1.6

EVALUATION OF GLOBAL PRECIPITATION IN REANALYSES

Michael G. Bosilovich and Junye Chen
NASA Goddard Space Flight Center, Global Modeling and Assimilation Office

Franklin R. Robertson
NASA Marshall Space Flight Center

Robert F. Adler
NASA Goddard Space Flight Center

1 INTRODUCTION

Observations that are assimilated into reanalysis systems and the model parameterizations that are used for the weather forecast each affect the resulting precipitation form the system. Additionally, the complex interactions between the model and observations also affects the reanalysis precipitation. Kalnay et al. (1996) classified precipitation as being very close to model simulated data, and subject to large uncertainty. Ultimately, when trying to understand global and regional water cycles the uncertainty in reanalysis precipitation can be a limiting factor. For example, Dirmeyer and Brubaker (1999) use the reanalysis evaporation and moisture transport, but observed precipitation to study the basin scale water budgets. While this may be acceptable for certain studies, other studies may require the reanalysis precipitation.

The precipitation in reanalyses is closely related to all the physical aspects of the system, but also the assimilation of data. This paper aims to better quantify the uncertainties in precipitation from reanalysis and data assimilation systems, and to provide a benchmark for system development.

1.1 Background

Janowiak et al. (1998) tested the NCEP-NCAR reanalysis precipitation with several statistical approaches. The first test of a new data is the mean difference from an observed data set; in this case, the GPCP merged precipitation data (Adler et al 2002). The mean differences to become apparent, and the larger the difference, then it can be assumed that the reanalysis requires some development there. However, for many of the differences, the merged data set's own uncertainties make the results less clear.

In addition to mean differences, Janowiak et al. (1998) used temporal correlations, EOF analysis and anomaly correlations. While these analysis techniques provide additional information on the reanalysis precipitation, they rely on the existence of a sufficiently long time series. When developing a new system, a long time series is generally not available. Also, these time series evaluations tend to assume one or another observed precipitation data set for comparison. However, there are differences in observed data sets relating to developing retrieval algorithms, input data, treatment of gage uncertainties and quality flags. Gruber et al. (2000) and Yin et al. (2004) compared GPCP and CMAP merged precipitation data sets (for 1979-2001) and found spatial correlation over land, but significantly low correlation over ocean.

1.2 Bias and Spatial Correlation

In this paper, we will compare several existing reanalyses precipitation with the GPCP merged data set. Both mean bias and spatial correlation will be tested in the reanalyses. Only the monthly mean precipitation fields will be used, as opposed to anomalies or analysis that relies on long time series. In that way, developmental systems and new operational analyses can be tested along side the reanalyses. The comparison between CMAP and GPCP will provide uncertainty estimates on the comparisons, both for bias and spatial correlation.

2 DATA AND METHODOLOGY

2.1 Observed and reanalysis data

Since GPCP (Adler et al. 2002) is comprised of observations with global coverage, correlations will be calculated against it. However, it is not clear that GPCP precipitation can be considered ground truth, so that we need to include a measure of uncertainty. The CMAP precipitation (Xie and Arkin, 1996) will also be compared to GPCP in an effort to represent uncertainty. The CMAP precipitation provides two products; one includes NCEP reanalysis information to fill missing data in the other. The CMAP observed time series will be used for comparison.

We evaluate 5 global atmospheric reanalyses for the period of 1979 through 2005 (if available). The Japanese 25 year Reanalysis (JRA-25) is the most recent, released for use in

[*] Corresponding Author Address: Michael G. Bosilovich, Global Modeling and Assimilation Office, NASA GSFC Code 610.1, Greenbelt, MD 20771 email: Michael.Bosilovich@nasa.gov

March 2006 (http://www.jreap.org). The 45 year ECMWF reanalysis (ERA40, Uppala et al. 2005), which stops in August 2002. The National Centers for Atmospheric Research (NCEP) has released two reanalyses labeled here as NR1 (NCEP-NCAR, Kalnay et al 1996) and NR2 (NCEP-DOE, Kanamitsu et al. 2002). We also include a reprocessing of this period using the NASA GEOS4 data assimilation system (Bloom et al. 2005). GEOS4 was the operational analysis for NASA from 2003 through 2006.

2.2 Spatial Correlation

We use monthly means from each of the reanalyses, specifically to evaluate the climatology, impact of changing observing systems and time series. Higher frequency data would be needed to evaluate statistics on precipitation frequency and intensity. In the climate system, the global pattern of precipitation is as important as the mean bias of precipitation. In other words, are the reanalyses producing precipitation, or the lack there of, in the right places. A small global or regional bias may mask the spatial variability of precipitation, for example, a shift in the tropical convergence. So it is important to weigh both the mean bias with the correlation.

All monthly means are regridded to 2.5°×2.5° resolution (box averaging for finer grids, bilinear interpolation for coarser grids). All spatial averaging for biases and correlations use area weighting in the calculation.

3 CORRELATION AND BIAS

Here, we present the comparisons of the time series of annual average spatial correlations and mean differences (also called bias in the text), the mean annual cycles of over the period for the globe and several continental and oceanic regions (Figure 1).

3.1 Global and Tropical Regions

Time series of global precipitation correlations generally show that the JRA25 has increasing correlations with time, with a notable increasing shift around the time that SSMI becomes available (Figure 2). The JRA global bias tends to be lower than most of the other reanalyses in the recent years as well. This is in large part due to improved tropical precipitation (Figure 3).

ERA40 generally show good correlation values, compared to the other reanalyses, but the tropical precipitation bias, greatly affects the time series. The increasing trend in precipitation is not apparent in any of the global observation data.

CMAP and GPCP tend to show better correlation to each other over land than ocean (Figure 2). While the JRA and ERA40 correlations over ocean seem to show improvement for the reanalysis products there, there has been little improvement over land (when taken in the global sense). However, JRA25 does have marginally higher correlations in the SSMI period than the other reanalyses.

The mean annual cycle of correlations at the global scale shows significant seasonal fluctuations (Figure 4). These can be explained more with regional analyses (following sections). However, the NCEP Reanalysis 2 seasonal bias of land continental is also apparent in the global average (Figure 5).

3.2 Continental Regions

We have evaluated the reanalyses over large continental regions (Figure 1). The spatial correlations over North America and Europe for JRA25 and ERA40 show superior skill in precipitation from compared to the others, with most of this higher skill realized over Europe. (Figure 6). The mean biases are different, however. In North America, Most of the reanalyses overestimate both observed datasets, except ERA40. Over Europe, the reanalyses are consistently lower than observations by about 0.4 mm/day (Figure 7). In South America, all of the reanalyses have low spatial correlations, but they seem to be increasing slightly with time. For Africa, there is slight decreasing trend of the matched correlation of the observational data sets (Figure 6). JRA has a sharp decrease of the correlation in 1998. The GEOS4 precipitation shows a sharp increase in correlation when SSMI becomes available. ERA40 has generally the most consistent and highest time series of spatial correlation. All of the analyses have significant biases in South America (Figure 7).

In North America, JRA25 has a distinct annual cycle of spatial correlation, where it is high in winter and spring, but drops in summer (Figure 8). The annual cycle of other reanalyses are similar to JRA25, with the exception that they are in general lower than JRA25. On the other hand, ERA40 has smaller amplitude of the annual cycle, and the summer correlations are higher than JRA25. Mean biases are generally larger in the summer season in many of the regions. However, ERA40 has smaller amplitude of the mean bias than the other analyses (Figure 9). The South American biases are generally large magnitude. There are also regional differences in the observations biases. The North American difference of precipitation is continuous, around 0.2 mm/day. In Europe, the differences are large in winter (0.8 mm/day) and smaller in summer (0.2 mm/day).

3.3 Oceanic Regions

The oceanic precipitation in JRA25 shows very good correlations in the tropical Pacific Ocean and Indian Monsoon regions, with the noticeable increase after SSMI is included (Figure 10). Likewise, ERA40 seems to have good correlations as well. All the reanalyses have too much precipitation in these tropical regions (Figure 11). ERA40 has better correlations over the North Atlantic Ocean, with small or low biases. The GEOS4 analysis performs fairly well in the tropical oceans, but over the northern oceans, there appears to be distinct deficiencies in both correlation and mean bias.

Low correlations between the observations in the northern oceans tend to occur in the winter season, when the mean difference is larger than any other time in the seasonal cycle (Figure 12 and Figure 13). In the mean annual cycle, the Indian monsoon correlations are very similar during the monsoon, and there is some stratification of the correlations early in the year (Feb-May).

4 UNMATCHED CORRELATIONS

In the previous evaluation of the spatial correlation, GPCP and CMAP monthly means were correlated for the same (matching) months. These should be high, but not equal to one owing to the different merging techniques and variations in sources of data. Monthly precipitation does have a fairly regular distribution. For example, the location of the ITCZ or arid regions of North America and Africa occur regularly in the same location, and negative correlations for certain regions would be difficult if not impossible. Given that the reanalyses can likewise reproduce some of these large scale patterns, and the uncertainty in the reanalyses is more subtle, we need to define a minimum of spatial correlation to determine the quality of reanalysis values.

We have correlated unmatched monthly means of the GPCP and CMAP data sets. For example, January 1979 of GPCP is correlated to all of 1980-2003 Januaries of CMAP. The accumulation of the unmatched correlations follows,

$$\sum_{m=jan}^{dec}\sum_{i=1979}^{2003}\sum_{j=1979}^{2003} corr\left(Pg_{im}, Pc_{jm}\right)\delta_{ij}$$

$$where\ \delta_{ij} = 1, \quad if \quad i \neq j.$$

For averaging, the N number of times is determined by summing δ_{ij}, and the seasonal cycle of the unmatched correlation can also be determined. Pg and Pc represent GPCP and CMAP precipitation, respectively. These values represent the mean of spatial correlations of different years. The values are generally positive, owing to the fact that precipitation patterns have some regularly occurring features. If the reanalyses cannot have a spatial correlation greater than this mean value, it may indicate a region where further study in the model performance is needed.

Figure 14 shows the grand mean of all the unmatched spatial correlations for the regions discussed earlier and also some latitudinal bands. For the global spatial correlation, 0.69 is the mean, so that most years, the reanalyses annual means are above this value (Figure 2). While this is also true for the ocean, the land unmatched correlation of 0.80 is in the middle of the annual means. So, while it was concluded earlier that the reanalyses precipitation over land were not improving, they also seem to be, generally, not able to produce very good spatial distributions for the land. None of the reanalyses precipitation exceeds the unmatched correlation in South America and Africa (though there are a few years where ERA40 is close in Africa).

Europe has a very low unmatched correlation (0.42) compared to other regions. In the mean annual cycle, the lowest values are February – May (not shown). All of the reanalyses have annual values that exceed this correlation, likely related to the large scale storm track in the region. In North America, the lowest unmatched correlations (as well as matched correlations) occur during July-September, during the end of the warm season, when land-atmosphere interactions and the soil water reservoir contribute to the precipitation.

For the Indian Monsoon region, the JRA25 is less than the unmatched correlation mean before SSMI becomes available. However, it is well above that when SSMI are being assimilated. Most of the reanalysis precipitation is above the unmatched correlations in the Indian monsoon region.

Lastly, the matched and unmatched correlations for high latitudes are lower than other latitude bands, which may be related to both a high degree of variability as well as uncertainty in the data. The uncertainty in the data is an issue, because the matched correlations are lower than other bands and the global mean. This identifies the high latitudes as a region of development for the next generations of merged precipitation data sets.

5 SUMMARY

In this paper, we are developing a metric to evaluate global and regional precipitation in reanalysis data products. Spatial correlations provide an estimate of the agreement in spatial patterns. By correlating two observed data sets at matched times, we can identify the uncertainty in those data. By correlating the observed data at unmatched times, we can estimate a minimum value of correlation that the reanalyses need to attain when matched with an observation data set. The unmatched correlation mean represents the background correlation that exists in the real climate system. We have used this to identify where and when existing reanalyses excel or fail.

While other methods, such as anomaly correlation and EOF analysis, can do the same thing, they require existing long periods of data. The current analysis requires only monthly means, and so, could be used in developing a new system. Gruber et al. (2000) used a filter to remove fine spatial structures from CMAP and GPCP, and found that the ocean anomaly correlations were substantially increased. As reanalyses and observed precipitation data sets move to finer spatial scales, this approach must be considered carefully. Some of the fine structure may be important in the evaluation of the fine scale reanalyses. Here, we have simply reduced the resolution of the reanalyses to match the GPCP monthly climatology.

6 REFERENCES

Adler, RF; Huffman, GJ; Chang, A; et al., 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). J. Hydromet., 4 (6): 1147-1167.

Bloom, S. C., A. da Silva and D. Dee, 2005: *Documentation and Validation of the Goddard Earth Observing System (GEOS) Data Assimilation System Version 4*. Vol. 26, *NASA Technical Memorandum 104606*, NASA Goddard Space Flight Center, 165 pp.

Dirmeyer, P. A., Brubaker, K. L., 1999: Contrasting evaporative moisture sources during the drought of 1988 and the flood of 1993. J. Geophys. Res., 104 (D16): 19383-19397.

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

Janowiak, JE; Gruber, A; Kondragunta, CR; et al., 1998: A comparison of the NCEP-NCAR reanalysis precipitation and the GPCP rain gauge-satellite combined dataset with observational error considerations. J. Clim., 11 (11): 2960-2979.

Kalnay E. and co-authors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. Bull. Amer. Meteor. Soc., 77, 437-431.

Kanamitsu M., W. Ebisuzaki, J. Woollen, S-K Yang, J.J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). Bull. Amer. Met. Soc., 83, 1631-1643.

Uppala SM, and co-authors, 2005: The ERA-40 re-analysis, QJRMS, 131 Part B, 2961-3012.

Xie P., and P. A. Arkin, 1996: Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull. Amer. Meteor. Soc., 78, 2539-2558.

Yin, XG; Gruber, A; Arkin, P, 2004: Comparison of the GPCP and CMAP merged gauge-satellite monthly precipitation products for the period 1979-2001. 5 (6): 1207-1222.

7 ACKNOWLEDGEMENTS

GPCP data provided by the Laboratory for Atmospheres, NASA Goddard Space Flight Center, from: http://precip.gsfc.nasa.gov/

CMAP data provide by the NOAA/NWS/CPC, from http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.cmap.html

NCEP/NCAR Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from http://www.cdc.noaa.gov/

NCEP-DOE_Reanalysis 2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from http://www.cdc.noaa.gov/

ERA-40 data provided by the European Centre for Medium-Range Weather Forecasts, from http://www.ecmwf.int/

JRA-25 data provided by the Japan Meteorological Agency (JMA) from http://www.jreap.org/

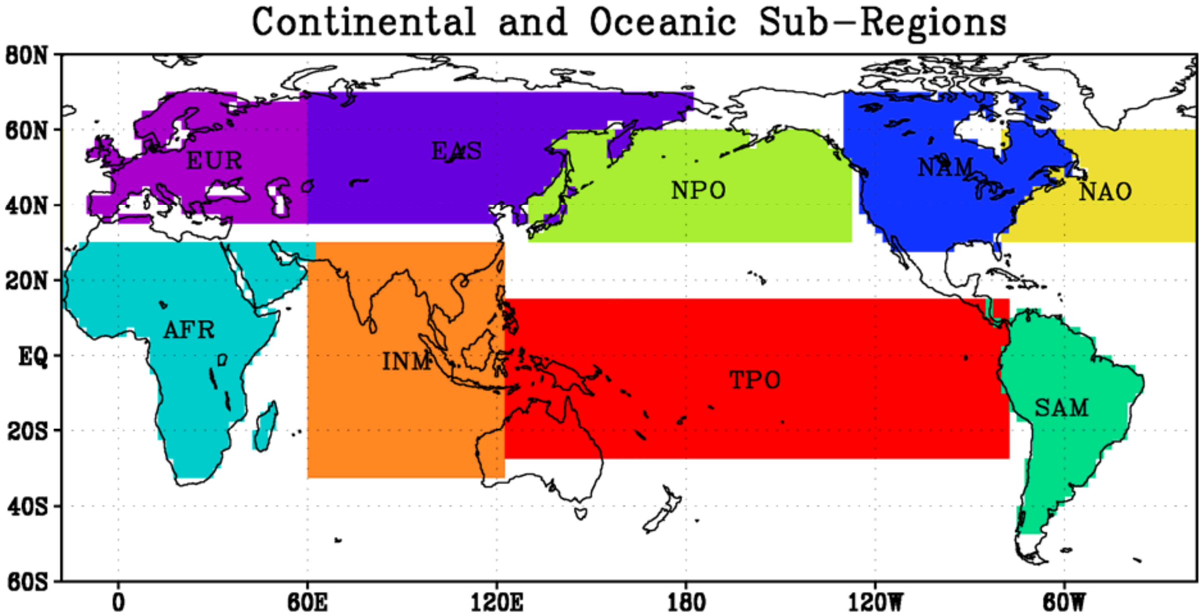

Figure 1 Regional areas for reanalysis and observation comparisons: EUR – Europe; EAS – Eurasia; NPO – North Pacific Ocean; NAM – North America; NAO – North Atlantic Ocean; AFR – Africa; INM – Indian Monsoon; TPO – Tropical Pacific Ocean; SAM – South America

1.6

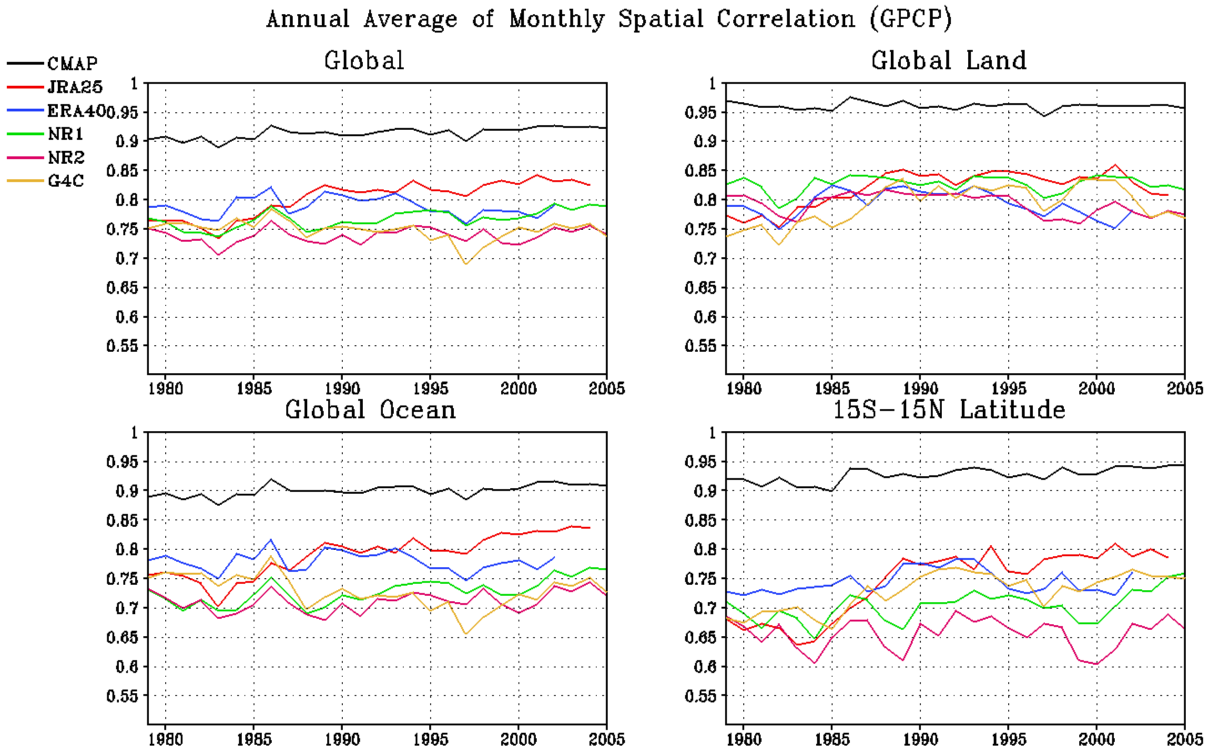

Figure 2 Annual average of monthly spatial correlations to GPCP for global areas.

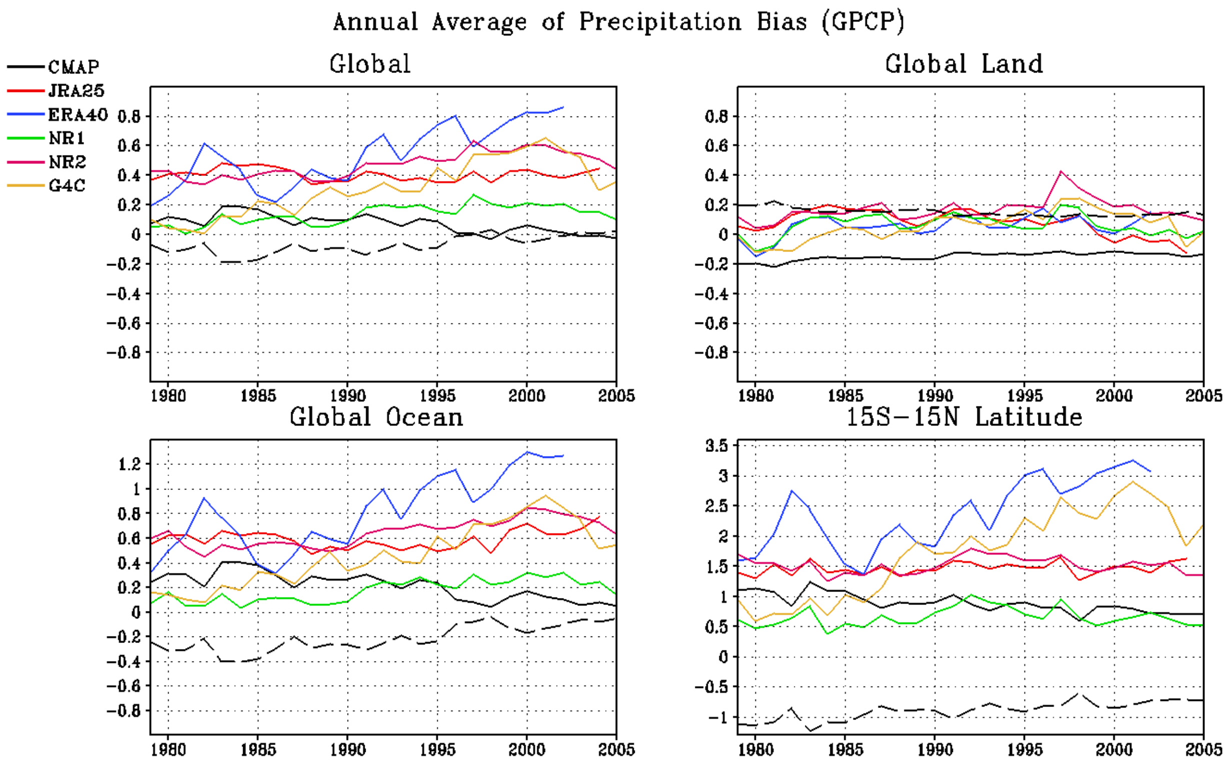

Figure 3 Annual average of monthly mean differences from GPCP for global regions. The dashed line indicates the difference GPCP minus CMAP, the opposite of the solid line, for a point of reference.

1.6

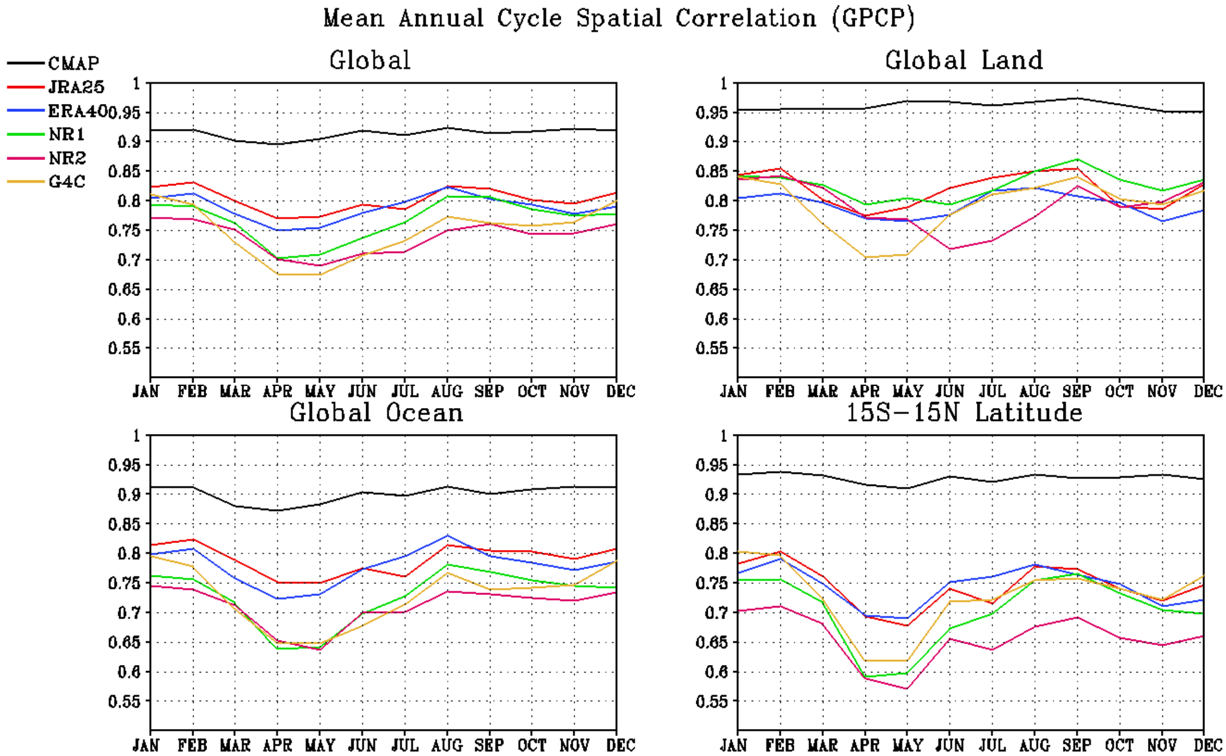

Figure 4 Mean annual cycle of spatial correlations for global regions.

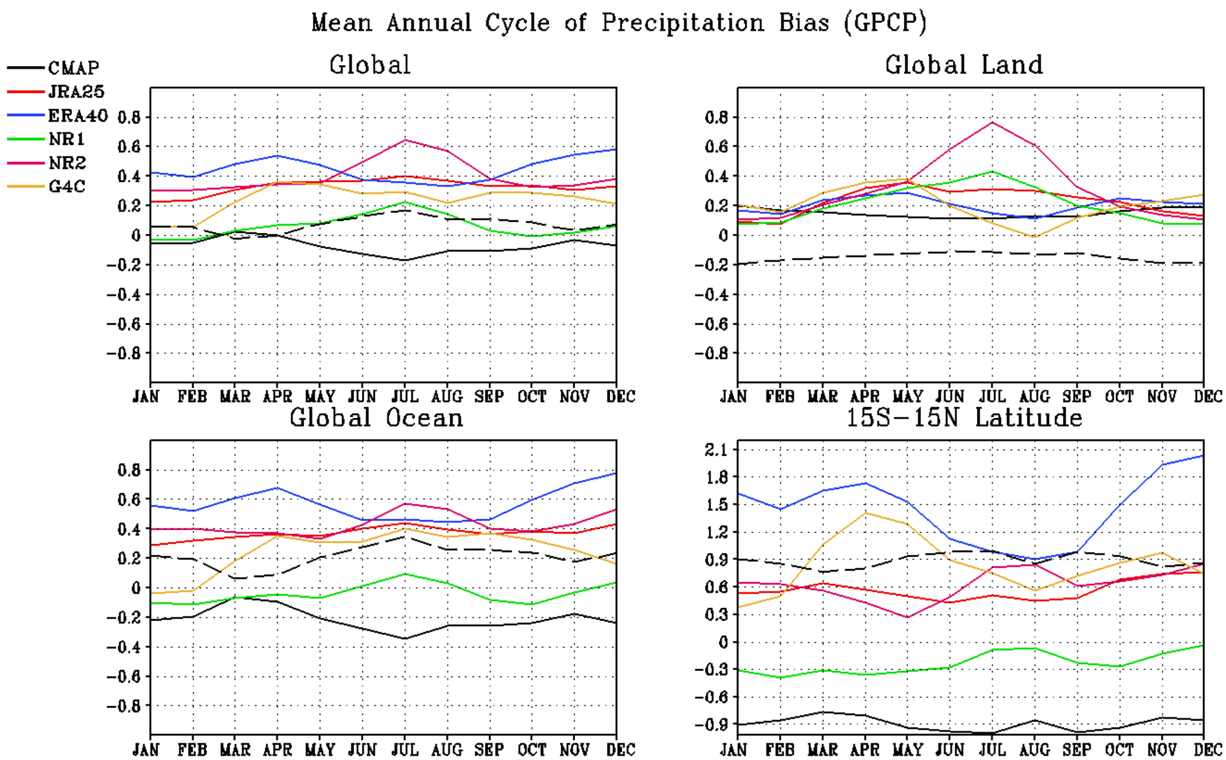

Figure 5 Mean annual cycle of mean differences from GPCP for global regions.

1.6

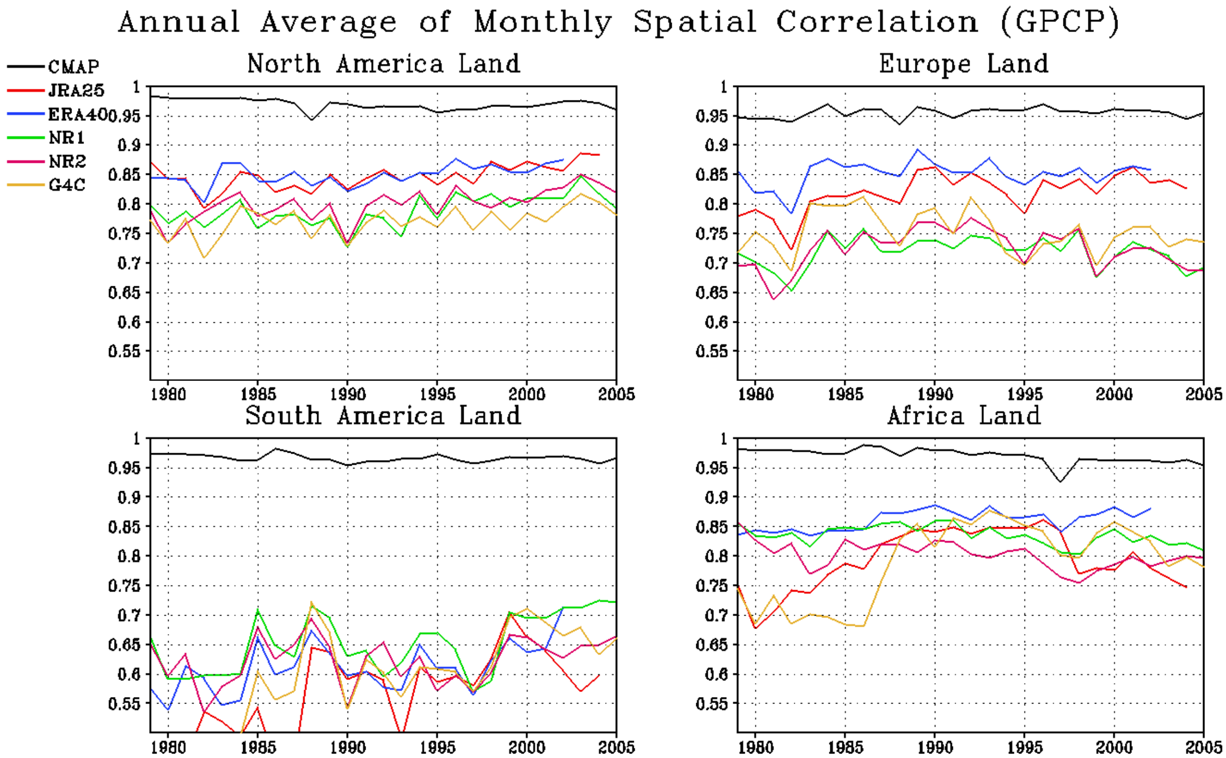

Figure 6 As in Figure 2, except for continental regions.

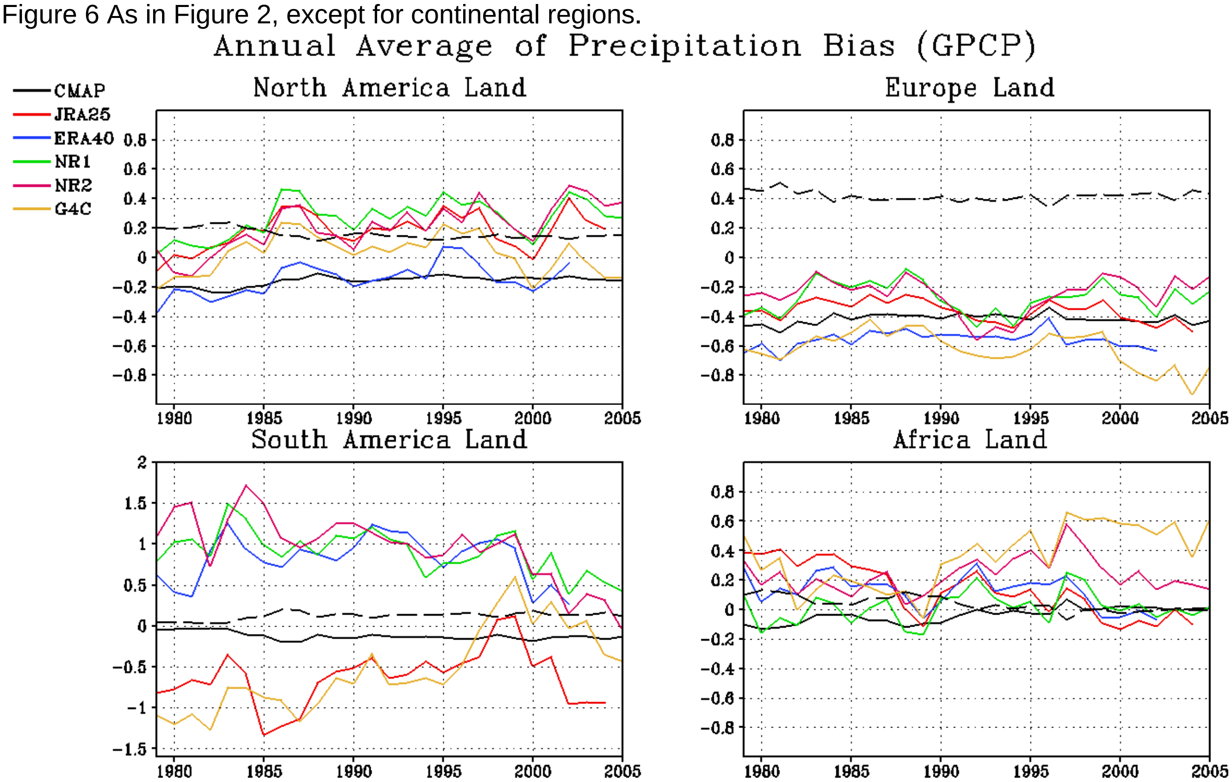

Figure 7 As in Figure 3, except for continental regions.

1.6

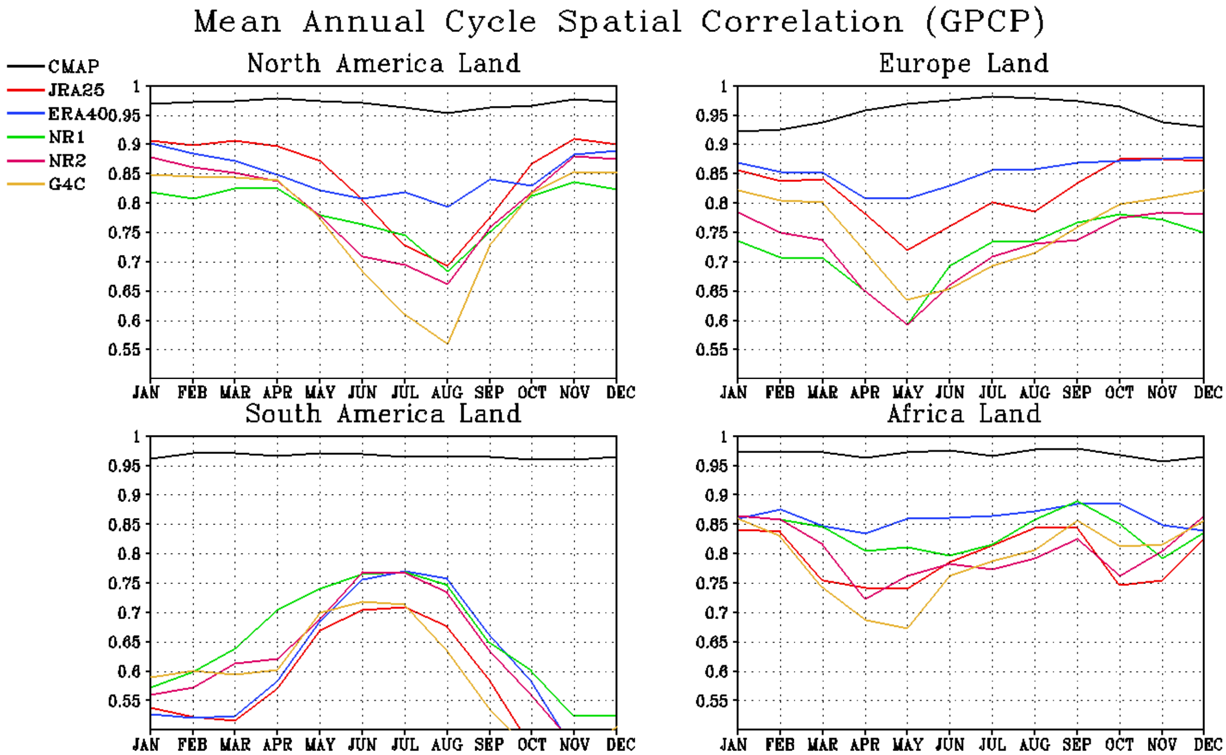

Figure 8 As in Figure 4, except for continental regions.

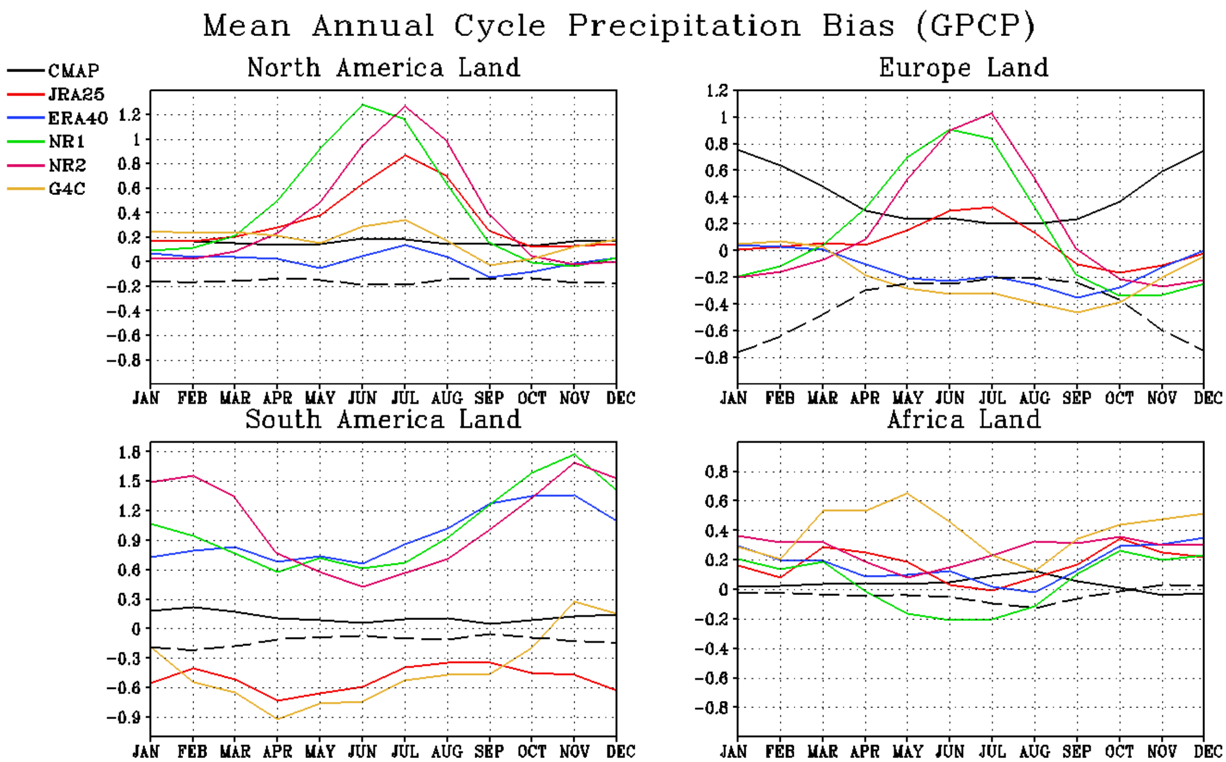

Figure 9 As in Figure 5, except for continental regions.

1.6

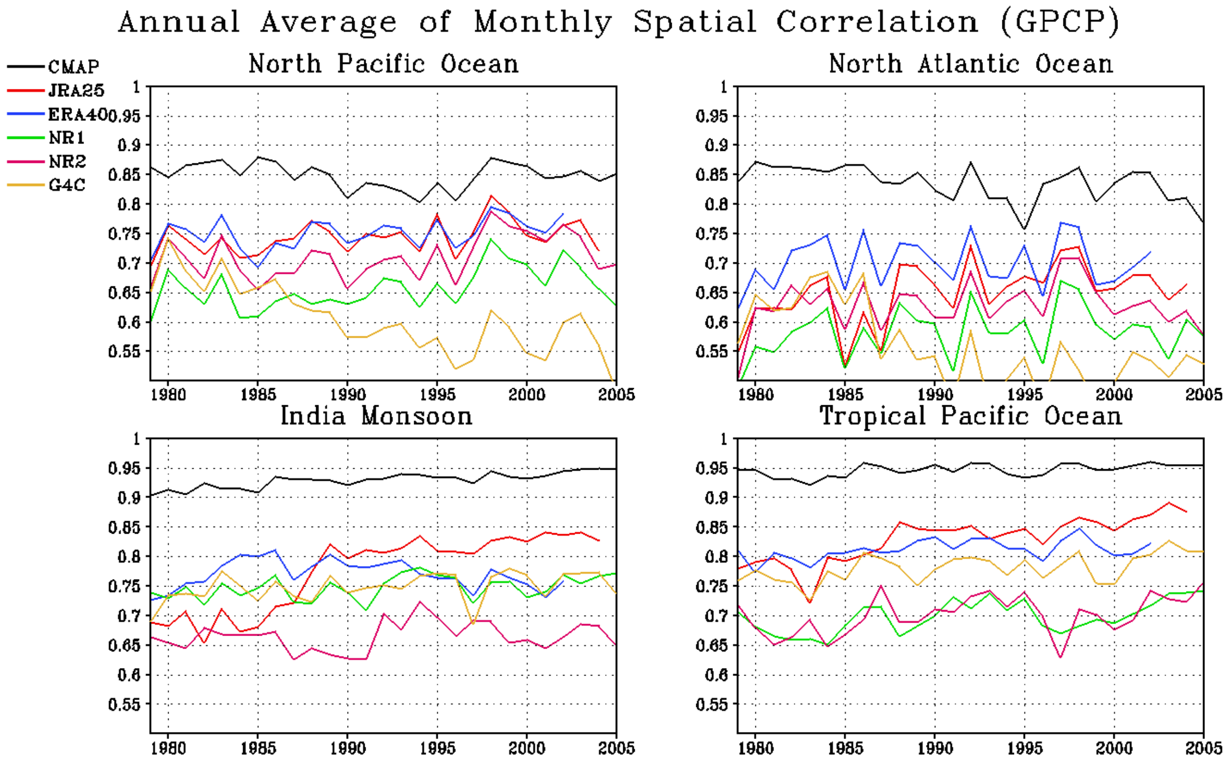

Figure 10 As in Figure 2, except for oceanic regions.

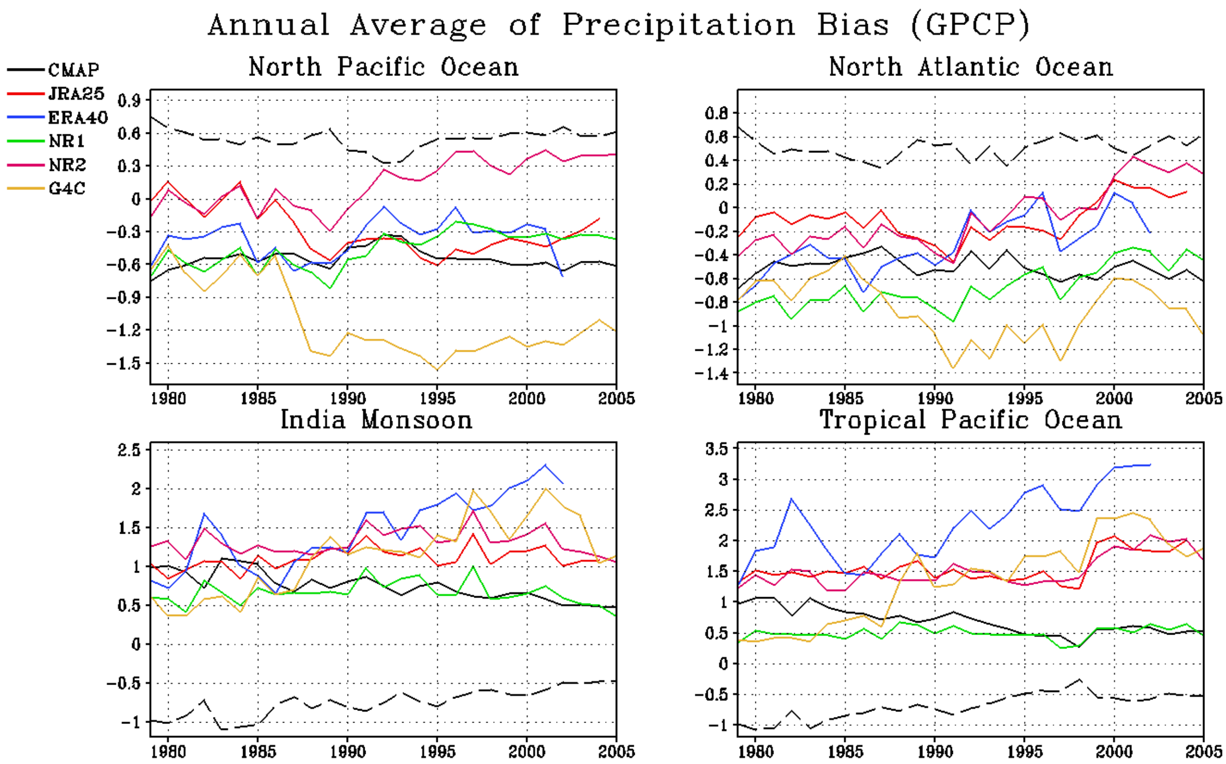

Figure 11 As in Figure 3, except for oceanic regions.

1.6

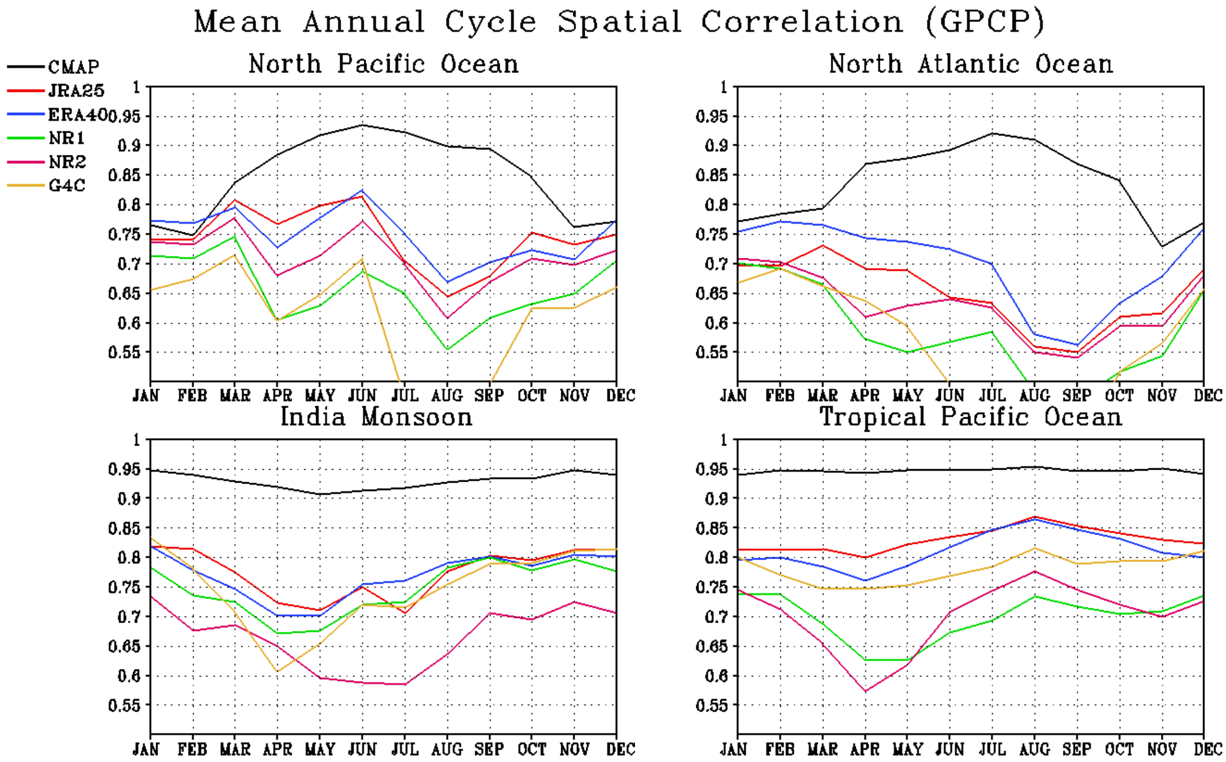

Figure 12 As in Figure 4, except for oceanic regions.

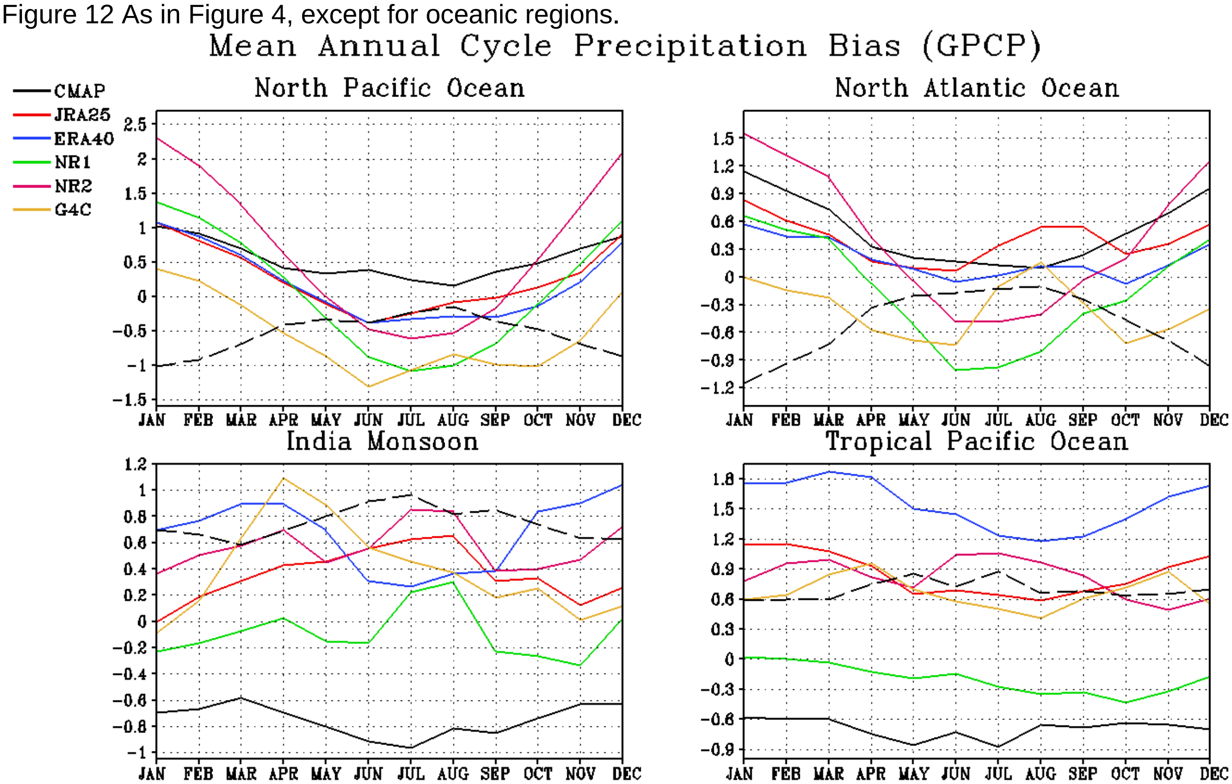

Figure 13 As in Figure 5, except for oceanic regions.

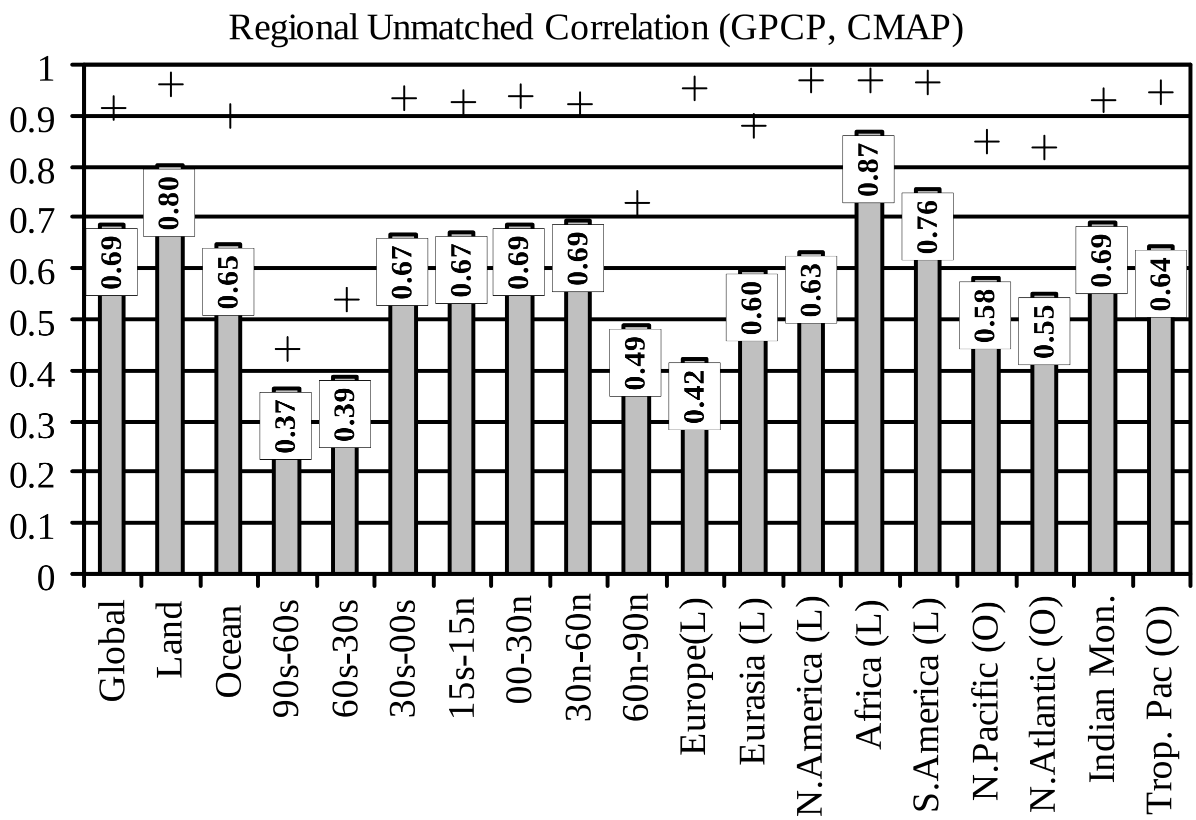

Figure 14 Average of all unmatched correlations between GPCP and CMAP for various regional domains over the period 1979-2003) in the bars. The crosses show the average of all matched correlations (used in the previous figures as time series of annual means or mean annual cycles). See the text for definitions.