Climate Dynamics (2006) 26: 145–165
DOI 10.1007/s00382-005-0067-7

K. D. Williams · M. A. Ringer · C. A. Senior
M. J. Webb · B. J. McAvaney · N. Andronova
S. Bony J.-L. Dufresne · S. Emori · R. Gudgel
T. Knutson · B. Li · K. Lo · I. Musat · J. Wegner
A. Slingo · J. F. B. Mitchell

Evaluation of a component of the cloud response to climate change in an intercomparison of climate models

Received: 1 April 2005 / Accepted: 5 August 2005 / Published online: 20 December 2005

Abstract Most of the uncertainty in the climate sensitivity of contemporary general circulation models (GCMs) is believed to be connected with differences in the simulated radiative feedback from clouds. Traditional methods of evaluating clouds in GCMs compare time–mean geographical cloud fields or aspects of present-day cloud variability, with observational data. In both cases a hypothetical assumption is made that the quantity evaluated is relevant for the mean climate change response. Nine GCMs (atmosphere models coupled to mixed-layer ocean models) from the CFMIP and CMIP model comparison projects are used in this study to demonstrate a common relationship between the mean cloud response to climate change and present-day variability. Although atmosphere–mixed-layer ocean models are used here, the results are found to be equally applicable to transient coupled model simulations. When changes in cloud radiative forcing (CRF) are composited by changes in vertical velocity and saturated lower tropospheric stability, a component of the local mean climate change response can be related to present-day variability in all of the GCMs. This suggests that the relationship is not model specific and might be relevant in the real world. In this case, evaluation within the proposed compositing framework is a direct evaluation of a component of the cloud response to climate change. None of the models studied are found to be clearly superior or deficient when evaluated, but a couple appear to perform well on several relevant metrics. Whilst some broad similarities can be identified between the 60°N–60°S mean change in CRF to increased CO_2 and that predicted from present-day variability, the two cannot be quantitatively constrained based on changes in vertical velocity and stability alone. Hence other processes also contribute to the global mean cloud response to climate change.

K. D. Williams (✉) · M. A. Ringer · C. A. Senior
M. J. Webb · J. F. B. Mitchell
Met Office, Hadley Centre for Climate Prediction and Research, FitzRoy Road, Exeter, EX1 3PB, UK
E-mail: keith.williams@metoffice.gov.uk
Tel.: +44-1392-886905
Fax: +44-1392-885681

A. Slingo · K. D. Williams
Environmental Systems Science Centre, University of Reading, Reading, UK

B. J. McAvaney
Bureau of Meteorology Research Centre, Sydney, Australia

N. Andronova · B. Li
Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA

S. Bony · J.-L. Dufresne · I. Musat
Institut Pierre Simon Laplace, Paris, France

S. Emori
National Institute for Environmental Studies, Ibaraki, Japan

R. Gudgel · T. Knutson
Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

K. Lo
Goddard Institute for Space Studies, New York, USA

J. Wegner
Max Planck Institute for Meteorology, Hamburg, Germany

1 Introduction

General circulation models (GCMs), which include representations of the complex physical processes in the climate system, are the primary tools used for climate change prediction. However, predictions of climate change for a given emissions scenario vary between different state-of-the-art GCMs (Cubasch et al. 2001). A leading order measure of the response of the climate system to an external forcing is the climate sensitivity, defined as the global mean surface temperature response due to a doubling of CO_2 once the climate system has reached an equilibrium state (e.g. Schlesinger and Mitchell 1987). Recent estimates using the instrumental

temperatures from the mid nineteenth century to present-day provide only a limited constraint on the climate sensitivity (Andronova and Schlesinger 2001; Forest et al. 2002; Gregory et al. 2002; Knutti et al. 2002). The likely range of climate sensitivity was first estimated by the National Academy of Sciences (1979) to be between 1.5 and 4.5 K. Based largely on the range of predictions from different GCMs, successive reports of the Intergovernmental Panel on Climate Change (IPCC) have not found sufficient evidence to change this estimate (Houghton et al. 1990, 1992, 1996, 2001), although some GCMs exist with a higher climate sensitivity (e.g. Murphy et al. 2004). Since climate change impacts are likely to be quite different for a climate sensitivity of 1.5 K versus 4.5 K (or higher), it is desirable that this uncertainty is reduced if policy makers are to have confidence in making potentially costly socio-economic decisions regarding climate change mitigation and adaption strategies.

Changes in clouds (amount, height and/or optical properties) have the ability to amplify or suppress the warming due to increased CO_2. This radiative feedback from clouds is believed to account for much of the range of climate sensitivity in contemporary GCMs (e.g. Cess et al. 1990; Senior and Mitchell 1993). Cloud radiative forcing (CRF), defined as the radiative effect of cloud on the radiation budget at the top of the atmosphere, provides a relatively simple measure of the radiative effect of cloud on the climate system (Charlock and Ramanathan 1985). Changes in CRF occurring in response to an external forcing may be used to compare cloud feedback, although processes not associated with cloud changes can also affect the change in CRF (e.g. Zhang et al. 1994; Soden et al. 2004).

Traditionally, evaluation of cloud in GCMs has been based on comparing climatological maps of cloud variables (typically CRF or total cloud amount) simulated by the GCM with observational data. More recently, process-based evaluation techniques have been developed which use compositing techniques in order to stratify the data (e.g. Bony et al. 1997; Ringer and Allan 2004; Tselioudis et al. 2000), with the aim of evaluating a GCM cloud simulation within a meteorological 'regime'. Whilst accurately simulating present-day climate is desirable, and essential for predicting regional climate change, it does not necessarily follow that a model which can provide a good present-day simulation of CRF will also accurately simulate the change in CRF in response to increased CO_2. As climate change continues, it may become possible to directly evaluate the climate change CRF response against observational data. However at present, the satellite observational record is not of sufficient length, and the changes observed to date are not of sufficient magnitude to detect and attribute a cloud change signal.

There have been many studies which have examined relationships between anomalies in cloud and other variables, either spatially or temporally, in the present-day climate (e.g. Hartmann and Michelsen 1993; Klein and Hartmann 1993; Norris and Leovy 1994). Some have also made a 'hypothetical' inference of the results to the possible mean cloud response to climate change (e.g. Norris and Weaver 2001; Hanson 1991; Ramanathan and Collins 1991), albeit with caveats since in many cases explicit demonstration of a causal mechanism is difficult. As an example, a negative correlation exists between low cloud and sea surface temperature in present-day climate variability, therefore it might be inferred that there will be a global reduction in low cloud under climate change. However, most GCM climate change simulations suggest an increase in low cloud in some regions, despite all areas showing a warming. This illustrates the importance of checking that a particular relationship found in present-day climate variability is applicable for the climate change problem, before insisting that all models must be able to reproduce it in order to have confidence in their climate change predictions. There have only been two studies which have attempted to 'directly' demonstrate that a particular form of evaluation is relevant for at least some aspects of the climate change cloud response, namely Bony et al. (2004) (hereafter B04) and Williams et al. (2003) (hereafter W03). Both of these studies aim to relate the change in cloud in a climate change simulation to processes associated with cloud variability in the present-day climate, which can then be evaluated. Both studies are limited to the tropical ocean regions and neither is able to provide a complete quantitative evaluation of the climate change cloud response for the region.

B04 composite monthly mean CRF from three GCM simulations and Earth Radiation Budget Experiment (ERBE) observations by 500 hPa vertical velocity (ω_{500}). By assuming that ω_{500} is a good measure of the large scale dynamics, compositing by this variable will isolate the 'dynamic' response (e.g. geographical shifts in cloud fields) and B04 refer to the change within each of these ω_{500} bins as the 'thermodynamic' response. B04 find that the net change in CRF due to changes in the dynamics is small, and most of the net cloud response to climate change is 'thermodynamic'. In addition, they find that most of the cloud response to climate change occurs in weakly subsiding regimes due to their high population. However, B04 acknowledge that evaluating the thermodynamic climate change cloud response is more challenging. Whilst they propose a possible methodology to investigate the thermodynamic response with idealised sensitivity experiments using cloud resolving and single column models, they do not pursue evaluation of the thermodynamic cloud response in the study.

W03 also use ω_{500} as a measure of the large scale dynamics, but they composite the change in cloud between control and $2\times CO_2$ slab model simulations by the change in ω_{500}, and present-day cloud variability by ω_{500} anomalies. In addition to using the change in ω_{500} as a measure of the large scale dynamics, W03 also composite by the change in sea surface temperature (SST) relative to the tropical mean warming in the climate change simulation and by the SST anomaly in the

present-day simulation. W03 composite both CRF and cloud amount diagnostics and find good agreement between the climate change cloud response and present-day simulation when composited by these variables. They find that the response of high-top cloud, particularly for thicker convective cloud, is mainly dependent on the change in ω_{500} whilst the low cloud response is mainly associated with the SST anomaly. Although the composited climate change response appears to be qualitatively similar to variability in the present-day simulation, especially in regions where the anomalies of ω_{500}/relative SST are largest, W03 note that most datapoints have only small changes in ω_{500}/relative SST. Hence a quantitative evaluation (not carried out by W03), must weight this region of weak changes in the ω_{500}-SST anomaly considerably higher than those regions with the larger cloud changes. In order to have confidence that the relationship between the composited climate change cloud response and present-day variability might operate in the real world, it should be shown to exist in several contemporary GCMs which contain considerable differences in model formulation (i.e. structural differences in parametrizations). This is a key requirement of the methodology for evaluation of cloud response to climate change proposed by W03 and will be the focus of this paper.

In this study, nine atmosphere–mixed-layer ocean models submitted to the Cloud Feedback Model Intercomparison Project (CFMIP) and to the Coupled Model Intercomparison Project (CMIP) IPCC 4th Assessment Report simulation database, are used in order to investigate a general relationship between the climate change CRF response and that simulated as part of present-day variability. The models and observational data are briefly described in the next section. The data from the models is composited as described in Sect. 3 in order to relate the mean cloud response to climate change to changes in present-day variability. The extent to which this relationship exists generally amongst the models will be explored in Sect. 4. It will be shown that the relationship can be seen in all of the models, hence a comparison of the composited present-day cloud variability with observational data will form an evaluation of cloud processes which have been *demonstrated* to be relevant to (at least some of) the cloud response to climate change. This evaluation will be pursued in Sect. 5. A summary and discussion are in Sect. 6.

2 Models and observational data

The experimental design of CFMIP is to initially repeat the atmosphere-only inverse climate change experiments of Cess et al. (1990) with SST uniformly increased and reduced by 2 K. These experiments are being repeated for backwards compatibility, however such an experimental design does not permit the SST to respond to local changes in surface fluxes and hence, feedback on the atmosphere. Therefore, CFMIP also requests parallel equilibrium control and $2\times CO_2$ atmosphere–mixed-layer ocean (slab model) experiments in order to provide an intercomparison of climate change simulations with a more realistic pattern of surface temperature response. The surface temperature of the ocean is maintained close to climatological values, in the absence of ocean currents, by use of a monthly varying heat flux. This is calculated in a calibration experiment (performed for each model prior to the main control and $2\times CO_2$ experiments) in which the SSTs are reset to climatological values at each timestep. Data from the slab model control and $2\times CO_2$ simulations are presented here.

The diagnostic requirements of CFMIP are specifically designed to allow a detailed intercomparison of cloud processes between GCMs; in particular, output from the International Satellite Cloud Climatology Project (ISCCP) simulator (Klein and Jakob1999; Webb et al. 2001; http://gcss-dime.gis.nasa.gov/simulator.html) is required. At this time, data are still being collected and high temporal resolution data from the ISCCP simulator is not currently available from many models. Analysis of such data will be pursued in future studies. In the present study, the more traditional diagnostics of monthly mean CRF from the CFMIP slab model experiments are used. Since these diagnostics are also requested for GCMs submitting to CMIP (which requests the same experimental design) the number of models used in this study can be increased (by 2) by also including slab models submitted to CMIP but not CFMIP.

The models used in this study are listed in Table 1 and together form an ensemble of GCMs with significant structural differences. The models cover almost a factor of three in horizontal resolution and a factor of two in vertical resolution. A mix of prognostic and diagnostic cloud liquid water parameterisations are present in the ensemble (together with many other parameterisation differences). Although there are three Hadley Centre models included, HadGSM1 and HadSM3 (slab model versions of HadGEM1 and HadCM3) have considerable structural differences including a different dynamic core, resolution and many different or revised physical parameterisations. Hence they may be considered as being as structurally different as models from two different centres. HadSM4 is a development model and contains aspects of both HadSM3 and HadGSM1.

Since this paper is more concerned with the methodology of demonstrating the relevance of the evaluation rather than the final evaluation, the models in Table 1 are randomly assigned letters A-I, and will simply be referred to by their letter for the remainder of the paper. (Note: the order of the letters does not correspond to the order of models in Table 1).

For the evaluation process, monthly mean CRF data from the Earth Radiation Budget Experiment (ERBE) (Barkstrom et al. 1990; Harrison et al. 1990) are used. The ERBE dataset is available on a 2.5° by 2.5° grid for the period November 1984–February 1990 (the period March 1985–February 1990 being used in this analysis). Since global observations of the compositing variables to be

Table 1 List of models used in this study (all are atmosphere–mixed-layer ocean configurations of the model). Horizontal resolution is pre-fixed by 'T' for the truncation of spectral models and 'N' for half the number of east-west points for grid-point models (this notation permits approximate comparison of the two model structures)

Model	Resolution	Nature of cloud scheme	Main references
ECHAM5	T63 L32	Prognostic	Roeckner et al. (2003)
GFDL AM2	N72 L24	Prognostic	GFDL GAMDT (2004)
GISS ER	N36 L20	Prognostic	Schmidt et al. (2005)
HadSM3	N48 L19	Diagnostic	Pope et al. (2000); Williams et al. (2001)
HadSM4	N48 L38	Diagnostic	Webb et al. (2001)
HadGSM1	N96 L38	Diagnostic	Martin et al. (2005); Johns et al. (2005)
IPSL CM4	N48 L19	Diagnostic	
MIROC	T42 L20	Diagnostic	K-1 model developers (2004)
UIUC	N36 L24	Prognostic	Yang et al. (2000)

The number of atmosphere levels is prefixed by 'L'. Also shown is whether the model uses prognostic or diagnostic cloud liquid water in its large-scale cloud scheme

introduced in Sect. 3 do not exist, model re-analysis data must be used to stratify the ERBE data. Two re-analysis datasets are utilised and the results compared in order to investigate possible errors in the choice of re-analysis model for the data compositing. The re-analyses are the European Centre for Medium Range Weather Forecasts (ECMWF) 40 year re-analysis (ERA-40) (http://www.ecmwf.int/products/data/archive/descriptions/e4) and the National Center for Atmospheric Research (NCAR)/National Centers for Environmental Prediction (NCEP) 40 year re-analysis (Kalnay et al. 1996).

3 Cloud compositing methodology

Several studies (e.g. Bony et al. 1997; Norris and Weaver 2001, W03) find an association between changes in cloud, particularly high-top cloud, and changes in vertical velocity, over both tropical and mid-latitude oceans. Whilst B04 find little net contribution to the climate change cloud response from changes in the dynamic regimes, the local change in dynamics is found to be one of the principle processes associated with the local response of high cloud in the GCMs. For example, the geographical distribution of the change in high-top cloud in response to a doubling of CO_2 and the change in ω_{500} show close agreement (e.g. Fig. 1a, b. Note that the colours used in the figures throughout this study are blue for more cloud/stronger CRF components/processes increasing cloud amount, and red for less cloud/weaker CRF components/processes reducing cloud amount).

Recently the lead-author has repeated the analysis of W03, extended to include weighting by the population of the bins (not shown). The analysis has been applied to several of the models submitted to CFMIP. Whilst the change in high cloud has generally been found to be associated with changes in vertical velocity, the relationship between low cloud and the local temperature response, relative to the mean response, appears to apply less well in other models and over other geographical regions. Hence, the relationship found by W03 is thought to be specific to the Hadley Centre model and an alternative variable which is more generally associated with low cloud is required.

In an observational study, Klein and Hartmann (1993) conclude that the variability of low cloud in the seasonal cycle is associated more with variations in lower tropospheric dry static stability, diagnosed in terms of the monthly mean difference between the potential temperature at 700 hPa and the surface, than with SST. They argue that this measure of stability can be used as a simple indicator of the frequency and intensity of lower tropospheric inversions. Weaver (1999) also identifies the importance of a lower tropospheric inversion for the formation of low cloud but in addition, presents the stability with respect to moist processes. The geographical pattern and sign of the change in low cloud in response to a doubling of CO_2 is now compared to mean change in three measures of lower tropospheric stability (Fig. 1c–f). The results are illustrated with one model (Model A), however several of the models have been investigated using these diagnostics and the results are qualitatively similar in each case.

Following Klein and Hartmann (1993), the difference between the potential temperature at 700 hPa and at the surface is used as a measure of dry stability (θ': the prime representing a difference between the two levels; Fig. 1d). These levels were chosen in order to test for the presence of a lower tropospheric inversion capping the boundary layer. 700 hPa is usually above the boundary layer, hence might be considered representative of the lower free troposphere. Weaver (1999) reports little sensitivity in choosing 700 or 500 hpa to represent the free troposphere. By subtracting the surface value, the environmental stability for an air parcel rising from the surface may be crudely estimated. Stable conditions suggest a strong capping inversion, hence rising air parcels will be capped and any cloud formed is likely to be low-topped stratocumulus. Unstable conditions are likely to generate less low cloud, either because shallow convection is not capped so strongly (and hence is less likely to spread into stratocumulus) or, in cases of strong instability throughout the troposphere, deep convective cloud with higher cloud tops may form which obscures the TOA radiative effect of any low cloud. An

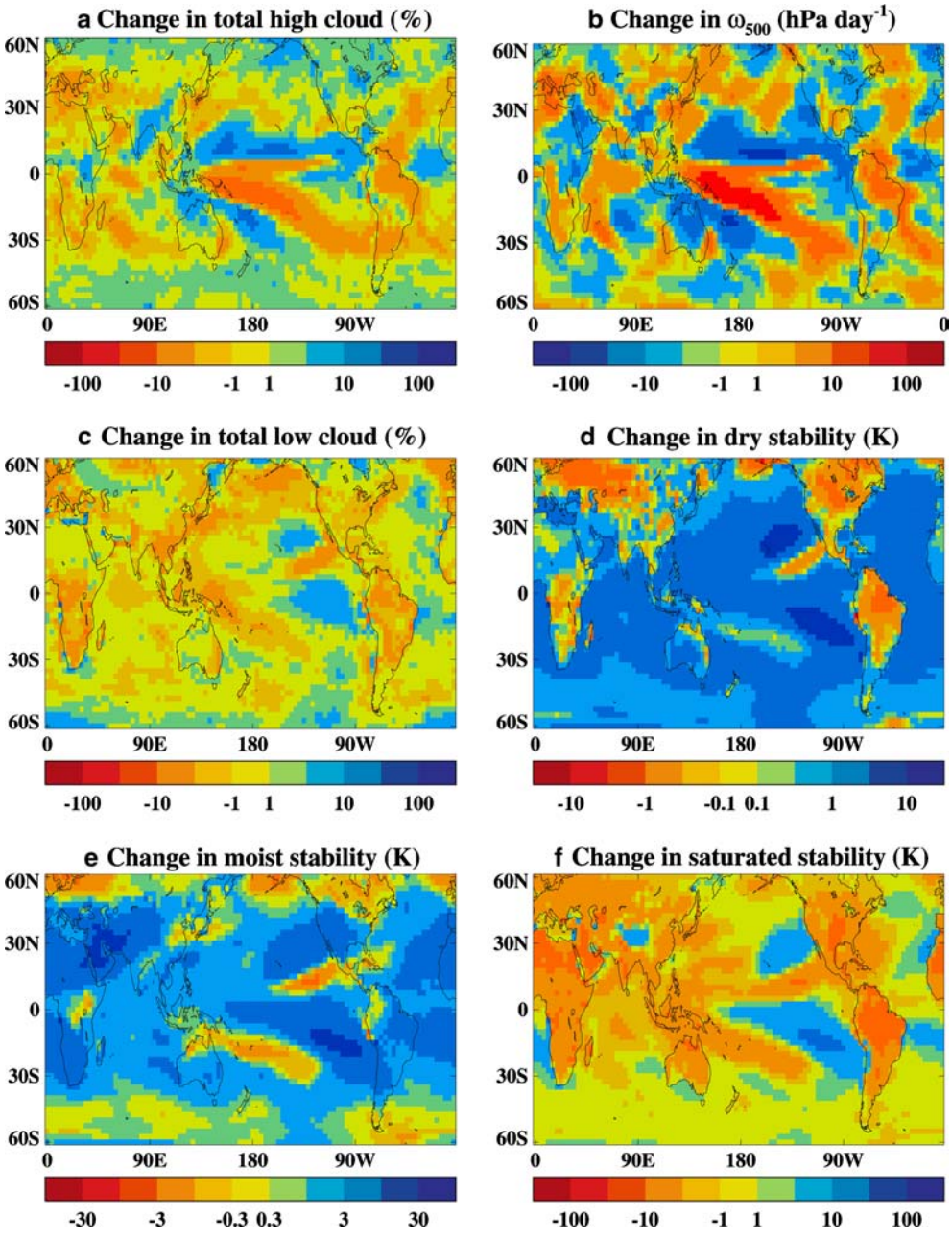

Fig. 1 Model A, 2×CO_2 simulation minus control. **a** Change in total high cloud amount (%) (as output from the ISCCP simulator). **b** Change in vertical velocity (ω_{500} hpa day^{-1}). **c** Change in total low cloud (%) (as output from the ISCCP simulator). **d** Change in dry lower tropospheric stability (θ', K). **e** Change in moist lower tropospheric stability (θ'_e, K). **f** Change in saturated lower tropospheric stability (θ'_{es}, K)

alternative measure of stability, taking moist processes into account, is the difference in the saturated equivalent potential temperature at 700 hpa and the equivalent potential temperature at the surface (θ'_e; Fig. 1e). Finally, the difference in saturated equivalent potential temperature between the two levels is considered as a measure of the stability to a saturated air parcel (θ'_{es}; Fig. 1f). Whilst the pattern of response to a doubling of CO_2 is broadly similar for each of these diagnostics, it can be seen that there is an increase in θ' and θ'_e in most regions (i.e. there is an increase in these measures of stability). In the regions of increased θ' and θ'_e, both increases and decreases in low cloud can be seen. The overall increase in dry, and to a lesser extent moist, stability is associated with the larger warming of the free troposphere than the surface. This is a common climate change signal in GCMs (e.g. Colman 2003). The pattern and sign of the change in θ'_{es} can be seen to be most closely related to the change in low cloud in Model A (Fig. 1c, f). The change in θ'_{es} tends to be towards more unstable conditions in many locations (compared with the general increase in stability for θ' and θ'_e) due to more latent heat release at lower levels i.e. for the same change in θ at two different levels, the change in θ'_{es} is larger nearer to the surface (illustrated by saturated adiabats narrowing at low levels on a standard 'tephigram'). The area-weighted spatial correlation for Fig. 1a, b is 0.85 and for Fig. 1c, f is 0.52[1], suggesting a

[1]The correlation for Fig. 1c, d is 0.41 and for Fig. 1c, e is 0.24

Table 2 Correlation between spatio-temporal monthly mean total low cloud amount in a grid-box obtained from ISCCP and three measures of monthly mean lower tropospheric stability (see main text for definitions)

	ERA-40	NCAR/NCEP
θ'	0.36	0.32
θ'_e	0.37	0.36
θ'_{es}	0.47	0.44

Correlations using ERA-40 and NCAR/NCEP re-analyses for the lower tropospheric stability are shown

relationship between cloud response and these two variables.

In order to test which measure of lower tropospheric stability may be most closely related to changes in low cloud in the real world, 5-years worth of monthly mean θ', θ'_e and θ'_{es} from the re-analyses are correlated against total low cloud amount from ISCCP for the region 60°S-60°N (which provides around 400,000 spatio-temporal data-points) (Table 2). Statistically significant correlations exist for all of the measures of lower tropospheric stability (although the correlations are not very high), however the correlation of low cloud amount with θ'_{es} is somewhat greater than for the other measures of stability when using either re-analysis. It is possibly surprising that the correlations are higher for θ'_{es} rather than θ'_e, which might be expected to be most closely related to the processes involved in lifting an unsaturated air parcel from the surface. Although it is possible that this is due to using monthly rather than daily data, a compositing method similar to B04 has been carried out using these various measures of stability as compositing variables on both daily and monthly data for two GCMs. Little difference is found when compositing by θ'_e or θ'_{es} between using monthly and daily data (other than in the range of θ'_e/θ'_{es} sampled). A possible physical explanation for the higher correlation with θ'_{es} is the following. The temperature profile of the lower free troposphere, particularly in the tropics, typically closely follows a saturated adiabat. The measure of a capping inversion should, therefore, be with respect to a saturated adiabat rather than a dry adiabat (hence θ'_{es} rather than θ'). In general, an increase in stability over a region where low cloud can form will lead to a stronger inversion and hence more low cloud, however θ'_e can also become more stable if the surface humidity decreases. In this case the boundary layer will be dryer and less low cloud form. Since θ'_e as a measure of stability does not discriminate between these two effects, the combination of the two opposing processes may lead to the lower correlation.

W03 composite the cloud response to climate change at each grid-point by the change in 500 hpa vertical velocity (ω_{500}) and the relative change in SST. They use monthly mean data so that each month from the $2\times CO_2$ simulation is differenced from the same month of the control. Such a methodology will include aspects of variability within the annual cycle which might dominate the association between the climate change response and present-day variability, and possibly lead to an artificially high correlation between the two composites. The long-term mean cloud response to climate change is the target for this study, hence the 20-year multi-annual mean climate change response of CRF for each slab model is composited i.e. one multi-annual mean difference field between the $2\times CO_2$ and control is composited for each variable. Each GCM grid-point is assigned to a bin based on the 20-year mean change in vertical velocity ($\Delta\ \omega_{500}$) and lower tropospheric stability ($\Delta\ \theta'_{es}$) at that grid-point (e.g. Fig. 2a). Hence the number of data values is the same as the number of grid-points in the model over the region studied. (The compositing has been repeated spatio-temporally for some of the GCMs using changes in monthly data, as done by W03, with little difference in the composite mean results other than in the range of $\Delta\ \omega_{500}$-$\Delta\ \theta'_{es}$ sampled). The mean θ'_{es} fields are calculated monthly from temperature and pressure fields and averaged to account for any non-linearities in determining stability between the monthly and multi-year time-scales (ideally the stability should be calculated at even higher temporal resolution, however these diagnostics were not available from all of the models). Each grid-point is assigned to a 2.5 hpa day$^{-1}\Delta\ \omega_{500}$ by 2.5 K$\Delta\ \theta'_{es}$ bin. Points which have changes in ω_{500} of greater than 51.25 hPa day^{-1} or less than -51.25 hpa day^{-1}, or changes in θ'_{es} of greater than 21.25 K or less than -41.25 K are included in the nearest bin of the composite (i.e. on the perimeter of the composite space). The range of $\Delta\ \theta'_{es}$ bins is not centred on zero as it has been found that all models studied show a slight shift towards more unstable θ'_{es} regimes under climate change. Hence, the compositing is centred on a $\Delta\ \theta'_{es}$ of -10 K. The population of each bin is area-weighted and normalised so that the population integrated across all the bins is one and the value in each bin is the fraction of the total area analysed. The mean change in shortwave, longwave and net CRF (ΔSCRF, ΔLCRF, ΔNCRF) at each GCM grid-point is assigned to the appropriate bin and the area weighted mean change in CRF for each bin calculated (e.g. Fig. 2b–d).

All land and ocean points in the tropics and mid-latitudes (60°S–60°N) are included in the compositing, except where the monthly mean surface pressure is below 800 hpa (since in these cases the difference in θ_{es} from 700 hpa is unlikely to be a good measure of lower tropospheric stability). Polar regions are excluded as ERBE CRF data are not believed to be reliable at high latitudes, restricting evaluation in this region. The compositing has been repeated with land and ocean regions separately and with the tropics and mid-latitudes separated. The resulting composites are similar to those presented here for the whole region, although the areas of the composite space populated differ in each case (e.g. the largest changes in stability are mainly over land and the largest changes in vertical velocity are mainly in the tropics).

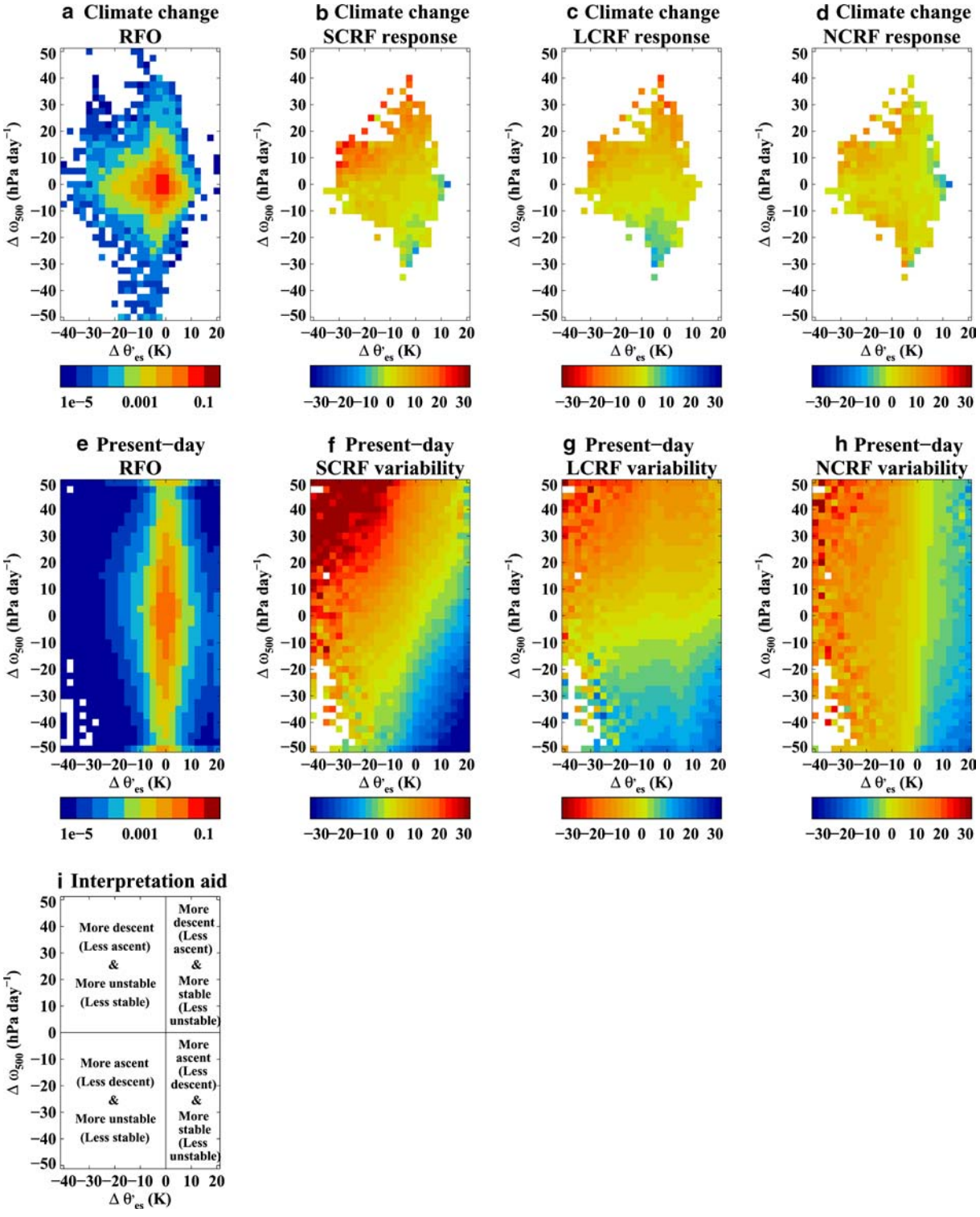

Fig. 2 **a** GCM-mean relative frequency of occurrence (RFO) of $\Delta\ \omega_{500}/\Delta\ \theta'_{es}$ bins for the climate change simulations ($2\times CO_2$ mean minus control mean) i.e. composites were produced for each model and the composites averaged. **b–d** GCM-mean change in SCRF, LCRF and NCRF in response to doubling CO_2 i.e. the geographic pattern of response is composited by $\Delta\ \omega_{500}/\Delta\ \theta'_{es}$ for each model and averaged. **e** GCM-mean RFO of spatio-temporal monthly-mean anomalies in ω_{500} and θ'_{es} from the control simulations. **f–h** GCM-mean spatio-temporal monthly-mean anomalies of SCRF, LCRF and NCRF from the model control simulations composited by $\Delta\ \omega_{500}/\Delta\ \theta'_{es}$. **i** An interpretation aid for the four quadrants of the composite space (for use with Figs. 2, 3, 4, 5, 6, 7, 8, 9)

Table 3 For each model: the RFO-mean change in θ'_{es}; unweighted correlations between the climate change and present-day variability ΔCRF composites (e.g. correlation between Fig. 2b, f); correlations between the climate change and present-day variability ΔCRF composites, weighted by the climate change RFO (e.g. correlation between Fig. 4a, d)

Model	Net$\Delta\ \theta'_{es}$ (K)	Correlations			
		Unweighted		Weighted	
		ΔSCRF	ΔLCRF	ΔSCRF	ΔLCRF
Model A	−2.7	0.75 ± 0.03	0.96 ± 0.01	0.79 ± 0.04	0.91 ± 0.02
Model B	−3.1	0.73 ± 0.03	0.86 ± 0.02	0.86 ± 0.04	0.46 ± 0.08
Model C	−3.1	0.81 ± 0.03	0.88 ± 0.02	0.60 ± 0.03	0.60 ± 0.05
Model D	−3.2	0.77 ± 0.03	0.69 ± 0.04	0.66 ± 0.05	0.84 ± 0.05
Model E	−3.8	0.59 ± 0.03	0.65 ± 0.02	0.57 ± 0.04	0.52 ± 0.04
Model F	−4.0	0.40 ± 0.04	0.77 ± 0.03	0.68 ± 0.07	0.25 ± 0.09
Model G	−2.9	0.65 ± 0.04	0.57 ± 0.04	0.85 ± 0.06	0.86 ± 0.03
Model H	−1.7	0.54 ± 0.04	0.90 ± 0.02	0.53 ± 0.09	0.87 ± 0.04
Model I	−2.6	0.54 ± 0.03	0.60 ± 0.02	0.70 ± 0.09	0.52 ± 0.07
GCM-mean	−2.7	0.80 ± 0.01	0.89 ± 0.01	0.88 ± 0.01	0.72 ± 0.03

Note that the 'GCM-mean' is the correlation between the mean climate change and present day composites, rather than the mean of the correlations

The composited mean climate change response is compared with spatio-temporal variability within the control simulation. Monthly anomalies in CRF are differenced from the 20 year mean for that month in the slab model control simulation, so that the number of data values used is equal to the number of GCM grid-points in the region multiplied by 240 months. The CRF anomalies are composited by anomalies in ω_{500} and θ'_{es} (e.g. Fig. 2e–h). Most of the control simulations use pre-industrial greenhouse gas emissions, however some use a CO_2 concentration more typical of the late twentieth century. Compositing anomalies in CRF means that similar results are obtained regardless of the period chosen. The control variability compositing was repeated for five of the models with parallel atmosphere-only simulations forced with observed SSTs and late twentieth century CO_2 concentrations. The resulting variability composites were correlated with those from the pre-industrial slab control simulations and for the region of composite space populated in both cases, the correlations of CRF anomalies was greater than 0.9 for all models. Note that the control variability will be referred to as 'present-day' variability in this study, even though the forcing is actually pre-industrial in most cases.

4 Relating cloud response to climate change to present-day variability

The GCM-mean change in CRF in response to doubling CO_2 and the present-day CRF variability, composited by $\Delta\ \omega_{500}-\Delta\ \theta'_{es}$, is shown in Fig. 2 i.e. composites were produced for each model and the composite average is shown. 'Postage stamp' composites for the individual models are in Fig. 3. Since each of the models populate different regions of $\Delta\ \omega_{500}-\Delta\ \theta'_{es}$ space, bins towards the edges of the composite may only be the mean of a few models. The GCM-mean ΔCRF composites are only shown for those bins which are populated by two or more models. The relative frequency of occurrence (RFO) of the bins for the climate change composite occupy both positive and negative values of $\Delta\ \omega_{500}$ and $\Delta\ \theta'_{es}$, with the centre of the RFO close to zero (note the logarithmic scale) (Fig. 2a). However it can be seen that there is also a large RFO in bins with reduced stability (up to −20K). Unlike the relative change in SST used by W03, there is no theoretical constraint on the RFO distribution in $\Delta\ \theta'_{es}$ space i.e. there is no prior requirement for the mean of the distribution to be near to zero for the climate change composite. The GCM-mean net change in stability is −2.7 K; the net change for each of the models is given in the left hand column of Table 3. All of the GCMs studied simulate a net reduction in saturated lower tropospheric stability in response to increased CO_2. The change in SCRF varies with both compositing variables: SCRF becoming stronger (more negative) where there is increased ascent/reduced descent due to more deep convective cloud being produced, and where the environment becomes more stable due to shallow convection being capped and spreading into stratocumulus. The change in LCRF can be seen to be mainly associated with the change in vertical velocity, with increased LCRF where there is increased ascent/reduced descent due to the increased amount of high cloud (as found by W03). The change in NCRF appears to mainly vary with the change in stability (Fig. 2d). This suggests that changes in SCRF and LCRF associated with $\Delta\ \omega_{500}$ are of similar magnitude and tend to cancel. The observed balance between SCRF and LCRF in deep convective regions has been noted in several studies (e.g. Hartmann et al. 2001), however it remains an open question as to whether the change in each component under climate change is likely to be similar. The composites from the GCMs used here suggest that in response to changes in vertical velocity, most models simulate changes in SCRF and LCRF of a similar magnitude.

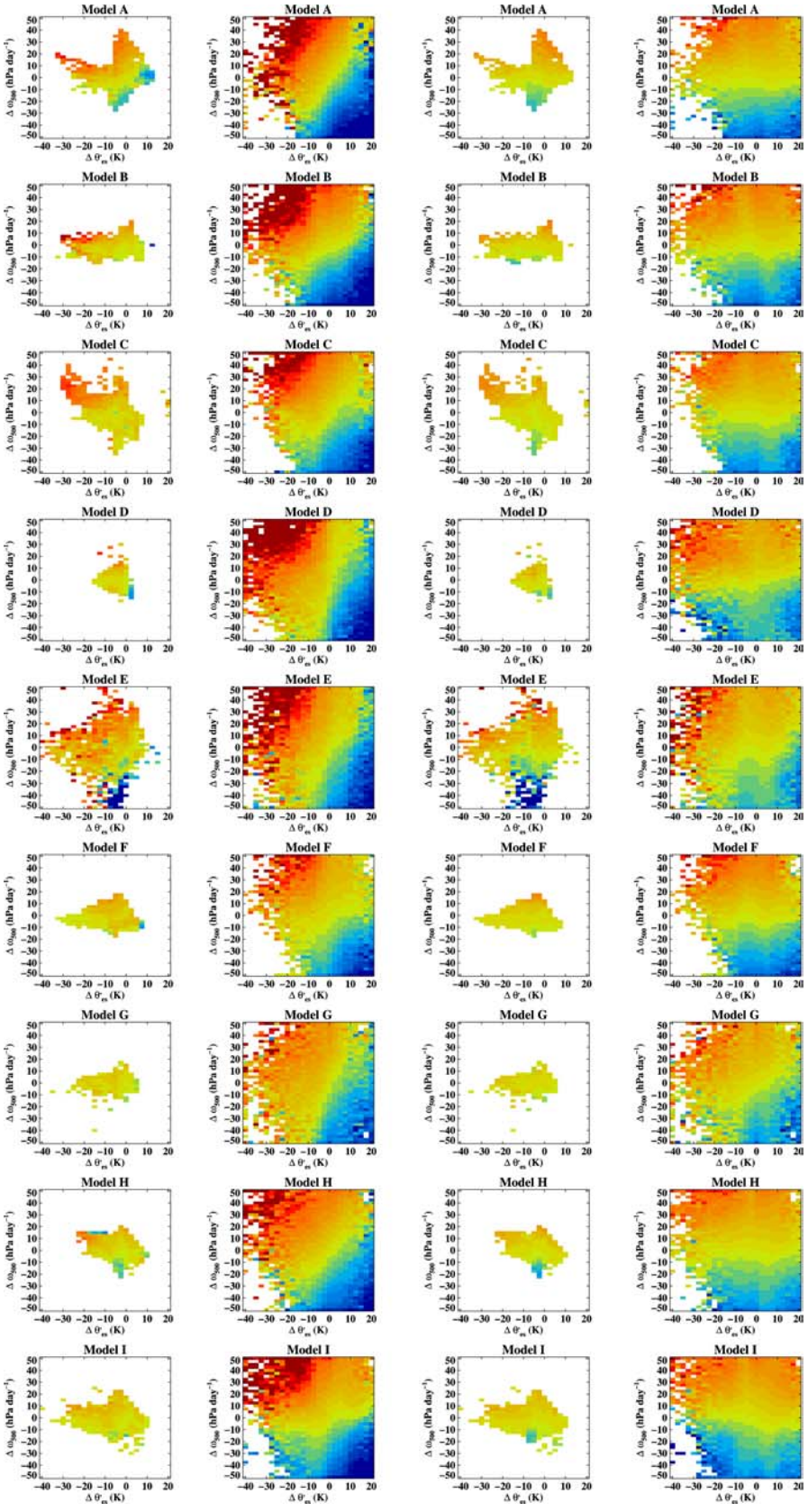

Fig. 3 For each model, the climate change and present-day variability composited $\Delta SCRF$ (*left columns*) and $\Delta LCRF$ (*right columns*) i.e. for each model from left to right, plots similar to Fig. 2b, c, f, g

Monthly variability in ω_{500} and θ'_{es} generally samples a larger region of the compositing space than the multi-annual mean climate change simulation (c.f. Fig. 2a, e). By construction, the present-day variability will be centred on zero. The changes in ω_{500} which occur as a mean response to doubling CO_2 can be seen to be well

Table 4 WMC for each model (e.g. sum across the bins in Fig. 4a, d, etc.) for the climate change and present-day variability composites

Model	WMC (Wm^{-2})			
	Climate change		Variability	
	ΔSCRF	ΔLCRF	ΔSCRF	ΔLCRF
Model A	1.7 ± 0.3	-0.4 ± 0.1	2.7 ± 0.4	-0.2 ± 0.1
Model B	2.2 ± 0.2	-1.3 ± 0.2	1.8 ± 0.2	0.1 ± 0.1
Model C	2.4 ± 0.2	-0.8 ± 0.1	2.1 ± 0.4	0.1 ± 0.1
Model D	0.9 ± 0.2	-1.5 ± 0.1	3.4 ± 0.3	-0.8 ± 0.2
Model E	4.5 ± 0.2	-1.7 ± 0.1	2.9 ± 0.3	-0.6 ± 0.2
Model F	3.3 ± 0.2	-1.3 ± 0.1	1.5 ± 0.3	0.2 ± 0.2
Model G	1.2 ± 0.1	-1.1 ± 0.1	2.4 ± 0.3	-1.0 ± 0.2
Model H	0.2 ± 0.1	-0.2 ± 0.1	1.4 ± 0.3	0.0 ± 0.2
Model I	1.1 ± 0.3	-1.1 ± 0.1	1.3 ± 0.3	0.3 ± 0.2
GCM-mean	1.8 ± 0.1	-1.0 ± 0.0	1.9 ± 0.1	-0.1 ± 0.1

sampled in present-day variability, however large changes in stability are sampled less well. In a couple of models there are a small number of grid-points where the reduction in θ'_{es} in the mean climate change field is outside that simulated in present-day variability. However for the GCMs in question, these points account for less than 3% of the net change in NCRF, so the lack of a present-day analogue for these extreme changes only imposes a limited restriction on the analysis. For the bins occupied by both composites, the anomalies in CRF due to present-day variability can be seen to be reasonably similar to the mean CRF changes under climate change (c.f. Fig. 2b–d, f–h). This similarity exists in each of the models with correlations between the two being mostly in excess of 0.5 (Fig. 3; Table 3 unweighted correlations). Uncertainty estimates have been added to each of the correlations in Table 3. These are based on compositing the climate change difference for each year of the 20 separately, similarly each of 20 years of the present-day simulation is composited separately and the correlations determined for each year. The uncertainty is calculated as two standard deviations of the inter-annual variability in the correlation divided by the square-root of the number of years. This provides approximately a 5% significance assuming a normal distribution and that each year is an independent sample. It can be seen that all of the models have a significant positive correlation for both CRF components.

The correlations between the climate change and present-day variability composites may be influenced by the larger CRF changes towards the edges of the composite space. These bins have a smaller RFO and hence, may be less important for the net change in CRF. For a quantitative comparison of the climate change and present-day variability composites, the ΔCRF in each bin should be multiplied by the normalised population of that bin (the RFO) from the climate change composite. By weighting both the climate change and present-day variability composites by the climate change RFO, the two composites can be directly compared (Fig. 4). Weighting by the population of the bins indicates that those bins closest to the centre of the composite are most important for the net cloud response i.e. the large number of points with small changes in ω_{500}, θ'_{es} (and CRF) provide a larger contribution to change in CRF than the few bins with large CRF changes towards the edge of the composite. As with the unweighted case, the climate change and variability population weighted composites appear to be very similar. For the GCM-mean, the correlation between the two for the change in SCRF and LCRF are 0.88 and 0.72, respectively. The correlations for the individual models are shown in Table 3. Whilst not always as high as for the ensemble mean, the correlations for all of the models are reasonable with the vast majority being greater than 0.5. The correlations are considerably higher than was obtained for composites with changes in other variables being used on the x-axis instead of $\Delta\ \theta'^{2}_{es}$[2]. Even the lowest correlation (0.25 for the change in LCRF in model F) has some qualitative similarity between the climate change and present-day variability composite (Fig. 5a, b) with LCRF reducing with reduced ascent/more descent and increasing in bins with the largest increases in ascent/reduced descent (this is also shown by the higher correlations for the unweighted composite). However, in the climate change response, the bins with near zero $\Delta\ \omega_{500}$ show a weak reduction in LCRF and the high population of these bins provides a significant contribution to the net change in LCRF. This suggests that factors other than the change in ω_{500} are affecting the high cloud response in this model (possibly affecting changes in cirrus). In contrast with model F, the change in LCRF in model A is almost completely determined by the change in vertical velocity, with little evidence for other processes affecting the climate change response (Fig. 5c, d).

In general, the weighted and unweighted correlations are of a similar size, hence the high unweighted correlations are not simply due to the large CRF changes in bins with a low RFO. The fact that the weighted climate change and present-day variability composites are significantly correlated demonstrates a mechanism which

[2]Other variables tested were absolute surface temperature, surface temperature relative to the local warming, different measures of stability and near surface relative humidity

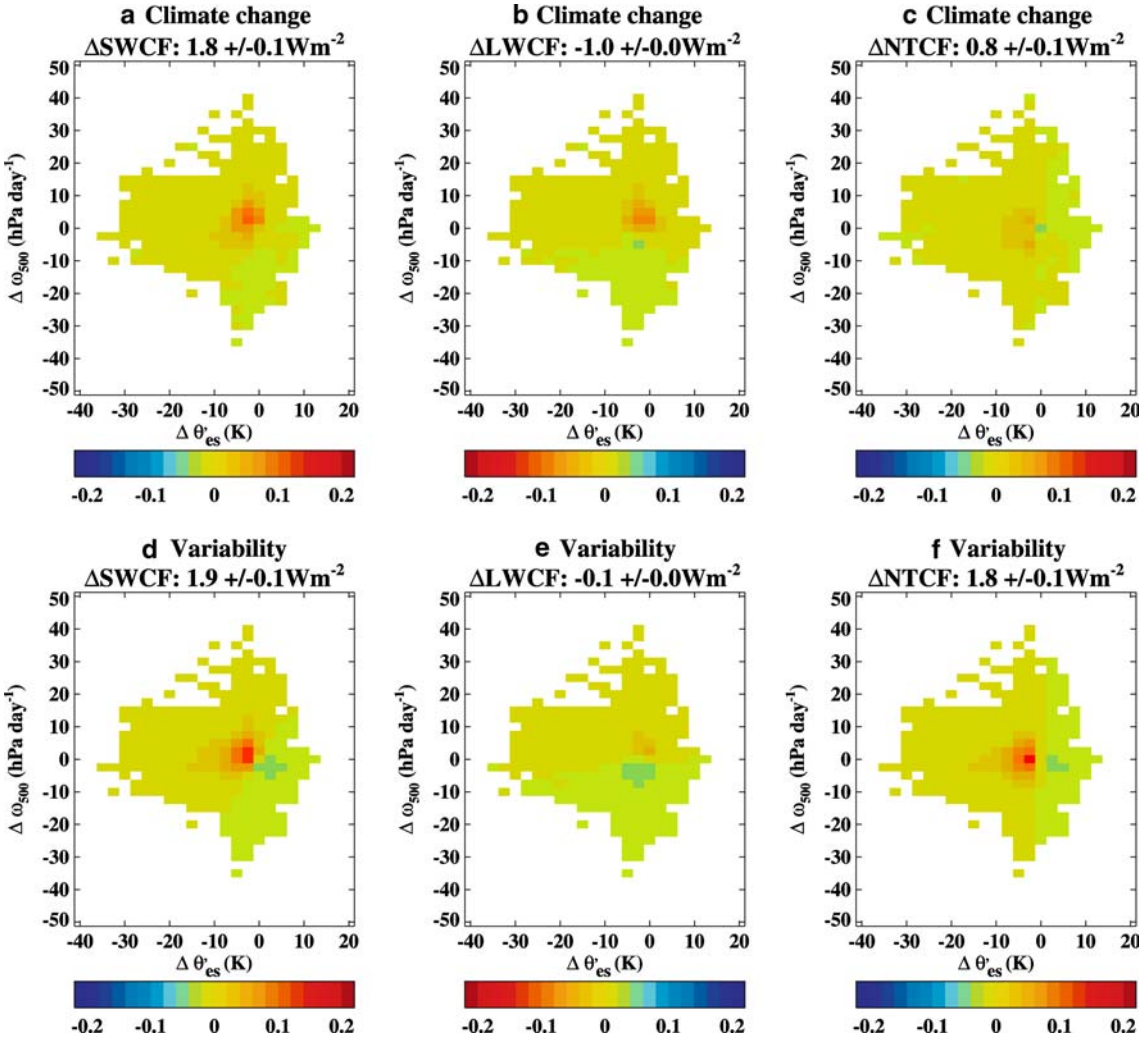

Fig. 4 As Fig. 2b–d, f–h, however both the mean climate change and present-day variability plots have been weighted by the GCM-mean RFO from the climate change simulation (Fig. 2a). The composites have been masked so that only bins with a non-zero RFO in both the climate change and present-day variability composites are shown. The WMCs are shown at the top of each figure and are approximately equal to the tropical and mid-latitude mean CRF response (this would be exact with infinitely small bins and if points with surface pressures below 800 hPa were included). The uncertainty estimates on each figure are estimates of the likely error due to variability in the composites

exists in both current climate variability and in the mean response to climate change. Changes in ω_{500} and θ'_{es} therefore determine at least a component of (and in models with a high correlation, most of) the climate change cloud response at a grid-point.

Integrating the RFO weighted changes in CRF from all the bins in the composite yields the weighted mean of the composite (WMC). For the climate change composite this is equal to the net change in CRF in response to doubling CO_2 over the region 60°N–60°S, hence if uncertainty in the WMC can be reduced then uncertainty in climate sensitivity may also be reduced. The climate change WMC is shown at the top of each panel in Fig. 4a–c. The WMC from the present-day variability CRF anomalies, weighted by the climate change population, is a prediction of the change in CRF due to a doubling of CO_2 based on the model's present-day CRF variability (after applying the model's climate change $\omega_{500}/\theta'_{es}$ response) (Fig. 4d–f). The increase in LCRF in regions of increased ascent/reduced descent tends to cancel with the reduction in LCRF where there is reduced ascent/increased descent. Therefore most of the local changes in LCRF are probably associated with shifts in high cloud which cancel globally, rather than, for example, regions of ascent becoming more concentrated. This is consistent with B04 who find little net change in CRF associated with changes in the large scale dynamics. As noted earlier, SCRF varies with both compositing variables. The overall reduction in stability results in a weakening of SCRF due to the more unstable regimes dominating the increase in SCRF at those points which have become more stable. The greater dependence of NCRF on stability rather than ω_{500} can be seen, with the shift to more unstable

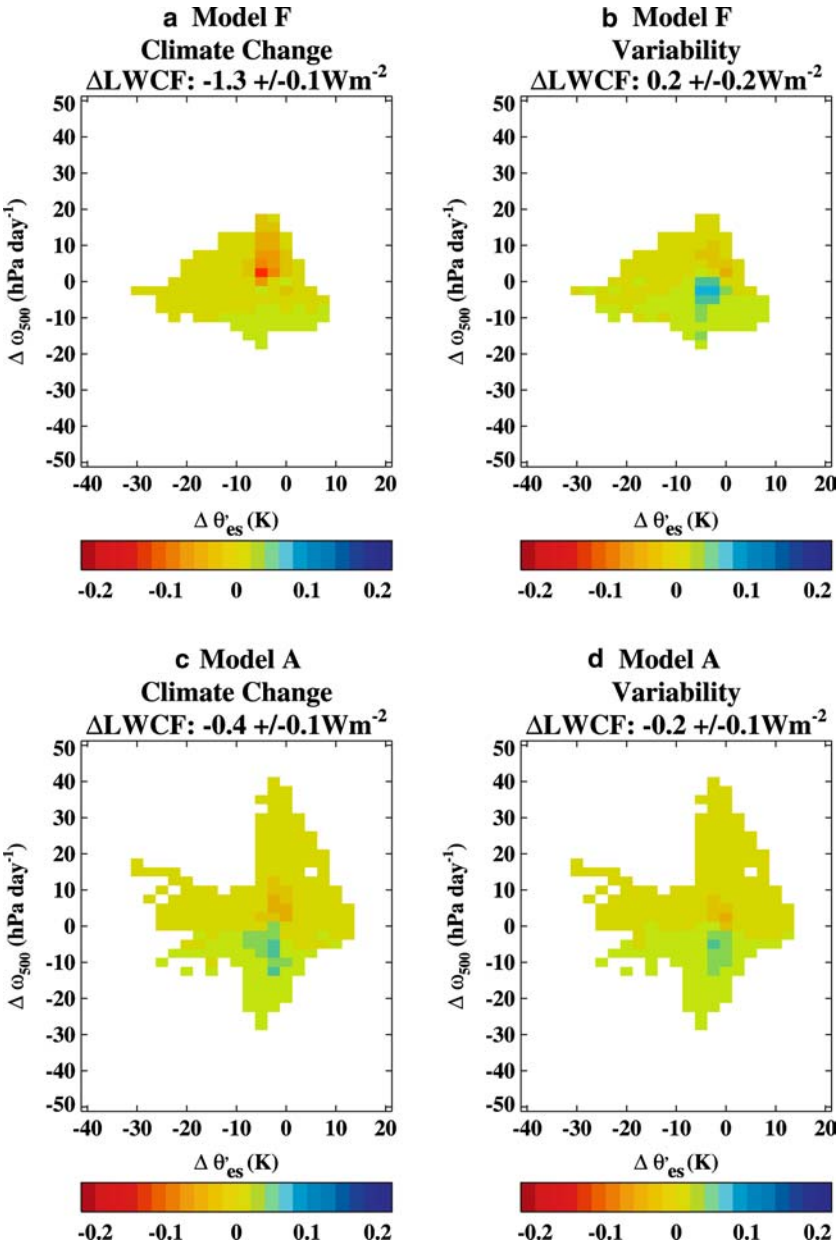

Fig. 5 Similar to Fig. 4b, e, except for Model F and Model A only

Table 5 For each model: the mean climate change weighted, RMS difference (Wm^{-2}) between the model and ERBE data composited by ERA-40 re-analysis and by NCAR/NCEP re-analysis

Model	ERBE/ERA		ERBE/NCEP	
	ΔSCRF	ΔLCRF	ΔSCRF	ΔLCRF
Model A	2.5	**1.0**	3.1	1.0
Model B	1.9	1.2	2.2	1.1
Model C	**1.2**	**1.0**	1.7	**0.9**
Model D	2.2	1.4	3.0	1.5
Model E	1.8	1.1	2.3	1.2
Model F	1.4	1.2	**1.3**	1.2
Model G	1.7	1.6	2.3	2.0
Model H	1.5	1.1	2.0	1.0
Model I	**1.2**	1.4	**1.3**	1.1

RMS difference between Figs. 2f and 9b for each model, with the RMS calculation being weighted by Fig. 2a. The model with the lowest RMS difference (i.e. most similar to observations) is highlighted in bold in each case

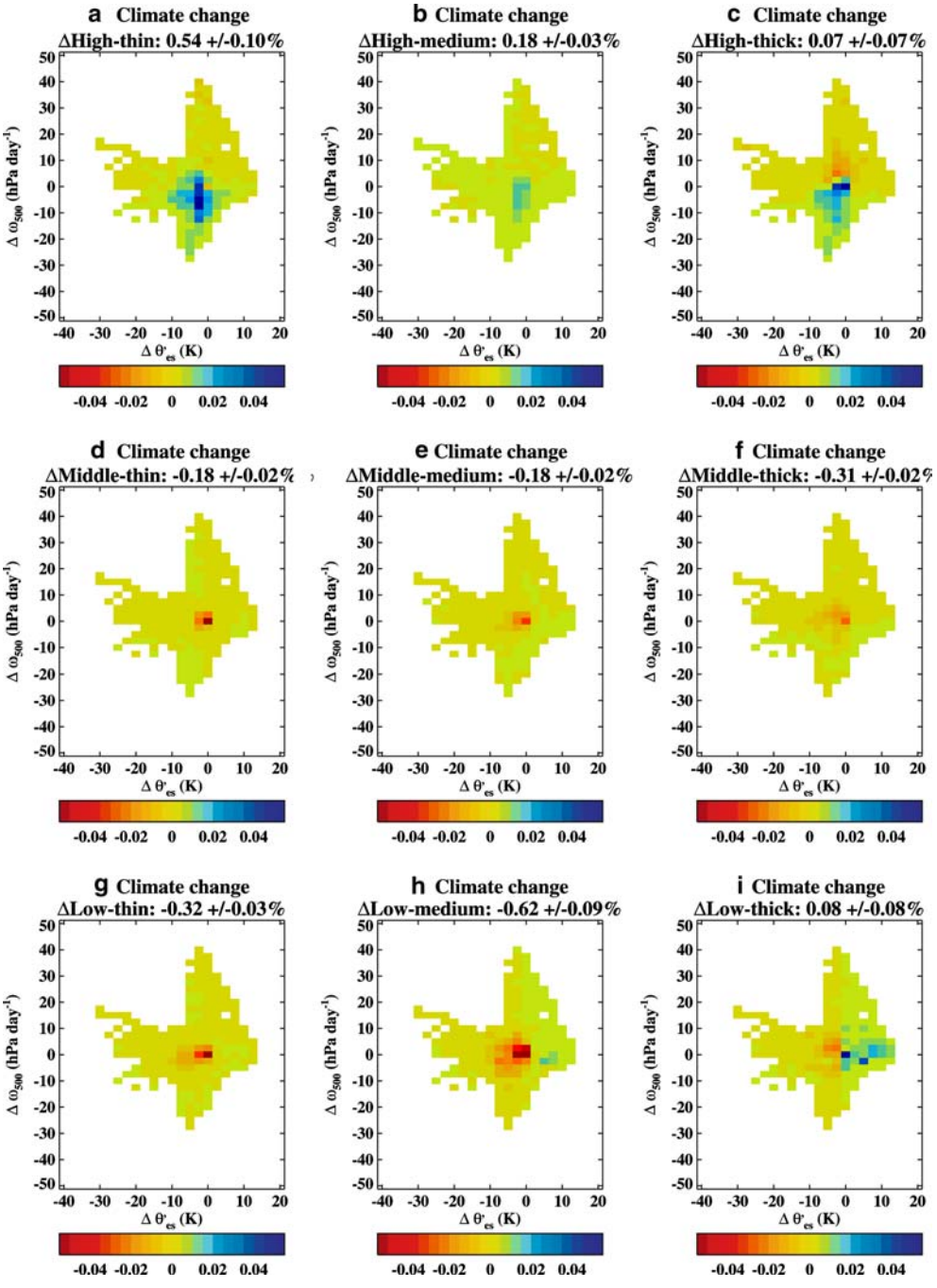

Fig. 6 Change in each ISCCP cloud type in response to doubling CO_2 as simulated by Model A and composited by $\Delta\omega_{500}-\Delta\theta'_{es}$. The composites are weighted by the RFO of the bins from Model A. WMCs and uncertainties due to variability are shown at the top of each figure

conditions resulting in a positive change in NCRF of 0.8 Wm^{-2}. The uncertainty estimates are calculated using similar method as that adopted for the correlation uncertainty estimates. Additionally, the RFO of the climate change composite is calculated separately and the variances added to those due to variability in CRF.

The climate change and variability WMC for each model are given in Table 4. The uncertainty estimates for the present-day variability tend to be slightly larger than the climate change composite due to the construction of the composite from monthly mean anomalies compared with a difference in a multi-annual mean. In general, the estimated error ranges from the climate change and present-day variability composites do not overlap, indicating that changes in ω_{500} and θ'_{es} alone are not sufficient to explain the global mean change in CRF in response to climate change. This is not surprising given the complexity and diversity of model parameterisations and interactions (cloud parameteri-sations are generally not a simple function of these two variables!). However, in all cases the climate change and present-day predicted change in SCRF is positive, and in most cases the climate change and present-day predicted change in LCRF is closer to zero (although this is always negative for the climate change response).

Table 6 For each model as a difference between El-Niño and a La-Niña: correlations and RMS differences for maps of $\Delta\ \omega_{500}$ (hPa day^{-1}) and $\Delta\ \theta'_{es}$ (K) between the model and re-analyses; correlations between the composited RFO of the model and re-analyses; the net change in θ'_{es} (for comparison, net $\Delta\ \theta'_{es}$ for ERA-40 is 0.2 K and for NCAR/NCEP is −1.0 K)

Model	Difference in ω_{500}				Difference in θ'_{es}				RFO		Net$\Delta\ \theta'_{es}$
	Corr.		RMSD		Corr.		RMSD		Correlation		
	ERA	NCEP	ERA	NCEP	ERA	NCEP	ERA	NCEP	ERA	NCEP	
Model A	0.65	0.62	15.3	14.6	0.69	0.66	2.7	2.8	0.94	0.91	−0.6
Model B	**0.69**	**0.69**	**13.0**	**11.0**	0.70	0.72	2.6	2.4	0.87	0.91	−0.8
Model C	0.68	0.62	15.9	16.6	**0.75**	**0.76**	**2.4**	**2.3**	**0.95**	**0.92**	−0.5
Model F	0.61	0.68	14.4	11.5	0.68	0.73	2.9	2.7	0.93	0.90	−0.5
Model G	0.62	0.58	13.9	12.0	0.64	0.65	2.8	2.6	0.91	**0.92**	−0.8

In each case, the model with the highest correlation/lowest RMS difference (i.e. closest to observations) is highlighted in bold

Similar weighted composites may be produced for other cloud variables: for example, cloud amount in each of the ISCCP D2 cloud types (Rossow and Schiffer 1991) as output by the ISCCP simulator. The climate change composite, weighted by the RFO of the bins, is shown for Model A in Fig. 6 and the climate change weighted present-day variability composite in Fig. 7. For some of the cloud types (high-top thick and thin cloud types, and low cloud with a medium optical thickness), the climate change and variability composites appear to be very similar (although the WMCs remain inconsistent), however mid-level-topped cloud and thin low cloud appear to have less similarity between the climate change and variability composites. At present, monthly ISCCP simulator data are only available for four of the models, but in all of these it is found that high and low-top cloud (particularly low cloud of intermediate thickness) show reasonable agreement between the climate change and variability composites, whereas middle-top clouds show less agreement. This suggests that processes affecting mid-level cloud are being excluded from the analysis presented in this study, possibly explaining some of the discrepancy in the WMCs between the climate change and variability composites in Table 3. Future work will investigate this further.

This study uses atmosphere models coupled to mixed-layer ocean models since this GCM configuration has the largest number of structurally different models available (with the required diagnostics) at the time of analysis. However, the slab model configuration is not a pre-requisite of the methodology. The technique could be applied equally well to transient coupled model simulations. This has been carried out for coupled model versions of four of the models used here and the results for the coupled and slab versions of Model A are shown in Fig. 8 for comparison. The difference between a 20-year period centred on the time of CO_2 doubling, in a coupled simulation with CO_2 increasing at 1%/year, and the same period in a pre-industrial control, are analysed as the climate change response. Since the warming at the time of CO_2 doubling in a transient simulation is less than the equilibrium warming due to the thermal inertia of the oceans, the shift to more unstable conditions is found to be smaller than in the slab composites, hence the WMC for SCRF is smaller (cf. Fig. 8a, e). However, for the region of the composite space occupied by both coupled and slab configurations, the correlation of the CRF changes is in excess of 0.8 for all of the models, suggesting the main processes determining the local cloud response are being simulated in both the slab and coupled models. In addition, the correlations of the composited transient climate change CRF with composited present-day variability CRF are similar to those in Table 3. This implies that the conclusions from this study are likely to also be valid for the cloud response in transient coupled model simulations.

The correlation between the climate change and present-day variability CRF composites in Table 3, combined with some broad similarity in the 60°N–60°S WMC, suggests that changes in ω_{500} and θ'_{es} are important for explaining at least a component of the cloud response to climate change, and that it can be related to present-day variability (although other processes also contribute to the WMC). Given this relationship, evaluation of the variability composite will also be an evaluation of an aspect of the cloud response to climate change in a GCM.

5 Evaluation of cloud composites

There are two stages to the evaluation of cloud response to climate change within the compositing framework. Firstly, an evaluation of whether the change in CRF associated with the compositing variables correct i.e. is the cloud response in each bin correct? The second stage is to determine whether the changes in ω_{500} and θ'_{es} are likely to be correct i.e. is the climate change RFO of the bins correct? These questions will be addressed in the subsequent two sub-sections and the relative importance of the two in reducing uncertainty between models is quantified in Sub-sect. 5.3.

5.1 Evaluation of composited CRF

The first part of the evaluation involves comparing the simulated present-day variability CRF composites with

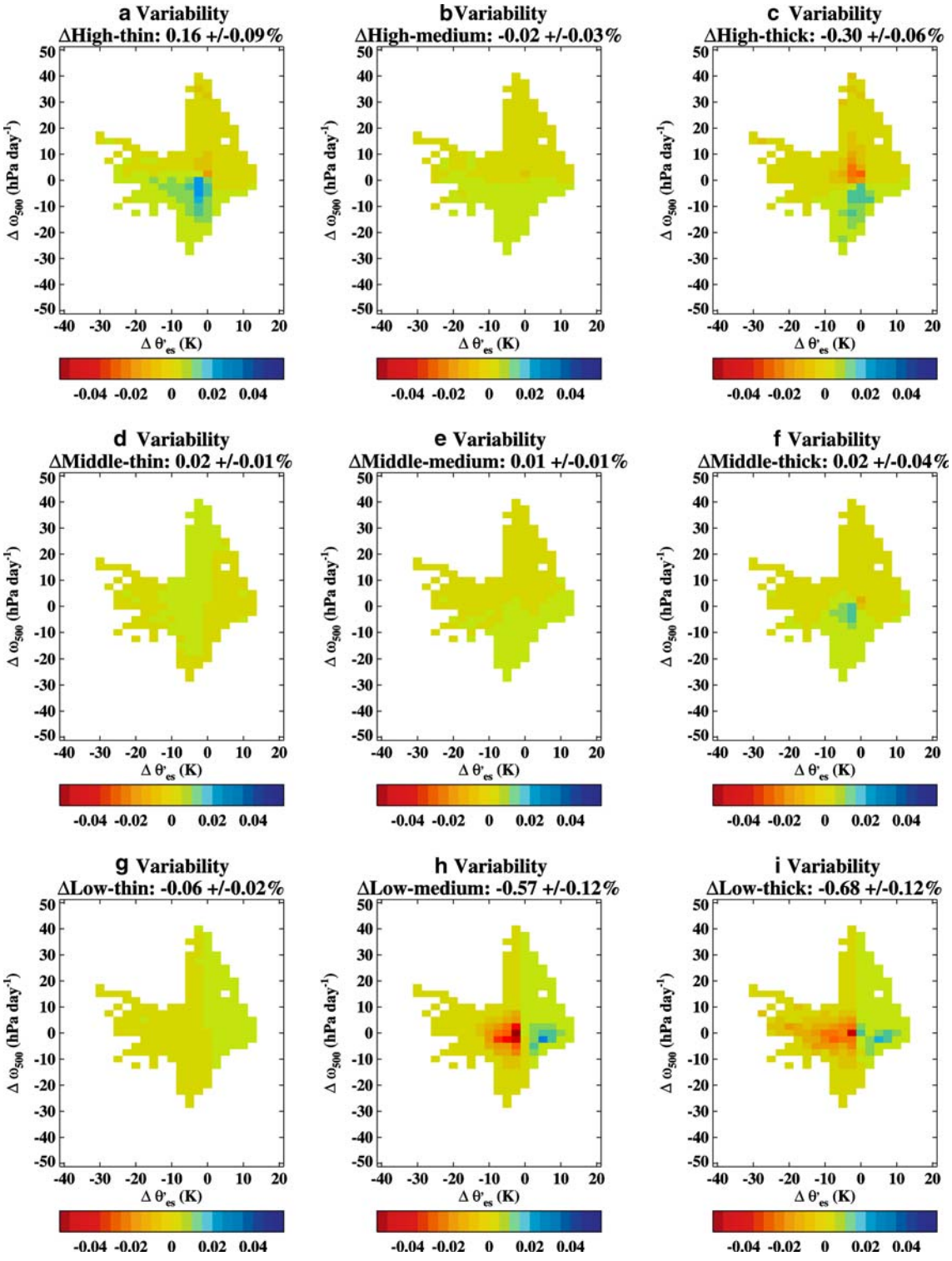

Fig. 7 Spatio-temporal anomalies in each ISCCP cloud type for the Model A control simulation, composited by anomalies in ω_{500} and θ'_{es}. The composites are weighted by the climate change RFO of the bins from model A for comparison with Fig. 6. WMCs and uncertainties due to variability are shown at the top of each figure

composited data from ERBE. The observational data are composited using anomalies in ω_{500} and θ'_{es} from both ERA-40 and NCAR/NCEP re-analyses for the period March 1985–February 1990. The two re-analyses are used to test for any sensitivity of the results to the re-analysis. Monthly ERBE datasets contain a reasonable amount of missing data. The approach adopted has been to simply include all non-missing data points from the observational dataset and compare with the model (which has no missing data other than those points with

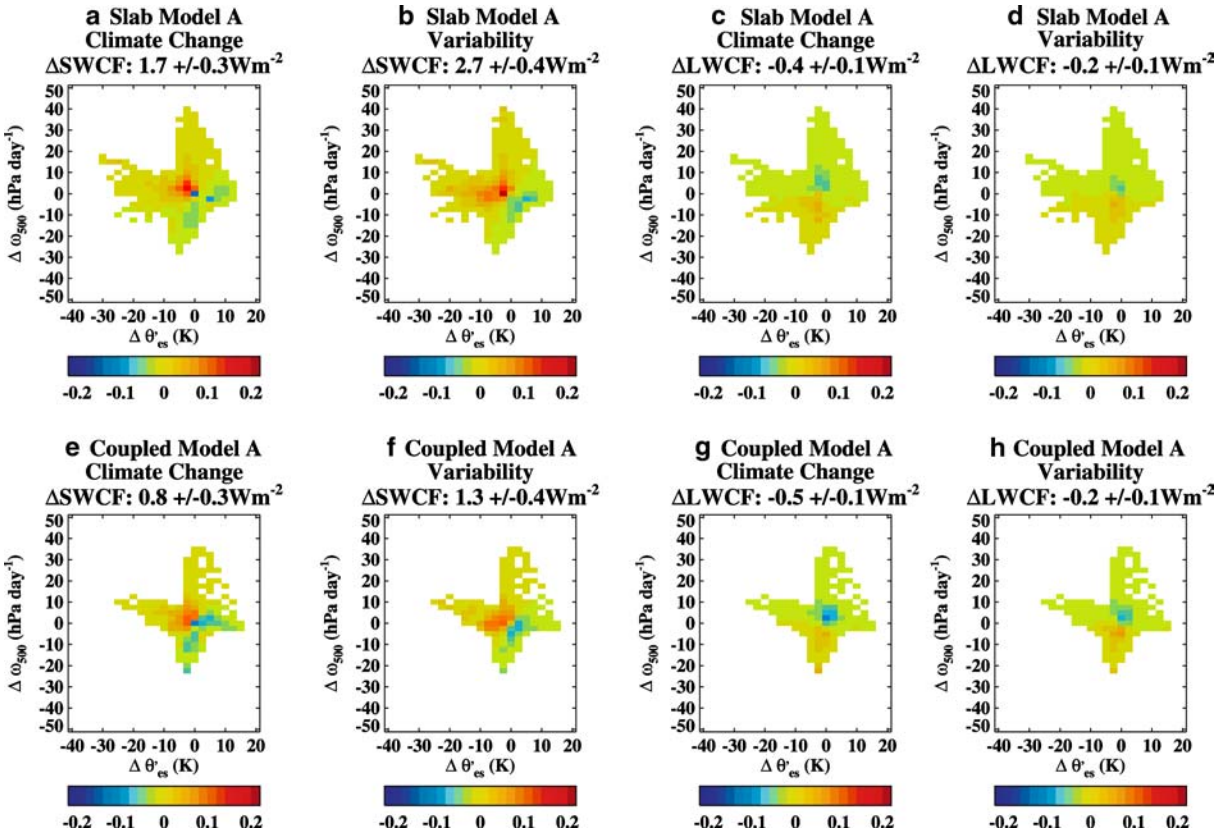

Fig. 8 Climate change and present-day variability $\Delta SCRF$ and $\Delta LCRF$, composited by $\Delta\ \omega_{500}$ and $\Delta\ \theta'_{es}$ and weighted by the climate change RFO. **a–d** Slab version of Model A with the climate change response being the equilibrium response to doubling CO_2. **e–h** Coupled version of Model A with the climate change response being the period around the time of CO_2 doubling in a transient simulation with CO_2 increasing at 1%/year

surface pressure below 800 hpa). Since more of the extreme stability anomalies are over land, where data are often missing, the range of stability anomalies sampled by the observational data are more constrained than from the models (Fig. 9a, e). Therefore, evaluation of points with the largest changes in stability under climate change is not possible.

Compositing the ERBE data by the two re-analyses gives very similar results, implying that this aspect of the evaluation is not sensitive to the choice of re-analysis (Fig. 9). A root-mean-square (RMS) difference of the observational and model present-day variability composites provides a simple scalar measure for evaluation. This could be performed on the unweighted composites, but differences in the large CRF anomalies towards the edges of the composite may considerably influence the results. The centre and slightly more unstable regions of the composite space are the most important for the mean climate change response (Fig. 2a), hence an RMS difference of the unweighted composites may not be a good measure of the model in those bins which determine the mean climate change response. Ideally there would be some independent method (i.e. not using models) for upweighting the bins which are most important for the climate change response. Unfortunately, this is not the case since observed climate change to date is not sufficient to detect a climate change signal of $\Delta\ \omega_{500}$ and $\Delta\ \theta'_{es}$. Hence, the observational and model variability composites from the GCMs are each weighted by the GCM-mean climate change RFO (Fig. 2a). In using this mean RFO, the simulated climate change $\Delta\ \omega_{500}$ and $\Delta\ \theta'_{es}$ from each model is treated as equally likely, although weighting by the RFO of the individual models does not greatly influence the results (see also Sect. 5.3). The RMS difference of the weighted model and observational composites are given in Table 5. In general there appears to be more variation in the RMS difference for SCRF anomalies than LCRF anomalies. Despite the RMS differences varying by more than a factor of two between the models, there are no clearly superior or deficient models, however Models C, F and I are in closest agreement with the observations for anomalies in both CRF components, when composited by either re-analyses. The main deficiency in the models which evaluate less favourably in the SCRF comparison (e.g. Models A and D), is that changes in SCRF are too sensitive to changes in θ'_{es} for locations with near-zero changes in vertical velocity i.e. the variation of ΔSCRF with $\Delta\ \theta'_{es}$ across the centre of the composite is too rapid. It is possible that further discrimination between the models might be

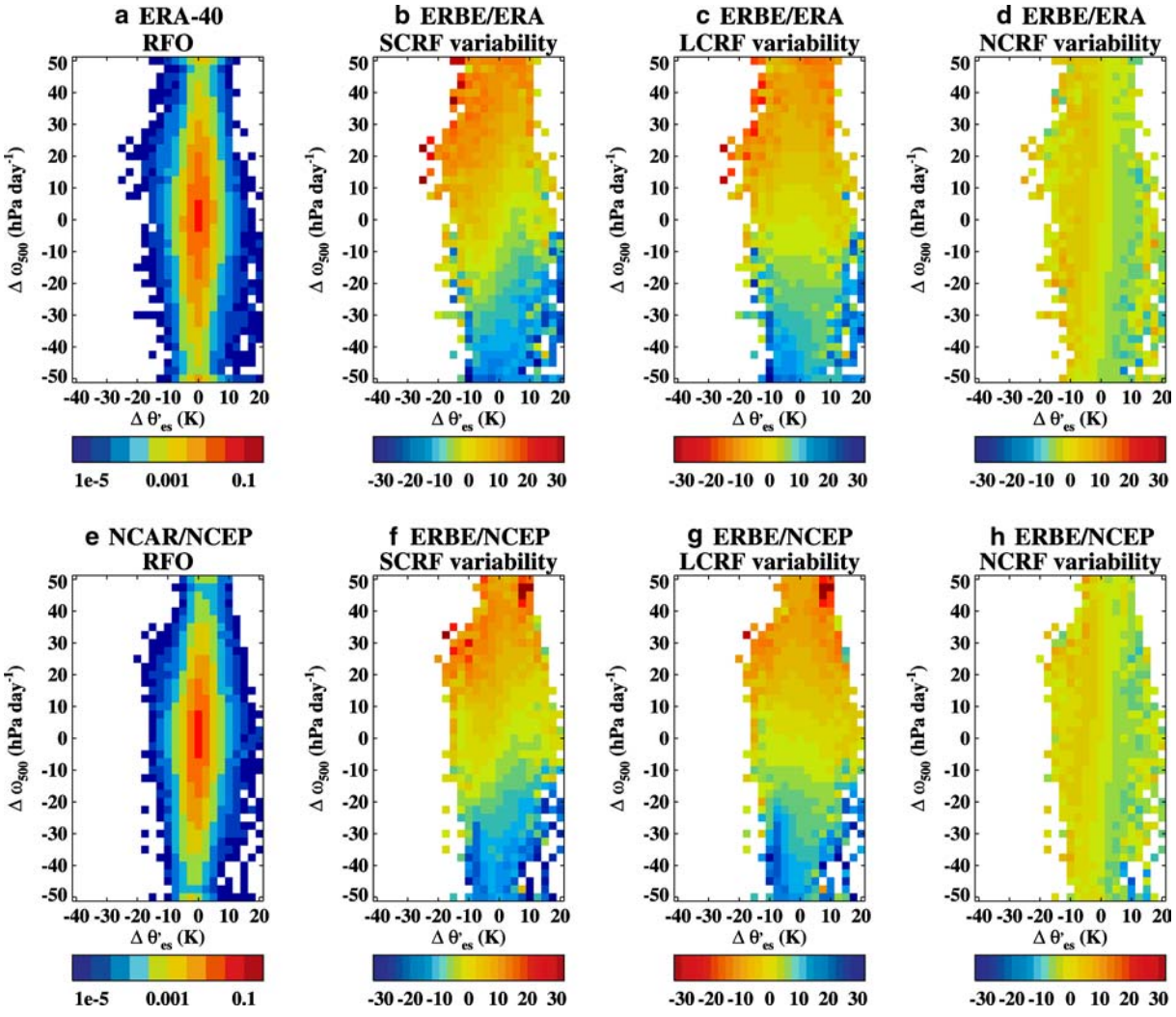

Fig. 9 a, e RFO of monthly spatio-temporal anomalies in ω_{500} and θ'_{es} from ERA-40 and NCAR/NCEP re-analyses. **b–d, f–h** Anomalies in SCRF, LCRF and NCRF from ERBE composited by anomalies in ω_{500} and θ'_{es} from the re-analyses

made easier if the observational data also sampled the bins with large reductions in saturated stability.

The WMC of the CRF composites from ERBE, when weighted by the climate change mean RFO from the GCMs, can be used as a prediction from observational data of the change in CRF (60°N–60°S) due to a doubling of CO_2, associated with changes in ω_{500} and θ'_{es}, and assuming the GCM-mean change in ω_{500} and θ'_{es} (Fig. 2a). The values for ERBE composited by ERA-40 and NCAR/NCEP, respectively are 1.3 ± 0.6 Wm^{-2} and 1.0 ± 1.0 Wm^{-2} for the change in SCRF and -0.2 ± 0.9 Wm^{-2} and 0.0 ± 1.1 Wm^{-2} for the change in LCRF. The error estimates take account of the uncertainty due to internal variability and in the different climate change RFO between GCMs. They do not include any estimate of uncertainty in the ERBE data. In addition, the prediction does not include any estimate of the contribution from the bins populated in the mean climate change response but not sampled by the observations. Comparing these observationally based WMCs with the model variability WMCs in Table 4, it can be seen that the ΔSCRF in several models (e.g. Models A, D and E) appears to be excessive.

5.2 Evaluation of the change in vertical velocity/stability

The second part of the evaluation process is to determine whether the climate change RFO composite is accurate. The climate change RFO cannot be directly evaluated, however confidence in a model's ability to simulate changes in ω_{500} and θ'_{es} may be obtained by evaluating changes in these variables as part of present-day variability. The El-Niño Southern Oscillation (ENSO) is a primary mode of natural variability in the climate system. Five of the models used in this study have parallel atmosphere-only simulations available, forced with observed SSTs (following the AMIP experimental design). The difference in mean $\omega_{500}/\theta'_{es}$ fields between an El-Niño and a La-Niña event from these models is used to assess the ability of the GCM to simulate tropical

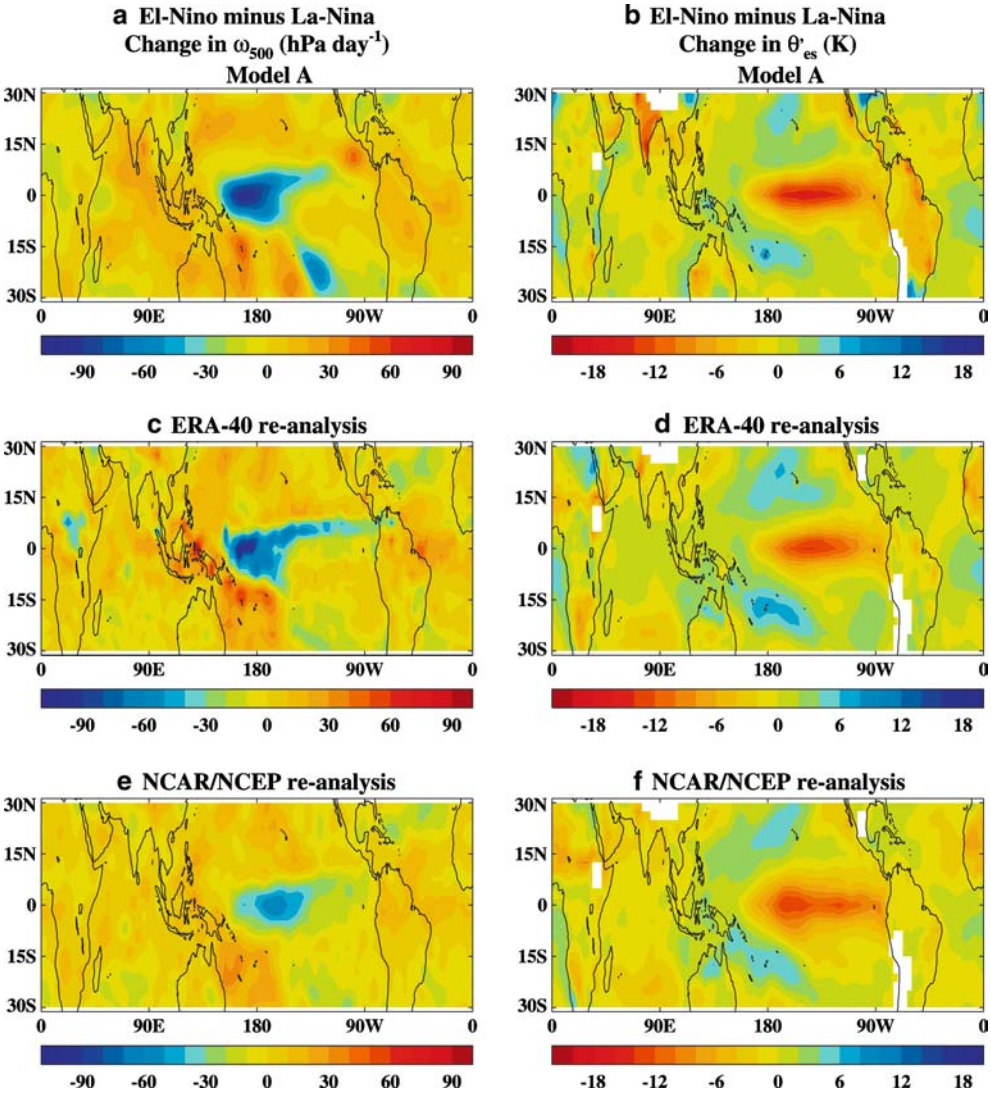

Fig. 10 The mean difference in ω_{500} (*left*) and θ'_{es} (*right*) between 12 months of an El-Niño event (January 1987–December 1987) and a La-Niña event (May 1988–April 1989). The mean difference is shown for: **a, b** Model A (atmosphere-only configuration); **c, d** ERA-40 re-analysis; **e, f** NCAR/NCEP re-analysis

variability in the compositing variables. The La-Niña period May 1988–April 1989 is differenced from the El-Niño period January 1987–December 1987, these periods being chosen as they are the 12 consecutive months in the late 1980s when the warm/cold anomaly in the Niño 3.4 region (5°N–5°S, 120°–170°W) was greatest.

Maps of differences in ω_{500} and θ'_{es} between an El-Niño and a La-Niña for one of the models (Model A), ERA-40 and NCAR/NCEP are shown in Fig. 10. Both re-analyses show a broadly similar pattern with increased ascent towards the central Pacific, a reduction in stability over the central and eastern Pacific and reduced ascent/increased stability over the warm pool and subtropical western Pacific. However, there are some differences, particularly in terms of the magnitude and extent of the areas of greatly increased ascent. These differences will affect the ability to differentiate between models. A correlation, or RMS difference, of the geographical map of $\Delta\ \omega_{500}/\Delta\ \theta'_{es}$ between the model and re-analyses provides a measure of the ability of the model to accurately simulate changes in these variables (Table 6). The correlations and RMS differences for both variables vary between GCMs by 20–30%, suggesting that the uncertainty in both variables is similar. For the change in ω_{500}, Model B scores highest for both the correlation and RMS difference (i.e. highest correlation and lowest RMS difference). For the change in θ'_{es}, Model C scores the most favourably.

Whilst the ability of the model to correctly simulate the geographical location of changes in ω_{500} and θ'_{es} is important for the regional cloud response to climate change, for the compositing analysis presented above it is the population of each $\Delta\ \omega_{500}/\Delta\ \theta'_{es}$ bin that is important. Therefore, as an additional measure of model variation in ω_{500} and θ'_{es}, RFO composites for the re-analyses and models are correlated for the difference between the El-Niño and a La-Niña event (Table 6). The mean change in θ'_{es} is also calculated (the mean change in ω_{500} being close to zero in all cases). The ENSO RFO composites for the two re-analyses are somewhat different with ERA-40 being centred on zero change in θ'_{es} for all changes in ω_{500}, whereas the NCAR/NCEP

re-analysis has a shift to more unstable conditions in regions of increased ascent/reduced descent (i.e. the axis of the RFO is tilted from top right to bottom left) (not shown). This suggests that the two compositing variables are related in the NCAR/NCEP model (although the methodology presented in this paper does not require them to be independent). The net change in θ'_{es} is -1.0 K for NCAR/NCEP and is slightly positive for ERA-40. All of the models show a shift to slightly more unstable conditions with net $\Delta\ \theta'_{es}$ values between those of the two re-analyses. In addition, all of the composite correlations are high—Model C correlating the best with both re-analyses. Overall, the models can be described as being within the envelope of the re-analyses, although we are unable to chose between them in terms of the composited ENSO RFO.

5.3 Relative importance of evaluating changes in CRF versus changes in RFO

In order to prioritise the development of the above evaluation methodologies, it is useful to know the relative importance of the two aspects to the evaluation. Given that the primary aim is to reveal uncertainty between GCM simulations, this may be determined by comparing the range of changes in CRF which occur from using different model CRF composites with the range which occurs from applying different model RFO distributions. Firstly, the GCM-mean climate change RFO (Fig. 2a) is applied to the present-day CRF variability composites from each GCM, and the variance in the composite integrated CRF components calculated. Secondly, the different GCM climate change RFO distributions are applied to the observational CRF anomaly composite (Fig. 9b–c) (alternatively the GCM-mean CRF composite could be used) and the variance in the composite integrated CRF components calculated. The relative variance of each component is then compared with the combined sum of the two. The combined variance is similar (but not identical) to the variance in the present-day predictions in Table 4. If the first component accounts for 100% of the combined variance, then the composited climate change $\Delta\ \omega_{500}/\Delta\ \theta'_{es}$ population is the same for all the models, and differences are entirely due to differences in the associated change in CRF. In this case the evaluation in Sub-sect. 5.1 is of primary importance. Alternatively, if the second component accounts for 100% of the combined variance, then the models all simulate the same change in CRF for a given change in $\Delta\ \omega_{500}/\Delta\ \theta'_{es}$, and the uncertainty in the climate change response is from differences in the RFO of the compositing variables. In this case, evaluation methodologies similar to those in Sub-sect. 5.2 require further development.

Applying this analysis, it is found that 63% of the combined variance in ΔSCRF and 84% of the combined variance in ΔLCRF is due to the first component (differences in the simulated change in CRF within the bin). Hence, the first part of the evaluation (Sub-sect. 5.1) is the most important for assessing differences between the GCM cloud response to climate change. Differences in the climate change RFO are less important for accounting for differences in response, which is fortunate since assessing confidence in vertical velocity and stability changes under climate change has been shown to be less straightforward. It should be noted, however, that this result assumes that the climate change response in the real world lies within the model ensemble. It is possible that all the model's are wrong and the actual change in RFO will be outside the model range.

6 Summary and discussion

This study has applied a compositing methodology in order to attempt to relate the cloud response to climate change to present-day variability in a number of contemporary GCMs. Reasonably high correlations are found between the multi-annual mean change in CRF in response to a doubling of CO_2 and anomalies of CRF as part of present-day variability, when composited by the change in 500 hpa vertical velocity and saturated lower tropospheric stability. This suggests that changes in ω_{500} and θ'_{es} are important for at least a component of the local cloud response to climate change. Since the relationship has been demonstrated to exist in several models with significant structural differences, we might have some confidence that the same relationship will exist in the real world (unless the parameterisations in *all* the models are wrong). The combined tropical and mid-latitude mean climate change ΔSCRF, and that predicted from present-day variability, both tend to be positive. The ΔLCRF from both composites tend to be closer to zero, although all models simulate a small negative change in LCRF in response to climate change. The quantitative 60°N–60°S mean ΔCRF in the climate change composites and the ΔCRF predicted from present-day variability are not consistent, suggesting that processes not accounted for in the present analysis are also responsible for determining the net cloud response to climate change. However, since the correlations and WMCs for the GCM-mean composites are amongst the highest/closest of the models, the processes not being represented by $\Delta\ \omega_{500} - \Delta\ \theta'_{es}$ appear to be largely specific to individual models.

It is possibly surprising that θ'_{es} appears to be more closely associated with the low cloud response than θ'_{e}, since the former contains no information about the humidity distribution. The full compositing analysis has been repeated with several of the GCMs using θ'_{e}, but the correlations between the climate change and present-day variability composites are poorer and the WMCs do not show even the broad similarities identified above. Further process based analysis, possibly involving simple models and high temporal-frequency diagnostics, is required in order to fully understand this result.

Since composited CRF anomalies in present-day climate variability have been shown to be important for a component of the cloud response to climate change, evaluation against observational data forms a 'direct' evaluation of an aspect of the climate change response. The evaluation process may be divided into two stages: an evaluation of the CRF anomalies within each bin and an evaluation of the climate change bin population. Differences in the former have been found to account for most of the difference in the net response between GCMs. The first stage of the evaluation involves comparison against composited observational data. The results are found to be relatively insensitive to the choice of re-analysis used for the compositing. This study used ERBE data for the observed CRF, with no account being taken of errors in this dataset. An extension of the evaluation method might include compositing an alternative CRF dataset such as from the Clouds and Earth's Radiant Energy System (CERES) (Wielicki et al. 1996) in order to investigate the effect of errors in the observational data. The second stage of the evaluation process is less straightforward since the changes in ω_{500} and θ'_{es} due to climate change cannot be directly evaluated. In this study, the ability of the GCMs to simulate changes in these variables in an ENSO cycle has been used to assess confidence in the likely ability of the model to simulate changes in these variables in response to doubling CO_2. Of course, processes governing changes in connection with ENSO may not be a good proxy for those resulting from an external forcing. If possible, the ability of the GCMs to simulate changes in ω_{500} and θ'_{es} associated with other forms of variability (the North Atlantic Oscillation, annual cycle, etc.) should be tested in order to increase confidence. Although the second part of the evaluation cannot be performed directly, it has been shown that more of the uncertainty between GCM simulations of cloud response occurs from differences in the CRF anomalies rather than from the climate change $\omega_{500}-\theta'_{es}$ response.

None of the models studied have been found to be clearly superior or deficient, however a couple of the GCMs (Model C and Model F) perform well in both stages of the evaluation, particularly in the more important evaluation of CRF anomalies within the composite bins. Since changes in ω_{500} and θ'_{es} do not fully account for the quantitative mean change in CRF, the evaluation proposed in this study cannot exclusively be used to discriminate between model predictions, and this is part of the reason why the models have been presented anonymously. Instead, it is proposed that evaluation of composited cloud anomalies be included as part of a basket of measures of GCM performance, similar to that used by Murphy et al. (2004).

Whilst changes in ω_{500} and θ'_{es} appear to account for a component (and in many models, a major component) of the local cloud response to climate change, analysis of output from the ISCCP simulator for several of the GCMs suggest that these variables are most important for particular cloud types (low cloud of an intermediate optical depth and high-top cloud). The response of mid-level cloud to increased CO_2 appears to be largely related to other processes, not accounted for by these variables. These other processes which are not directly constrained by ω_{500} and θ'_{es} result in the lack of quantitatively similar WMCs in the climate change and variability composites, and hence compositing by these two variables alone does not directly constrain the global mean ΔCRF or climate sensitivity. However the qualitative similarity of the WMCs and the composite correlations correlations highlight the importance of the ω_{500} and θ'_{es} in determining the local physico-dynamical regimes. This suggests that processes related to ω_{500} and θ'_{es} must be isolated if more subtle mechanisms contributing to the mean cloud response to climate change are to be identified. Future work will aim to refine compositing methods, similar to those presented here, for example by applying the compositing techniques to particular cloud regimes (such as defined by Jakob and Tselioudis 2003). This will possibly permit identification of other contributing processes and further our understanding of the differences in the cloud response to climate change between GCMs.

Acknowledgements This work was funded under the U.K. Government Meteorological Research programme and by the U.K. Department of the Environment, Food and Rural Affairs under contract PECD 7/12/37. Thanks go to David Sexton for assistance with demonstrating the statistical significance of the correlations and to William Ingram for supplying useful comments on an early draft of the paper. We also thank two anonymous reviewers for some useful comments which have improved the manuscript. ERBE data were obtained from the NASA Langley Research Center Atmospheric Sciences Data Center. NCEP Reanalysis data provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their web site at http://www.cdc.noaa.gov/. ERA-40 data were obtained from ECMWF. We acknowledge IPSL for providing an FTP server for the CFMIP project and the Met Office for hosting the CFMIP website. We also acknowledge the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the data used here from the MPI and GISS models. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy.

References

Andronova NG, Schlesinger ME (2001) Objective estimation of the probability density function for climate sensitivity. J Geophys Res 106:22605–22612

Barkstrom BR, Harrison EF, Lee RB et al (1990) Earth radiation budget experiment (ERBE) archival and April 1985 results. Bull Am Meteorol Soc 70:1254–1262

Bony S, Lau KM, Sud YC (1997) Sea surface temperature and large scale circulation influences on tropical greenhouse effect and cloud radiative forcing. J Climate 10:2055–2077

Bony S, Dufresne JL, Le Treut H, Morcrette JJ, Senior CA (2004) On dynamic and thermodynamic components of cloud changes. Clim Dyn 22:71–86. DOI 10.1007/s00382–003–0369–6

Cess RD et al (1990) Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. J Geophys Res 95:16601–16615

Charlock TP, Ramanathan V (1985) The albedo field and cloud radiative forcing produced by a general circulation model with internally generated cloud optics. J Atmos Sci 42:1408–1429

Colman R (2003) A comparison of climate feedbacks in general circulation models. Clim Dyn 20:865–873
Cubasch U, Meehl GA, Boer GJ, Stouffer RJ, Dix M, Noda A, Senior CA, Raper SCB, Yap KS (2001) Projections of future climate change. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden P, Dai X, Maskell K, Johnson CI (eds) Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp 525–582
Forest CE, Stone PH, Sokolov AP, Allen MR, Webster MD (2002) Quantifying uncertainties in climate system properties with the use of recent climate observations. Science 295:113–117
GFDL Global Atmospheric Model Development Team (2004) The new GFDL global atmosphere and land model AM2-LM2: evaluation with prescribed SST simulations. J Climate 17(24):4641–4673
Gregory JM, Stouffer RJ, Raper SCB, Stott PA, Rayner NA (2002) An observationally based estimate of the climate sensitivity. J Climate 15:3117–3121
Hanson HP (1991) Marine stratocumulus climatologies. Int J Climatol 11:147–164
Harrison EP, Minnis P, Barkstrom BR, Ramanathan V, Cess RD, Gibson GG (1990) Seasonal variation of cloud radiative forcing derived from the Earth radiation budget experiment. J Geophys Res 95:18687–18703
Hartmann DL, Michelsen ML (1993) Large-scale effects on the regulation of tropical sea surface temperature. J Climate 6:2049–2062
Hartmann DL, Moy LA, Fu Q (2001) Tropical convection and the energy balance at the top of the atmosphere. J Climate 14:4495–4511
Houghton JT, Jenkins GJ, Ephraums JJ (1990) Climate change: the IPCC scientific assessment. Cambridge University Press, Cambridge
Houghton JT, Callander BA, Varney SK (1992) Climate change 1992. The supplementary report to the IPCC scientific assessment. Cambridge University Press, Cambridge
Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (1996) Climate change 1995—the science of climate change. Cambridge University Press, Cambridge, p 572
Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden P, Dai X, Maskell K, Johnson CI (2001) Climate change 2001: the scientific basis. Contribution of working group 1 to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
Jakob C, Tselioudis G (2003) Objective identification of cloud regimes in the tropical Western Pacific. Geophys Res Lett 30(21). DOI 10,1029/2003GL018367
Johns TC, Durman CF, Banks HT, Roberts MJ, McLaren AJ, Ridley JK, Senior CA, Williams KD, Jones A et al (2005) The new Hadley centre climate model HadGEM1: evaluation of coupled simulations in comparison to previous models. J Climate Submitted
K-1 model developers (2004) K-1 coupled model (MIROC) description. K-1 technical report 1. H Hasumi, S Emori (eds), Center for Climate System Research, University of Tokyo
Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77(3):437–471
Klein SA, Hartmann DL (1993) The seasonal cycle of low stratiform clouds. J Climate 6(8):1587–1606
Klein SA, Jakob C (1999) Validation and sensitivities of frontal clouds simulated by the ECMWF model. Mon Weather Rev 127(10):2514–2531
Knutti R, Stocker TF, Joos F, Plattner GK (2002) Constraints on radiative forcing and future climate change from observations and climate model ensembles. Nature 416:719–723
Martin GM, Ringer MA, Pope VD, Jones A, Dearden C, Hinton T (2005) The physical properties of the atmosphere in the new Hadley centre global environmental model, HadGEM1. Part 1: Model description and global climatology. J Climate (in press)
Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430:768–772
National Academy of Sciences (1979) Carbon dioxide and climate: a scientific assessment. National Research Council, Washington
Norris JR, Leovy CB (1994) Interannual variability in stratiform cloudiness and sea surface temperature. J Climate 7:1915–1925
Norris JR, Weaver CP (2001) Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3. J Climate 14:2540–2550
Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Centre climate model—HadAM3. Clim Dyn 16:123–146
Ramanathan V, Collins W (1991) Thermodynamic regulation of ocean warming by cirrus clouds deduced from observations of the 1987 El Niño. Nature 351(6321):27–32
Ringer MA, Allan RP (2004) Evaluating climate model simulations of tropical cloud. Tellus 56:308–327
Roeckner E, Bauml G, Bonaventura L, Brokopf R, Esch M, Giorgetta M, Hagemann S, Kirchner I, Kornblueh L, Manzini E, Rhodin A, Schlese U, Schulzweida U, Tompkins A (2003) The atmospheric general circulation model ECHAM 5. Part I: Model description. Technical Report 349, Max Planck Institute for Meteorology
Rossow WB, Schiffer RA (1991) ISCCP cloud data products. Bull Am Meteorol Soc 72:2–20
Schlesinger ME, Mitchell JFB (1987) Climate model simulations of the equilibrium climate response to increased carbon dioxide. Rev Geophys 25:760–798
Schmidt GA et al (2005) Present day atmospheric simulations using GISS ModelE: comparison to in-situ, satellite and reanalysis data. J Climate (in press)
Senior CA, Mitchell JFB (1993) Carbon dioxide and climate: the impact of cloud parameterization. J Climate 6:393–418
Soden BJ, Broccoli AJ, Hemler RS (2004) On the use of cloud forcing to estimate cloud feedback. J Climate 17(19):3661–3665. DOI 10.1175/1520–0442(2004)017
Tselioudis G, Zhang Y, Rossow WB (2000) Cloud and radiation variations associated with northern midlatitude low and high sea level pressure regimes. J Climate 13(2):312–327
Weaver CP (1999) The interactions among cyclone dynamics, vertical thermodynamic structure and cloud radiative forcing in the North Atlantic summertime storm track. J Climate 12:2625–2642
Webb M, Senior C, Bony S, Morcrette JJ (2001) Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models. Clim Dyn 17:905–922
Wielicki B et al (1996) Clouds and the Earth's radiant energy system (CERES): an earth observing system experiment. Bull Am Meteorol Soc 77(5):853–868
Williams KD, Senior CA, Mitchell JFB (2001) Transient climate change in the Hadley centre models: the role of physical processes. J Climate 14(12):2659–2674
Williams KD, Ringer MA, Senior CA (2003) Evaluating the cloud response to climate change and current climate variability. Clim Dyn 20:705–721. DOI 10.1007/s00382–002–0303–3
Yang F, Schlesinger ME, Rozanov EV (2000) Description and performance of the UIUC 24-layer stratosphere-troposphere general-circulation model. J Geophys Res 105(D14):17925–17954
Zhang MH, Cess RD, Hack JJ, Kiehl JT (1994) Diagnostic study of climate feedback processes in atmospheric GCMs. J Geophys Res 99:5525–5537