LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-216730

Evaluation of Continental Precipitation in 20th-Century Climate Simulations: The Utility of Multi-Model Statistics

T. J. Phillips, P. J. Gleckler

November 1, 2005

Water Resources Research

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Evaluation of Continental Precipitation in 20th-Century Climate Simulations: The Utility of Multi-Model Statistics

Thomas J. Phillips and Peter J. Gleckler
Program for Climate Model Diagnosis and Intercomparison (PCMDI)
Lawrence Livermore National Laboratory
Livermore, California 94551

Revised submission to *Water Resources Research*
Accepted for publication on October 29, 2005

Corresponding Author Information:

Thomas J. Phillips
Program for Climate Model Diagnosis and Intercomparison (PCMDI)
Lawrence Livermore National Laboratory, Mail Code L-103
P.O. Box 808
Livermore, California 94551
Fax: 925-422-7675
E-mail: phillips14@llnl.gov

Abstract

At the request of the Intergovernmental Panel on Climate Change (IPCC), simulations of 20th-century climate have been performed recently with some 20 global coupled ocean-atmosphere models. In view of its central importance for biological and socio-economic systems, model-simulated continental precipitation is evaluated relative to three observational estimates at both global and regional scales. Many models are found to display systematic biases, deviating markedly from the observed spatial variability and amplitude/phase of the seasonal cycle. However, the point-wise ensemble mean of all the models usually shows better statistical agreement with the observations than does any single model. Deficiencies of current models that may be responsible for the simulated precipitation biases as well as possible reasons for the improved estimate afforded by the multi-model ensemble mean are discussed. Implications of these results for water-resource managers also are briefly addressed.

1. Introduction

Since its founding in 1988, the Intergovernmental Panel on Climate Change (IPCC) has issued periodic reports that assess the state of scientific knowledge of climate change and its biological and socio-economic impacts (see online information at http://www.ipcc.ch). Presently, the IPCC is composing its Fourth Assessment Report (FAR), which is scheduled for publication in 2007.

An important part of this process is the IPCC's coordination of an unprecedented climate modeling effort that involves the participation of international centers that have developed some 20 global coupled ocean-atmosphere general circulation models (OAGCMs). In addition to climate-change simulations based on various greenhouse-emission projections, IPCC-mandated "control" runs include those designed to reproduce the climate of the 20th century (designated as 20C3M simulations) with historically increasing greenhouse-gas concentrations. The Program for Climate Model Diagnosis and Intercomparison [PCMDI, 2005] is coordinating the distribution of the associated model output data in a standardized format to more than 300 analysis projects that are intensively studying different facets of these IPCC simulations.

In view of its central importance for human and other life forms, there is a need to analyze the simulated terrestrial hydrological cycle, since it is here that greenhouse-induced climate change may have especially severe consequences. However, in order to judge the relative likelihood of this projected climate change, the respective IPCC simulations of historical 20th-century land hydrology must first be assessed. Total continental precipitation (i.e. rainfall plus snowfall) is central to such an evaluation, since other hydrological components such as runoff/streamflow, evapotranspiration, snowpack, and soil moisture all depend on this basic forcing. Moreover, unlike some components of land hydrology, the IPCC 20C3M simulations of

continental precipitation can be validated by direct comparison with historic observational records. (See also evaluations of continental runoff/streamflow and of water storage in the IPCC 20C3M simulations by Milly et al. [2005] and Swenson and Milly [2005].)

The representation of precipitation in present-day global climate model is typically quite rudimentary. Stratiform precipitation (e.g. forming in frontal layer clouds) is computed at model grid-scale (typical resolution ~ 1-3 º latitude/longitude) either from a humidity-dependent condensation criterion or from a simple representation of cloud water content. Whether the precipitate falls as rain or as snow is commonly determined by a temperature criterion that is applied at the atmospheric model level where condensation occurs, and snow usually only accumulates at ground level when the surface temperature is below freezing. Further, the cloud microphysics that regulate raindrop formation, the fraction of precipitation that forms in convective towers, and the evaporation of falling precipitate all are inherently sub-grid scale processes that must be parameterized in terms of computable grid-scale quantities (e.g. atmospheric humidity, temperature, wind speed and vertical motion). The details of these moisture parameterizations vary considerably among different climate models that are implementing the IPCC-mandated simulations (e.g. see http://www-pcmdi.llnl.gov/ ipcc/ model_ documentation/ipcc_model_documentation.php for summary documentation).

In view of the many simplifications inherent in such parameterizations, today's global climate models might generally be expected to poorly represent the magnitudes and patterns of continental precipitation. In fact, it will be shown that the models implementing IPCC 20C3M simulations range widely in their abilities to replicate these aspects of the historical record. Further, it will be demonstrated that point-wise averaging over the multi-model ensemble of simulations usually provides a statistically superior estimate of continental precipitation.

The global OAGCMs analyzed in this study are listed in Table 1, along with their institutional sponsors and countries. The horizontal and vertical resolution of the atmospheric components of these OAGCMs also are listed. The typical horizontal resolution is ~ 2º x 2º latitude/longitude, although there are several instances of model grids as coarse as 4º x 5º or as fine as ~1º x 1º. The number of atmospheric model vertical levels varies between 12 and 56, with the typical model having more than 20 levels.

Because the inclusion/specification of 20th-century variations in solar irradiance, land cover, and concentrations of greenhouse gases and ozone, and of naturally occurring or anthropogenic aerosols were not mandated by the IPCC, these modeling choices also are indicated in Table 1. It is seen that although the models all account to some extent for historical variations in greenhouse gases and sulfate aerosols, they otherwise differ widely in their inclusion of additional natural and anthropogenic forcings.

In addition, Table 1 supplies information on the surface ocean-atmosphere flux adjustments [e.g. Sausen et al. 1988] which a few models employed to constrain "coupled climate drift". This drift results from coupling atmosphere and ocean models which display inconsistencies in surface fluxes of heat, momentum, and fresh water when in their respective decoupled equilibrium states. It is indicative of a general technical advance in coupled model development that most of today's OAGCMs no longer employ such flux adjustments, and those models that still do so often adjust surface fluxes in only a spatially or temporally restricted way.

Although some models ran multiple 20C3M runs starting from different initial conditions, only a single simulation by each model was considered for purposes of this study. In evaluating model performance, however, we cannot expect any 20C3M simulation to replicate the observed sequence of 20th-century climatic events (e.g. timing and strengths of historical El Niños, heat

waves, droughts and floods, etc.) which manifest the chaotic internal variability of the global climate system.

Instead of considering specific climate fluctuations, climate statistics should be compared. This study evaluates climatological means of simulated continental precipitation relative to those observed for the years 1980-1999. Several global terrestrial precipitation datasets overlap during this period, thus allowing multiple validations of a model as well as the estimation of current observational uncertainties. Validation of the latter part of a 20th- century simulation also more stringently tests a coupled model, which is prone to drift increasingly from the observations the longer the run proceeds.

Different global-scale observational datasets for validating simulated continental precipitation are described in Section 2, and their use in evaluating model climatologies is elaborated in Section 3. Selective comparisons of the annual cycle of model precipitation with these observational datasets on sub-continental and regional scales are presented in Section 4. In Section 5 the main findings and interpretations of this study are summarized, and their implications for water resource managers are briefly discussed.

2. Datasets for Precipitation Evaluation

In evaluating continental precipitation in the IPCC 20C3M simulations, global-scale datasets are of paramount importance. For example, scientists with a particular interest in the simulation of land-surface climate commonly use the 0.5º x 0.5º gauge-based Climatic Research Unit dataset of 20th-century global terrestrial precipitation [New et al., 2000] for model validation.

Other global precipitation estimates that are in common use for model validation over periods after 1979 include the 2.5º x 2.5º datasets of the Global Precipitation Climatology Project (GPCP) [Huffman et al., 1997] and of the Climate Prediction Center's Merged Analysis of

Precipitation (CMAP) [Xie and Arkin, 1997]. Although these three datasets are not totally independent in their origins, they do draw on somewhat different networks of precipitation gauge measurements. The datasets also differ in their choice of algorithms for interpolation of station observations to the dataset grid. In addition, GPCP precipitation reflects corrections applied to the raw gauge data in order to account for systematic measurement errors related to evaporation, wind-blown rain/snow, etc. In contrast to CRU, the GPCP and CMAP datasets also include satellite-based estimates of precipitation over the oceans. Further information on such details may be found in the cited references.

A map of the GPCP annual-mean continental precipitation is shown in Figure 1a. Note, for example, the precipitation maxima (~ 8-10 mm day^{-1}) over Amazonia and the Equatorial West Pacific islands, the large magnitudes (~ 5-6 mm day^{-1}) in Equatorial Africa and Southern Asia, but also the scant precipitation (< 1 mm day^{-1}) prevailing in the subtropical deserts that extend across stretches of Northern and Southern Africa, Central Asia, and the Southwest U.S., as well as Greenland and Antarctica.

CMAP-GPCP differences in annual-mean precipitation are shown in Figure 1b. The largest positive differences (~ 2 mm day^{-1}) occur over the Himalayas, while the most negative ones (~ -1 mm day^{-1}) are found in a small portion of Northern Eurasia; otherwise, the CMAP precipitation differences are within ±1 mm day^{-1} of those in the GPCP dataset.

After mapping the 0.5º x 0.5º CRU to the 2.5º x 2.5º GPCP grid, annual-mean CRU-GPCP precipitation differences are comparatively greater and more spatially extensive (Figure 1c). Relatively large positive differences (~ 2-4 mm day^{-1}) are found over Northern Brazil and the Andean regions, Greenland, and Southern Asia in the vicinity of the Himalayas. Less spatially extensive negative differences (~ -1 mm day^{-1}) are seen in Northern Eurasia and Southern Alaska.

A concise comparison of the spatial variability statistics of these 3 continental precipitation datasets is provided by the Taylor plot [Taylor, 2001] shown in Figure 2a. The radial dimension of this polar plot indicates the amplitude of the aggregated area-weighted spatial variability of the continental CMAP and CRU climatologies (points O2 and O3) about their respective global means, relative to the normalized aggregated spatial variability of the GPCP data (point O1), which is denoted by the quarter circle of unit radius. The aggregated area-weighted similarity of the spatial variability patterns of the CMAP or CRU climatologies to that of the GPCP reference, expressed as the cosine of their respective pattern correlations, is indicated in the azimuthal dimension of the Taylor plot. It also can be shown [Taylor, 2001] that the area-weighted root-mean-square (RMS) differences between the CMAP or CRU data and the GPCP reference, computed after subtracting their respective global means, are proportional to the straight-line distances O2—O1 or O3—O1 in Figure 2a. These *centered* RMS differences are seen to result from combined differences in variability amplitude and pattern, indicated respectively by radially and azimuthally directed displacements on the Taylor plot.

The spatial variability of the annual-mean CMAP continental precipitation therefore exhibits a high degree of aggregate statistical similarity (in both amplitude and pattern) with the GPCP reference data, consistent with the relatively small CMAP-GPCP differences mapped in Figure 1b. (The CMAP and GPCP datasets show less similarity over the oceans, where different algorithms for estimating precipitation from satellite retrievals are applied—see Gruber et al. 2000 for details.) In contrast, the aggregate spatial variability of the annual-mean CRU terrestrial precipitation data shown in Figure 1c is about 30% greater in magnitude than that of the GPCP land reference data, and its spatial pattern correlates less closely with GPCP than does the corresponding CMAP data (CRU correlation ~ 0.80 vs. ~ 0.95 for CMAP).

3. Large-Scale Evaluation of Precipitation Simulations

In the Taylor plot of Figure 2b, the spatial variabilities of 20-year annual-mean climatologies of continental precipitation from 19 IPCC 20C3M runs (see the legend for corresponding model names) also are shown relative to the annual-mean spatial variability of the GPCP reference data (point O1). Here, all the model climatologies are remapped to the 2.5º x 2.5º GPCP grid and are area-weighted accordingly in computing the aggregated variability amplitude ratios and pattern correlations. The corresponding multi-model ensemble mean, defined as the point-wise arithmetic average over all individual model climatologies, is denoted by point MM. (Note, in this context, "ensemble" signifies a collection of different model climatologies, rather than a set of simulations with different initial conditions.)

The simulation climatologies also display a wide range of spatial pattern correlations--from ~ 0.55 to ~ 0.85. Most of these climatologies also show higher amplitudes of annual-mean spatial variability than the GPCP reference data, as indicated by the generally greater radial displacements of the simulation points 1, 2, …,19 relative to point O1. (That is, most points lie outside the plot's unit-radius quarter-circle that denotes the same spatial variability amplitude as reference point O1 ≡ GPCP .) Many simulation climatologies nonetheless display variability amplitudes that fall within a range bounded by the GPCP reference (point O1) and the alternative CRU validation data (point O3), which serves as a rough estimate of current observational uncertainties. A number of the simulated annual-mean climatologies also display RMS differences from GPCP that are less than that between CRU and this reference (i.e. distance O3—O1 in Figure 2b).

The collective performance of present-day OAGCMs relative to that of previous-generation coupled models is of considerable interest to the IPCC. The OAGCMs considered in this study

appear to have annual-mean precipitation statistics (Figure 2b) that are marginally improved over those of older coupled models. For instance, Covey et al. [2003] display a Taylor plot of previous-generation OAGCM control simulations from the Coupled Model Intercomparison Project (CMIP) in which the spatial pattern correlations of global precipitation with Xie-Arkin [1997] reference data ranged between ~0.4-0.7, as compared with global correlations of ~0.5-0.8 for the current IPCC runs (figure not shown). It should be noted, however, that the IPCC 20C3M runs with transient forcings are not directly comparable to the CMIP control runs, which used time-invariant forcings. More precise comparisons of the overall performance of today's OAGCMs, as exhibited in other IPCC simulations that resemble the CMIP control runs, are currently in progress [personal communication, K. Taylor].

In order to investigate possible seasonal dependencies in the IPCC 20C3M precipitation simulations, Taylor diagrams also are plotted for 20-year mean January and July climatologies in Figures 2c and 2d. The biases in variability amplitudes of most annual-mean model climatologies in Figure 2b are less pronounced in January, when the maximum ratio of model to GPCP spatial variability amplitude is ~ 1.4 as compared with ~1.6 in the annual-mean cases. However, this model amplitude bias is larger in a relative sense in January, since the range in observational annual-mean climatologies (bounded by O1 and O3 in Figure 2b) is much-reduced in Figure 2c. In July (Figure 2d), the model amplitude biases are generally larger than in January, but the O1—O3 range of observational amplitude ratios also is greater. In contrast, the spatial variability amplitude of the multi-model ensemble-mean statistic MM is only slightly lower than that of GPCP in all cases.

Many simulations show substantial pattern errors as well; these are generally larger in July (Figure 2d), when the spatial correlations range from ~0.5 to ~0.8. The performance of a global

model in simulating continental precipitation often falls off in boreal summer relative to winter. This is probably a consequence of the heightened importance of convective mechanisms in summer precipitation, and thus the greater need to rely on sub-grid scale parameterizations instead of model-resolved hydrodynamics to simulate it.

The aggregated statistics displayed in Figure 2 also imply that overall model performance in simulating precipitation is not determined solely by resolution: some models with coarser grids outperform those with finer ones (refer to Table 1). This is most clearly demonstrated by comparing simulations 12 and 13 in Figure 2, which correspond to versions of the MIROC3.2 model with the same physical parameterizations, but implemented at "high" versus "medium" resolution (MIROC3.2 (hires) at 1.1º x 1.1º L56 resolution vs. simulation 13 with MIROC3.2 (medres) at 2.8º x 2.8º L20). The centered RMS differences from the GPCP reference in Figure 2 are substantially lower for the medium-resolution simulation 13 than for the high-resolution simulation 12, principally because of better agreement in spatial-variability amplitude of the former with the GPCP data. The better precipitation simulation of the MIROC3.2 (medres) model may be due to a better match of the model dynamics with the operative scale of the moisture parameterizations [e.g. Phillips et al., 1995].

There also does not appear to be an obvious relationship between model performance and the inclusion of natural and anthropogenic forcings, apart from greenhouse gases and sulfate aerosols (see Table 1). The impact of some models' inclusion of surface flux adjustments is less clear-cut, however. For example, the generally superior statistical characteristics of precipitation displayed by the MRI-CGCM2.3.2 model may be partly due to its application of tropical adjustments in the surface fluxes of heat, fresh water, and momentum.

It is especially noteworthy that the RMS difference between the multi-model ensemble mean MM and the GPCP reference (proportional to the straight-line distance MM—O1 in Figure 2) is usually less than that of any single model simulation. (In January, however, the RMS difference of simulation 15 corresponding to model MRI-CGCM2.3.2 is essentially identical to that of MM.) The superior agreement of the ensemble-mean statistic with the GPCP reference is most apparent in July (Figure 2d) when the RMS differences of individual simulations from the GPCP reference clearly fall outside the envelope of observational uncertainties implied by differences among the alternative validation data (i.e. as denoted by the triangle O1—O2—O3). Note, however, that the RMS difference between the ensemble-mean climatology and the GPCP reference (distance MM—O1) is still considerably larger than that between the CMAP and GPCP datasets (distance O2—O1). Nonetheless, the apparent biases of the individual models are reduced by computing a point-wise average MM of the entire ensemble of simulations.

There probably are several reasons for this outcome, which has been noted in other model intercomparisons [e.g. Lambert and Boer, 2001; PCMDI, 2004]. First, to the extent that the point-wise RMS difference between a simulated climate field and an observationally based estimate is symptomatic of random error, averaging over many model simulations is likely to reduce this RMS difference. Moreover, this averaging smoothes out point-wise variations in a climatic field, thereby filtering simulated phenomena on regional scales, where current climate simulations are known to be especially problematical. However, further investigation is needed to more deeply understand the statistical characteristics of the multi-model ensemble mean MM [personal communication, K. Taylor].

This regional filtering may be especially important for reducing errors in precipitation, which displays such rich fine-scale structure. Although the regional-scale shortcomings in GCM

precipitation simulations are partly due to numerical truncation errors associated with coarse model resolution, they may be due even more to flaws in parameterizations (e.g. in sub-grid scale representations of cloud microphysics, convection, etc.) which are manifested as precipitation errors.

In spite of the superior estimate of precipitation statistics afforded by the multi-model ensemble mean, nontrivial errors that are indicative of general model shortcomings remain. These errors are illustrated in Figure 3 by differences between the MM and respective GPCP climatologies. The largest positive MM-GPCP differences (~ 6 to 8 mm day^{-1}) mostly occur in the vicinity of orography (notably the Andes and Himalayas, but also mountainous regions of Mexico and Southern Africa), which is poorly resolved by the typically coarse horizontal resolution of the models (Table 1). These orographic effects also appear to be magnified in the summer hemisphere (i.e. during January in the Southern Hemisphere and during July in the Northern Hemisphere) where there are higher amounts of available atmospheric moisture.

Other relatively large positive MM-GPCP precipitation differences (~ 4 to 5 mm day^{-1}) occur in January along the Pacific coast of North America and the Atlantic coast of Brazil, as well as the central U.S. in July. These differences, together with the relatively large negative MM-GPCP departures (~ -2 to -4 mm day^{-1}) in regions with strong monsoonal circulations (e.g. in southern Asia and sub-Saharan Africa in July and in Amazonia in January), suggest there are general model shortcomings in simulating ocean-land moisture transports, and possibly also continental convection.

Both the magnitude and pattern of precipitation differences between the MM and GPCP annual/seasonal climatologies are similar to those which Lambert and Boer [2001] noted in the ensemble means of the CMIP models. Unlike these previous-generatin models, however, most

of today's OAGCMs do not employ flux adjustments to constrain their surface ocean states close to observations (Table 1).

4. Finer-Scale Evaluation of Precipitation Simulations

In order to compare the entire annual cycle of the multi-model ensemble-mean field with that of the GPCP reference precipitation, the respective monthly climatologies were subjected to harmonic analysis [e.g. Kirkyla and Hameed 1989]. In this technique, the time series of 12 monthly climatological fields in different continental sectors is decomposed into 6 orthogonal harmonics that are associated with progressively higher (e.g. annual, semi-annual, etc.) frequencies.

For example, the amplitude and phase of the annual harmonic of precipitation in the North American sector are shown in Figure 4 for the GPCP reference (top panel), the MM statistic (middle panel), and the MM-GPCP difference field (bottom panel). The ensemble-mean precipitation is seen to be too scant in the Great Lakes and Gulf Coast regions, but is excessive along the U.S. West Coast (consistent with the MM-GPCP differences seen in Figure 3b). Moreover, the transition from a U.S. West-Coast precipitation maximum to a Southwest minimum is not well-represented. The phase of the MM annual harmonic also differs markedly from that of the GPCP reference in the Southwestern U.S.

In Figure 5a, the annual cycles of individual IPCC precipitation simulations in the Southwestern U.S are compared with those of the GPCP, CMAP, and CRU observations. Prominent observational features include the sharp onset of summer rainfall associated with the North American Monsoon and the scant precipitation during other seasons. Most models fail to capture this seasonal variation, as implied by the anomalous semi-annual component in the simulation ensemble mean (solid black line). Except in summer, simulated precipitation

magnitudes are also generally too high.

Summer monsoonal precipitation is replicated more realistically over India (Figure 5b). However, the models tend to under-simulate its observed magnitude (consistent with the ensemble-mean differences seen in Figure 3c) and to lag in the timing of its onset. Harmonic analysis of the MM climatological cycle over the Asian subcontinent and neighboring ocean (not shown) indicates that the observed monsoonal rainfall maximum near the Ghats mountains in western India is only weakly simulated, probably due to the coarse rendering of orography by most models. The simulated ensemble-mean rainfall maximum over the Arabian Sea also occurs later than observed, consistent with the tendency for Indian monsoon onset to lag in the models.

The seasonal variation of precipitation over the semi-arid Sahel region of Africa is also credibly simulated by many models (Figure 5c), although several substantially under-simulate its summer maximum. Harmonic analysis of the Northern African sector (not shown) indicates that the phase of the MM annual harmonic generally agrees well with that of the GPCP reference, but amplitude differences are as large as $\sim \pm 1$ mm day^{-1}.

In the South American sector, the amplitude of the MM annual harmonic is under-simulated in the Intertropical Convergence Zone (ITCZ) over northern Brazil, while it is over-simulated in the Andes and in eastern Brazil, where the phase of the annual harmonic also is poorly represented (not shown). For example, in the Nordeste coastal region of Brazil, most models roughly capture the timing of the peak rainfall that is observed in austral autumn (Figure 5c). However, the peak magnitude is generally over-simulated (as also seen in the MM-GPCP differences of Figure 3b), while Nordeste precipitation during the drier part of the year is mostly under-simulated.

In general, the MM statistic (solid black line in Figures 5a-d) qualitatively captures the observed annual cycle of precipitation in the selected regions, but with some lapses in amplitude or phase that reflect shortcomings of the individual simulations in attenuated form. Thus, the apparent error-reduction resulting from ensemble averaging and aggregating the simulations on regional scales seems less dramatic than what is obtained globally (e.g. point MM in Figures 2b-d), where the aggregation is over many more grid points.

5. Discussion

Many differences of IPCC 20C3M simulations of continental precipitation from the GPCP, CMAP, and CRU observational estimates appear to reflect systematic model biases. In particular, most models display aggregated precipitation variability magnitudes that are larger than observed, a phenomenon that is especially pronounced in climatological January when there is closer agreement among alternative observational estimates.

Aggregated precipitation pattern errors appear to be only modestly reduced relative to those of previous-generation OAGCMs [Covey et al., 2003]. However, such incremental improvements are rendered more impressive for having been achieved by models that mostly no longer employ flux adjustments to constrain climate drift.

On finer scales, model precipitation is generally excessive in some regions (e.g. U.S. Southwest, Brazilian Nordeste), but is markedly under-simulated in others (e.g. India). Present-day OAGCMs also have difficulty simulating the timing of regional peak precipitation. Thus, the problematical simulation of regional-scale precipitation by previous-generation OAGCMs noted in the last IPCC assessment [Houghton, 2001] appears to persist in today's coupled models. However, this study suggests that the precipitation climatologies of some regions (e.g. the African Sahel) are inherently more predictable than those of others (e.g. U.S. Southwest).

It should be emphasized that the spatially aggregated precipitation variability of the multi-model ensemble-mean (MM) climatology usually shows more agreement with observational data than that of any single model. There are, however, nontrivial regional biases in MM precipitation magnitudes relative to the GPCP reference data that are indicative of systematic model biases. These discrepancies seem to result mainly from problematical representations of orographic effects and monsoonal circulations which are probably related to the OAGCMs' generally coarse horizontal resolution [e.g. Mo et al. 2005].

Precipitation pattern errors in individual 20C3M model simulations also persist to lesser degree in the ensemble-mean climatology. These pattern errors are especially apparent in monsoonal and coastal regions, suggesting that some discrepancies result from shortcomings in simulating ocean-land moisture transports, and possibly also continental convection. Thus, qualitative improvement in the simulation of precipitation by OAGCMs probably will require improvements in moist parameterizations in addition to increases in resolution.

In view of the noted model deficiencies, simulation of precipitation intensity, frequency, and duration is also likely to be problematical [e.g. Trenberth et al., 2003]. Further analysis of the hydrological cycle in the IPCC simulations should therefore include evaluation of these additional precipitation characteristics, especially since these may be sensitive to greenhouse-induced climate change.

The results of this study would seem to present mixed messages to water-resource managers. On the one hand, the substantial precipitation errors present in many 20C3M climatologies indicate that the typical global climate model must be significantly improved before it can serve as a practical planning tool, especially on regional scales. On the other hand, the enhanced approximation of observed precipitation provided by the multi-model ensemble-mean

climatology is a strong argument for utilizing multiple climate models in water-resource applications, rather than just a single "good" one.

There is also some evidence [e.g. Koster et al. 2002, 2004] that a multi-model approach is necessary when considering interactions between continental precipitation and surface processes (e.g. evapotranspiration, soil moisture, and runoff) that vary according to model-specific land-atmosphere coupling strengths. Use of multiple models also seems advisable for determining maximum-likelihood projections of climate change in different greenhouse-emission scenarios, as these pertain to water-resource planning. In this case, it also may be possible to obtain a better estimate of maximum likelihood than can be obtained from the multi-model ensemble-mean statistic by weighting each climate-change simulation according to the associated model's relative performance in replicating 20th-century climatologies.

Acknowledgments

This paper exemplifies studies currently being conducted by the IPCC Global Hydroclimates Analysis Project, whose other participants include Q. Duan, K.A. Dunne, P.C.D. Milly, C. A. Schlosser, C.S. Swenson, and J. Wahr. Information on model inclusions of 20th-century variations in natural and anthropogenic forcings or surface flux adjustments was contributed by B. Santer and K. Sperber of PCMDI. We also gratefully acknowledge the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP), the Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy (USDOE). This work was performed under the auspices of the USDOE Office of Science, Biological and Environmental Research (BER) program by the University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

References

Covey, C., K. 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 (CMIP). *Global Planet. Change*, *37*, 103-133.

Gruber, A. X. Su, M. Kanamitsu, and J. Schemm (2000), The comparison of two merged rain gauge-satellite precipitation datasets. *Bull. Amer. Meteor. Soc*., *81*, 2631-2644.

Huffman, G.J. and coauthors (1997), The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc*., *77*, 891-905.

Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.) (2001), *Climate Change (2001), The Scientific Basis*, *Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. 881 pp., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kirkyla, K.I., and S. Hameed (1989), Harmonic analysis of the seasonal cycle in precipitation over the United States: A comparison between observations and a general circulation model. *J. Climate*, *2*, 1463-1475.

Koster, R.D., P.A. Dirmeyer, A.N. Hahmann, R. Ijpelaar, L. Tyahla, P. Cox, and M.J. Suarez (2002), Comparing the degree of land-atmosphere interaction in four atmospheric general circulation models. *J. Hydrometeor.*, *3*, 363-375.

Koster, R.D., P.A. Dirmeyer, Z. Guo, G. Bonan, E. Chan, P. Cox, C.T. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C-H. Lu, S. Malyshev, B. McAvaney, K. Mitchell, D. Mocko, T. Oki, K. Oleson, A. Pitman, Y.C. Sud, C.M. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada (2004), Regions of strong coupling between soil moisture and precipitation. *Science*, *305*, 20 August 2004, 1138-1140.

Lambert, S.J., and G.J. Boer (2001), CMIP1 evaluation and intercomparison of coupled climate models. *Climate Dyn*., *17*, 83-106.

Milly, P.C.D., K.A. Dunne, and A.V. Vecchia (2005), The global fingerprint of forced hydro-climatic change on land during the 20th and 21st centuries. Accepted for publication in *Nature*.

Mo, K.C., J-K. Schemm, H.M.H. Juang, R.W. Higgins, and Y. Song (2005), Impact of model resolution on the prediction of summer precipitation over the United States and Mexico. *J. Climate, 18*, 3910-3927.

New, M., M. Hulme, and P.D. Jones (2000), Representing twentieth-century space-time climate variability. Part II: Development of 1901-1996 monthly grids of terrestrial surface climate. *J. Climate*, *13*, 2217-2238.

PCMDI (2004), An appraisal of coupled climate model simulations. Editor: D. Bader, contributing authors: K. AchutaRao, C. Covey, C. Doutriaux, M. Fiorino, P. Gleckler, T. Phillips, K. Sperber, and K. Taylor. Report UCRL-TR-202550, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California, 183 pp. Accessible online at http://www-pcmdi.llnl.gov/model_appraisal.pdf .

PCMDI (2005), Data archive of IPCC simulations and related information, accessible from http://www-pcmdi.llnl.gov/ about_ipcc.php .

Phillips, T.J., L. C. Corsetti, and S.L. Grotch (1995), The impact of horizontal resolution on moist processes in the ECMWF model. *Climate Dyn*., *11*, 85-102.

Sausen, R., K. Barthels, and K. Hasselmann (1988), Coupled ocean-atmosphere models with flux correction. *Climate Dyn*., *2*, 145-163.

Swenson, S.C., and P.C.D. Milly (2005), Climate-model biases in seasonality of continental water storage revealed by satellite gravimetry. *Water Resources Res.* (submitted).

Taylor, K.E. (2001), Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res., 106*, 7183-7192.

Trenberth, K.E., A. Dai, R.M. Rasmussen, and D.B. Parsons (2003), The changing character of precipitation. *Bull. Amer. Meteor. Soc*., *84*, 1205-1217.

Xie, P., and P. Arkin (1997), Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc*., *78*, 2539-2558.

Table 1: Model designations and related sponsors for 19 coupled ocean-atmosphere simulations analyzed in this study. The resolution of each model also is listed, where horizontal resolution is expressed as degrees latitude x degrees longitude (approximated in models using spectral representations of climate fields) and vertical resolution is denoted by the number of levels L. Also indicated are forcings varying on inter-annual and longer scales that are accounted for in each 20th-century climate simulation. These include temporal variations in solar irradiance (**I**) and land cover (**L**), in the concentrations of greenhouse gases (**G)** and ozone (**O**), of volcanic (**V**) and other naturally occurring aerosols (**N**), and of sulfate (**S**) and other anthropogenic (**A**) aerosols. In the last column, adjustments in ocean surface fluxes of heat (**H**), freshwater (**FW**), or momentum (**M**) that are employed by some models to constrain coupled climate drift also are indicated. Further documentation of model/experimental attributes may be accessed online at http://www-pcmdi.llnl.gov/ipcc/ model_documentation/ ipcc_model_documentation.php .

Model Name	Model Sponsor(s) (Country)	Atmospheric Resolution	Variable Forcings	Surface Flux Adjustments
CGCM3.1	Canadian Centre for Climate Modelling and Analysis (Canada)	3.8ºx3.8º L31	G,S	Annual H, FW
CNRM-CM3	Meteo-France/Centre National de Recherches Meteorologiques (France)	2.8ºx2.8º L45	G,O,S,A	
CSIRO-Mk3.0	CSIRO Atmospheric Research (Australia)	1.9ºx1.9º L18	G,S	
ECHAM5/MPI-OM	Max Planck Institute for Meteorology (Germany)	1.9º x 1.9º L31	G,O,S	
FGOALS-g1.0	LASG/Institute of Atmospheric Physics (China)	2.8ºx2.8º L26	G,S	
GFDL-CM2.0	U.S. Department of Commerce/ NOAA Geophysical Fluid Dynamics Laboratory (USA)	2.0º x2.5º L24	I,L,G,O, V, S,A	
GFDL-CM2.1		2.0º x2.5º L24	I,L,G,O, V,S,A	
GISS-AOM	NASA/Goddard Institute for Space Studies (USA)	3.0ºx4.0º L12	I,G,N,S	
GISS-EH		4.0ºx5.0º L12	I,L,G,O, V,N,S,A	
GISS-ER		4.0ºx5.0º L12	I,L,G,O, V,N,S,A	
INM-CM3.0	Institute for Numerical Mathematics (Russia)	4.0ºx5.0º L21	I,G,S	Annual high-latitude FW
IPSL-CM4	Institute Pierre Simon Laplace (France)	2.5º x 3.8º L19	G,S	
MIROC3.2(hires)	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC) (Japan)	1.1º x 1.1º L56	I,L,G,O, V,N,S,A	
MIROC3.2(medres)		2.8º x 2.8º L20	I,L,G,O, V,N,S,A	
MRI-CGCM2.3.2	Meteorological Research Institute (Japan)	2.8º x 2.8º L30	I,G,S	Tropical monthly H, FW, M
NCAR-CCSM3	National Center for Atmospheric Research (USA)	1.4º x 1.4º L26	I,G,O, V,S,A	
NCAR-PCM		2.8º x 2.8º L26	I,G,O,V,S	
UKMO-HadCM3	Hadley Centre for Climate Prediction and Research / Met Office (United Kingdom)	2.5º x 3.8º L19	G,O,S	
UKMO-HadGEM1		1.3º x 1.9º L38	I,L,G,O, V,S,A	

Figure Captions

Figure 1: Annual-mean continental precipitation climatology for the period 1980-1999, as estimated from the GPCP dataset in a), and differences from this GPCP climatology displayed by the CMAP dataset in b), and by the CRU terrestrial dataset (not including Antarctic data) in c). All units are in mm day^{-1}.

Figure 2: The Taylor diagram in a) indicates the amplitude ratio and pattern correlation of the aggregated spatial variability of the CMAP (point O2) and CRU (point O3) 1980-1999 annual-mean observational climatologies about their respective global means, relative to the normalized aggregated spatial variability of the GPCP observational reference data (point O1). The variability amplitude and pattern correlation relative to GPCP data also are shown for precipitation climatologies of model simulations 1, 2, …,19 in b) annual-mean, c) January-mean, and d) July-mean cases. The point MM indicates the corresponding aggregated spatial variability of the point-wise multi-model ensemble mean. The aggregated centered RMS differences between the GPCP reference climatology and each other observational and model climatology is proportional to the straight-line distance between the point O1 and the respective other points. See text for further details.

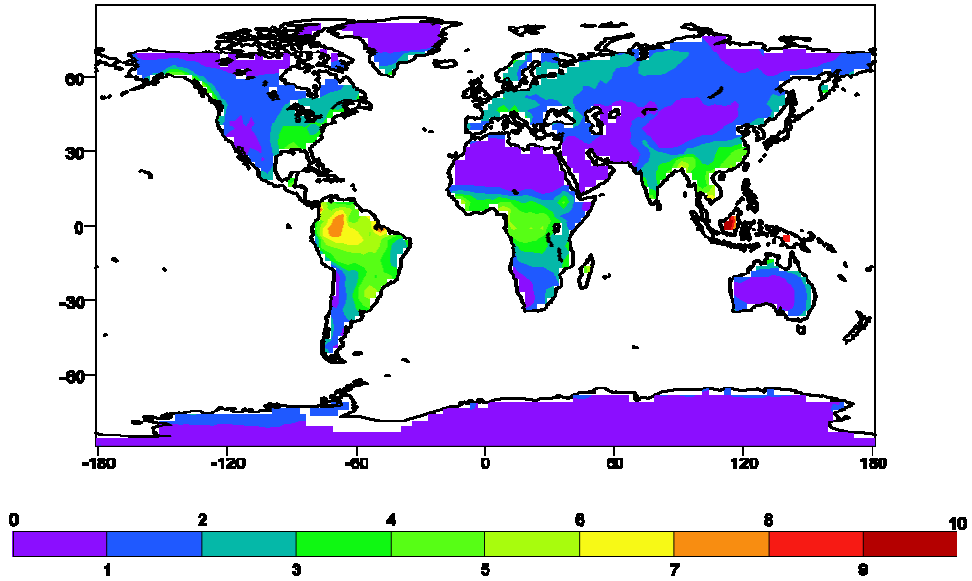
Figure 3: Geographical distribution of differences in 20-year continental precipitation climatologies of the multi-model ensemble mean (point MM in Figures 2b-d) from the GPCP observational reference for a) annual, b) January, and c) July climatologies (all in units of mmday^{-1}).

Figure 4: First harmonic of the climatological annual cycle of precipitation for the GPCP reference dataset (top panel), the multi-model ensemble mean MM (middle panel), and the difference field MM-GPCP (bottom panel). The amplitude of the first harmonic (in mm day^{-1}) in each case is indicated by shading. The spatial variation of the phase of the harmonic is expressed as the month of the annual-cycle maximum, indicated by the orientation of the included vectors (shown only for every other grid box). In the top and middle panels, a January maximum is indicated by a northward-pointing vector (12 o'clock orientation), an April maximum by an eastward-pointing vector (3 o'clock orientation), etc.; in the bottom panel, a zero difference in phase between MM and GPCP is indicated by a 12 o'clock orientation, while MM lagging GPCP by 3 months is indicated by a 3 o'clock orientation, etc.

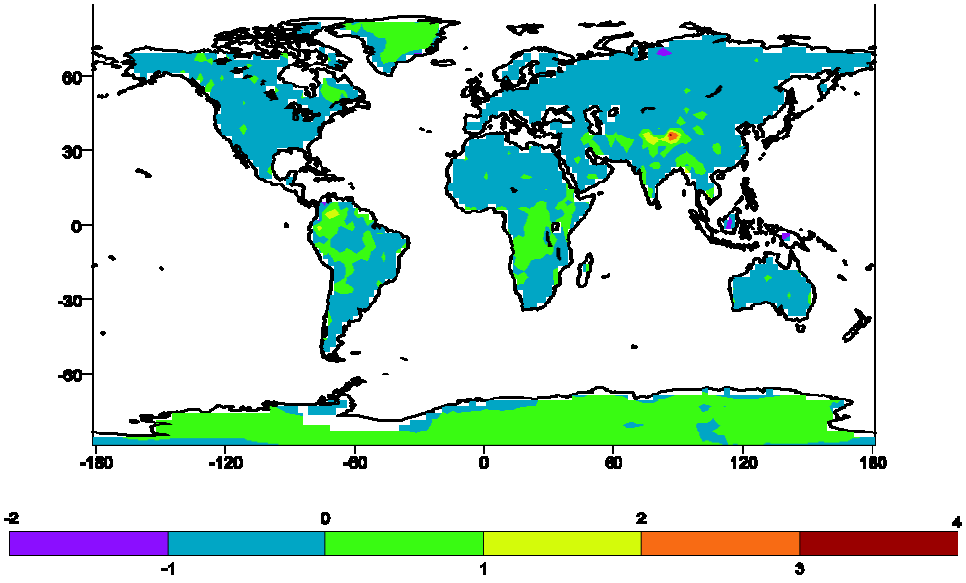
Figure 5: Regional-scale climatological annual cycle of IPCC model simulations (dashed lines) and of the multi-model ensemble mean (solid black line) compared with that of the GPCP, CMAP, and CRU observational estimates (colored solid lines) in a) Southwest U.S., b) India, c) African Sahel, and d) Brazilian Nordeste regions, where the respective geographical boundaries are shown at the top of each panel. All units are in mm day^{-1}.

Figure 1: Precipitation Datasets

a) Annual-mean GPCP precipitation (mm day^{-1})



b) Annual-mean CMAP – GPCP precipitation differences (mm day^{-1})



c) Annual-mean CRU – GPCP precipitation differences (mm day^{-1})

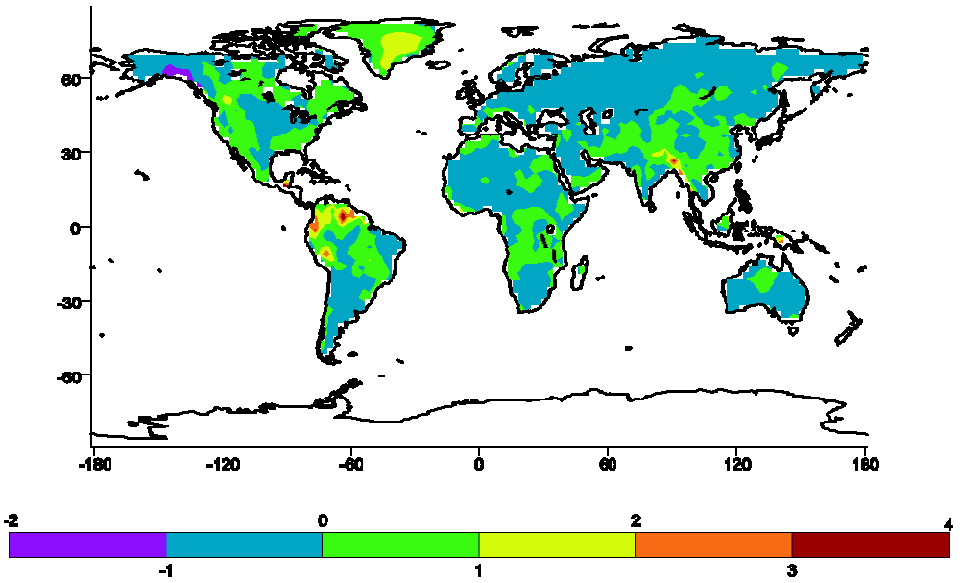
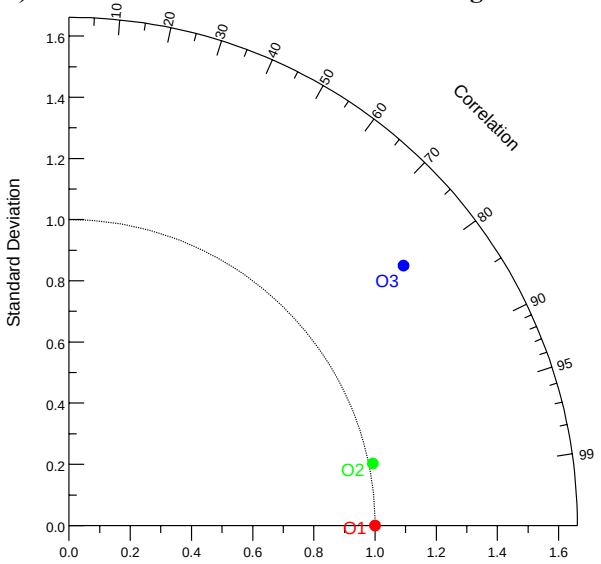
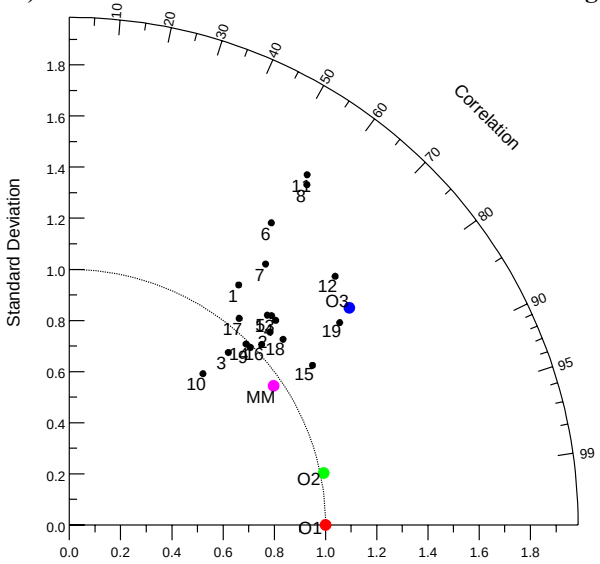


Figure 2: Taylor Diagrams

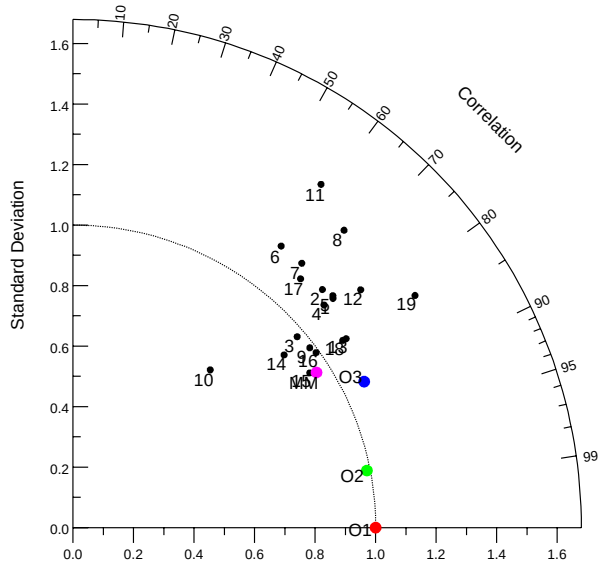
a) Annual-mean observational climatologies



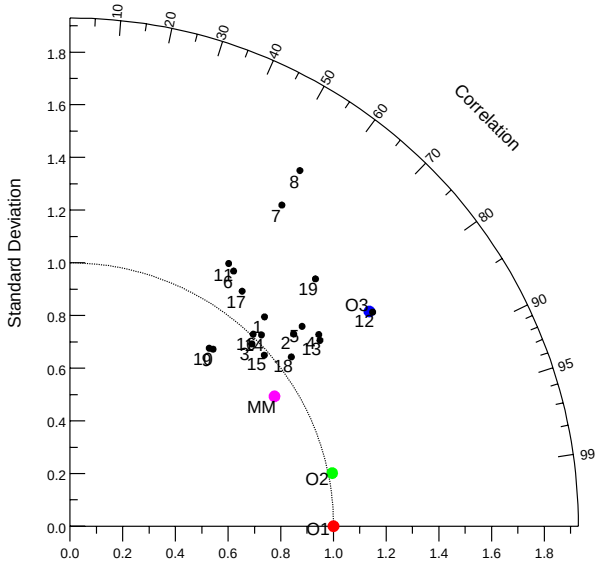
b) Annual-mean model vs. observational climatologies



c) January-mean model vs. observational climatologies



d) July-mean model vs. observational climatologies



Legend

O1: GPCP dataset
O2: CMAP dataset
O3: CRU dataset
1: Model CGCM3.1
2: Model CNRM-CM3
3: Model CSIRO-Mk3.0
4: Model GFDL-CM2.0
5: Model GFDL-CM2.1
6: Model GISS-AOM
7: Model GISS-EH
8: Model GISS-ER
9: Model FGOALS-g1.0
10: Model INM-CM3.0
11: Model IPSL-CM4
12: Model MIROC3.2 (hires)
13: Model MIROC3.2 (medres)
14: Model ECHAM/MPI-OM
15: Model MRI-CGCM2.3.2
16: Model NCAR-CCSM3
17: Model NCAR-PCM
18: Model UKMO-HadCM3
19: Model UKMO-HadGEM1
MM: Multi-Model Mean

Figure 3: MM – GPCP Precipitation Differences

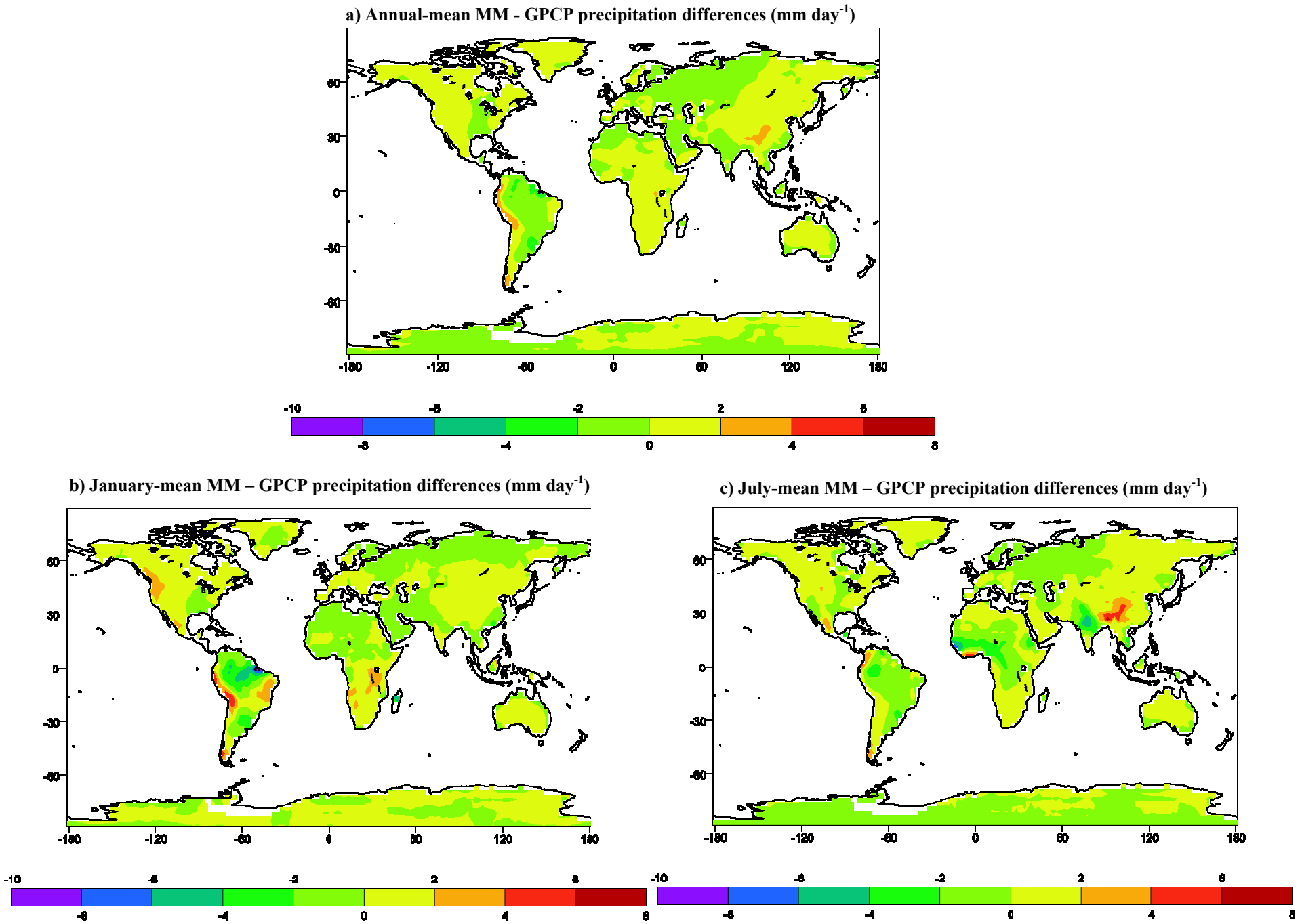


Figure 4 (Rotated): Harmonic Analysis of Precipitation in the North American Sector

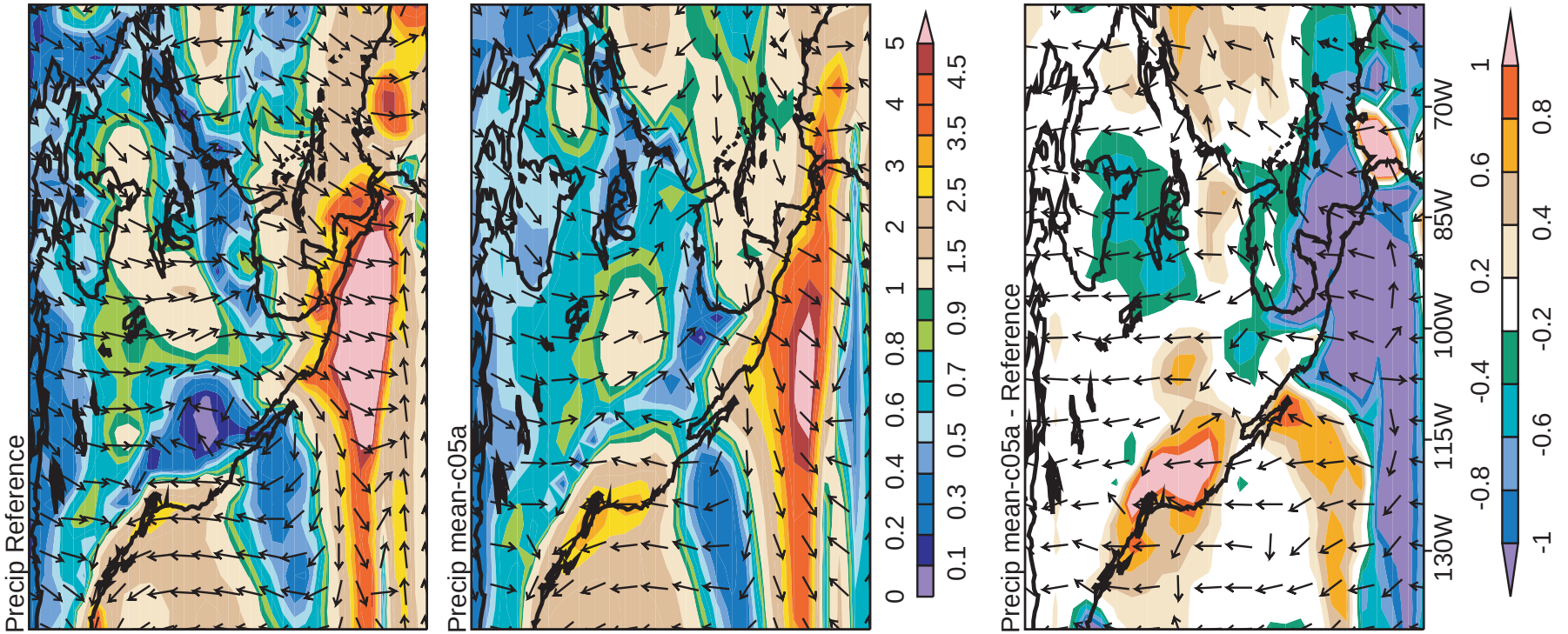


Figure 5: Regional Precipitation Comparisons