

## A Multimodel 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, Maryland*

JOHN O. ROADS AND ALEX RUANE<sup>†</sup>

*Experimental Climate Prediction Center, Scripps Institute of Oceanography, University of California, San Diego, La Jolla, California*

(Manuscript received 8 September 2008, in final form 11 March 2009)

### 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 datasets. The present study evaluated the global precipitation, outgoing longwave radiation (OLR) at the top of the atmosphere, and basin-scale precipitation over the United States. In addition, a multimodel ensemble was created from a linear average of the available data, as close to the analysis time as each system permitted. The results show that the monthly global precipitation and OLR from the multimodel ensemble compares generally better 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 comparisons have seasonality, when the members of the ensemble exhibit generally more skill, during winter. There is notably higher skill of the summertime basin precipitation by the ensemble. Using the global precipitation and OLR, the sensitivity was tested to selectively choose the members with the best statistical comparisons to the reference data. Only small improvements in the statistics were found when comparing a selective ensemble to the full ensemble. Additionally, terms of the global energy budget were compared among the ensemble and to other estimates. The ensemble data and the variance of the ensemble should make a useful point of comparison for the development of model and assimilation components of global analyses.

### 1. Introduction

Ensemble means of simulations using different models have been shown to provide a result better than the mean of the skill of the contributing members. At weather scales, improved hurricane predictions have been found through such superensembles (Krishnamurti et al. 2001). In climate simulations and predictions, a multimodel approach also tends to provide the better result (Phillips

and Gleckler 2006). Additionally, ensembles of stand-alone land process models (constrained by observations and analyses as prescribed forcing) show 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 require 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 also 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.

---

\* Additional affiliation: Science Applications International Corporation, San Diego, California.

<sup>†</sup> Current affiliation: NASA Oak Ridge Associated Universities Postdoctoral Fellow, NASA Goddard Institute for Space Studies, New York, New York.

---

*Corresponding author address:* Michael G. Bosilovich, Global Modeling and Assimilation Office, Code 610.1, NASA Goddard Space Flight Center, Greenbelt, MD 20771.  
E-mail: michael.bosilovich@nasa.gov

The Coordinated Enhanced Observing Period (CEOP; Koike 2004) has collected concurrent observations and operational analyses for the period October 2002–December 2004, in which 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. Although a suggested variable list was included in the request, data structures were not strictly provided. The contributed data are on various grids, and each center provides 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 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 have 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, whereas 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 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; by January 2008, seven centers provided 27 months of data for the CEOP Enhanced Observing Periods 3 and 4 (EOP-3 and EOP-4)

periods (October 2002–December 2004). Two separate model contributions from one center gave a total of eight analyses. The contributing centers are as follows:

- Bureau of Meteorology Research Centre (BMRC; Rikus 2007),
- Centro de Previsão de Tempo e Estudos Climáticos [The Center for Weather Forecasts and Climate Studies (CPTEC); Chou et al. 2007],
- Experimental Climate Prediction Center (ECPC)—Reanalysis II (RII) and Seasonal Forecast Model (SFM) (Ruane and Roads 2007b),
- Japan Meteorological Agency (JMA; Hirai et al. 2007),
- Meteorological Services of Canada (MSC; Côté et al. 1998; Bélair et al. 2005, 2009),
- National Centers for Environmental Prediction (NCEP; GCWMB 2003),
- Met Office (UKMO; Milton and Earnshaw 2007).

Each of the contributions was from operational numerical weather prediction centers, except for ECPC, which is a research institution. ECPC executed two experiments: one using the NCEP–Department of Energy Global Reanalysis 2 (Kanamitsu et al. 2002b) system, except with high temporal resolution (called ECPC-RII), and another with the NCEP SFM (Kanamitsu et al. 2002a; here called ECPC-SFM). Ruane and Roads (2007a,b) discuss the specific design of the CEOP experiments.

In general, comparisons of the analyses from the NWPCs have primarily been through the single-point Model Output Location Time Series (MOLTS) collocated with CEOP reference sites or through 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; 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 as well as 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. Last, a synthesis of the model output would facilitate the use of the data in the broader science community where increased use of the data should expose strengths and 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 gridded binary (GRIB1) format] by the Model and Data group at the

TABLE 1. Grid spacing of individual NWPC model output as archived at MPI showing the number of points ( $X \times Y$ ). Also the spatial resolution in degrees longitude by latitude, the number of levels in the model vertical coordinate, and the documented pressure at the top (hPa).

	1) Surface		2) Atmosphere pressure levels		3) Vertical grid	
	Points	Resolution	Points	Resolution	Levels	Top (hPa)
BMRC	$480 \times 240$	$0.75 \times \sim 0.75$	$480 \times 240$	$0.75 \times \sim 0.75$	29	10
CPTEC	$384 \times 192$	$0.9375 \times 0.9375$	$384 \times 192$	$0.9375 \times 0.9375$	28	2.5
ECPC-RII	$192 \times 94$	$1.875 \times \sim 1.915$	$144 \times 73$	$2.5 \times 2.5$	28	2.5
ECPC-SFM	$192 \times 94$	$1.875 \times \sim 1.915$	$144 \times 73$	$2.5 \times 2.5$	28	2.5
JMA	$640 \times 320$	$0.5625 \times \sim 0.5625$	$288 \times 145$	$1.25 \times 1.25$	40	0.4
MSC	$800 \times 600$	$0.45 \times 0.30$	$800 \times 600$	$0.45 \times 0.30$	58	10
NCEP	$360 \times 181$	$1.0 \times 1.0$	$360 \times 181$	$1.0 \times 1.0$	64	0.266
UKMO	$288 \times 145$	$1.25 \times 1.25$	$288 \times 145$	$1.25 \times 1.25$	38	3

Max Planck Institute (MPI) for Meteorology in Hamburg, Germany. Aside from the format, there are few similarities in the contributed data. For example, Table 1 shows the spatial resolution and grid structure of the data held in the CEOP archive. Each center provided various analysis/forecast data in their time series. Many provided their analysis and a 6-h forecast, but not all provided an analysis or forecast beyond six hours. As a rule, the data closest to the analysis provided by each center (either the analysis or nearest forecast data) were used for that center's time series in this comparison. In these systems, the model error grows in time, so weather will diverge from that which actually occurred. In this evaluation, we prefer to use the data closest to the analysis, so the weather patterns should be more highly correlated among the analyses. Contrary to this is the effect of spinup of the forecast, in which the model is adjusting to the analysis initial conditions (e.g., in the water cycle; Uppala et al. 2005). Data further into the forecast cycle (perhaps 24–36 h) would have less spinup error but more of the model's background error. The analysis/forecast data contributions varied among centers and details on the location of each center's data relative to the analysis/forecast cycle are provided in the appendix.

Not all the centers provided the same output variables. Forty-eight 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 ( $1.25^\circ$  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 the centers, the mean, and the standard deviation at the 6-hourly, daily, and monthly times in the final formats of Network Common Data Form (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 to the source data, check 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 October 2002–December 2004.

### 3. Evaluation of the ensemble and members

#### a. Monthly time scale

Precipitation from analyses can be a useful quantity but the uncertainties have to be understood (Trenberth et al. 2007; Bosilovich et al. 2008, and the citations therein). Many of the CEOP science objectives relate to precipitation. At monthly time scales, there are many similarities among the analyses provided here. Figures 1, 2 show the difference of each of the analyses to the Global Precipitation Climatology Project (GPCP; Adler et al. 2003) for July 2004. Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP; Xie and Arkin 1996) 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

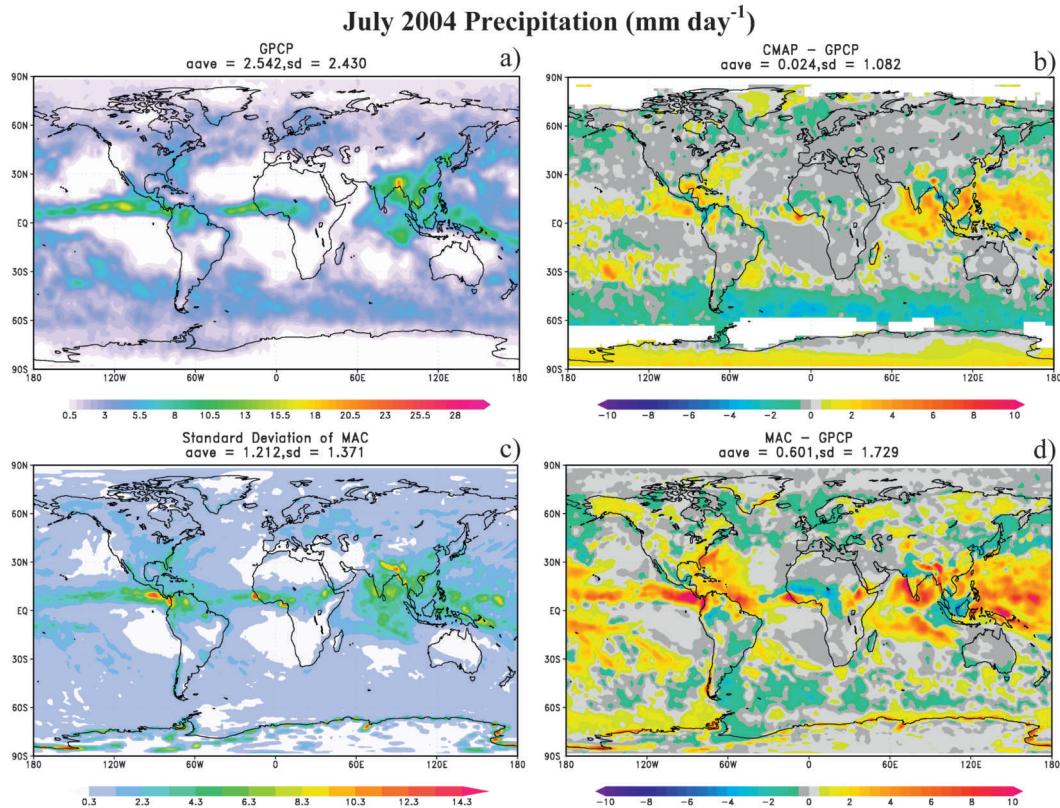


FIG. 1. July 2004 monthly-mean (a) GPCP precipitation, (b) difference of GPCP and CMAP precipitation, (c) standard deviation of the ensemble of monthly-mean MAC ensemble members, and (d) difference of GPCP and MAC precipitation ( $\text{mm day}^{-1}$ ). For reference, the area average (aave) and the standard deviation (sd) of the fields in each map are shown in the title of each panel.

degree the South Pacific convergence zone (SPCZ). Also note that CMAP is likewise biased slightly higher than GPCP in these areas (related to the implementation of atoll gauge observations; Yin et al. 2004), and the variance of the ensemble is largest here as well (Fig. 1c). 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, pointing to a key consideration of the MAC ensemble average; systematic errors and biases among the contributing members will persist into the resulting ensemble average.

Although 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 (Fig. 2). Summary statistics of global mean bias and standard deviation of the difference field are included in the titles of the figure. With a standard deviation of  $1.7 \text{ mm day}^{-1}$ , the MAC ensemble average has a lower error in this field than any of the contrib-

uting members. This suggests that the uncorrelated errors are being reduced in the ensemble. Figure 2 is representative of the large-scale variance among the analyses, in a visual sense. Subtle variations can be obscured in the contour intervals and color shades. Taylor diagrams (Taylor 2001) provide a quantitative measure of the skill in the map fields—in this case, lending to comparison of July 2004 precipitation to GPCP observations (Fig. 3). Taylor diagrams compare the variance of a field with their correlation relative to a reference data field, where the distance to the reference point is a measure of skill. 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 Fig. 3) shows how closely a model approximates the reference data (see also Phillips 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 4 extends this discussion across the entire October 2002–December 2004 period. The time series

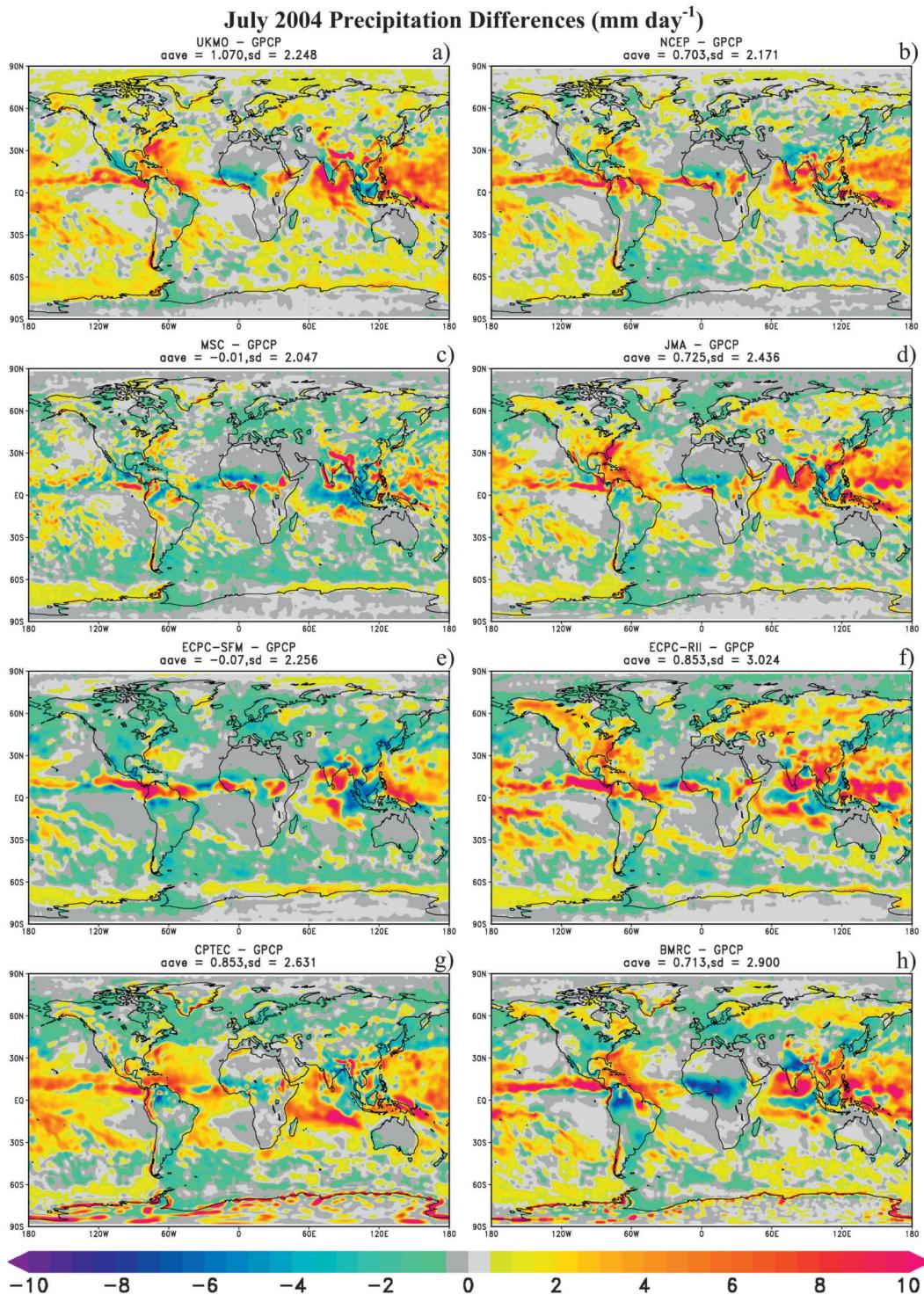


FIG. 2. As in Fig. 1, but for the July 2004 monthly precipitation differences for each of the ensemble members from GPCP merged precipitation data for (a) UKMO, (b) NCEP, (c) MSC, (d) JMA, (e) ECPC-SFM, (f) ECPC-RII, (g) CPTEC, and (h) BMRC.

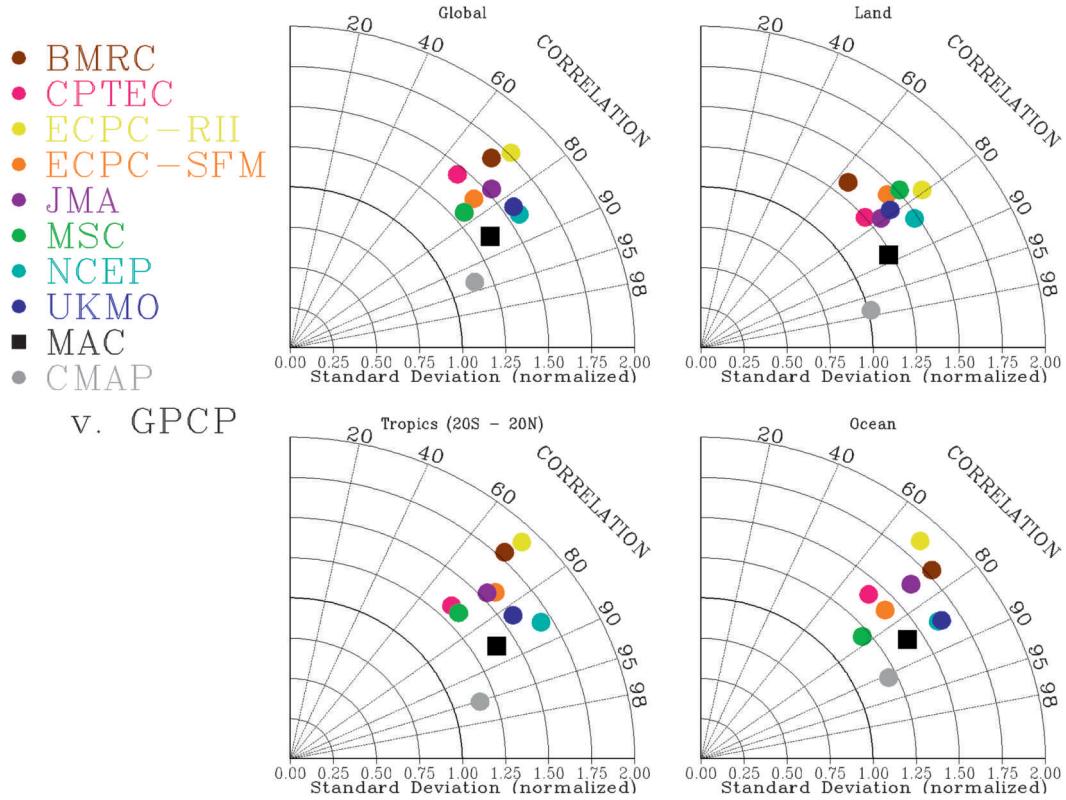


FIG. 3. Taylor diagrams (Taylor 2001) for global, global land, global ocean, and tropics precipitation. GPCP merged precipitation is the reference dataset. The diagrams compare spatial correlation (to GPCP) of the analysis to standard deviation normalized by the reference dataset for July 2004 monthly means. Linear distance to the (1, 1) point is a measure of skill in reproducing the reference dataset. CMAP (gray dot) is also provided as a measure of uncertainty in the observations.

of standard deviation (Fig. 4a) of the monthly-mean precipitation difference 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 several analyses in June 2003, in which a substantial increase in the standard deviation may also affect the MAC ensemble statistics. However, the shift of the MAC is a lower magnitude than these individual systems. 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 generally higher than any of the contributing members (Fig. 4b). (Anomaly correlations would have been more desirable but difficult to interpret with just two years of data to generate anomalies to the mean annual cycle.) However, there are a few months in the boreal winter of 2004 when NCEP's precipitation correlation is nearly identical to the MAC ensemble average, and the UKMO correlations are also generally higher than other analyses.

Outgoing longwave radiation (OLR) is a critical climate diagnostic, and global observations are readily available. However, uncertainty persists in both the observations and model representation of OLR. Trenberth et al. (2009) summarize the uncertainty of OLR observations and the comparison against existing long reanalyses. The reanalyses vary greatly compared to the observations. Figures 5, 6 compare the July 2004 monthly differences of the ensemble members and mean OLR compared with surface radiation budget (SRB; Cox et al. 2006; Lin et al. 2008) merged observational data product [and Clouds and the Earth's Radiant Energy System (CERES) Earth Radiation Budget Experiment (ERBE)-like OLR is provided as a reference of another observed data product (Wielicki et al. 1996; Loeb et al. 2001)]. As opposed to the precipitation comparison, wide variations among the models are apparent (Fig. 6). The JMA system exhibits systematic positive biases, whereas the UKMO shows systematic negative bias. NCEP, UKMO, and MSC biases are strongest in the tropics but of different signs. Although these latter three have the lowest standard deviations for this month, the MAC ensemble

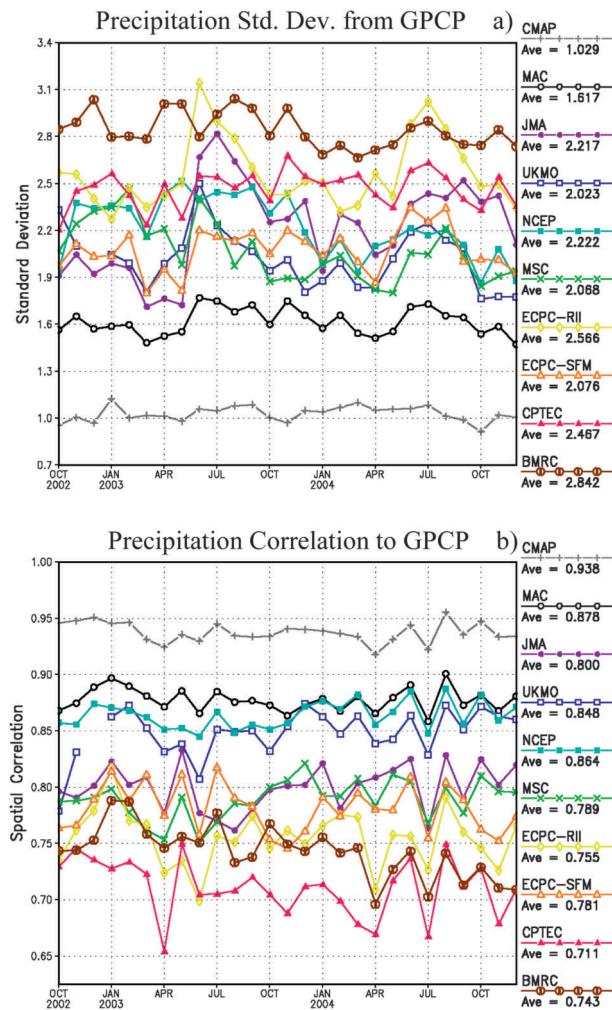


FIG. 4. Standard deviation of the monthly global differences from (a) GPCP and correlation to (b) GPCP for the MAC ensemble mean, each ensemble member, and CMAP merged observations (gray).

produces the lowest standard deviation of the difference from SRB than any single contributor.

Figure 7 shows the time series of global standard deviation and spatial correlation of the monthly differences between each member and the ensemble mean with SRB OLR. Spatial correlation values show how well the patterns match (note that seasonality generally adds to these correlation values). 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 error in summer. Further, the UKMO, MSC, and JMA OLR data are all generally close to the MAC ensemble mean. In contrast to the precipitation statistics (Fig. 4) that 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 selectively choosing the statistically better systems improve the ensemble mean?

#### b. Selective ensembles

Using Global Soil Wetness Project (GSWP-2) 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; however, 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 and included more members than the present study. 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 analysis here. The distribution of error evaluated by Guo et al. (2007) varied evenly across the ensemble members. This is somewhat different from the error we see generated in the three-dimensional operational analyses, where Fig. 7 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 (Fig. 4).

Here, we test the MAC ensemble average against a selective member ensemble, determined from each systems statistics in Fig. 4 and Fig. 7, 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). Figures 8a,b shows the time series of statistics, including an ensemble mean of the five 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 were 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); however, all

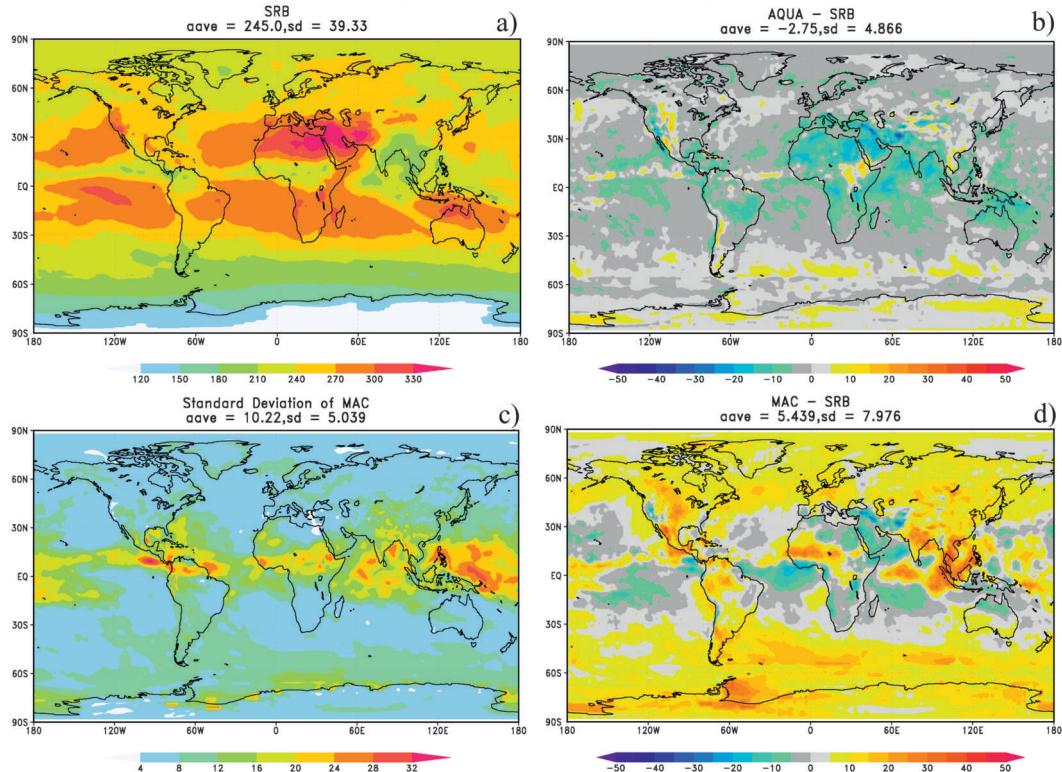
July 2004 TOA OLR ( $\text{W m}^{-2}$ )

FIG. 5. July 2004 monthly TOA OLR ( $\text{W m}^{-2}$ ) from (a) SRB, (b) the difference of *Aqua* CERES ERBE-like OLR from SRB, (c) the standard deviation of monthly-mean MAC ensemble members, and (d) the difference of the MAC ensemble from SRB. The mean and standard deviation of the field shown in each map are summarized above the maps.

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. Removing the lowest three systems from the ensemble does not increase the skill of the ensemble spatial correlation (Fig. 8d). However, in standard deviation (Fig. 8c), 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 of the ensembles is reduced on average from 8.1 to 7.4  $\text{W m}^{-2}$  (a difference of  $\sim 0.6 \text{ W m}^{-2}$ ); therefore, for some cases, a selective ensemble may provide better results. Although 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 to what degree 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; Fig. 8a). Because 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 are identified in specific process studies, selective ensembles may be justified.

The JMA OLR statistics show a general improvement over the period (Fig. 7). The JMA OLR is also included in the selective ensemble. However, in neither of the ensembles is an improvement of the statistics in time noticeable (Figs. 8c,d). The ensembles' statistics are steadier compared to the individual analyses. This suggests that the individual effect 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 more favorably to global datasets. Because the analyses assimilate observations, 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

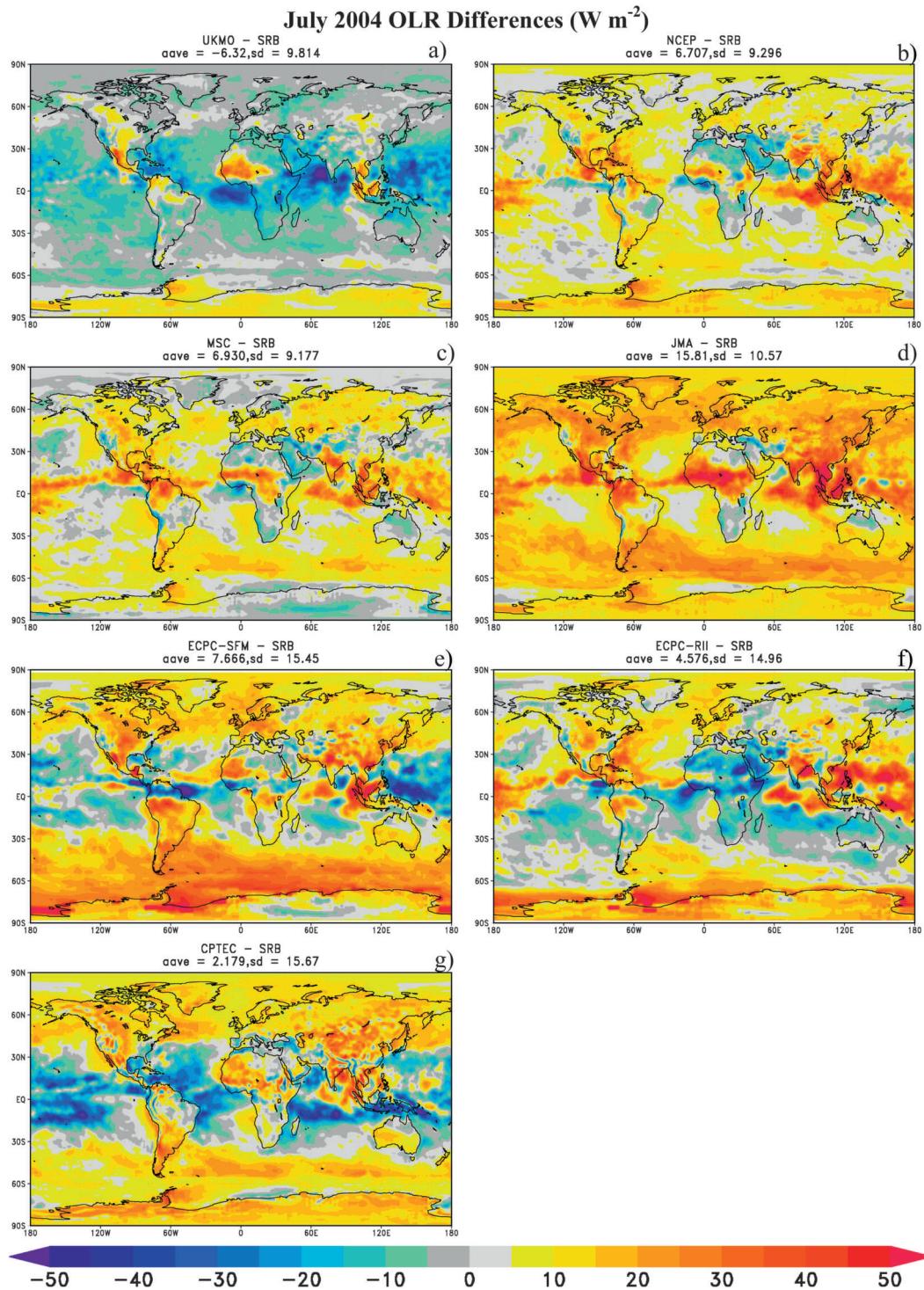


FIG. 6. As in Fig. 5, but for the July 2004 monthly TOA OLR differences for each of the ensemble members from SRB OLR for (a) UKMO, (b) NCEP, (c) MSC, (d) JMA, (e) ECPC-SFM, (f) ECPC-R11, and (g) CPTEC. (BMRC did not provide OLR.)

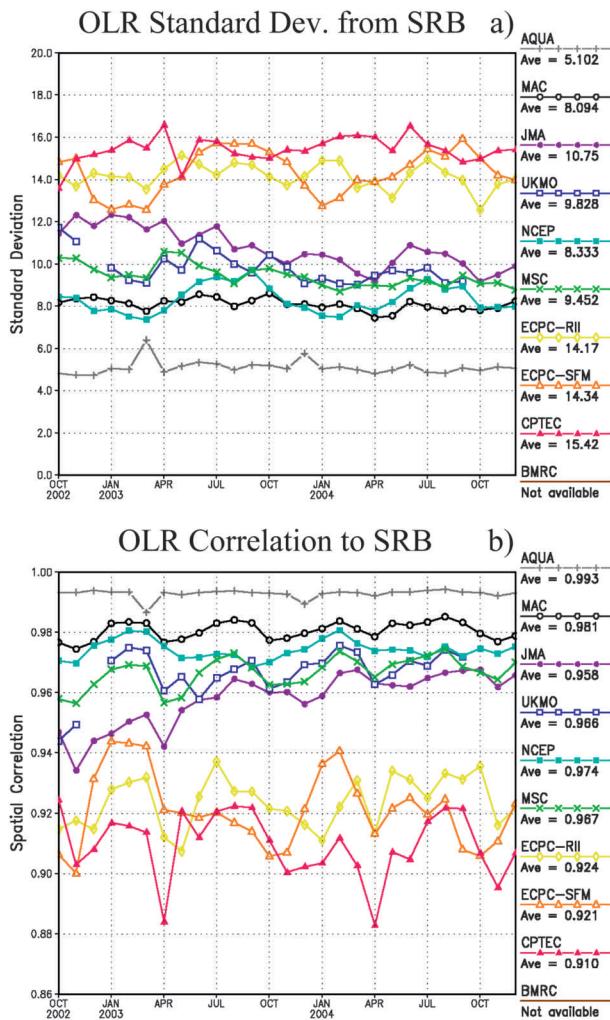


FIG. 7. Standard deviation of the monthly global differences from (a) SRB OLR and (b) correlation to SRB OLR for the MAC ensemble mean, each ensemble member providing the quantity, and the CERES ERBE-like OLR retrieved from *Aqua* observations (gray).

scales. The CPC provides a 1/4° daily gridded gauge-only precipitation dataset for the United States (Shi et al. 2003). We evaluate the daily time series of precipitation over the Mississippi River basin (MRB) and subbasins for the period January 2003 through December 2004.

Figure 9 shows the MRB on the MAC grid, on which we will focus this evaluation. Five subbasins 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 10 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. ECPC-RII and ECPC-SFM tend to be some of the coarser source datasets (Table 1), which may have some influence on their results. However, it is difficult to conclude much about the influence of resolution in this evaluation because the BMRC data have more fine resolution than ECPC but also lower skill.

Table 2 shows the statistics of the area-averaged daily time series for the MRB and each of the subbasins. 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 subbasins, the MAC ensemble time series has some of the highest temporal correlation to the observations, and the standard deviation of the time series differences is always the lowest. These temporal statistics suggest that the MAC ensemble precipitation is generally consistent with the weather-scale observations.

The previous statistics show the area-average comparison of the time series to the gauge precipitation record. However, one possibility is that the simple linear ensemble dilutes the gradients in the occurrence of measurable precipitation across the basin. One way to test this is to compute spatial statistics for each day, between the models and the gauge observations. Spatial correlation and standard deviation of the difference from the gauge observations across the MRB domain are computed each day of the 2003/04 period. The daily time series are quite variable, so to compensate, Fig. 11 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 value indicates small squared difference from the observations). This is most apparent in the warm season, when the errors are generally larger than in the other seasons. In the winter season, 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 (Fig. 11b). Some of the analyses that perform better for the whole period (see Table 2) are closer to the high values of the MAC. In some months (generally in the winter), the MAC ensemble is not the highest spatial correlation. In the winter, dynamics and initial conditions

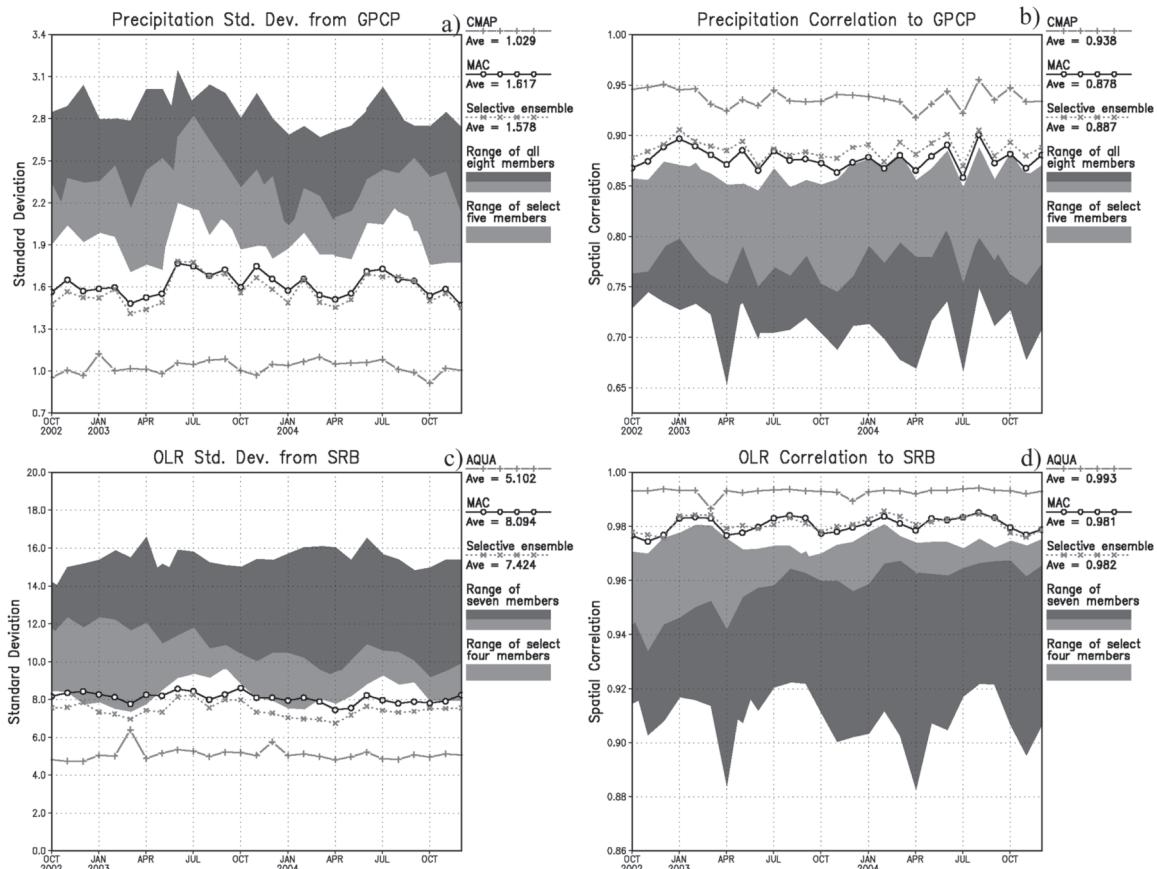


FIG. 8. Comparison of the statistics of the MAC ensemble (black line) compared to a selective ensemble (gray dashed line), and the range of values contributing to the full ensemble and the selective ensemble. Standard deviation from (a) GPCP, (b) correlation to GPCP, (c) standard deviation from SRB OLR, and (d) correlation to SRB OLR. CMAP precipitation and CERES ERBE-like OLR from *Aqua* are included on the respective panels for reference. Note that UKMO monthly precipitation is missing from December 2002, leaving only a 4-member selective ensemble in that month. Also, UKMO monthly OLR is missing from December 2002 and October–December 2004, leaving only a 3-member selective ensemble in those months.

of the data analysis provide skillful forecasts that allow more accurate precipitation occurrence. In the summer, physical processes and mesoscale systems govern the observed precipitation, thus the forecasts and analyses have more uncertainty. The significance of Fig. 9 is that the daily basin-area spatial correlation represents the scale at which synoptic weather is producing measurable precipitation, so weather information (occurrence and timing) is not lost in the simple linear ensemble of the analyses, at least at the daily scale. Given that the summer precipitation patterns are more governed by the physical parameterizations than the dynamical forcing, it seems reasonable that the overwhelming source of uncertainty in the analyses is the essentially randomness of the convective precipitation. The ensemble result reduces the random errors and compares better with observations.

Figures 12, 13 compare precipitation of two individual days, one summer (10 July 2003) and one winter

(23 December 2004), 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; it also underestimates the intensity at the core of the event. UKMO is chosen as a member of the ensemble, with reasonable statistics. In the July case, UKMO does produce a larger amount of precipitation 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 have a higher correlation to and a smaller variance than the UKMO data for the MRB but not without some deficiency.

In the December example (Fig. 13), a strong frontal system extends across the United States, west of the

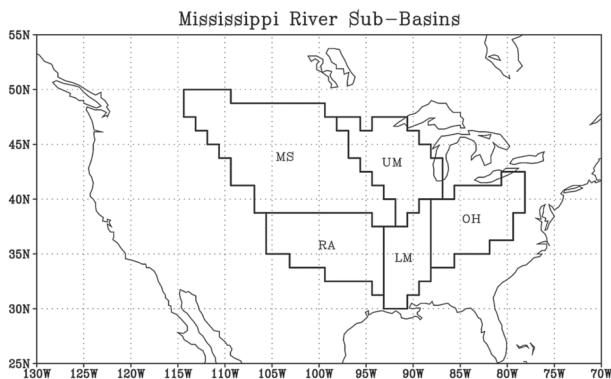


FIG. 9. Five subbasins of the MRB: Red-Arkansas (RA), Missouri (MS), Upper Mississippi (UM), Ohio (OH), and the Lower Mississippi (LM). The outlines also show the MAC  $1.25^\circ$  grid. The collective region is referred to as the MRB.

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 southeastern maxima, whereas the UKMO system does have more precipitation there (though no closed contours are 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. Table 3 shows the standard deviation and spatial correlation for the MAC and the individual members in the MRB on these two dates.

Roads and Betts (2000) compared the European Centre for Medium-Range Weather Forecasts (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 because not all data are available (notably, a soil moisture storage diagnostics among the systems is not uniform and 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 (2007a,b) evaluate the reanalyses' diurnal cycles globally and regionally, finding significant deficiencies in both the phase and amplitude of diurnal precipitation. Although 6-hourly data are

available in this multimodel dataset, 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. Trenberth et al. (2009, henceforth TFK) recently revised assessments of the energy budget components based on newly available data and models. The assessment used GPCP, International Satellite Cloud Climatology Project (ISCCP) and CERES observations. The TFK period of interest was from March 2000 to May 2004. We compare the 2-yr (January 2003–Dec 2004) globally averaged energy budget from the ensemble members and mean to the representative values developed by TFK.

Table 4 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) separated for the globe, land, and ocean. 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. There is considerable variability of the land surface turbulent fluxes, and the ECPC-RII data appears to be an outlier.

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, whereas the upward longwave radiation is consistent, related to the use of prescribed SSTs. TFK estimated the net heating of the global surface to be  $0.9 \text{ W m}^{-2}$ . 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. Note that net TOA heating is provided from the ensemble; however, with fewer systems contributing to that quantity, total heating of the atmosphere may have additional uncertainty. The ocean imbalance at the surface is a combination of prescribed SST acting as a heat source/sink for the atmosphere. Also, the net surface heating over land is large in some models. This may be the result of either not enough of

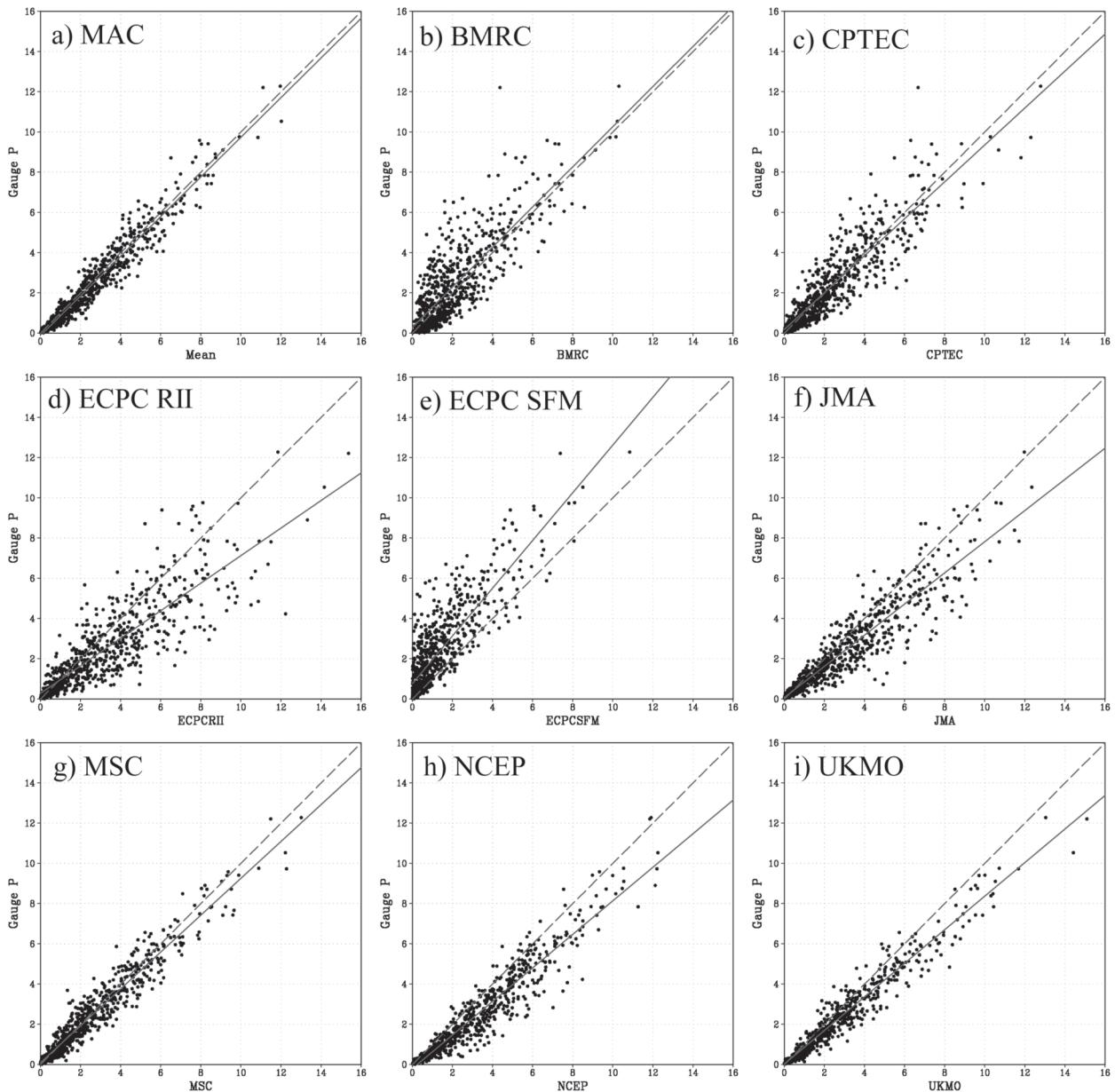


FIG. 10. Scatter diagrams comparing area-averaged MRB daily observed precipitation ( $\text{mm day}^{-1}$ ) with the MAC ensemble and with all of the contributing member data, by panel. The solid line indicates the linear fit of the data points and the dashed line indicates 1:1.

the output data considered in the analysis (such as flux of heat related to snow), inconsistencies in the single land/sea mask generated for this purpose, or the analysis's effect on the fluxes. The main point is that operational analyses demonstrate substantial bias and variability in many global water and energy budget terms.

#### 4. Summary and conclusions

The Multimodel Analysis for CEOP (MAC) comprises eight 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 and compare among the different analyses. A similar effort has begun to take shape at NCEP, primarily focusing on the state variables (Ebisuzaki et al. 2007). 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 are as skillful as or better than the most skillful

TABLE 2. Statistics from the 2-yr aave time series of daily precipitation of each analysis system and the ensemble (MAC) compared to gauge observations. Mean bias is the mean difference from observations ( $\text{mm day}^{-1}$ ). Std dev is the standard deviation of the difference from observations ( $\text{mm day}^{-1}$ ). Correlation is the temporal correlation to the observed time series.

Mean bias	MRB	RA	MS	UM	OH	LM
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
ECPC-RII	0.66	0.63	0.61	0.50	0.77	0.92
ECPC-SFM	-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	0.12	0.04	0.14	0.10	0.38	0.10

Std dev	MRB	RA	MS	UM	OH	LM
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
ECPC-RII	1.35	2.43	1.46	2.21	3.03	4.07
ECPC-SFM	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

Correlation	MRB	RA	MS	UM	OH	LM
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
ECPC-RII	0.86	0.85	0.87	0.85	0.85	0.78
ECPC-SFM	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

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. Testing of the data focused on precipitation and outgoing longwave radiation, variables that are predominantly driven by the model physics but also have reliable global observations datasets 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; however, 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 closer to GPCP than any of the members. Comparing the members and ensemble to the global energy budget estimate (Trenberth et al. 2009) shows which terms in

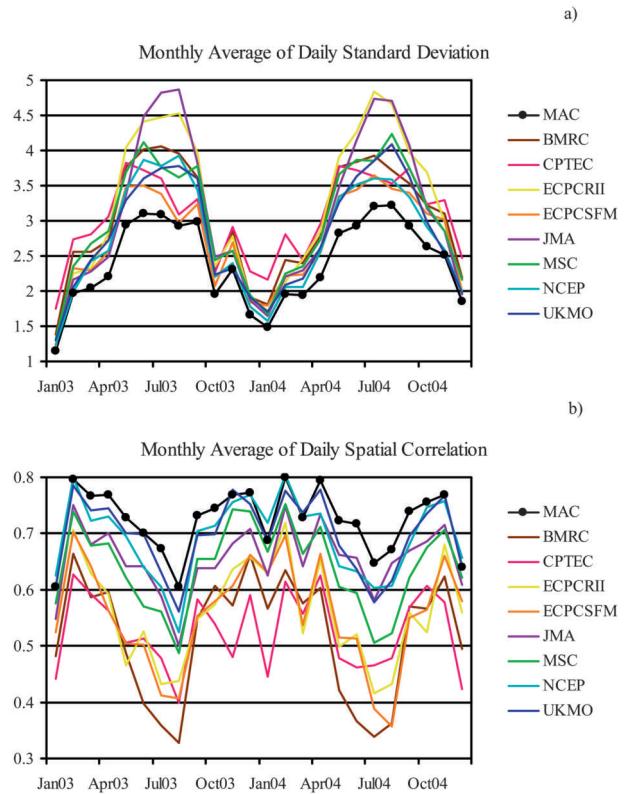


FIG. 11. Time series of monthly means of the (a) daily standard deviation of MRB area differences ( $\text{mm day}^{-1}$ ) and (b) daily spatial correlation of the MRB-area precipitation from each model system to the gauge observations.

the analyses are similar and which vary. Although the global net surface energy from any given system may show imbalance, the ensemble net surface energy is in-line with the TFK estimate. The ensemble also shows 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 subbasins.

The effect 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 that the number of members is not large. 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.

Although 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

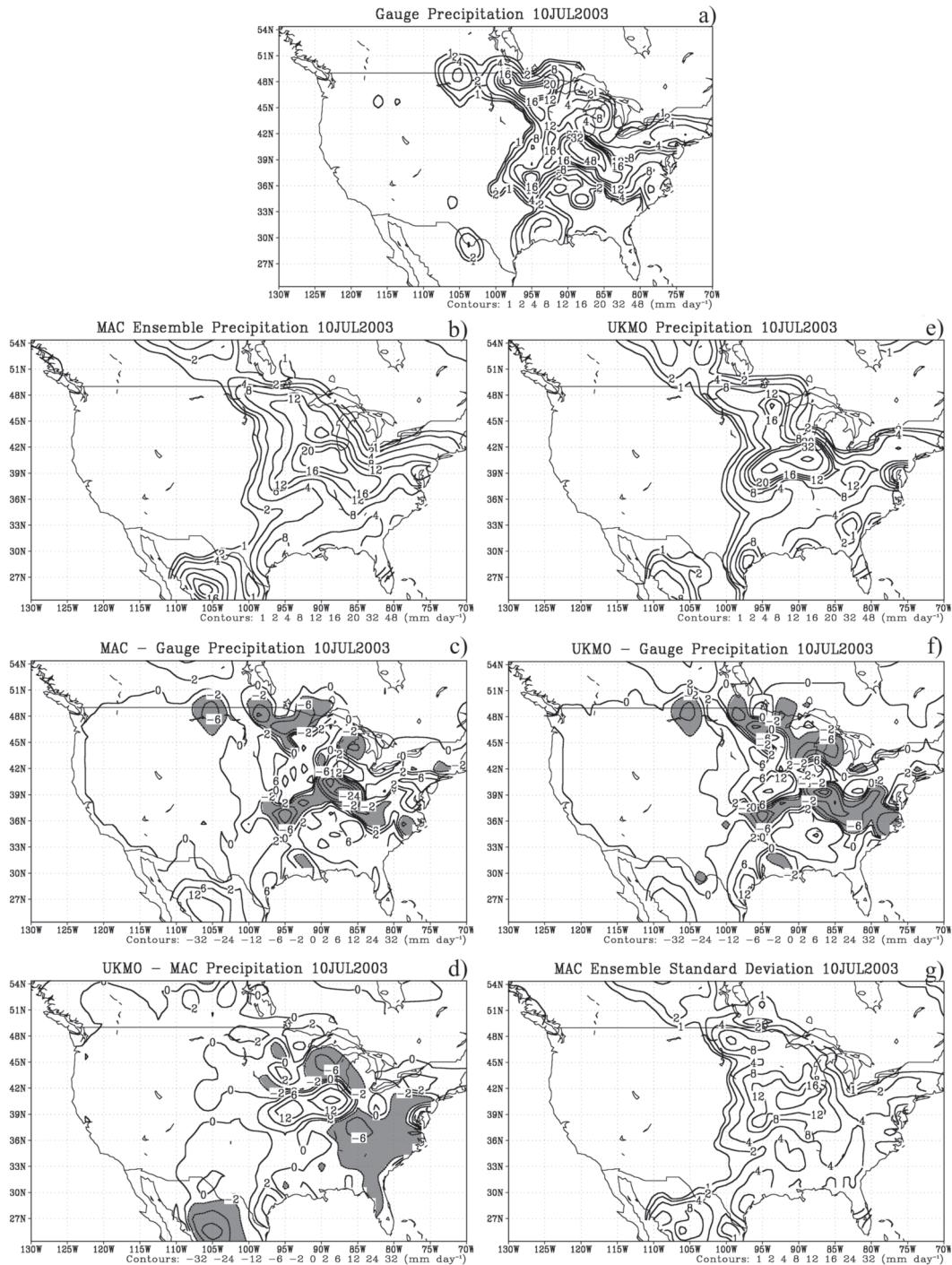


FIG. 12. Daily precipitation ( $\text{mm day}^{-1}$ ) for 10 Jul 2003 (a) gauge observations, (b) MAC ensemble, (c) MAC minus gauge, (d) UKMO minus MAC, (e) UKMO daily mean, (f) UKMO minus gauge, and (g) the standard deviation of the ensemble mean.

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

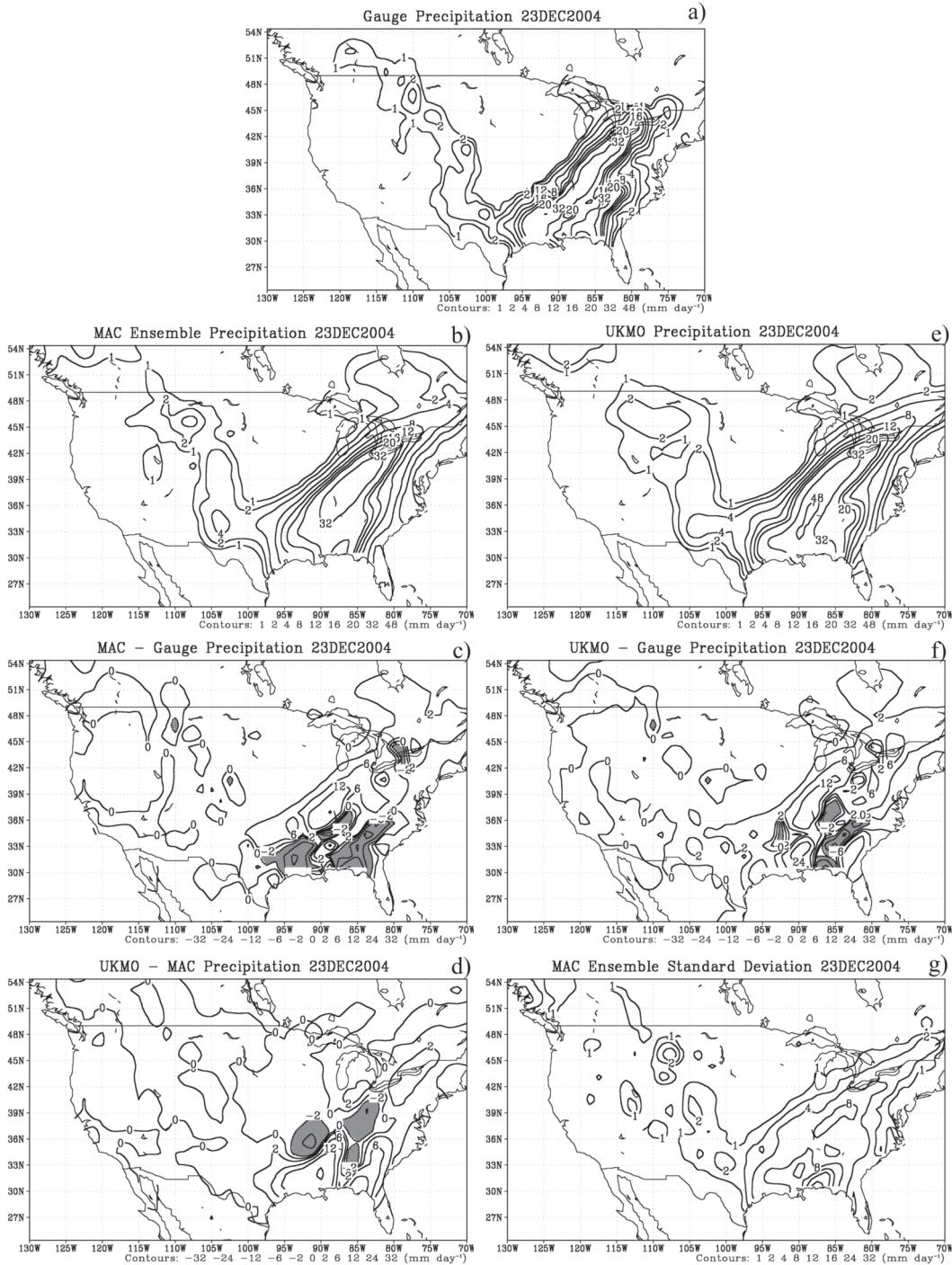


FIG. 13. As in Fig. 12, but for 23 Dec 2004.

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 ob-

servations (e.g., apparent in the JMA OLR) do not appear to be reflected in the MAC ensemble statistics.

Since preparing this initial evaluation of the data and procedures, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Global Modeling and Assimilation Office (GMAO) have provided data

TABLE 3. Daily statistics for the MRB precipitation of each analysis and the MAC ensemble compared to gauge observations. For each day, the (top) correlations and the (bottom) standard deviations are shown.

	MAC	BMRC	CPTEC	ECPC-RII	ECPC-SFM	JMA	MSC	NCEP	UKMO
10 Jul 2003	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
23 Dec 2004	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 4. Global, land, and ocean ave of energy budget components for each of the analyses, the ensemble mean (MAC), and the values reported by TFK. Objectively Analyzed Air–Sea Fluxes (OASFlux) (ocean; Yu 2007; Yu and Weller 2007; Yu et al. 2008), SRB, and GPCP global average of January 2003–December 2004 are also included in the row labeled “Observations.” MAC indicates the global average of the gridded ensemble data. The std dev is the standard deviation of the global averages of the members’ values in the table. Units are  $W m^{-2}$ , except for precipitation (Precip,  $mm day^{-1}$ ). NetET<sub>sfc</sub> and Net<sub>toa</sub> indicates the net downward heat flux at the surface and TOA, respectively. Note that the net TOA heating in the ensemble is computed from fewer models than net surface. The radiation components (R) are shown for shortwave (S) and longwave (L) directed upward (u) and downward (d) at the surface (subscript sfc) and TOA (subscript toa). Also provided are are latent heat (LH) and sensible (SH) at the surface, directed positive upward.

Global	LH	SH	RLd <sub>sfc</sub>	RLu <sub>sfc</sub>	RSd <sub>sfc</sub>	RSu <sub>sfc</sub>	RLu <sub>toa</sub>	RSd <sub>toa</sub>	RSu <sub>toa</sub>	Net <sub>sfc</sub>	Net <sub>toa</sub>	Precip
BMRC	92	18	350	400	167	25	—	—	—	−19.3	—	3.32
CPTEC	99	22	347	401	206	24	244	—	100	7.4	−2.6	3.40
ECPC-RII	96	7	337	400	202	24	247	341.4	92	11.9	2.9	3.21
ECPC-SFM	84	17	333	400	207	27	249	341.4	91	13.0	1.5	2.47
JMA	90	17	320	398	204	25	257	341.4	88	−6.2	−3.8	3.15
MSC	91	20	334	397	201	28	249	342.0	94	−0.6	−0.5	2.61
NCEP	95	9	332	398	208	29	248	—	87	8.5	6.0	3.26
UKMO	95	16	345	399	180	22	235	341.5	105	−7.2	1.5	3.58
MAC	92	16	337	399	197	26	247	341.5	93	1.1	0.9	3.12
Std dev	4.6	5.3	9.8	1.3	15.2	2.4	6.6	0.2	6.5	11.3	3.3	0.38
TFK	80	17	333	396	184	23	238.5	341.3	101.9	0.9	0.9	2.76
SRB/GPCP	—	—	343	398	182	22	241	341.4	—	—	—	2.63
Land	LH	SH	RLd <sub>sfc</sub>	RLu <sub>sfc</sub>	RSd <sub>sfc</sub>	RSu <sub>sfc</sub>	RLu <sub>toa</sub>	RSd <sub>toa</sub>	RSu <sub>toa</sub>	Net <sub>sfc</sub>	Net <sub>toa</sub>	Precip
BMRC	51	46	326	401	212	40	—	—	—	01	—	2.40
CPTEC	55	54	333	401	224	44	253	—	106	2.3	−15.3	2.46
ECPC-RII	87	−3	322	393	208	42	246	343.2	105	12.4	−7.5	2.81
ECPC-SFM	54	42	313	396	232	45	255	343.2	89	7.4	−0.6	1.95
JMA	53	27	301	391	219	46	260	343.1	95	1.7	−11.9	2.45
MSC	49	43	319	392	206	43	250	342.0	103	−2.6	−10.5	2.34
NCEP	65	23	322	392	208	44	249	—	101	5.8	−7.2	2.84
UKMO	58	31	328	396	196	38	240	343.4	104	1.2	−0.9	2.64
MAC	59	33	320	395	214	43	250	343.0	100	3.5	−7.8	2.48
Std dev	12.2	18.0	9.9	3.9	11.3	2.7	6.3	0.5	6.1	4.8	5.5	0.28
TFK	39	27	304	383	185	40	232	330.2	113	0.0	−15.6	—
SRB/GPCP	—	—	329	402	192	35	243	343.1	—	—	—	2.30
Ocean	LH	SH	RLd <sub>sfc</sub>	RLu <sub>sfc</sub>	RSd <sub>sfc</sub>	RSu <sub>sfc</sub>	RLu <sub>toa</sub>	RSd <sub>toa</sub>	RSu <sub>toa</sub>	Net <sub>sfc</sub>	Net <sub>toa</sub>	Precip
BMRC	109	7	359	399	149	20	—	—	—	−27.3	—	3.69
CPTEC	117	9	352	400	199	15	240	—	98	9.5	2.9	3.79
ECPC-RII	99	11	343	403	200	17	247	340.6	86	11.6	7.2	3.38
ECPC-SFM	95	6	341	402	197	19	247	340.6	91	15.4	2.4	2.69
JMA	105	12	327	401	198	17	256	340.6	85	−9.5	−0.5	3.43
MSC	108	11	341	399	199	22	248	341.8	90	0.3	3.6	2.72
NCEP	106	4	336	400	208	23	248	—	81	9.5	11.8	3.43
UKMO	110	10	352	400	173	16	233	340.6	105	−10.7	2.5	3.96
MAC	106	9	344	401	191	19	246	340.9	90	0.1	4.4	3.38
Std dev	6.5	2.9	10.2	1.4	19.7	2.9	7.2	0.6	8.2	14.7	4.0	0.47
TFK	97	12	343	401	184	17	241	345	98	1.0	6.8	—
Observations	92	10	349	396	178	17	240	341	—	—	—	2.78

covering the period evaluated here. These are currently being processed, so that version two of this dataset will include 10 members. At this point, the existing long reanalyses from NCEP and JRA-25 have been excluded to mainly focus on operational systems. However, these could also be included as some point in the future.

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 shifts that would be major changes to any individual system 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. However, older reanalyses eventually end [e.g., 40-yr ECMWF Re-Analysis (ERA-40) is not available later than August 2002]. The notion of a multimodel 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 Intergovernmental Panel on Climate Change (IPCC) projections, systems to handle these data and to make them available to the community are tractable. To accomplish such a repository would also require an investment of time and resources from already overburdened numerical weather prediction centers. However, the results presented here suggest that the investment would be useful to both the research community and to 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, or from the international collaborative environment fostered by CEOP and the Global Energy and Water Cycle Experiment (GEWEX). Petra Koudelova and Sam Benedict provided the organizational support that kept the CEOP model group to a schedule. Dr. Per Kållberg and two anonymous reviewers provided thoughtful comments and suggestions that greatly improved the final form of the paper. The NASA Modeling, Analysis and Prediction program supported this project. The MAC data are available for download from the NASA Goddard Data

Information Services Center (DISC) and from the Max Planck Institute Model and Data Center. The SRB and CERES ERBE-like observations were obtained from the NASA Langley Research Center Atmospheric Science Data Center. GPCP precipitation and CMAP data are available online (<http://precip.gsfc.nasa.gov/> and <http://www.cdc.noaa.gov>, respectively).

## APPENDIX

### Data Processing

In section 2, seven steps of the process for reformatting the data are outlined. The following subsections provide further details on the procedures and methods. The source data for the models is documented at the CEOP Data Management Center (available online at [http://www.eol.ucar.edu/projects/ceop/dm/model/model\\_chars.html](http://www.eol.ucar.edu/projects/ceop/dm/model/model_chars.html)) and also at the Max Planck Institute Model and Data Center (available online at <http://www.mad.zmaw.de/projects-at-md/ceop/>).

#### *a. Six-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 typically available was the analysis (0-h forecast) for the instantaneous variables and the 0–6 hourly forecast for the average/accumulation (ave/acc) variables. Some significant exceptions include the following:

- The CPTEC data, which was only run once a day starting at 1200 UTC, and no data were available for 12 h. Thus, the data is a 12-hourly forecast at 0000 UTC, an 18-h forecast at 0600 UTC, a 24-h forecast at 1200 UTC, and a 36-h forecast at 1800 UTC. Similarly, the ave and acc variables from 0000 to 0600 UTC are a 12–18 hourly forecast, and so on.
- The MSC data, which were only run once a day starting at 1200 UTC. The instantaneous surface variable data at 1200 UTC are an analysis/0-h forecast, at 1800 UTC a 6-h forecast, at 0000 UTC a 12-h forecast, and at 0600 UTC an 18-h forecast. The upper-air data, however, was not available at 0600 and 1800 UTC; the 1200 UTC data are an analysis/0-h forecast and the 0000 UTC data are a 12-h forecast. The MSC ave and acc variables from 1200 to 1800 UTC are a 0–6 hourly forecast, and so on.
- Several NCEP and ECPC RII and SFM instantaneous surface variables are a 6-h forecast rather than an analysis/0-h forecast.

TABLE A1. Location in the forecast cycle of each NWPC's variables included in the MAC, including variable description and units. Abbreviations are defined as follows: forecast (fcst) and analysis (anl).

Description	Units	BMRC	CPTEC	ECPC-RII	ECPC-SFM	JMA	MSC	NCEP	UKMO
Surface pressure	Pa	analysis	12-h fcst+	6-h fcst	6-h fcst	analysis	anl/6-h fcst+	analysis	analysis
Mean sea level pressure	Pa	analysis		analysis	analysis				analysis
Surface air temperature	K	analysis	12-h fcst+	6-h fcst	6-h fcst	analysis	anl/6-h fcst+	6-h fcst	analysis
Surface skin temperature	K	analysis	12-h fcst+	6-h fcst	6-h fcst		anl/6-h fcst+	6-h fcst	analysis
Surface air moisture	kg kg <sup>-1</sup>	analysis	12-h fcst+	6-h fcst	6-h fcst	analysis	anl/6-h fcst+	6-h fcst	analysis
Surface eastward wind	m s <sup>-1</sup>	analysis	12-h fcst+	6-h fcst	6-h fcst	analysis	anl/6-h fcst+	6-h fcst	analysis
Surface northward wind	m s <sup>-1</sup>	analysis	12-h fcst+	6-h fcst	6-h fcst	analysis	anl/6-h fcst+	6-h fcst	analysis
Precipitation	kg m <sup>-2</sup> s <sup>-1</sup>	0-6-h avg	12-h fcst+	0-6-h avg	0-6-h avg	6-h fcst	3-h fcst+	0-6-h avg	0-6-h acc
Convective precipitation	kg m <sup>-2</sup> s <sup>-1</sup>			0-6-h avg	0-6-h avg				0-6-h acc
Surface runoff	kg m <sup>-2</sup>		12-h fcst+	0-6-h avg	0-6-h avg		3-h fcst+	0-6-h acc	0-6-h acc
Liquid equivalent snow depth	kg m <sup>-2</sup>	analysis		6-h fcst	6-h fcst		anl/6-h fcst+	6-h fcst	
LH Flux	W m <sup>-2</sup>	0-6-h avg	12-h fcst+	0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+	0-6-h avg	0-6-h avg
SH Flux	W m <sup>-2</sup>	0-6-h avg	12-h fcst+	0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+	0-6-h avg	0-6-h avg
Surface incoming shortwave	W m <sup>-2</sup>	0-6-h avg	12-h fcst+	0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+	0-6-h avg	0-6-h avg
Surface incoming longwave	W m <sup>-2</sup>	0-6-h avg	12-h fcst+	0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+	0-6-h avg	0-6-h avg
Surface reflected shortwave	W m <sup>-2</sup>	0-6-h avg	12-h fcst+	0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+	0-6-h avg	0-6-h avg
Surface outgoing longwave	W m <sup>-2</sup>	0-6-h avg	12-h fcst+	0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+	0-6-h avg	0-6-h avg
TOA longwave outgoing	W m <sup>-2</sup>		12-h fcst+	0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+	0-6-h avg	0-6-h avg
TOA shortwave incoming	W m <sup>-2</sup>			0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+		0-6-h avg
TOA shortwave outgoing	W m <sup>-2</sup>		12-h fcst+	0-6-h avg	0-6-h avg	0-6-h avg	3-h fcst+	0-6-h avg	0-6-h avg
Total cloud cover	(0-1)		12-h fcst+	0-1-h avg	0-1-h avg	analysis	anl/6-h fcst+	0-6-h avg	analysis
Total column water vapor	kg m <sup>-2</sup>	analysis	12-h fcst+	6-h fcst	6-h fcst		anl/6-h fcst+	analysis	
Total column condensed water	kg m <sup>-2</sup>					analysis	anl/6-h fcst+	analysis	
Q850	kg kg <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis		anl/12-h fcst+	analysis	analysis
T850	K	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
U850	m s <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
V850	m s <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
H850	m	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
Q700	kg kg <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis		anl/12-h fcst+	analysis	analysis
T700	K	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
U700	m s <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
V700	m s <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
H700	m	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
Q500	kg kg <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
T500	K	analysis	12-h fcst+	analysis	analysis		anl/12-h fcst+	analysis	analysis
U500	m s <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
V500	m s <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
H500	m	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
Q300	kg kg <sup>-1</sup>	analysis	12-h fcst+	analysis	analysis		anl/12-h fcst+	analysis	analysis
T300	K	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis

TABLE A1. (Continued)

Description	Units	BMRC	CPTEC	ECPC-RII	ECPC-SFM	JMA	MSC	NCEP	UKMO
U300	$\text{m s}^{-1}$	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
V300	$\text{m s}^{-1}$	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
H300	m	analysis	12-h fcst+	analysis	analysis	analysis	anl/12-h fcst+	analysis	analysis
Q200	$\text{kg kg}^{-1}$	analysis	12-h fcst+	analysis	analysis	analysis		analysis	analysis
T200	K	analysis	12-h fcst+	analysis	analysis	analysis		analysis	analysis
U200	$\text{m s}^{-1}$	analysis	12-h fcst+	analysis	analysis	analysis		analysis	analysis
V200	$\text{m s}^{-1}$	analysis	12-h fcst+	analysis	analysis	analysis		analysis	analysis
H200	m	analysis	12-h fcst+	analysis	analysis	analysis		analysis	analysis

Further details, including descriptions of the forecast times and missing variables, are included in Table A1.

### b. Interpolation (step 2)

To compare and produce an ensemble, a common grid must be defined. Because most operational analyses are near or going to  $\sim 100$ -km spatial scales, a grid on the order of  $1^\circ$  latitude and longitude was desirable. Also, many data products (GPCP and the existing climate reanalyses data) use a regular latitude–longitude coarse grid ( $2.5^\circ$ , in the case of GPCP). 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  $1.25^\circ$  latitude  $\times$   $1.25^\circ$  longitude ( $144 \times 288$  grid points), with the (1, 1) center point located at  $89.375^\circ\text{S}$ ,  $179.375^\circ\text{W}$ .

The native grid from each of the NWPCs supplied to CEOP was interpolated to the common grid using a freely available routine (available online at <http://www.opengrads.org/re/> function). In the cases where the native grid is finer than  $1.25^\circ \times 1.25^\circ$ , box averaging was used. In the cases where the native grid is coarser than  $1.25^\circ \times 1.25^\circ$ , bilinear interpolation was used. No other filtering or screening of the gridded data was applied (except for some belowground heights; details in the appendix, section 1c). At the end of this step, the data from each NWPC were 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. Because not all of the centers provided all variables, the ensemble averaging was done with those centers that did provide the given variable. If any data were 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 850 and 700 hPa, 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 850 hPa or less than 700 hPa, respectively). This masking was also performed for the BMRC 6-hourly data but not for the other individual NWPCs. The flowchart of decisions used for each variable during the creation of the MAC ensemble at each of the 3292 6-hourly times is shown in Fig. A1.

The individual center's interpolated variable is also provided with the MAC, so that it will be apparent when data are 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 time step 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 in the same format as the individual NWPCs described at the end of the appendix (step 2).

### d. Daily and monthly averages (steps 4–6)

The daily average of the ensemble mean was the simple average of the 0000, 0600, 1200, and 1800 UTC 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 were only available at 0000 and 1200 UTC, and the daily average is just the average of these 2 times. Also, for each dataset, if at least 1 of the 4 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 Fig. 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 height field was missing for

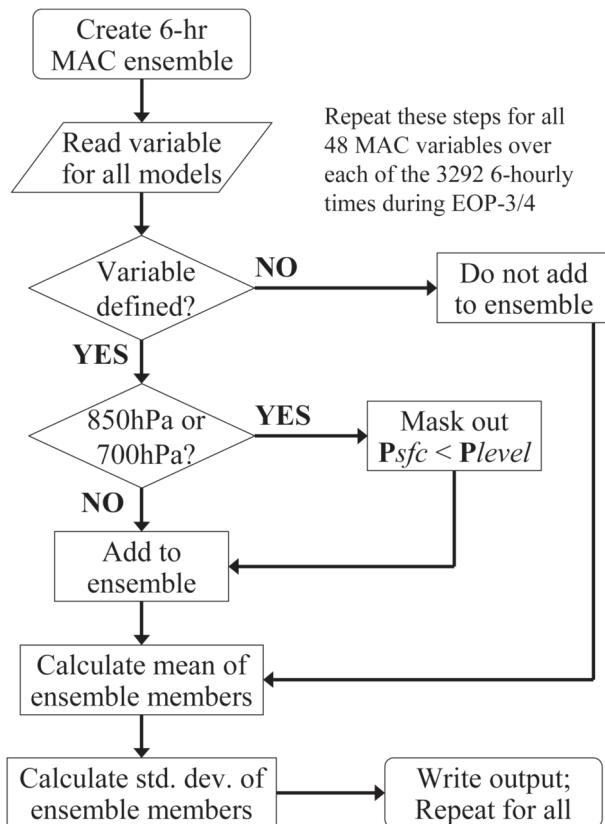


FIG. A1. Flowchart depicting the decisions and calculations performed in the generation of the MAC 6-h time series.

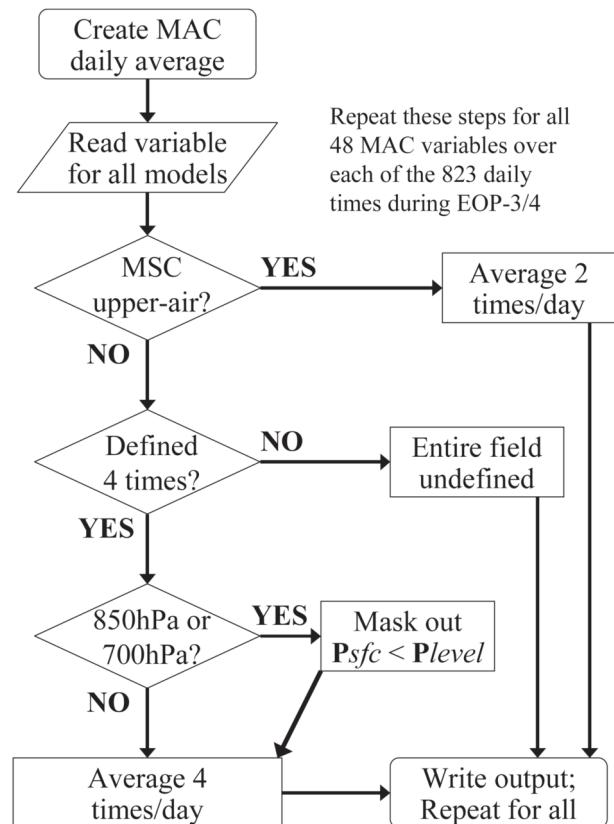


FIG. A2. As in Fig. A1, but for the creation of daily averages.

1200 UTC only for one center. The 6-hourly ensemble means will include the 0000, 0600, and 1800 UTC times for this center; therefore, 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 standard deviation.

The monthly average of the ensemble mean was the simple average of all the times in the month. For the individual NWPCs, the monthly average was calculated differently. First, all the 0000 UTC times during the month were averaged and then the times of 0600, 1200, and 1800 UTC. Next, these 4 times were summed and divided by four. This method was done to minimize the effect of an individual missing time on the monthly average. For example, if a 0600 UTC time was missing for a variable such as downward surface radiation on a single date, then this missing time would have a noticeable effect on the monthly average. If the similar times were averaged first, this problem is reduced; however, averaging 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 un-

defined (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), the CPTEC data (only for May 2003, as a result of missing data), and the MSC data (only 0000 and 1200 UTC data were available). Note the because of numerous missing forecasts during December 2002 for UKMO, the monthly values are all undefined for this month only. The flowchart for the monthly averages is shown in Fig. 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 out the gridded data 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 to the community. The resulting

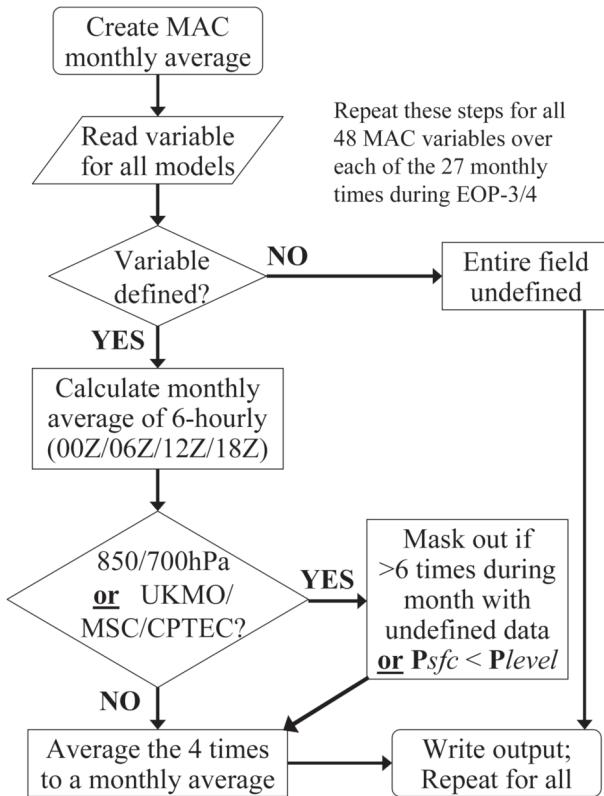


FIG. A3. As in Fig. A1, but for the generation of monthly averages.

binary (or NetCDF) output size is roughly 284 GB (about 134 GB in GRIB1). Each file contains the variables 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 MAC data are available from the NASA Goddard Data Information Services Center (DISC) and the Max Planck Institute Model and Data Center.

#### REFERENCES

- Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167.
- Bélair, S., J. Mailhot, C. Girard, and P. Vaillancourt, 2005: Boundary layer and shallow cumulus clouds in a medium-range forecast of large-scale weather system. *Mon. Wea. Rev.*, **133**, 1938–1960.
- , M. Roch, A.-M. Leduc, P. A. Vaillancourt, S. Laroche, and J. Mailhot, 2009: Medium-range quantitative precipitation forecasts from Canada's new 33-km deterministic global operational system. *Wea. Forecasting*, **24**, 690–708.
- 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. Meteor. Soc.*, **83**, 1495–1499.
- , 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. *J. Meteor. Soc. Japan*, **85A**, 205–228.
- , J. Chen, F. R. Robertson, and R. F. Adler, 2008: Evaluation of global precipitation in reanalyses. *J. Appl. Meteor. Climatol.*, **47**, 2279–2299.
- Chou, S. C., J. A. Marengo, C. P. Derezynski, 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. *J. Meteor. Soc. Japan*, **85A**, 25–42.
- Compo, G. P., J. S. Whittaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, **87**, 175–190.
- Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1393.
- Cox, S. J., P. W. Stackhouse Jr., S. K. Gupta, J. C. Mikovitz, T. Zhang, L. Hinkelmann, M. Wild, and A. Ohmura, 2006: The NASA/GEWEX surface radiation budget project: Overview and analysis. Preprints, *12th Conf. on Atmospheric Radiation*, Madison WI, Amer. Meteor. Soc., 10.1. [Available online at <http://ams.confex.com/ams/pdfpapers/112990.pdf>.]
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, 2006: GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.*, **87**, 1381–1395.
- Ebisuzaki, W., S. Worley, and C.-F. Shih, 2007: *NCEP-NCAR Annual Analyses*. Vol. 1, University Corporation for Atmospheric Research, DVD. [Available online at [http://dss.ucar.edu/pub/reanalysis/annual\\_dvd.html](http://dss.ucar.edu/pub/reanalysis/annual_dvd.html).]
- GCWMB, 2003: The GFS atmospheric model. NCEP Office Note 442, 14 pp.
- Guo, Z., P. A. Dirmeyer, X. Gao, and M. Zhao, 2007: Improving the quality of simulated soil moisture with a multi-model ensemble approach. *Quart. J. Roy. Meteor. Soc.*, **133**, 731–747, doi:10.1002/qj.48.
- 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. *J. Meteor. Soc. Japan*, **85A**, 1–24.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kanamitsu, M., and Coauthors, 2002a: NCEP dynamical seasonal forecast system 2000. *Bull. Amer. Meteor. Soc.*, **83**, 1019–1037.
- , W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002b: NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643.
- Kato, H., M. Rodell, F. Beyrich, H. Cleugh, E. van Gorsel, H. Z. Liu, and T. P. Meyers, 2007: Sensitivity of Land Surface Simulations to Model Physics, Land Characteristics, and Forcings, at four CEOP Sites. *J. Meteor. Soc. Japan*, **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 Coauthors, 2001: Real-time multianalysis-multimodel superensemble forecasts of precipitation using

- TRMM and SSM/I products. *Mon. Wea. Rev.*, **129**, 2861–2883.
- Lin, B., P. W. Stackhouse Jr., P. Minnis, B. A. Wielicki, Y. Hu, W. Sun, T.-F. Fan, and L. M. Hinkelman, 2008: Assessment of global annual atmospheric energy balance from satellite observations. *J. Geophys. Res.*, **113**, D16114, doi:10.1029/2008JD009869.
- Loeb, N. G., K. J. Priestley, D. P. Kratz, E. B. Geier, R. N. Green, B. A. Wielicki, P. O. Hinton, and S. K. Nolan, 2001: Determination of unfiltered radiances from the clouds and the earth's radiant energy system instrument. *J. Appl. Meteor.*, **40**, 822–835.
- Meinke, I., J. Roads, and M. Kanamitsu, 2007: Evaluation of RSM-Simulated Precipitation during CEOP. *J. Meteor. Soc. Japan*, **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. *J. Meteor. Soc. Japan*, **85A**, 43–72.
- Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *J. Meteor. 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. *J. Meteor. Soc. Japan*, **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. Hydrometeorol.*, **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**. [Available online at <http://EarthInteractions.org>.]
- , and —, 2007b: The diurnal cycle of water and energy over the continental United States from three reanalyses. *J. Meteor. Soc. Japan*, **85A**, 117–143.
- Shi, W., E. Yarosh, R. W. Higgins, and R. Joyce, cited 2003: Processing daily rain-gauge precipitation data for the Americas at the NOAA Climate Prediction Center. [Available online at [http://ams.confex.com/ams/annual2003/techprogram/paper\\_56719.htm](http://ams.confex.com/ams/annual2003/techprogram/paper_56719.htm).]
- 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 the National Meteorological Center and the European Centre for Medium Range Weather Forecasts. *Bull. Amer. Meteor. Soc.*, **69**, 1047–1057.
- , 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. Hydrometeorol.*, **8**, 758–769.
- , J. T. Fasullo, and J. Kiehl, 2009: Earth's global energy budget. *Bull. Amer. Meteor. Soc.*, **90**, 311–323.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee III, G. L. Smith, and J. E. Cooper, 1996: Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. *Bull. Amer. Meteor. Soc.*, **77**, 853–868.
- 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 Coauthors, 2007: Initial CEOP-based Review of the Prediction Skill of Operational General Circulation Models and Land Surface Models. *J. Meteor. Soc. Japan*, **85A**, 99–116.
- Yin, X., A. Gruber, and P. Arkin, 2004: Comparison of the GPCP and CMAP merged gauge–satellite monthly precipitation products for the period 1979–2001. *J. Hydrometeorol.*, **5**, 1207–1222.
- Yu, L., 2007: Global variations in oceanic evaporation (1958–2005): The role of the changing wind speed. *J. Climate*, **20**, 5376–5390.
- , and R. A. Weller, 2007: Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). *Bull. Amer. Meteor. Soc.*, **88**, 527–539.
- , X. Jin, and R. A. Weller, 2008: Multidecade global flux datasets from the Objectively Analyzed Air–Sea Fluxes (OAFlux) Project: Latent and sensible heat fluxes, ocean evaporation, and related surface meteorological variables. Woods Hole Oceanographic Institution OAFlux Project Tech Rep. OA-2008-01, 64 pp. [Available online at [http://oafux.whoi.edu/pdfs/OAFlux\\_TechReport\\_3rd\\_release.pdf](http://oafux.whoi.edu/pdfs/OAFlux_TechReport_3rd_release.pdf).]