



Copyright © 2009, Paper 13-009; 57,648 words, 10 Figures, 0 Animations, 1 Tables.
<http://EarthInteractions.org>

Consistency in Global Climate Change Model Predictions of Regional Precipitation Trends

Bruce T. Anderson*

Department of Geography and Environment, Boston University, Boston, Massachusetts

Catherine Reifen and Ralf Toumi

Space and Atmospheric Physics Group, Department of Physics, Imperial College of Science, Technology, and Medicine, London, United Kingdom

Received 11 July 2008; accepted 6 July 2009

ABSTRACT: Projections of human-induced climate change impacts arising from the emission of atmospheric chemical constituents such as carbon dioxide typically utilize multiple integrations (or ensembles) of numerous numerical climate change models to arrive at multimodel ensembles from which mean and median values and probabilities can be inferred about the response of various components of the observed climate system. Some responses are considered reliable in as much as the simulated responses show consistency within ensembles and across models. Other responses—particularly at regional levels and for certain parameters such as precipitation—show little intermodel consistency even in the sign of the projected climate changes. The authors' results show that in these regions the consistency in the sign of projected precipitation variations is greater for intramodel runs (e.g., runs from the same model) than intermodel runs (e.g., runs from different models), indicating that knowledge of the internal “dynamics” of the climate system can provide additional skill in making projections of climate change. Given the consistency provided by the governing dynamics of the model, the authors test whether persistence from an

* Corresponding author address: Bruce T. Anderson, Department of Geography and Environment, Boston University, 675 Commonwealth Ave., Boston, MA 02215-1401.

E-mail address: brucea@bu.edu

individual model trajectory serves as a good predictor for its own behavior by the end of the twenty-first century. Results indicate that, in certain regions where intermodel consistency is low, the short-term trends of individual model trajectories do provide additional skill in making projections of long-term climate change. The climate forcing for which this forecast skill becomes relatively large (e.g., correct in 75% of the individual model runs) is equivalent to the anthropogenic climate forcing imposed over the past century, suggesting that observed changes in precipitation in these regions can provide guidance about the direction of future precipitation changes over the course of the next century.

KEYWORDS: Regional climate change; Climate models; Anthropogenic forcing

1. Introduction

Understanding the future response of the global climate system to human emissions of radiatively active gases such as CO₂ and methane [termed greenhouse gases (GHGs)] has become a timely and compelling concern. This interest is particularly acute at the regional level, which is where impacts upon natural and socioeconomic systems will be realized (Parry et al. 2007). At these scales, it is well known that climate forecasts based upon the use of multiple simulations—or *ensembles*—of model predictions, which are then averaged to produce *ensemble means*, provide better forecast skill than any one individual forecast (e.g., Tebaldi and Knutti 2007); the same holds true for *multimodel ensemble means* generated using multiple simulations from multiple models. However, in certain regions, the (in)consistency between individual model forecasts for the end of the twenty-first century limits the utility of multimodel ensemble-mean forecasts because the uncertainty in the mean value of the forecast (as determined from the spread of the individual model forecasts) does not discount the possibility of no change or even a change of opposite sign (Giorgi and Francisco 2000; Räisänen 2001; Covey et al. 2003; DelSole 2004; Neelin et al. 2006; Räisänen 2007). Effectively, this problem arises because the individual model realizations of climate change for these regions are considered equally plausible outcomes, and differences between these outcomes are large. This problem is particularly evident for estimates of how the regional hydrologic cycle may vary with global-scale climate change over the course of the twenty-first century (Allen and Ingram 2002; Murphy et al. 2004; Held and Soden 2006; Sun et al. 2007; Solomon et al. 2007).

At the same time, the actual climate system is one realization of its own “internal” model system; that is, there is an underlying (albeit unobtainable) model system that is appropriate for the actual climate evolution. For this reason, previous researchers have used statistical methods to identify which models best represent the actual climate system; through various weightings based upon the agreement between historical observations and simulations (Krishnamurti et al. 2000; Giorgi and Mearns 2002; Robertson et al. 2004; Shukla et al. 2006), the aim has been to improve the consistency in forecasts by statistically identifying the “correct” model system. Unfortunately, given finite historical and future observed measures of the actual climate system, it may be extremely difficult to identify which numerical modeling system best captures the internal behavior of the actual climate system (Judd and Smith 2004). However, we can still quantify the *internal* consistency within a given

model system, and across model systems, to determine how strongly the evolution of individual realizations from a given model is constrained by the internal dynamics of the model itself, and by extension how strongly the observed climate system may also be constrained by its own internal dynamics (Räisänen 2001).

If we do find that the individual realizations of a given model system are constrained by the internal dynamics of the underlying model (e.g., there is internal consistency between the forecasts from a given model system), we can then determine whether this constrained behavior is self-contained within the evolution of the individual realizations themselves. In particular, we can test whether model persistence from an individual model trajectory—which by definition is governed by the same model dynamics throughout its evolution—serves as a good predictor for its own behavior by the end of the twenty-first century. While we may not know the true model system, much less the single realization, that the actual climate system is following, we can still test whether information from a single realization of the climate system (simulated or observed) can be used to predict its own behavior based upon how each of the individual realizations from the various model systems performs in predicting its own behavior. In this sense, we are interested in testing if persistence is a good predictor for the long-term behavior of the model systems, under the assumption that the evolution of the observed climate system—which is effectively a single realization governed by its own internal dynamics—has similar persistence as that found in the numerical modeling systems.

Model long-memory persistence studies have previously been performed for regional historical temperature (e.g., Syroka and Toumi 2001) and precipitation (e.g., Tomsett and Toumi 2001). Here we expand upon these to examine projections of regional precipitation variations using model-generated climate simulations forced by anthropogenic emissions of radiatively active chemical constituents. Because the spread of the individual model forecasts of regional precipitation tend to be large, we follow the lead of the Solomon et al. (Solomon et al. 2007) consistency analyses and focus on the sign of these projected precipitation changes, not necessarily the magnitude. Section 2 describes the model systems and datasets used in this study while section 3 discusses the skill metric used throughout this paper. Section 4 examines how the skill of regional precipitation projections changes as a function of location, time, and predictor. Section 5 summarizes the results of this study.

2. Data

For this study, we use coupled atmosphere–ocean–land surface model output produced from seven different numerical coupled-climate model systems, forced by projected changes in greenhouse gas concentrations and anthropogenic aerosols over the next 100 years (2000–2100) that stabilize at an (equivalent) CO₂ concentration of 720 ppm (parts per million) by the year 2100. These seven models are chosen because they each have three or more individual simulations—or *ensemble members*—forced by the projected changes in greenhouse gas concentrations and anthropogenic aerosols. Generally, for a given model system, a long-run (multi-century) control simulation of the coupled-climate model is performed in which the radiatively active chemical constituents (including CO₂ and other greenhouse gas concentrations, sulfate aerosols, and volcanic particulates) and solar activity are fixed at their preindustrial levels (generally designated as 1860). Then multiple

Table 1. Name and characteristics of model simulations used in this analysis.

Name	No. in ensemble	Horizontal resolution (atmosphere only)	Vertical levels
CGCM3	5	T47 (about 3.75°)	31
CGCM2	5	T42 (about 2.8°)	30
CCSM3	4	T85 (about 1.4°)	26
PCM1	4	T42 (about 2.8°)	26
ECHAM	4	T63 (about 1.875°)	31
GISS-EH	3	Lat–lon: 4° × 5°	20
MIROC(medres)	3	T42 (about 2.8°)	20

integrations of the same coupled-climate model are initialized using different time periods taken from the control simulation and forced by the same historical, twentieth-century (generally 1860–2000) changes in atmospheric chemical constituents and solar activity. (The time periods chosen from the control simulations to initialize the twentieth-century simulations differ for each model system; for instance, some models use time periods that are 20 years apart while others use time periods that are 100 years apart.) The individual twentieth-century simulation output at year 2000 is then used to initialize an individual simulation (using the same coupled-climate model) forced by projected changes in greenhouse gas concentrations and anthropogenic aerosols over the next 100 years (2000–2100). We will be using model data in which the future changes in radiatively active chemical constituents follows the A1B emissions projection—termed an *emissions scenario*—from the Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios (SRES; Nakićenović et al. 2000), which as mentioned corresponds to a stabilization of CO₂ concentrations at 720 ppm by the year 2100.

The models used in this study include the Canadian Centre for Climate Modelling and Analysis's T47-resolution Coupled General Circulation Model, version 3 (CGCM3; 5 ensemble members); the Meteorological Research Institute's Coupled General Circulation Model version 2 (CGCM2; 5 ensemble members); the Parallel Climate Model version 1 (PCM1; 4 ensemble members); the Community Climate System Model, version 3 (CCSM3; 4 ensemble members); ECHAM (4 ensemble members); the Goddard Institute for Space Studies Model E/Hybrid Coordinate Ocean Model (HYCOM) (GISS-EH; 3 ensemble members); and the Model for Interdisciplinary Research on Climate 3, medium-resolution version [MIROC(medres); 3 ensemble members] (Solomon et al. 2007). All data from the model runs are taken from the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and are made available through the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project (CMIP3) multimodel dataset. Details about each model are provided online (http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php) and summarized in Table 1. For this investigation, we will be examining annual- and seasonal-mean precipitation values and their long-term variations. Hence, we first compute the annual and 3-month means for each field at each grid point. We then apply a 20-year running mean to each field at each grid point. All figures are based upon these 20-year running mean gridpoint values unless noted otherwise. In addition, in order to compare across model systems, we interpolate all gridpoint values to the highest-resolution model grid (CCSM3), which has a T85 (approximately 1.4°) resolution.

3. Methods

Throughout this paper, skill of model forecasts—whether concurrently or lagged—are based upon a modified version of the Hanssen–Kuipers skill score (Hanssen and Kuipers 1965). This measure of skill is based upon a two-dimensional contingency table of binomial outcomes between predicted and observed events. In its traditional formulation, it gives the difference between the number of “hits” (i.e., forecasted events that occurred) and “false alarms” (i.e., forecasted events that did not occur). It does not account for the number of observed nonevents that were correctly predicted or the number of observed events that were missed. For our study, however, we consider below-normal precipitation amounts (i.e., negative precipitation trends) to be an equally valid forecast as above-normal precipitation amounts (i.e., positive precipitation trends). Hence, we determine the consistency between model forecasts based upon the total number of correct predictions (either of positive or negative precipitation trends) minus the total number of incorrect predictions (again, either of positive or negative precipitation trends). For a traditional contingency table, this calculation would be equivalent to taking the total number of “hits” H and “correct nonevents” N then subtracting off the total number of “false alarms” F and “misses” M and normalizing by the number of predictions:

$$\text{Skill} = \frac{(H + N) - (F + M)}{H + N + F + M}. \quad (1)$$

This measure of skill can range from -1 to 1 , with numbers above 0 representing fractional improvement upon chance. Numerically, the skill score for a given number (n) of forecasts and observations can be calculated as

$$\text{Skill} = \frac{\left[\sum_{i=1}^n \text{sign}(F_i) \times \text{sign}(O_i) \right]}{n}, \quad (2)$$

where F_i and O_i are the i th pair of forecasted and observed values, respectively, and $\text{sign}(\)$ indicates the sign of the operand (e.g., “1” for positive values and “ -1 ” for negative values).

In addition, because we are dealing with finite numbers of predictions—in some cases as small as 28—it is necessary to devise a method for testing for significance of the results. To do this, we construct a stochastic model in which 28 pairs of values are randomly selected from a normal distribution centered on zero, one representing the “forecasted” value and the other representing the “observed” value. The test-statistic skill score is calculated for these 28 pairs as

$$\text{Skill}_{\text{rand}} = \frac{\left[\sum_{i=1}^{28} \text{sign}(F_i^{\text{rand}}) \times \text{sign}(O_i^{\text{rand}}) \right]}{28}, \quad (3)$$

where F_i^{rand} and O_i^{rand} are the i th pair of randomly selected “forecasted” and “observed” values, and $\text{sign}(\)$ is the same as above. A similar estimate is made

for 10 000 different sets of pairs. The probability distribution function of skill scores from these sets of 28 “forecasts” and “observations” is then calculated to determine the likelihood certain skill scores could arise by chance. From these calculations, we find that skill scores above 0.3 are significantly different from those expected by chance at the 95% confidence level. Similarly, if we randomly select 28 “forecast” values and use them to “predict” a single (random) “observed” value, we also find that that skill scores above 0.3 are significantly different from those expected by chance at the 95% confidence level. Hence, for all gridpoint estimates of skill (in which only 28 sets of ensemble members are available), the minimum significant skill score is set to 0.3, although values above 0 are shown as well.

To test the difference in skill scores between two predictors, a similar methodology is used except that two randomly distributed “forecast” values are used as separate predictors for the same randomly distributed “observed” value. The skill score for the two separate sets of 28 “forecasts”–“observation” pairs is calculated and then we find the difference; as above we repeat the analysis for 10 000 different sets. We find that differences in skill scores above ± 0.25 are significantly different from those expected by chance at the 95% confidence level; those above ± 0.21 are significant at the 90% level. Hence, for gridpoint estimates the minimum significant (absolute) skill-score difference is set to 0.25, although values above 0.20 are shown as well.

4. Results

4.1. Intermodel and intramodel projections of precipitation trends

Previous studies have investigated coupled global climate models’ ability to simulate historical regional precipitation variations (Zhang et al. 2007) and found that observed historical trends along certain latitudinal bands can be reproduced by multimodel ensemble-mean estimates generated from 10 different models (5 of which are included in this study). Other studies have compared simulated trends with observations across different regions and time periods, some of which appear reproducible (Bhend and von Storch 2008; Barnett et al. 2008) while others are not (Lambert et al. 2005; Allan and Soden 2007). While this ability (or inability) of models to simulate past regional precipitation variations can hamper the direct use of these same models to project future changes, below we argue that we can still obtain *information* from the simulated output about the possible future behavior of the observed climate system, even if the simulations show large disparities in their *projections* of historical (and future) precipitation changes at regional scales (Allen and Ingram 2002).

To start, Figure 1 shows the projected changes in ensemble mean (EM) gridpoint precipitation values between the periods 2080–2100 and 2000–20, as found in the seven different model simulations, as well as in the multimodel ensemble mean (MMEM). These trends are qualitatively consistent with those found in other MMEM projections (Murphy et al. 2004; Zhang et al. 2007; Sun et al. 2007; Solomon et al. 2007) including decreasing trends across most of the subtropics in the Northern and Southern Hemispheres and increasing trends over the mid- and high latitudes of both hemispheres. In addition, there are overall increases in the equatorial and tropical regions of most models. At the same time, there are

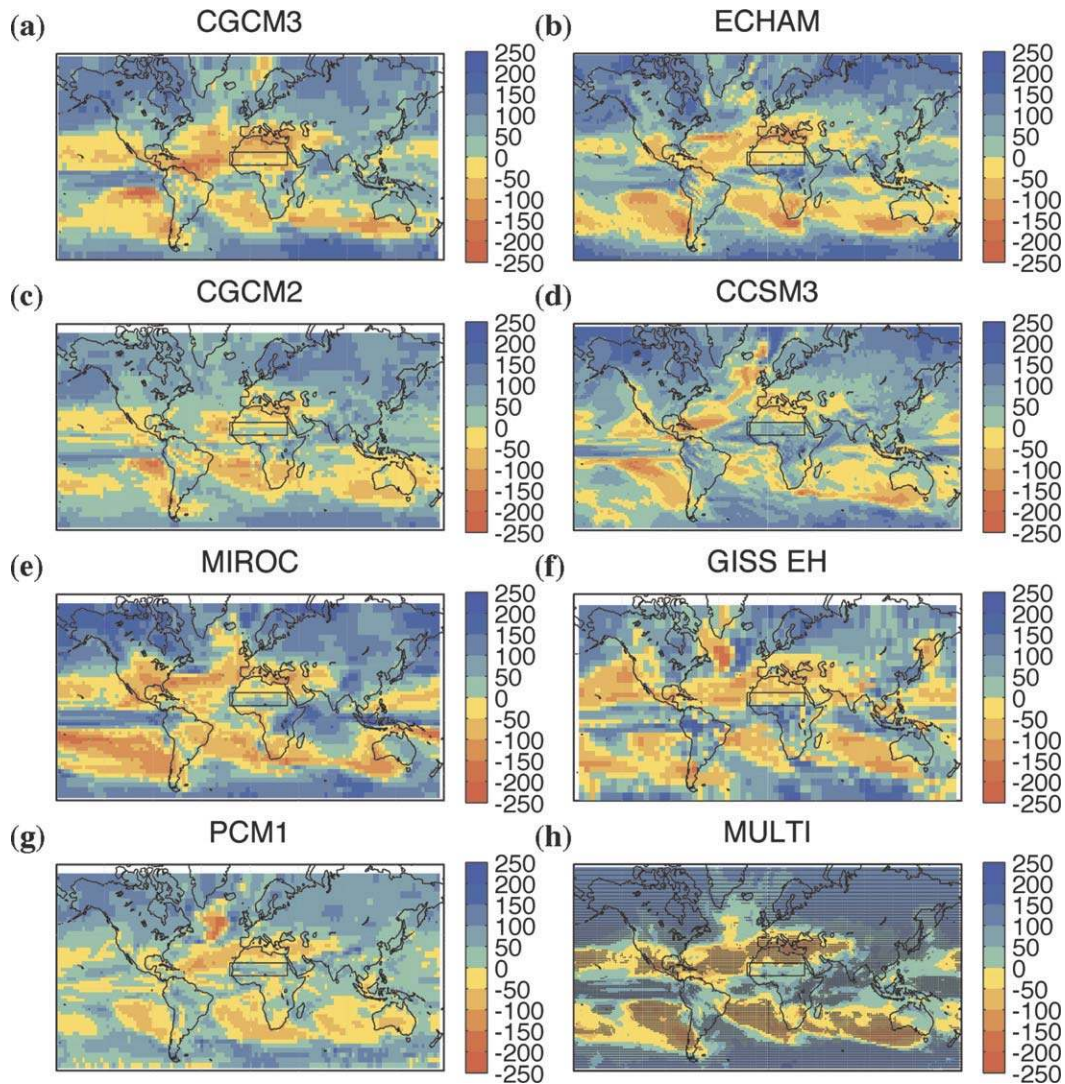


Figure 1. Projected changes in ensemble mean annual precipitation under the A1B emissions scenario for the (a) CGCM3.1, (b) ECHAM, (c) CGCM2.1, (d) CCSM3, (e) MIROC(medres), (f) GISS-EH, and (g) PCM1.0 model systems, plotted on their native grid. (h) Also shown is the multimodel ensemble mean, plotted on a common T128 resolution grid. Changes are calculated as the difference between the ensemble-mean precipitation amounts averaged from 2080 to 2100 and from 2000 to 2020. Values are presented as a fraction of the interannual standard deviation of the 20-year running mean gridpoint values for the full period (2000–2100). Stippled regions in (h) represent grid points in which 6 of 7 (85%) of the models show the same sign change in 2080–2100 precipitation. The box represents area-averaging domain used in Figure 2.

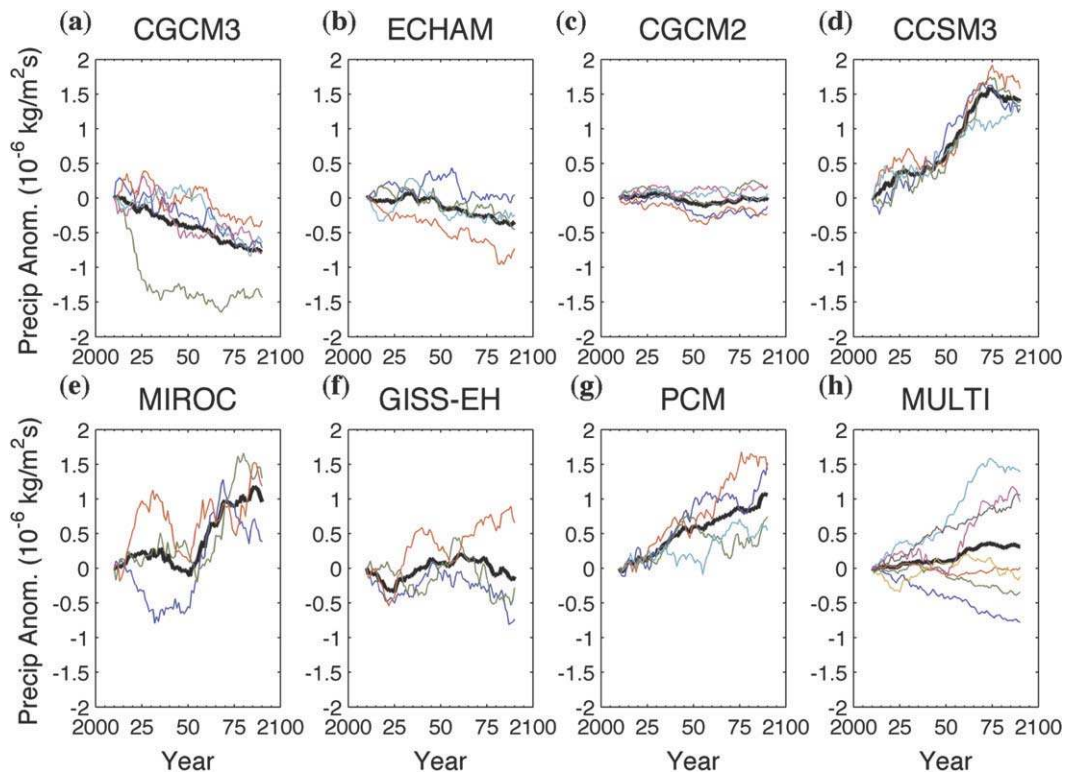


Figure 2. Time evolution of projected changes in area-average annual precipitation for central Africa under the A1B emissions scenario for the (a) CGCM3.1, (b) ECHAM, (c) CGCM2.1, (d) CCSM3, (e) MIROC(medres), (f) GISS-EH, and (g) PCM1.0 model systems. See Figure 1 for location of area-averaging region; only land-based grid points are considered. Colored lines represent individual ensemble members from the given model system; thick black lines represent ensemble mean for the given model system. Units are $\text{kg m}^{-2} \text{s}^{-1}$. Values represent 20-year running means centered on the given date; all lines shifted such that initial values start at 0. (h) Time evolution of ensemble-mean area-average annual precipitation for central Africa under the A1B emissions scenario. Colored lines represent ensemble means from individual model systems (equivalent to thick, black lines in previous panels); the thick black line represents multimodel ensemble mean for all model systems.

model-based differences in the sign of projected precipitation trends in many regions—such as much of Africa, Australia, the Amazon, and the eastern United States—again in agreement with MMEM estimates of consistency between model predictions (Murphy et al. 2004; Sun et al. 2007; Solomon et al. 2007).

To highlight the difficulty in predicting the long-term trends in precipitation for these regions, Figure 2 shows the evolution of the area-average precipitation amounts over central Africa (see Figure 1 for area-averaging region) from the seven different model systems. While three of the EM area-average precipitation amounts are positive [CCSM3, MIROC(medres), and PCM1], two show only small

changes (CGCM2 and GISS-EH) and another two indicate decreasing precipitation amounts (CGCM2 and ECHAM) over this region. In addition, of 28 separate forecasts (from all seven model systems), only 16 (57%) have the same sign as the MMEM value. In all, it is apparent that the MMEM projections of precipitation changes across this region are of little value when considering the range of plausible scenarios as provided by the different model systems.

At the same time, four of the model systems show high intraensemble consistency [CGCM3, CCSM3, MIROC(medres), and PCM1], with all ensemble members showing the same sign change in precipitation by the 2080–2100 period; in addition, three of the four ensemble members from the ECHAM model system show the same sign change in precipitation. To better quantify whether a similar improvement in consistency holds in other regions, we estimate the gridpoint consistency of 2080–2100 precipitation variations using the modified skill score described in section 3. Here, we consider each of the model realizations at each grid point to be one equally plausible outcome of anthropogenic forcing of climate change; from these we get estimates of what the possible “observed value” of gridpoint precipitation changes for the 2080–2100 period may be (Räisänen and Palmer 2001). We then calculate the MMEM projection of 2080–2100 precipitation changes at each grid point and use this as the “forecast” value. We can then estimate how large the forecasted value skill is, given the range of plausible realizations of the climate system:

$$\text{Skill}_{\text{MMEM}} = \frac{\left[\sum_{i=1}^{28} \text{sign}(O_i) \times \text{sign}(F_{\text{MMEM}}) \right]}{28}, \quad (4)$$

where O_i is the “observed” 2080–2100 value provided by the i th individual model realization (at a given grid point), F_{MMEM} is the forecasted 2080–2100 value provided by the MMEM value (at the same grid point), and $\text{sign}(\)$ is as above. Because the forecast value is the same for each of the forecast/observation comparisons, the skill score is a measure of how much the “observations,” that is, the plausible realizations of future climate change, differ from one another.

Figure 3a shows the results of this comparison. In many regions, particularly the high latitudes of the Northern Hemisphere, along with most of northern Eurasia, the skill score is 1.0, indicating that every model realization (out of 28) is producing the same sign change in precipitation for the period 2080–2100. In other regions, the skill score falls below the 95% confidence interval ($\text{skill} < 0.3$) and actually becomes negative. The first implication of these results is that using the MMEM forecast is a poor predictor for determining the sign of precipitation trends, given the range of plausible realizations drawn from the individual model runs. Again, this lack of skill arises solely from the lack of consistency in the individual realizations (since the “forecast” is the same for each of these forecast/observation comparisons).

We see that consistency is low along the storm-track regions of the North and South Atlantic and Pacific Ocean basins, as well as at the boundaries between the tropics (subtropics) where the expansion (contraction) of the ITCZ (subtropical highs) can produce differing trends in precipitation across models (Neelin et al. 2006; Allan and Soden 2007). In addition, there are regions of low consistency over

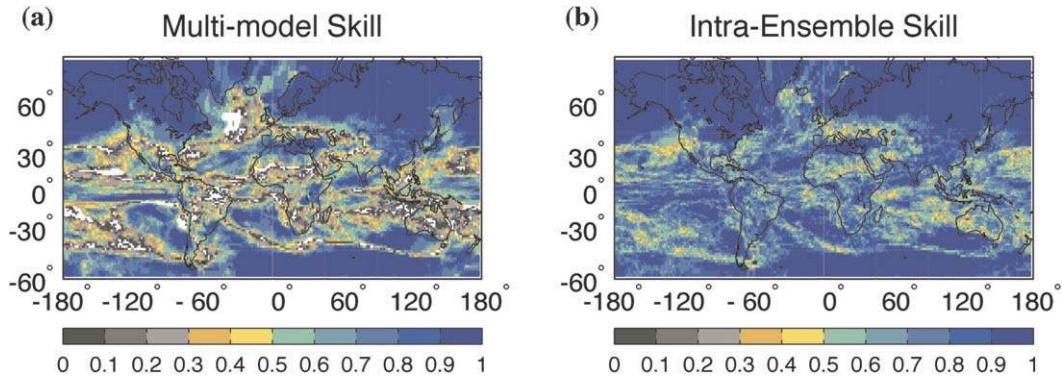


Figure 3. (a) The skill of using the sign of the 2080–2100 multimodel ensemble-mean precipitation anomalies to predict the sign of the 2080–2100 precipitation anomalies in the 28 individual model realizations; anomalies calculated as deviations from the 2000–20 values. Skill based upon a modified Hanssen-Kuipers skill score (see text for details). Only values greater than 0 shown here; skill scores that are significantly different from chance at the 95% level are shaded in color. (b) Same as in (a), except when using the 2080–2100 ensemble-mean precipitation anomalies from a given model system to predict the sign of the 2080–2100 precipitation anomalies in the individual model realizations from that model system only.

much of the tropical landmasses, including the Amazon in South America, the Sahel in Africa, and across Indonesia. There is also low consistency over southern Africa and Australia. Qualitatively, regions of low consistency, as determined by the low skill scores in Figure 3a, match those derived from the MMEM projections (Figure 1h; also Figure 2 from Murphy et al. 2004; Figure 9c from Sun et al. 2007; Figure 10.12 from Solomon et al. 2007).

As mentioned, in these regions the low skill score between the MMEM projection of precipitation trends and the range of plausible realizations of these trends (as found in the individual model runs) suggest that the MMEM projection of precipitation is of little value. However, the MMEM projection is not the only projection available. In Figure 3b, we compare the EM precipitation trends from each model with its own individual model realizations. We then calculate the overall skill score for the “intraensemble” predictions/observations:

$$\text{Skill}_{\text{EM}} = \frac{\left[\sum_{j=1}^7 \sum_{i=1}^{n_j} \text{sign}(O_i^j) \times \text{sign}(F_j^{\text{EM}}) \right]}{28}, \quad (5)$$

where O_i^j is the “observed” 2080–2100 value provided by the i th individual model realization from the j th model (at a given grid point), F_j^{EM} is the forecasted 2080–2100 value provided by the EM value from the j th model (at the same grid point), n_j is the number of individual realizations provided by the j th model, and $\text{sign}(\)$ is as above. In this case, there are still 28 individual realizations of the climate system, but now the projections for each realization are not the same but are based upon a priori

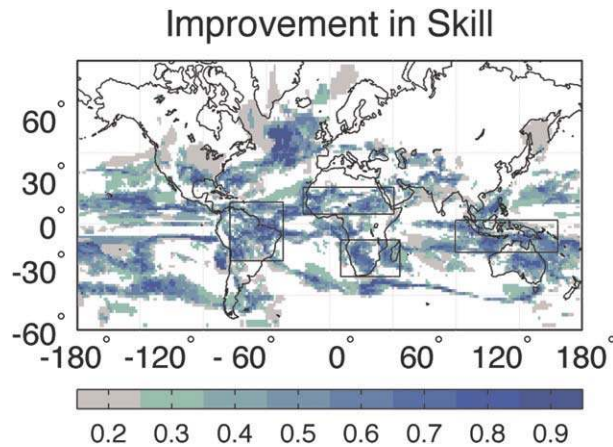


Figure 4. Difference in the skill of using the sign of the 2080–2100 ensemble-mean precipitation anomalies from a given model system to predict the sign of the 2080–2100 precipitation anomalies in the individual model realizations from that model system only, compared with the skill of using the sign of the 2080–2100 multimodel ensemble-mean precipitation anomalies. Skill is based upon a modified Hanssen–Kuipers skill score (see text for details). Only values greater than 0.2 (90% confidence level) are shown here; values in color are significant at 95% confidence level (>0.25). The boxes indicate regions that are analyzed further.

knowledge of the model system generating a given realization, hence there are seven different “forecast” values (one from each model system). Here we are testing whether, given the right model system, there is consistency in the relation between EM model projections and the individual realizations of the given model system.

As is evident in Figure 3a, the skill improves markedly across the globe and is significant (at the 95% confidence interval) everywhere. In addition, in many regions where the MEM projection had little skill in forecasting any given realization of the climate system—for instance over the Amazon or southern Africa—the intraensemble model skill is significantly greater. In these regions, then, there is little consistency in projections of future climate change between model systems; however, there is much greater consistency of these projections within the individual model systems.

To quantify the improvement in consistency that can be gained by using the “correct” model system to estimate changes of climate for a given realization, as opposed to using the MEM projection, Figure 4 shows the difference in skill between the skill scores in Figures 3b and 3a. As mentioned, there is significant improvement (>0.25) in skill over the 1) Amazon basin in South America, over 2) central and 3) southern Africa, and over the 4) island states of Indonesia. There is also improvement over many ocean regions, such as the storm track of the North Atlantic and the subtropical regions of the Pacific. Our focus here, however, will be upon the four regions mentioned previously.

Figure 5 shows the evolution of the EM area-average precipitation amounts for these four regions, taken from the seven different model systems; in each case we

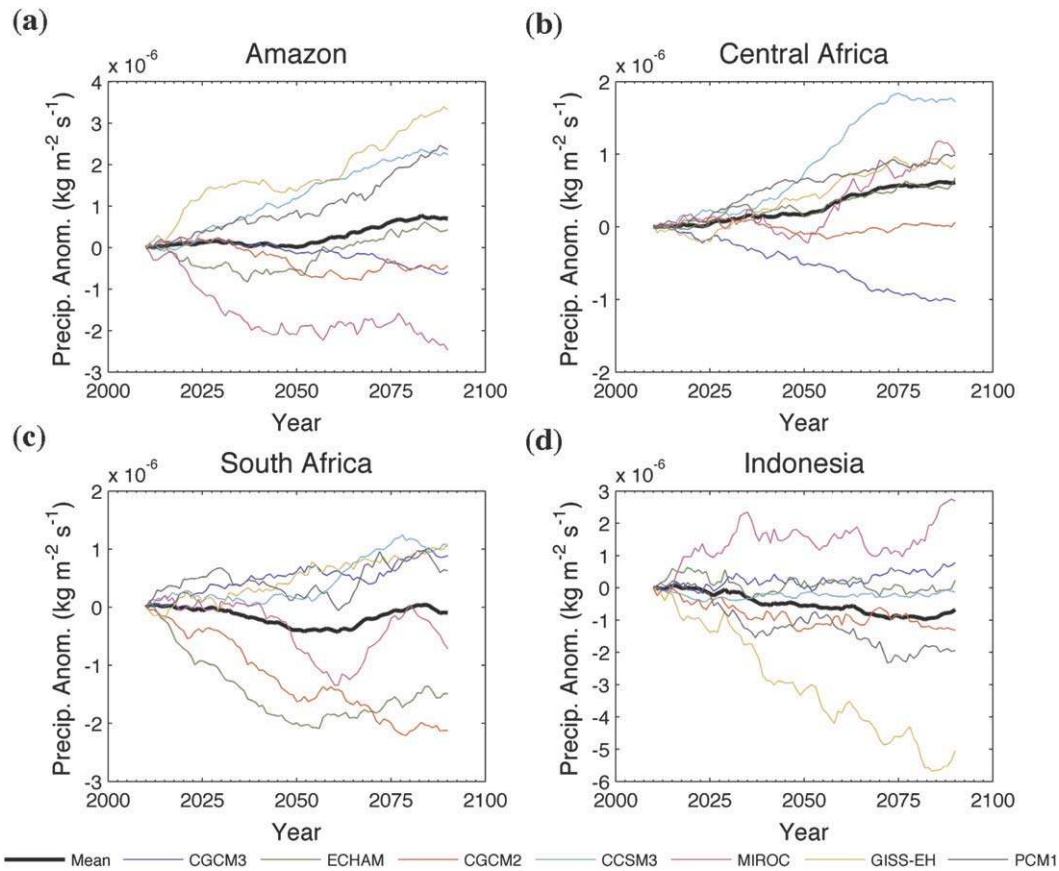


Figure 5. Time evolution of projected changes in area-average ensemble-mean annual precipitation for (a) Amazon basin, (b) central Africa, (c) South Africa, and (d) Indonesia under the A1B emissions scenario. See Figure 4 for location of area-averaging regions. For all regions except Indonesia, only land-based grid points are considered. Colored lines represent individual model system ensemble means; the thick black lines represent multimodel ensemble mean. Units are $\text{kg m}^{-2} \text{s}^{-1}$. Values represent 20-year running means centered on the given date; all lines shifted such that initial values start at 0.

only use land-based grid points, except for Indonesia where all grid points within the box are included in the area average. In all four regions there is a significant spread in both the magnitude and sign of the projected changes across the model systems (with the exception of central Africa, where only one model projects a decrease in area-average precipitation). At the same time, it appears that many of the model systems follow a quasi-monotonic evolution through time such that the signs of the initial trends in the evolution of the system match the signs of the final trends. While this generalization does not hold for all models and all regions [see the ECHAM projections over the Amazon for instance or the MIROC(medres) projections over central and southern Africa], it does suggest that it is possible to use

the intervening trends in the evolution of the climate system to project the sign of the longer-term trends, as represented by the state of the system during 2080–2100.

To test this hypothesis, we take each of the EM gridpoint model estimates at a given time and use the sign of the precipitation anomaly (compared with the initial state) as a prediction for the 2080–2100 gridpoint precipitation anomaly for each realization of that model system:

$$\text{Skill}_j^{\text{EM}}(t) = \frac{\left\{ \sum_{i=1}^{n_j} \text{sign}[O_i^j(2080 - 2100)] \times \text{sign}[O_j^{\text{EM}}(t)] \right\}}{n_j}, \quad (6)$$

where $O_i^j(2080-2100)$ is the “observed” 2080–2100 value provided by the i th individual model realization from the j th model (at a given grid point), $O_j^{\text{EM}}(t)$ is the EM value from the j th model at time t , n_j is the number of individual realizations provided by the j th model, and $\text{sign}(\)$ is as above. To calculate the skill of this prediction system for a given region, all gridpoint prediction/observation pairs within the region are included in the “hits”/“misses” statistics, hence there are significantly more predictions included in these estimates (generally n_j is on the order of 3–5 predictions at 300–600 separate grid points). In this sense, we are determining whether persistence of short-term EM gridpoint precipitation trends serve as good predictors for the individual long-term realizations of the given model system. We can also test the skill of the predictions if no a priori knowledge is available regarding which model system produced the given realization; in this case the MMEM gridpoint precipitation anomaly is used as the predictor for each of the model realizations:

$$\frac{\text{Skill}^{\text{MMEM}}(t) = \left\{ \sum_{j=1}^7 \sum_{i=1}^{n_j} \text{sign}[O_i^j(2080 - 2100)] \times \text{sign}[O_{\text{MMEM}}(t)] \right\}}{\sum_{j=1}^7 n_j}, \quad (7)$$

where all variables are the same as in Equation (6) except for $O_{\text{MMEM}}(t)$, which is the MMEM value at time t (at the given grid point).

Figure 6 indicates that, in the four regions considered here, short-term trends do have significant predictive skill for determining the sign of long-term trends of the various plausible realizations of precipitation, but only if it is known what model system to use as the predictor; without this a priori knowledge, the MMEM values at intervening periods show only slight improvement (above chance) in predicting the plausible long-term trends in precipitation given by the individual ensemble members. This result highlights that a priori information about the appropriate model system (or the observed climate system) can improve the forecast capabilities for these regions. It is also important to highlight that the improvement in model forecast capability is not the same across models. For instance, the CGCM2 EM does only about as well as the MMEM at predicting the end state of its own ensemble members over central Africa. This result does not indicate that the CGCM2 is a worse-performing model but instead simply indicates that the internal consistency among ensemble members is lower than for other models.

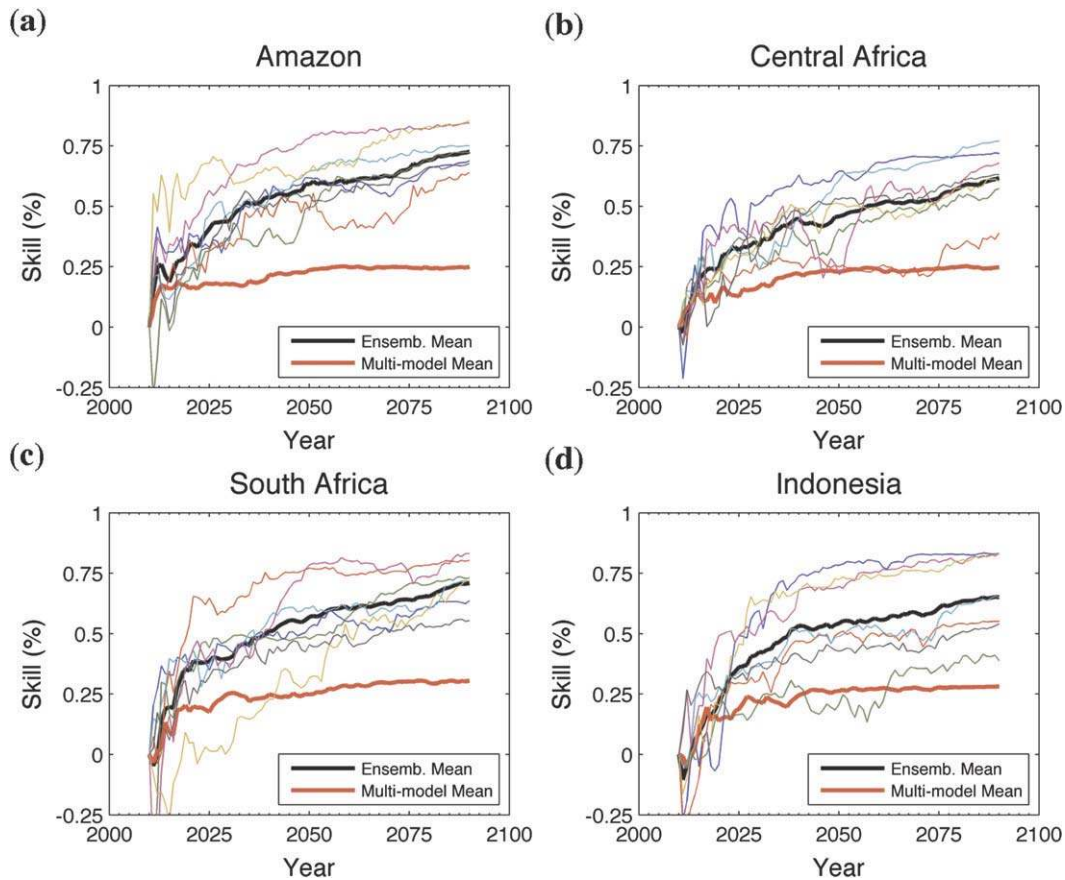


Figure 6. Skill of using the persistence of the sign of intervening gridpoint precipitation anomalies at the given time as predictors for the sign of the 2080–2100 gridpoint precipitation anomalies. Thin, colored lines represent the skill of the ensemble-mean values from a given model system in predicting the sign of the 2080–2100 values of individual model realizations from that model system only. Line colors are the same as in Figure 5. The thick black lines represent the average of all the model systems' ensemble-mean skills. The thick red lines represent the skill of the multi-model ensemble-mean values in predicting the sign of the 2080–2100 values of individual model realizations from all model systems. Skill calculated separately for (a) Amazon basin, (b) central Africa, (c) South Africa, and (d) Indonesia. See Figure 4 for location of area-averaging regions. For all regions except Indonesia, only land-based grid points are considered. Predictor values represent the difference between the 20-year mean value centered on the given date and the initial 20-year mean value; by construction all lines start at 0.

4.2. Individual model projections of precipitation trends

We note that, in each of the cases shown in Figure 6, the forecast estimate was derived from an ensemble of model realizations. However, in the actual climate system, the observed evolution only represents one realization of its own internal

“dynamics.” Determining which model climate system the actual climate system maps onto may be extraordinarily difficult (and may actually change from region to region). In addition, as results show, different model systems can have differing internal consistency in the behavior of their individual realizations.

At the same time, it would be of interest to determine whether the short-term evolution of individual realizations of the climate system can be used as predictors for their own evolution. In this sense, the a priori knowledge of the governing model system is self-contained within the evolution of the individual model realization itself. To test for this ability, for each model system we take each of the gridpoint anomalies from an individual model realization at a given time and use the sign of the precipitation anomaly (compared with the initial state) as a prediction for its own 2080–2100 gridpoint precipitation anomaly:

$$\text{Skill}_j^{\text{indiv}}(t) = \frac{\left\{ \sum_{i=1}^{n_j} \text{sign}[O_i^j(2080 - 2100)] \times \text{sign}[O_i^j(t)] \right\}}{n_j}, \quad (8)$$

where $O_i^j(2080-2100)$ is the “observed” 2080–2100 value provided by the i th individual model realization from the j th model (at a given grid point), $O_i^j(t)$ is the value from the i th individual model realization from the j th model at time t , n_j is the number of individual realizations provided by the j th model, and $\text{sign}(\)$ is as above. As before, the skill is calculated using all grid points within a given region, so n_j is much larger than 28. Results are shown in Figure 7. As before, the evolution of individual model realizations, and their consistency in use as predictors for the final state of the model realization, depends upon the underlying model system. As an extreme example, individual realizations of gridpoint precipitation over Indonesia from the ECHAM model system appear to be very poor predictors of their final states. Again, this does not indicate that the ECHAM model is a poor one, simply that the short-term model trends in this region do not necessarily serve as good predictors for the final end state.

However, if we calculate the average skill score for all model realizations, we find that in all four regions it lies above 0.5 by 2050 and in some regions (the Amazon, southern Africa, and Indonesia once the ECHAM model is removed) it is above 0.5 by 2040. Because these time series are plotted at the center of the 20-year averaging period, these results suggest that short-term trends (e.g., during the 2030–50 period) can be used as predictors for the sign of the trend in these regions during 2080–2100.

While these results do not sound promising, it should be noted that during the 2030–50 period, the greenhouse gas concentrations, as represented by the concentrations of CO₂ in the A1B scenario, are expected to be about 90 ppm higher than the initial 2000–20 period (e.g., an increase from 390 to 484 ppm). This increase of 90 ppm is nearly equivalent to the observed increase in CO₂ concentrations during the period 1900–2000 (i.e., from 280 to 370 ppm). In this sense, these results suggest that the observed trends in precipitation over the last 100 years may serve as predictors for the sign of future climate change over the next 100 years. However, to confirm this hypothesis it will be necessary to compare historical realizations of the climate system with future realizations to see whether similar increases in skill are found, a study that is beyond the scope of this paper (but is presently being carried out).

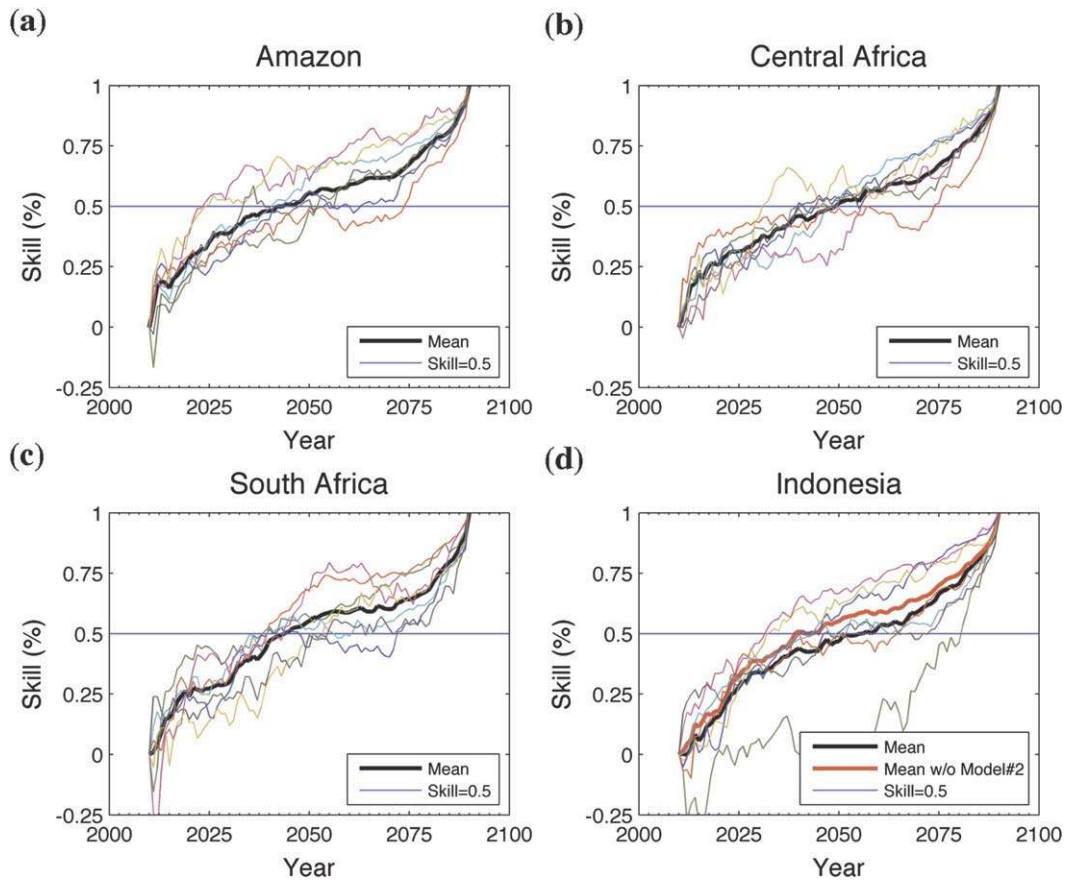


Figure 7. Skill of using the persistence of the sign of intervening gridpoint precipitation anomalies from individual model realizations at the given time as predictors for the sign of 2080–2100 gridpoint precipitation anomalies from the same realization. Colored lines represent the skill averaged over each realization from a given model system; line colors are the same as in Figure 5. The thick black lines represent the mean of all the model systems' skills. Skill is calculated separately for (a) Amazon basin, (b) central Africa, (c) South Africa, and (d) Indonesia. See Figure 4 for location of area-averaging regions. For all regions except Indonesia, only land-based grid points are considered. The thick red line in (d) represents the mean of all the model systems' skills after removing the ECHAM model. Predictor values represent the difference between the 20-year mean value centered on the given date and the initial 20-year mean value; by construction all lines start at 0 and end at 1.

Next, we perform a similar analysis for all grid points in order to determine what regions of the globe show consistency in their short-term and long-term trends of precipitation. To do so requires selecting a specific time period upon which to base the predictions, at which point it is possible to determine the skill of these predictions. Based upon Figure 7, we select the 2030–50 period as the prediction period. We then use the sign of the MMEM gridpoint estimates during the 2030–50

period as the same predictor for the 2080–2100 gridpoint precipitation anomalies taken from the individual realizations of all the model systems (Figure 8a). For comparison, we also use the EM gridpoint model estimates during the 2030–50 period as a prediction for the 2080–2100 gridpoint precipitation anomalies taken from the individual realizations of the given model system (Figure 8b). Finally, we use the individual model realization gridpoint anomaly estimates during the 2030–50 period as a prediction for their own 2080–2100 gridpoint precipitation anomalies (Figure 8c).

As expected, in the regions we have been studying short-term trends from the MMEM gridpoint estimates serve as poor predictors for the long-term trend of precipitation, given the range of plausible individual realizations of the long-term climate evolution. In other regions, however, the skill is significant and matches that of the actual MMEM gridpoint predictions from 2080 to 2100 (see Figure 3a). In comparison, the skill provided by the short-term EM anomalies in predicting the individual model realizations of its own ensemble members is greater than the MMEM predictions almost everywhere. In particular the improvement is large over the regions of interest here, namely, the Amazon basin, central and southern Africa, and Indonesia. As before, these results suggest that if a priori knowledge is available regarding which model system to use as the predictor for a given realization, the short-term trends in the EM precipitation for these regions can serve as significant predictors for the longer-term evolution of individual ensemble members.

As before, there is only one observed evolution of the actual climate system itself. However, based upon Figure 8c, it appears that, for many regions, short-term individual realizations of climate change can be used as predictors for their own long-term evolutions. Even in regions where fully coupled model predictions of trends in precipitation show little intermodel consistency, the short-term evolution of a single realization appears to provide additional information about its future state. While not perfect, by construction the skill associated with these predictions averages about 0.5 in the regions examined earlier; that is, about 75% of the model predictions capture the correct sign of future trends in precipitation. In addition, even outside regions examined here there appears to be skill found in the short-term evolution of the climate system, which provides information about its long-term behavior. For example, over Australia, western India, and the western United States—regions in which MMEM climate projections indicate inconsistency in the overall trend of precipitation—short-term trends can provide some guidance regarding the longer-term evolution.

Figure 9 shows the difference in skill provided by the sign of the 2030–50 precipitation anomalies from individual model realizations in predicting their own end state, as compared with the skill provided by the sign of the 2030–50 precipitation anomalies from the MMEM values. As mentioned, short-term trends in individual model realizations appear to provide enhanced skill over much of Africa, the Amazon basin, Australia, and the western United States. At the same time, it is apparent that in certain regions—particularly the high-latitude regions of North America and Eurasia—the MMEM provides a much better estimate of the long-term behavior of the climate system than that afforded by any one model realization. Hence, in many regions care must be placed in projecting observed short-term trends into the future, particularly in the presence of nonlinear and/or nonstationary behavior (DelSole 2005).

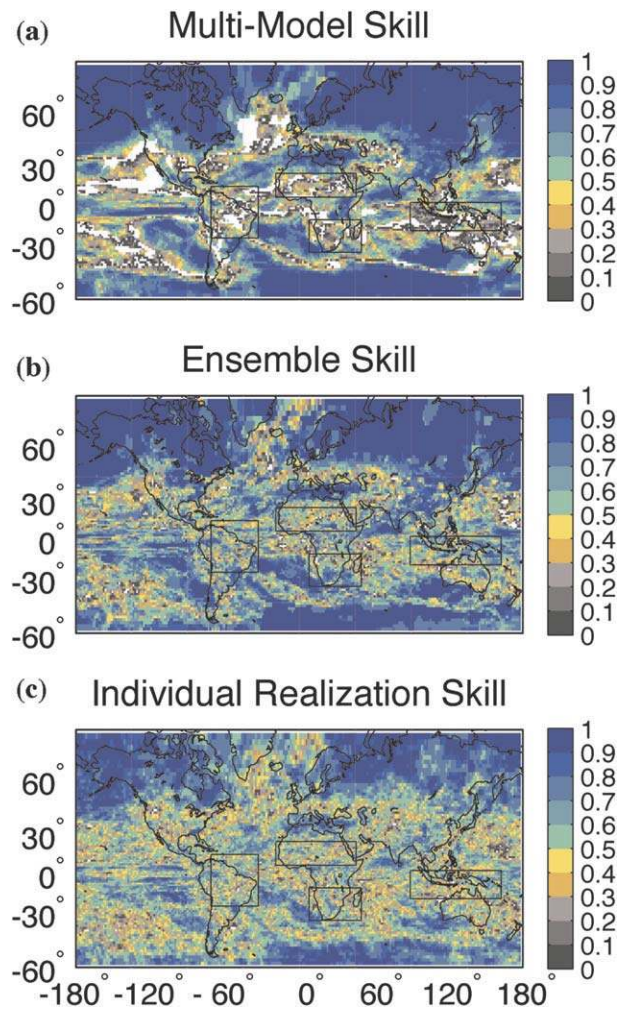


Figure 8. Skill of using the sign of the 2030–50 precipitation anomalies to predict the sign of the 2080–2100 precipitation anomalies in the 28 individual model realizations; anomalies calculated as deviations from the 2000–20 values. Sign of 2030–50 gridpoint precipitation anomalies determined from (a) multimodel ensemble-mean gridpoint value, (b) ensemble-mean gridpoint value from the model system that generated the individual model realization, and (c) the individual model realization itself. Skill is based upon a modified Hanssen–Kuipers skill score (see text for details). Only values greater than 0 shown here; skill scores that are significantly different from chance at the 95% level are shaded in color. The boxes indicate regions that are analyzed in previous figures.

4.3. Projections of seasonal-mean precipitation trends

Given that the signs of future anthropogenic-induced climate changes can vary between seasons (Solomon et al. 2007), it is of interest to see how seasonally based results may differ. Generally, the skill of the MMEM precipitation anomalies from

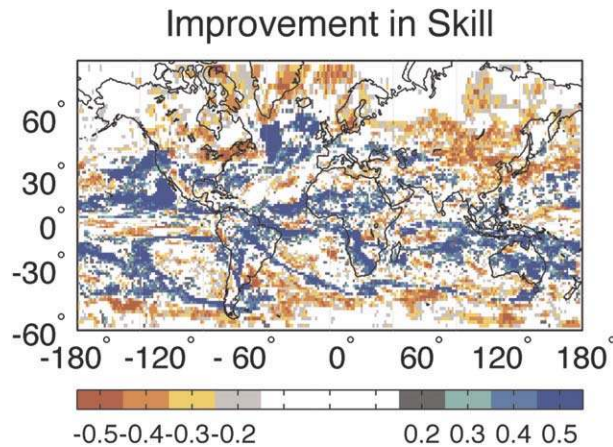


Figure 9. Difference in the skill of using the sign of the 2030–50 precipitation anomalies from individual model realizations to predict the sign of the 2080–2100 precipitation anomalies from the same realization, compared with the skill of using the sign of the 2030–50 multimodel ensemble-mean precipitation anomalies; anomalies calculated as deviations from the 2000–20 values. Skill is based upon a modified Hanssen–Kuipers skill score (see text for details). Only values greater than ± 0.2 (90% confidence level) are shown here; values in color are significant at 95% confidence level (± 0.25).

the 2080–2100 period in predicting the sign of any one of the individual model realizations is worse for seasonal-mean values [here December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON)] than for the annual means (not shown). In addition, the MMEM skill (and hence intermodel consistency) seems to decrease the most over high-latitude regions during hemispheric summer, including over most of Eurasia and North America during June–August and southern Africa and Australia during December–February. This discrepancy most likely arises because of differences in convective schemes and in the strength of land–atmosphere coupling in the various model systems.

If we compare the intermodel consistency with the intramodel consistency—derived by using the EM precipitation anomalies from the 2080–2100 period as the predictor for the individual ensemble members themselves—we find that the skill improves if a priori knowledge of the appropriate model system is incorporated into the prediction. This is particularly true over North America, the Sahel, and southwestern Asia (including northern India, Pakistan, and Afghanistan) during the June–August period; southern Africa and Australia during the December–February period; the Amazon basin during the March–May and September–November periods; and Indonesia and northern Australia during all four seasons. In addition, we find that the intramodel consistency of 2080–2100 precipitation trends is significant at all grid points, regardless of season (not shown).

Finally, we can determine whether the persistence of using short-term trends from individual model realizations provides additional skill in forecasting the sign of long-term trends in seasonal-mean precipitation as compared with using the

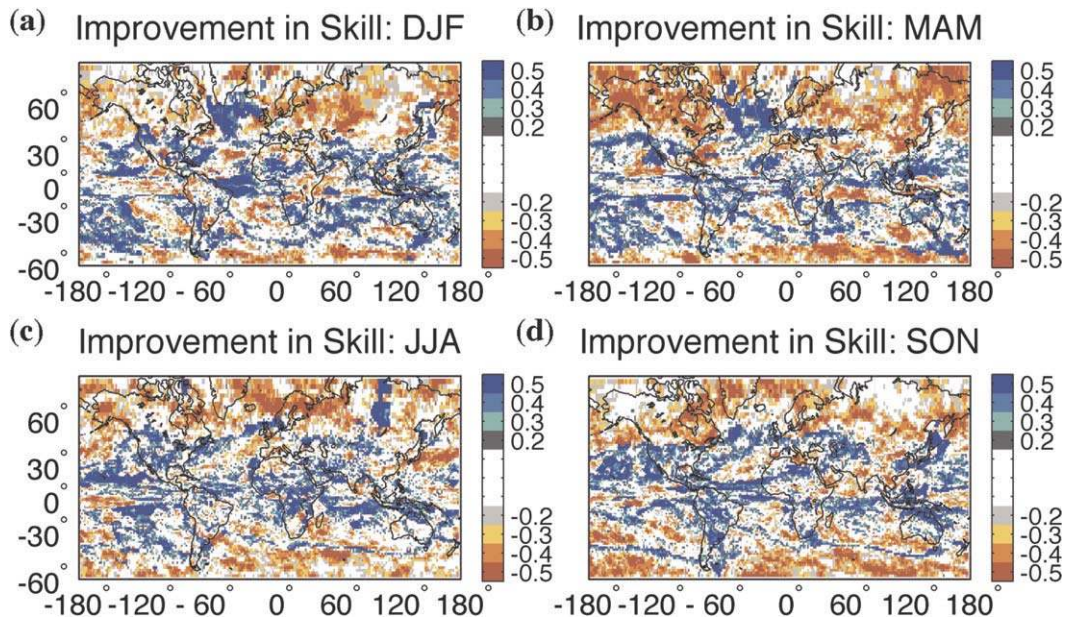


Figure 10. (a) Difference in the skill of using the sign of the 2030–50 DJF precipitation anomalies from individual model realizations to predict the sign of the 2080–2100 DJF precipitation anomalies from the same realization, compared with the skill of using the sign of the 2030–50 DJF multimodel ensemble-mean precipitation anomalies; anomalies are calculated as deviations from the 2000–20 values. Skill is based upon a modified Hanssen–Kuipers skill score (see text for details). Only values greater than ± 0.2 (90% confidence level) are shown here; values in color are significant at 95% confidence level (± 0.25). (b)–(d) Same as in (a), except for MAM, JJA, and SON precipitation anomalies.

MMEM estimates (Figure 10). As before the short-term trends are derived from the various predictor fields during the 2030–50 time period. While the short-term trends from the individual realizations do not provide as much skill in predicting their own long-term seasonal-mean evolution as compared with the skill found in annual-mean precipitation (not shown), they do provide better skill than the MMEM values over many regions, particularly over the tropics and subtropics. In addition, the improvement in skill tends to follow the seasonal cycle in precipitation. For instance, there is added skill in forecasting December–February precipitation changes over the southern portions of Africa, when precipitation tends to be greatest; similarly, over central Africa, there is added skill during June–August, again when the seasonal cycle of precipitation peaks in this region. Over the Amazon, improved skill is found during the onset (September–November) and retreat (March–May) of the monsoon rains. Over the southern portions of North America and Europe, skill appears to increase most during summer (June–August) and into fall (September–November). As before, results indicate that the MMEM estimates from 2030 to 2050 still serve as the best predictors for precipitation in the high-latitude regions of North America and Eurasia, particularly during March–May. However, in other

regions it appears that these MMEM forecasts can be augmented, or constrained, by historical and future observations of short-term trends.

5. Summary

Here we have analyzed the consistency in global climate change model predictions of regional precipitation trends. Results indicate that, while certain regions show high inter- and intramodel consistency in projections of regional precipitation responses to anthropogenic-induced climate change, other regions show little consistency in even the sign of these changes. In these regions, the large spread of individual model realizations, each of which is considered an equally plausible evolution of the actual climate system, precludes the use of multimodel ensemble means to make predictions of the response of the climate system to anthropogenic emissions of radiatively active chemical constituents. However, given a priori knowledge of the underlying (model) climate system, the skill of the projections is markedly improved, indicating that intramodel consistency is robust, even if intermodel consistency is not. This result suggests that the evolution of the system is constrained by the internal dynamics of the model itself; this result also suggests that the time-dependent behavior of a single realization of the climate system, which by definition is governed by the same model dynamics throughout its evolution, may be able to provide information about its own long-term time evolution.

To test this hypothesis we examine four specific regions—the Amazon basin in South America, the central and southern portions of Africa, and Indonesia—where intramodel consistency between ensemble members significantly improves compared with intermodel consistency. First, we quantify the skill of using the sign of the gridpoint precipitation anomalies derived from the ensemble-mean value of a given model system at some intervening time as the predictor for the sign of the precipitation anomalies at the end of the simulation period (e.g., 2080–2100) from individual realizations of the same model. We find this skill is significantly improved compared with the skill of using intervening gridpoint precipitation anomalies derived from the multimodel ensemble-mean value. In addition, short-term trends of *individual* model realizations also provide improved skill in predicting their own state by the end of the simulation period—in the four regions considered here, approximately 75% of the individual model gridpoint trends during the 2030–50 period correctly predict the sign of their own gridpoint trends during the final 20 years of the simulation period. While these results suggest that only after 50+ years of climate forcing does the short-term trend consistently match the long-term trend in these regions, the climate forcing associated with this 50-year period (captured by an approximate 100-ppm increase in carbon dioxide concentrations) is similar to that imposed over the last 100 years (1900–2000). These results suggest that precipitation trends during this historical period may provide guidance regarding the sign of future precipitation trends over the next 100 years; however, care must be taken in using short-term precipitation trends as predictors for longer-term trends because results are sensitive to the model system, geographic location, and time of year. At the same time, they do suggest that the actual evolution of the climate system, vis-à-vis precipitation changes, does inherently contain information about its own future evolution that can be used to augment model-based climate change projections.

Acknowledgments. Dr. Anderson’s research was supported by a Visiting Scientist appointment to the Grantham Institute for Climate Change, administered by Imperial College of Science, Technology, and Medicine. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP’s Working Group on Coupled Modeling (WGCM), for their roles in making available the WCRP CMIP3 multimodel dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.

References

- Allan, R., and B. Soden, 2007: Large discrepancy between observed and simulated precipitation trends in the ascending and descending branches of the tropical circulation. *Geophys. Res. Lett.*, **34**, L18705, doi:10.1029/2007GL031460.
- Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419**, 224–232.
- Barnett, T. P., and Coauthors, 2008: Human-induced changes in the hydrology of the western United States. *Science*, **319**, 1080–1083.
- Bhend, J., and H. von Storch, 2008: Consistency of observed winter precipitation trends in northern Europe with regional climate change projections. *Climate Dyn.*, **31**, 17–28.
- Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor, 2003: An overview of results from the Coupled Model Intercomparison Project. *Global Planet. Change*, **37**, 103–133.
- DelSole, T., 2004: Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.*, **61**, 2425–2440.
- , 2005: Predictability and information theory. Part II: Imperfect forecasts. *J. Atmos. Sci.*, **62**, 3368–3381.
- Giorgi, F., and R. Francisco, 2000: Evaluating uncertainties in the prediction of regional climate change. *Geophys. Res. Lett.*, **27**, 1295–1298.
- , and L. O. Mearns, 2002: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging” (REA) method. *J. Climate*, **15**, 1141–1158.
- Hanssen, A. W., and W. J. A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Meded. Verh.*, **81**, 2–15.
- Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699.
- Judd, K., and L. A. Smith, 2004: Indistinguishable states II: The imperfect model scenario. *Physica A*, **196**, 224–242.
- Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216.
- Lambert, F. H., N. P. Gillett, D. A. Stone, and C. Huntingford, 2005: Attribution studies of observed land precipitation changes with nine coupled models. *Geophys. Res. Lett.*, **32**, L18704, doi:10.1029/2005GL023654.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.
- Nakićenović, N., and Coauthors, 2000: *Special Report on Emissions Scenarios*. Cambridge University Press, 570 pp.
- Neelin, J. D., M. Münnich, H. Su, J. E. Meyerson, and C. E. Holloway, 2006: Tropical drying trends in global warming models and observations. *Proc. Natl. Acad. Sci. USA*, **103**, 6110–6115.

- Parry, M. L., O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, Eds., 2007: *Climate Change 2007: Impacts, Adaptation and Vulnerability*. Cambridge University Press, 976 pp.
- Räisänen, J., 2001: CO₂-induced climate change in CMIP2 experiments: Quantification of agreement and role of internal variability. *J. Climate*, **14**, 2088–2104.
- , 2007: How reliable are climate models? *Tellus*, **59**, 2–29.
- , and T. N. Palmer, 2001: A probability and decision model analysis of a multimodel ensemble of climate change simulations. *J. Climate*, **14**, 3212–3226.
- Robertson, A. W., U. Lall, S. E. Zebiak, and L. Goddard, 2004: Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Wea. Rev.*, **132**, 2732–2744.
- Shukla, J., T. DelSole, M. Fennessy, J. Kinter, and D. Paolino, 2006: Climate model fidelity and projections of climate change. *Geophys. Res. Lett.*, **33**, L07702, doi:10.1029/2005GL025579.
- Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.
- Sun, Y., S. Solomon, A. G. Dai, and R. W. Portmann, 2007: How often will it rain? *J. Climate*, **20**, 4801–4818.
- Syroka, J., and R. Toumi, 2001: Scaling and persistence in observed and modelled surface temperature. *Geophys. Res. Lett.*, **28**, 3255–3258.
- Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc. London*, **A365**, 2053–2075.
- Tomsett, A. C., and R. Toumi, 2001: Annual persistence in observed and modelled UK precipitation. *Geophys. Res. Lett.*, **28**, 3891–3894.
- Zhang, X., F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa, 2007: Detection of human influence on twentieth-century precipitation trends. *Nature*, **448**, 461–465.

Earth Interactions is published jointly by the American Meteorological Society, the American Geophysical Union, and the Association of American Geographers. Permission to use figures, tables, and *brief* excerpts from this journal in scientific and educational works is hereby granted provided that the source is acknowledged. Any use of material in this journal that is determined to be “fair use” under Section 107 or that satisfies the conditions specified in Section 108 of the U.S. Copyright Law (17 USC, as revised by P.L. 94-553) does not require the publishers’ permission. For permission for any other form of copying, contact one of the copublishing societies.
