top of atmosphere)

## 3.4 Air-sea fluxes

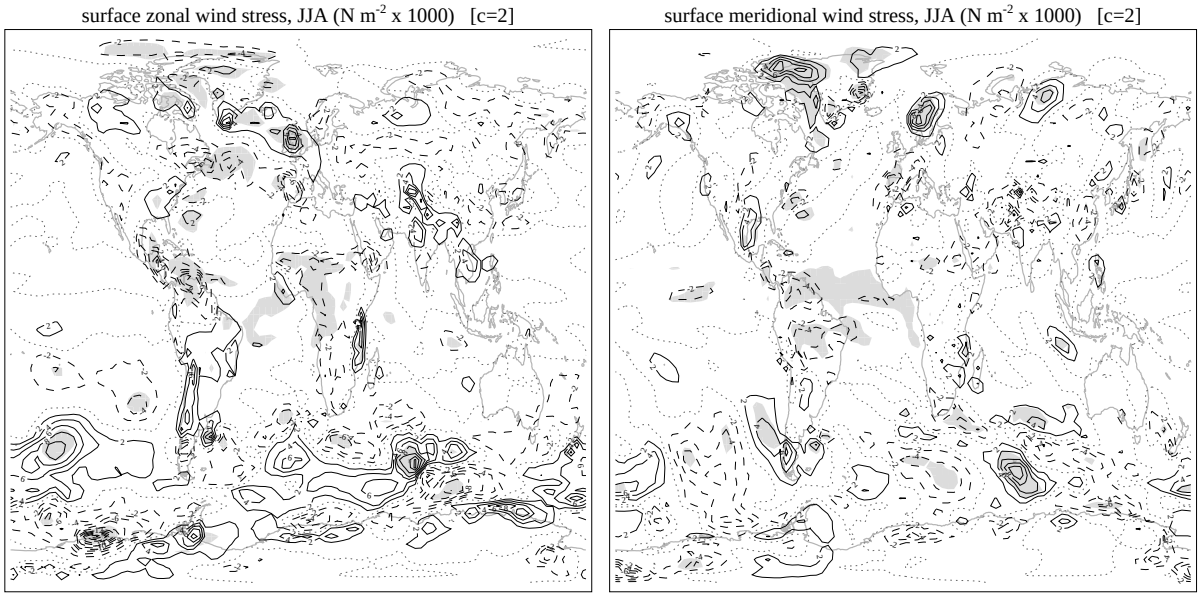


Figure 14: Surface wind stress anomalies for JJA.

Figure 14 shows anomalous wind stresses for JJA. It can be seen that the differences are more significant in the meridional direction than the zonal direction, and are well correlated with the 10m anomalies in meridional wind (figure 7). The sea-air heat flux anomalies (figure 15; positive is upwards for B45 / downwards for C44) are also consistent with the stresses and the surface downward longwave radiation anomalies. The latter is expected since enhanced evaporation due to enhanced downwelling longwave gives a negative latent heat flux anomaly. Although the latent heat flux anomalies are larger over a greater area of the southern ocean the sensible heat flux is also well correlated.

## 3.5 Ocean fields

As already seen in the timeseries in figure 1, the ocean shows small but readily detectable differences between the two integrations, which are evident from the start, and hence due at least in part to the differences in initial conditions. Given the long persistence of any initial anomalies, and the relatively small interannual variability, it would therefore be expected to find that even small anomalies in the deep ocean have high statistical significance.

Figures 16 and 17 show anomalies in ocean potential temperature and salinity between the integrations, at the 5 m (i.e. top) and 447 m levels. Temperature anomalies are mostly small, except for some well-defined regions in the north Atlantic and around 40–50 °S. The ocean temperatures are well correlated between the 447 m level and the top ocean level, as well as with surface air temperature (SAT, figure 6), and are doubtless a key factor in the SAT anomalies. The 5 m ocean temperature field also shows the effect of forcing of the ocean by the atmosphere (due to anomalies in wind stress and radiative fluxes), in that the larger area of non-significance at 5 m compared to 447 m is indicative of atmosphere-induced variability. B45 has generally fresher water than C44, with differences up to 0.5 PSU, mostly in the Atlantic. At depth, differences are smaller ($\sim 0.1$ PSU) but as with potential temperature the statistical significance is high.

Figure 18 shows mixed-layer depth anomalies. There are differences of order tens of metres in winter mixed-layer depth, in the main regions where atmospheric fields show anomalies already discussed (north Atlantic, southern ocean); in the generally shallower mixed layer in summer, any anomalies are also correspondingly smaller.

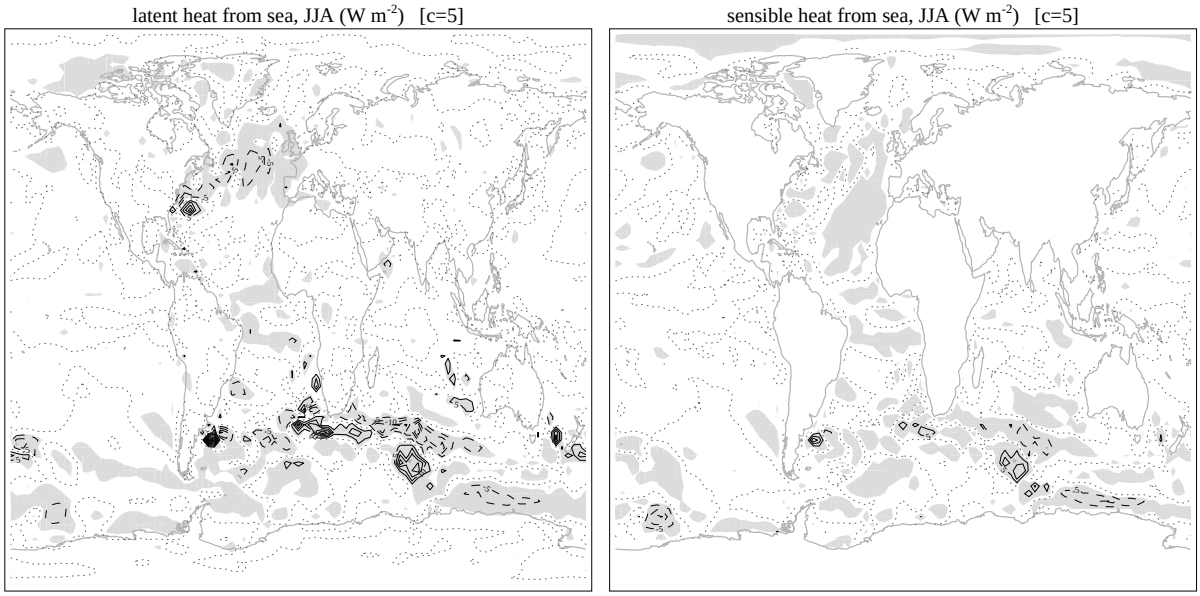


Figure 15: Sea-air heat flux anomalies for JJA.

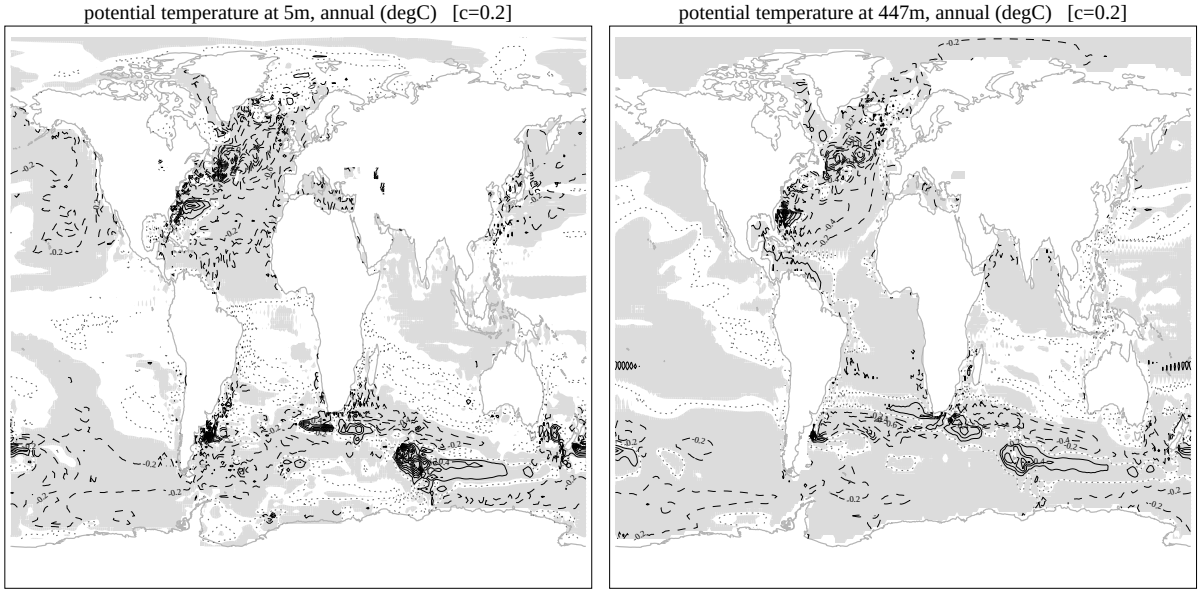


Figure 16: Annual mean ocean potential temperature at 5m (left) and at 447m (right).

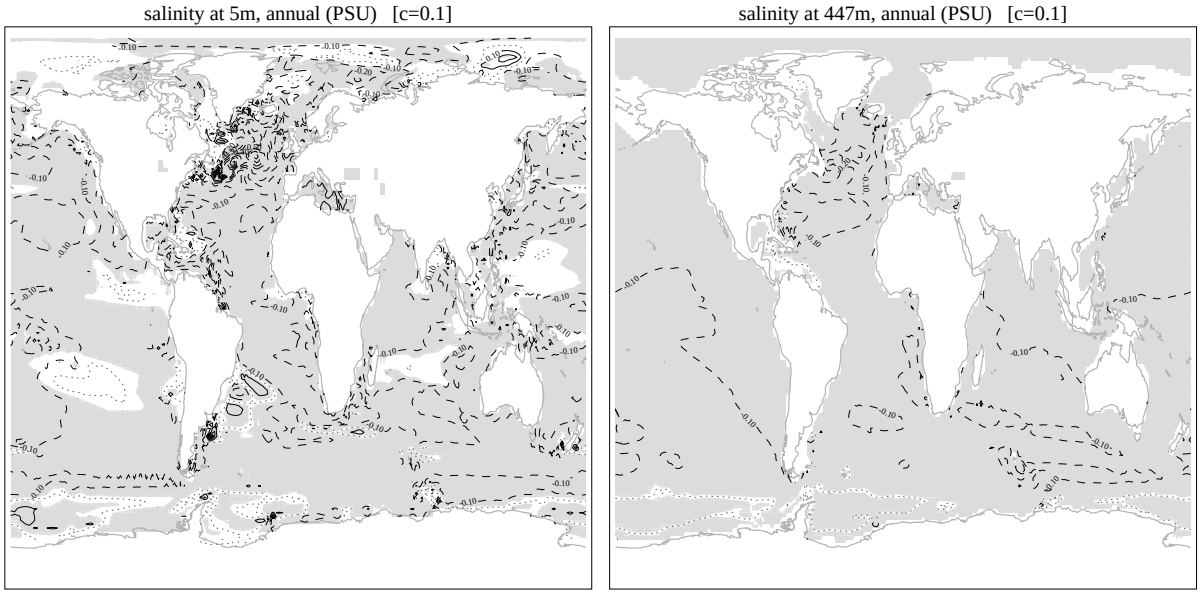


Figure 17: Annual mean salinity at 5m (left) and at 447m (right).

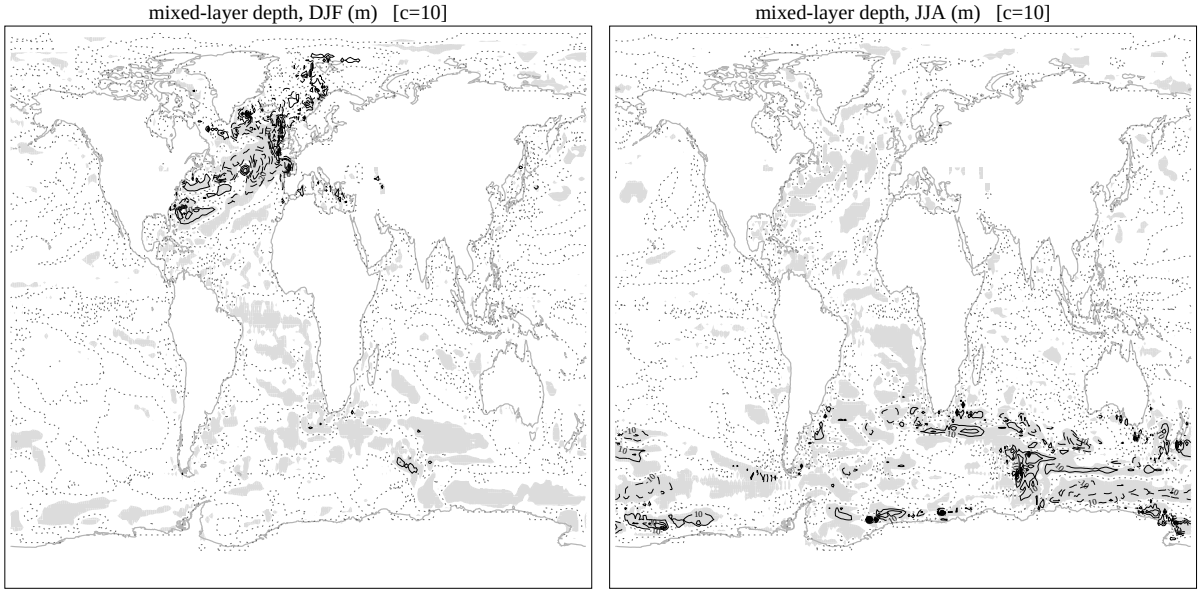


Figure 18: Mixed layer depth

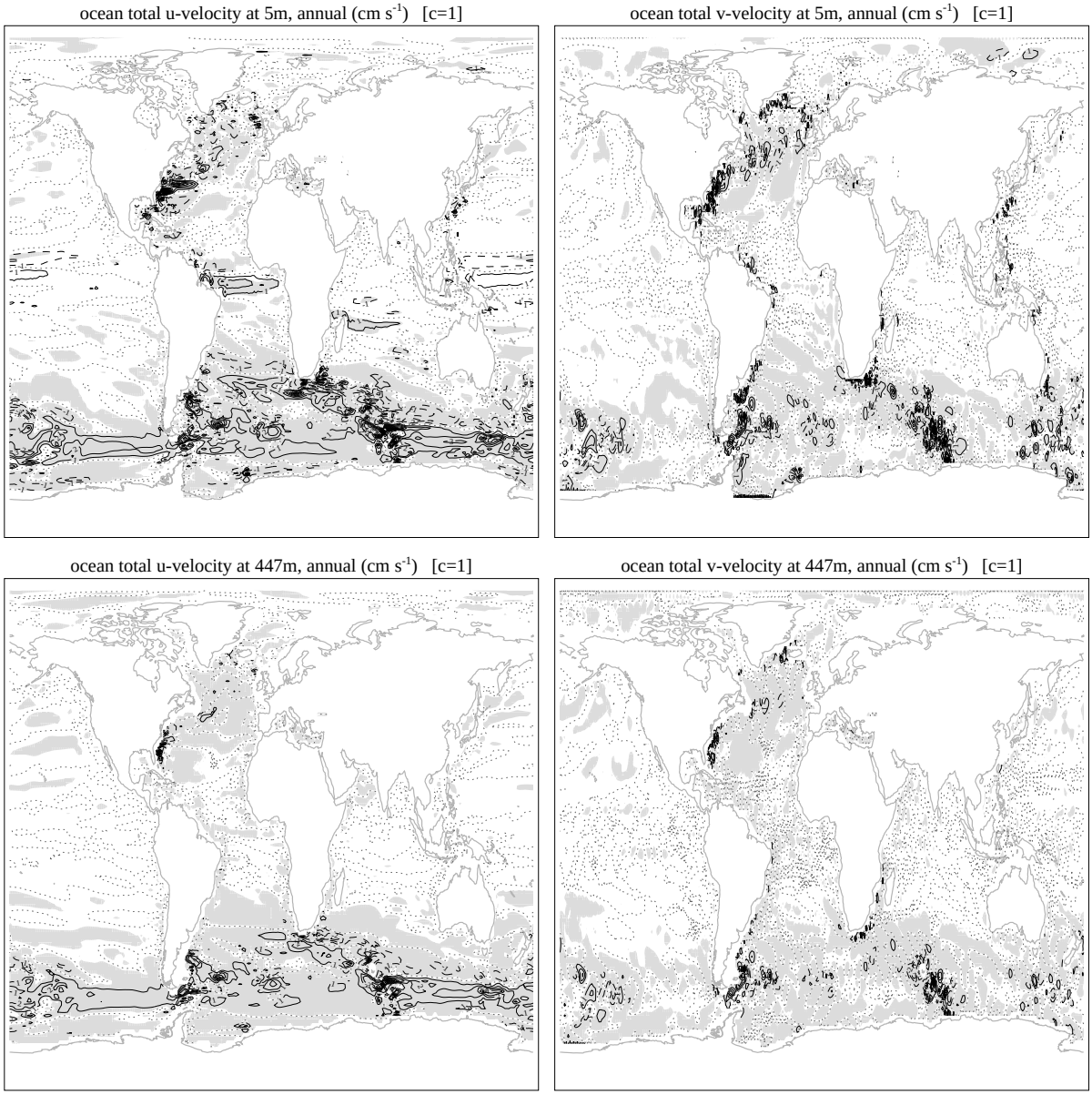


Figure 19: Ocean horizontal velocity (top level and 447 m)

Figure 19 shows horizontal velocity anomalies at the top level and at 447 m. The most notable features are a 1–2 cm s $^{-1}$ eastward anomaly in $u$-velocity in the southern ocean and in the tropical Atlantic.

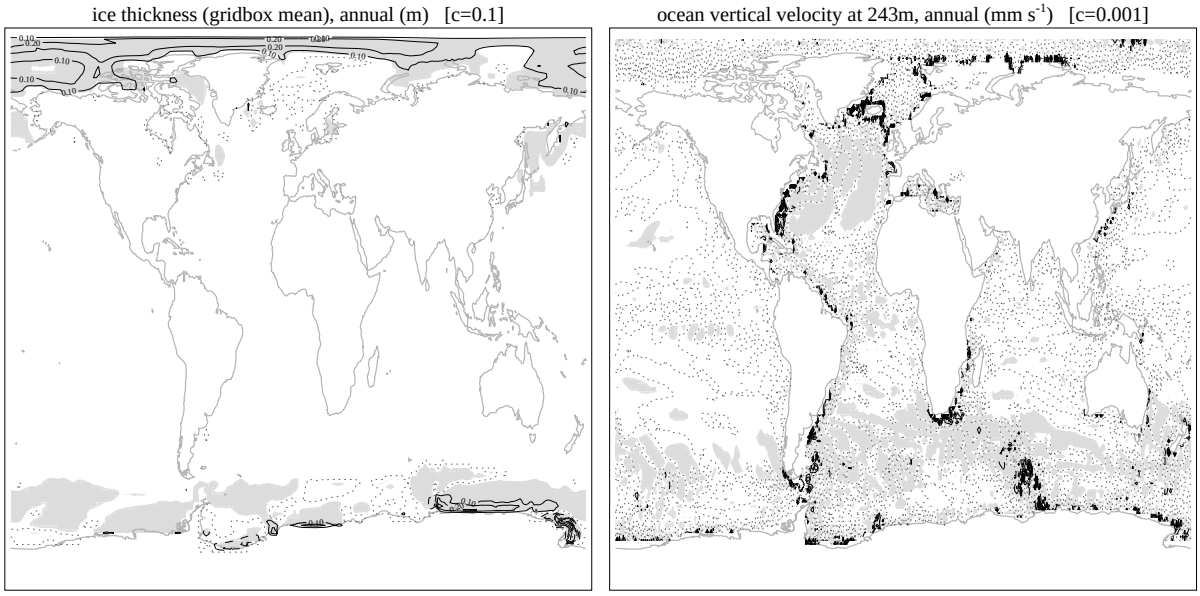


Figure 20: Annual mean differences in ice thickness (left) and vertical velocity at 243m (right).

Finally, although not related, ice depth and oceanic vertical velocity anomalies are displayed together in figure 20 for conciseness. The ice depth in B45 is approximately 0.2m deeper than in C44 in both polar regions throughout the year. The largest differences in oceanic vertical velcity at (243m) are around the continental boundaries (presumably related to small horizontal differences). There appears to be more rising (or less sinking) in the north Atlantic and southern oceans in B45, although the numbers are small.

# 4 Discussion

The detailed comparison in the section 3 outlined a number of differences between the results of integrations B45 and C44. Here it is discussed whether these differences are statistically significant, and if so whether they are important.

## 4.1 Statistical significance of anomalies

Although we have seen that many of the fields of anomalies between the two integrations have areas where the anomaly passes a $t$-statistic significance test at the 5% level, it is difficult to deduce from this that these detected differences are truly statistically significant, as even under the null hypothesis of identical model climate, some proportion of the local significance tests will show differences to be significant at any given probability cutoff level. [3] Furthermore, because of the underlying model physics, any such locally significant differences are likely also to exist in a consistent way across a number of physically related fields, and to show some spatial correlation corresponding to the scale of the simulated features.

In principle, it is possible to establish significance with greater rigour by using a part of the dataset for exploratory comparison, in order to choose carefully a small number (say $n$) of statistical indices for formal hypothesis testing, and use the remainder of the dataset to test the significance of these indices formally (each at probability level $1-(1-p)^{1/n}$, so that the overall probability of a false positive is no greater than $p$).

It is not the intention here to follow this methodology completely. However, as an example, let us consider how it would work in the case of one of the most substantial differences already noted, namely the difference in top-of-atmosphere outgoing shortwave radiative flux (hereafter OSR) in the North Atlantic in summer (JJA), associated with positive cloud cover anomalies (B45 minus C44).

Let us suppose that in integration C44, the last 40 years were used for exploratory anomaly plots, and the first 40 years were then used for formal testing, and that in B45 the last and first 240 years respectively were used. (This gives an unused 20-year gap between the two periods, to help them to be reasonably independent even if there is some decadal variability.)

The OSR anomaly plot in this case is shown in figure 21 (this is as for the OSR shown in figure 13, but as calculated using the sub-timeseries mentioned). In this case, the OSR anomaly pattern still shows an obvious significant anomaly over the North Atlantic. Although this does not itself establish that this feature would reasonably be chosen for formal testing (as anomaly plots for some other fields could appear to show features which are more important, when recalculated using these sub-timeseries), let us nonetheless suppose that a hypothesis test is to be performed on an index which is the average OSR value over the region bounded with a dashed box in figure 21 (unweighted average of 66 grid points).

This box-average JJA OSR has the following values in the formal-testing periods from the two integrations: 154.2 W m$^{-2}$ (with interannual standard deviation 4.5 W m$^{-2}$) in the first 240 years of B45, and 150.1 W m$^{-2}$ (with interannual s.d. 4.4 W m$^{-2}$) in the first 40 years of C44. This gives a 4.1 W m$^{-2}$ anomaly with an associated $t$-statistic probability of $10^{-7}$.

This impressively low probability shows with virtual certainty that there is a difference in JJA OSR between B45 and

[3] It is, in principle, possible to test instead for "field significance" on the overall difference field (or here, the overall set of heterogeneous difference fields) between the two integrations. This works as follows. If $P$ is the proportion of points which pass a local significance test when comparing two timeseries, then the fields will be significantly different at the 95% level if the value of $P$ is beyond the 95% point of the distribution of $P$ which is obtained using many randomly selected pairs of timeseries from a long control dataset.

However, this requires a control dataset which is very much longer than the actual pair of timeseries to be compared, and even if most of the 500-year integration were used as the control, it is very questionable whether this condition would be achieved. (This is because to establish with any accuracy the 95% cutoff value of the distribution of $P$, there must be more than just one or two points above this value; this suggests the need for at least of order 100 truly independent pairs of timeseries, each as long as the timeseries to be compared between integrations.)

In any event, all the field significance would show is whether *some* difference exists between the two integrations (but without indicating that differences in any particular fields or regions are significant). Given the differences of model version, this in itself would not be a particularly informative result.

C44, and is robust even if this statistic were tested as part of a large set of candidate indices.

## 4.2 "Importance" of anomalies

Let us stay with the example of Atlantic OSR. We have shown reasonably conclusively that there is a statistically significant difference between the integrations. However, is an anomaly of $4.1\,\mathrm{W\,m^{-2}}$ enough to be considered "important"? We have already seen that it is of similar magnitude to the interannual variability in the integrations. Here are comparisions with the magnitude of some other variations.

- **Spatial and temporal variability:** OSR has a large annual cycle: in B45, the average OSR over this box drops from $154\,\mathrm{W\,m^{-2}}$ in JJA to $84\,\mathrm{W\,m^{-2}}$ in SON. It also has large spatial variations: see figure 22(a) for full fields (not anomalies) for JJA OSR from B45. These mean that a $4\,\mathrm{W\,m^{-2}}$ difference in OSR is equivalent to a $1^\circ$ offset in latitude or a 5-day offset in time.

- **Comparison with observational data:** These integrations are forced with pre-industrial levels of $CO_2$, which limits the usefulness of comparisons with observations of recent climate. Nonetheless, a comparison with observations is useful for rough comparison of the magnitudes of between-model anomalies and model-data anomalies, even if the causes of model-data anomalies are difficult to separate into the effects of model limitations versus those of different forcings.

  Figure 22(b) shows the JJA average OSR for the period 1985-1989, derived from measurements from the NASA Earth Radiation Budget Satellite (ERBS). Although the patterns are broadly similar to those shown in figure 22(a), there are model-data anomalies of a few tens of $\mathrm{W\,m^{-2}}$ in the Atlantic, so the model-model anomalies are comparatively small.

  For the specific box under consideration, the area-average OSR is $137.1\,\mathrm{W\,m^{-2}}$, (c.f. $154.2\,\mathrm{W\,m^{-2}}$ and $150.1\,\mathrm{W\,m^{-2}}$ for the two models as above). The associated inter-annual standard deviation in the observations is $4.5\,\mathrm{W\,m^{-2}}$. Hence in this one particular variable the model appears to estimate variability correctly, even though models are in general known to underestimate variability; that said, the observed timeseries is a small sample, and too short a timeseries to contain decadal variability.)

Overall, the inter-model anomalies, although statistically significant, appear fairly small compared to other variability. Additionally, even though only one variable has been discussed in detail here, for the other fields shown in this report the magnitudes of the anomalies have also been similarly compared with the spatial variability in the full fields, and although the full fields are not shown here, it can be reported that the differences due to model-model anomalies are generally small compared to spatial variations.

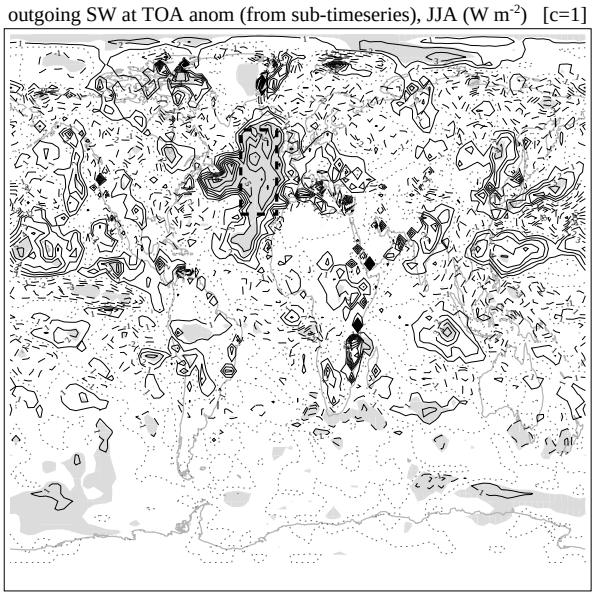


Figure 21: Top of atmosphere outgoing shortwave for JJA, as in figure 13, but calculated from subset timeseries, see text

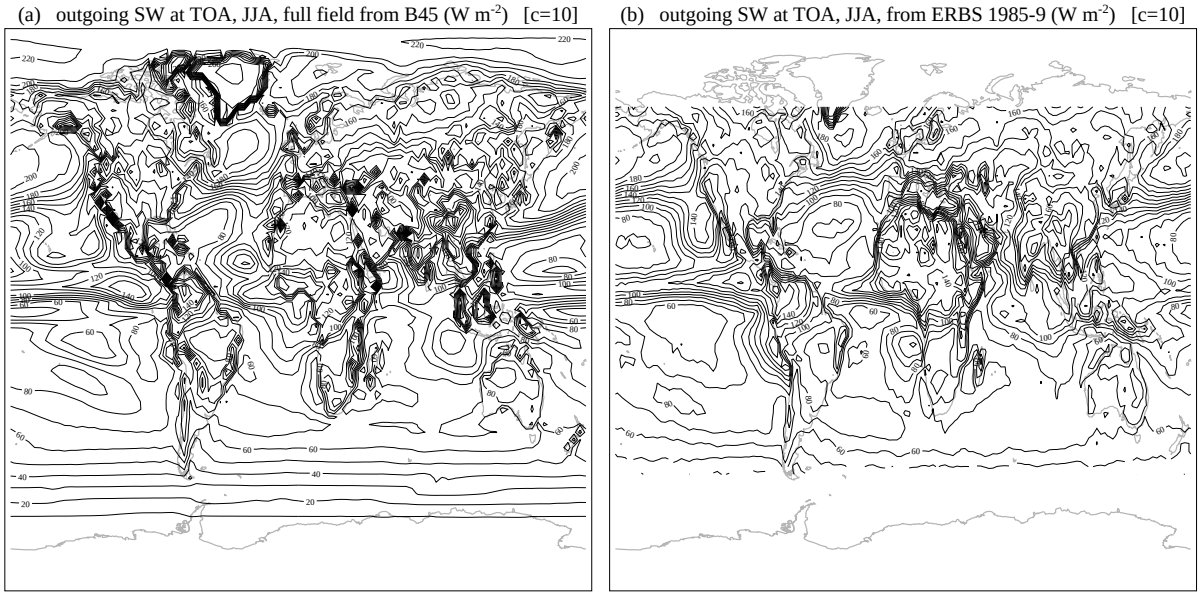


Figure 22: Top of atmosphere outgoing shortwave for JJA, full fields (not anomalies): (a) from run B45, and (b) from ERBS observations

# 5 Conclusions

Results from two integrations of HadCM3 have been compared: The first (C44) was a 100-year run performed on a Cray T3E using Unified Model version 4.4. The second (B45) was a 500-year run performed on a PC Beowulf Cluster using UM version 4.5. Both assumed pre-industrial $CO_2$ concentrations, and were run at 64-bit precision.

A detailed comparison of model output from the two integrations has been undertaken. This has shown a few small but discernible differences between the two integrations, most notably slight differences in geopotential and surface pressure, stratospheric water vapour, and in a set of physically related atmospheric and oceanic fields in the Atlantic and Southern Ocean (cloud cover, radiation, precipitation, soil moisture, air-sea fluxes, ocean temperature and ocean zonal velocity).

Although the statistical methods used have largely consisted of ad-hoc inspection of local significance levels, from a more careful case-study it appears that at least some of these differences are genuinely statistically significant.

The fact that statistically significant differences exist between the B45 and C44 model integrations exist underlines the fact that when performing a controlled experiment, it is always highly inadvisable to "mix and match" control and perturbation runs from two different hardware platforms or model versions. (Here there is no separation of the effects of model version versus hardware platform, although a separate comparison of the same model on Beowulf and Cray also showed differences to be small.)

However, the magnitude of the differences is small in comparison with typical levels or variability in the relevant fields (in the case of the pressure and geopotential anomalies, this is particularly evident when considering spatial derivatives which are more dynamically important than their absolute values), or they relate to stratospheric water values which are known to be unphysical anyway.

Provided that the models are used appropriately (i.e. conducting control and perturbation experiments on the same platform), the evidence here strongly suggests that both combinations of model version and hardware platform examined here are valid for the uses expected from climate models. Any differences between them are relatively minor, and given the simulations are in any event for pre-industrial conditions, it is not possible here to use observational data as an arbiter of these minor differences.