



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Can Top-of-Atmosphere Radiation Measurements Constrain Climate Predictions? Part I: Tuning

Citation for published version:

Tett, SFB, Mineter, MJ, Cartis, C, Rowlands, DJ & Liu, P 2013, 'Can Top-of-Atmosphere Radiation Measurements Constrain Climate Predictions? Part I: Tuning' *Journal of Climate*, vol 26, no. 23, pp. 9348-9366. DOI: 10.1175/JCLI-D-12-00595.1

Digital Object Identifier (DOI):

[10.1175/JCLI-D-12-00595.1](https://doi.org/10.1175/JCLI-D-12-00595.1)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Journal of Climate

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Can Top Of Atmosphere Radiation Measurements Constrain Climate Predictions? Part 1: Tuning.

SIMON F. B. TETT *

MICHAEL J. MINETER

School of GeoSciences, University of Edinburgh

CORALIA CARTIS

School of Mathematics, University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK

DANIEL J. ROWLANDS

Atmospheric, Oceanic & Planetary Physics, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

PING LIU

School of Mathematics, University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK

ABSTRACT

Perturbed physics configurations of the HadAM3 atmospheric model driven with observed sea surface temperatures (SST) and sea ice were tuned to outgoing radiation observations using a Gauss-Newton line-search optimisation algorithm to adjust the model parameters. Four key parameters, that previous research found affected climate sensitivity, were adjusted to several different target values including two sets of observations. The observations used were the global average Reflected Shortwave Radiation (RSR) and Outgoing Longwave Radiation (OLR) from the Clouds and Earth's Radiant Energy System instruments combined with observations of ocean heat content. Using the same method, configurations were also generated that were consistent with the earlier Earth Radiation Budget Experiment results. Many, though not all, tuning experiments were successful with about 2500 configurations being generated and the changes in simulated outgoing radiation largely due to changes in clouds. Clear sky radiation changes were small largely due to a cancellation between changes in upper-tropospheric relative humidity and temperature. Changes in other climate variables are strongly related to changes in OLR and RSR particularly on large scales. There appears to be some equifinality with different parameter configurations producing OLR and RSR values close to observed values. These models have small differences in their climatology with the one group being similar to the standard configuration and the other group drier in the tropics and warmer everywhere.

1. Introduction

There have been several attempts to constrain climate sensitivity using observations. Some of these compare diagnostics from an ensemble of multiple model simulations (e.g. Knutti and Hegerl (2008)) while others systematically vary parameters in a single model (e.g. Stainforth et al. (2005)). Varying parameters in a single model typically represents the ocean as a “slab” which, though computationally less expensive than using a full ocean model, requires simulations of a few decades in length both to estimate the model climatology, for comparison with observations, and to estimate the climate sensitivity. Randomly perturbing parameters generates model configurations that have large differences from observations. Such configurations will have low likelihoods and so contribute little

to observationally constrained distributions. An approach which preferentially generates “plausible” model configurations would reduce the computational cost of generating an ensemble of models to explore future uncertainty in climate change. The computational cost could be reduced further if atmospheric model simulations of a few years could be used.

In part 1 of this paper we use an optimisation method to show that it is possible to “tune” an atmospheric model to different outgoing radiation targets through modifying parameters in the models parametrisation schemes. Unlike others who have considered several observational datasets we use several different targets but only two observations: the global-mean outgoing longwave radiation (OLR) and reflected shortwave radiation (RSR). An important reason to focus on global-mean outgoing radiation is that an at-

mosphere/ocean climate model with a poor simulation of net radiation will drift until the model is back in equilibrium leading to significant surface temperature errors. Such models with large surface temperature errors will have distorted climate feedbacks. We chose targets that are close to two different observational values *and* sampled around the edge of the “plausible” region of model-observational agreement whose construction we describe in part 2 of this paper. We also explore the physical mechanisms by which the model makes these changes and if different climatologies are possible for similar outgoing radiation values. Our results suggest that automatic tuning of an atmospheric model is possible.

In part 2 we use the data from the simulations we generated to first investigate the relationship between climate feedbacks and outgoing radiation in HadAM3 finding there is a strong relationship between outgoing radiation and climate sensitivity for the parameters we varied. Using these simulations and an uncertainty estimate for model-observational difference we make an observationally constrained estimate of climate sensitivity based on the structure of HadAM3. We find that that the range of plausible climate sensitivities is considerably narrower than the prior range when using the CERES record of Loeb et al. (2009). These results suggest that the CERES data provides a significant observational constraint on climate sensitivity.

Turning to the question of “tuning” which is the focus of part 1. Current techniques to “tune” climate models are slow, expensive and require expert judgement (e.g. Mauritsen et al. (2012)). Several authors have attempted to automatically “tune” climate models in a variety of ways. Gregoire et al. (2011) carried out 100 evaluations of the FAMOUS AOGCM and selected those configurations that were in objective good agreement with present observations and Last Glacial Maximum reconstructions. Medvigy et al. (2010) sampled uniformly across a plausible range of two parameters for the Ocean-Land-Atmosphere Model (Walko and Avissar 2008) and found the configurations that maximised the likelihood of multiple weighted observations. These approaches which require sampling uniformly across parameter space are fairly inefficient particularly as the dimensionality of the problem increases.

To estimate the multivariate probability distribution of parameters, Jackson et al. (2004) proposed using instead of pure MCMC sampling, an approximate Bayesian stochastic inversion algorithm based on multiple very fast simulated annealing that ‘focuses’ the sampling so as to also decrease a cost function of the error between model output and observations; this algorithm is one to two orders of magnitude more efficient in terms of evaluations count than classical samplers (e.g. Järvinen et al. (2010)). Though possibly not yielding equally good sampling of the parameter space, optimisation methods that locally minimise model-observational difference have the potential to be even more

efficient in terms of model evaluations required.

Jones et al. (2005) used an optimisation algorithm based on coordinate search directions, to tune multiple characteristics of the low resolution FAMOUS coupled atmosphere/ocean climate model to the climatology of its higher resolution HadCM3 parent (Gordon et al. 2000). To avoid overfitting they used a skill score that summed over several climate variables. Their algorithm recursively updates one parameter at a time, by maximising a quadratic approximation of the skill score determined by the current best guess of the solution and two perturbations along the same coordinate that span the allowed range of the parameter.

Severijns and Hazeleger (2005) employed the — “often unreliable” Wright (1995) — Nelder-Mead simplex to tune a fast climate model to observations. Like Jones et al. (2005) they maximised a skill score based on several climate variabilities. Neelin et al. (2010) defined a weighted error function as a weighted sum of root-mean-square errors for model-observation differences. Their innovative approach was to emulate the skill score for different observational datasets. This then allows post-optimisation selection of minima based on which datasets were considered most reliable or important.

Schirber et al. (2013) used ensemble Kalman filter data assimilation to simultaneously estimate six model parameters and atmospheric states on timescales of a few days. However, they found the their model performed worse on climatological time scales. Ollinaho et al. (2012) also used an ensemble forecast technique but with each forecast using different parameter values. After reweighting they estimated a best estimate parameter combination and covariances.

In contrast to Jones et al. (2005); Neelin et al. (2010); Gregoire et al. (2011); Severijns and Hazeleger (2005) who include multiple variables in their error function we focus on the global average of the top-of-atmosphere reflected shortwave and outgoing longwave radiation. We do this partly for simplicity and transparency, and partly because estimating the covariance structure over multiple climate variables is likely to be difficult and very subjective. Our principal focus is on the Clouds and Earth’s Radiant Energy System (CERES) estimates (Loeb et al. 2009) adjusted for the small net-flux imbalance found in ocean observations. We also consider optimising to the older ERBE results (Fasullo and Trenberth 2008).

Our experimental design uses many six-year atmospheric only simulations of the HadAM3 model (Pope et al. 2000) driven by observed Sea Surface Temperature (SST) and sea-ice as well as a package of climate forcings. We vary four parameters that have been found to have most impact on climate sensitivity (Knight et al. 2007). Though this model is no longer state-of-the-art it is still in widespread use and a good test case. In part 1 of the paper we aim to explore if:

- i. simple optimisation methods can be used to tune the atmospheric model component of a climate model against radiation observations.
- ii. the climatologies of the models that result are related to the outgoing global-mean fluxes.
- iii. how sensitive the climatologies obtained are to different observed estimates from Fasullo and Trenberth (2008) and Loeb et al. (2009).

The rest of the paper is structured as follows. In the following section we describe our experimental design, then describe how we optimised the model so that it produced specified outgoing longwave radiation (OLR) and reflected shortwave radiation (RSR) showing that we could, in most cases regardless of starting parameter values, adjust model parameters to observations. We then present results from our optimisation experiments before showing that other diagnostics are strongly correlated with global-mean RSR and OLR. We explore issues of equifinality (Beven and Freer 2001) showing that model configurations with similar OLR/RSR values can have a wide variety of parameter values and somewhat different surface climatologies before concluding.

2. Methods

a. Experimental Design

Our experimental design is a “perturbed physics” ensemble of simulations of HadAM3 (Pope et al. 2000) driven with the HadISST SST & Sea-Ice analysis (Rayner et al. 2003) and a updated package of natural and human forcings (Tett et al. 2007). We use the standard N48 resolution ($3.75^\circ \times 2.5^\circ$) of HadAM3.

The anthropogenic forcings include well mixed greenhouse gases, tropospheric and stratospheric ozone, the direct effect of aerosols, and a relatively crude parameterisation of the indirect effect of aerosols. The differences from Tett et al. (2007) are the use of greenhouse gases that are close to observed values to 2010, the use of Total Solar Irradiance (TSI) values from recent observations (Kopp and Lean 2011) and a fix to an error in the shortwave atmospheric (Rayleigh) scattering in HadAM3. The effect of the reduction in TSI is to reduce RSR by 0.3 Wm^{-2} while fixing the Rayleigh scattering error increases RSR by 1.3 Wm^{-2} .

Our simulations all started 11th December 1998 and ran to 1st March 2005. Simulated values of 5-year average OLR and RSR, for the period March 2000 to February 2005 are compared with the observed CERES values of Loeb et al. (2009) for the same period.

Previous work (Knight et al. 2007; Murphy et al. 2004; Sanderson et al. 2008) has found that four parameters, which affect the vertical mixing of water and cloud formation, largely control climate sensitivity in HadAM3. The

range of acceptable parameters values for these parameters was reported by Murphy et al. (2004) based on expert judgement and we use this range (see Table 1). The four parameters we focus on are:

- entcoef** Parameter in the model’s convection scheme that controls the rate at which bulk air is entrained into the convective plumes.
- vf1** A parameter in the model’s large scale precipitation scheme which is the speed of falling ice.
- ct** Another parameter in the model’s large scale precipitation scheme which is the rate at which cloud droplets turn into rain.
- rhcrit** A parameter in the model’s large scale cloud scheme which gives the critical value, at each model level, of relative humidity for clouds to form in a grid box. We specified a single value for **rhcrit** and assigned it to each of the 19 model levels as follows: Levels one and two always set **rhcrit** to 0.95 & 0.9 respectively. Level three had a value of two thirds the specified value + 0.3 but with a minimum value of 0.85. Levels 4-19 used the specified value.

The parameters in the large-scale precipitation scheme affect cloud as they are one mechanism by which cloud ice and liquid water are removed but through evaporation of precipitation they also moisten layers below clouds.

We also carried out an ensemble of 19 simulations of the standard configuration of HadAM3 started with different initial conditions and use this to assess the impact of internal climate variability on our results. In doing this we are assuming that parameter changes have no significant impact on internal variability. Though unlikely to be strictly true as internal variability is small (see below) our neglect of its changes is not likely to be important.

b. Optimisation techniques

In this section we describe the optimisation algorithm we use with the aim of finding parameter combinations that reduce model-observational difference, where the latter is measured by the root-mean-square error of the difference between the observations and simulated OLR and RSR. The essential challenges that this (non-linear non-convex) optimisation problem poses and that the algorithm must face are: the *computational expensiveness* of each function evaluation (here, each function evaluation requires a 5 year simulation of HadAM3); *noise* present in the function values (estimated to be of order 0.1 Wm^{-2} , due to the chaotic nature of the climate model) and the *unavailability of derivatives* of the error function (no adjoint exists) that we aim to minimise.

State-of-the-art optimisation approaches that are designed to tackle such challenges are known as *derivative free*

optimisation algorithms (Conn and Vicente (2008); Powell (1998)), and significant research continues into such algorithms, especially in the last decade. Deciding which variant within this class of methods is best for a given problem is not straightforward, but some general purpose software packages are available (such as DFO, BOBYQA, BFO, etc.) for testing. However, in our case, due to the need to integrate the optimisation approach with HadAM3, it was crucial to have easy access to the source code of the optimisation algorithm, control over its parameters and possibility to do changes so as to improve performance, which constrained some of our choices. We experimented with BOBYQA and derivative-free gradient methods (Liu 2010), but found that what was most efficient was an *implicit filtering*-type approach (Kelley 2011) based on the classical Gauss-Newton (Nocedal and Wright 2006) method, that we implemented and customised appropriately.

The key ingredients of this iterative technique involve computing approximate derivatives of the model; then, based on these, choosing a regularised direction of change of our current best estimate of a solution that takes into account parameter constraints and deciding how far along this direction to go to ensure the error function is decreased; finally, deciding what to do when no progress can be made and relatedly, when to terminate the algorithm. Next, we describe each of these aspects in greater detail.

Firstly, for the “base”/current set of parameter values, we estimate the model derivatives (i.e., Jacobian) using forward finite differences by increasing each parameter value by a fixed amount (Table 1). The magnitude of the perturbation was approximately 5-10% of the parameter range and chosen to be large enough to avoid making derivative evaluations sufficiently close together to generate strongly noise-contained estimates of the derivatives. In our early attempts we modified the parameter values by 1% of their range and found that convergence was difficult.

We perform an initial scaling of the parameters that brings their values within similar magnitudes, so as to avoid algorithmic inefficiencies due to ill-conditioning. As our inverse problem is ill-posed, the Jacobian is singular and so it requires some regularisation to make the problem well-posed and allow us to compute an iterative improvement. We used a Tikhonov-type regularisation (Nocedal and Wright 2006) of the Jacobian, namely a perturbation of the Jacobian by a small multiple of the identity matrix, to ensure the resulting matrix is full rank. The effect of this is to prefer solutions in which the four parameters, after the initial scaling, all have similar magnitudes. For example preferring solutions of the form (1, 1, 1, 1) to (10, 0.1, 1, 1).

To compute an improvement to the current best guess of optimal parameters values, we form a Gauss-Newton direction of search using the regularised approximate derivatives we have computed. To keep all parameters within their

expert-defined bounds (Murphy et al. 2004) (Table 1), we project the search vector along the boundary of allowed parameters values whenever such a boundary is encountered by our search. The projection keeps the points, that are within the bounds, unchanged and inexpensively computes the closest point, within bounds, to any points outside the allowed bounds.

Finally, to ensure the error is decreased, we carry out a linesearch along this vector by evaluating the error at the full feasible vector, at 90% and at 30% of the feasible vector. Then by comparing the errors at these points with the current best value of the error, we update our current set of parameter values to the values, if any, that gave most decrease in the error.

We terminate the optimisation algorithm when the minimum linesearch root-mean-square error is less than 1 Wm^{-2} (this error level is meant to keep the runs well above noise levels) or when an iteration yields no or insignificant decrease in the error over previous iterations. If an iteration yields no decrease in the error, it is customary in implicit-filtering techniques to shrink the perturbations used to compute the model derivatives (while still keeping their size above noise levels). This is meant to give a more “accurate” approximation of model derivatives; indeed, if we had exact derivatives, a sufficiently small move along the Gauss-Newton direction would be guaranteed to decrease the error locally. Until now, we have only marginally experimented with this additional heuristic, see Section 3.

The use of finite differences implies that $n+1$ function evaluations are required for a problem in n variables plus an additional 3 for the linesearch giving a total of $n+4$ evaluations per each iteration of the algorithm. This cost of $n+4$ evaluations is computationally feasible for problems, like ours, where the dimensionality is small.

3. Optimisation Results

In this section we focus on results from our optimisation calculations. We first report on results which targeted the observed CERES values (Loeb et al. 2009) and the ERBE estimates (Fasullo and Trenberth 2008) starting from two different parameter combinations. We then report on results in which we initialised the optimisation algorithm with all 16 combinations of extreme parameter choices and optimised to six different target values.

We found parameter combinations using an emulator (see part 2), for the four parameters being considered, that gave a model configuration with the smallest (Low) or largest (High) climate sensitivities and started our optimisation algorithm from them (Table 2). We first optimised both these initial parameter choices to the CERES values and then to the ERBE values (Fig. 1). The high sensitivity case did not converge to the CERES target though the remaining three cases converged. For the case when con-

vergence failed, we tried further halving and quartering the perturbations used to compute the Jacobian, following an implicit-filtering strategy described earlier. No significant improvement was found suggesting we have reached a local error-minimum.

We then carried out optimisation studies where we started the algorithm with the extreme values for the four variables, and targeted the CERES observations, giving 16 cases in all. Twelve of these cases failed not because of a failure to converge but because the model failed or “blew up” indicating unrealistic model physics. Examining the annual mean of OLR and RSR for 1999 (Fig. 2) we see mainly, though not exclusively, that failure occurs for model configurations with OLR and RSR values very different from observations or the standard configuration of HadAM3. The four configurations that did not fail converged to the CERES observations with between 3 and 6 iterations.

We then repeated our attempt with parameters at 75% of the difference between the default and limiting values (Table 1). No failures occurred and 15 cases converged to the CERES observed values taking from one to eight iterations to do so (Table 3). These final configurations have a broad range of parameter values (see below). We then repeated this for the ERBE target, and due to a setup error, a target value near the ERBE values. Twelve cases successfully converged to the ERBE target while 13 successfully converged to the Near ERBE target. For the cases that failed six, in total, had a model failure while one, in total, failed to converge.

The ERBE value lies at the edge of the plausible region around the CERES values (see part 2 for details). We then carried out further experiments, targeting three values spread approximately equally around the edge of the uncertainty region (Fig. 1). For one target we only had 4 successful optimisation cases with 11 failing to converge, and one model failure. This target appears to be difficult to generate model configurations that match with the optimisation algorithm, likely getting trapped in local minima though further experimentation is required to confirm this. The other two target values appeared to behave similarly to the ERBE targets with a total of four model failures and two cases, in total, when the algorithm failed to converge.

To provide a statistical summary of our optimisation results we focus our analysis on five targets all started with initial parameter values at 75% of the difference between the default and limiting values (* in Table 3) giving 60 “sample” cases in all. About 30% of these failed with the remainder converging. The failure cases took, on average, about 28 model evaluations to either “blow up” or terminate due to failure to reduce the error. The convergent cases took, on average, about 24 evaluations. One algorithm to find an optimal solution would be to start with some initial parameter choice and apply our algorithm. If the optimisation fails to converge or the model “blows up”

then generate another configuration and keep trying until convergence happens. The number of model evaluations required, taking account of the possibility of repeated failure, is:

$$E = E_c + \sum_{i=1}^{\infty} E_f f^i \quad (1)$$

where E is the total number of evaluations, $E_c(E_f)$ the number of evaluations for convergence (failure) and f the probability of failure. This gives $E = E_c + E_f \frac{f}{1-f}$. For our values this means that such an algorithm would take, on average, about 36 model-evaluations, each of 6.25 years, requiring 225 simulated years.

4. Changes in Other Climate Variables

So far we have focused our attention on global-mean Reflected Shortwave Radiation (RSR) and Outgoing Longwave Radiation (OLR). We did this as it allows a relatively simple and reproducible tuning strategy allowing us to generate model configurations that agree well with the global mean outgoing longwave and reflected shortwave. However, in so doing, we might produce configurations that have poor climatologies for other variables including OLR and RSR themselves. In this section we examine climatologies in land temperature and precipitation, and several other climate variables to see what relationship there is between OLR and RSR and those variables. If there are strong relationships between those variables and OLR and RSR then that supports our focus on OLR/RSR while if there is little relationship then we should include them in our minimisation algorithm to produce model configurations that are in better agreement with observations.

Some of our model configurations produce OLR or RSR values that are tens of watts different from the observations. In order to avoid our analysis being overdetermined by these cases we only consider model configurations that have reflected SW and outgoing longwave values *both* within 10 Wm^{-2} of the observed values. As we show in part 2 of this paper that is well outside the observed uncertainty range. This gives us 1849 HadAM3 configurations, from a total of 2640 configurations, to analyse in this section.

a. Clear or Cloudy effects?

We first consider those variables that are closely related to RSR and OLR with the aim of determining if the changes in RSR and OLR arise from clear sky effects or changes in cloud. Examining first the fractional cloud (Fig. 3(a)) we see systematic changes in fractional cloud area with values ranging from 0.38 to 0.57. The systematic variation is a reduction in cloud fraction as OLR increases and an increase in cloud fraction as reflected shortwave radiation increases. This is what we would expect as reducing cloud amount increases OLR (more radiation from

lower in the atmosphere) and reduces RSR (less reflected radiation) (see below). The range of clear sky RSR values are small ranging from 52.6 to 53.4 Wm^{-2} suggesting our parameter changes are not greatly affecting the surface albedo even though the land temperatures (see below) change somewhat.

The range for clear-sky OLR (Fig. 3(b)) at 258.4 to 262.1 Wm^{-2} is larger than the clear-sky RSR range but much smaller than the $\pm 10 \text{ Wm}^{-2}$ range in all-sky OLR and RSR. The simulated clear sky OLR of around 260 Wm^{-2} is also much smaller than the Loeb et al. (2009) value of 269.1 Wm^{-2} . Simulated clear sky OLR is computed by running the radiation scheme with clouds removed but no other changes made to the model state. This is different from the observed values which sample those few pixels where there are no clouds. This difference can lead to quite large biases (John et al. 2011). So the reason for model-data discrepancy in clear sky OLR may be due to this bias or reflect some systematic problem with HadAM3 that no amount of model adjustment can rectify.

Changes in upper tropospheric water vapour and temperature may explain the changes in clear sky OLR. We find (Fig. 3(c) and (d)) that 250 hPa temperatures are greatest when 250 hPa relative humidity (RH) is also largest. That suggests a compensation between these two variables such as to keep the clear sky OLR near constant. When the atmosphere is warmer and so emits more outgoing radiation it is also wetter. The moister upper troposphere means the radiation that reaches space is coming from higher in the atmosphere where it is colder. This compensation between lapse rate and moisture changes has been found for many models in response to climate change (Randall et al. 2007).

We now investigate, in more detail, the small changes in clear sky radiation and the dominance of cloud changes on the outgoing radiation. One possibility is that our results are an artifact of the four parameters we tuned or because we are using results from “tuning” experiments. Rowlands et al. (2012) used 14,001 perturbed physics simulations of HadCM3 perturbing the same 10 parameters as Sander et al. (2008). We randomly sampled 100 parameter-configurations from these 14,001 configurations and carried out atmospheric model simulations using these parameter values (“randomly sampled” simulations). Two of the simulations failed through model “blow up” leaving 98. For the clear sky RSR ten parameters gives a larger scatter than just modifying four. However, the range of clear sky RSR values is still small at less than 3 Wm^{-2} compared to range of 100 Wm^{-2} in all sky RSR (Fig. 4(a)). For clear sky OLR the range is about 5 Wm^{-2} which is small compared to the range of OLR values in both the “tuned” and “randomly sampled” simulations (Fig. 4(b)). The “tuned” and “randomly sampled” simulations appear to behave similarly with RSR increasing as cloud fraction increases and

OLR decreasing as cloud fraction decreases (Fig. 4(c-d)). Thus, at least for the parameters we consider, changes in outgoing radiation are driven by changes in cloud rather than changes in clear sky conditions. Changes in clear sky OLR are small likely because upper tropospheric temperature (UTT) and humidity (UTH) change to partially compensate the impact each one has on OLR. These results do not appear to be an artifact of our focus on four parameters or “tuning” as similar results hold for a set of 100 randomly sampled parameter configurations.

b. Other climatologies

In this section we investigate how changes in the climatology of other variables are related to global-mean OLR and RSR. If these variables are independent of OLR and RSR then including them as separate variables in our analysis might have benefit – see for example the climate prediction index of Sexton and Murphy (2011). In this section we focus on large-scale averages over a many of our simulations. In Section 5 we consider, in more detail, those configurations that are close to the CERES and ERBE observations. In this section we carry out a weighted regression between global-mean OLR and RSR with several variables. The weights we use are the areas of the Voronoi polygons for each model configuration in OLR & RSR.

A Voronoi polygon is the polygon enclosing the region of space closer to a point than to any other point (Aurenhammer 1991). Using this weighting gives smallest weight to those individual configurations that are close, in terms of their simulated OLR/RSR values, to other configurations. We set a maximum weight of π for each polygon to stop large polygons on the boundary biasing our results. We chose this factor to give a disk of radius 1 corresponding to our precision for tuning. The effect of this maximum weight is to reduce the variance explained in our results below. We computed the Voronoi polygons for all our results and then selected the points and weights only for those where the RSR and OLR was within 10 Wm^{-2} of the CERES observations.

We first focus on changes in global-mean or land-average values. We find that land-temperature, global-mean precipitation, fractional cloud, 250 hPa RH and land precipitation are all strongly related to global-mean all-sky OLR and RSR with these two variables explaining more than 67% of the variance in the other variables (Fig. 5). In this figure the spread along the x-axis is a measure of the variability unexplained by OLR and RSR. For global-mean precipitation, cloud and 250 hPa relative humidity more than 90% of the variance is explained. Only for clear sky OLR is a small amount (38%) of the variance explained.

For these variables we also repeated the analysis but using land-temperature as a predictor. These regression models are much poorer than regressions using OLR and RSR. Internal climate variability could affect these regres-

sions and using the ensemble of 19 HadAM3 default configurations found that the impact is small for the variables investigated here (Fig. 5).

As already discussed clear sky OLR is less than the Loeb et al. (2009) estimate. We can also investigate if land-mean precipitation and temperatures are consistent with other observations. For observed land-precipitation we use the CRU TS3.1 dataset. This dataset is an update of the CRU TS2.1 data (Mitchell and Jones 2005) but without including any new corrections for changes in observing practise. We compute average area-weighted precipitation from it from all land-points north of 60° S. For 2000-2005 period average this precipitation is 2.14 mm/day. This value is smaller than many, but far from all, of the parameter configurations considered (Fig. 5(c)). We can also compare the simulated temperature records for 2001-2005 with the CRUTS3.1 dataset. None of the configurations have simulated land-surface temperatures greater than the observed estimate of 286.9K.

Considering only the configurations that have a RMS error of less than a Wm^{-2} with the ERBE and CERES observations (Fig. 5) we find that the scatter for these configurations is approximately the same for both cases. This scatter is the variability unexplained by changes in OLR and RSR and is about 1K (land temperature), 0.1mm/day (global-mean precipitation), 0.2 mm/day (land precipitation), 5% (cloud fraction), 2 Wm^{-2} (OLR) and 5% (RH). These ranges are much larger than that generated by chaotic climate variability suggesting that there is some scope for tuning *both* outgoing radiation and other facets of the simulated climate. In Section 5 we explore the surface climatologies of those configurations.

We now repeat the regression calculations on 5-year averages of zonal-mean temperature and humidity. As before we only use those configurations that are within 10 Wm^{-2} of the CERES observations and weight them by the area of Voronoi polynomials. We first consider the variance explained by OLR and RSR in zonal-mean temperature (Fig. 6(a)). Over almost all of the model troposphere more than 50% of the variance is explained by OLR and RSR and in much of the mid to upper troposphere more than 75% of the temperature variance is explained. There are only a few regions where the fraction of variance explained drops below 50%. Carrying out a similar analysis for relative humidity (Fig. 6(b)) we find over almost all of the troposphere more than 50% of the variance can be linearly explained by changes in OLR and RSR. There are regions in the tropical troposphere at around 300 hPa and in the extra-tropical upper troposphere where more than 75% of the variance is related to changes in outgoing radiation. Water vapour in these regions is particularly important as it has impact on the outgoing radiation (see earlier discussion). In summary for zonal-mean temperature and relative humidity much of the intra-configuration variability is

strongly correlated with changes in global mean OLR and RSR.

We repeated these calculations on the gridded 5-year averages for fields from the different configurations weighting, as before, by the area of the Voronoi polynomials. We find that for large parts of the world global-mean RSR and OLR explain more than 75% of the variance in RSR and OLR (Fig. 7(a) & (b)). For RSR over much of the tropics and off the coast of Antarctica more than 75% of the variance can be explained by changes in global-mean RSR and OLR. For OLR there are large parts of the extra-tropics where more than 90% of the variance can be explained by global-mean OLR and RSR. In only a very few regions is the variance explained in OLR less than 50%. RSR in the northern extra-tropics is not well explained by changes in global-mean OLR and RSR suggesting that some inclusion of spatial information may improve our optimisation algorithm. With the exception of this our results suggest that the parameter changes are coherently affecting outgoing radiation across the planet supporting our focus on global-mean OLR and RSR.

Clear sky OLR has a much weaker relationship with global average all-sky OLR and RSR than all sky OLR (Fig. 7(c)). Though there are regions in the tropics and polar regions where more than 50% of the variance in clear sky OLR can be explained by global-average OLR and RSR over most of the mid-latitudes the variance explained is 25-50 %.

Near-surface temperature has some parts of the world where more than 50% of the variance in grid-scale values can be explained by the TOA all-sky fluxes (Fig. 7(d)). These regions are much of Asia, inland North America and Antarctica.

Precipitation does not show such a clear relationship, as the other variables we have considered, with global-mean OLR and RSR with a quite noisy appearance. Nevertheless, this linear relationship explains more than 50% of the variance across parts of the world mainly in the tropics. Over much of the world more than 50% of the variance in mean sea level pressure can be explained by OLR and RSR. This is particularly apparent in the tropics where there are large regions where more than 75% of the variance can be explained by OLR and RSR. As the parameters we adjust affect the model convection, cloud and rainfall scheme then we speculate that they are affecting the atmospheric heating rates and thus the atmospheric circulation particularly in the tropics.

Overall there are very strong relationships between global-mean OLR and RSR and other climatologies as we perturb model parameters. This supports our focus on OLR and RSR rather than optimising on several variables as others have done. However, there is still variability unexplained by changes in OLR and RSR which is larger than that generated by internal climate variability. This suggests we might have been able to tune surface temperature and pre-

precipitation to bring them closer to observations as well as OLR and RSR. We examine the surface climatologies of those configurations in the next section.

5. Equi-finality

Equi-finality (Beven and Freer 2001) is when different parameter values lead to similar outputs and frequently arises in complex environmental models. In this section we examine in more detail those configurations whose simulated RSR and OLR are in good agreement CERES and ERBE observations. We examine the parameter values for the configurations to see if 1) the parameter values are constrained by OLR and RSR values and 2) if the simulated temperature and precipitation climatologies are related to particular parameter values. We consider the 37 (41) configurations that have a RMS error of less than 1 Wm^{-2} from the CERES (ERBE) observations of which 20 (14) are the best fit cases arising from earlier optimisation trials.

First considering land, north of 60°S , temperature (Fig 5(a)) For configurations close to the CERES values that there are two groups with a gap between them while for configurations close to the ERBE values there is an approximately 1K range of land temperatures but no obvious gap.

To simplify further analysis we group the model configurations into five groups and average within each group. These groups are: 1) The standard configuration of HadAM3 with different initial conditions (Standard); 2) the configurations close to CERES that are colder than 285.3 K (CERES cold group); 3) the configurations close to CERES that are warmer than 285.3K (CERES warm group); 4) the configurations close to ERBE that are colder than 285.6 K (ERBE cold group); 5) the configurations close to ERBE that are warmer than 285.6K (ERBE warm group). These values were chosen to keep the groups approximately equally sized (Table 4) and allow us to see if there is any relationship between surface temperatures and other facets of the simulated climatologies.

The warm CERES and ERBE groups are warmer, by construction, than the Standard configuration and so are in better agreement with observed large-scale temperatures. The CERES cold group has similar temperatures to the Standard configuration though with larger temperature variability than internal climate variability. The global-average precipitation is not greatly different in the two CERES groups though this is a result of a broad spread (Fig. 5(b)) across the two groups. The warm CERES group, on average, has 0.1 mm/day greater land precipitation (Table 4) than the cold CERES group and Standard configuration. However, there is significant variability within the two groups with the two groups overlapping. For global and land average precipitation both ERBE groups overlap each other (Fig. 5(b-c)) and have similar average land-precipitation

(Table 4).

Both CERES groups are well separated for cloud, clear sky OLR and 250 hPa Relative Humidity (Fig. 5)(d-f)). In contrast, the two ERBE groups are not strongly separated. Average values of the groups show that each ERBE group, compared to the equivalent CERES group, has more cloud, less land precipitation, similar clear-sky OLR and larger 250 hPa relative humidities. Compared to the equivalent warm groups the cold groups have more cloud, less clear sky OLR and larger 250 hPa relative humidity (Table 4). In general the standard configuration is most similar to the cold CERES group.

For the configurations close to the ERBE and CERES observations individual parameter values(Fig. 8) can have values across their valid range. However, this requires correlation between different parameter values. For example high values of `vf1` are associated with high values of `rhcrit`(Fig. 8(c)). While, high values of `ct` are associated with high values of `entcoef` (Fig. 8(d)). Jackson (2009) following Jackson et al. (2008) examined the probability distribution of several parameters and, like us, found multiple parameter values could lead to similar model simulations.

The `ct` parameter appears to be responsible for the behaviour of the warm vs cold groups with both warm groups having values of around 10^{-4} . The cold groups have more scatter with the cold CERES group having `ct` values around $3 - 4 \times 10^{-4}$ while the cold ERBE group has `ct` values around $2 - 4 \times 10^{-4}$ and less separation between the two groups. Of the other three parameters only `entcoef` shows some sensitivity to group with the cold CERES group have generally lower values than the warm group. The cold ERBE group has, on average, lower values of `entcoef` there is some overlap between the two groups.

The correlation exhibited between the parameter values suggests that very different parameter choices can lead to similar surface and outgoing radiation climatologies leading support to the idea that HadAM3 can exhibit equi-finality. However, as we've already seen different parameter choices can modify the surface temperature, and other variables, and still produce similar values of outgoing radiation. Thus, if we had used more parameters and observables in our analysis we might have been able to reduce this equi-finality.

We now compare the standard configuration, and averages from the four groups with observations focusing on zonal-mean values. As before we compare with the CRUTS3.1 dataset. In very broad terms all configurations capture the broad features of the observed zonal-mean temperatures (Fig. 9(a)) with largest temperatures in the tropics and cooling towards the extra-tropics. This is not greatly surprising – some of this will be driven by the SSTs and some by resolved flows. Examining the dif-

ferences between the various simulated land temperatures and observations (Fig. 9(b)) we can see that the cold ERBE and CERES groups and Standard configurations have similar temperature differences from observations. The patterns of zonal-mean error in all five groups appear to be similar with a simple shift explaining most of the differences. Both warm groups have smallest differences from observations near the equator and largest at high northern latitudes though still cooler than the observations. This suggests that other parameters, probably closely related to the land surface scheme, need adjusting or better representation of land surface processes are needed to remove these temperature biases while producing outgoing radiation values that are comparable with observations.

Examining simulated and observed zonal-mean land precipitation now (Fig. 9(c)) we see that the simulations capture the major observed features but are all more similar to one another than they are to the observations. The major differences from observations is that all model configurations have too much precipitation at 45°S. This is due to excessive precipitation over the southern Andes likely due to excessive diffusion of moisture over the steep orography in this region. All model configurations also have too much precipitation north of about 30°N and produce too little precipitation in the equatorial region with this being particularly apparent for the warm group. Looking at the differences between observations and model configurations (Fig. 9(d)) we see that the Standard configuration and both cold groups have similar errors: about a mm/day too little precipitation near the equator and about half a mm/day too much precipitation in northern extra-tropics. Both warm groups are too dry by about one and a half mm/day in the equatorial regions but elsewhere have similar errors to the Standard and cold group.

We now focus on the average difference between the CERES warm and cold groups. Near surface temperature differences (Fig. 10(a)) show spatial heterogeneity with some regions being 1K warmer and others close to zero. Most of the land is between 0.5-1K warmer. Regions with the least warming appear in the wet tropics while largest warming is at high northern latitudes, India and parts of East Africa. Precipitation differences (Fig. 10(b)) appear to be important only in the tropics with a reduction in land precipitation and an increase in ocean precipitation. Perhaps these precipitation differences are water vapour transports from ocean to land being affected by the parameter choices. MSLP differences are also apparent with increases in pressure in the Indian and North Atlantic Oceans and reductions in the mid-latitude north and south Pacific. Earlier we found the changes in cloud were the dominant contributor to changes in RSR and OLR. Cloud fractions are about 30 % less off the coast of Peru, 20 % smaller off the coast of Namibia and 5-10% less in the Arctic in the warm group than in the cold group. Only in the tropical Pacific and

Atlantic is there an increase in fractional cloud of 5-10% in the warm group relative to the cold group.

We repeated this calculation for ERBE warm - cold and found similar patterns to that for the two CERES groups though with a greater magnitude (not shown). This suggests that the perturbed climatologies that result from perturbing the model physics have differences that are large scale and systematic.

To summarise our analysis has found evidence of some equi-finality with different parameter combinations that produce similar outgoing radiation also producing similar surface climatologies. The values for individual parameters have a broad range though that requires correlation between the parameters. We defined two groups based on their land temperatures and found the warm group, regardless of use of ERBE or CERES observations, was warmer over the land everywhere but drier in the tropics. The main differences between the groups arose from changes in the τ parameter which controls the rate at which cloud water turns to precipitation. When it is high HadAM3 has a warmer land but produces less rain over the tropical land. The impact of this is to make the land-temperature simulation warmer and in better agreement with observations but the simulation of tropical land-precipitation worse.

6. Summary and Conclusions

Our results show that tuning atmosphere-only models and exploring parameter space is straightforward using an optimisation method. It is computationally efficient requiring short simulations of an atmospheric only model. Our results suggest that, if other atmospheric models behave like HadAM3, one approach to developing a climate model would be to automate a manual tuning process (Mauritsen et al. 2012) by first automatically optimising the atmospheric model to have near-radiative balance before coupling it to an ocean model.

We systematically explored the reliability of our optimisation algorithm by changing the target values and starting from all 16 combinations of extreme parameters values. We found convergence, to a RMS error of 1 Wm^{-2} or less, took a median of four iterations. Failure to converge, for most targets, occurred about once in 16 cases. For one target we found many more failures to converge. This suggests that convergence depends on the target reflecting the local nature of the optimisation algorithm we are using.

Though HadAM3 has been superseded and so it is no longer state-of-the-art the application of our techniques to more computationally complex models should be feasible. The algorithm has two steps: the first in which finite differences are computed and the second in which the line-search is performed. For the four parameters we considered the first step required 4 model simulations and the second step required 3 model simulation. In each step each

simulation can be run in parallel. Modelling centres with more resources than were available to us should be able to apply the techniques we used to more state-of-the-art climate models as each optimisation would, on average, require about 225 simulated atmospheric years. This assumes, on average, about 36 model-evaluations each of 6.25 years in length. We expect that the optimisation algorithm we used should scale to order 10 parameters with a linear increase in the number of model evaluations per evaluation though would need to explore how the number of iterations behaves. It may be possible to reduce the length of the simulation to 18 months as internal climate variability in OLR and RSR is small which would reduce the number of simulated years to 54.

Looking in more detail at the model climatologies we found that the changes in several large scale variables was strongly related to changes in reflected shortwave radiation (RSR) and outgoing longwave radiation (OLR). We found little change in clear sky RSR and OLR largely due, for the latter, due to a cancellation between changes in upper tropospheric water and atmospheric temperatures which may be due to the HadAM3 model structure. As a consequence of this cancellation changes in outgoing radiation were largely due to changes in model cloud.

We also explored the climatologies for the configurations that had a root-mean-square error of less than 1 Wm^{-2} to either the CERES or ERBE observations. Parameters varied across a broad range but with correlation between the parameters. This led to some equi-finality with different parameter combinations leading to similar climatologies. The main impact on the surface climatologies was due to the parameter (ct) that controls the rate at which water cloud droplets are converted to precipitation. When ct was high configurations had climatologies that were warmer but had less precipitation in the tropics than the standard configuration.

Earlier evaluation of HadAM3 (Pope et al. 2000) found that the model was too cool by about 2K and too wet by about 1 mm/day. We still find that the model is too cool but that model precipitation was about correct. As part of the development of our configuration we investigated if changes in total solar irradiance or fixing the Rayleigh scattering error could explain this difference. We found these had little impact on land precipitation. Interestingly, changing the Sea Surface Temperatures used to drive the atmospheric model from HadISST (Rayner et al. 2003) to Reynolds et al. (2002) which is a blend of *in situ* and satellite data reduced the land precipitation by 0.2 mm/day. However, that does not bring our results into agreement with that of Pope et al. (2000). Remaining differences might be the use of a different precipitation dataset, a different validation period or the use of Aerosols and non CO_2 gases in our experiments.

Looking at the model climatologies we obtained, in

more detail, we found a group of HadAM3 configurations which were warmer (and so less in error) but drier in the tropics. It might be that some existing GCM errors have arisen from tuning to a poor minimum and that tuning from extreme parameter values might lead to better simulation of current climate.

Others have also explored optimisation techniques though examining different problems to us. Jones et al. (2005) tuned the low-resolution FAMOUS atmosphere/ocean model by iteratively adjusting six variables. Each iteration required thirteen simulations each of 100 years for a grand total of 6500 simulated years. Jackson et al. (2004) required an average of 80 model evaluations for estimating the posterior distribution of four parameters and all the evaluations ran in serial rather than partially in parallel as ours do. Ollinaho et al. (2012), like us, used four variables but, unlike us, carried out a set of 10-day ensemble atmospheric forecasts with each ensemble member perturbing a parameter. After reweighting these forecasts they had generated best estimate parameters and a distribution for them assuming that the parameters were multi-normally distributed. They also compared their best estimate results with those of Loeb et al. (2009) and found zero error in the total outgoing radiation but RMS errors of about 1.6 Wm^{-2} larger than our convergence criteria. Their approach used 180 50-member ensembles with each forecast being of length 10 days for a total of about 250 simulated years.

Finally, answering the questions we posed in the introduction. It is feasible to use optimisation methods to tune atmospheric mode to radiation observations. Many facets of the model climatologies are related to these radiation observations and there appears to be little difference in the surface climatologies that result when we tuned to the older ERBE results of Fasullo and Trenberth (2008) rather than the recent CERES results of Loeb et al. (2009). However, we did find that model configurations with similar outgoing radiation values could have different surface temperature and precipitation climatologies.

Acknowledgments.

We thank the two anonymous reviewers whose comments improved this paper. SFBT was supported by the National Centre for Earth Observations (NERC grant NE/F001436/1). Support for MJM and computer time on the Edinburgh Computing and Data Facility was provided by the Centre for Earth Dynamics which is part of the Scottish Alliance for GeoScience, Environment and Society. CC, SFBT and MJM also acknowledge support from “Bridging the Gaps” and “Maximaths” funding. DJR was supported by a NERC PhD studentship with a CASE award from CEH Wallingford.

REFERENCES

- Aurenhammer, F., 1991: Voronoi Diagrams - A Survey Of A Fundamental Geometric Data Structure. *Computing Surveys*, **23** (3), 345–405, doi:ISI:A1991GM03300003.
- Beven, K. and J. Freer, 2001: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, **249** (1-4), 11–29.
- Conn, K. S., A. R. and L. N. Vicente, 2008: *Introduction to Derivative Free Optimization*. MPS-SIAM Series on Optimization., SIAM, Philadelphia.
- Fasullo, J. T. and K. E. Trenberth, 2008: The annual cycle of the energy budget. Part I: Global mean and land-ocean exchanges. *J. Clim.*, **21** (10), 2297–2312, doi:10.1175/2007JCLI1935.1.
- Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.*, **16**, 147–168.
- Gregoire, L., P. Valdes, A. Payne, and R. Kahana, 2011: Optimal tuning of a GCM using modern and glacial constraints. *Climate Dynamics*, **37**, 705–719, doi:10.1007/s00382-010-0934-8.
- Jackson, C., 2009: Use of Bayesian inference and data to improve simulations of multi-physics climate phenomena. *Journal of Physics: Conference Series*, IOP Publishing, Vol. 180, 012029.
- Jackson, C., M. K. Sen, and P. L. Stoffa, 2004: An efficient stochastic bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *Journal of Climate*, **17** (14), 2828–2841.
- Jackson, C. S., M. K. Sen, G. Huerta, Y. Deng, and K. P. Bowman, 2008: Error Reduction and Convergence in Climate Prediction. *J. Clim.*, **21** (24), 6698–6709, doi:10.1175/2008JCLI2112.1.
- Järvinen, H., P. Räisänen, M. Laine, J. Tamminen, A. Ilin, E. Oja, A. Solonen, and H. Haario, 2010: Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmos. Chem. Phys.*, **10**, 9993–10 002.
- John, V. O., G. Holl, R. P. Allan, S. A. Buehler, D. E. Parker, and B. J. Soden, 2011: Clear-sky biases in satellite infrared estimates of upper tropospheric humidity and its trends. *J. Geophys. Res.*, **116**, D14 108, doi:10.1029/2010JD01535.
- Jones, C. D., J. M. Gregory, R. B. Thorpe, P. M. Cox, J. M. Murphy, D. M. H. Sexton, and P. Valdes, 2005: Systematic optimisation and climate simulation of FAMOUS, a fast version of HadCM3. *Clim. Dyn.*, **25** (2-3), 189–204, doi:10.1007/s00382-005-0027-2.
- Kelley, C. T., 2011: *Implicit Filtering*. SIAM, Philadelphia.
- Knight, C. G., et al., 2007: Association of parameter, software, and hardware variation with large-scale behavior across 57,000 climate models. *Proc. Natl. Acad. Sci. U. S. A.*, **104** (30), 12 259–12 264, doi:10.1073/pnas.0608144104.
- Knutti, R. and G. Hegerl, 2008: The equilibrium sensitivity of the Earth’s temperature to radiation changes. *Nature Geoscience*, **1** (11), 735–743.
- Kopp, G. and J. L. Lean, 2011: A new, lower value of total solar irradiance:evidence and climate significance. *Geophys. Res. Lett.*, L01706, doi:10.1029/2010GL045777.
- Liu, P., 2010: Parameter estimation for climate modelling. M.S. thesis, School of Mathematics, University of Edinburgh.
- Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the earth’s top-of-atmosphere radiation budget. *J. Clim.*, **22**, 748–765, doi:10.1175/2008JCLI2637.1.
- Mauritsen, T., et al., 2012: Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, **4**, M00A01, doi:10.1029/2012MS000154.
- Medvigy, D., R. Walko, M. Otte, and R. Avissar, 2010: The ocean-land-atmosphere model: Optimization and evaluation of simulated radiative fluxes and precipitation. *Monthly Weather Review*, **138** (5), 1923–1939.
- Mitchell, T. and P. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, **25** (6), 693–712.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.
- Neelin, J. D., A. Bracco, H. Luo, J. C. McWilliams, and J. E. Meyerson, 2010: Considerations for parameter optimization and sensitivity in climate models. *Proc. Natl. Acad. Sci. U. S. A.*, **107** (50), 21 349–21 354, doi:10.1073/pnas.1015473107.

- Nocedal, J. and S. J. Wright, 2006: *Numerical Optimization*. 2d ed., Springer Verlag.
- Ollinaho, P., M. Laine, A. Solonen, H. Haario, and H. Järvinen, 2012: Nwp model forecast skill optimization via closure parameter variations. *Quarterly Journal of the Royal Meteorological Society*, doi:10.102/qj.2044.
- Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parametrizations in the Hadley Centre climate model – HadAM3. *Clim. Dyn.*, **16**, 123–146.
- Powell, M. J. D., 1998: Direct search algorithms for optimization calculations. *Acta Numerica*, **7**, 287–336.
- Randall, D. A., et al., 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. C. Marquis, K. B. Averyt, M. Tignor and H. L. Miller (eds.)]*, 589–662.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108 (D14)**, 10.1029/2002JD002670.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Q. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15 (13)**, 1609–1625.
- Rowlands, D., et al., 2012: Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nature Geoscience*, **5 (4)**, 256–260.
- Sanderson, B. M., et al., 2008: Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. *J. Clim.*, **21 (11)**, 2384–2400, doi:10.1175/2008JCLI1869.1.
- Schirber, S., D. Klocke, R. Pincus, J. Quaas, and J. L. Anderson, 2013: Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *Journal of Advances in Modeling Earth Systems*, 58–70, doi:10.1029/2012MS000167.
- Severijns, C. and W. Hazeleger, 2005: Optimizing parameters in an atmospheric general circulation model. *J. Clim.*, **18 (17)**, 3527–3535.
- Sexton, D. M. H. and J. M. Murphy, 2011: Multivariate probabilistic projections using imperfect climate models. Part II: robustness of methodological choices and consequences for climate sensitivity. *Clim. Dyn.*, doi:10.1007/s00382-011-1209-8.
- Stainforth, D. A., et al., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.
- Tett, S. F. B., et al., 2007: The impact of natural and anthropogenic forcings on climate and hydrology. *Clim. Dyn.*, **28 (1)**, 3–34, doi:10.1007/s00382-006-0165-1.
- Walko, R. and R. Avissar, 2008: The ocean-land-atmosphere model (OLAM). Part I: Shallow-water tests. *Monthly Weather Review*, **136 (11)**, 4033–4044.
- Wright, M. H., 1995: Direct search methods: once scorned, now respectable. *Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis*, D. F. Griffiths and G. A. Watson, Eds., Addison-Wesley Longman, Harlow, Essex, 191–208.

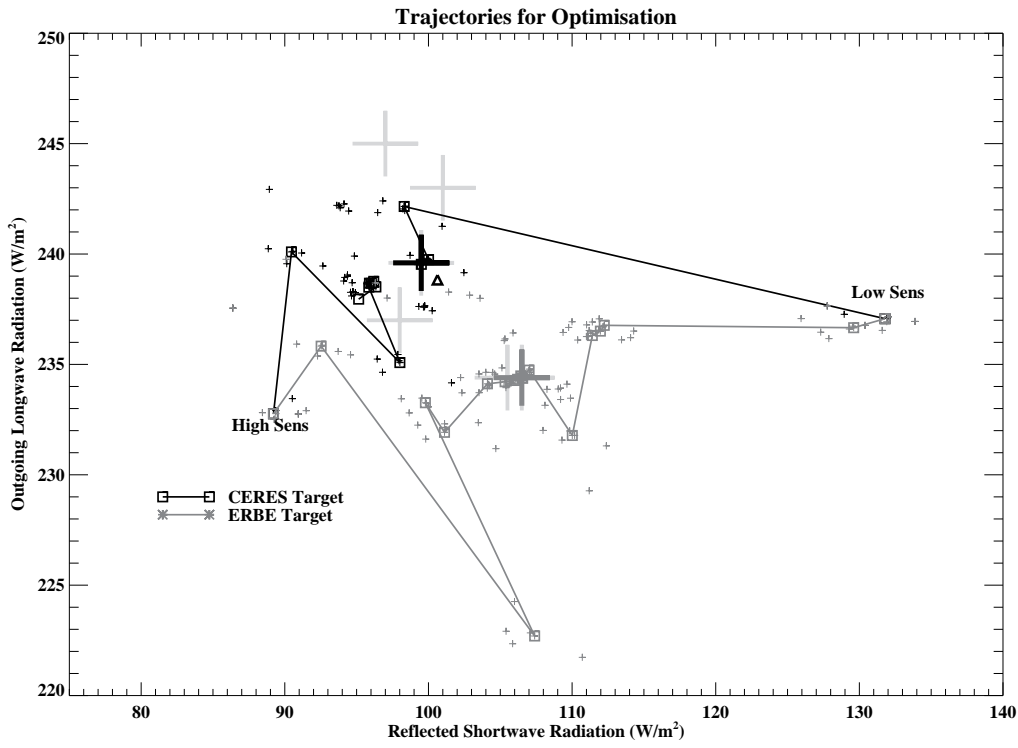


FIG. 1. Optimisation trajectories beginning from high and low climate sensitivities to CERES (black lines) and ERBE (gray lines) targets. Crosses show the observed targets. Small pluses are values from the individual simulations. The black triangle shows the average values for the default HadAM3 configuration while the light gray crosses show four other targets used.

TABLE 1. Parameters, their default value, allowed range and step size used in derivative calculations. Bracketed values are 75% of range used for many of our optimisation experiments.

Parameter	Default Value	Range	Step Size
entcoef	3	0.5–10 (1.1–8.3)	0.5
vf1	1	0.01–2.5 (0.26–2.1)	0.2
ct	1.0×10^{-4}	$0.5 - 4 (1-3.2) \times 10^{-4}$	2×10^{-5}
rhcrit	0.7	0.6 – 0.9 (0.63–0.85)	0.05

TABLE 2. Optimisation simulations. Shown are initial and final configurations from CERES and ERBE targets with both starting from configurations with largest and smallest estimated equilibrium climate sensitivity (see text). Also show are parameter choices and RMS error (Wm^{-2}) for initial choice and final state of the optimisation. Error is computed against the original observed values. Also shown are the number of iterations and the number of simulated years. Parameter choices are shown in order: **entcoef**, **vf1**, **ct** $\times 10^{-4}$ and **rhcrit**. Also shown is Standard configuration of HadAM3.

Name	Initial		Final		Iter. (Yrs.)
	Parameter	Error (Wm^{-2})	Parameter	Error (Wm^{-2})	
HadAM3	(3, 1, 1, 0.7)	0.7	-	-	-
CERES Low	(9.00, 2.00, 0.50, 0.70)	32.4	(4.16, 2.32, 0.86, 0.89)	0.1	3 (137)
CERES High	(0.60, 0.50, 1.28, 0.85)	12.4	(0.95, 0.66, 0.50, 0.65)	4.7	8 (364)
ERBE Low	(9.00, 2.00, 0.50, 0.70)	25.4	(4.15, 0.83, 0.66, 0.83)	0.1	9(410)
ERBE High	(0.60, 0.50, 1.28, 0.85)	17.4	(2.05, 0.57, 0.50, 0.60)	0.3	8(364)

TABLE 3. Summary of optimisation experiments each one consisting of 16 optimisation cases. All cases except CERES (100%) were started from parameter values set to 75% of their extreme conditions (see Table 1 for values). Shown is the name of the optimisation, the target (RSR, OLR) in Wm^{-2} , how many successfully converged, how many cases there was a model failure and how many optimisations were terminated because they did not converge. The minimum and maximum iterations for the convergent cases are shown as a range with the median number of iterations shown bracketed. Asterisk cases are the “sample” cases (see main text).

Name	Target	Successful	Model Failed	Converge Failed	Iterations
CERES (100%)	(99.5, 239.6)	4	12	0	3-6 (4)
CERES (75%)*	(99.5, 239.6)	15	0	1	1-8 (4)
ERBE*	(106.5, 234.4)	12	4	0	3-7 (3)
Near ERBE	(105.5, 234.4)	13	2	1	2-8 (4)
TGT#1*	(97, 245)	4	1	11	2-9 (5)
TGT#2*	(98, 237)	13	2	1	1-6 (3)
TGT#3*	(101, 243)	12	2	1	2-8 (4)

TABLE 4. Summary of tuned simulations. Shown is name, number of simulations, RSR (Wm^{-2}), OLR (Wm^{-2}), land precipitation (mm/day), land temperature, area Cloud (%), clear sky OLR (Wm^{-2}) and 250 hPa relative Humidity (%).

Name	No.	RSR	OLR	Land Temp.	Land Precip	Cloud	OLRC	UTH
Standard	19	100.6	238.8	285.1	2.3	46.9	259.9	49.3
CERES Cold	17	99.5	239.6	284.9	2.4	47.1	259.1	50.0
CERES Warm	20	99.8	239.5	285.7	2.3	45.1	259.9	47.7
ERBE Cold	23	106.4	234.4	285.2	2.2	52.4	259.3	51.2
ERBE Warm	18	106.8	234.3	286.0	2.2	51.7	259.9	49.5

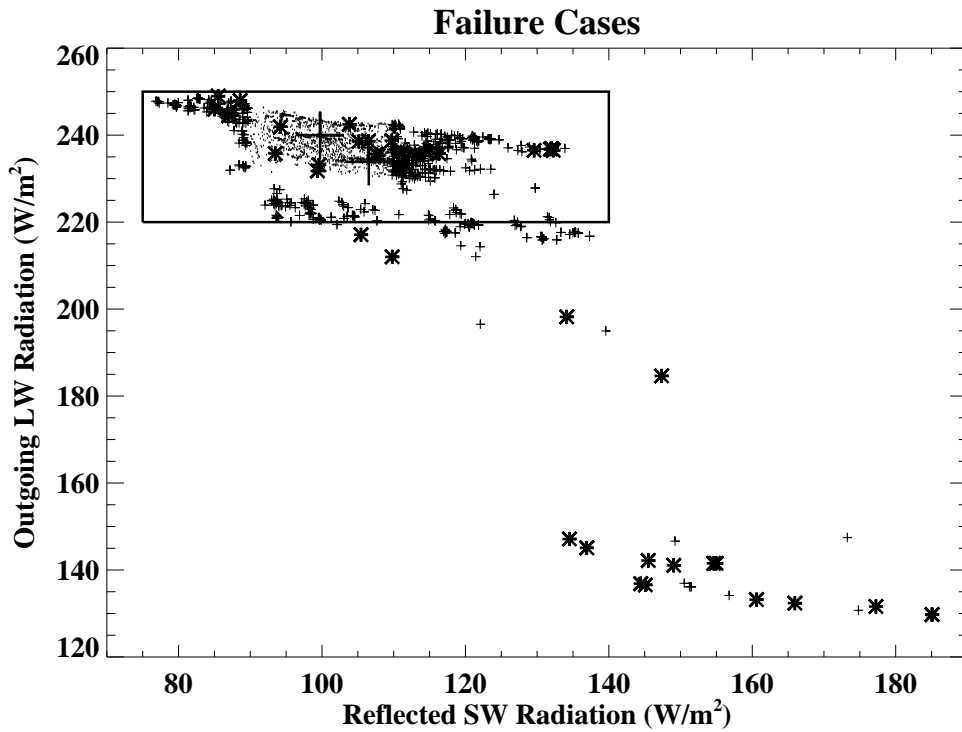


FIG. 2. OLR and RSR values for 1999 from all model configurations. Large asterisks are from model configuration that failed at some point. Dots are values from configurations that did not fail and have RSR and OLR values within 10 Wm^{-2} of the CERES values while small pluses are the remaining successful configurations. The failure cases are about 1% of the total cases. Large pluses show the CERES (upper-left-hand value) and ERBE (lower-right-hand value) while box shows the region shown in Fig. 1.

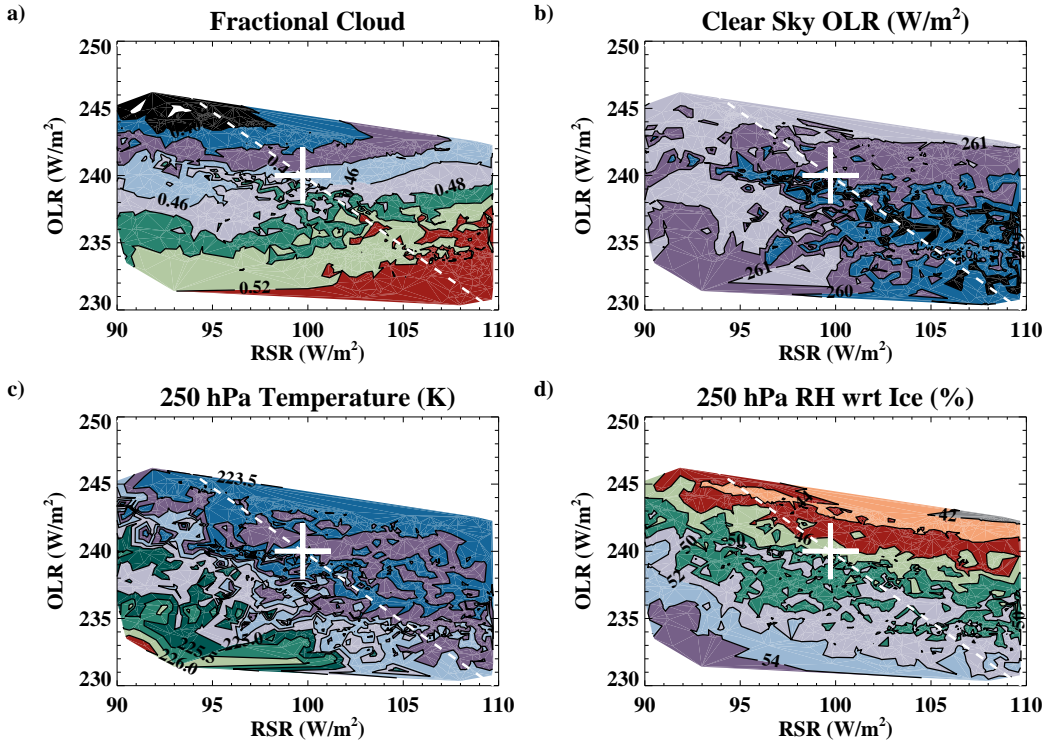


FIG. 3. Total cloud fraction (a), Clear Sky OLR (a), 250 hPa temperature (c) and relative humidity wrt Ice (d) as interpolated functions of RSR and OLR. Contour intervals vary between sub-plots and are 2%, 1 Wm⁻², 1K and 4% in plots (a), (b), (c) and (d) respectively. Only values within 10 Wm⁻² of the CERES observations (white cross) are shown. Dashed diagonal line shows total outgoing radiation agreeing with observed near radiative balance.

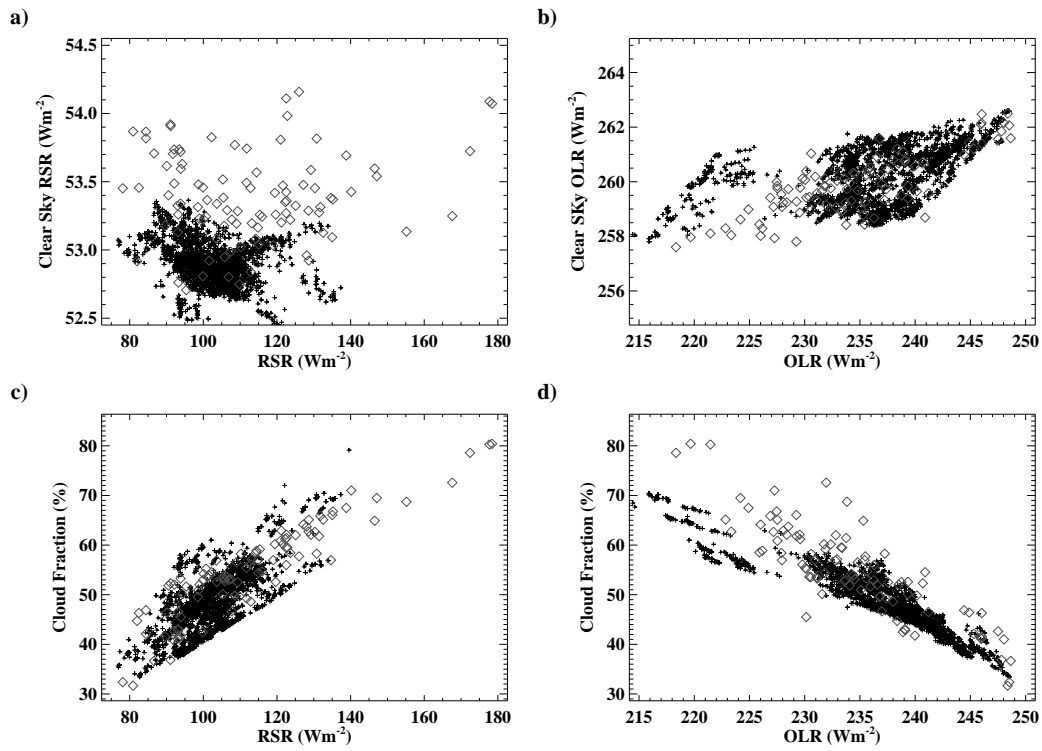


FIG. 4. a) RSR vs Clear Sky RSR. Small crosses are samples from 4-parameter tuning simulations and gray diamonds are from random parameter simulations. b) as a) but for OLR and clear sky OLR. c) as a) but for RSR vs Cloud fraction (%). d) as c) but OLR vs Cloud fraction (%).

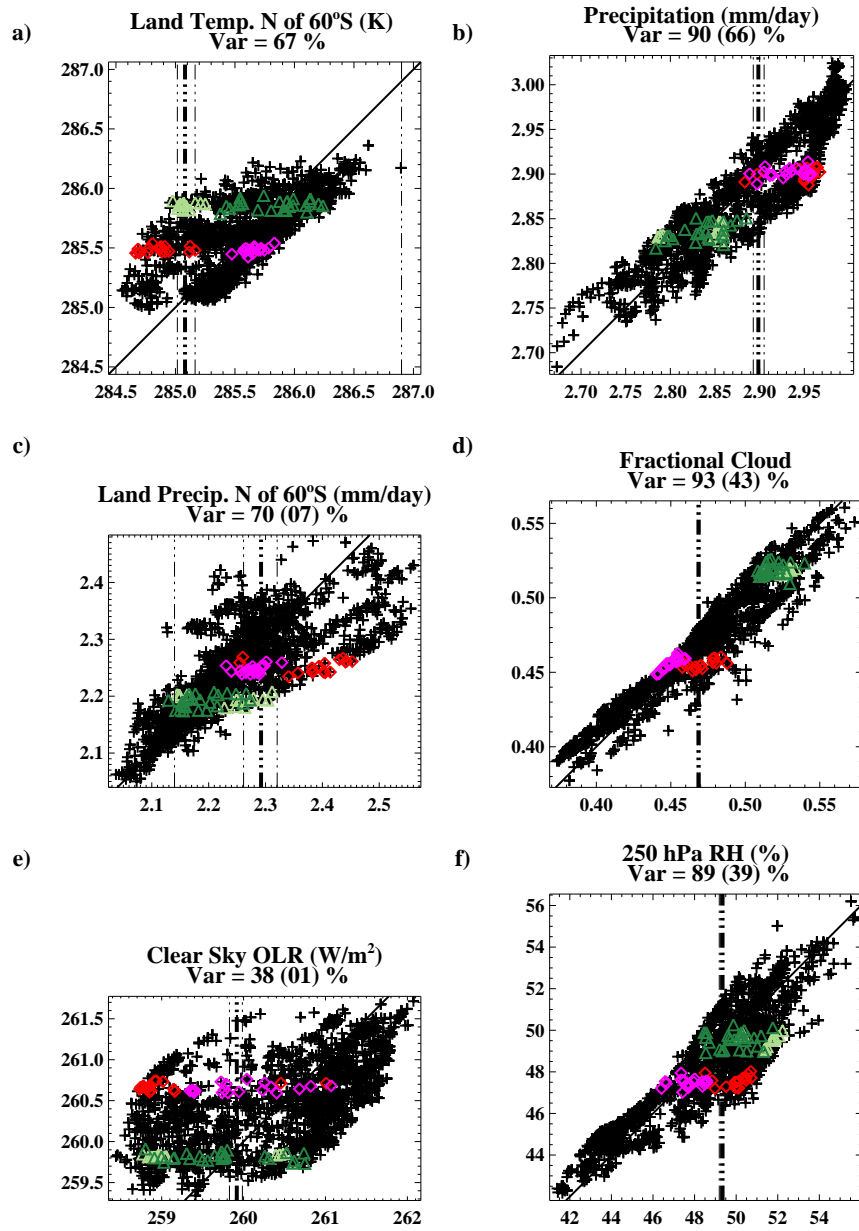


FIG. 5. Scatter (black asterisks) between simulated values (x-axis) and that predicted from regression coefficients of RSR & OLR (y-axis) for average land temperature (a), global-average precipitation (b), average land precipitation (c), global-average fractional cloud (d), global-average clear sky OLR (e) and global-average 250 hPa relative humidity wrt ice(f). Thick vertical thick dot-dashed lines show the mean values from the standard configuration of HadAM3 while the thin dot-dashed lines show the minimum and maximum values from the initial condition ensemble. Dashed lines in plots a) and c) show observed estimates. Title shows variance explained by linear regression on OLR and RSR with bracketed values showing the variance explained by average land-temperature. Red (light green) triangles show the CERES (ERBE) cold group, magenta triangles (dark green) show the CERES (ERBE) warm group – see text for details.

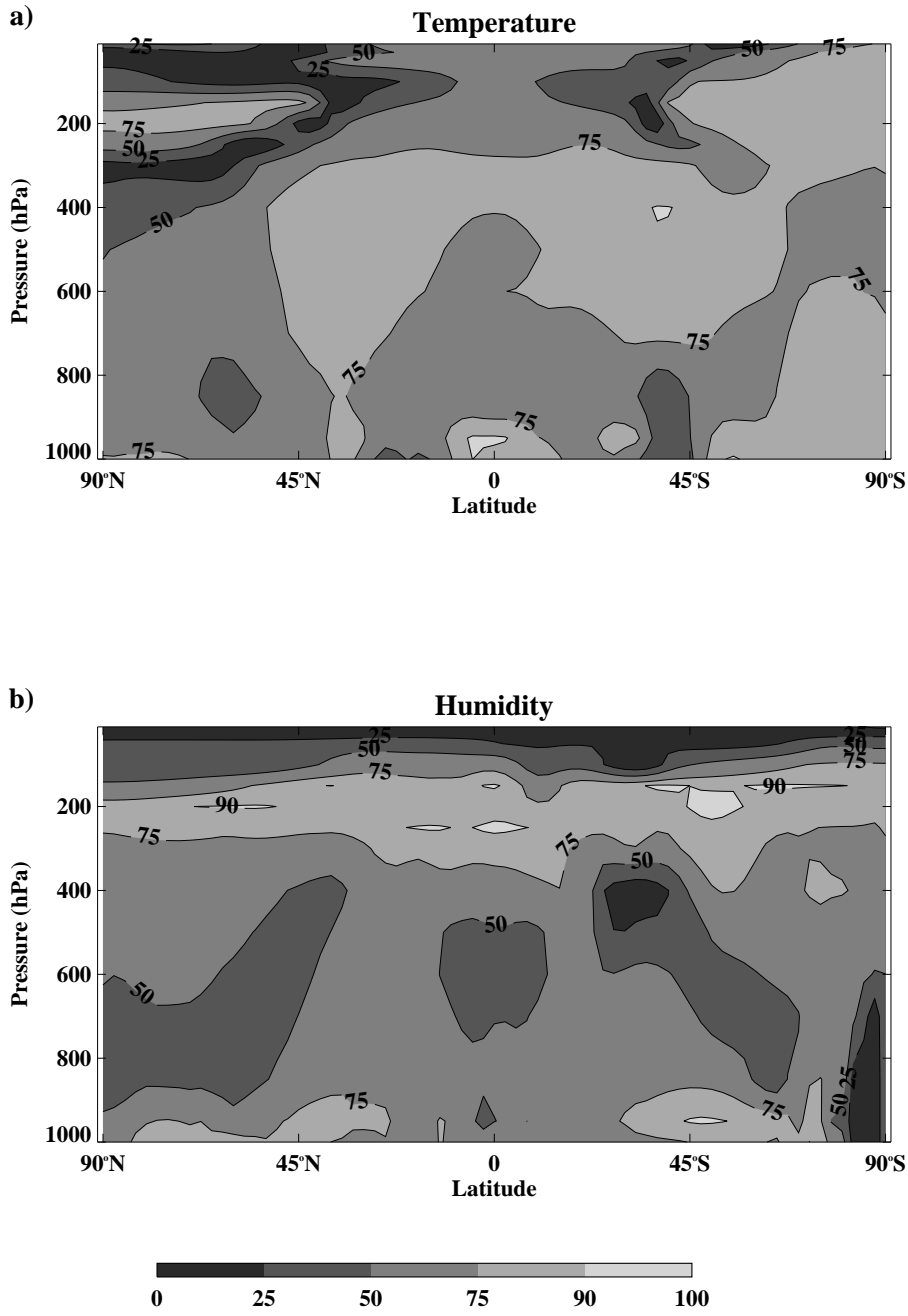


FIG. 6. % Variance explained in zonal-mean Temperature (a) and Relative Humidity (b) explained by global-mean RSR and OLR. Grey shades/contours are at 0, 25, 50, 75 & 90 % of the variance.

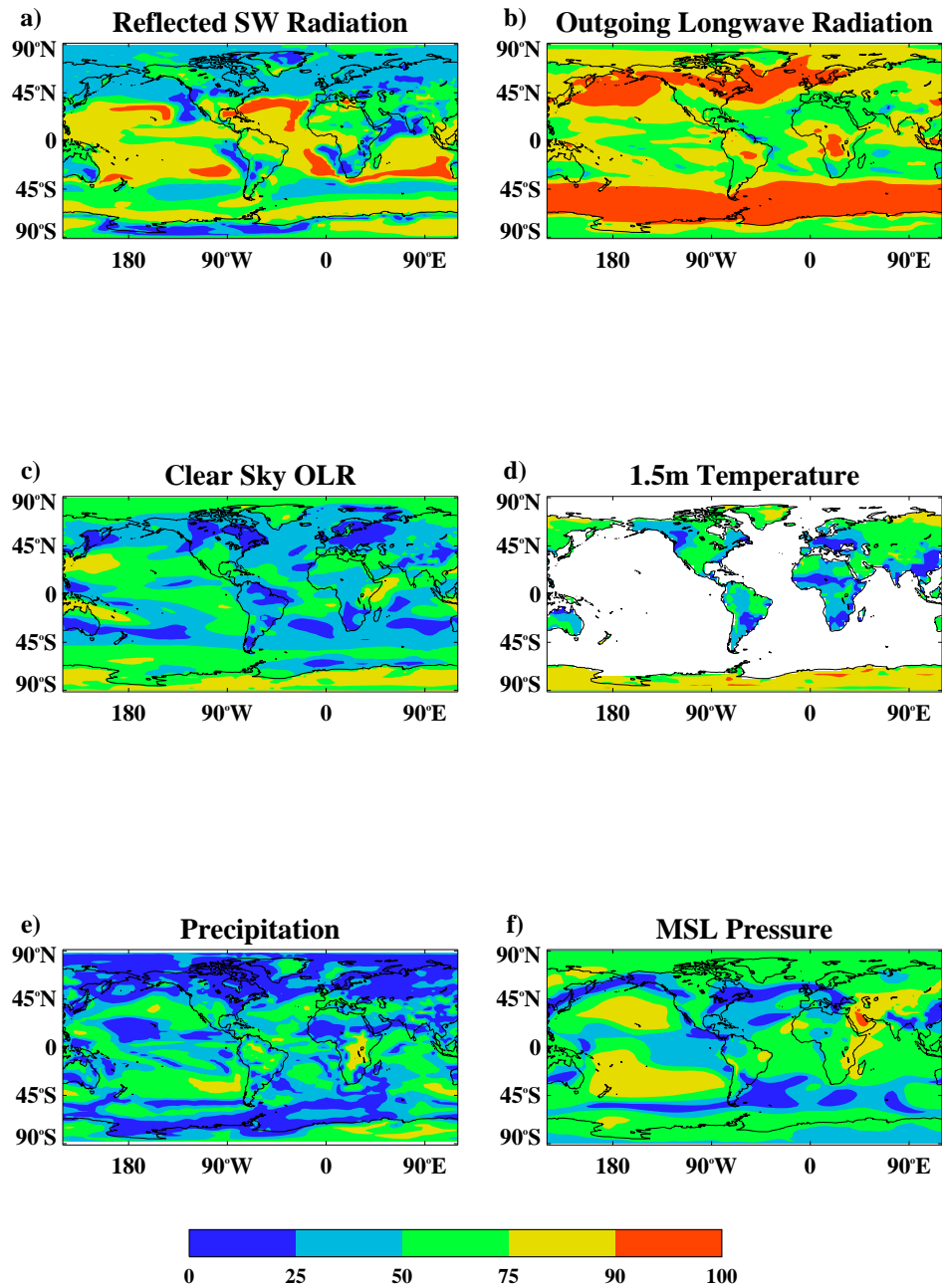


FIG. 7. % Variance explained at each grid-point by global-mean RSR & OLR for RSR (a), OLR (b), clear sky OLR (c), land near-surface temperature (d), Precipitation (e) and MSL Pressure (f). Colours are at 0, 25, 50, 75 & 90 % of the variance.

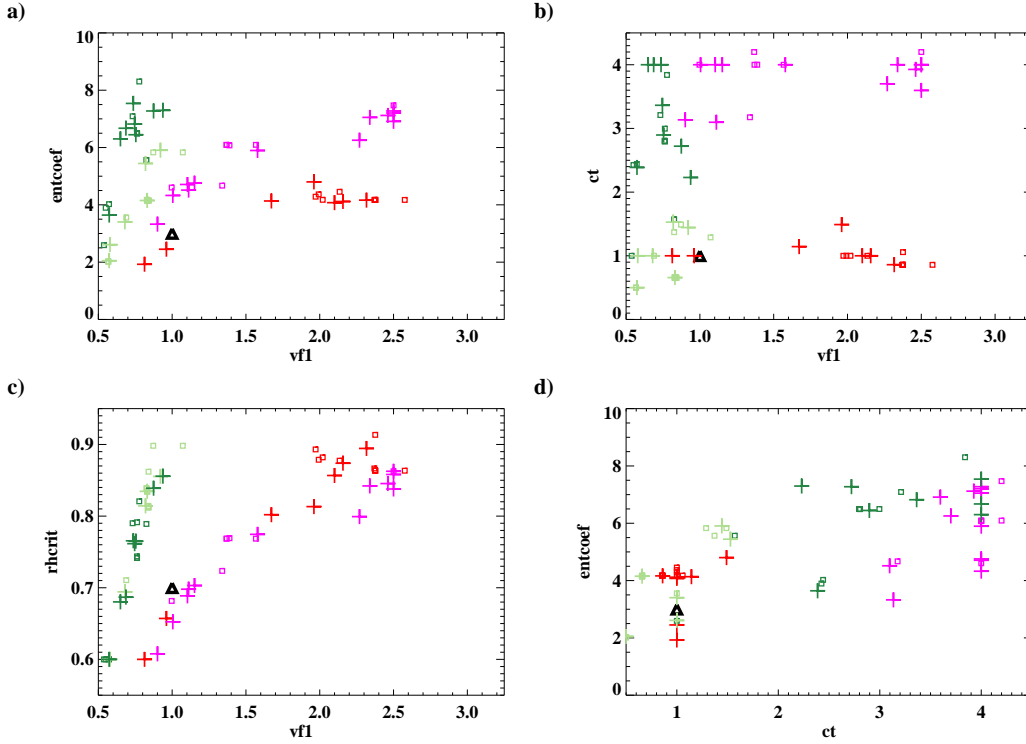


FIG. 8. Scatter plots showing parameters for groups. Values from final iteration of successful “extremes” optimisation experiments (crosses) are coloured according to group they are (see main text) with red (light green) for CERES (ERBE) cold group, magenta (dark green) for CERES (ERBE) warm group. Small coloured squares show values for all simulations with a root-mean-square error of less than 1 Wm^{-2} relative to the particular target. Black triangle shows standard values for HadAM3.

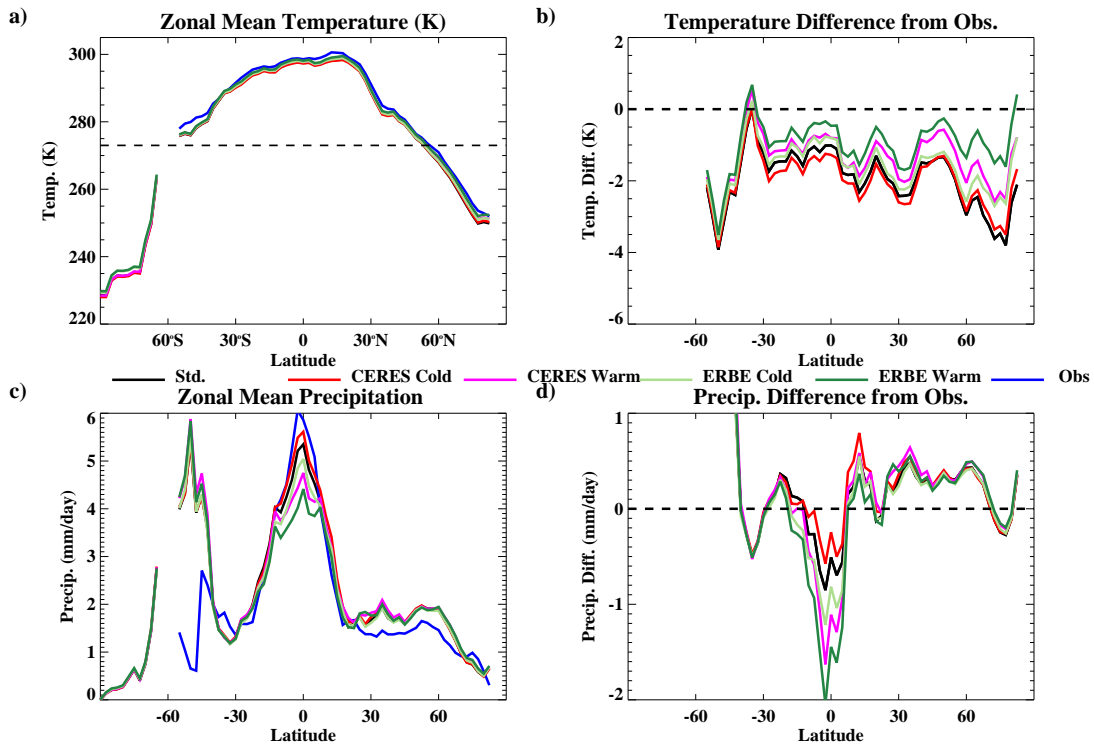


FIG. 9. (a): Zonal-mean 2000–2004 average of land-temperatures (K) from standard HadAM3 configuration (black), CRUTS3.1 observations (blue) and average of CERES cold group (red); CERES warm group (magenta); ERBE cold group (light green); ERBE warm group (dark green). (b): Difference from observations. Horizontal dashed line is zero difference from observations. c) & d) as a) & b) but for land-precipitation (mm/day).

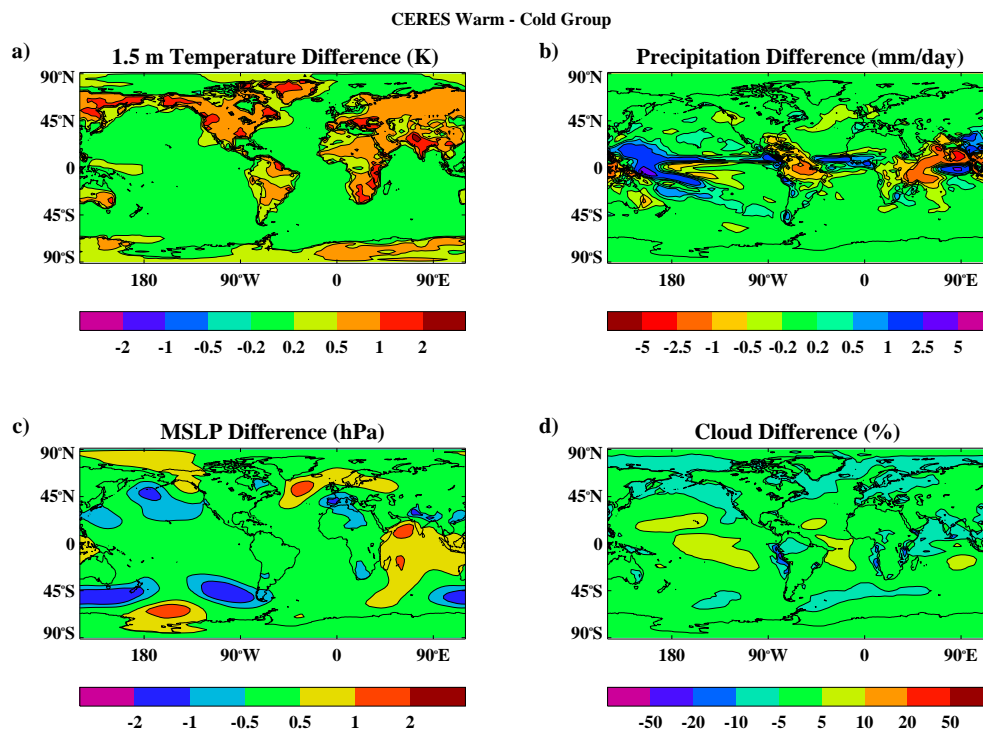


FIG. 10. Difference of 5-year (2000-2004) average between CERES warm and cold groups for temperature(a), precipitation (b), mean sea level pressure (c) and percentage cloud (d). Contours for plot a) are at ± 2 , ± 1 , ± 0.5 and ± 0.2 K. For plot b) they are at ± 5 , ± 2.5 , ± 1 , ± 0.5 and ± 0.2 mm/day. For plot c) they are at ± 2 , ± 1 and ± 0.5 hPa while in plot d) they are at ± 50 , ± 20 , ± 10 , ± 5 %.