LLNL-JRNL-635873

LAWRENCE LIVERMORE NATIONAL LABORATORY

Evaluation of CMIP5 Continental Precipitation Simulations Relative to Satellite-Based Gauge-Adjusted Observations

A. Mehran, A. AghaKouchak, T. J. Phillips

April 30, 2013

Journal of Geophysical Research Atmospheres

Disclaimer

LLNL-JRNL-635873

Evaluation of CMIP5 Continental Precipitation Simulations Relative to Satellite-Based Gauge-Adjusted Observations

A. Mehran[1], A. AghaKouchak[1], and T.J. Phillips[2]

[1]University of California, Irvine, CA

[2]Lawrence Livermore National Laboratory, Livermore, CA

Ali Mehran, University of California Irvine, E4130 Engineering Gateway Irvine, CA 92697-2175, USA (a.mehran@uci.edu).

Amir AghaKouchak, University of California Irvine, E4130 Engineering Gateway Irvine, CA 92697-2175, USA (amir.a@uci.edu).

Thomas J. Phillips, Lawrence Livermore National Laboratory, 7000 East Avenue Livermore, CA 94550, USA (phillips14@llnl.gov).

Abstract

Numerous studies have emphasized that climate simulations are subject to various biases and uncertainties. The objective of this study is to cross-validate 34 Coupled Model Intercomparison Project Phase 5 (CMIP5) historical simulations of precipitation against the Global Precipitation Climatology Project (GPCP) data, quantifying model pattern discrepancies and biases for both entire data distributions and their upper tails. The results of the Volumetric Hit Index (*VHI*) analysis of the total monthly precipitation amounts show that most CMIP5 simulations are in good agreement with GPCP patterns in many areas, but that their replication of observed precipitation over arid regions and certain sub-continental regions (e.g., northern Eurasia, eastern Russia, central Australia) is problematical. Overall, the *VHI* of the multi-model ensemble mean and median also are superior to that of the individual CMIP5 models. However, at high quantiles of reference data (e.g., the 75th and 90th percentiles), all climate models display low skill in simulating precipitation, except over North America, the Amazon, and central Africa. Analyses of total bias (*B*) in CMIP5 simulations reveal that most models overestimate precipitation over regions of complex topography (e.g. western North and South America and southern Africa and Asia), while underestimating it over arid regions. Also, while most climate model simulations show low biases over Europe, inter-model variations in bias over Australia and Amazonia are considerable. The Quantile Bias (*QB*) analyses indicate that CMIP5 simulations are even more biased at high quantiles of precipitation. It is found that a simple mean-field bias removal improves the overall *B* and *VHI* values, but does not make a significant improvement in these model performance metrics at high quantiles of precipitation.

1. Introduction

Numerous studies have emphasized that water resources are sensitive to climate change, and thus water resources management and planning strategies should be adjusted accordingly [e.g., *Seager et al*., 2007; *Stoll et al*., 2011; *Sivakumar,* 2011; *Cayan et al*., 2008; *Wood et al*., 1997]. One of the key climate variables is precipitation, which plays a dominant role in the hydrologic cycle. Developing future water resources management and planning strategies thus requires estimation of current and future precipitation magnitude and variability [*Wehner et al.*, 2013].

In the past several decades, global climate models have been used to estimate future projections of precipitation [*IPCC,* 2007]. However, these projections are inherently uncertain and often are difficult for decision makers to interpret [e.g., *Liepert and Previdi,* 2012; *Reichler and Kim,* 2008; *Brekke and Barsugli,* 2012; *Schubert and Lim*, 2013; *Feddema et al*., 2005; *Min et al*., 2007]. Quantification of biases and uncertainties in climate simulations of precipitation thus are fundamental to understanding the reliability of climate simulations for future water resources management. *Gleckler et al.*, [2008] introduced several metrics for performance analysis of climate models and emphasized the need to go beyond the mean statistics for comprehensive analysis of climate model performance. *Moise and Delage* [2011] and *Schaller et al.* [2011] subsequently presented alternative approaches and metrics for evaluating seasonal precipitation simulations. *AghaKouchak and Mehran* [2013] introduced a number of volumetric indicators for validation and verification of climate model simulations.

Since the inception of the of the Coupled Model Intercomparison Project Phase 3 (CMIP3) by the World Climate Research Programme (WCRP) Working Group on Coupled Modelling (WGCM), evaluation of coupled ocean-atmosphere simulations of historical climate relative to available observational data has become an especially strong scientific focus [e.g., *Bony et al.*, 2006]. Indeed, future developments and improvements in global climate models (GCMs) rely heavily on their rigorous and informative validation.

Generally, climate model simulations are known for poor representation of frontal, convective and mesoscale processes [*Van Weverberg et al.*, 2013]. Numerous studies have evaluated various aspects of precipitation in the CMIP3 model simulations. *Phillips and Gleckler* [2006], for example, evaluated CMIP3 simulations of seasonal-mean continental precipitation amounts, and concluded that many of these differed markedly from several observational estimates. They also noted that the ensemble-mean model precipitation was generally closer to the observations than that of any individual CMIP3 model. Several studies focused on the common errors and/or frequency and intensity of CMIP3 daily precipitation simulations [e.g. *Dai,* 2006; *Sun et al*., 2007; *Stephenson et al.*, 2010; *Brown et al*., 2010]. *Dai*, 2006 evaluated the mean spatial patterns, precipitation intensity, frequency, and diurnal cycle. The results showed that many climate models simulated unrealistic double Intertropical Convergence Zone precipitation patterns, though most models captured the overall precipitation pattern. Furthermore, *Dai*, [2006] showed that the CMIP3 simulations produce light rain (1 to 10 mm) too frequently (see *Brown et al*., [2010]; *Stephens et al*., [2010]; *Wilcox and Donner*, [2007]; *Sun et al*., [2007]).

Brown et al., [2010] demonstrated that models capture the synoptic regimes well, and concluded that uncertainties in precipitation simulations are due to problems in simulating the characteristics of precipitation within different synoptic regimes. In a recent study, *Catto et al.*, 2013, argued that climate models often underestimate frontal precipitation estimates. Biases have been reported in model precipitation simulations from warm clouds, associated with unrepresentative microphysical parameterizations [*Lebsock et al.*, 2013]. *Wehner et al*., [2010] showed that many climate models underestimated 20-year return values of precipitation, and suggested increasing the horizontal resolution could improve estimation of extremes. *Ghan et al*., 2002 showed that improving representation of subgrid variability and surface topography has a significant positive impact on model precipitation simulations (see also *Qian et al*., [2010]).

Following upon CMIP3, the current Coupled Model Intercomparison Project Phase 5 (CMIP5) includes an unprecedented suite of coordinated simulations of historical and future-climate scenarios [*Taylor et al.*, 2012] that are designed to facilitate consideration of the wide range of scientific issues to be addressed in the forthcoming Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report. The CMIP5 climate simulations are archived by institutional participants in the Global Organization for Earth System Science Portals that is coordinated by the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison (PCMDI). A comprehensive description of the detailed numerical, dynamical, and physical properties of the CMIP5 models is now in progress [*Guilyardi et al.*, 2013].

In a recent study, *Liu et al.* [2012] evaluated the variability of CMIP5 precipitation simulations and their response to temperature using satellite data and showed that there is generally good agreement (correlation) between model simulations and satellite-inferred observed precipitation anomalies over land, both in the Tropics and globally. In addition, *Sillman et al*. [2013] evaluated models’ performance in simulating precipitation extremes at 1-5 day time scales. *Hirota et al*. [2013] investigated reproducibility of observed precipitation distribution in CMIP5 relative to CMIP3 simulations over the tropical oceans. They showed ensemble mean of CMIP5 simulations exhibited slightly higher skill score compared to CMIP3 ensemble mean. *Gaetani and Mohino*, [2013] studied the decadal predictability of CMIP5 simulations of the Sahel precipitation, and concluded that predictive skills of CMIP5 precipitation simulations varies significantly from model-to-model. Kumar et al., 2013 showed that the CMIP5 ensemble mean precipitation matched very well with that of ground-based observations, while there were substantial biases in the simulation of regional precipitation trends. A number of other studies assessed future changes in precipitation based on CMIP5 simulations at regional or global scales (e.g., *Chadwick et al*., [2013]; *Joetzjer et al*., [2013]).

The present study evaluates CMIP5 historical simulations of continental precipitation against the Global Precipitation Climatology Project (GPCP) monthly mean observational estimates [*Adler et al.*, 2003] using several quantitative statistical measures. This model evaluation focuses on the years 1979-2005, a period for which long-term and gauge-adjusted satellite observations are available. The remainder of this paper is organized into three sections. The observational data sets are briefly introduced in Section 2, while Section 3 is devoted to methodology and results. Summary remarks and conclusions are included in Section 4.

2. Data Sets

The GPCP reference data set [*Adler et al*., 2003] is derived from merged satellite-precipitation data that are bias-corrected using thousands of continental rain-gauge observations. The GPCP land algorithm is primarily based on microwave-scattering estimates from low-Earth-orbit satellites, and Geosynchronous Infrared (IR) estimates. Additional precipitation information is integrated based on data from the Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS), and outgoing longwave radiation (OLR) measurements [*Adler et al*., 2003]. The data set is available as a monthly time series from 1979 onward. GPCP data sets have been validated and widely used in numerous studies [e.g., *Bolvin et al*., 2009; *Huffman et al*., 2009]). The ground-based data are assembled by the Global Precipitation Climatology Centre (GPCC; *Rudolf et al.*, [1994]) of the Deutscher Wetterdienst and by the National

Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) and used for bias correction of the final product (see *Easterling 2013* for the distribution of gauges).

In this study, 34 CMIP5 historical simulations of monthly mean precipitation for the period 1979-2005, as well as their ensemble mean and median, are validated against the GPCP observations. Because of lack of reference gauge data across oceans, GPCP data over oceans are not bias-adjusted. One cannot evaluate biases in CMIP5 simulations with a reference data set having unknown bias. For this reason, this study is limited to evaluation of CMIP5 simulations overland where GPCP data are bias-adjusted using thousands of ground-based gauge data [*Adler et al.*, 2003].

All CMIP5 precipitation simulations and GPCP data are re-gridded onto a common 2 × 2-degree grid. Table 1 summarizes the CMIP5 models considered in this study. In addition, the results for both the multi-model ensemble mean and median are provided, since the latter is less sensitive to statistical outliers than the former.

It should be noted that the model simulations that are designated as "_esm" are historical simulations of climate with atmospheric CO_2 *emissions* specified in coupled earth systems models (ESMs) that include a prognostic carbon cycle, but with model-specific dynamical vegetation schemes "turned off". All the other climate simulations are performed with coupled ocean-atmosphere models in which the historical time series of global atmospheric CO_2 *concentrations* are prescribed. Although the CO_2 concentration time series in the "_esm" runs are not identical to these prescribed values (owing to model-specific differences in converting CO_2 emissions to concentrations), they are effectively constrained to be very similar.

3. Methodology and Results

Several statistical measures are employed to assess climate model-based historical precipitation simulations. Figure 1 displays the overall biases of climate model simulations relative to observations. The bias *B* is defined as the sum of monthly precipitation amount P for each CMIP5 model divided by the sum of the corresponding GPCP observations in each 2x2-degree grid box:

$$B = \frac{\sum_{i=1}^{n} P_{CMIP5}}{\sum_{i=1}^{n} P_{GPCP}} \quad (1)$$

Here *n* is the number of exceedances of a specified local monthly (daily) precipitation threshold *t* (which in this case is set equal to 1 mm/day), while, for simplification, the index *i* signifying each month in the 1979-2005 time series is not included (i.e., $P_{CMIP5} = P_{CMIP5_i}$ and $P_{GPCP} = P_{GPCP_i}$). A bias value above (below) 1 thus indicates an aggregate model overestimation (underestimation) of the monthly GPCP precipitation amounts for a particular grid box.

In the Figure 1 mappings of *B*, the color green indicates where there is little simulation bias, while red (blue) indicates large positive (negative) bias relative to GPCP data (white areas indicate no data in either observations or model simulations). This pertains to most model simulations of precipitation over the eastern United States and northern Europe and Asia; but many show a large positive bias (as high as ~ 2) in regions of complex topography such as western North and South America, and southern Africa and Asia, as was also noted for CMIP3 models by *Phillips and Gleckler* [2006]. In contrast, most models underestimate precipitation over the Saharan and central Asian deserts. Australia and Amazonia are other locations where there are substantial variations in both the sign and magnitude of the biases across individual simulations. On the other hand, the GPCP data may be subject to biases due to the limited availability of ground-based gauge data for bias-correction, as well as the limitations of satellite data in estimating orographic precipitation [*Sorooshian et al.*, 2011; *Mehran and AghaKouchak*, 2013].

In Figure 1, it is also noteworthy that the spatial patterns of the biases of the ensemble mean and median simulations are similar (bottom right panels), and over several regions (e.g., Australia and the Americas) are of lower magnitude than the biases of most of the individual model simulations. From comparing global-average biases *B* summarized in Table 1, however, the overall global bias of a number of the CMIP5 simulations is seen to be closer to the optimum value of 1 than is that of the ensemble mean (B = 0.89) and median (B=0.85).

It is well known that the bias is not necessarily constant throughout a distribution function, but may change at different quantile levels. To further investigate precipitation biases in the CMIP5 simulations, the quantile bias (QB_t), defined as the ratio of monthly precipitation amounts in each simulation to that of the GPCP observations above a specified threshold t (e.g., the 75th percentile of all the local monthly values), can be calculated in each 2x2-degree grid box:

$$QB_t = \frac{\sum_{i=1}^{n}(P_{CMIP5}|P_{CMIP5} \geq t)}{\sum_{i=1}^{n}(P_{GPCP}|P_{GPCP} \geq t)} \qquad (2)$$

Here, QB_t = 1 indicates no bias in the simulations, whereas a value above (below) 1 corresponds to a climate model's overestimation (underestimation) of precipitation amount above the specified threshold t, with respect to that of the GPCP observations. Figure 2 displays QB_t values, computed for the 75th percentile (precipitation values above t = 75% of the reference data), in the CMIP5 simulations. This figure indicates that the climate model biases apparent in Figure 1 are generally accentuated in the upper tail (i.e., > 75% quantile) of the GPCP precipitation distribution. While individual CMIP5 models behave somewhat differently from one another, most of their simulations underestimate heavier precipitation amounts over large areas (e.g. Eurasia, Middle East, northern China), while overestimating them only in certain limited regions (e.g., Amazonia, central Africa, United States). Given this general behavior, it is not surprising that the multi-model ensemble mean and median also show large negative biases for observed precipitation amounts > 75% quantile of the distribution. Such negative biases are even more pronounced for P amounts > 90% of the distribution (figure not shown for brevity).

Bias and Quantile Bias describe the overall ratio of simulations over observations, and do not provide information on the grid-scale matching of simulated precipitation relative to missed precipitation based on reference observations. The Volumetric Hit Index (*VHI*; *AghaKouchak et al*., 2011), which measures the volume of precipitation above the threshold (t) detected correctly by climate models with respect to the total simulated and missed precipitation (based on GPCP), can provide such a measure of model performance [*Mehran and AghaKouchak*, 2013]. For t = 0 and t > 0, the *VHI* in each 2x2-degree grid box is defined as (*AghaKouchak et al*., 2011):

$$\text{for t=0} \quad VHI = \frac{\sum_{i=1}^{n}(P_{CMIP5}|P_{CMIP5}>0 \ \& \ P_{GPCP}>0)}{\sum_{i=1}^{n}(P_{CMIP5}|P_{CMIP5}>0 \ \& P_{GPCP}>0)+\sum_{i=1}^{n}(P_{GPCP}|P_{GPCP}>0 \ \& \ P_{CMIP5}=0)}$$

(3)

$$\text{for t >0} \quad VHI = \frac{\sum_{i=1}^{n}(P_{CMIP5}|P_{CMIP5}\geq t \ \& \ P_{GPCP}\geq t)}{\sum_{i=1}^{n}(P_{CMIP5}|P_{CMIP5}\geq t \ \& \ P_{GPCP}\geq t)+\sum_{i=1}^{n}(P_{GPCP}|P_{GPCP}\geq t \ \& \ P_{CMIP5}<t)}$$

where P_{CMIP5} = CMIP5 simulations, P_{GPCP} = GPCP observations, n = number of exceedances above threshold t. The ideal *VHI* score is 1, indicating perfect simulation skill, while 0 corresponds to no skill. In this study, *VHI* is computed for the entire distribution of precipitation (t = 0) and for values above the 75th percentile of the observations ($t \geq 75$ percentile of GPCP). Figure 3 presents the *VHI* when all precipitation data are included in the analysis (t = 0) – white areas indicate no data below the choice of threshold t in either climate model simulations or observations. There is generally good agreement between model simulations and GPCP observations over many areas, especially in moist tropical regions such as Amazonia and southern Africa, and in temperate latitudes of Eurasia and North American, consistent with the findings of *Liu et al*. [2012]. However, there are obvious discrepancies over arid regions, especially northern Africa and the Middle East, but also the southwestern U.S. and Australia. From Figure 3, the ensemble mean and median appear to be superior to the individual climate models in reproducing the main GPCP precipitation patterns (see also Table 1 for global-average *VHI* values of the CMIP5 models and their ensemble mean and median).

It is acknowledged that the *VHI* is a skill score which assumes that a “perfect” model will be able to recreate the observations at each grid. However, the CMIP5 historical runs have large internal variability in precipitation, and one cannot expect the models to precisely reproduce precipitation observations, since they are not forced with prescribed, historical sea surface temperatures. Nevertheless, the *VHI* still provides valuable information as to what extent model simulations can capture historical satellite-based gauge-adjusted observations. Here, the main purpose of using *VHI* is to show whether climate models, relative to each other, are consistent with observations. As shown, over many regions, many climate model simulations exhibit high *VHI* scores indicating reasonable consistency with observations.

Figure 4 displays *VHI* for the 75th percentile threshold of the observations (hereafter, VHI_{75}). When considering only data above this threshold, the performance of most models is seen to decrease substantially, indicating the presence of systematic biases in the CMIP5 simulations at higher quantiles. The VHI_{75} maps show that except over parts of high-latitude Eurasia, temperate North America, the lower Amazon, southeast Asia and central Africa, the model simulations lack skill above the 75% quantile. This is even more so for VHI at the 90^{th} percentile threshold (figure not shown). It should be emphasized, however, that a low value of *VHI* (or VHI_{75}) does not necessarily imply the absence of simulated precipitation, but only that the models are simulating amounts below the local threshold (here, the 75th percentile) of the GPCP reference observations. The results of both the *VHI* and QB_t analyses confirm that there are biases in climate simulations of precipitation at higher quantiles, implying that more effort should be focused on improving precipitation physics in climate models, so as to more realistically simulate extreme values.

As an alternative to model physics improvements, bias adjustment algorithms have been developed to bring climate simulations closer to observational reference data [e.g., *Li et al.*, 2010; *Christensen et al.*, 2008; *Haerter et al.*, 2011; *Dosio and Paruolo*, 2011; *Xu et al.*, 2012]. Bias-adjustment of global model simulations is necessary in order to supply more realistic estimates of precipitation (or other climate variables) to regional-scale models that can assess the impacts of climate on hydrology or agriculture, for example. The most common approach involves the removal of the mean-field bias of climate simulations relative to a given observed data set. Figure 5 plots global average global bias values of all 35 evaluated CMIP5 models for the thresholds of t=0 (all data), t=75% (QB_{75}) and t=90% (QB_{90}) percentiles of observations, both before and after removal of the mean-field bias. The models are sorted based on their overall bias values for better visualization. Here the mean-field bias is removed by multiplying the inverse of Equation (1) by the original CMIP5 simulations. Considering all the data, the models exhibit global-average biases B between 0.75 to 1.25 (solid black line in Figure 5), and after mean-field adjustment, this overall bias can be removed (dashed black line in Figure 5). However, while the mean-field bias adjustment eliminates the overall bias, the figure indicates that this adjustment does not necessarily reduce the bias associated with a particular high or low quantile. Figure 5 instead confirms that a simple mean-field bias adjustment only marginally reduces such quantile biases.

In order to compare the spatial patterns of bias after mean-field bias adjustment, Figure 6 displays QB_{75} after removing the mean-field bias of model simulations with respect to GPCP data (the bias-adjusted version of Figure 2). One can see that, while some improvements in QB can be achieved through such adjustment, this is not the case over several areas such as portions of

Australia, Africa, Eurasia, and North America. Thus, on average, CMIP5 models underestimate high quantiles of precipitation even after mean-field bias adjustment. This result underscores the importance of developing more sophisticated precipitation bias-adjustment techniques, such as *Watanabe et al.* [2012] and *Mehrotra and Sharma* [2012], that go beyond consideration of only the mean statistics.

Figure 7 displays the globally averaged *VHI*, VHI_{75} and VHI_{90} of climate model simulations against GPCP data before and after removal of the mean-field bias. The models are ranked on the x-axis based on their overall *VHI* for better visualization. Similar to the case of *QB*, it is seen that *VHI* values drop substantially as the threshold increases, but that the mean-field bias adjustment significantly improves the *VHI* of the ensemble median. Unlike the bias metrics of Figure 5, the *VHI* values of the ensemble median in Figure 7 are consistently higher than those of the ensemble mean.

Figure 8 displays the spatial patterns of VHI_{75} of climate model simulations against GPCP data after mean-field bias removal (i.e. Figure 8 depicts results similar to those shown in Figure 4, but after bias adjustment). Compared with Figure 4, the *VHI* values in Figure 8 improve over certain areas, such as the United States, Amazonia, and Southeast Asia. In these regions, models such as FGOALS_g2_hist, GFDL_ESM2M_hist, CESM1-BGC_esm and MIROC show high (~ 1) values of VHI_{75}. Overall, the VHI_{75} values of the ensemble mean and median are superior to those of individual models.

4. Summary and Conclusions

Reliable estimates of precipitation are essential for both research and practical applications. CMIP5 climate simulations provide both historical simulations and future projections of climate variables. Numerous studies have highlighted that climate simulations are subject to various biases and uncertainties (e.g., *Brekke and Barsugli* [2012]). The objective of this study is to cross-validate CMIP5 historical simulations of precipitation relative to GPCP reference data, quantifying model pattern discrepancies (*VHI* metric) and biases (*B* and *QB* metrics) for both entire data distributions and their upper tails. It is acknowledged that observational data sets, like model simulations, are also subject to uncertainties, including systematic and random sampling errors [*AghaKouchak et al*., 2012]; over land, however, the GPCP data set is bias-adjusted using thousands of rain gauges [*Adler et al*., 2003], and hence it should serve as an a suitable reference for evaluation of continental precipitation in climate models.

From the results of the Volumetric Hit Index (*VHI*) pattern analysis of the total monthly precipitation amounts, it is found that most CMIP5 simulations are in fairly good agreement with GPCP observations in many areas, but model replication of observed precipitation patterns over deserts and certain sub-continental regions (e.g. northern Eurasia, central Australia) is problematical. The *VHI* of the multi-model ensemble mean and median also are found to be superior to most CMIP5 model simulations overall.

Analyses of total biases (*B*) in CMIP5 simulations reveal that most models overestimate precipitation in regions of steep topography, while underestimating it leeward of the mountains, as well as over many other arid regions. Moreover, while most climate model simulations show low *B* values over Europe, there are considerable inter-model variations in bias over Australia and Amazonia.

At high quantiles (> 75% and > 90%) of the distribution of monthly precipitation, the Quantile Bias (*QB*) analyses indicate that CMIP5 simulations show more glaring discrepancies in precipitation amounts with respect to the GPCP satellite observations. While continuing to overestimate precipitation in regions of steep topography, the models generally underestimate it in tropical locations such as Amazonia, central Africa, and southern Asia, as well as in broad swaths of the extra-tropics such as Australia, the arid regions of northern Africa and central Asia, and northern China, Russia, and Canada.

At high precipitation quantiles also, the CMIP5 models show substantially reduced agreement with the patterns of the GPCP reference data (e.g., the VHI_{75} metric for precipitation above the 75 % quantile << *VHI* over the entire distribution). Except over North America, Amazonia, and central Africa, most CMIP5 simulations are lacking in predictive skill ($VHI_{75} \sim 0$) for the higher tail of the precipitation distribution. Note, however, that a low VHI_{75} does not necessarily imply the absence of locally simulated precipitation, only that its amount falls below the given reference data's threshold value. In addition, the ensemble-mean and median precipitation at the higher quantiles are found to be superior to the individual climate model simulations when evaluated by the VHI_{75}, and VHI_{90} pattern measures, but not by the the QB75 and QB90 bias metrics.

These results thus suggest that, while today's climate model simulations are generally in agreement with satellite-based gauge-adjusted estimates of total monthly precipitation in many areas, they presently are not well-suited for simulating upper quantiles of the precipitation distribution. Such distribution errors, which have persisted across CMIP phases [e.g. *Dai* ,2006; *Sun et al*., 2007; *Stephenson et al*. , 2010; *Sillmann et al*., 2013], are often characterized by a general tendency for the models to precipitate too frequently in light amounts, but too rarely in the intense downbursts that are occasionally observed. In focusing on the upper tails of the precipitation distributions, the present study reveals such model intensity errors in a particularly stark way (e.g. in Figures 2 and 4).

The persistence of these upper-tail errors in all the evaluated CMIP5 simulations is indicative of the presence of *general* deficiencies in the models. For instance, these systematic precipitation distribution errors do not seem to be very sensitive to inter-model differences in horizontal resolution (e.g. the MRI-ESM1_esm model, with a 160x320 grid, does not clearly outperform other coarser-resolution models in Table 1). Substantive differences in error structure also are not apparent between the "_esm" historical simulations with prescribed CO_2 *emissions* and those with prescribed CO_2 *concentrations.* Thus, it is likely that these systematic precipitation errors are due more to general model shortcomings in representing the dynamics or physics of climatic phenomena than to inter-model differences in greenhouse forcings or horizontal resolution.

For example, an ongoing preoccupation of model developers is to improve sub-grid scale parameterizations of convection, since precipitation errors tend to be especially large in the tropics [e.g. *IPCC*, 2007]. It is perhaps less widely appreciated that intense precipitation also often originates in frontal systems [*Catto et al.*, 2012; *Pfahl and Wernli*, 2012] and that representative climate models tend to underestimate these extra-tropical precipitation events in spite of current abilities to adequately simulate the interaction of dynamics and moisture at model grid scale [*Catto et al.* , 2010, 2013].

An underestimation or incorrect placement of intense tropical and extra-tropical precipitation is also clearly displayed by the CMIP5 simulations analyzed in the present study (e.g. in Figures 2 and 4). Such ubiquitous precipitation errors suggest that improvements, not only in model convective parameterizations, but also in their representation of sub-grid scale cloud microphysical processes that regulate droplet auto-conversion, accretion, and through-fall [*Lebsock et al.*, 2013; *Van Weverberg et al.*, 2013], may be essential for better simulation of the observed global precipitation distributions.

Once such enhancements of model physics are in place, increases in model resolution are also likely to contribute to more realistic simulation of precipitation [*Wehner et al*., 2010; *Champion et al*, 2011]. This may be especially true in mountainous regions, where an accurate representation of the interaction of complex dynamics and steep moisture gradients is difficult to achieve solely through parameterization [e.g. Ghan et al., 2002; Qian et al., 2010].

Finally, this study demonstrates that, while a simple mean-field bias removal enhances the overall *B* and *VHI* values, it does not yield much improvement at high quantiles (i.e., QB_{75}, QB_{90}, VHI_{75}, and VHI_{90}). Thus, for purposes of climatic impacts studies, it is important to develop more sophisticated techniques for adjusting the upper-tail biases of global precipitation simulations in order to better replicate observed extreme values.

Acknowledgments. The financial support for authors AM and AA was made available from the United States Bureau of Reclamation (USBR) Award No. R11AP81451, and National Science Foundation Award No. OISE-1243543. The contributions of author TJP were performed under the auspices of the Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate-modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and leads the development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Also, the authors acknowledge the Global Precipitation Climatology Project (GPCP) team for providing and distributing The GPCP data sets.

References

Adler, R., G. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, and P. Arkin, 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979 present). Journal of Hydrometeorology, 4, 1147-1167.

AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite-retrieved extreme precipitation rates across the central United States. Journal of Geophysical Research-Atmospheres, 116(D2), D02115.

AghaKouchak, A., A. Mehran, H. Norouzi, and A. Behrangi, 2012: Systematic and random error components in satellite precipitation data sets. Geophysical Research Letters, 39(9), L09406.

AghaKouchak A., Mehran A., 2013, Extended Contingency Table: Performance Metrics for Satellite Observations and Climate Model Simulations, Water Resources Research, 49, 1-6, doi:10.1002/wrcr.20498.

Bolvin, D. T., R. F. Adler, G. J. Huffman, E. J. Nelkin, and J. P. Poutiainen, 2009: Comparison of GPCP monthly and daily precipitation estimates with high-latitude gauge observations. Journal of Applied Meteorology and Climatology, 48(9), 1843–1857.

Bony, S., R. Colman, V. M. Kattsov, R. P. Allan, C. S. Bretherton, J.-L. Dufresne, A. Hall, S. Hallegatte, M. M. Holland, W. Ingram, D. A. Randall, B. J. Soden, G. Tselioudis, and M. J. Webb, 2006: How well do we understand and evaluate climate change feedback processes? Journal of Climate, 19(15), 3445–3482.

Brekke, L. and J. Barsugli, 2012: Uncertainties in projections of future changes in extremes. In AghaKouchak, A., D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian, eds., Extremes in a Changing Climate: Detection, Analysis and Uncertainty, Springer, DOI: 10.1007/978-94-007-4479-0 11.

Brown, J. R., C. Jakob, and J. M. Haynes, 2010: An evaluation of rainfall frequency and intensity over the Australian region in a global climate model. J. Climate, 23, 6504–6525.

Catto, J.L., L.C. Shaffrey, and K.I. Hodges, 2010: Can climate models capture the structure of extratropical cyclones? Journal of Climate, 23, 1621-1635.

Catto, J.L., C. Jakob, G. Berry, and N. Nicholls, 2012: Relating global precipitation to atmospheric fronts. Geophysical Research Letters, 39, L10805, doi: 10.1029/2012GL051736.

Catto, J.L., C. Jakob, and N. Nicholls, 2013: A global evaluation of fronts and precipitation in the ACCESS model. Australian Meteorological and Oceanographic Society Journal, 63, 191-203.

Cayan, D. R., E. P. Maurer, M. D. Dettinger, M. Tyree, and K. Hayhoe, 2008: Climate change scenarios for the California region. Climatic Change, 87(1), S21–S42.

Chadwick, Robin, Ian Boutle, Gill Martin, 2013: Spatial Patterns of Precipitation Change in CMIP5: Why the Rich Do Not Get Richer in the Tropics. J. Climate, 26, 3803–3822.

Champion, A.J., K.I. Hodges, L.O. Bengtsson, N.S. Keenlyside, and M. Esch, 2011: Impact of increasing resolution and a warmer climate on extreme weather from Northern Hemisphere extratropical cyclones. Tellus, 63A, 893-906.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher, 2008: On the need for bias correction of regional climate change projections of temperature and precipitation. Geophysical Research Letters, 35(20), L20.

Dai, A, 2006, Precipitation characteristics in eighteen coupled climate models. Journal of Climate, 19 (18), 4605-4630.

Dosio, A. and P. Paruolo, 2011: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate. Journal of Geophysical Research-Atmospheres, 116(D16), D16106.

Easterling, D. R. (2013). Global Data Sets for Analysis of Climate Extremes. In Extremes in a Changing Climate (pp. 347-361). Springer Netherlands.

Feddema, J., K. Oleson, G. Bonan, L. Mearns, W. Washington, G. Meehl, and D. Nychka, 2005: A comparison of a GCM response to historical anthropogenic land cover change and model sensitivity to uncertainty in present-day land cover representations. Climate Dynamics, 25(6), 581–609.

Gaetani, Marco, Elsa Mohino, 2013: Decadal Prediction of the Sahelian Precipitation in CMIP5 Simulations. J. Climate, 26, 7708–7719.

Ghan, S.J., X. Bian, A.G. Hunt, and A. Coleman, 2002: The thermodynamic influence of subgrid orography in a global climate model. Climate Dynamics, 20, 31-44.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. Journal of Geophysical Research-Atmospheres, 113(D6), D06104.

Guilyardi, E., V. Balaji, B. Lawrence, S. Callaghan, C. Deluca, S. Denvil, M. Lautenschlager, M. Morgan, S. Murphy, and K.E. Taylor, 2013: Documenting climate models and their simulations. Bulletin of the American Meteorological Society, 94(5), 623-629.

Haerter, J. O., S. Hagemann, C. Moseley, and C. Piani, 2011: Climate model bias correction and the role of timescales. Hydrology and Earth System Sciences, 15(3), 1065–1079.

Hao Z., AghaKouchak A., Phillips T., 2013, Changes in Concurrent Monthly Precipitation and Temperature Extremes, Environmental Research Letters, 8(4), 034014.

Hirota, Nagio, and Yukari N. Takayabu. 2013, "Reproducibility of precipitation distribution over the tropical oceans in CMIP5 multi-climate models compared to CMIP3." Climate Dynamics: 1-12.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP Version 2.1. Geophysical Research Letters, 36(17), L17808.

IPCC, 2007: Climate Change 2007: *Impacts, Adaptation, and Vulnerability*. Exit EPA Disclaimer Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Parry, Martin L., Canziani, Osvaldo F., Palutikof, Jean P., van der Linden, Paul J., and Hanson, Clair E. (eds.)]. Cambridge University Press, Cambridge, United Kingdom.

Joetzjer, E., et al., 2013, "Present-day and future Amazonian precipitation in global climate models: CMIP5 versus CMIP3." Climate Dynamics: 1-16.

Kelley, Colin, et al., 2012, Mediterranean precipitation climatology, seasonal cycle, and trend as simulated by CMIP5. Geophysical Research Letters 39.21.

Kharin, V. V., et al., 2013, "Changes in temperature and precipitation extremes in the CMIP5 ensemble." Climatic Change: 1-13.

Kumar, Sanjiv, Venkatesh Merwade, James L. Kinter, Dev Niyogi, 2013: Evaluation of Temperature and Precipitation Trends and Long-Term Persistence in CMIP5 Twentieth-Century Climate Simulations. J. Climate, 26, 4168–4185.

Lebsock, M., H. Morrison, and A. Gettelman, 2013: Microphysical implications of cloud-precipitation covariance derived from satellite remote sensing. Journal of Geophysical Research, 18, 6521–6533.

Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. Journal of Geophysical Research-Atmospheres, 115(D10), D10101.

Liepert, B. G. and M. Previdi, 2012: Inter-model variability and biases of the global water cycle in CMIP3 coupled climate models. Environmental Research Letters, 7(1), 014006.

Liu, C., R. P. Allan, and G. J. Huffman, 2012: Co-variation of temperature and precipitation in CMIP5 models and satellite observations. Geophysical Research Letters, 39(13), L13803.

Mehran, A., A. AghaKouchak, 2013, Capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations. Hydrological Processes, doi: 10.1002/hyp.9779.

Mehrotra, R., & Sharma, A., 2012: An improved standardization procedure to remove systematic low frequency variability biases in GCM simulations. Water Resources Research, 48(12), W12601.

Min, S.-K., D. Simonis, and A. Hense, 2007: Probabilistic climate change predictions applying Bayesian model averaging. Philosophical Transactions of the Royal Society A – Mathematical Physical and Engineering Sciences, 365(1857), 2103–2116.

Moise, A. F. and F. P. Delage, 2011: New climate model metrics based on object-orientated pattern matching of rainfall. Journal of Geophysical Research-Atmospheres, 116(D12), D12108.

Pfahl, S., and H. Wernli, 2012: Quantifying the relevance of cyclones for precipitation extremes. Journal of Climate, 25, 6770-6780.

Phillips, T. J. and P. J. Gleckler, 2006: Evaluation of continental precipitation in 20th century climate simulations: The utility of multimodel statistics. Water Resources Research, 42(3), W03202.

Polson, Debbie, et al., 2013, "Causes of robust seasonal land precipitation changes." Journal of Climate, 26, 6679–6697.

Qian, Y., S.J. Ghan, and L. R. Leung, 2010: Downscaling hydroclimatic changes over the Western

US based on CAM subgrid scheme and WRF regional climate simulations. International Journal of Climatology, 30, 675-693.

Reichler, T. and J. Kim, 2008: Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model. Journal of Geophysical Research-Atmospheres, 113(D5), D05106.

Rudolf, B., H. Hauschild, W. Rüth, and U. Schneider, 1994: Terrestrial precipitation analysis: Operational method and required density of point measurements. Global Precipitation and Climate Change, M. Desbois and F. Desalmand, Eds., NATO ASI Series I, Vol. 26, Springer–Verlag, 173–186.

Schaller, N., I. Mahlstein, J. Cermak, and R. Knutti, 2011: Analyzing precipitation projections: A comparison of different approaches to climate model evaluation. Journal of Geophysical Research-Atmospheres, 116(D10), D10118.

Schubert, S. D., & Lim, Y. K. (2013). Climate Variability and Weather Extremes: Model-Simulated and Historical Data. In Extremes in a Changing Climate (pp. 239-285). Springer Netherlands.

Seager, R., M. Ting, I. Held, Y. Kushnir, J. Lu, G. Vecchi, H.-P. Huang, N. Harnik, A. Leetmaa, N.-C. Lau, C. Li, J. Velez, and N. Naik, 2007: Model projections of an imminent transition to a more arid climate in southwestern North America. Science, 316(5828), 1181–1184.

Sillmann, J., V.V. Kharin, Z. Zhang, F.W. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, Journal of Geophysical Research: Atmospheres, 118 (4), 1716–1733.

Sivakumar, B., 2011: Global climate change and its impacts on water resources planning and management: assessment and challenges. Stochastic Environmental Research and Risk Assessment, 25(4), 583–600.

Sorooshian, S., A. AghaKouchak, P. Arkin, J. Eylander, E. Foufoula-Georgiou, R. Harmon, J. M. H. Hendrickx, B. Imam, R. Kuligowski, B. Skahill, and G. Skofronick-Jackson, 2011: Advanced concepts on remote sensing of precipitation at multiple scales. Bulletin of the American Meteorological Society, 92(10), 1353–1357.

Stoll, S., H. J. H. Franssen, M. Butts, and W. Kinzelbach, 2011: Analysis of the impact of climate change on groundwater related hydrological fluxes: A multi-model approach including different downscaling methods. Hydrology and Earth System Sciences, 15(1), 21–38.

Stephens, G.L., T. L'Ecuyer, R. forbes, A. Gettlemen, J-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes, 2010: Dreary state of precipitation in global models, Journal of Geophysical Research: Atmospheres, 115: D24211.

Sun, Y., S. Solomon, A. Dai, R.W. Portmann , 2007, How often will it rain? Journal of Climate, 20, (19), 4801-4818.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. Bulletin of the American Meteorological Society, 93(4), 485–498.

Van Weverberg, K., A.M. Vogelmann, W. Lin, E.P. Luke, A. Cialella, P. Minnis, M. Khaiyer, E.R. Boer, and M.P. Jensen, 2013: The role of cloud microphysics parameterizations in the simulation of mesoscale convective system clouds and precipitation in the Tropical Western

Pacific. Journal of the Atmospheric Sciences, 70, 1104-1128.

Watanabe, S., Kanae, S., Seto, S., Yeh, P. J. F., Hirabayashi, Y., & Oki, T., 2012: Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models. Journal of Geophysical Research, 117(D23), D23114.

Wehner, M.F., R.L. Smith, G. Bala, and P. Duffy, 2010: The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. Climate Dynamics, 34, 241-247.

Wehner, M. (2013). Methods of Projecting Future Changes in Extremes. In Extremes in a Changing Climate (pp. 223-237). Springer Netherlands.

Wilcox, E. M. and L. J. Donner, 2007: The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model. J. Climate, 20, 53–69.

Wood, A., D. Lettenmaier, and R. Palmer, 1997: Assessing climate change implications for water resources planning. Climatic Change, 37(1), 203–228.

Xu, Z. and Z.-L. Yang, 2012: An improved dynamical downscaling method with GCM bias corrections and its validation with 30 Years of climate simulations. Journal of Climate, 25(18), 6271–6286.

List of Figures and Tables

Figure 6. Quantile Bias (t=75th percentile, QB_{75} with optimal value =1) maps after bias-adjustment of CMIP5 precipitation simulations, with respect to GPCP observations.

Figure 7. Averaged Volumetric Hit Index (VHI) and 75th and 90th percent quantile values VHI_{75} and VHI90 (with optimal values = 1) for selected CMIP5 precipitation simulations, and for their ensemble mean and median, both before bias adjustment (solid lines) and after (dashed lines), all with respect to GPCP observations.

Figure 8. Volumetric Hit Index for t = 75th percentile (VHI_{75}, with optimal value = 1) after bias-adjustment of selected CMIP5 precipitation simulations, all with respect to GPCP observations.

Table 1 – CMIP5 climate models and summary statistics of global Bias (B) and Volumetric Hit Index (VHI) before and after mean-field bias adjustment. Optimal values of these metrics are all equal to 1.

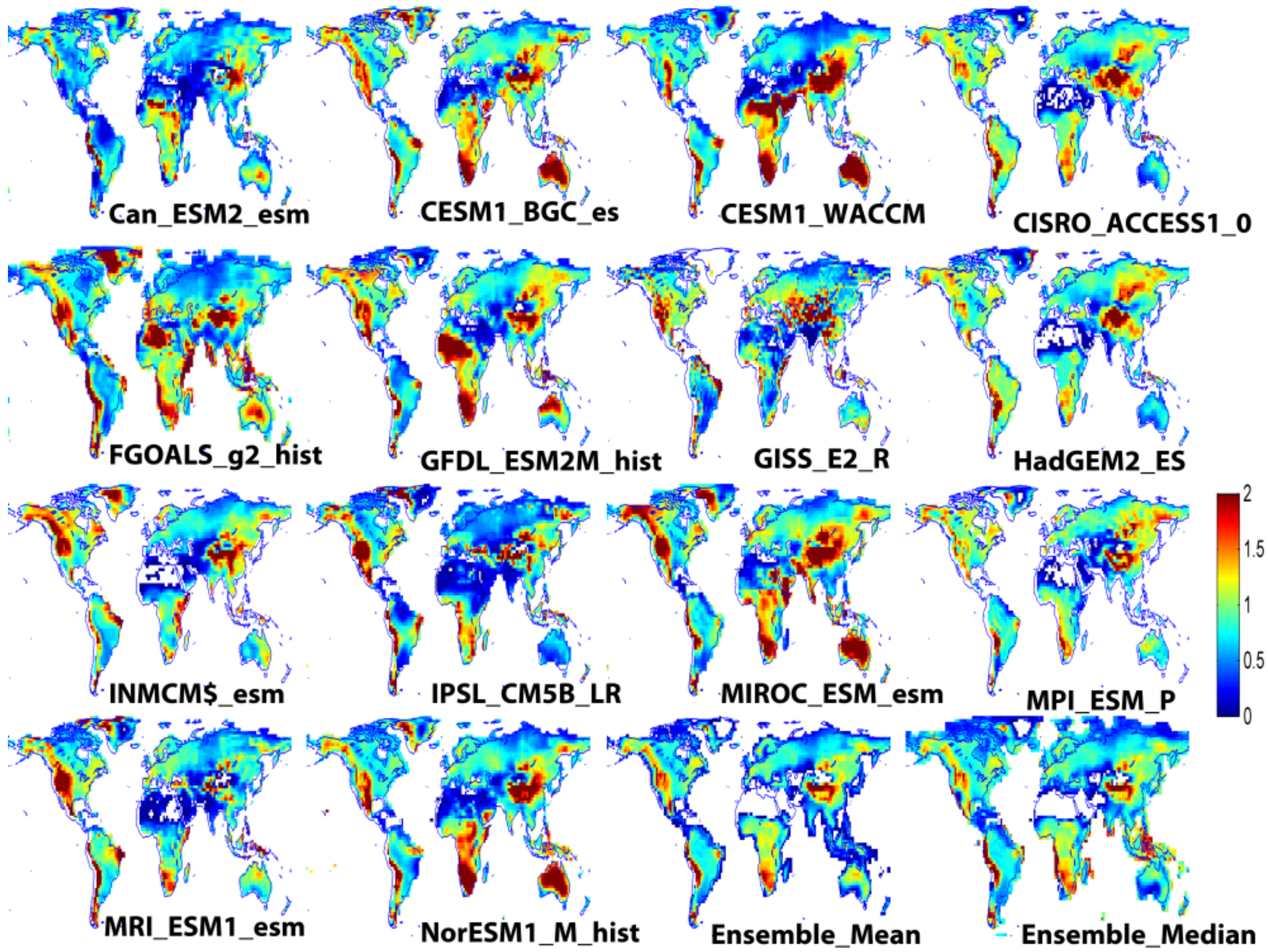

Figure 1. Bias (with optimal value = 1) of selected CMIP5 precipitation simulations, and of their ensemble mean and median, all with respect to GPCP observations. White-colored areas indicate “NaN”—undefined or no values.

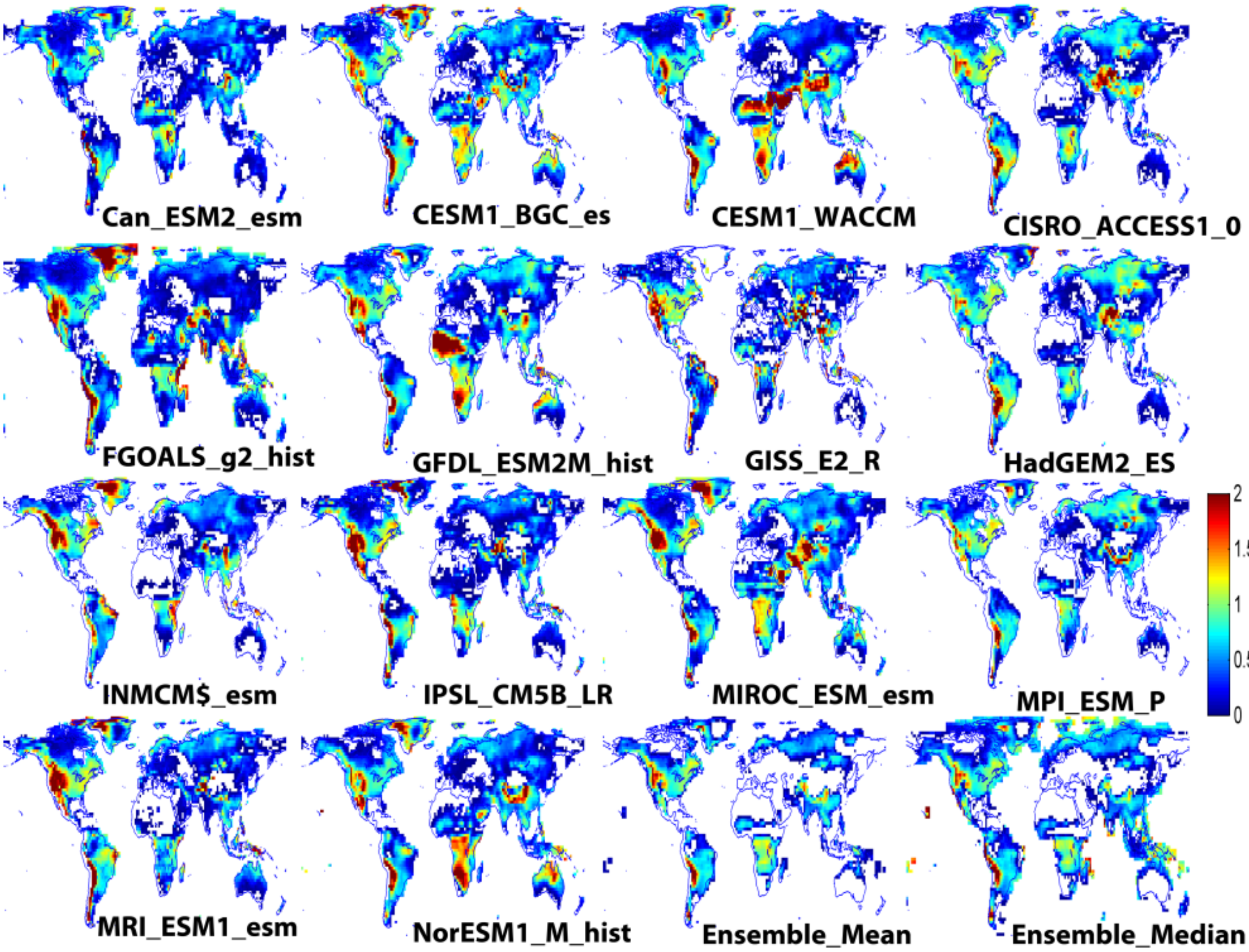

Figure 2. Quantile Bias (t = 75 percentile, QB_{75} with optimal value = 1) maps for selected CMIP5 precipitation simulations, and for their ensemble mean and median, all with respect to GPCP observations.

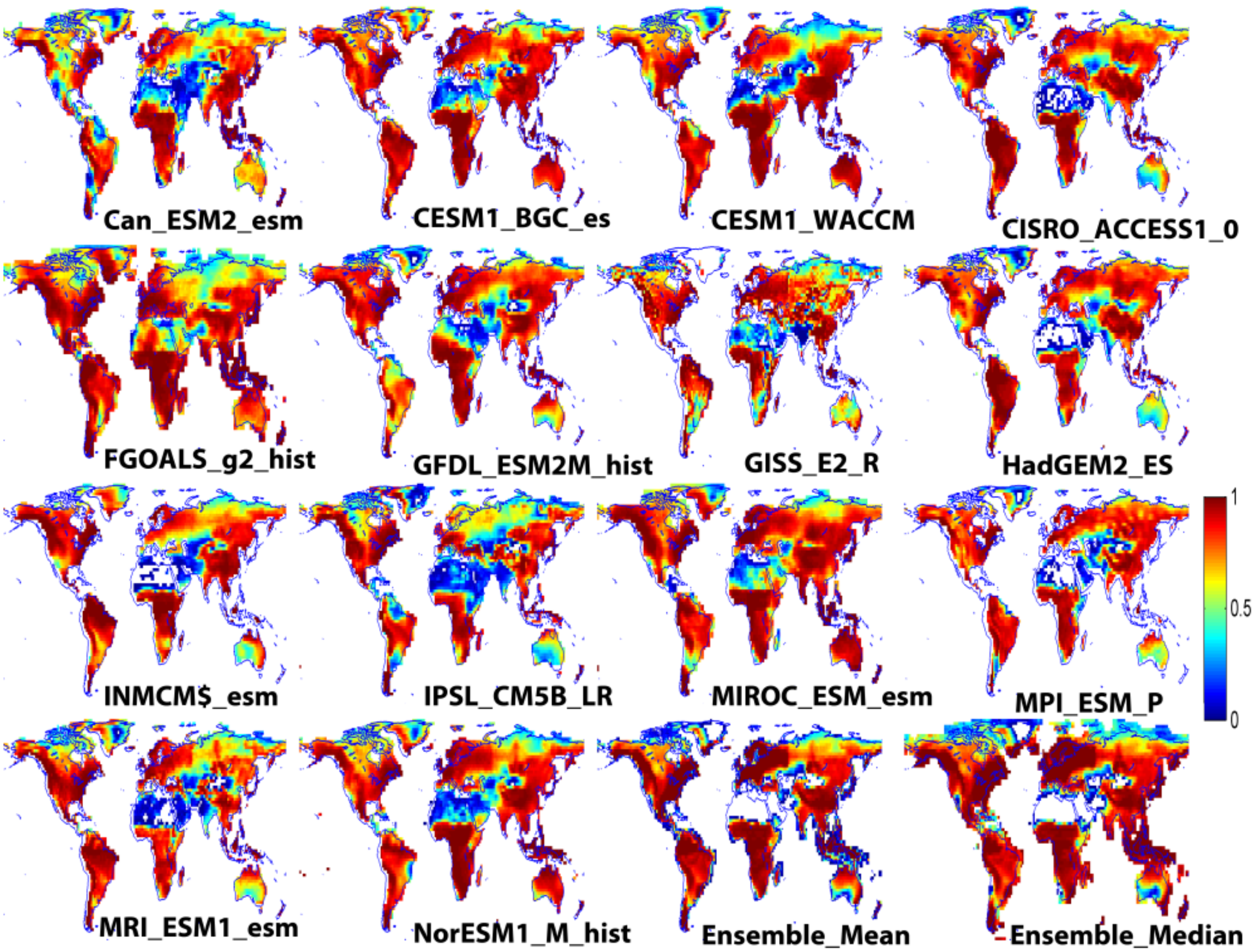

Figure 3. Volumetric Hit Index (*VHI*) for $t = 0$ (with optimal value = 1) maps for selected CMIP5 precipitation simulations, and for their ensemble mean and median, all with respect to GPCP observations.

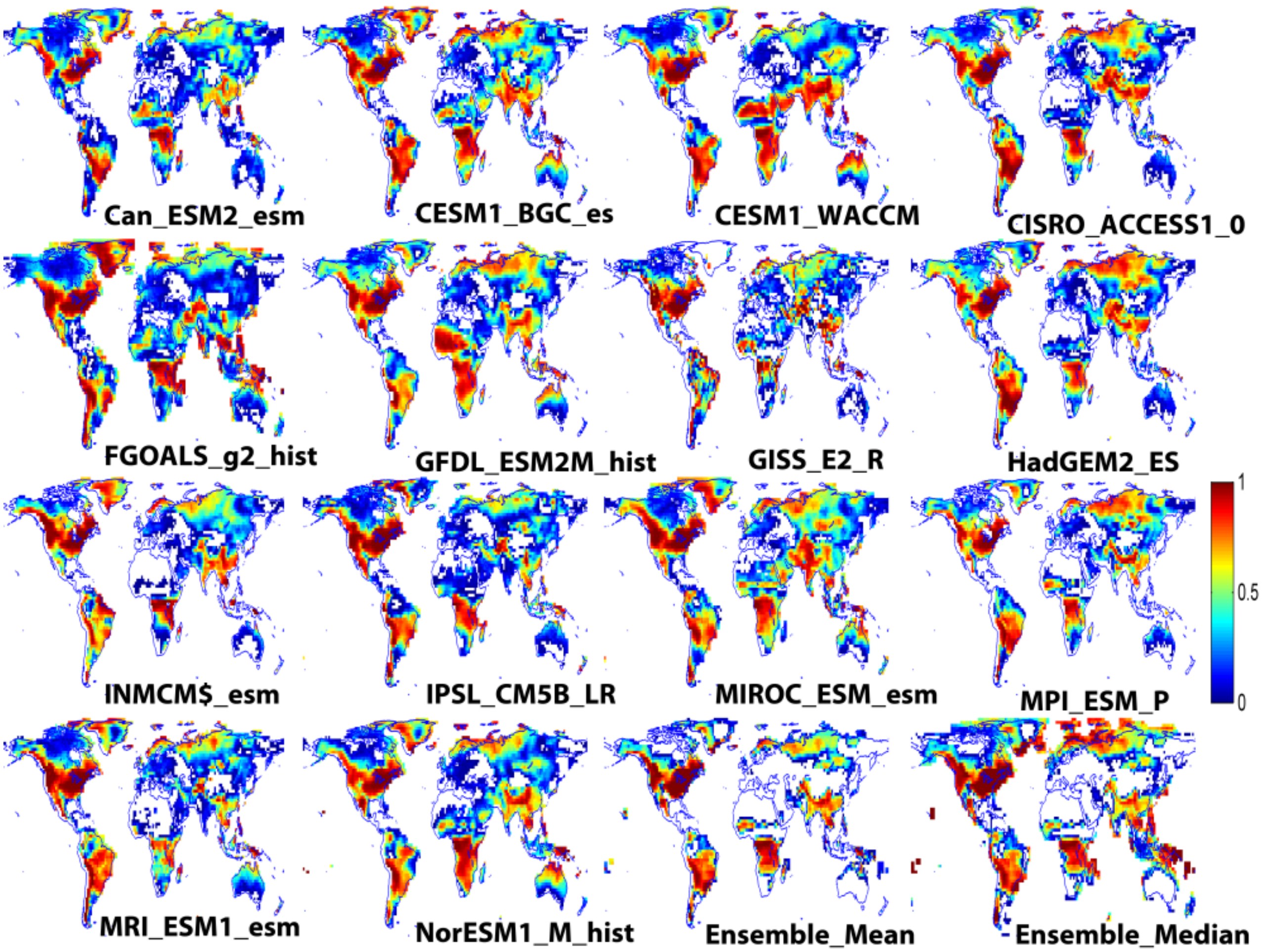

Figure 4. Volumetric Hit Index for t = 75 percentile VHI_{75} (with optimal value = 1) for selected CMIP5 precipitation simulations, and for their ensemble mean and median, with respect to GPCP observations.

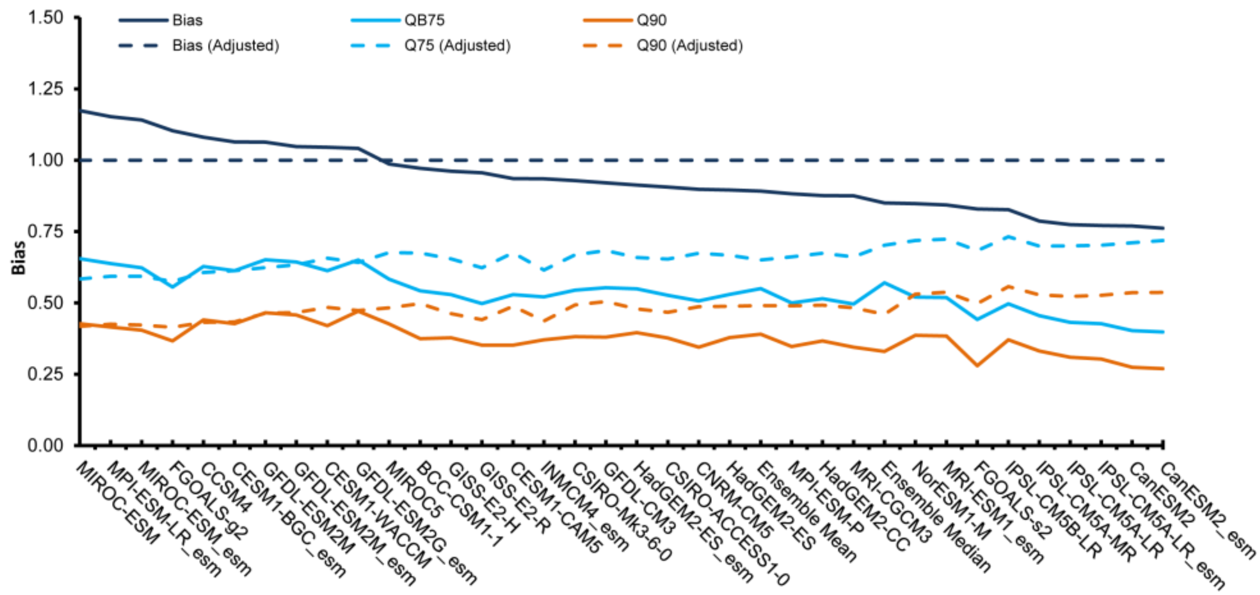

Figure 5. Averaged Bias (*B)* and 75th and 90th Quantile Bias values (*QB*$_{75}$ and *QB*$_{90}$, all with optimal values = 1) for selected CMIP5 precipitation simulations, and for their ensemble mean and median, both before bias adjustment (solid lines) and after (dashed lines), all with respect to GPCP reference data.

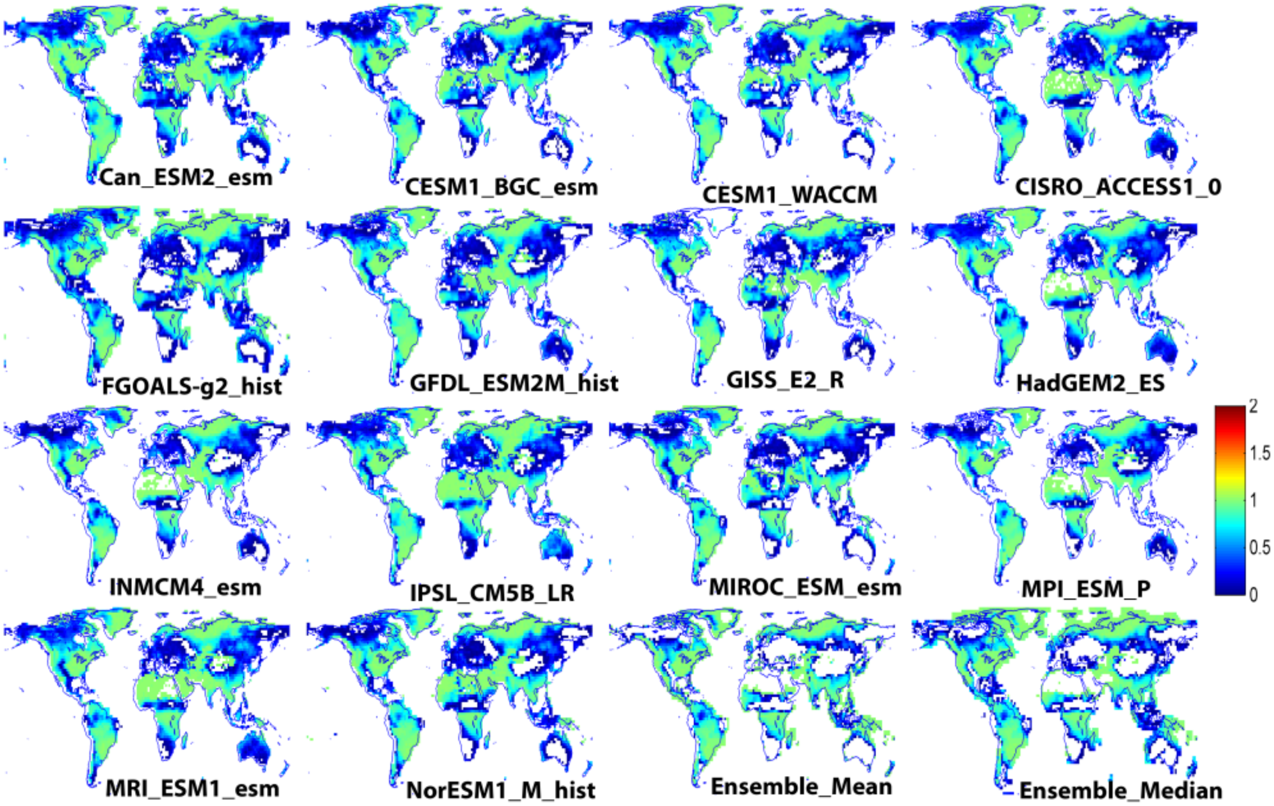

Figure 6. Quantile Bias (t=75th percentile, QB_{75} with optimal value =1) maps after bias-adjustment of CMIP5 precipitation simulations, with respect to GPCP observations.

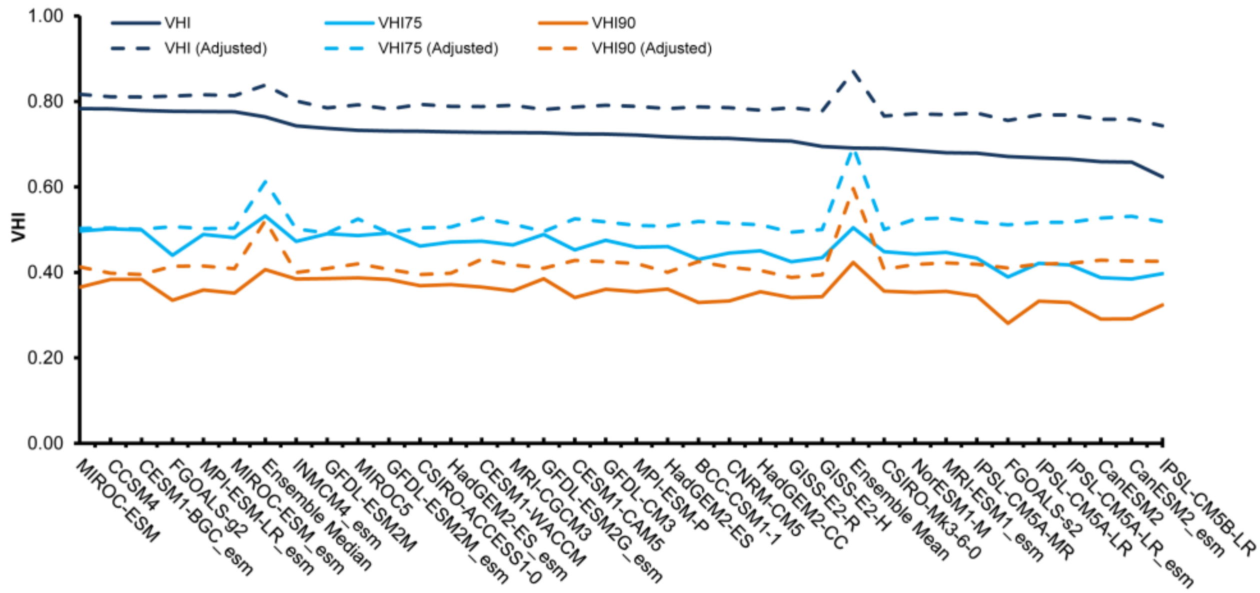

Figure 7. Averaged Volumetric Hit Index (*VHI*) and 75th and 90th percent quantile values VHI_{75} and VHI_{90} (with optimal values = 1) for selected CMIP5 precipitation simulations, and for their ensemble mean and median, both before bias adjustment (solid lines) and after (dashed lines), all with respect to GPCP reference data.

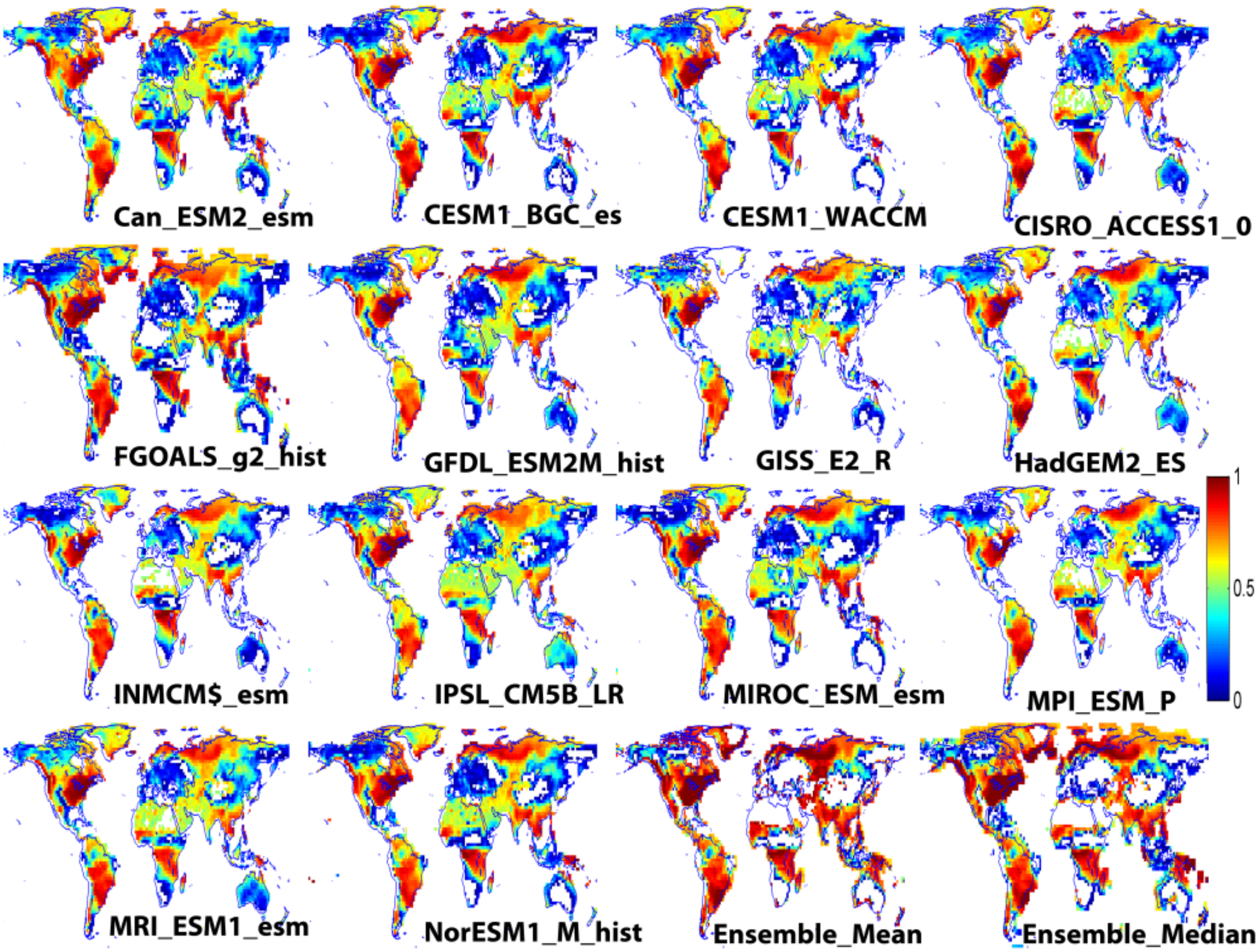

Figure 8. Volumetric Hit Index for t = 75th percentile (VHI_{75}, with optimal value = 1) after bias-adjustment of selected CMIP5 precipitation simulations, all with respect to GPCP observations.

Table 1 – CMIP5 climate models and summary statistics of global Bias (*B*) and Volumetric Hit Index (*VHI*) before and after mean-field bias adjustment (after bias adjustment *B*=1 for all models). Optimal values of these metrics are all equal to 1.

Climate Models	**Original Data**		**After Bias Adjustment**
	B	***VHI***	***VHI***
BCC-CSM1-1	0.97	0.71	0.79
CanESM2_esm	0.76	0.66	0.76
CanESM2	0.77	0.66	0.76
CCSM4	1.08	0.78	0.81
CESM1-BGC_esm	1.06	0.78	0.81
CESM1-CAM5	0.94	0.72	0.79
CESM1-WACCM	1.04	0.73	0.79
CNRM-CM5	0.90	0.71	0.79
CSIRO-ACCESS1-0	0.91	0.73	0.79
CSIRO-Mk3-6-0	0.93	0.69	0.77
FGOALS-g2	1.10	0.78	0.81
FGOALS-s2	0.83	0.67	0.76
GFDL-CM3	0.92	0.72	0.79
GFDL-ESM2G_esm	1.04	0.73	0.78
GFDL-ESM2M_esm	1.05	0.73	0.78
GFDL-ESM2M	1.06	0.74	0.78
GISS-E2-H	0.96	0.69	0.78
GISS-E2-R	0.96	0.71	0.79
HadGEM2-CC	0.88	0.71	0.78
HadGEM2-ES_esm	0.91	0.73	0.79
HadGEM2-ES	0.90	0.72	0.78
INMCM4_esm	0.93	0.74	0.80
IPSL-CM5A-LR_esm	0.77	0.67	0.77
IPSL-CM5A-LR	0.77	0.67	0.77
IPSL-CM5A-MR	0.79	0.68	0.77
IPSL-CM5B-LR	0.83	0.62	0.74
MIROC5	0.99	0.73	0.79
MIROC-ESM_esm	1.14	0.78	0.81
MIROC-ESM	1.17	0.78	0.82
MPI-ESM-LR_esm	1.15	0.78	0.82
MPI-ESM-P	0.88	0.72	0.79
MRI-CGCM3	0.88	0.73	0.79
MRI-ESM1_esm	0.84	0.68	0.77
NorESM1-M	0.85	0.69	0.77
Ensemble Mean	**0.89**	**0.69**	**0.87**
Ensemble Median	**0.85**	**0.76**	**0.84**