7/30/08

# A Multi-model Analysis for the Coordinated Enhanced Observing Period (CEOP)

Michael G. Bosilovich *

and

David Mocko

NASA/Global Modeling and Assimilation Office

Goddard Space Flight Center Greenbelt MD

John O. Roads

Experimental Climate Prediction Center

Scripps Institute for Oceanography, La Jolla CA

Submitted to JHM

* *Corresponding address*: Global Modeling and Assimilation Office, Code 610.1 NASA Goddard Space Flight Center, Greenbelt, MD 20771, Michael.Bosilovich@nasa.gov

# A Multi-model Analysis for the Coordinated Enhanced Observing Period (CEOP)


ABSTRACT

A collection of eight operational global analyses over a 27 month period have been processed to common data structures to facilitate comparisons among the analyses and global observational data sets. In the present study, we evaluate the global precipitation, outgoing longwave radiation (OLR) at the top of the atmosphere and basin-scale precipitation over the United States. In addition, a multi-model ensemble is created from a linear average of the available data, as close to the analysis time as each system permits. The results show that the monthly global precipitation and OLR from the multi-model ensemble has generally better comparison to the observations than any single analysis. Likewise, the daily precipitation from the ensemble exhibits better statistical comparison (in space and time) to gauge observations over the Mississippi River Basin. However, the statistics have seasonality, where the members of the ensemble are much improved, and closer to the ensemble mean, during winter. Using the global precipitation and OLR, the sensitivity is tested to selectively choosing the members with the best statistical comparisons to the reference data. Only small improvements in the statistics are found when comparing a selective ensemble to the full ensemble. The ensemble data and the variance of the ensemble should make a useful point of comparison for the development of global analyses.

## 1. Introduction

Ensembles means of simulations using different models have been shown to provide a result better than any of the contributing members. At weather scales, improved hurricane predictions have been found through such super ensembles (Krishnamurti et al. 2003). In climate simulations and predictions, a multi-model approach also tends to provide the better result (Philips and Gleckler 2006). Additionally, ensembles of stand-alone land process models (using observations and analyses as prescribed forcing) show the smaller biases and errors than the contributing members (Dirmeyer et al 2006). In retrospective-analyses (or reanalyses) of the weather and climate, numerous diagnostic fields are classified as being related to the model uncertainties, as opposed to fields closely related to assimilated observations, and therefore lower quality and requiring further validation when evaluated (Kalnay et al, 1996). Compo et al. (2006) demonstrated that assimilating only surface pressure with an ensemble filter approach can produce reasonable weather patterns. It stands to reason then that an ensemble based on operational analyses diagnostics (assimilating large amounts of satellite and radiosonde observations) can produce not only reasonable weather systems, but an improved representation than any single analysis. A difficulty to this point is that a collection of analyses would be required to provide enough members of such an ensemble.

The Coordinated Enhanced Observing Period (CEOP, Koike et al. 2004) has collected concurrent observations and operational analyses for the period October 2002 – December 2004, where a primary objective was to quantify the uncertainty of analyses (Bosilovich and Lawford 2002). Requests were sent to numerical weather prediction centers for contributions to the CEOP model data archive. As of January 2008, eight analyses for the full period have been submitted. While a suggested variable list was included in the request, data structures were not strictly provided. So that, the contributed data was on various grids and each center provided its default analysis and forecast cycle data. In short, the data are not immediately comparable among the various centers. This paper presents the first results of the Multi-model Analysis for CEOP (MAC). The purpose of MAC is to homogenize the data files, providing a common spatio-temporal grid of the analyses, using as many of the most common variables in order to facilitate comparisons among the analyses and observational data. In this framework, we can then assess the current state of uncertainties among the analyses.

In addition to homogenizing the data structures, a mean and variance of the data has been produced. We hypothesize that the extensive use of observations in modern analysis/forecast systems will provide commonality and the uncorrelated model errors that exist in the analyses can be reduced in an ensemble average of the analyses. If so, the ensemble of analyses can provide a baseline for comparison of physical quantities not easily observed or have no independent source. Comparing an individual analysis to the ensemble will show the uncorrelated error in that system, while comparing the ensemble to observed data will show the correlated errors common among analyses (in combination with observed uncertainty). In this manuscript, we summarize the homogenization of the data, the development of the ensemble and variance, and present comparisons of key model diagnostics at monthly and daily time scales.

## 2. Data and Methods

Early in the formulation of the Coordinated Enhanced Observing Period (CEOP), the need for global model analysis data to support science objectives became apparent. Additionally, the observations being developed for CEOP would be very useful to the validation of model analyses and forecasts. Invitations were sent to the major international Numerical Weather Prediction and data assimilation Centers (NWPCs). Ten centers responded favorably, and by the end of 2006, seven centers provided 27 months of data for the CEOP EOP-3 and -4

(Oct 2002-Dec2004). Two separate model contributions from one center gave a total of eight analyses. The contributing centers are:

- Bureau of Meteorology Research Centre (BMRC)
- Centro de Previsão de Tempo e Estudos Climáticos (CPTEC, The Center for Weather Forecasts and Climate Studies)
- Experimental Climate Prediction Center (ECPC)
- Japan Meteorological Agency (JMA)
- Meteorological Services of Canada (MSC)
- National Centers for Environmental Prediction (NCEP)
- United Kingdom Meteorological Office (UKMO)

In general, comparisons of the analyses from the NWPCs has primarily been through the single-point Model Output Location Time Series (MOLTS) co-located with CEOP reference sites or considering only one model system (Yang et al. 2007; Chou et al. 2007; Rikus 2007, Milton and Earnshaw, 2007; Hirai et al. 2007; Meinke et al. 2007; Kato et al. 2007 and Bosilovich et al. 2007). To get at the comparison of global grids, an ensemble of the analyses was developed for several purposes. First, the variance of the analyses can provide a measure of uncertainty in analyses. It also provides a range of the state-of-the-art analyses. Second, this ensemble may make a better benchmark for comparing individual analyses than simply differencing any one against another. Lastly, a synthesis of the model output would facilitate use of the data in the broader science community where, increased use of the data should expose weaknesses in individual systems and the ensemble, leading to eventual improvements in the models.

There are several major differences in the structures of the model output data, that users of the original CEOP contributions would need to address. The original structure of the model data from the NWPCs participating in CEOP is archived (in Grib1 format) by the Model & Data group at the Max Planck Institute (MPI) for Meteorology in Hamburg, Germany. Aside from the format, there are few similarities in the contributed data. Each center provided various analysis/forecast data in their time series. Many provided their analysis to 6 hour forecast. As a rule, the data closest to the analysis provided by each center were used for that center's time series in this comparison. Details on the location of each center's data relative to the analysis/forecast cycle are provided in the appendix. In general, the model spatial grids were also different. Not all centers provided the same output variables. 42 of the most common meteorology and flux fields were selected and included in the data processing (list is in the appendix). The steps to create the ensemble and variance are as follows:

1. Generate a 6-hourly dataset for all centers, using consistent units and timing
2. Interpolate the 6-hourly data from each center to a common grid (1.25° lat-lon)
3. Create an ensemble mean and standard deviation of the 6-hourly time series
4. Create daily-averages and monthly-averages from the 6-hourly ensemble mean
5. Create daily and monthly-averages of the individual centers
6. Create daily and monthly standard deviations between the individual centers
7. Write the interpolated data for all centers, the mean, and the standard deviation at the 6-hourly, daily, and monthly times in the final formats of NetCDF and grib1

Sanity checks were performed along the way, to ensure that coding errors in the calculations were not being introduced (e.g. compare back to the source data, checking the incoming solar radiation for timing). The appendix discusses issues and decisions made at each step in transforming the output data and generating the ensemble mean. The final data includes eight different analyses located at the same time with consistent grid, as well as ensemble mean and variance of the members at 6 hourly, daily and monthly frequencies for the period of Oct 2002 – Dec 2004.

## 3. Evaluation of the ensemble and members

*a. Monthly time scale*

Precipitation from analyses can be a useful quantity, but uncertainties have to be understood (Trenberth et al. 2007; Bosilovich et al. 2008, and the reviews therein). Many of the CEOP science objectives relate to precipitation. At monthly time scales, there are many similarities among the analyses provided here. Figure 1 shows the difference of July 2004 of each of the analyses to GPCP. CMAP precipitation is also provided as a reference for observational uncertainty. Most of the analyses show high precipitation biases in the tropical Pacific Ocean, Intertropical Convergence Zone (ITCZ) and to a lesser degree the South Pacific Convergence Zone (SPCZ). Note also, that CMAP is likewise biased slightly higher than GPCP in these areas, related to the implementation of atoll gauge observations (Yin et al., 2004). The MSC and ECPC-SFM means are much closer to GPCP than the other analyses in this respect. However, the predominance of these biases across the members leads to a similar bias pattern apparent in the MAC ensemble average. This points to a key consideration of the MAC ensemble average; systematic errors and biases among the contributing members will persist into the resulting ensemble average.

While some large scale similarities are apparent, there are still many differences in the monthly precipitation of the contributing analyses. For example, continental precipitation anomalies among the members vary greatly (Figure 1). Summary statistics of mean bias and standard deviation of the difference field are included in the title of the figure. With a standard deviation of 1.7, the MAC ensemble average has lower error in this field than any of the contributing members. This then suggests that the uncorrelated errors are being reduced in the averaging of the ensemble. This result is also revealed in Taylor diagrams (Taylor, 2001) of the precipitation (Figure 2). Taylor diagrams compare the variance of a field with their correlation relative to a reference data field. In this example, GPCP provides the reference field, and CMAP is also included as a data point. The linear distance from the reference point (1,1 in Figure 2) shows how closely a model approximates the reference data (see also Philips and Gleckler, 2006 and Bosilovich et al. 2008). This shows that the MAC ensemble average is closer to GPCP than any of the ensemble members, generally, with high correlation and variance closer to that of GPCP for July 2004.

Figure 3 extends this discussion across the entire Oct 2002-Dec 2004 period. The time series of standard deviation of the monthly mean precipitation shows that the MAC ensemble average is clearly lower than any of the individual analyses. The seasonal variability of the standard deviation is also fairly small compared to large seasonal variations in some of the analyses (such as MSC). Also, there is a large shift of the JMA data beginning in June 2003, where a substantial increase in the standard deviation may also affect the MAC ensemble statistics. ECPC-RII shows a strong annual cycle with much larger standard deviations in boreal summer than otherwise. Correlation between the MAC ensemble precipitation and GPCP is also high, but there is less separation between the highest individual analyses. There are a few months in the boreal winter of 2004 where NCEP's precipitation correlation is nearly identical to the MAC ensemble average.

Outgoing Longwave Radiation (OLR) observations are generally high quality and provide a critical climate diagnostic. Figure 4 compares the July 2004 monthly differences of the ensemble members and mean OLR compared with Surface Radiation budget (SRB, Stackhouse et al, 200?) merged observational data product. As opposed to the precipitation comparison, wide variations among the models are apparent. The JMA system exhibits systematic positive biases while the UKMO shows systematic negative bias. NCEP, UKMO and MSC biases are strongest in the tropics but of different signs. While these latter three have the lowest standard deviations for this month, the MAC ensemble produces the lowest standard deviation of the difference from SRB than any single contributor.

Figure 5 shows the time series of global spatial correlation and standard deviation of the monthly differences between each member and the ensemble mean with SRB OLR. Spatial correlation shows how well the pattern matches. In spatial correlation, the ensemble mean has higher correlations than any individual model. However, the standard deviation is less clear. During Boreal winter NCEP OLR has less error than the MAC ensemble, but slightly more in error in summer. Further, the UKMO, MSC and JMA OLR data are all quite close to the MAC ensemble mean. In contrast to the precipitation statistics (Figure 3) which show that the systems are distributed along a range of values, the OLR statistics show two distinct clusters in the systems. The ECPC systems and CPTEC have markedly higher error in OLR. This result begs the question: Will choosing the statistically better systems improve the ensemble mean?

*b. Selective Ensembles*

Using Global Soil Wetness Project (GSWP2) offline land models, Guo et al. (2007) tested the sensitivity of the ensemble average to the soil moisture quality of the ensemble members. It was shown that adding better (higher correlation, lower error) members to an ensemble average reduced the error of the ensemble, but adding data with lower skill did not significantly degrade the ensemble while the better systems were in place. There are several differences between that study and the present data. First, Guo et al. (2007) were using the long reanalysis data, but these were few compared to the number of GSWP2 offline land models considered. The offline models all used similar and prescribed atmospheric forcing. The prescribed forcing likely reduces the degrees of freedom in the simulated realizations, compared to the three dimensional data assimilation data in the present analysis here. The distribution of error evaluated by Guo et al (2007) varies evenly across the ensemble members. This is somewhat different from the error we see generated in the three dimensional operational analyses here. Figure 5 shows that the OLR error delineates a subset of analyses that are less skillful than the rest. Precipitation error, on the other hand, does show a more uniform distribution across the members (Figure 3). Guo et al. (2007) also had a larger number of ensemble members in that study.

Here, we test the MAC ensemble average against a selective member ensemble, determined from each systems statistics in Figure 3 and Figure 5, using the 27-month means of the statistics to rank the systems. In each analysis we remove the lowest three scores from the comparison, for precipitation, BMRC, CPTEC and ECPC-RII are eliminated and for OLR, ECPC-SFM, ECPC-RII and CPTEC are eliminated (BMRC did not provide OLR). Figure 6a and b shows the time series of statistics including an ensemble mean of the 5 most skillful analyses precipitation for the selective ensemble. The spatial correlation of precipitation does indicate an apparent improvement (by approximately 0.01) on average for the whole period. However, this seems quite small compared to the low values of the data that was excluded. The MAC ensemble standard deviation was already a standout compared to the ensemble members. A selective ensemble does reduce the error, but only by a small margin. Also in June and July 2003, the selective ensemble standard deviation is slightly higher than that of the full MAC ensemble.

For OLR spatial correlation, there are two distinct clusters of systems (higher and lower), though all are fairly high correlations with the lowest average of monthly correlations being 0.91. The MAC ensemble at 0.98 correlation to SRB is higher than any individual system, but being so high, there is little room for improvement. Removing the lowest 3 systems from the ensemble does not increase the skill of the ensemble spatial correlation. However, in standard deviation, where NCEP winter months OLR error shows better skill than the MAC ensemble, the selective ensemble shows some improvement beyond the NCEP skill. The standard deviation or the ensembles is reduced on average from 8.6 $Wm^{-2}$ (a difference of 1 $Wm^{-2}$). So that, for some cases, a selective ensemble may provide better results, but as a rule the full ensemble will be just as accurate. While this is an improvement, and the selective ensemble skill is higher than any contributing member, the improvement is generally small in comparison to the full ensemble. The difficulty with selective ensembles is that there is no way to tell the degree that any single member may contribute to the ensemble or in what variable at any given time (e.g. seasonality as in the precipitation standard deviation Figure 6a). Since any improvements may be small, the full ensemble will be more reliable in most cases. However, if the ensemble size increases, and generally error prone members identified, selective ensembles could be justified.

The JMA OLR statistics show a general improvement over the period (Figure 5). The JMA OLR is also included in the selected ensemble. However, in neither of the ensembles is an improvement of the statistics in time noticeable (Figure 6 c and d). The ensembles' statistics are more steady compared to the individual analyses. This suggests that the individual impact of the improvement of any one operational analysis in time has limited effect on the ensemble of analyses.

### *c. Synoptic Time Scale*

The previous analysis shows that the MAC ensemble can generally produce monthly data that compares favorably to global data sets. Because the analyses use data near the observation time, weather patterns should also be resolved in the ensemble. However, the averaging of the ensemble may smooth fields, such as precipitation, at the 6 hourly and daily time scales. Precipitation should be a difficult quantity to compare with at these time scales. The Climate

Prediction Center (CPC) provides 1/4 degree daily gridded gauge only precipitation data set for the United States (Shi et al. 2003). We evaluate the daily time series of precipitation over the Mississippi River basin and sub-basins for the period of Jan 2003 to Dec 2004.

Figure 7 shows the Mississippi River Basin (MRB) on the MAC grid, which we will focus this evaluation. Five sub-basins contribute to the MRB drainage into the Gulf of Mexico. The gauge data were box-averaged up to the MAC grid for comparison purposes. For each daily basin average of the CPC observations over two years, the data are compared to the corresponding daily basin average of each model and the ensemble. Figure 8 shows the scatter diagram of the two years of daily time series data. It seems quite remarkable that the MAC ensemble average is clustered so closely to the observations at daily time scales. In some models, clear biases are apparent, positive (ECPC RII, JMA, NCEP) and negative (ECPC SFM). Most individual models have noticeable scatter in the data points. BMRC, ECPC RII and ECPC SFM tend to be some of the coarser source data sets (GET a table of resolution of the source data), which may have some influence on their results.

T1 shows the statistics of the area averaged daily time series for the MRB and each of the sub basins. The MAC ensemble bias tends to be small, but positive, following the consensus of the ensemble members. Note that the MAC area weighted bias cannot be computed by linear averaging the area weighted bias of each member, in part because of missing data in the source data (see the Appendix). Even in the smaller sub-basins, the MAC ensemble time series has some of the highest temporal correlation to the observations, and the standard deviation of the time series differences are always the lowest. These temporal statistics suggest that the MAC ensemble precipitation is generally consistent with the weather scale observations.

To pursue this further, we computed spatial statistics (correlation and standard deviation of the difference from the gauge observations across the domain) for the MRB each day of the

2003-2004 period. Figure 9 shows the monthly average of the daily spatial statistics. In the standard deviation of the difference, the MAC ensemble is nearly always the lowest value (low indicates small squared difference from the observations). This is most apparent in the warm season when the errors are generally larger than other seasons. In the winter seasons, the analyses start to group together and their values are generally smaller, but the ensemble still tends to be the smallest (or nearly the smallest) error.

In spatial correlation, there is more separation among the different analyses throughout the annual cycle (Figure 9b). Some of the analyses that perform better for the whole period (see T1) are closer to the high values of the MAC. In some months, the MAC ensemble is not the highest spatial correlation. These are generally the winter months. In the winter, dynamics and initial conditions of the data analysis provide skillful forecasts that allow more accurate precipitation. In the summer, physical processes and mesoscale systems govern the observed precipitation, the forecasts and analyses have more uncertainty. While this is not a surprising result, the fact that the MAC ensemble has noticeably more skill at the daily spatial statistics than the member analyses during the summer is a very interesting result.

Figure 10 and F8 compare two individual days, one summer (10 July 2003) and one winter (23 Dec 2004), precipitation from the MAC to observations. These two days were chosen because they exhibit some of the largest precipitation amounts when averaged over the MRB. For 10 July 2003, the primary maximum of precipitation is reasonably well located in the central United States, but the MAC Ensemble contours do not show the detailed structure apparent in the observations, and also underestimates the intensity at the core of the event. UKMO is chosen for as a member of the ensemble, with reasonable statistics. In the July case, UKMO does produce a larger amount in the core, but misses the southward extent of the event. The standard deviation of the daily MAC precipitation resembles the mean with the largest values near the core of the rain event. Statistically speaking, the MAC data has a higher correlation to and smaller variance than the UKMO data for the MRB, but not without some deficiency.

In the December example (F8a), a strong frontal system extends across the United States, west of the Appalachian Mountain chain. A secondary maximum of precipitation is evident in the southeastern United States. The MAC ensemble seems to locate the main frontal precipitation well, but the width of the core is wider than apparent in the gauge observations. The UKMO has more precipitation than observed along the southern extent of the core. The MAC ensemble has little resemblance to the secondary south eastern maxima, while the UKMO system does have more precipitation there (though no closed contours evident). Even with these apparent differences, there is little statistical difference between the MAC ensemble and the UKMO data, for this case. The standard deviations of the ensemble precipitation are generally related to the occurrences of precipitation and the systems that generate the precipitation.

Roads and Betts (2000) compared the ECMWF and NCEP reanalysis water budgets over the MRB. All components of the regional water budgets had differences between them. The level of that evaluation cannot be duplicated with the present analyses as not all data are available (notably, a soil moisture storage diagnostics among the systems is not uniform, total runoff/drainage water would also be required). Yet, the results at the synoptic time scale suggest that much may be learned about regional water budgets from an ensemble of analyses approach. The evaluation here does not address the diurnal cycle of precipitation in analyses. Ruane and Roads (2007 a&b) evaluate reanalyses diurnal cycles globally and regionally, finding significant deficiencies in both the phase and amplitude of diurnal precipitation. While 6 hourly data are available in this multi-model data set, only 27 months may limit certain statistics.

*d. Global Energy Budget*

Nearly all of the systems provide the components to evaluate the Earth's global energy budget components. Recently, Trenberth et al. (2008, henceforth TFK) revised assessments of the energy budget components based on newly available data and models. The assessment utilized GPCP, International Satellite Cloud Comparison Project (ISCCP) and Clouds and the Earth's Radiant Energy System (CERES) observations. The TFK period of interest was March 2000 to May 2004. We compare the 2 year (Jan 2003-Dec 2004) globally averaged energy budget from the ensemble members and mean to the representative values developed by TFK.

Table 3 shows the major components of the Earth's global energy budgets at the surface and top of the atmosphere (TOA), as well as precipitation (representative of atmospheric latent heat). At the top of the atmosphere, all models use much the same solar forcing. The analyses tend to overestimate OLR and underestimate reflected shortwave compared to the TFK estimates, which suggests that the clouds or the effect of clouds on the radiation is underestimated in the ensemble of analyses. The surface latent heating in the ensemble members is generally higher than the TFK estimate, and it would appear that this is driven by too much downward shortwave radiation at the surface. However, the ensemble average downward and upward longwave radiations at the surface exceed the TFK estimates by roughly the same amounts. Both have smaller biases than the surface latent heat and downward shortwave radiation. The downward longwave radiation at the surface has a substantial amount of variability while upward longwave radiation is consistent, related to the use of prescribed SSTs. TFK estimate the net heating of the global surface at 1.2 W $m^{-2}$. The variability of net surface heating across each of the member analyses is substantial, ranging from strong warming to strong cooling of the surface. However, the MAC ensemble average net surface heating is comparable to the TFK estimate.

## 4. Summary and conclusions

The Multi-model Analysis for CEOP (MAC) comprises 8 operational global analyses as well as their ensemble mean and standard deviation. The method to unify the data structures in space and time has been discussed here. The goal of the project is to simplify the comparisons of the analyses with existing observations, as well as compare among the different analyses. We hypothesized, based on the results of several previous studies using multiple model simulation results, that the ensemble mean of the analyses should provide data that is as skillful as or better than the most skillful contributing member. This is founded on the use of similar observations in each of the analyses, so that random errors in each system will be minimized when ensemble averaged. We focus the testing of the data on precipitation and outgoing longwave radiation, variables that are predominantly driven by the model physics, but also have reliable global observations data sets for verification.

At monthly time scales, we find that the statistics produced by the ensemble of the analyses are similar to or better than the best contributing member. This is generally true for the duration of the period, though in OLR, the NCEP analysis shows slightly better results compared to the full ensemble during boreal winter. The global precipitation of the ensemble mean is much better than any of the members. Comparing the members and ensemble to global energy budget estimeate (from Trenberth et al 2008) shows where the analyses are similar and where they vary. While the global net surface energy from any given system may show imbalance, the ensemble net surface energy is inline with the estimate. The ensemble also show the best or nearly so temporal and spatial statistics compared to daily gauge observations in the well instrumented Mississippi River Basin, even to the level of the sub-basins.

The impact of selectively choosing the most skillful ensemble members to create a new ensemble is also tested. Generally speaking, the improvement that is gained by choosing the most skillful members is small, and may be exaggerated in the present tests given a fairly small number of members. In one case, where there is a clear separation in skill between two sets of analyses, the selective ensemble did produce noticeably smaller error. However, even that improvement is relatively small when comparing the range of error in the members that were not included in the selective ensemble.

While the ensemble mean of the analyses does compare well with the observations presented here, it does so by retaining the information in each analysis that is correlated. The higher skill of the ensemble average indicates that the random or system-specific errors in the contributing members are minimized. However, biases that are correlated, such as high tropical precipitation, high global incoming shortwave radiation at the surface and high surface evaporation, are retained in the ensemble average. Presumably, as the systems improve, and these systematic biases are reduced, the ensemble of analyses would converge to a less biased depiction of reality before any one analysis might. It is also interesting to note that the large monthly variations of standard deviation or correlation in some of the members are minimized in the ensemble mean. Also, trends in the statistical comparison of individual analyses to the observations (e.g. apparent in the JMA OLR) do not appear to be reflected in the MAC ensemble statistics.

Long retrospective-analyses were developed to address the issue of the changing modeling and data assimilation systems in operational analyses, so that climate studies could be undertaken (Bengtsson and Shukla 1988; Trenberth and Olson, 1988). In this short time series, we see that what would be major changes to any individual system is minimized in the ensemble mean so that that a long consistent climate record might be formed through an ensemble of analyses (which could also include reanalyses). However, there were no major changes to the operational observing system during this period of investigation. Reanalyses are inherently resource intensive projects, so that only a small number have been completed (Kalnay et al. 1996; Uppala et al. 2005; Onogi et al, 2007). New reanalyses are being prepared and should continue well into the future. The notion of a multi-model ensemble of operational analyses would allow the large number of meteorological agencies worldwide to contribute their data to a climate record, regardless of variations in the system being used. Given that large numbers of models are already being contributed to support IPCC projections, systems to handle these data already exist. To accomplish such a repository would also require investment of time and resources from already over burdened numerical whether prediction centers. However, the results presented here suggest that the investment would be useful to both the research community, and the contributing centers in their own system development efforts.

*Acknowledgments*

This work would not have been possible without the contributions of data and time from the collaborating modeling centers. Petra Koudelova and Sam Benedict provided the organizational support which kept the CEOP model group to a schedule. This project was supported by the NASA Modeling, Analysis and Prediction program. The data is available for download from the NASA Goddard Data Information Services Center (DISC) and the Max-Planck Institute Model and Data Center.

## 5. Appendix

In Section 2, seven steps of the process to reformat the data are outlined. The following sections provide further details on the procedures and methods.

*a. 6-hourly dataset (Step 1)*

For each NWPC dataset, a Grib table was used to identify and locate the subset of high-priority variables listed in the Table A1. The minimum forecast time available for each variable of the center was then pulled from the raw model data using "wgrib". The minimum forecast time available typically was the analysis (0-hour forecast) for the instantaneous variables, and the 0-6 hourly forecast for the average/accumulation (ave/acc) variables. Some major exceptions include the CPTEC data, the MSC data, and some variables from the ECPC data. The CPTEC data at 00Z is a 12-hourly forecast, at 06Z an 18-hr forecast, at 12Z a 24-hr forecast, and at 18Z a 36-hr forecast. Similarly, the average and accumulated (ave/acc) variables from 00Z to 06Z are a 12-18 hourly forecast, and so on. The MSC instantaneous surface variable data at 12Z is an analysis/0-hr forecast, at 18Z a 6-hr forecast, at 00Z a 12-hr forecast, and at 06Z an 18-hr forecast. The upper air data, however, was not available at 06Z and 18Z; the 12Z data is an analysis/0-hr forecast and the 00Z data is a 12-hr forecast. The MSC ave/acc variables from 12Z to 18Z are a 0-6 hourly forecast, and so on. Several ECPC RII and SFM instantaneous surface variables are a 6-hr forecast rather than an analysis/0-hr forecast. Further details, including descriptions of the forecast times and missing variables are included in Tables A1 and A2.

*b. Interpolation (Step 2)*

In order to compare and produce an ensemble, a common grid must be defined. Since most operational analyses are near or going to ~100km spatial scales, a grid on the order of 1 degrees latitude and longitude was desirable. Also, many data products (GPCP and the existing climate reanalyses data) use a regular latitude-longitude coarse grid (2.5 degrees). Thus, a regular latitude-longitude grid that is near the spatial scale of the observational analyses, but also can be related easily to the reanalyses coarse grid, was chosen. The resolution is 1.25º longitude by 1.25º latitude (288×144 gridpoints), with the 1,1 center point located at 179.375W, 89.375S.

The native grid from each of the NWPCs supplied to CEOP was interpolated to the common grid using a freely available routine (http://opengrads.org/ re() funnction). In the cases where the native grid is finer than 1.25ºx1.25º, box averaging was used. In the cases where the native grid is coarser than 1.25ºx1.25º, bilinear interpolation was be used. No other filtering or screening of the gridded data was applied (except for some below-ground heights – details in Sections 6c-f). At the end of this step, the data from each NWPC was on the common grid at a 6-hourly time interval, with common variable names and units. A list of the available variables for each center can be found in Table A1.

*c. Ensemble Average (Step 3)*

The ensemble average is the straight average of all of the available variables from each NWPC at each 6-hourly time interval. As all centers did not provide all variables, the ensemble averaging was done with those centers that did provide the given variable. If any data was missing from one or more of the NWPCs at a given time, the ensemble average was the average of the remaining data available. For the upper-air data at 850hPa and 700hPa, a masking to the MAC ensemble was applied for areas where the surface pressure at the given time was less than the pressure of the level (less than 850hPa or than 700hPa). This masking was also performed for the BMRC 6-hourly data, but not for the other individual NWPCs. The flowchart decisions used for each variable during the creation of the MAC ensemble at each of the 3292 6-hourly times is shown in Figure A1.

The individual center's interpolated variable is also provided with the MAC, so that it will be apparent when data is included in the ensemble average or not. Also, a separate dataset is provided that enumerates the number of ensemble members for each variable for each time. Similarly, the standard deviation at each 6-hourly timestep was computed from the available data that made up the ensemble average. The ensemble mean and standard deviation are provided as separate datasets on the same grid and same format as the individual NWPCs described at the end of Section 3b (Step 2).

*d. Daily- and Monthly-Averages (Steps 4, 5, and 6)*

The daily average of the ensemble mean was the simple average of the 00Z, 06Z, 12Z, and 18Z data on the given date. For the individual NWPCs, the daily average was the same, except that if an individual variable was missing or unavailable for at least one time during the date, that variable was considered to be undefined for that center on that day. The one exception to this is the MSC upper-air data, which was only available at 00Z and 12Z, and the daily average is just the average of these two times. Also, for each dataset, if at least one of the four times of the day had a point masked out because the surface pressure was less than the pressure of the upper-air level, then that point was also masked out for the entire day. The flowchart for the daily averages is shown in Figure A2. The daily standard deviation was then calculated between the centers that had valid daily averages for each variable. Note that the daily ensemble mean may include more data/centers than the daily ensemble standard deviation. An example of this is to suppose the 500 hPa heights was missing for 12Z only for one center. The 6-hourly ensemble means will include the 00Z, 06Z, and 18Z times for this center, and thus the daily ensemble mean will proportionally include this data. However, the daily mean for this center/variable will be considered undefined, and will not be included in the daily ensemble mean.

The monthly average of the ensemble mean was the simple average of all times in the month. For the individual NWPCs, the monthly average was calculated differently. First, all the 00Z times during the month were averaged, then the 06Z times, the 12Z times, and then the 18Z times. Next, these four times were summed and divided by four (4). This method was done to minimize the effect of an individual missing time on the monthly average. For example, if a 06Z time was missing for a variable such as downward surface radiation on a single date, this missing time would have a noticeable effect on the monthly average. If the similar times were averaged first, this problem is reduced, but does give a little extra weight to the other dates where the variable was available. No more than 6 times during the month were allowed to be undefined (out of a typical 120 or 124 6-hourly periods). If more than 6 times were undefined, the variable for that month was undefined. Similarly, if a given point had more than 6 times masked because the surface pressure was less than the pressure of the upper-air level, the point was also masked. The exceptions to this were for the UKMO data (numerous missing times), for the CPTEC data (only for May 2003, due to missing data), and for the MSC data (only 00Z and 12Z data available). The flowchart for the monthly averages is shown in Figure A3. The monthly-average standard deviation was then calculated between the individual centers' monthly averages. Again, because of the different methods of the monthly-average calculations, the monthly-average standard deviation will not be exactly centered about the ensemble mean monthly-average.

*e. Write the gridded data out for the MAC (Step 7)*

Data were written to binary output, and then converted to the NetCDF and Grib1 formats for release to the contributors and community. The resulting binary (or NetCDF) output size is roughly 284Gb (about 134Gb in Grib1). Each file contains each variable listed in Table A1 (with the common naming convention). Common utilities, ncdump and wgrib, can be used to identify the vital information needed to access the data. A grib table common to all processed centers and the MAC is also provided. The data will be sent to the NASA Goddard Data Information Services Center (DISC) and the Max-Planck Institute Model and Data Center.

## 6. References


Adler, R. F., G. J. Huffman, GJ; A. Chang, and co-authors., 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J. Hydromet*., **4**, 1147-1167.

Bengtsson, L. and J. Shukla, 1988: Integration of space and in situ observations to study global climate change, *Bull. Amer. Meteor. Soc.,* **69**, 1130- 1143.

Bosilovich, M. G., and R. Lawford, 2002: Report on the Coordinated Enhanced Observing Period (CEOP) International Workshop, *Bull. Amer. Met. Soc*., **83**, 1495-1499.

Bosilovich, M. G., J. D. Radakovich, A. da Silva, R. Todling and F. Verter, 2007: Skin Temperature Analysis and Bias Correction in a Coupled Land-Atmosphere Data Assimilation System. *JMSJ*, **85A**, 205-228.

Bosilovich, M. G., J. Chen, F. R. Robertson and R. F. Adler, 2008: Evaluation of Global Precipitation in Reanalyses. *J. Appl. Meteor. Clim*. In Press.

Chou, S. C., J. A. Marengo, C. P. Dereczynski, P. V. Waldheim and A. O. Manzi, 2007: Comparison of CPTEC GCM and Eta Model Results with Observational Data from the Rondonia LBA Reference Site, Brazil. *JMSJ*, **85A**, 25-42.

Compo, G. P., J. S.Whittaker, and P. Shardeshmukh, 2006: Feasibility of a 100-Year Reanalysis using only surface pressure data, *Bull. Amer. Met. Soc*., DOI:10.1175/BAMS-87-2-175. 175-190.

Dirmeyer, P. A. and co-authors, 2006: Multi-model Analysis and Implications for our perception of the land surface, *Bull. Amer. Met. Soc*., **87**, 1381-1395.

Guo, Z., P. A. Dirmeyer, X. Gao and M. Zhao, 2007: Improving the quality of simulated soil moisture with a multi-model ensemble approach. *Q. J. R. Meteorol. Soc.*, **133**, 731–747.

Hirai, M., T. Sakashita, H. Kitagawa, T. Tsuyuki, M. Hosaka and M. Oh'Izumi, 2007: Development and Validation of a New Land Surface Model for JMA's Operational Global Model Using the CEOP Observation Dataset. *JMSJ*, **85A**, 1-24.

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

Kato, H., M. Rodell, F. Beyrich, H. Cleugh, E. van Gorsel, H. Liu and T. P. Meyers, 2007: Sensitivity of Land Surface Simulations to Model Physics, Land Characteristics, and Forcings, at Four CEOP Sites. *JMSJ*, **85A**, 187-204.

Koike, T., 2004: The Coordinated Enhanced Observing Period (CEOP): an initial step for integrated global water cycle observation. WMO Bull., **53**, 115-121.

Krishnamurti, T. N., and co-authors, 2003: Real-time multianalysis-multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. *Mon. Wea. Rev*., **129**, 2861-2883.

Meinke, I., J. Roads and M. Kanamitsu, 2007: Evaluation of RSM-Simulated Precipitation During CEOP. *JMSJ*, **85A**, 145-166.

Milton, S. F. and P. Earnshaw, 2007: Evaluation of Surface Water and Energy Cycles in the Met Office Global NWP Model Using CEOP Data. *JMSJ*, **85A**, 43-72.

Onogi, K. and co-authors, 2007: The JRA-25 Reanalysis. *J. Met. Soc. Japan*, **85**, 369-432.

Phillips, T. J., and P. J. Gleckler, 2006: Evaluation of continental precipitation in 20th century climate simulations: The utility of multimodel statistics. *Water Resour. Res.*, **42**, W03202, doi:10.1029/2005WR004313.

Rikus, L., 2007: Validating Basic Surface Variables in the Australian Bureau of Meteorology Model with CEOP EOP3 In-situ Data. *JMSJ*, **85A**, 73-97.

Roads, J. O. and A. K. Betts, 2000: NCEP-NCAR and ECMWF reanalysis surface water and energy budgets for the Mississippi River basin. *J. Hydromet*., **1**, 88-94.

Ruane, A. C. and J. O. Roads, 2007a: 6-hour to 1-year variance of five global precipitation sets. *Earth Interactions*, **11**, 1-29, DOI: 10.1175/EI225.1.

Ruane, A. C. and J. O. Roads, 2007b: The diurnal cycle of water and energy over the Continental United States from three reanalyses. *J. Met. Soc. Japan*, **85A**, 117-143.

Shi, W., R. W. Higgins, E. Yarosh and R. Joyce, 2003: Processing Daily Rain-Gauge Precipitation Data For the Americas at the NOAA Climate Prediction Center. 19th International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology. Long Beach, CA, 9-13 Feb, 2003.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res*., **106**, D7, 7183-7192.

Trenberth, K. E., and J. G. Olson, 1988: An evaluation and intercomparison of global analyses from NMC and ECMWF. *Bull. Amer. Meteor. Soc*., **69**, 1047-1057.

Trenberth, K., E. L. Smith, T. Qian, A. Dai, and J. Fasullo, 2007: Estimates of the global water budget and its annual cycle using observational and model data. *J. Hydromet*., **8**, 758-769.

Trenberth, K. E., J. T. Fasullo and J. Kiehl, 2008: Earth’s global energy budget. Under revision for the *Bull. Amer. Met. Soc*.

Uppala, S. M., 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.

Yang, K. and co-authors, 2007: Initial CEOP-based Review of the Prediction Skill of Operational General Circulation Models and Land Surface Models. *JMSJ*, **85A**, 99-116.

Yin, X. G., A. Gruber, and P. Arkin, 2004: Comparison of the GPCP and CMAP merged gauge satellite monthly precipitation products for the period 1979-2001. *J. Hydromet*. **5**, 1207-1222.

## 7. List of Tables

## 8. Tables

Table 1 Statistics from the 2 year area average time series of daily precipitation compared to gauge observations. a) the mean difference from observations (mm day$^{-1}$), b) the standard deviation of the difference from observations (mm day$^{-1}$) and c) temporal correlation to the observed time series.

| a) Mean Bias | MRB | Red Ark | Missouri | Upper Miss | Ohio | Lower Miss |
|---|---|---|---|---|---|---|
| BMRC | -0.29 | -0.77 | -0.09 | -0.26 | 0.14 | -0.85 |
| CPTEC | 0.03 | -0.01 | 0.19 | 0.01 | 0.22 | -0.80 |
| ECPCRII | 0.66 | 0.63 | 0.61 | 0.50 | 0.77 | 0.92 |
| ECPCSFM | -1.02 | -1.31 | -0.81 | -0.98 | -0.77 | -1.66 |
| JMA | 0.57 | 0.83 | 0.34 | 0.39 | 0.64 | 1.16 |
| MSC | 0.11 | 0.12 | -0.02 | 0.22 | 0.21 | 0.26 |
| NCEP | 0.70 | 0.48 | 0.50 | 0.58 | 1.30 | 1.07 |
| UKMO | 0.46 | 0.38 | 0.45 | 0.37 | 0.50 | 0.72 |
| MAC | 0.15 | 0.04 | 0.14 | 0.10 | 0.38 | 0.10 |

| b) Std Dev | MRB | Red Ark | Missouri | Upper Miss | Ohio | Lower Miss |
|---|---|---|---|---|---|---|
| BMRC | 1.16 | 2.04 | 1.03 | 2.45 | 2.55 | 3.57 |
| CPTEC | 0.91 | 1.92 | 1.05 | 1.78 | 2.47 | 3.25 |
| ECPCRII | 1.35 | 2.43 | 1.46 | 2.21 | 3.03 | 4.07 |
| ECPCSFM | 1.14 | 2.19 | 1.06 | 2.23 | 2.23 | 3.53 |
| JMA | 0.98 | 1.77 | 0.94 | 1.47 | 2.11 | 3.82 |
| MSC | 0.57 | 1.54 | 0.73 | 1.40 | 1.80 | 2.68 |
| NCEP | 0.76 | 1.33 | 0.80 | 1.38 | 1.99 | 2.80 |
| UKMO | 0.65 | 1.37 | 0.83 | 1.34 | 1.59 | 2.44 |
| MAC | 0.52 | 1.11 | 0.54 | 1.14 | 1.47 | 2.32 |

| c) Corr | MRB | Red Ark | Missouri | Upper Miss | Ohio | Lower Miss |
|---|---|---|---|---|---|---|
| BMRC | 0.84 | 0.81 | 0.80 | 0.72 | 0.88 | 0.82 |
| CPTEC | 0.90 | 0.84 | 0.82 | 0.85 | 0.87 | 0.85 |
| ECPCRII | 0.86 | 0.85 | 0.87 | 0.85 | 0.85 | 0.78 |
| ECPCSFM | 0.85 | 0.79 | 0.78 | 0.78 | 0.90 | 0.84 |
| JMA | 0.93 | 0.90 | 0.91 | 0.92 | 0.93 | 0.84 |
| MSC | 0.97 | 0.91 | 0.92 | 0.93 | 0.94 | 0.91 |
| NCEP | 0.95 | 0.94 | 0.93 | 0.93 | 0.95 | 0.90 |
| UKMO | 0.97 | 0.92 | 0.92 | 0.91 | 0.94 | 0.94 |
| MAC | 0.97 | 0.95 | 0.95 | 0.95 | 0.96 | 0.92 |

Table 2 Daily statistics for the MRB precipitation of each analyses and the MAC ensemble compared to gauge observations. For each day, the correlations are on the upper row and the standard deviations are on the lower row.

| | MAC | BMRC | CPTEC | ECPCRII | ECPCSFM | JMA | MSC | NCEP | UKMO |
|---|---|---|---|---|---|---|---|---|---|
| 10JUL2003 | 0.78 | 0.22 | 0.56 | 0.52 | 0.45 | 0.79 | 0.74 | 0.50 | 0.75 |
| | 3.75 | 5.68 | 4.45 | 5.65 | 4.91 | 4.39 | 4.11 | 5.05 | 4.29 |
| 23DEC2004 | 0.91 | 0.84 | 0.82 | 0.90 | 0.88 | 0.84 | 0.90 | 0.93 | 0.92 |
| | 3.22 | 4.34 | 5.48 | 3.84 | 3.82 | 3.46 | 3.97 | 3.09 | 3.39 |

Table 3 Global average of energy budget components for each of the analyses, the ensemble mean (MAC) and the values reported by Trenberth et al. (2008) (TFK). MAC indicates the global average of the gridded ensemble data, not the average of the members' global averages. The standard deviation (Sdev) is the standard deviation of the global averages of the members' values in the table. Units are W $m^{-2}$, except for precipitation (mm $day^{-1}$). $NET_{Sfc}$ indicates the net downward heat flux at the surface.

| Center | LH | SH | $RLd_{sfc}$ | $RLu_{sfc}$ | $RSd_{sfc}$ | $RSu_{sfc}$ | $RLu_{toa}$ | $RSd_{toa}$ | $RSu_{toa}$ | $NET_{Sfc}$ | Precip |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BMRC | 92 | 18 | 350 | 400 | 167 | 25 | | | | -19.3 | 3.32 |
| CPTEC | 99 | 22 | 347 | 401 | 206 | 24 | 244 | | 100 | 7.4 | 3.40 |
| ECPC-RII | 96 | 7 | 337 | 400 | 202 | 24 | 247 | 341.4 | 92 | 11.9 | 3.21 |
| ECPC-SFM | 84 | 17 | 333 | 400 | 207 | 27 | 249 | 341.4 | 91 | 13.0 | 2.47 |
| JMA | 90 | 17 | 320 | 398 | 204 | 25 | 257 | 341.4 | 88 | -6.2 | 3.15 |
| MSC | 91 | 20 | 334 | 397 | 201 | 28 | 249 | 342.0 | 94 | -0.6 | 2.61 |
| NCEP | 95 | 9 | 332 | 398 | 208 | 29 | 248 | | 87 | 8.5 | 3.26 |
| UKMO | 95 | 16 | 345 | 399 | 180 | 22 | 235 | 341.5 | 105 | -7.2 | 3.58 |
| MAC | 92 | 16 | 337 | 399 | 197 | 26 | 247 | 341.5 | 93 | 1.1 | 3.12 |
| Sdev | 4.6 | 5.3 | 9.8 | 1.3 | 15.2 | 2.4 | 6.6 | 0.2 | 6.5 | 11.3 | 0.38 |
| TFK | 80 | 17 | 333 | 396 | 184 | 23 | 238.5 | 341.3 | 101.9 | 1.2 | 2.73 |

## 9. List of Figures

Standard deviation from SRB OLR (c) and Correlation to SRB OLR (d). CMAP precipitation is included on the precipitation panels for reference. Note that UKMO monthly OLR is missing from Dec 2002 and Oct-Dec 2004, leaving only a 3 member selective ensemble in those months.

Figure 7 Five sub-basins of the Mississippi River basin – Red-Arkansas (RA), Missouri (MS), Upper Mississippi (UM), Ohio (OH) and the Lower Mississippi (LM). The outlines also show the MAC 1.25 degree grid. The collective region is referred to as the Mississippi River Basin (MRB).

Figure 8 Scatter diagrams comparing area averaged MRB daily observed precipitation with the MAC ensemble and all of the contributing member data. The solid line indicates the linear fit of the data points and the dashed line indicates 1:1. Units mm day$^{-1}$.

Figure 9 Time series of monthly means of the daily (a) standard deviation of MRB area differences (mm day$^{-1}$) and (b) spatial correlation of the MRB area precipitation from each model system to the gauge observations.

Figure 10 Daily precipitation for 10 July 2003 gauge observations (a), MAC Ensemble (b), MAC – gauge differences (c), UKMO minus MAC differences (d), the UKMO daly mean (e), UKMO minus gauge observations (f), and the standard deviation of the ensemble mean (g),. Units are mm day$^{-1}$.

Figure 11 As in Figure 10, except for 23 December 2004.

**10. Figures**

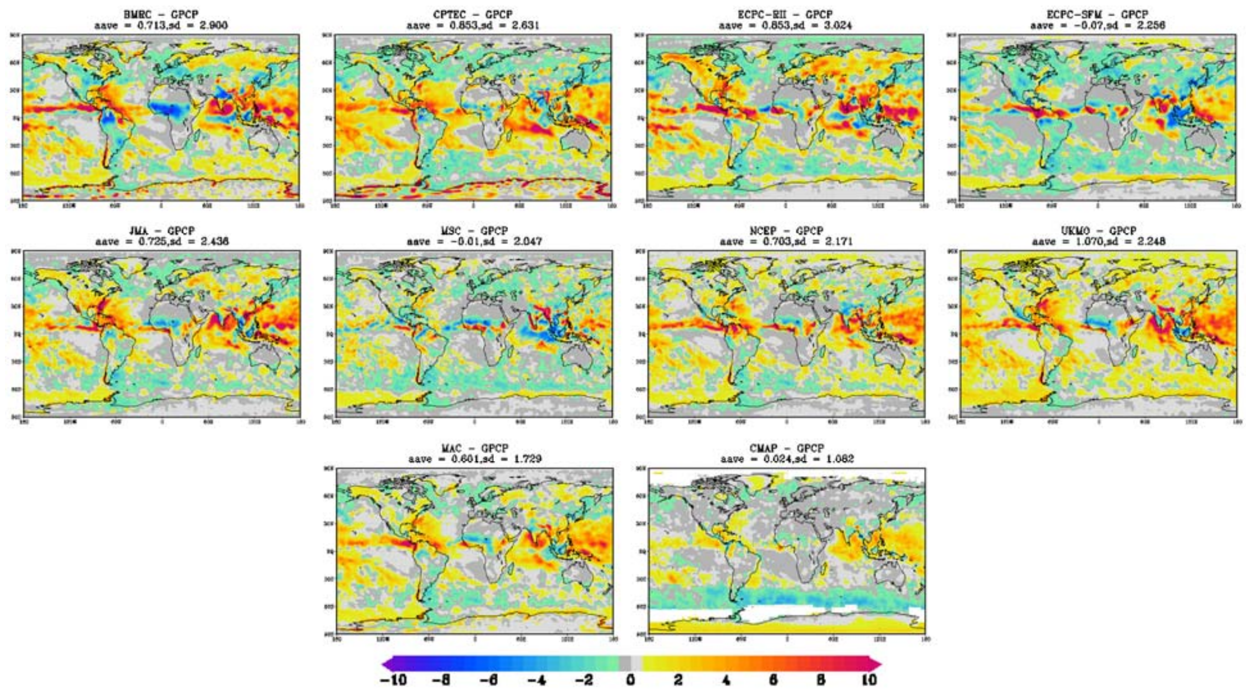



Figure 1 Monthly precipitation differences for each of the ensemble members, the ensemble mean, and CMAP from GPCP merged precipitation data.

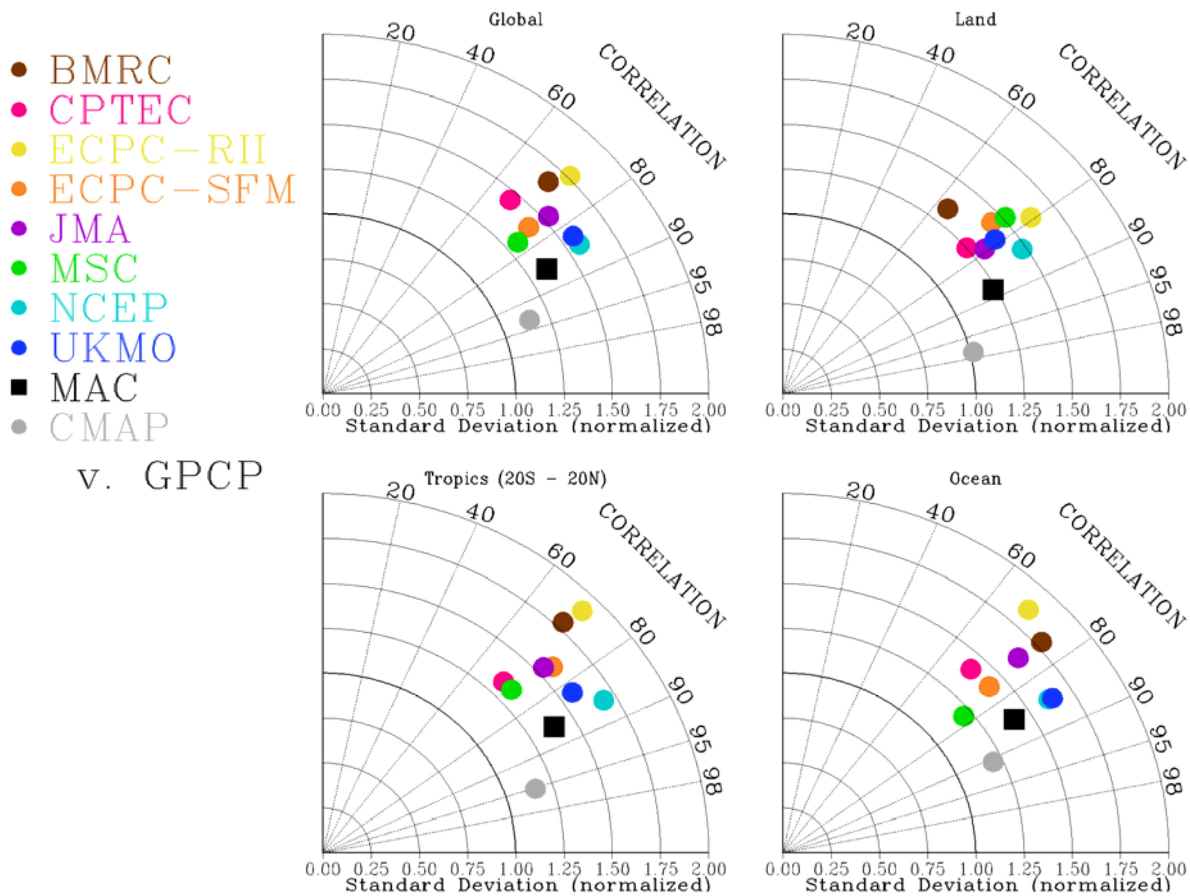



Figure 2 Taylor diagrams (Taylor 2001) for global, global land, global ocean and tropics precipitation. GPCP merged precipitation is the reference data set. The diagrams compare spatial correlation (to GPCP) of the analysis to standard deviation normalized by the reference data set. Linear distance to the 1,1 point is a measure of skill in reproducing the reference data set. CMAP (grey dot) is also provided as a measure of uncertainty in the observations.

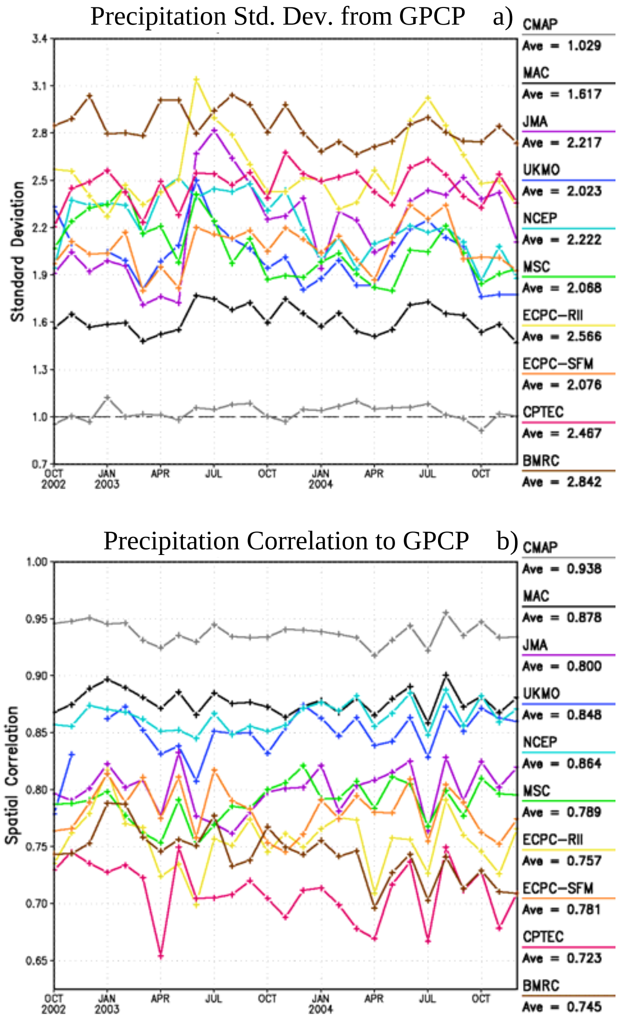



Figure 3 Standard deviation of the monthly global differences from GPCP (a) and correlation to GPCP (b) for the MAC ensemble mean, each ensemble member and CMAP merged observations.

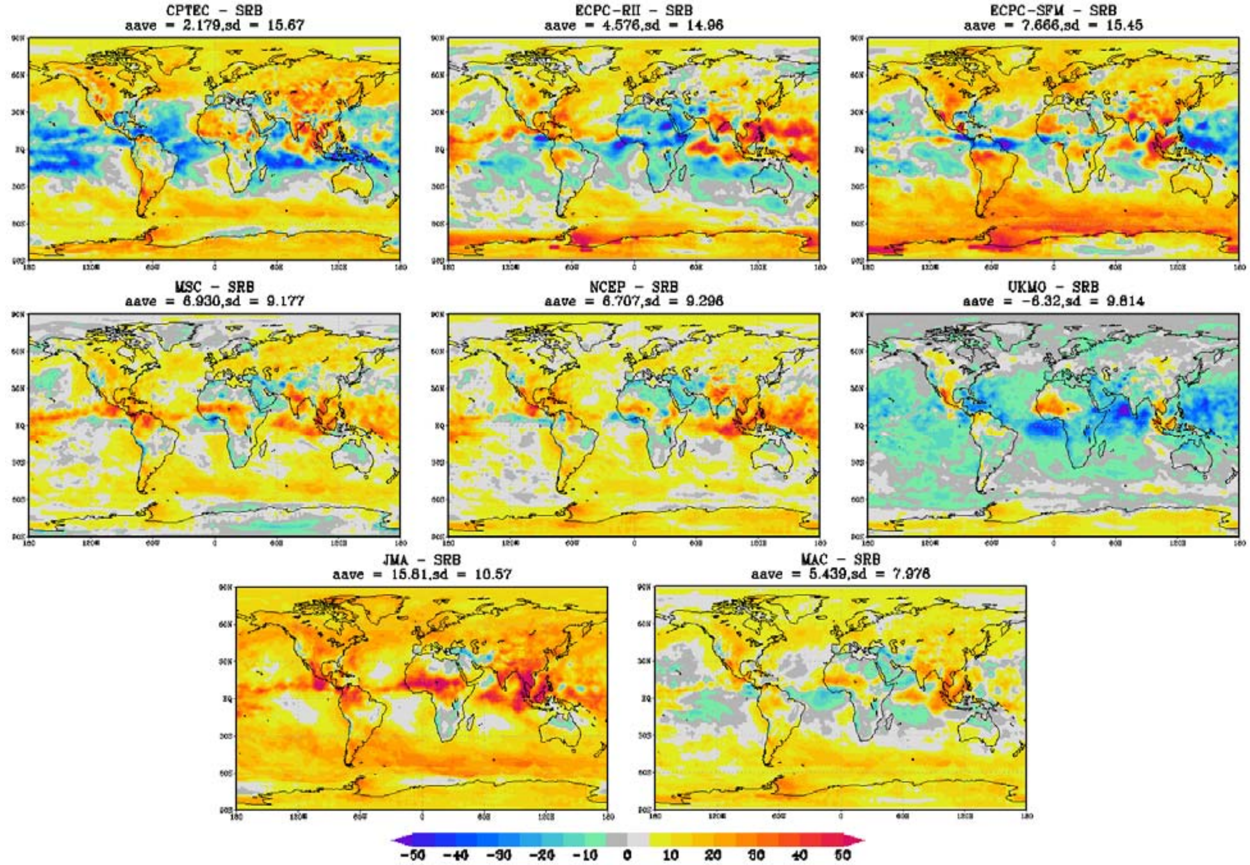
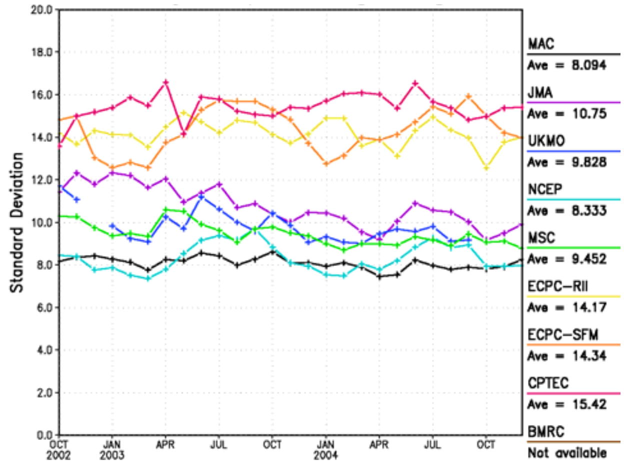



Figure 4 Monthly OLR differences each of the ensemble members and the ensemble mean from Surface Radiation Budget (SRB) OLR (BMRC did not provide OLR). The mean and standard deviation of the differences shown in each map are summarized above the maps.

OLR Standard Deviation from SRB a)




OLR Correlation to SRB b)

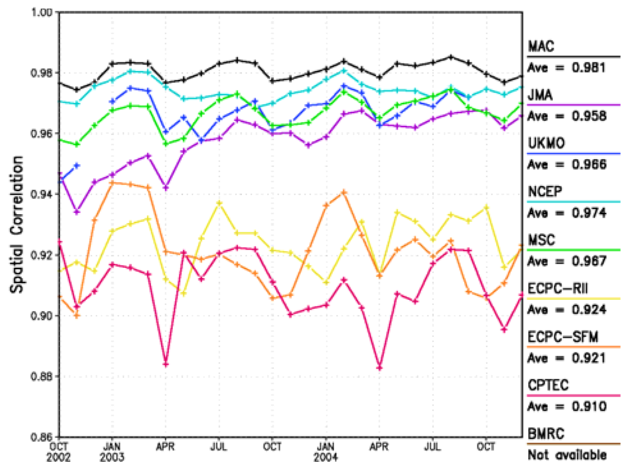



Figure 5 Standard deviation of the monthly global differences from SRB OLR (a) and correlation to SRB OLR (b) for the MAC ensemble mean, each ensemble member.

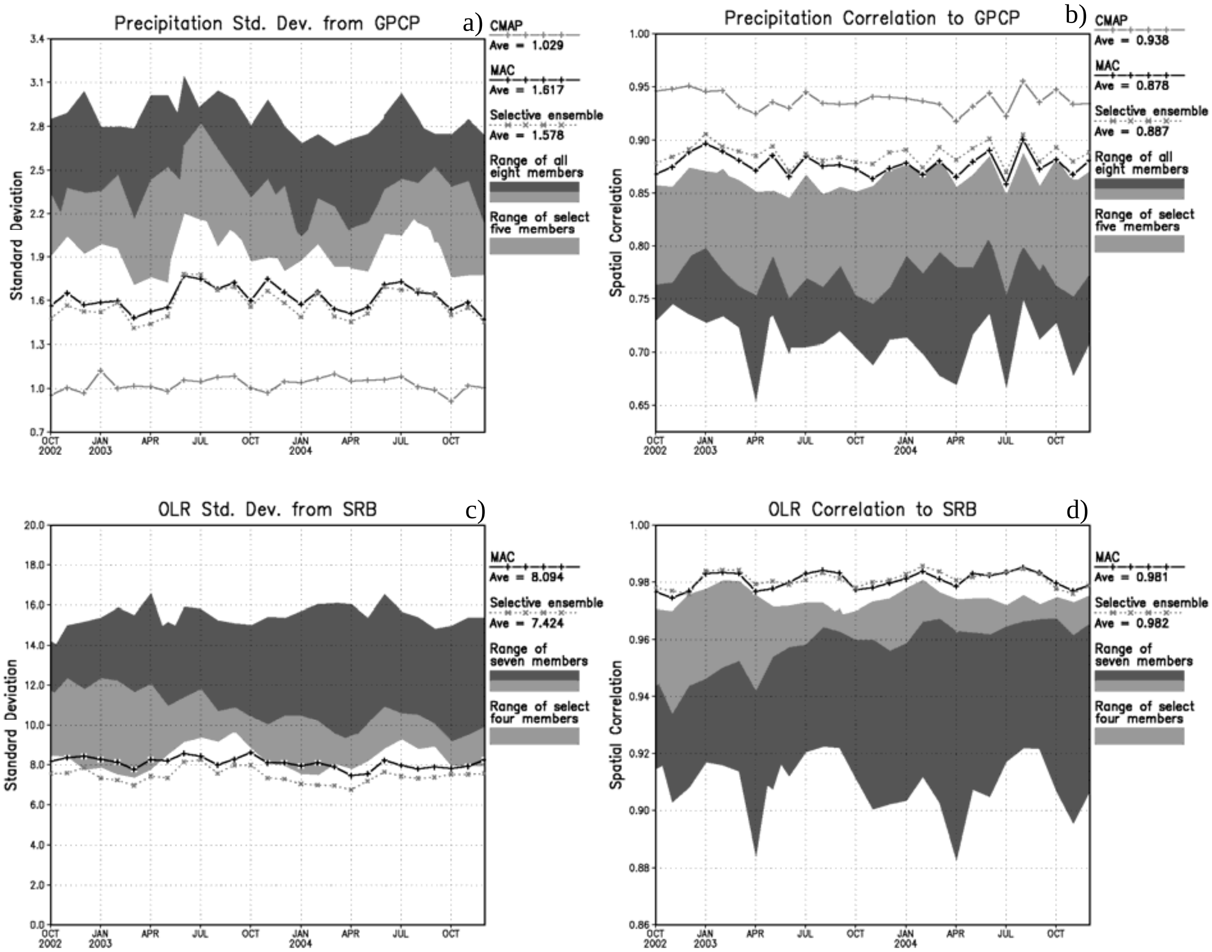



Figure 6 Comparison of the statistics of the MAC ensemble (black line) compared to a selective ensemble (grey dashed line) and the range of values contributing to the full ensemble and the selective ensemble. Standard deviation from GPCP (a), Correlation to GPCP (b), Standard deviation from SRB OLR (c) and Correlation to SRB OLR (d). CMAP precipitation is included on the precipitation panels for reference. Note that UKMO monthly OLR is missing from Dec 2002 and Oct-Dec 2004, leaving only a 3 member selective ensemble in those months.

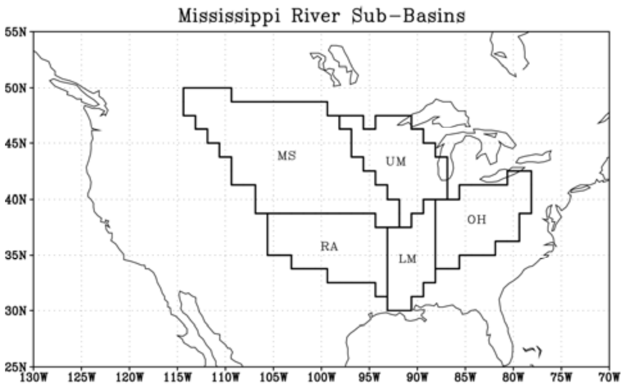



Figure 7 Five sub-basins of the Mississippi River basin – Red-Arkansas (RA), Missouri (MS), Upper Mississippi (UM), Ohio (OH) and the Lower Mississippi (LM). The outlines also show the MAC 1.25 degree grid. The collective region is referred to as the Mississippi River Basin

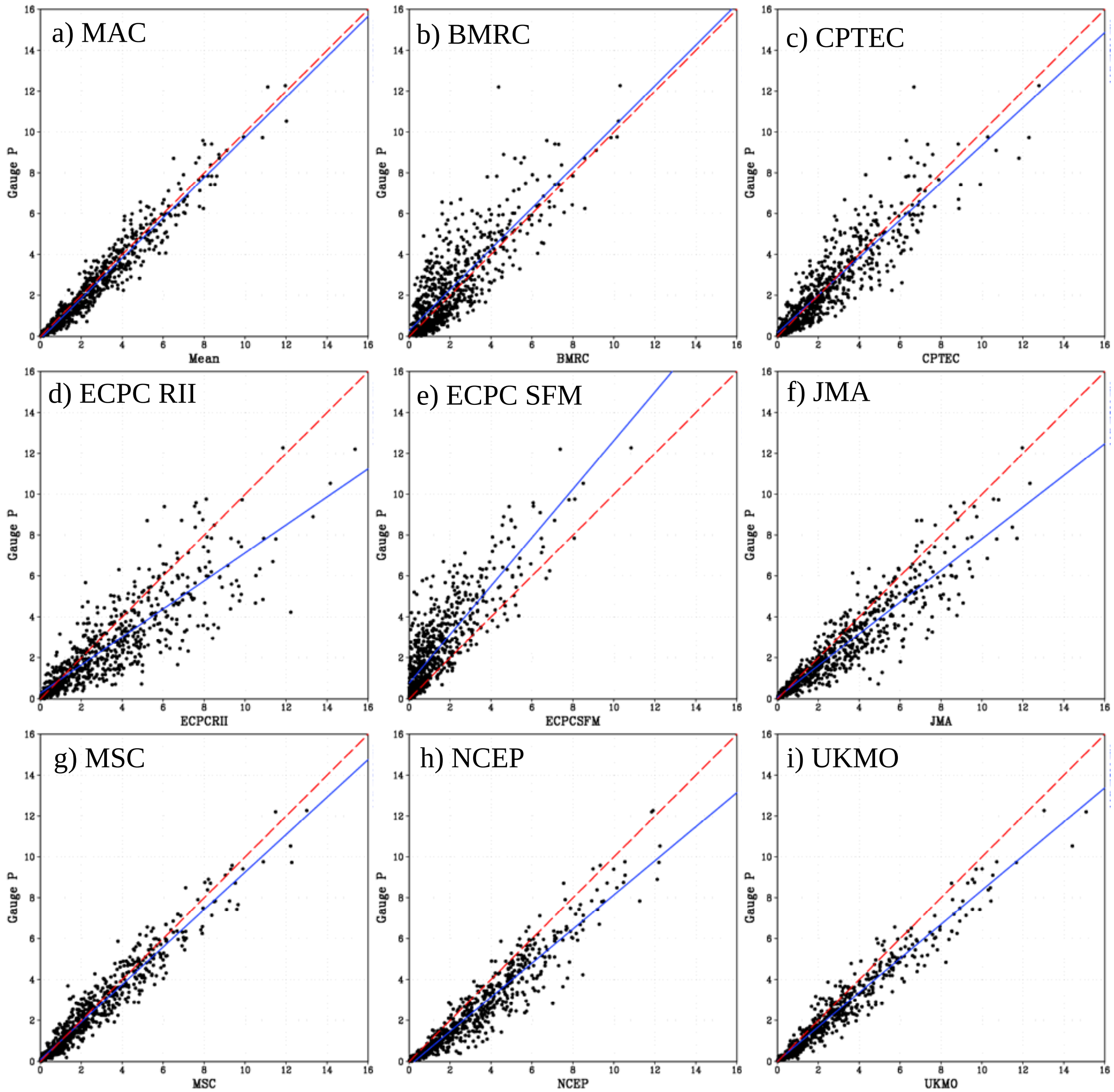
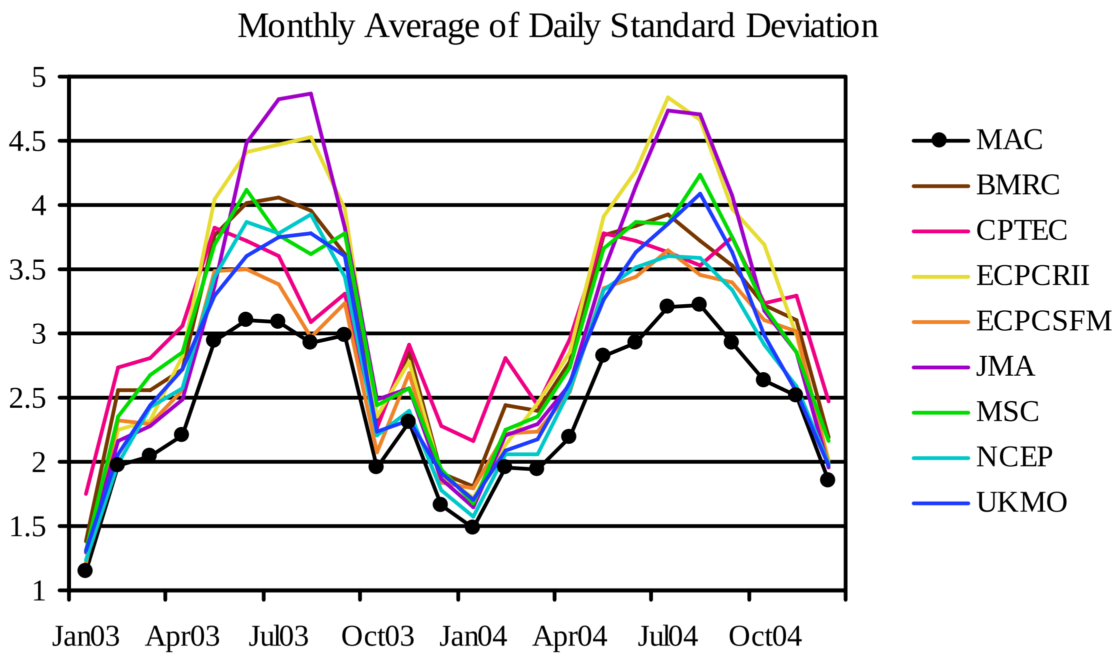



Figure 8 Scatter diagrams comparing area averaged MRB daily observed precipitation with the MAC ensemble and all of the contributing member data. The solid line indicates the linear fit of the data points and the dashed line indicates 1:1. Units mm day$^{-1}$.

a)




b)

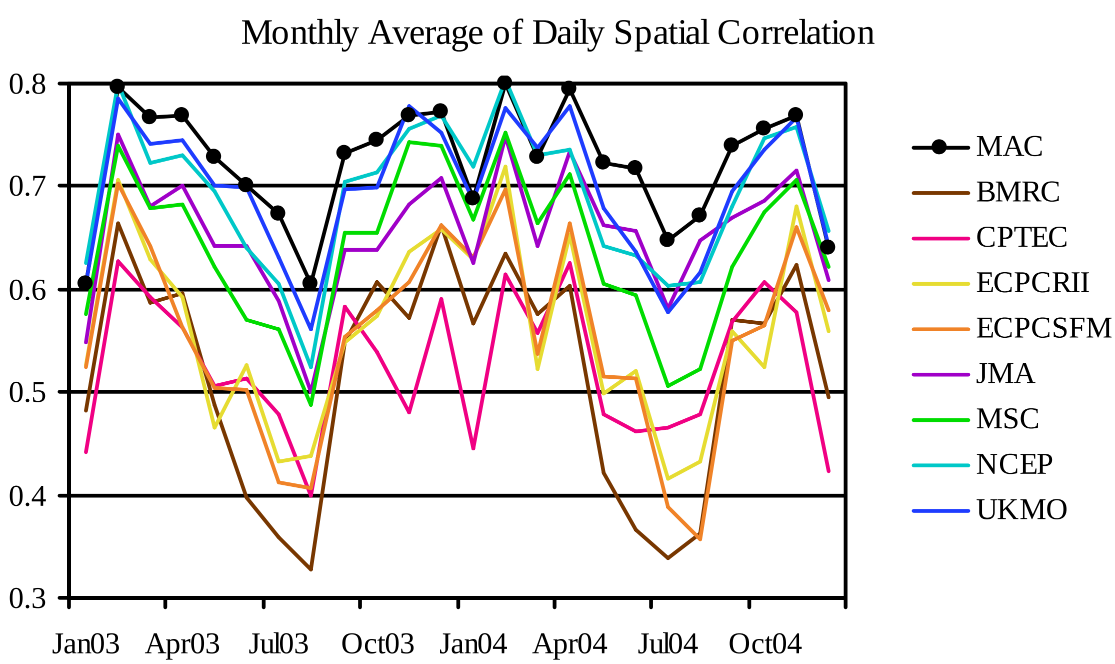



Figure 9 Time series of monthly means of the daily (a) standard deviation of MRB area differences (mm day$^{-1}$) and (b) spatial correlation of the MRB area precipitation from each model system to the gauge observations.

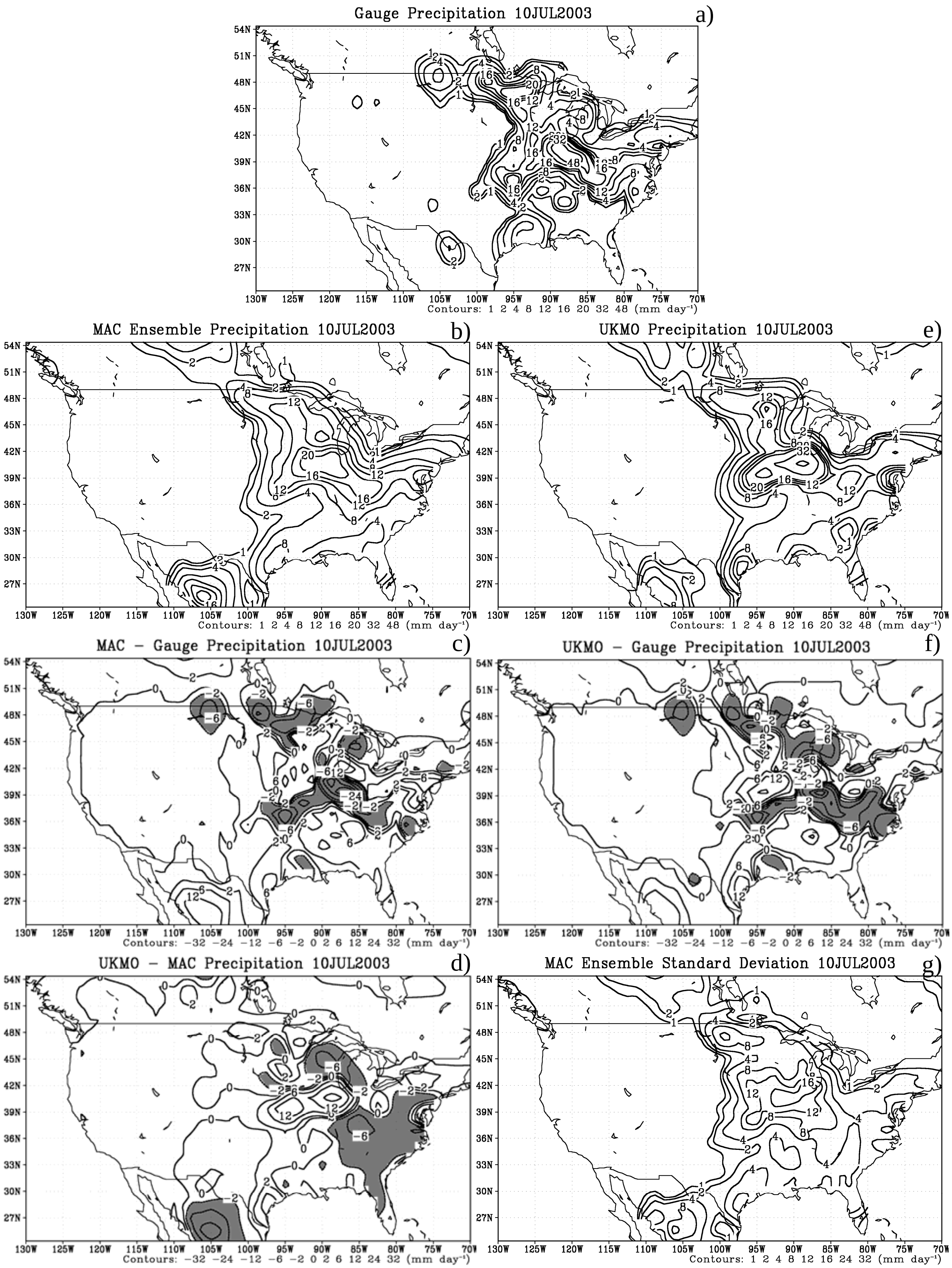



Figure 10 Daily precipitation for 10 July 2003 gauge observations (a), MAC Ensemble (b), MAC – gauge differences (c), UKMO minus MAC differences (d), the UKMO daly mean (e), UKMO

minus gauge observations (f), and the standard deviation of the ensemble mean (g),. Units are mm $day^{-1}$.

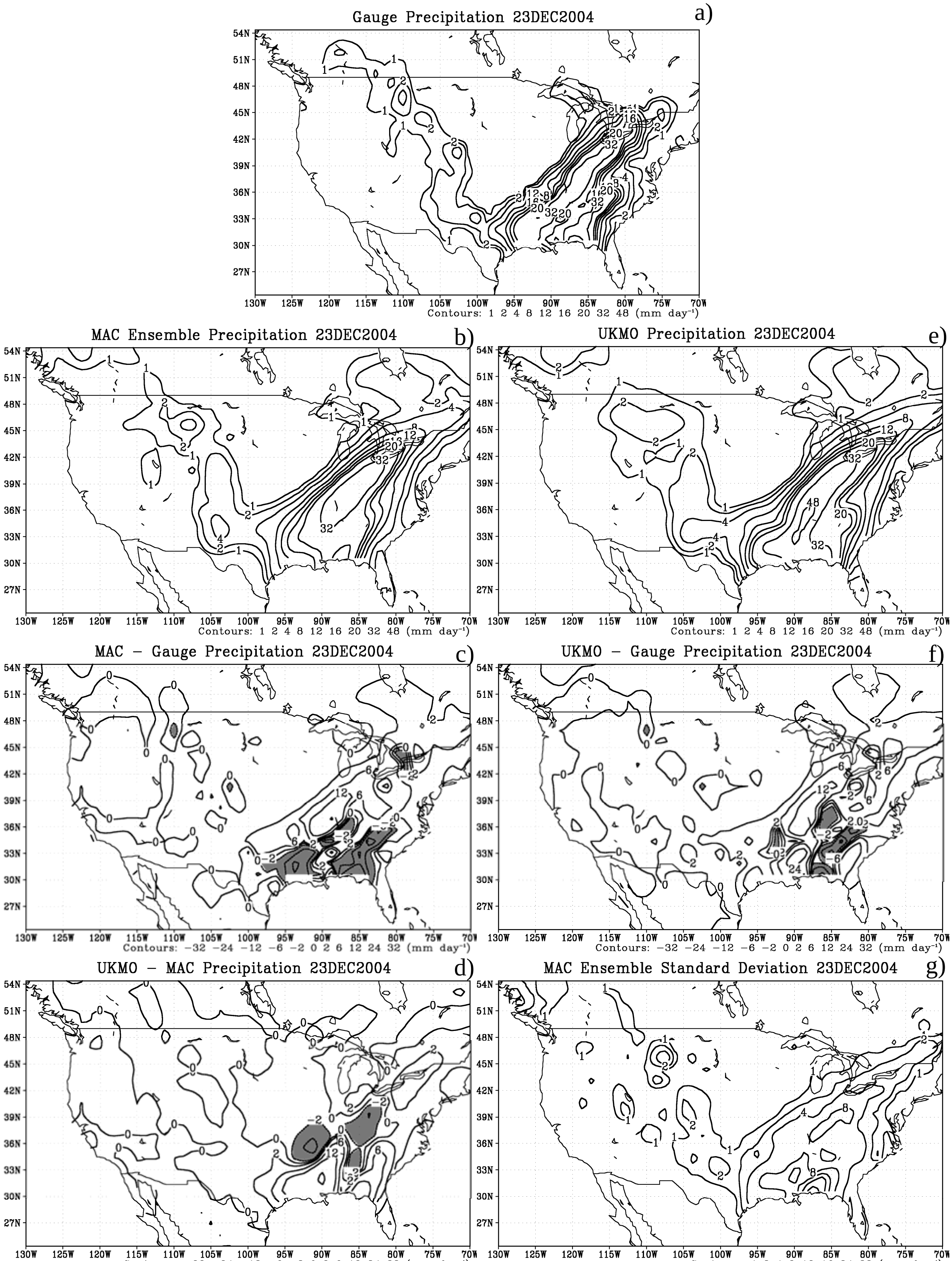

a)
Gauge Precipitation 23DEC2004
Contours: 1 2 4 8 12 16 20 32 48 (mm day⁻¹)
b)
MAC Ensemble Precipitation 23DEC2004
Contours: 1 2 4 8 12 16 20 32 48 (mm day⁻¹)
e)
UKMO Precipitation 23DEC2004
Contours: 1 2 4 8 12 16 20 32 48 (mm day⁻¹)
c)
MAC − Gauge Precipitation 23DEC2004
Contours: −32 −24 −12 −6 −2 0 2 6 12 24 32 (mm day⁻¹)
f)
UKMO − Gauge Precipitation 23DEC2004
Contours: −32 −24 −12 −6 −2 0 2 6 12 24 32 (mm day⁻¹)
d)
UKMO − MAC Precipitation 23DEC2004
Contours: −32 −24 −12 −6 −2 0 2 6 12 24 32 (mm day⁻¹)
g)
MAC Ensemble Standard Deviation 23DEC2004
Contours: 1 2 4 8 12 16 24 32 (mm day⁻¹)
54N 51N 48N 45N 42N 39N 36N 33N 30N 27N
130W 125W 120W 115W 110W 105W 100W 95W 90W 85W 80W 75W 70W

Figure 11

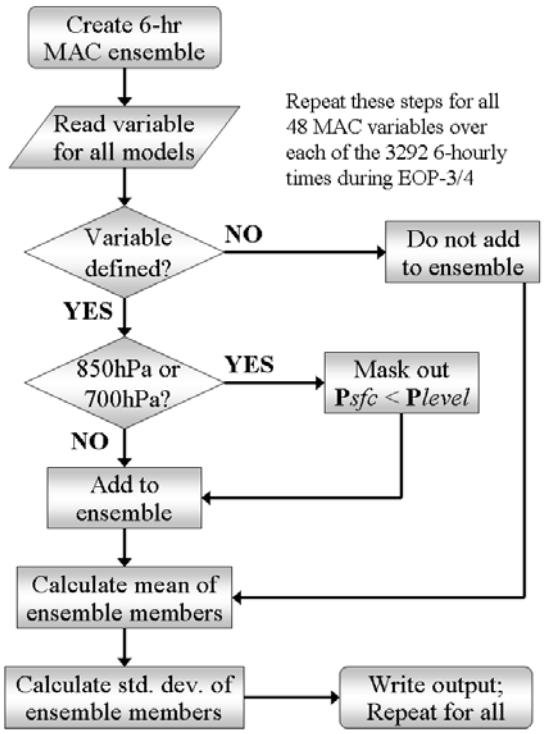



Figure A1

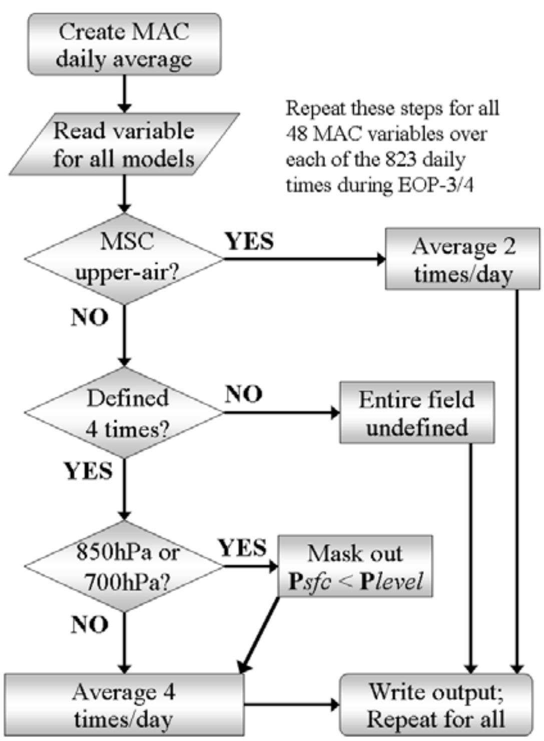



Figure A2

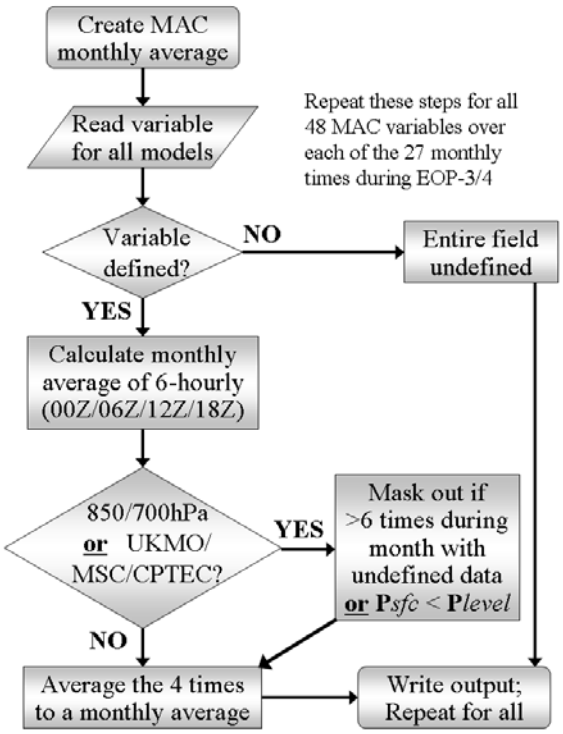

Create MAC monthly average
Read variable for all models
Repeat these steps for all 48 MAC variables over each of the 27 monthly times during EOP-3/4
Variable defined?
NO
Entire field undefined
YES
Calculate monthly average of 6-hourly (00Z/06Z/12Z/18Z)
850/700hPa or UKMO/ MSC/CPTEC?
YES
Mask out if >6 times during month with undefined data or Psfc < Plevel
NO
Average the 4 times to a monthly average
Write output; Repeat for all

Figure A3

Table A1 List of variables, descriptions/names, and availability by center of data used in the MAC.

| Description | Units | BMRC | CPTEC | ECPCRII | ECPCSFM | JMA | MSC | NCEP | UKMO |
|---|---|---|---|---|---|---|---|---|---|
| | | **Centers** | | | | | | | |
| Surface Pressure | Pa | SURFPsfc | PRESsfc | PRESsfc | PRESsfc | PRESsfc | SURFPsfc | PRESsfc | SURFPsfc |
| Mean Sea Level Pressure | Pa | MSLPsfc | | PRMSLmsl | PRESmsl | | | | PRMSLmsl |
| Surface Air Temperature | K | TMP2m | TMP2m | TMP2m | TMP2m | TMP2m | TTSUsfc | TMP2m | LOWT2m |
| Surface Skin Temperature | K | SURFTsfc | SURFTsfc | TMPsfc | TMPsfc | | SURFTsfc | TMPsfc | SURFTsfc |
| Surface Air Moisture | kg kg$^{-1}$ | SPFH2m | RH2m | SPFH2m | SPFH2m | SPFHhbl | HUSUsfc | SPFH2m | LOWSH2m |
| Surface Eastward Wind | m s$^{-1}$ | UGRD10m | UGRD10m | UGRD10m | UGRD10m | UGRD10m | UUSUsfc | UGRD10m | TENUS10m |
| Surface Northward Wind | m s$^{-1}$ | VGRD10m | VGRD10m | VGRD10m | VGRD10m | VGRD10m | VVSUsfc | VGRD10m | TENVS10m |
| Precipitation | kg m$^{-2}$ s$^{-1}$ | APCPsfc | APCPsfc | PRATEsfc | PRATEsfc | PRATEsfc | PRsfc | PRATEsfc | APCPsfc |
| Convective Precipitation | kg m$^{-2}$ s$^{-1}$ | | | CPRATsfc | CPRATsfc | | | | ACPCPsfc |
| Surface Runoff | kg m$^{-2}$ | | WATRsfc | WATRsfc | WATRsfc | | N0sfc | WATRsfc | WATRsfc |
| Liquid equivalent snow depth | kg m$^{-2}$ | SNODsfc | | WEASDsfc | WEASDsfc | | I5sfc | WEASDsfc | |
| Latent Heat Flux | W m$^{-2}$ | LHTFLsfc | LHTFLsfc | LHTFLsfc | LHTFLsfc | LHTFLsfc | AVsfc | LHTFLsfc | LHTFLsfc |
| Sensible Heat Flux | W m$^{-2}$ | SHTFLsfc | SHTFLsfc | SHTFLsfc | SHTFLsfc | SHTFLsfc | AHsfc | SHTFLsfc | SHTFLsfc |
| Surface Incoming Shortwave | W m$^{-2}$ | DSWRFsfc | DSWRFsfc | DSWRFsfc | DSWRFsfc | DSWRFsfc | N4sfc | DSWRFsfc | TDSWSsfc |
| Surface Incoming Longwave | W m$^{-2}$ | DLWRFsfc | DLWRFsfc | DLWRFsfc | DLWRFsfc | DLWRFsfc | ADsfc | DLWRFsfc | TDLWSsfc |
| Surface Reflected Shortwave | W m$^{-2}$ | USWRFsfc | USWRFsfc | USWRFsfc | USWRFsfc | USWRFsfc | N4sfc-ASsfc | USWRFsfc | TUSWSsfc |
| Surface Outgoing Longwave | W m$^{-2}$ | ULWRFsfc | ULWRFsfc | ULWRFsfc | ULWRFsfc | ULWRFsfc | ADsfc-AIsfc | ULWRFsfc | TULWSsfc |
| TOA Longwave Outgoing | W m$^{-2}$ | | ULWRFtoa | ULWRFtoa | ULWRFtoa | ULWRFtoa | ARsfc | ULWRFtoa | TULWTtoa |
| TOA Shortwave Incoming | W m$^{-2}$ | | | DSWRFtoa | DSWRFtoa | DSWRFtoa | ABsfc | | TDSWTtoa |
| TOA Shortwave Outgoing | W m$^{-2}$ | | USWRFtoa | USWRFtoa | USWRFtoa | USWRFtoa | AUsfc | USWRFtoa | TUSWTtoa |
| Total Cloud Cover | (0-1) | | TCDCclm | TCDCclm | TCDCclm | TCDCsfc | TCDCsfc | TCDCclm | TCDCsfc |
| Total Column Water Vapor | kg m$^{-2}$ | PWATclm | PWATclm | PWATclm | PWATclm | | IHsfc | PWATclm | |
| Total Column Condensed Water | kg m$^{-2}$ | | | | | CWATprs | IEsfc | CWATclm | |
| Q850 | kg kg$^{-1}$ | SPFHprs | SPFHprs | SPFHprs | SPFHprs | | SPFHprs | RHprs | RHprs |
| T850 | K | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs |
| U850 | m s$^{-1}$ | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs |
| V850 | m s$^{-1}$ | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs |
| H850 | m | HGTprs | HGTprs | HGTprs | HGTprs | GPprs | HGTprs | HGTprs | GPprs |
| Q700 | kg kg$^{-1}$ | SPFHprs | SPFHprs | SPFHprs | SPFHprs | | SPFHprs | RHprs | RHprs |
| T700 | K | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs |
| U700 | m s$^{-1}$ | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs |
| V700 | m s$^{-1}$ | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs |
| H700 | m | HGTprs | HGTprs | HGTprs | HGTprs | GPprs | HGTprs | HGTprs | GPprs |
| Q500 | kg kg$^{-1}$ | SPFHprs | SPFHprs | SPFHprs | SPFHprs | | SPFHprs | RHprs | RHprs |
| T500 | K | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs |
| U500 | m s$^{-1}$ | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs |
| V500 | m s$^{-1}$ | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs |
| H500 | m | HGTprs | HGTprs | HGTprs | HGTprs | GPprs | HGTprs | HGTprs | GPprs |
| Q300 | kg kg$^{-1}$ | SPFHprs | SPFHprs | SPFHprs | SPFHprs | | SPFHprs | RHprs | RHprs |
| T300 | K | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs |
| U300 | m s$^{-1}$ | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs |
| V300 | m s$^{-1}$ | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs |
| H300 | m | HGTprs | HGTprs | HGTprs | HGTprs | GPprs | HGTprs | HGTprs | GPprs |
| Q200 | kg kg$^{-1}$ | SPFHprs | SPFHprs | SPFHprs | SPFHprs | | | RHprs | RHprs |
| T200 | K | TMPprs | TMPprs | TMPprs | TMPprs | TMPprs | | TMPprs | TMPprs |
| U200 | m s$^{-1}$ | UGRDprs | UGRDprs | UGRDprs | UGRDprs | UGRDprs | | UGRDprs | UGRDprs |
| V200 | m s$^{-1}$ | VGRDprs | VGRDprs | VGRDprs | VGRDprs | VGRDprs | | VGRDprs | VGRDprs |
| H200 | m | HGTprs | HGTprs | HGTprs | HGTprs | GPprs | | HGTprs | GPprs |

Table A2 Location in the forecast cycle of each NWPC's variables included in the MAC.

| Description | Units | BMRC | CPTEC | ECPCRII | ECPCSFM | JMA | MSC | NCEP | UKMO |
|---|---|---|---|---|---|---|---|---|---|
| Surface Pressure | Pa | analysis | 12hr fcst+ | 6hr fcst | 6hr fcst | analysis | anl/6hr fcst+ | analysis | analysis |
| Mean Sea Level Pressure | Pa | analysis | | analysis | analysis | | | | analysis |
| Surface Air Temperature | K | analysis | 12hr fcst+ | 6hr fcst | 6hr fcst | analysis | anl/6hr fcst+ | 6hr fcst | analysis |
| Surface Skin Temperature | K | analysis | 12hr fcst+ | 6hr fcst | 6hr fcst | | anl/6hr fcst+ | 6hr fcst | analysis |
| Surface Air Moisture | kg $kg^{-1}$ | analysis | 12hr fcst+ | 6hr fcst | 6hr fcst | analysis | anl/6hr fcst+ | 6hr fcst | analysis |
| Surface Eastward Wind | m $s^{-1}$ | analysis | 12hr fcst+ | 6hr fcst | 6hr fcst | analysis | anl/6hr fcst+ | 6hr fcst | analysis |
| Surface Northward Wind | m $s^{-1}$ | analysis | 12hr fcst+ | 6hr fcst | 6hr fcst | analysis | anl/6hr fcst+ | 6hr fcst | analysis |
| Precipitation | kg $m^{-2}$ $s^{-1}$ | 0-6hr ave | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 6hr fcst | 3hr fcst+ | 0-6hr ave | 0-6hr acc |
| Convective Precipitation | kg $m^{-2}$ $s^{-1}$ | | | 0-6hr ave | 0-6hr ave | | | | 0-6hr acc |
| Surface Runoff | kg $m^{-2}$ | | 12hr fcst+ | 0-6hr ave | 0-6hr ave | | 3hr fcst+ | 0-6hr acc | 0-6hr acc |
| Liquid equivalent snow depth | kg $m^{-2}$ | analysis | | 6hr fcst | 6hr fcst | | anl/6hr fcst+ | 6hr fcst | |
| Latent Heat Flux | W $m^{-2}$ | 0-6hr ave | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | 0-6hr ave | 0-6hr ave |
| Sensible Heat Flux | W $m^{-2}$ | 0-6hr ave | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | 0-6hr ave | 0-6hr ave |
| Surface Incoming Shortwave | W $m^{-2}$ | 0-6hr ave | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | 0-6hr ave | 0-6hr ave |
| Surface Incoming Longwave | W $m^{-2}$ | 0-6hr ave | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | 0-6hr ave | 0-6hr ave |
| Surface Reflected Shortwave | W $m^{-2}$ | 0-6hr ave | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | 0-6hr ave | 0-6hr ave |
| Surface Outgoing Longwave | W $m^{-2}$ | 0-6hr ave | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | 0-6hr ave | 0-6hr ave |
| TOA Longwave Outgoing | W $m^{-2}$ | | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | 0-6hr ave | 0-6hr ave |
| TOA Shortwave Incoming | W $m^{-2}$ | | | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | | 0-6hr ave |
| TOA Shortwave Outgoing | W $m^{-2}$ | | 12hr fcst+ | 0-6hr ave | 0-6hr ave | 0-6hr ave | 3hr fcst+ | 0-6hr ave | 0-6hr ave |
| Total Cloud Cover | (0-1) | | 12hr fcst+ | 0-1hr ave | 0-1hr ave | analysis | anl/6hr fcst+ | 0-6hr ave | analysis |
| Total Column Water Vapor | kg $m^{-2}$ | analysis | 12hr fcst+ | 6hr fcst | 6hr fcst | | anl/6hr fcst+ | analysis | |
| Total Column Condensed Water | kg $m^{-2}$ | | | | | analysis | anl/6hr fcst+ | analysis | |
| Q850 | kg $kg^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | | anl/12hr fcst | analysis | analysis |
| T850 | K | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| U850 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| V850 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| H850 | m | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| Q700 | kg $kg^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | | anl/12hr fcst | analysis | analysis |
| T700 | K | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| U700 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| V700 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| H700 | m | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| Q500 | kg $kg^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | | anl/12hr fcst | analysis | analysis |
| T500 | K | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| U500 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| V500 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| H500 | m | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| Q300 | kg $kg^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | | anl/12hr fcst | analysis | analysis |
| T300 | K | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| U300 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| V300 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| H300 | m | analysis | 12hr fcst+ | analysis | analysis | analysis | anl/12hr fcst | analysis | analysis |
| Q200 | kg $kg^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | | | analysis | analysis |
| T200 | K | analysis | 12hr fcst+ | analysis | analysis | analysis | | analysis | analysis |
| U200 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | | analysis | analysis |
| V200 | m $s^{-1}$ | analysis | 12hr fcst+ | analysis | analysis | analysis | | analysis | analysis |
| H200 | m | analysis | 12hr fcst+ | analysis | analysis | analysis | | analysis | analysis |