

Tuning the climate of a global model

Thorsten Mauritsen,¹ Bjorn Stevens,¹ Erich Roeckner,¹ Traute Crueger,¹ Monika Esch,¹ Marco Giorgetta,¹ Helmuth Haak,¹ Johann Jungclaus,¹ Daniel Klocke,² Daniela Matei,¹ Uwe Mikolajewicz,¹ Dirk Notz,¹ Robert Pincus,^{3,4} Hauke Schmidt,¹ and Lorenzo Tomassini¹

Received 27 January 2012; revised 22 June 2012; accepted 2 July 2012; published 7 August 2012.

[1] During a development stage global climate models have their properties adjusted or *tuned* in various ways to best match the known state of the Earth's climate system. These desired properties are observables, such as the radiation balance at the top of the atmosphere, the global mean temperature, sea ice, clouds and wind fields. The tuning is typically performed by adjusting uncertain, or even non-observable, parameters related to processes not explicitly represented at the model grid resolution. The practice of climate model tuning has seen an increasing level of attention because key model properties, such as climate sensitivity, have been shown to depend on frequently used tuning parameters. Here we provide insights into how climate model tuning is practically done in the case of closing the radiation balance and adjusting the global mean temperature for the Max Planck Institute Earth System Model (MPI-ESM). We demonstrate that considerable ambiguity exists in the choice of parameters, and present and compare three alternatively tuned, yet plausible configurations of the climate model. The impacts of parameter tuning on climate sensitivity was less than anticipated.

Citation: Mauritsen, T., et al. (2012), Tuning the climate of a global model, *J. Adv. Model. Earth Syst.*, 4, M00A01, doi:10.1029/2012MS000154.

1. Introduction

[2] Model tuning is an integral part of the model development process, but is not extensively discussed in the literature. Although models and their configuration are well-documented, the process through which a particular model configuration comes into being is not, and as a result, the process of selecting a model configuration is shrouded in mystery. In this contribution we address ourselves to this gap in the literature.

[3] Model tuning is not a well-defined term. Often, model calibration or model tuning is associated with the last step of a broader model development cycle, after structural enhancements, improved parameterizations and refined boundary conditions have been implemented, wherein selected parameters are adjusted so as to better match the model results with some targeted features of the climate system [Randall and Wielicki, 1997]. The idea that models need to be harmonized with

observations is of course applicable to the model development process as a whole, as parameterizations and grid configurations are usually selected based on their ability to improve the representation of some aspect of the climate system [Jakob, 2010]. Only seldom do we implement model changes that degrade the performance of a climate model; improved aspects of the model results following a changed parameterization are frequently used as proof of concept. Because this model development happens over generations, and is difficult to describe comprehensively, in this paper the discussion is focused on how a final parameter configuration of a model is selected with an external set of goals in mind.

[4] The need to tune models became apparent in the early days of coupled climate modeling, when the top of the atmosphere (TOA) radiative imbalance was so large that models would quickly drift away from the observed state. Initially, a practice to input or extract heat and freshwater from the model, by applying flux-corrections, was invented to address this problem [Sausen et al., 1988]. As models gradually improved to a point when flux-corrections were no longer necessary [Colman et al., 1995; Guilyardi and Madec, 1997; Boville and Gent, 1998; Gordon et al., 2000], this practice is now less accepted in the climate modeling community. Instead, the radiation balance is controlled primarily by tuning cloud-related parameters at most climate modeling centers [e.g., Watanabe et al., 2010; Donner et al., 2011; Gent et al., 2011; HadGEM2 Development Team, 2011; Hazeleger

¹Max Planck Institute for Meteorology, Hamburg, Germany.

²European Centre for Medium Range Weather Forecasts, Reading, UK.

³Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado, USA.

⁴Physical Sciences Division, Earth System Research Laboratory, NOAA, Boulder, Colorado, USA.

et al., 2012], while others adjust the ocean surface albedo [Hourdin *et al.*, 2012] or scale the natural aerosol climatology to achieve radiation balance [Voldoire *et al.*, 2012]. Tuning cloud parameters partly masks the deficiencies in the simulated climate, as there is considerable uncertainty in the representation of cloud processes. But just like adding flux-corrections, adjusting cloud parameters involves a process of error compensation, as it is well appreciated that climate models poorly represent clouds and convective processes. Tuning aims at balancing the Earth's energy budget by adjusting a deficient representation of clouds, without necessarily aiming at improving the latter.

[5] Arguably, the most basic physical property that we expect global climate models to predict is how the global mean surface air temperature varies naturally, and responds to changes in atmospheric composition and solar insolation. We usually focus on temperature anomalies, rather than the absolute temperature that the models produce, and for many purposes this is sufficient. Figure 1 instead shows the absolute temperature evolution from 1850 till present in realizations of the coupled climate models obtained from the Coupled Model Intercomparison Project phase 3 (CMIP3) [Meehl *et al.*, 2007] and phase 5 (CMIP5) [Taylor *et al.*, 2012] multi-model datasets available to us at the time of writing, along with two temperature records reconstructed from

observations [Brohan *et al.*, 2006]. There is considerable coherence between the model realizations and the observations; models are generally able to reproduce the observed 20th century warming of about 0.7 K, and details such as the years of cooling following the volcanic eruptions, e.g., Krakatau (1883) and Pinatubo (1991), are found in both the observed record and most of the model realizations.

[6] Yet, the span between the coldest and the warmest model is almost 3 K, distributed equally far above and below the best observational estimates, while the majority of models are cold-biased. Although the inter-model span is only one percent relative to absolute zero, that argument fails to be reassuring. Relative to the 20th century warming the span is a factor four larger, while it is about the same as our best estimate of the climate response to a doubling of CO₂, and about half the difference between the last glacial maximum and present. To parameterized processes that are non-linearly dependent on the absolute temperature it is a prerequisite that they be exposed to realistic temperatures for them to act as intended. Prime examples are processes involving phase transitions of water: Evaporation and precipitation depend non-linearly on temperature through the Clausius-Clapeyron relation, while snow, sea-ice, tundra and glacier melt are critical to freezing temperatures in certain regions. The models in CMIP3

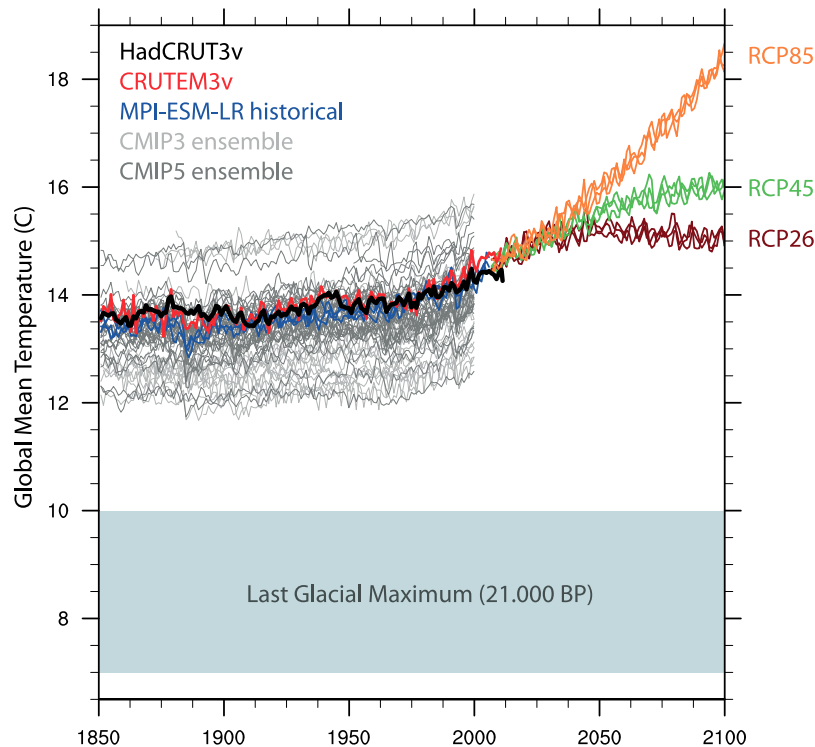


Figure 1. Absolute temperatures from climate model historical realizations and future scenarios. Black line is the HadCRUT3v blended land and ocean temperature dataset and red line is CRUTEM3v land-only temperatures [Brohan *et al.*, 2006]. Blue lines are three historical realizations, while orange, green and brown are future RCP-scenario realizations with the MPI-ESM-LR model, and light gray lines are the first historical realization from each model found in the CMIP3 dataset [Meehl *et al.*, 2007] and dark gray lines the corresponding CMIP5 historical realizations [Taylor *et al.*, 2012]. Some model realizations were started later than 1850. The estimated Last Glacial Maximum temperature range of 4–7 K below present is from *Intergovernmental Panel on Climate Change* [2007].

were frequently criticized for not being able to capture the timing of the observed rapid Arctic sea-ice decline [e.g., *Stroeve et al.*, 2007]. While unlikely the only reason, provided that sea ice melt occurs at a specific absolute temperature, this model ensemble behavior seems not too surprising when the majority of models do start out too cold.

[7] In addition to targeting a TOA radiation balance and a global mean temperature, model tuning might strive to address additional objectives, such as a good representation of the atmospheric circulation, tropical variability or sea-ice seasonality. But in all these cases it is usually to be expected that improved performance arises not because uncertain or non-observable parameters match their intrinsic value – although this would clearly be desirable – rather that compensation among model errors is occurring. This raises the question as to whether tuning a model influences model-behavior, and places the burden on the model developers to articulate their tuning goals, as including quantities in model evaluation that were targeted by tuning is of little value. Evaluating models based on their ability to represent the TOA radiation balance usually reflects how closely the models were tuned to that particular target, rather than the models intrinsic qualities.

[8] These issues motivate our present contribution where we both document and reflect on the model tuning that accompanied the preparation of a new version of our model system for participation in CMIP5. Through the course of preparation we took note of the decision-making process applied in selecting and adjusting parameters, and these notes are elaborated upon in Section 2. Because a number of the authors were new to model development the tuning exercise served as a learning process, one in which questions emerged that might have been taken for granted by the more experienced of the model developers, but are nonetheless of interest. As decisions were made, often in the interest of expediency, a nagging question remained unanswered: To what extent did our results depend on the decisions we had just made? Although the idea of a perturbed physics ensemble, through which an ensemble of simulations with different parameter settings is explored, was introduced partly to address this very question [*Stainforth et al.*, 2005], such an ensemble tends to produce models with unlikely parameter settings [*Rodwell and Palmer*, 2007], whereas during the tuning-process we adjust parameters in a more goal-oriented way.

[9] After the point when our model was frozen and the CMIP5 production runs were initiated we therefore revisited some of our earlier decisions, and asked how our model might have differed had a slightly different path been followed. In so doing we created a small number of alternative worlds; model configurations that were tuned by following a different branch in our tuning strategy. Using these alternative “worlds” as plausible configurations of our model that could have emerged in the development process, we explore how sensitive our model system is to the details of its configuration. These simulations help answer the question as to how much

tuning really improves a model, and what aspects of its critical behavior, for instance its patterns of variability or its climate sensitivity, depend on the tuning. Our hope is that this discussion will help to demystify the climate model tuning process.

2. Tuning the Model Climate

[10] A few model properties can be tuned with a reasonable chain of understanding from model parameter to the impact on model representation, among them the global mean temperature. It is comprehensible that increasing the models low-level cloudiness, by for instance reducing the precipitation efficiency, will cause more reflection of the incoming sunlight, and thereby ultimately reduce the model’s surface temperature. Likewise, we can slow down the Northern Hemisphere mid-latitude tropospheric jets by increasing orographic drag, and we can control the amount of sea ice by tinkering with the uncertain geometric factors of ice growth and melt. In a typical sequence, first we would try to correct Northern Hemisphere tropospheric wind and surface pressure biases by adjusting parameters related to the parameterized orographic gravity wave drag. Then, we tune the global mean temperature as described in Sections 2.1 and 2.3, and, after some time when the coupled model climate has come close to equilibrium, we will tune the Arctic sea ice volume (Section 2.4). In many cases, however, we do not know how to tune a certain aspect of a model that we care about representing with fidelity, for example tropical variability, the Atlantic meridional overturning circulation strength, or sea surface temperature (SST) biases in specific regions. In these cases we would rather monitor these aspects and make decisions on the basis of a weak understanding of the relation between model formulation and model behavior.

[11] Formulating and prioritizing our goals is challenging. To us, a global mean temperature in close absolute agreement with observations is of highest priority because it sets the stage for temperature-dependent processes to act. For this, we target the 1850–1880 observed global mean temperature of about 13.7°C [*Brohan et al.*, 2006]. Beyond that, we prioritize having globally averaged TOA shortwave absorption and outgoing longwave radiation in good agreement with satellite observations, along with a representation of important climate variability modes. We would accept a model if the global mean cloud cover is above 60 percent in present-day climate, even if satellite-estimates are generally higher, and global mean liquid water paths only in the range 50–80 gm^{-2} , which is consistent with estimates over the oceans from microwave instruments onboard satellites [e.g., *O’Dell et al.*, 2008] while the bulk of observational estimates would allow a broader range.

[12] We further put emphasis on having a consistent overall strategy for tuning our model, whereby most parameters are set to published values and changes among different model versions and resolutions are kept at a minimum. Although this may mean that sub-optimal parameter settings are used in the short term,

we believe that this strategy increases the utility of the model (for instance when results are compared across different configurations), and may facilitate the long-term development of the model. If model parameters vary significantly from one model version to the next it is not easy to know from where deficiencies or improvements arise.

[13] The experiments presented below were conducted by modifying the Max Planck Institute Earth System Model at base-resolution (MPI-ESM-LR (M. Giorgetta et al., Climate variability and climate change in MPI-ESM CMIP5 simulations, manuscript in preparation, 2012)), which consists of ECHAM6 version 6.0, at T63 spectral resolution with 47 vertical levels (B. Stevens et al., The Atmospheric Component of the MPI-M Earth System Model: ECHAM6, manuscript in preparation, 2012), including the JSBACH land model (V. Brovkin et al., Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations, submitted to *Journal of Advances in Modeling Earth Systems*, 2012), coupled to the MPIOM ocean model at 1.5 degree resolution with 40 vertical levels (J. Jungclaus et al., MPIOM: Characteristics of the ocean simulations, manuscript in preparation, 2012).

2.1. The Tuning Process

[14] We tune the radiation balance with the main target to control the pre-industrial global mean temperature by balancing the TOA net longwave flux via the greenhouse effect and the TOA net shortwave flux via the albedo affect. The methodology of tuning the radiation balance may vary between model development groups, and is usually adapted to the specific goals and constraints of the exercise. After a problem has been identified in the coupled climate model, we iterate the following steps until a satisfactory solution is found:

- [15] 1. Short runs of single months, or if possible one or more years, with prescribed observed SST's and sea ice concentration; first with reference parameter settings, and then altered parameter settings.
- [16] 2. A longer simulation with altered parameter settings obtained in step 1 and observed SST's, currently 1976–2005 from the Atmospheric Model Intercomparison Project (AMIP), is compared with the observed climate.
- [17] 3. Implement the changes in the coupled model setup to run under pre-industrial conditions and evaluate the altered climate. Frequently, we make small parameter changes in this step to fine-tune the climate, without first revisiting steps 1 and 2.

[18] Our tuning process resembles the protocol described by *Gent et al.* [2011] as used in the preparation of CCSM4. They tune the individual model components first in uncoupled mode (step 2), and after coupling (step 3) they allow only changing one cloud parameter to adjust the radiation balance, and the sea ice albedo in order to adjust the Arctic sea ice volume. A somewhat different approach was taken by *Watanabe et al.* [2010] who tuned a number of parameters related to the cloud,

convection, turbulence, aerosol, and sea ice schemes iteratively every 5 years, while running their model (MIROC5) in coupled mode for about thousand years. Below we explain some of the most important parameters that we use to tune the radiation balance.

2.2. Cloud Processes and the Radiation Balance

[19] ECHAM6 predicts cloud fraction based on the relative humidity [*Sundqvist et al.*, 1989], distinguishes liquid and ice clouds [*Lohmann and Roeckner*, 1996], and accounts for vertical transport by shallow and deep convective clouds [*Tiedtke*, 1989; *Nordeng*, 1994]. The major uncertain climate-related cloud processes which are frequently used for tuning ECHAM are illustrated in Figure 2, some of which will be explored below, while Figure 3 shows the influence of five model parameters on globally averaged model properties, including the TOA net, longwave and shortwave fluxes, cloud cover, cloud liquid water- and water vapor paths.

2.2.1. Cloud Inhomogeneity

[20] Most climate models represent clouds with their fractional coverage in each grid cell. One can picture them as ‘cloud-boxes’ stacked vertically, under assumptions of how they overlap spatially to yield the total cloud cover. Each cloud-box contains cloud liquid and/or ice and we usually have no information on how that condensate is distributed horizontally. The radiative properties of clouds depend non-linearly on their thickness: A twice as thick cloud is less than twice as reflective, all other things being equal, and only very thin clouds do not behave as nearly ideal black body absorbers in the infrared spectrum. In reality, cloud thickness varies significantly on the spatial scale of a typical climate model grid cell, and therefore the mean cloud radiative effect is smaller than the radiative effect of the mean model cloud. In ECHAM6 the effect of cloud inhomogeneity is modeled by multiplying the cloud liquid and ice contents by respective homogeneity factors before radiation is calculated [*Cahalan et al.*, 1994]. A perfectly homogeneous cloud has a homogeneity factor of unity, while inhomogeneous clouds have factors less than one. At increasing resolution the homogeneity factor should in principle increase, as one begins to resolve some of the cloud inhomogeneities. There is some evidence that the inhomogeneity parameters can be replaced by making assumptions about the sub-grid scale distribution of cloud water [*Tompkins*, 2002] and incorporating this information into radiation calculations [*Pincus et al.*, 2006]. Figure 3, right column, shows how the homogeneity factor for liquid clouds barely influences any of the global quantities that we monitor, but the TOA net shortwave flux and thereby the TOA imbalance. This makes the parameter convenient as a tuning parameter for closing the radiation balance. The cloud homogeneity factor for liquid clouds was used as the only parameter by *Hazeleger et al.* [2012] to tune the EC-EARTH model.

2.2.2. Moist Shallow Convective Processes

[21] Vertical transport by convective updrafts, and sometimes downdrafts, is parameterized in global climate models mostly by means of mass-flux schemes of

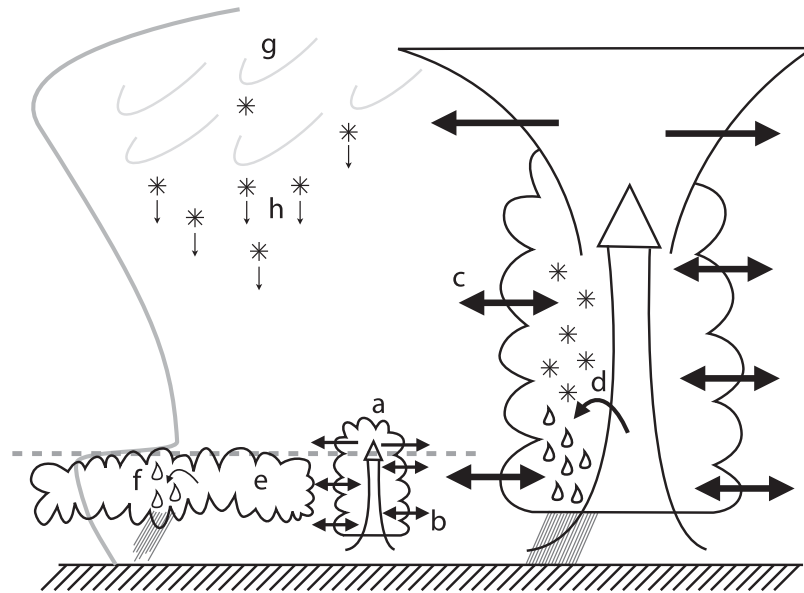


Figure 2. Illustration of the major uncertain climate-related cloud processes frequently used to tune the climate of the ECHAM model. Stratiform liquid and ice clouds, and shallow and deep convective clouds are represented. The grey curve to the left represents tropospheric temperatures and the dashed line is the top of the boundary layer. Parameters are a) convective cloud mass-flux above the level of non-buoyancy, b) shallow convective cloud lateral entrainment rate, c) deep convective cloud lateral entrainment rate, d) convective cloud water conversion rate to rain, e) liquid cloud homogeneity, f) liquid cloud water conversion rate to rain, g) ice cloud homogeneity, and h) ice particle fall velocity.

varying complexity. In essence, these schemes diagnose the mass-flux of the updrafts, the entrainment of air into the updraft from the surroundings and detrainment of air out of the updraft as functions of height. The three are connected by mass conservation to provide two independent updraft properties. More mass-flux leads naturally to more transport, while the role of entrainment is more complicated. Typically, more entrainment will act to reduce the buoyancy of the updraft making the convection less vigorous and thereby less efficient. In ECHAM6 a modified version of the Tiedtke convection-scheme [Tiedtke, 1989; Nordeng, 1994] is applied, which distinguishes between shallow, deep and mid-level convection in its formulation (Figure 2). This scheme uses a single updraft to effectively represent the real-world spectrum of convective cloud updrafts of varying sizes and strengths.

[22] With respect to the radiation balance, the main effect of shallow convection is to export cloud water from the boundary layer to the free atmosphere where it tends to evaporate, thereby reducing the thickness and extent of the boundary-layer clouds. The strength of this process is influenced mainly by two model parameters. The convective mass-flux above the level of non-buoyancy (leftmost column of Figure 3) is representing the most vigorous fraction of the updraft ensemble that ‘overshoots’ from the level where the mean updraft loses its buoyancy, to the next model level. The overshooting parameterization is conceptually unsatisfactory, and in the future we hope to replace it by a formulation involving the vertical updraft velocity. Increasing the parameter leads to less, and thinner boundary-layer

clouds, which increases surface temperature because more sunlight is absorbed by the system. Increasing instead the entrainment rate for shallow convection (second column of Figure 3) has the opposite effect on the cloud fields and the radiation balance; increased entrainment dilutes the updrafts, making them weaker and thereby more cloud liquid water is retained in the boundary-layer clouds.

2.2.3. Deep Convective Processes

[23] The parameterization of deep convection plays a more complex role in a climate model than shallow convection. Deep convective processes control basic features of the Tropical mean circulation, and are responsible for most of the Tropical rainfall. They are central to Tropical variability and help determine the vertical temperature structure in the Tropics.

[24] In relation to the radiation balance, the lateral entrainment rate for deep convection acts much like that for shallow convection, with the important difference that the low-level cloud-cover increase with increasing entrainment is to some extent compensated by loss of high-level ice clouds from the outflow of deep convection. There is also increasing amounts of water vapor with increasing entrainment rate, because more water is mixed into the free troposphere as less water rains directly out from the weakening updrafts. Associated with convective cloud water detrainment is a cooling of the upper troposphere due to evaporation on one hand, and radiative warming from the formation of cirrus clouds.

[25] Increasing the conversion rate of cloud water to rain in convective systems generally leads to less cloud

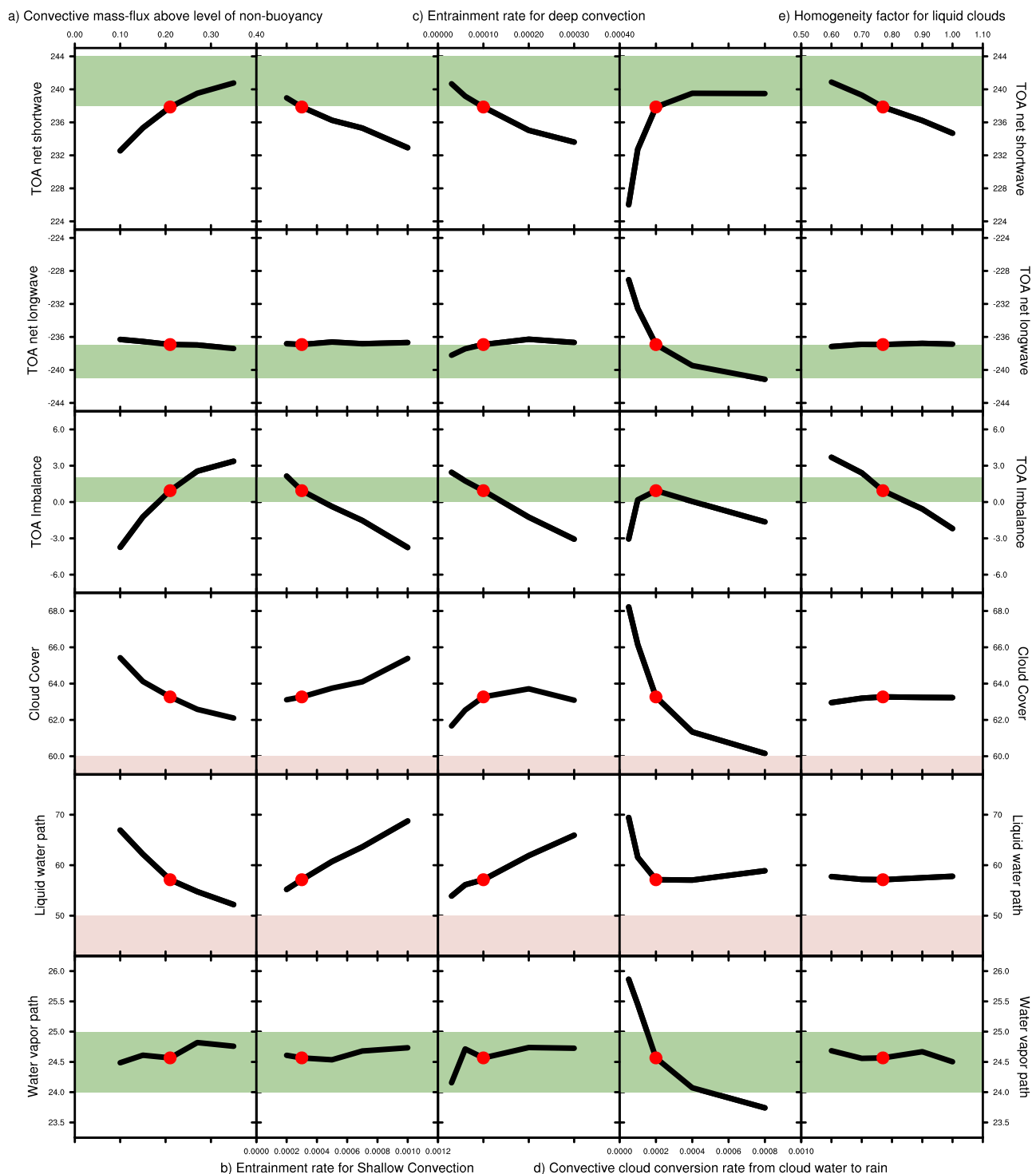


Figure 3. Overview of the influence of five major tuning parameters on various globally averaged model properties. Each run is one year (1976) with prescribed SST's and sea ice. The red dots show results for the default ECHAM6. Fluxes are positive downward. Green shadings show our acceptable ranges supported by observations, sometimes for the specific year, while pink shadings indicate values we find unacceptable. The range of $0\text{--}2\text{ Wm}^{-2}$ for TOA imbalance is typical for single-year simulations, while a narrower range would be required for a longer simulation. The water vapor path acceptable range is set lower for 1976 than for the full AMIP simulations (1976–2005), based on an increasing trend in reanalyses.

cover and less atmospheric water vapor, as more water is deposited directly from deep convective systems to the surface. By this parameter, both TOA net shortwave and net longwave fluxes both increase in magnitude, while approximately maintaining TOA radiation balance. This makes the parameter useful for adjusting the level of the TOA net shortwave and longwave fluxes.

2.3. Controlling the Global Mean Surface Temperature and Climate Drift

[26] A particular problem when tuning a coupled climate model is that it takes thousands of years for the deep ocean to be equilibrated. In many cases, it is not computationally feasible to redo such long simulations several times. Therefore it is valuable to estimate the equilibrium temperature with good precision long before equilibrium is actually reached. Ideally, one would like to think that if we tune our model to have a TOA radiation imbalance that closely matches the observed ocean heat uptake in simulations where SST's are prescribed to the present-day observed state with all relevant forcings applied, then the coupled climate model attains a global mean temperature in reasonable agreement with the observed. Recent studies suggest that the ocean heat uptake is of $0.5\text{--}0.7\text{ Wm}^{-2}$ when averaged over the

Earth's total surface area, indicating that the present-day climate is out of balance [Hansen *et al.*, 2011; Stevens and Schwartz, 2012]. There are at least three reasons why abiding to this ideal need not be successful:

- [27] 1. Climate models may not exactly conserve energy.
- [28] 2. The climate sensitivity of the model to the various forcings may not match the real climate system, and the forcings themselves may be erroneous.
- [29] 3. Local SST biases in the coupled model may influence the atmospheric state, for example cloudiness, and thereby shift the global mean temperature.

[30] To investigate whether climate models leak energy, Figure 4 shows the relation between TOA energy imbalance and global mean temperature for MPI-ESM-LR (blue) and the CMIP3 (light gray) and CMIP5 (gray) multi-model ensembles from control simulations of pre-industrial climate. Climate drift is indicated by the trails, and most models have fairly low drift during the typically 500-year long control runs. Some models drift considerably, up to 1 K. Models will relax slowly towards their equilibrium state approximately along slopes corresponding to their climate

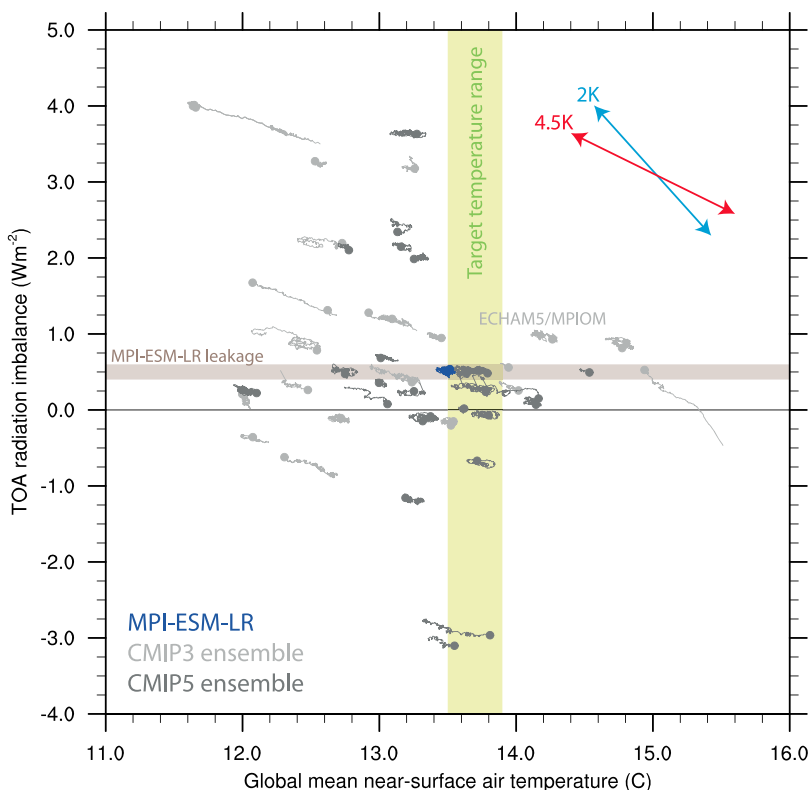


Figure 4. Drift in global mean temperature and radiation imbalance for pre-industrial control simulations. Blue is a 1000-year control simulation with MPI-ESM-LR, while light gray symbols are simulations from the CMIP3, including the predecessor ECHAM5/MPI-OM, and gray from the CMIP5 model ensembles. A large circle is the mean of the last 50 years of each simulation, while the trailing lines are 50-year running means for the entire simulations. Our target pre-industrial temperature range is marked as 13.7 ± 0.2 C. The blue and red arrows show the approximate slopes that models with a climate sensitivity to a CO_2 -doubling of 2 K and 4.5 K will follow. Note that the CMIP5 models displayed here are not exactly the same as those in Figure 1.

sensitivity, as indicated by the blue and red arrows [e.g., Gregory *et al.*, 2004]. Among the model simulations whose data were available at the time of this analysis, there is a tendency for drift in the CMIP5 models to be less pronounced than in some of the CMIP3 models, and there is a reduction in the number of warm and cold biased models in CMIP5. Only a few models are close to zero imbalance, or likely to relax to near-zero imbalance. If a model equilibrates at a positive radiation imbalance it indicates that it leaks energy, which appears to be the case in the majority of models, and if the equilibrium balance is negative it means that the model has artificial energy sources. We speculate that the fact that the bulk of models exhibit positive TOA radiation imbalances, and at the same time are cold-biased, is due to them having been tuned without account for energy leakage.

[31] We investigated the leakage of energy in MPI-ESM-LR of about 0.5 Wm^{-2} and found that it arises for the most part from mismatching grids and coastlines between the atmosphere and ocean model components. Further, some energy is lost due to an inconsistent treatment of the temperature of precipitation and river runoff into the ocean, and a small leakage of about 0.05 Wm^{-2} occurs in a not yet identified part of the atmosphere. When run with prescribed SST's and sea ice during present-day conditions and forcings (1976–2005), ECHAM6 has a TOA imbalance of 0.53 Wm^{-2} , which is barely enough to compensate the coupled models energy leakage. This would indicate that the model should be too cold when run in coupled mode because it is effectively well below the present-day observed ocean heat uptake ($0.5\text{--}0.7 \text{ Wm}^{-2}$). Yet, the coupled model arrives relatively close to the pre-industrial temperature. We shall see more examples of this behavior in Section 3.

[32] For these reasons we fine-tune the global mean temperature in coupled mode. Once we know the energy leakage of the coupled climate model system, it is relatively easy to estimate the equilibrium temperature based on short simulations. Let R be the TOA net radiation imbalance, L the model energy leakage, T the current global mean temperature, and $\lambda = \partial R / \partial T$ the climate feedback factor, then the equilibrium temperature (T_{eq}) can be derived from the energy balance equation:

$$R - L = \lambda(T - T_{eq}) \quad (1)$$

The imbalance and global mean temperature should be averaged over a long period, at least 10 years, sometimes longer, while the feedback can be estimated by regression, if it is not known *a priori*. If thereby the equilibrium temperature is outside the target range, measures can be taken to adjust the radiation balance as described in Section 2.1. In our experience it is possible to successfully tune our model's equilibrium temperature in about hundred years of coupled simulation, well before equilibrium is reached.

2.4. Tuning the Arctic Sea Ice

[33] The decline of Arctic sea ice extent in recent decades has certainly caught the attention of the scientific

community, as well as the public, and in the past it has been difficult for models to simulate this decline [e.g., Stroeve *et al.*, 2007; Rampal *et al.*, 2011]. As argued in the introduction, a key to simulating the evolution of the observed sea ice extent is to have a reasonable surface temperature, while also the mean thickness of the sea ice and the parameterization of surface albedo are important factors determining the susceptibility to external forcing [e.g., Holland and Bitz, 2003]. Unfortunately, sea ice thickness is a challenging quantity to observe; most of our knowledge is based on submarine records from the central parts of the Arctic and only recently has it been possible to obtain sporadic observations from satellite altimetry [Kwok and Rothrock, 2009]. Based on what we currently know, we aim at having an annual mean Arctic ice volume of about $20\text{--}25 \cdot 10^{12} \text{ m}^3$ in our pre-industrial climate, which corresponds to a mean sea ice thickness of about 2–2.5 m in the Northern Hemisphere given a sea ice area of about $10 \cdot 10^{12} \text{ m}^2$.

[34] In the past, the parameterization of snow and sea ice albedo was often used to tune the sea ice volume. Eisenman *et al.* [2007] argue that sea ice thickness is very sensitive to even small changes made to the model ice albedo using an energy balance model, while DeWeaver *et al.* [2008] show that in their fully coupled climate model several compensating processes limit the effectiveness of tuning sea ice with albedo. Either way, detailed and accurate observations of the snow and sea ice albedos are now available, and in ECHAM6 we apply a scheme that is better constrained by these empirical observations, including processes such as aging of snow and a representation of melt ponds on the top of sea ice [Pedersen *et al.*, 2009; E. Roeckner *et al.*, Impact of melt ponds on Arctic sea ice in past and future climates as simulated by MPI-ESM, manuscript in preparation, 2012]. For these reasons, we have for now abandoned the strategy of tuning the sea ice with the surface albedo parameters.

[35] The sea ice in the MPI-ESM-LR model is represented by a fractional coverage and a mean sea ice thickness in every grid-cell [Hibler, 1979]. As new ice is formed it is not readily known if the new ice primarily acts to thicken the existing ice, or if it mainly increases the fraction of the grid cell that is ice covered. Likewise, as the sea ice melts, it is not known if it does so from the top and bottom, or from the sides. In MPI-ESM-LR the geometry of melting and freezing processes are controlled by two non-dimensional parameters, c_{melt} and c_{freeze} , which can be varied between zero and one.

[36] Changing these two parameters has little impact on Antarctic sea ice (Figure 5). In the Arctic, varying c_{freeze} up and down from the MPI-ESM-LR model default value (2/3) allows changing the sea ice volume moderately by $\pm 10^{12} \text{ m}^3$. Increasing c_{freeze} permits more open ocean to exist during freeze-up, which enhances the ocean heat-loss to the atmosphere and thereby allows more sea ice to form. This is because even a thin layer of ice is effective in insulating the upper ocean, thereby reducing heat loss and inhibiting further sea ice formation. Even though the process is only effective in the beginning of the winter, the signature is

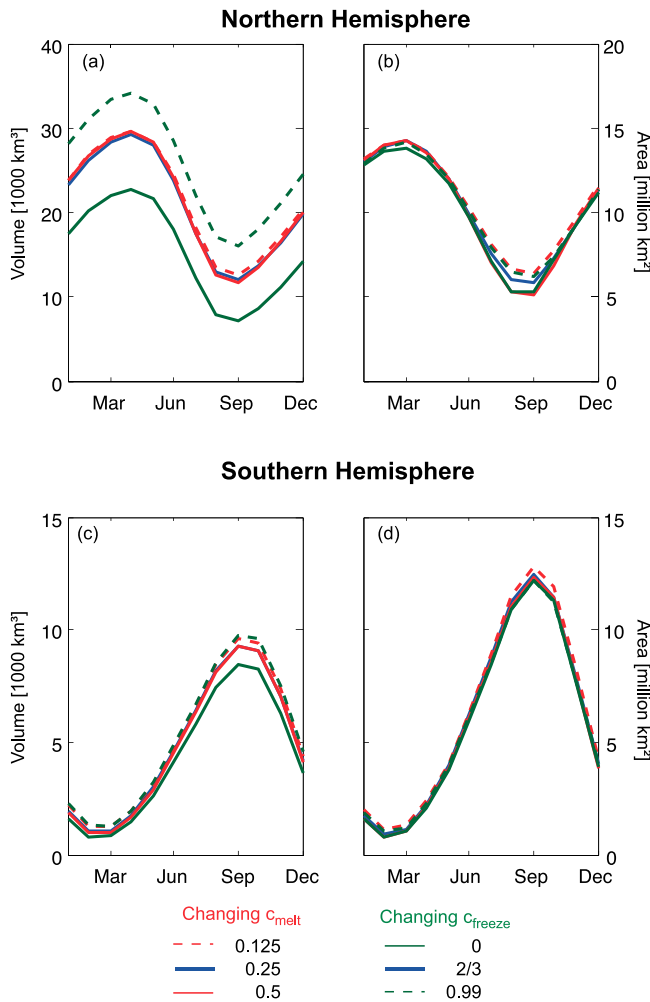


Figure 5. Seasonal cycles of sea ice volume and area for the Northern and Southern Hemispheres in different experiments with different geometries of sea ice melt and formation. The simulations are conducted with pre-industrial forcings.

seen in the ice volume throughout the year, while there is almost no impact on the sea ice area. The c_{melt} parameter had a small impact on sea ice area during the Arctic melt-season, but no impact in other seasons, nor on the sea ice volume.

2.5. The Final Tuning of MPI-ESM-LR for Participation in CMIP5

[37] The model development and final parameter tuning process leading up to the new MPI-ESM-LR model was targeted at deficiencies identified in the predecessor coupled climate model ECHAM5/MPIOM, here we shall focus on the parts where tuning played a central role: The pre-industrial control run global mean temperature was warm-biased at about 14.3°C , and the control simulation exhibited a weak drift of about 0.2 K in 500 years (Figure 4), while at the same time the model had too thick Arctic sea ice with a maximum that was centered near the North pole, rather than north of the Canadian Archipelago as is observed. Further, the radiation balance

of ECHAM5 was tuned to be close to the ERBE satellite estimates of OLR, while newer CERES estimates indicate a higher level of OLR.

[38] ECHAM6 incorporates a range of structural model improvements over ECHAM5, including increased atmospheric vertical resolution better representing the stratosphere, a new formulation of the snow and ice albedo, a minor change to the convection excess buoyancy, updated ozone and aerosol climatologies, an improved coupling between the atmosphere and land-surface, and several bug fixes. After these changes were introduced, the first parameter change was a reduction in two non-dimensional parameters controlling the strength of orographic wave drag from 0.7 to 0.5. This greatly reduced the low zonal mean wind- and sea-level pressure biases in the Northern Hemisphere in atmosphere-only simulations, and further had a positive impact on the global to Arctic temperature gradient and made the distribution of Arctic sea-ice far more realistic when run in coupled mode. In a second step the conversion rate of cloud water to rain in convective clouds (Section 2.2.3), was doubled from $1.0 \cdot 10^{-4}\text{ s}^{-1}$ to $2.0 \cdot 10^{-4}\text{ s}^{-1}$ in order to raise the OLR to be closer to the CERES satellite estimates.

[39] At this point it was clear that the new coupled model was too warm compared to our target pre-industrial temperature. Different measures using the convection entrainment rates, convection overshooting fraction and the cloud homogeneity factors were tested to reduce the global mean temperature. In the end, it was decided to use primarily an increased homogeneity factor for liquid clouds from 0.70 to 0.77 combined with a slight reduction of the convective overshooting fraction from 0.22 to 0.21, thereby making low-level clouds more reflective to reduce the surface temperature bias. Now the global mean temperature was sufficiently close to our target value and drift was very weak. At this point we decided to increase the Arctic sea ice volume from $18 \cdot 10^{12}\text{ m}^3$ to $22 \cdot 10^{12}\text{ m}^3$ by raising the c_{freeze} parameter from $1/2$ to $2/3$. ECHAM5/MPIOM had this parameter set to $4/5$. These three final parameter settings were done while running the model in coupled mode.

3. Parallel Worlds

[40] Usually there are multiple routes to reach a certain goal during the tuning process, even within the rather limited parameter set that we use. In building MPI-ESM-LR we chose one such route. To subsequently explore how this choice influenced our results, we have produced three alternatively tuned models here named World 1, 2 and 3. Each alternative World was created by first perturbing one parameter, and then adjusting one other parameter until the TOA imbalance and liquid water path are again close to that of the default model. This was done by iterating the first two steps in the tuning process outlined in the previous section. For the purpose of demonstration, this process was slightly simplified over the normal tuning procedure, which would involve using more parameters and observational targets, and long simulations in coupled mode. The alternative worlds were motivated by pre-

vious studies by *Klocke et al.* [2011], *Bender* [2008] and *Stainforth et al.* [2005]:

- [41] 1. *Klocke et al.* [2011] studied a perturbed physics ensemble using ECHAM5, and found that the parameter controlling the convective mass-flux above the level of non-buoyancy explained most of the variability in the climate sensitivity within their ensemble. Here we increase this parameter by about 50 percent, and, as in their study we also increase the shallow convection entrainment rate to compensate the loss of cloudiness. The resulting entrainment rate is closer to estimates from large-eddy simulations by *Siebesma and Cuijpers* [1995]. Extending the results of *Klocke et al.* [2011], World 1 should have had a climate sensitivity to a doubling of CO₂ increased from 3 K to 4–5 K, which, as we shall see, turned out not to be the case.
- [42] 2. *Bender* [2008] lowered the planetary albedo of the CAM3 climate model, from close to the ERBE satellite dataset to instead resemble the CERES data. This was done, partly, by increasing a parameter analogous to the convective cloud conversion rate from cloud water to rain (see Figure 3). *Bender* [2008] found only a small increase from 2.26 K to 2.50 K of the models climate sensitivity to a doubling of CO₂. In World 2 we lower the

planetary albedo by about 1 percent by doubling the conversion rate, and compensate the resulting shift in the TOA imbalance by lowering the liquid cloud homogeneity parameter.

- [43] 3. *Stainforth et al.* [2005] created a large perturbed physics ensemble based on the HadAM3 climate model, finding climate sensitivities to a doubling of CO₂ ranging between 2 and 11 K. They found that the ensemble members with sensitivities above 8 K were all related to perturbations of the convective cloud entrainment rate. In World 3 we lower the entrainment rate for deep convection to less than one third of ECHAM6's standard value. As in World 2 we compensate the radiation balance with the liquid cloud homogeneity parameter.

[44] Table 1 shows the final parameter settings along with key global mean properties from simulations with prescribed sea surface temperature and sea ice concentration. The tabulated simulations have prescribed historical SST's and sea ice concentrations (AMIP). Next, we shall couple these alternative worlds to the ocean model component, then assess the alternative models performance in representing the mean state and the Tropical variability, and finally we shall study climate sensitivity to a doubling of CO₂ when coupled to a mixed-layer ocean.

Table 1. Overview of Parameter Settings and Global Mean Properties of the Standard ECHAM6 and Three Alternative Models

Parameters		ECHAM6	World 1	World 2	World 3
Cloud mass-flux above level of non-buoyancy		0.21	0.32		
Entrainment rate for shallow convection [m ⁻¹]		3.0·10 ⁻⁴	8.0·10 ⁻⁴		
Entrainment rate for deep convection [m ⁻¹]		1.0·10 ⁻⁴			0.3·10 ⁻⁴
Conversion rate to rain in convective clouds [s ⁻¹]		2.0·10 ⁻⁴		4.0·10 ⁻⁴	
Homogeneity of liquid clouds		0.77		0.70	0.92
Properties	Observed	ECHAM6	World 1	World 2	World 3
Total cloud cover [percent]		63.4	62.9	61.4	61.6
Water vapor path [kg/m ²]		25.3	25.7	24.8	24.9
Liquid water path [g/m ²]		59.3	60.4	59.1	55.1
Ice water path [g/m ²]		31.1	31.4	28.4	27.8
Total precipitation [mm/d]		2.97	2.99	3.03	2.99
Surface downwelling shortwave [W/m ²]	175–190	186.0	186.4	189.4	187.5
Surface downwelling longwave [W/m ²]	337–354	343.9	344.3	342.9	342.8
Surface net shortwave [W/m ²]	152–167	161.2	161.5	164.2	162.5
Surface net longwave [W/m ²]	–(40–57)	–54.9	–54.9	–55.7	–55.6
Surface sensible heat flux [W/m ²]	–(16–19)	–19.7	–19.4	–19.8	–19.6
Surface latent heat flux [W/m ²]	–(75–87)	–85.8	–86.3	–87.7	–86.5
Energy to melt snow [W/m ²]		–0.21	–0.20	–0.21	–0.20
Greenhouse effect [W/m ²]		161.9	161.9	158.9	160.3
Reflected shortwave TOA [W/m ²]		–103.0	–102.5	–100.1	–101.7
Planetary albedo [percent]		30.2	30.1	29.4	29.9
Shortwave cloud radiative effect at TOA [W/m ²]		–47.4	–47.5	–45.0	–46.5
Longwave cloud radiative effect at TOA [W/m ²]		24.9	24.8	22.6	23.7
Shortwave net at TOA [W/m ²]	238–244	237.4	237.9	240.3	238.8
Longwave net at TOA [W/m ²]	–(237–241)	–236.9	–237.3	–239.6	–238.1
Imbalance at TOA [W/m ²]		0.53	0.58	0.64	0.65

^aValues are for AMIP runs with prescribed SSTs and sea-ice evolution, averaged over the years 1976–2005. Empty fields means a parameter is set to default. Reported observed values are from *Stevens and Schwartz* [2012].

3.1. Coupled Model Climate

[45] After having tuned the models to provide similar TOA radiation imbalances in AMIP mode, we coupled the various versions of ECHAM6 to the MPIOM ocean component and simulated 200 years under pre-industrial conditions (Figure 6). The first interesting point is that the alternative worlds relatively quickly approach equilibrium global mean surface air temperatures, but the

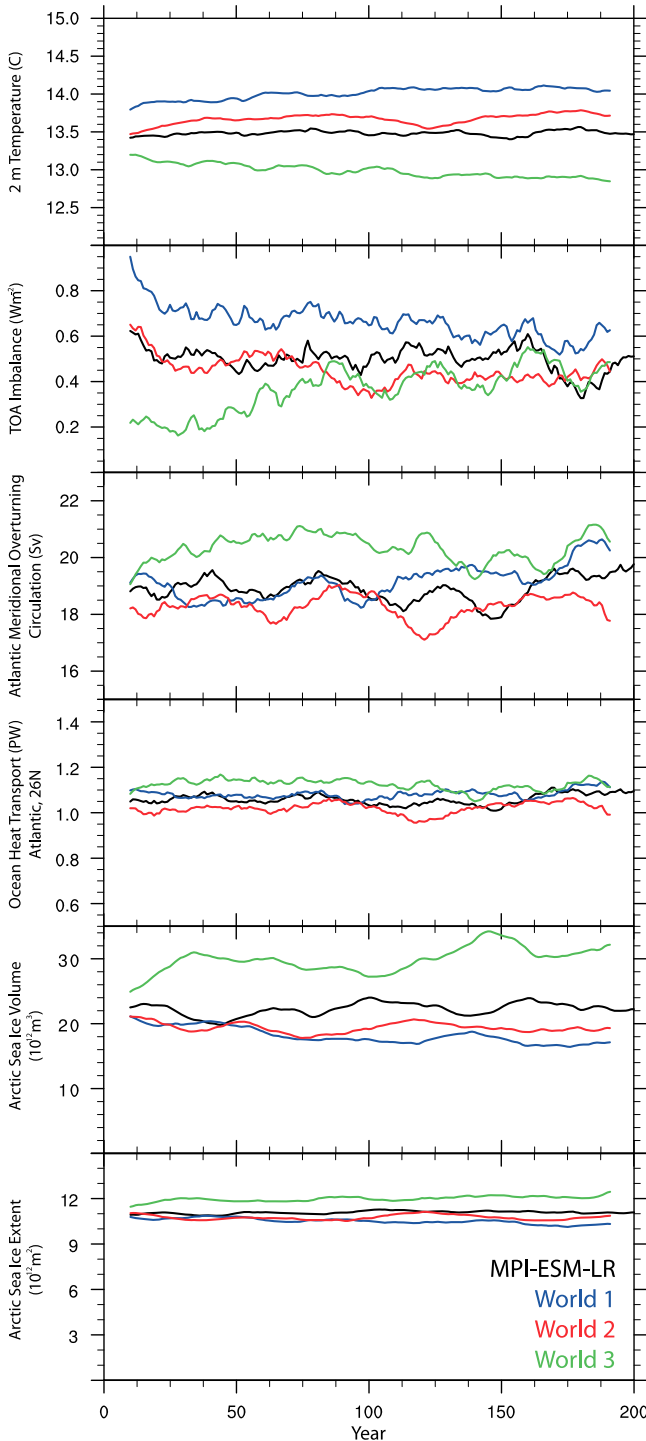


Figure 6. Coupled climate simulation time series. A 20-year running mean is applied to the time series.

surface temperature at which they equilibrate is very different, with more than one degree from the coldest (World 3) to the warmest (World 1). After 200 years World 1 is still warming, and we estimate it will warm further by 0.1 to 0.2 K, while the other two have come into balance with a model system energy leakage of about 0.5 Wm^{-2} (Section 2.3). This demonstrates the point made earlier with controlling climate drift in coupled mode, because the pre-industrial temperature cannot be predicted with sufficient precision using the AMIP simulations.

[46] The mean ocean circulation was only weakly impacted by the parameter changes studied here. A significant impact was found in World 3 in the Atlantic meridional overturning circulation which was raised by 1–2 Sv. The associated northward ocean energy transport likewise increased by 5–10 percent. The colder high latitudes lead to dense water formation, which increases the ocean deep convection. This response we often find during cooling transients, and usually the enhanced circulation relaxes as the model approaches equilibrium.

[47] The response of the annual mean Arctic sea ice volume and extent is correlated with the global mean temperature. In particular, the colder World 3 exhibits a rapid growth of sea ice volume to more than $30 \cdot 10^{12} \text{ m}^3$ towards the end of the simulation. In Section 2.4 we explained how we can tune the Arctic sea ice volume by adjusting geometric factors controlling the growth of ice. However, even with extreme settings it would not have been possible to reduce the sea ice volume in World 3 to our target value of about $20\text{--}25 \cdot 10^{12} \text{ m}^3$, hence the necessity of tuning the temperature in order to be able to represent sea ice.

3.2. Evaluating the Climate

[48] Biases relative to observations and reanalyses in our standard evaluation for the AMIP simulations with fixed SST's are remarkably similar, both in geographical and vertical structure and in magnitude. If anything, pressure biases in the Southern Hemisphere are slightly reduced in World 2 and 3, and upper tropospheric temperatures are best represented by World 1 and 3. During the development of MPI-ESM-LR, we regularly referred to a set of diagnostics proposed by Reichler and Kim [2008] to get an overview of the model performance and to quickly compare different model versions against each other. Following Stevens et al. (manuscript in preparation, 2012), a selection of these indices is shown in Figure 7. The indices are normalized relative to the average performance of the CMIP3 coupled model ensemble in representing the corresponding variables.

[49] It is no surprise that the AMIP simulations perform generally better than the coupled simulations in terms of representing the mean climate, as they are helped by the prescribed SST's and sea ice. Among the AMIP simulations World 1 is performing slightly worse than the rest, and it does so consistently across most of the individual indices, while the tables turn when inspecting the coupled simulations. The latter is in our

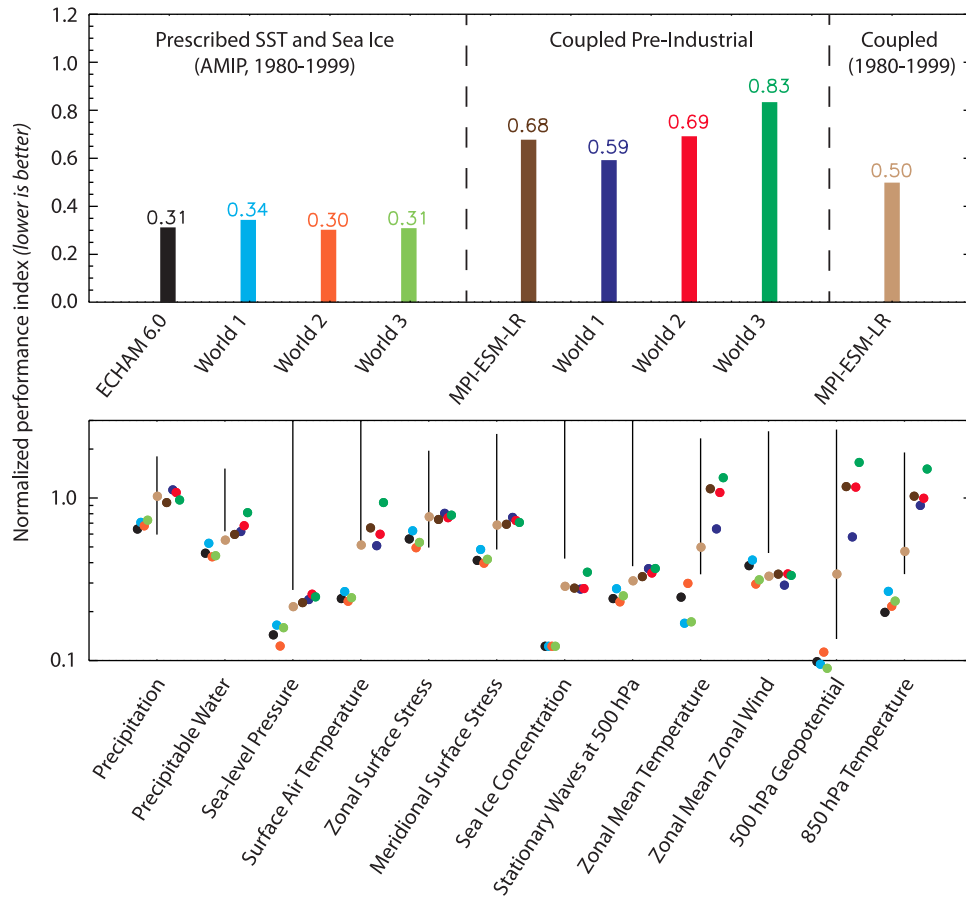


Figure 7. A selection of the global annual mean climate performance indices by *Reichler and Kim* [2008] - lower is better. The indices are normalized such that the average model of the CMIP3 coupled model ensemble is unity. (top) The mean of the indices for each model version in different configurations, and (bottom) the individual indices. In Figure 7 (bottom) the sea ice concentration performance of the AMIP simulations are all the same because sea ice is prescribed. Thin vertical bars in the lower panel shows the range of CMIP3 models.

experience mainly an artifact of comparing pre-industrial simulations to present-day observations combined with the World 1 coupled simulation being the warmest (Figure 6), thereby providing better scores in the temperature-sensitive indices. This temperature-effect can also be seen in the cold World 3 simulation, and when comparing MPI-ESM-LR scores for pre-industrial and present-day conditions. While the *Reichler and Kim* [2008]-diagnostics are certainly helpful to quickly get an overview of a models performance in representing the mean state climate, they do little to elucidate processes responsible for better or worse model performance.

[50] An interesting and challenging issue in MPI-ESM is the Tropical precipitation distribution over land versus ocean. The model prefers precipitating in the ocean, whereas observations indicate a stronger preference to precipitate on land. The problem is interesting because it impinges on our understanding of deep convective processes, and it is becoming a major issue with the inclusion of dynamic vegetation in the Earth system model as underestimated precipitation on land leads to biases in the vegetation which may ultimately degrade the representation of the carbon cycle [Collins *et al.*, 2011; Brovkin *et al.*, submitted manuscript, 2012].

[51] The alternative worlds exhibit significant differences in the distribution of precipitation over Maritime Southeast Asia (Figure 8). Whereas standard ECHAM6 and to an even larger extent World 1 prefer to precipitate over the ocean, World 2 and in particular World 3 do produce precipitation over the larger islands of Sumatra and Borneo, in better agreement with observations. Note that these detailed differences are not visible in the *Reichler and Kim* [2008] precipitation performance index.

3.3. Tropical Variability

[52] Tropical inter-annual variability is dominated by the El Niño-Southern Oscillation (ENSO), while intra-seasonal variability is dominated by the Madden and Julian Oscillation (MJO). Here we investigate the influence of the parameter changes on the model representation of these phenomena.

[53] Earlier versions of our climate model, e.g., ECHAM5/MPIOM, had a too pronounced ENSO variability, while the MPI-ESM-LR is much closer to the observations (Figure 9). ENSO was not significantly impacted by the changes made to convective parameters in World 1 and 3, but the variability was consistently

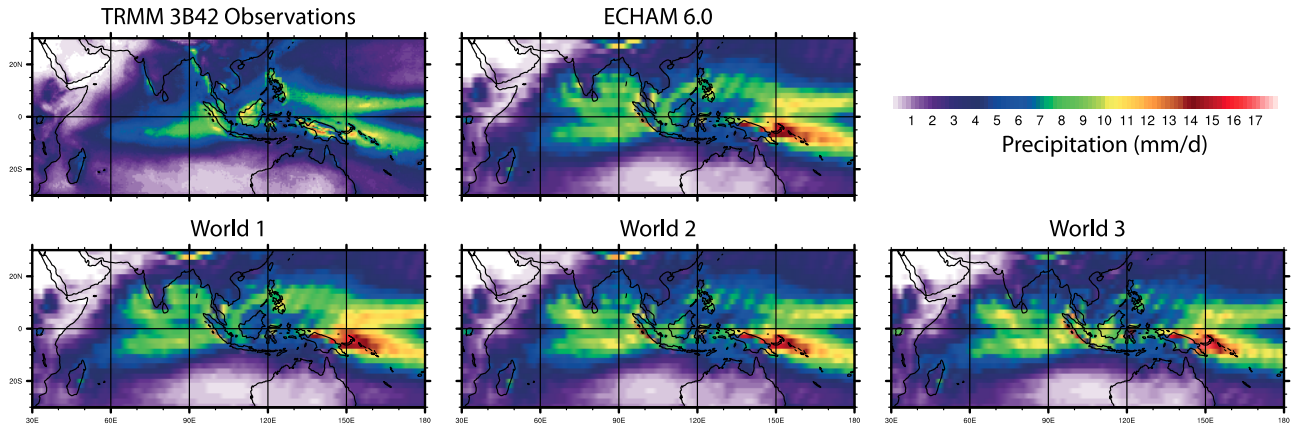


Figure 8. Annual mean precipitation over Maritime Southeast Asia in simulations with prescribed SST's (AMIP).

reduced in World 2, in which we doubled the convective cloud conversion rate to rain from $2 \cdot 10^{-4}$ to $4 \cdot 10^{-4}$. This is interesting because the same parameter was also doubled from ECHAM5/MPIOM to MPI-ESM-LR, from $1 \cdot 10^{-4}$ to $2 \cdot 10^{-4}$, in order to increase OLR in closer agreement with the more recent CERES satellite observations. This means that at least part of the improvement in ENSO could have been achieved simply by chance. Other features of ENSO, such as the shape of the frequency spectrum, annual cycle and the skewness were insensitive to the changes made in the alternative worlds. Other studies have shown sensitivity of ENSO characteristics to the representation of atmospheric convection [e.g., Neale *et al.*, 2008; Kim *et al.*, 2011a], although a direct connection to our findings is difficult to establish due to the different model formulations.

[54] Tropical intraseasonal variability, which is dominated by the MJO, is also poorly represented by most state-of-the-art GCMs. T. Crueger *et al.* (The Madden-Julian Oscillation in ECHAM6 and the introduction of a MJO metric, submitted to *Journal of Climate*, 2012) proposed to quantify the MJO in ECHAM by characterizing: 1) The strength of the convective signal and 2) the eastward propagation. The former is derived from a multivariate empirical orthogonal function analysis of 20–100 day band-pass filtered tropospheric winds and outgoing longwave radiation (OLR). OLR is used as an indicator of deep convective clouds and precipitation. The eastward propagation is derived from the relative strength of eastward- to westward propagating waves in this same band-pass filtered data set.

[55] MJO characteristics of the parallel worlds are shown in Figure 10. All model versions exhibit a weaker convective signal than observed (26%), and they all have less distinct eastward wave propagation than observed (3.5). The coupled model simulations (diamonds) show roughly twice the convective strength of the AMIP simulations, and an enhancement of the eastward propagation (Crueger *et al.*, submitted manuscript, 2012). In both cases, World 3 shows practically the same strength of westward and eastward propagating waves, in contrast to what one observes. The weak eastward propagation is also an issue in World 2. This is likely a consequence of the parameter changes making the deep

convection less sensitive to environmental conditions [Kim *et al.*, 2011a]. The two models, World 2 and 3, that perform the least in terms of MJO are the best in representing the distribution of tropical precipitation between land and ocean in Maritime Southeast Asia (Figure 8).

3.4. Climate Sensitivity

[56] Equilibrium climate sensitivity to a doubling of CO_2 is a standard measure of models' sensitivity to external forcing. We have performed simulations of the parallel worlds coupled to a 50 meter deep mixed-layer ocean with pre-industrial levels of CO_2 and with doubled CO_2 concentration. In the mixed-layer ocean, no currents are explicitly resolved, instead the ocean energy transports are prescribed as inferred from the AMIP simulations. The simulations are run for 50 years, while the presented differences are evaluated over the last 30 years of the simulations.

[57] The distributions of the temperature change exhibit significant similarities across the model versions (Figure 11). Warming over land is stronger than over the oceans, peaks at the poles, and is stronger in the Northern Hemisphere than in the Southern Hemisphere. In particular, the temperature response in World 2 closely resembles the standard ECHAM6 response. The zonally averaged temperature response (Figure 12) further quantifies the similarities among the model versions, and, World 3 does stand out with 0.5 to 1 K more warming than the others throughout the Tropics and Sub-tropics. In the Arctic all three parallel worlds exhibit slightly more warming than the standard ECHAM6 model.

[58] A simple analysis of how the parallel worlds approach equilibrium after the instantaneous CO_2 -doubling can provide some hints as to why the sensitivity is different among the models (Figure 13). Gregory *et al.* [2004] argue that the extrapolated intercept with the y-axis is a measure of the atmospheric adjusted forcing from the CO_2 -doubling, while the intersection with the x-axis is a good estimate of the equilibrium climate sensitivity. The slope is therefore related to the total climate system feedback. World 2 and 3 both have adjusted CO_2 -forcings which are 10–20 percent higher than the standard ECHAM6 model. This is related

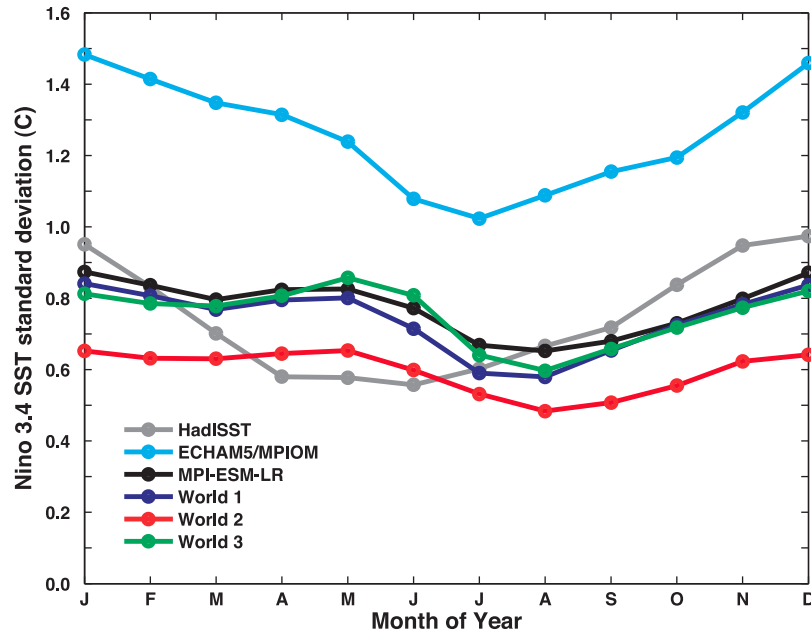


Figure 9. Annual cycle of ENSO variability exemplified by the Nino-3.4 sea surface temperature variability.

primarily to a fast reduction of cloudiness as a direct response to the increase in CO_2 [Gregory and Webb, 2008], which is more pronounced in World 2 and 3. The

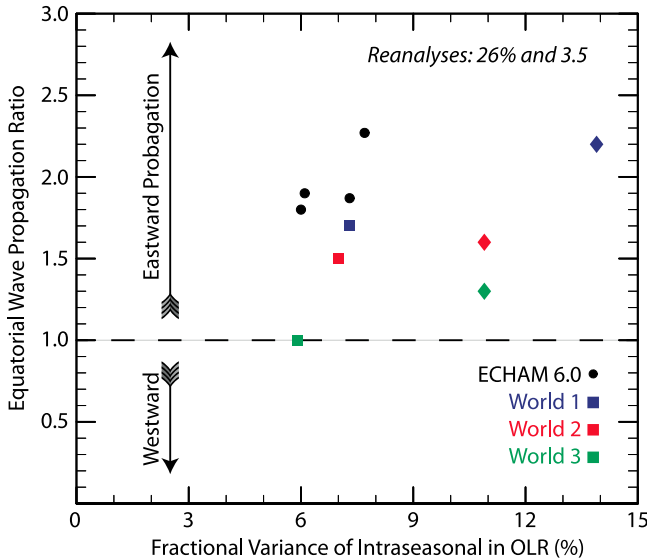


Figure 10. Intraseasonal tropical variability analysis. The four black circles are four realizations of the AMIP simulation with ECHAM6, squares are AMIP simulations with the parallel worlds, while the diamonds are coupled simulations. The x-axis is a fractional variance of the intraseasonal variability of OLR from the first two multivariate EOF of OLR and zonal winds at 850 hPa and 200 hPa, while the y-axis shows the relative strength of eastward to westward propagating Tropical waves (wavenumber 1–3). The underlying data is band pass filtered (20–100 days). Tabulated values for reanalyses are averaged from ERA-40, ERA-Interim and NCEP.

slope is steeper in World 1 and 2 than it is in World 3 and the standard ECHAM6 model. Together, this indicates that the higher sensitivity of World 3 relative to the standard model is mainly a consequence of a stronger adjustments to CO_2 -forcing, less so temperature-dependent feedbacks.

4. Discussion

[59] Parameter tuning is the last step in the climate model development cycle, and invariably involves making sequences of choices that influence the behavior of the model. Some of the behavioral changes are desirable, and even targeted, but others may be a side effect of the tuning. The choices we make naturally depend on our preconceptions, preferences and objectives. We choose to tune our model because the alternatives - to either drift away from the known climate state, or to introduce flux-corrections - are less attractive. Within the foreseeable future climate model tuning will continue to be necessary as the prospects of constraining the relevant unresolved processes with sufficient precision are not good.

[60] Climate model tuning has developed well beyond just controlling global mean temperature drift. Today, we tune several aspects of the models, including the extratropical wind- and pressure fields, sea-ice volume and to some extent cloud-field properties. By doing so we clearly run the risk of building the models' performance upon compensating errors, and the practice of tuning is partly masking these structural errors. As one continues to evaluate the models, sooner or later these compensating errors will become apparent, but the errors may prove tedious to rectify without jeopardizing other aspects of the model that have been adjusted to them. To aid the longterm development of our model we choose a tuning-strategy with only a small number of

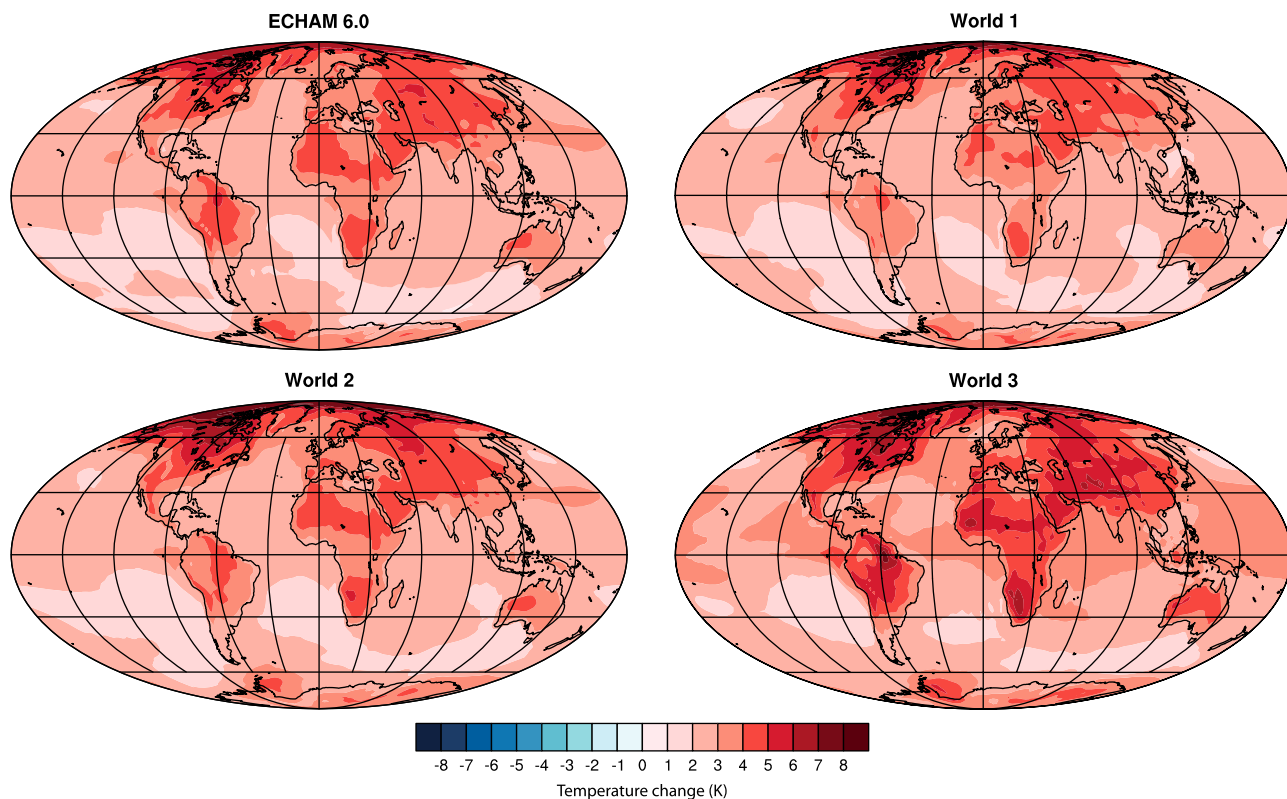


Figure 11. Maps of surface temperature change from a doubling of CO₂.

parameter changes between different model versions and resolutions, such that it will be easier to identify and understand how the model formulation can be improved.

[61] Often, we are confronted with trade-offs when tuning a coupled climate model. For example, two of the

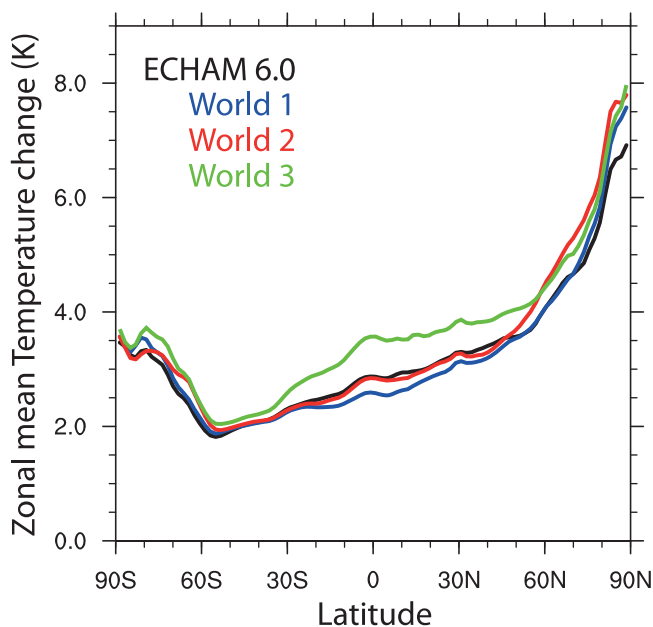


Figure 12. Zonally averaged surface temperature change from a doubling of CO₂.

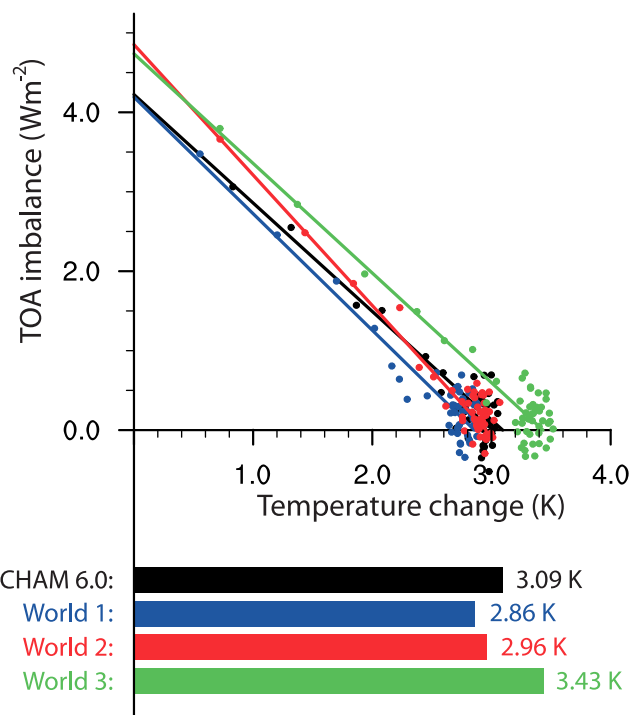


Figure 13. Global mean temperature versus TOA radiation imbalance after an instantaneous doubling of CO₂. Lower part shows the estimated equilibrium climate sensitivity based on the linear regression.

parallel worlds showed improvements in the precipitation distribution between land and ocean in the Tropics, but they both exhibit weakened modes of tropical variability. This particular trade-off is becoming increasingly difficult now that the model is coupled to interactive vegetation. The problem suggests that underlying structural errors in the model prohibit us from achieving fidelity with both observed aspects at the same time by means of parametric tuning alone. Thereby the practice helps us to identify the processes that could improve the model by the development of new or improved parameterizations, and as such is integral to the overall model development cycle [Jakob, 2010].

[62] The model tuning process at our institute is artisanal in character, in that both the adjustment of parameters at each tuning iteration and the evaluation of the resulting candidate models are done by hand, as is done at most other modeling centers. It is, however, at least conceptually possible to automate this process and find optimal sets of parameters with respect to certain targets. When considering model biases that appear on long time-scales (months to years), one option is to use the full model and search through parameter-space seeking areas in which errors are minimized [Jackson *et al.*, 2008; Järvinen *et al.*, 2010]. Alternatively, one can use a relatively small number of model runs to build a statistical model, or emulator, of the error as a function of parameter space to obtain parameter sets that minimize model error [Neelin *et al.*, 2010]. On the other hand, many of the parameters we adjust have relatively fast impacts on the model state, making it appealing to look at errors in very short (hours) forecasts starting from realistic initial conditions. If a data assimilation system is available for the model one can use the average change to the model introduced with each new set of observations as a measure of a specific model configuration's disagreement with observations [Rodwell and Palmer, 2007], and further, such a data-assimilation system can be exploited to find parameter sets that minimize short-term errors [Annan *et al.*, 2005; Järvinen *et al.*, 2012]. We are experimenting with several of these strategies, but have yet to find something significantly better or faster than tuning by hand, nor are we aware of any other climate modeling center that operates differently. We suspect that this relates to the difficulty in making unsupervised compromises: Any such objective tuning algorithm requires a subjective choice of a cost function and this involves weighting trade-offs against one another, which is difficult to do ahead of time.

[63] The climate sensitivity of the worlds we considered spanned only one quarter of the range of the climate sensitivity of the CMIP3 models. World 1 shows a small decrease, contrary to findings by Klocke *et al.* [2011], and the increase in climate sensitivity in World 3 is only modest compared to the very high climate sensitivities found by Stainforth *et al.* [2005]. Obviously, a perturbed physics ensemble with only four members is not going to reveal the full range of parameter-dependent uncertainty spanned in more systematic studies, and in particular certain combinations of parameters may yield larger impacts than found here. It is possible, however, that

perturbed physics studies, which out of necessity are performed on low-resolution climate models, could be extra sensitive to the changes of parameters. Tests with ECHAM do confirm that the lack of an increased climate sensitivity in World 1 was due to the increase from 19 to 47 vertical levels, relative to that used by Klocke *et al.* [2011]. This could indicate that complex interactions between the resolved and the parameterized model components can introduce erratic behavior at low resolutions.

[64] One of the few tests we can expose climate models to, is whether they are able to represent the observed temperature record from the dawn of industrialization until present. Models are surprisingly skillful in this respect [Räisänen, 2007], considering the large range in climate sensitivities among models - an ensemble behavior that has been attributed to a compensation with 20th century anthropogenic forcing [Kiehl, 2007]: Models that have a high climate sensitivity tend to have a weak total anthropogenic forcing, and vice-versa. A large part of the variability in inter-model spread in 20th century forcing was further found to originate in different aerosol forcings. It seems unlikely that the anti-correlation between forcing and sensitivity simply happened by chance. Rational explanations are that 1) either modelers somehow changed their climate sensitivities, 2) deliberately chose suitable forcings, or 3) that there exists an intrinsic compensation such that models with strong aerosol forcing also have a high climate sensitivity. Support for the latter is found in studies showing that parametric model tuning can influence the aerosol forcing [Lohmann and Ferrachat, 2010; Golaz *et al.*, 2011]. Understanding this complex is well beyond our scope, but it seems appropriate to linger for a moment at the question of whether we deliberately changed our model to better agree with the 20th century temperature record.

[65] The MPI-ESM was not tuned to better fit the 20th century. In fact, we only had the capability to run the full 20th Century simulation according to the CMIP5-protocol after the point in time when the model was frozen. Yet, we were in the fortunate situation that the MPI-ESM-LR performed acceptably in this respect, and we did have good reasons to believe this would be the case in advance because the predecessor was capable of doing so. During the development of MPI-ESM-LR we worked under the perception that two of our tuning parameters had an influence on the climate sensitivity, namely the convective cloud entrainment rate and the convective cloud mass flux above the level of non-buoyancy, so we decided to minimize changes relative to the previous model. The results presented here show that this perception was not correct as these parameters had only small impacts on the climate sensitivity of our model.

[66] Climate models ability to simulate the 20th century temperature increase with fidelity has become something of a show-stopper as a model unable to reproduce the 20th century would probably not see publication, and as such it has effectively lost its purpose as a model quality measure. Most other observational datasets sooner or later meet the same destiny, at least

beyond the first time they are applied for model evaluation. That is not to say that climate models can be readily adapted to fit any dataset, but once aware of the data we will compare with model output and invariably make decisions in the model development on the basis of the results. Rather, our confidence in the results provided by climate models is gained through the development of a fundamental physical understanding of the basic processes that create climate change. More than a century ago it was first realized that increasing the atmospheric CO₂ concentration leads to surface warming [Arrhenius, 1896], and today the underlying physics and feedback mechanisms are reasonably understood (while quantitative uncertainty in climate sensitivity is still large). Coupled climate models are just one of the tools applied in gaining this understanding [Oreskes et al., 1994; Bony et al., 2012].

[67] In this paper we have attempted to illustrate the tuning process, as it is being done currently at our institute. Our hope is to thereby help de-mystify the practice, and to demonstrate what can and cannot be achieved. The impacts of the alternative tunings presented were smaller than we thought they would be in advance of this study, which in many ways is reassuring. We must emphasize that our paper presents only a small glimpse at the actual development and evaluation involved in preparing a comprehensive coupled climate model - a process that continues to evolve as new datasets emerge, model parameterizations improve, additional computational resources become available, as our interests, perceptions and objectives shift, and as we learn more about our model and the climate system itself.

[68] **Acknowledgments.** We are grateful for constructive comments and suggestions from Jin-Song von Storch, Marc Salzmann, Frida Bender, Akos Horváth, Christian Jakob and an anonymous reviewer. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for making available the CMIP3 and CMIP5 multi-model datasets. The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7/2007–2013) under grant agreement 244067. This work was supported by the Max Planck Gesellschaft (MPG), and computational resources were provided by Deutsche Klima Rechen Zentrum (DKRZ).

References

- Annan, J. D., D. J. Lunt, J. C. Hargreaves, and P. J. Valdes (2005), Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter, *Nonlinear Processes Geophys.*, *12*, 363–371, doi:10.5194/npg-12-363-2005.
- Arrhenius, S. (1896), On the influence of carbonic acid in the air upon the temperature of the ground, *Philos. Mag. J. Sci.*, *41*, 237–275, doi:10.1080/14786449608620846.
- Bender, F. A.-M. (2008), A note on the effect of GCM tuning on climate sensitivity, *Environ. Res. Lett.*, *3*, 014001, doi:10.1088/1748-9326/3/1/014001.
- Bony, S., B. Stevens, I. Held, J. Mitchell, J.-L. Dufresne, K. Emanuel, P. Friedlingstein, S. Griffies, and C. Senior (2012), Carbon dioxide and climate: Perspectives on scientific assessment, report, World Meteorol. Organ., Geneva, Switzerland.
- Boville, B. A., and P. R. Gent (1998), The NCAR climate system model, version one, *J. Clim.*, *11*, 1115–1130, doi:10.1175/1520-0442(1998)011<1115:TNCMSV>2.0.CO;2.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.
- Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider (1994), The albedo of fractal stratocumulus clouds, *J. Atmos. Sci.*, *51*(16), 2434–2455, doi:10.1175/1520-0469(1994)051<2434:TAOFSC>2.0.CO;2.
- Collins, W. J., et al. (2011), Development and evaluation of an Earth-system model—HadGEM2, *Geosci. Model Dev.*, *4*, 1051–1075, doi:10.5194/gmd-4-1051-2011.
- Colman, R. A., S. B. Power, B. J. McAvaney, and R. R. Dahni (1995), A non-flux corrected transient CO₂ experiment using the BMRC coupled atmosphere/ocean GCM, *Geophys. Res. Lett.*, *22*, 3047–3050, doi:10.1029/95GL01727.
- DeWeaver, E. T., E. C. Hunke, and M. M. Holland (2008), Comment on “On the reliability of simulated Arctic sea ice in global climate models” by I. Eisenman, N. Untersteiner, and J. S. Wettlaufer, *Geophys. Res. Lett.*, *35*, L04501, doi:10.1029/2007GL031325.
- Donner, L. J., et al. (2011), The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3, *J. Clim.*, *24*, 3484–3519, doi:10.1175/2011JCLI3955.1.
- Eisenman, I., N. Untersteiner, and J. S. Wettlaufer (2007), On the reliability of simulated Arctic sea ice in global climate models, *Geophys. Res. Lett.*, *34*, L10501, doi:10.1029/2007GL029914.
- Gent, P. R., et al. (2011), The Community Climate System Model Version 4, *J. Clim.*, *24*, 4973–4991, doi:10.1175/2011JCLI4083.1.
- Golaz, J.-C., M. Salzmann, L. J. Donner, L. W. Horowitz, Y. Ming, and M. Zhao (2011), Sensitivity of the aerosol indirect effect to subgrid variability in the cloud parameterization of the GFDL atmosphere general circulation model AM3, *J. Clim.*, *24*, 3145–3160, doi:10.1175/2010JCLI3945.1.
- 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, doi:10.1007/s003820050010.
- Gregory, J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, J. A. Lowe, T. C. Johns, and K. D. Williams (2004), A new method for diagnosing radiative forcing and climate sensitivity, *Geophys. Res. Lett.*, *31*, L03205, doi:10.1029/2003GL018747.
- Gregory, J., and M. Webb (2008), Tropospheric adjustment induces a cloud component in CO₂ forcing, *J. Clim.*, *21*, 58–71, doi:10.1175/2007JCLI1834.1.
- Guilyardi, E., and G. Madec (1997), Performance of the OPA/ARPEGE-T21 global ocean-atmosphere coupled model, *Clim. Dyn.*, *13*, 149–165, doi:10.1007/s003820050157.
- HadGEM2 Development Team (2011), The HadGEM2 family of Met Office Unified Model climate configurations, *Geosci. Model Dev.*, *4*, 723–757, doi:10.5194/gmd-4-723-2011.
- Hansen, J., M. Sato, P. Kharecha, and K. von Schuckmann (2011), Earth's energy imbalance and implications, *Atmos. Chem. Phys.*, *11*, 13421–13449, doi:10.5194/acp-11-13421-2011.
- Hazeleger, W., et al. (2012), EC-Earth V2.2: Description and validation of a new seamless earth system prediction model, *Clim. Dyn.*, doi:10.1007/s00382-011-1228-5, in press.
- Hibler, W. D. (1979), Dynamic thermodynamic sea ice model, *J. Phys. Oceanogr.*, *9*, 815–846, doi:10.1175/1520-0485(1979)009<0815:ADTSIM>2.0.CO;2.
- Holland, M. M., and C. M. Bitz (2003), Polar amplification of climate change in coupled models, *Clim. Dyn.*, *21*, 221–232, doi:10.1007/s00382-003-0332-6.
- Hourdin, F., et al. (2012), Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model, *Clim. Dyn.*, in press.
- Intergovernmental Panel on Climate Change (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., Cambridge Univ. Press, Cambridge, U. K.
- Jackson, C. S., K. M. Sen, G. Huerta, Y. Deng, and K. P. Bowman (2008), Error reduction and convergence in climate prediction, *J. Clim.*, *21*, 6698–6709, doi:10.1175/2008JCLI2112.1.
- Jakob, C. (2010), Accelerating progress in global atmospheric model development through improved parameterizations, *Bull. Am. Meteorol. Soc.*, *91*, 869–875, doi:10.1175/2009BAMS2898.1.
- 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, 9993–10002, doi:10.5194/acp-10-9993-2010.
- Järvinen, H., M. Laine, A. Solonen, and H. Haario (2012), Ensemble prediction and parameter estimation system: The concept, *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.923, in press.
- Kiehl, J. T. (2007), Twentieth century climate model response and climate sensitivity, *Geophys. Res. Lett.*, 34, L22710, doi:10.1029/2007GL031383.
- Kim, D., Y.-S. Jang, D.-H. Kim, Y.-H. Kim, M. Watanabe, F.-F. Jin, and J.-S. Kug (2011a), El Niño–Southern Oscillation sensitivity to cumulus entrainment in a coupled general circulation model, *J. Geophys. Res.*, 116, D22112, doi:10.1029/2011JD016526.
- Klocke, D., R. Pincus, and J. Quaas (2011), On constraining estimates of climate sensitivity with present-day observations through model weighting, *J. Clim.*, 24(23), 6092–6099, doi:10.1175/2011JCLI4193.1.
- Kwok, R., and D. A. Rothrock (2009), Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008, *Geophys. Res. Lett.*, 36, L15501, doi:10.1029/2009GL039035.
- Lohmann, U., and S. Ferrachat (2010), Impact of parameter uncertainties on the present-day climate and on the anthropogenic aerosol effect, *Atmos. Chem. Phys.*, 10, 11373–11383, doi:10.5194/acp-10-11373-2010.
- Lohmann, U., and E. Roeckner (1996), Design and performance of a new cloud microphysical scheme developed for the ECHAM general circulation model, *Clim. Dyn.*, 12, 557–572, doi:10.1007/BF00207939.
- Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor (2007), The WCRP CMIP3 multi-model dataset: A new era in climate change research, *Bull. Am. Meteorol. Soc.*, 88, 1383–1394, doi:10.1175/BAMS-88-9-1383.
- Neale, R. B., J. H. Richter, and M. Jochum (2008), The impact of convection on ENSO: From a delayed oscillator to a series of events, *J. Clim.*, 21, 5904–5924, doi:10.1175/2008JCLI2244.1.
- 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, 21349–21354, doi:10.1073/pnas.1015473107.
- Nordeng, T. E. (1994), Extended versions of the convection parameterization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics, *ECMWF Tech. Memo.* 206, 41 pp., Eur. Cent for Medium-Range Weather Forecasts, Reading, U. K.
- O'Dell, C. W., F. J. Wentz, and R. Bennartz (2008), Cloud liquid water path from satellite-based passive microwave observations: A new climatology over the global oceans, *J. Clim.*, 21, 1721–1739, doi:10.1175/2007JCLI1958.1.
- Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), Verification, validation, and confirmation of numerical models in the Earth sciences, *Science*, 263, 641–646, doi:10.1126/science.263.5147.641.
- Pedersen, C. A., E. Roeckner, M. Luthje, and J. Winther (2009), A new sea ice albedo scheme including melt ponds for ECHAM5 general circulation model, *J. Geophys. Res.*, 114, D08101, doi:10.1029/2008JD010440.
- Pincus, R., R. Hemler, and S. A. Klein (2006), Using stochastically generated subcolumns to represent cloud structure in a large-scale model, *Mon. Weather Rev.*, 134, 3644–3656, doi:10.1175/MWR3257.1.
- Räisänen, J. (2007), How reliable are climate models? *Tellus, Ser. A*, 59, 2–29, doi:10.1111/j.1600-0870.2006.00211.x.
- Rampal, P., J. Weiss, C. Dubois, and J.-M. Campin (2011), IPCC climate models do not capture Arctic sea ice drift acceleration: Consequences in terms of projected sea ice thinning and decline, *J. Geophys. Res.*, 116, C00D07, doi:10.1029/2011JC007110.
- Randall, D. A., and B. A. Wielicki (1997), Measurements, models and hypotheses in the atmospheric sciences, *Bull. Am. Meteorol. Soc.*, 82, 283–294.
- Reichler, T., and J. Kim (2008), How well do climate models simulate today's climate? *Bull. Am. Meteorol. Soc.*, 89, 303–311, doi:10.1175/BAMS-89-3-303.
- Rodwell, M. J., and T. N. Palmer (2007), Using numerical weather prediction to assess climate models, *Q. J. R. Meteorol. Soc.*, 133, 129–146, doi:10.1002/qj.23.
- Sausen, R., K. Barthel, and K. Hasselmann (1988), Coupled ocean-atmosphere models with flux correction, *Clim. Dyn.*, 2, 145–163, doi:10.1007/BF01053472.
- Siebesma, A. P., and J. W. M. Cuijpers (1995), Evaluation of parametric assumptions for shallow convection, *J. Atmos. Sci.*, 52, 650–666, doi:10.1175/1520-0469(1995)052<0650:EOPAFS>2.0.CO;2.
- Stainforth, D. A., et al. (2005), Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, 433, 403–406, doi:10.1038/nature03301.
- Stevens, B., and S. E. Schwartz (2012), Observing and modeling Earth's energy flows, *Surv. Geophys.*, doi:10.1007/s10712-012-9184-0, in press.
- Stroeve, J., M. M. Holland, W. Meier, T. Scambos, and M. Serreze (2007), Arctic sea ice decline: Faster than forecast, *Geophys. Res. Lett.*, 34, L09501, doi:10.1029/2007GL029703.
- Sundqvist, H., E. Berge, and J. E. Kristjansson (1989), Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model, *Mon. Weather Rev.*, 117, 1641–1657, doi:10.1175/1520-0493(1989)117<1641:CACPSW>2.0.CO;2.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, 93, 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Tiedtke, M. (1989), A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Mon. Weather Rev.*, 117, 1779–1800, doi:10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.
- Tompkins, A. M. (2002), A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover, *J. Atmos. Sci.*, 59, 1917–1942, doi:10.1175/1520-0469(2002)059<1917:APPFTS>2.0.CO;2.
- Voldoire, A., et al. (2012), The CNRM-CM5.1 global climate model: Description and basic evaluation, *Clim. Dyn.*, doi:10.1007/s00382-011-1259-y, in press.
- Watanabe, M., et al. (2010), Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity, *J. Clim.*, 23, 6312–6335, doi:10.1175/2010JCLI3679.1.

Corresponding author: T. Mauritsen, Max Planck Institute for Meteorology, Bundesstrasse 53, D-20146 Hamburg, Germany. (thorsten.mauritsen@zmaw.de)