

# An Assimilated Dataset for Earth Science Applications

Siegfried D. Schubert,  
Richard B. Rood, and  
James Pfaendtner  
Data Assimilation Office,  
Laboratory for Atmospheres,  
NASA/Goddard Space Flight Center,  
Greenbelt, Maryland

---

## Abstract

The Data Assimilation Office at NASA's Goddard Space Flight Center is currently producing a multiyear gridded global atmospheric dataset for use in climate research, including tropospheric chemistry applications. The data, which are being made available to the scientific community, are well suited for climate research since they are produced by a fixed assimilation system designed to minimize the spinup in the hydrological cycle. By using a nonvarying system, the variability due to algorithm change is eliminated and geophysical variability can be more confidently isolated.

The analysis incorporates rawinsonde reports, satellite retrievals of geopotential thickness, cloud-motion winds, and aircraft, ship, and rocketsonde reports. At the lower boundary, the assimilating atmospheric general circulation model is constrained by the observed sea surface temperature and soil moisture derived from observed surface air temperature and precipitation fields. The available output data include all prognostic variables and a large number of diagnostic quantities such as heating rates, precipitation, surface fluxes, cloud fraction, and the height of the planetary boundary layer. These variables were chosen to assure a complete budget of the energy and moisture cycles. The assimilated data should also be useful for estimating transport by cumulus processes. The analysis increments (observation minus first guess) and the estimated analysis errors are provided to help the user assess the quality of the data. All quantities are made available every 6 h at the full resolution of the assimilating general circulation model. Selected surface quantities are made available every 3 h.

## 1. Introduction

Climate and weather research has benefited greatly over the past decade from global analyses generated at operational weather prediction centers such as the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Meteorological Center (NMC). These analyses are a tremendous improvement over ungridded and/or single variable data, since one can more easily study the interrelationships between fields while at the same time being spared the difficulties of dealing with irregularly spaced data. The assimilation system performs all the data preprocessing (consistency checks, gridding, quality control) that would otherwise fall on the shoulders of the individual users. The analyses also provide dynamically consis-

tent estimates of the atmosphere in regions with little data coverage as well as estimates of difficult-to-observe quantities such as vertical motion and precipitation.

On the other hand, the assimilated data may suffer from deficiencies in the assimilation system. For example, bias in the atmospheric general circulation model (AGCM) may carry over into the assimilated data in regions of sparse observations. The quality of quantities such as precipitation and heating rates depends on both the veracity of the parameterizations and the quality of the data. Aside from the model bias, inconsistencies between the observations and model introduce imbalances that often appear, at the data insertion times, as an initial spike followed by a "spinup" or "spindown" in the evaporation, precipitation, and radiative fields. The adjustment process generates divergence and vertical motion fields that are unrepresentative of the large-scale motion field. This profoundly affects the quantification of features such as the Hadley circulation. The Hadley circulation and related wave-driven mean meridional circulations are among the most difficult atmospheric features to represent accurately, and they must be quantified if assimilated datasets are to be useful in interseasonal and interannual variability studies.

Recently, efforts have been announced by NMC (Kalnay and Jenne 1991) and ECMWF (Bengtsson and Shukla 1988) to reanalyze historical data using fixed versions of their operational data assimilation systems. These efforts are important in that they will allow researchers to study atmospheric variability and potential short-term climate change without the additional burden of first trying to take into account spurious changes resulting from changes to the assimilation system. The Data Assimilation Office (DAO) in the Goddard Laboratory for Atmospheres (GLA) has as its primary mission the development of the tools necessary to produce research-quality assimilated datasets (National Research Council 1991). The mission of the DAO is unique because it is the quality and the utility of the assimilated data, rather than the forecast, that measures the success of the effort. Ultimately, the assimilation system developed in this effort will be

used to assimilate the satellite and other air- and surface-based measurements of the earth system, which will become available at the turn of the century from the Earth Observing System (EOS) program. The system will combine these measurements into a comprehensive and consistent climate dataset for use by the general research community.

As part of this effort, we are currently producing a "control" or benchmark dataset for the years 1985–89 using version 1 of the Goddard EOS (GEOS-1) assimilation system. The assimilation is currently in production and it is anticipated that the full five years will be complete in the summer of 1994. The assimilation will likely continue beyond 1989 depending on user demand and computer resources. Our objectives in producing these data are twofold. First, we believe that the absence of spinup in the hydrological cycle and the use of a fixed assimilation system will make these data extremely useful for a wide range of climate studies. Second, by making the data available to the larger scientific community we hope to get valuable feedback on the quality and limitations of our assimilated data products. This feedback will help guide future development. Section 2 gives a brief description of the assimilation system including the GEOS-1 AGCM. Section 3 and the Appendix describe the output data. Other ongoing work and our future plans are highlighted in section 4.

## 2. The assimilation system

The main components of the assimilation system are the GEOS AGCM (described below) and a three-dimensional, multivariate optimal interpolation (OI) scheme. An early version of this scheme is documented in Baker et al. (1987). The current version of the system is summarized below.

### *a. The OI scheme*

The tropospheric version of the OI analysis scheme being used for the control assimilation is carried out at a horizontal resolution of  $2^\circ$  latitude by  $2.5^\circ$  longitude at 14 upper-air pressure levels (20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 1000 mb) and at sea level. The analysis increments are computed every 6 h using observations from a  $\pm 3$ -h data window centered on the analysis times (0000, 0600, 1200, and 1800 UTC). The innovation vector (observation minus background forecast) used as input to the OI is computed using a single forecast valid at the analysis time.

The input observational database is one that has been accumulated over the years at GLA. Data for times prior to July 1987 were mostly acquired from the National Center for Atmospheric Research (NCAR).

Since that time, data have been obtained directly from NMC and do not include data that came in after the cutoff time for the operational NMC system. In addition to these two sources, some TOVS (Tiros Operational Vertical Sounder) temperature soundings have come directly from the National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite, Data and Information Service (NESDIS), and gaps have been filled with data from the National Climatic Data Center (NCDC) in Ashville, North Carolina. For the global sea level pressure and near-surface wind analysis over the oceans, data from surface land synoptic reports (sea level pressure only), ships, and buoys are used. The upper-air analyses of height, wind, and moisture incorporate the data from rawinsondes, dropwindsondes, rocketsondes, aircraft winds, cloud tracked winds, and thicknesses from the historical TOVS soundings produced by NOAA NESDIS. The satellite heights are computed using a reference level that depends on the analyzed sea level pressure. The only bogus data used are 1000-mb height observations that are generated above pressure reports from ships. These serve to further couple the surface and upper-air analyses.

The OI scheme is multivariate in geopotential height and winds, and employs a damped cosine function for the horizontal correlation of model prediction error. The height–wind cross-correlation model is geostrophic and scaled to zero at the equator. The multivariate surface analysis scheme over the oceans adopts an Ekman balance for the pressure–wind analysis. The moisture analysis for mixing ratio employs only rawinsonde moisture data. All gridpoint analyses are conducted using up to 75 nearby observations from within a circular data selection cylinder of 1600-km radius.

The assimilation system does not include an initialization scheme and relies on the damping properties of a Matsuno time-differencing scheme to control initial imbalances generated by the insertion of observations. However, the initial imbalances and spinup have been greatly reduced over earlier versions by the introduction of an incremental analysis update (IAU) procedure (Bloom et al. 1991). As shown in Fig. 1, in the IAU procedure, standard OI analysis increments are computed at the analysis times (0000, 0600, 1200, 1800 UTC). The increments are then inserted gradually into the AGCM by rerunning the forecast and adding a fraction of the increment at each model time step. Over the 6-h period centered at the analysis time the full effect of the increment is realized. The assimilation thus effectively consists of a continuous AGCM forecast (heavy solid lines in Fig. 1) with additional heat, momentum, moisture, and mass source terms updated every 6 h from observations. An important difference between the IAU scheme and the usual

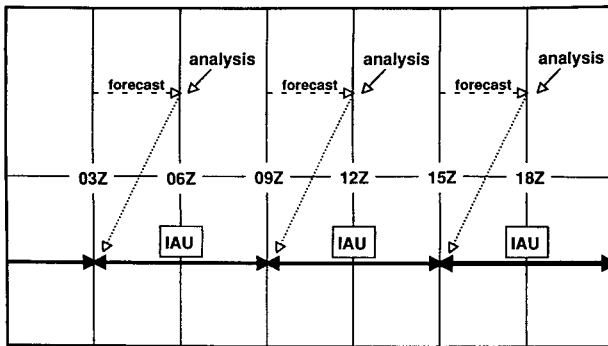


FIG. 1. Schematic of the incremental analysis update (IAU) scheme employed in the GEOS-1 assimilation system. The dashed lines indicate the forecasts that produce a first guess at the analysis times (0000, 0600, 1200, 1800 UTC). Once the OI analysis is performed, the GCM is restarted three hours prior to the analysis time and run for six hours forced by the constant IAU increments (analyzed fields minus first guess). The IAU increments are scaled such the full effect of the increments is introduced over the 6-h interval centered on the analysis times (heavy solid lines).

Newtonian nudging procedure is that the IAU forcing terms are held *constant* over the insertion period, while in Newtonian nudging they are proportional to the difference between a target analysis and the instantaneous current model state. These IAU increments are saved as part of the AGCM diagnostic fields described in section 2b. The quality of the balance achieved when using IAU is illustrated in Fig. 2; the globally averaged precipitation and evaporation fields show a strong diurnal cycle with no evidence of model imbalances (see also Bloom et al. 1991, Fig. 1).

#### b. The GEOS model

The current tropospheric version of the model (GEOS-1) uses the potential enstrophy and energy-conserving horizontal differencing scheme on a C grid developed by Sadourney (1975) and further described by Burrige and Haseler (1977). An explicit leapfrog scheme is used for the time differencing, applying an Asselin (1972) time filter to damp out the computational mode. An eighth-order Shapiro filter is applied to the wind, potential temperature, and specific humidity to avoid nonlinear computational instability. The filter is applied at every step in such a way that the amplitude of the two-grid interval wave would be reduced by half in 2 h. Applying the filter weakly at each time step eliminates the shock that occurred in earlier assimilations by intermittent application of the filter. The model also uses a polar Fourier filter to avoid linear instability due to violation of the Courant–Friedrichs–Lewy condition for the Lamb wave and internal gravity waves. This polar filter, however, is applied only to the tendencies of the winds, potential temperature, specific hu-

midity, and surface pressure. The model's vertical finite-differencing scheme is that of Arakawa and Suarez (1983). The aforementioned dynamics routines are organized into a plug-compatible module called the ARIES/GEOS "dynamical core" (Suarez and Takacs 1993).

The infrared and solar radiation parameterizations follow closely those described by Harshvardhan et al. (1987). In the longwave, water vapor absorption is parameterized as in Chou (1984), the 15- $\mu\text{m}$  band of  $\text{CO}_2$  as in Chou et al. (1983), and ozone absorption as in Rodgers (1968), with modifications suggested by Rosenfield et al. (1987). The shortwave follows Davies (1982), as described in Harshvardhan et al. (1987). Shortwave absorption by water vapor uses a k-distribution approach as in Lacis and Hansen (1974). Cloud albedo and transmissivity for the model layers are obtained from specified single-scattering albedo and cloud optical thickness using the delta-Eddington approximation (Joseph et al. 1976; King and Harshvardhan 1986).

The penetrative convection originating in the boundary layer is parameterized using the relaxed Arakawa–Schubert (RAS) scheme (Moorthi and Suarez 1992), which is a simple and efficient implementation of the Arakawa–Schubert (1974) scheme. Unlike the Arakawa–Schubert scheme, which solves an adjustment problem by considering simultaneous interaction among all possible cloud types, RAS considers only one cloud at a time, and rather than adjusting fully every hour or two, it does a series of partial adjustments that tend to relax the state toward equilibrium. The AGCM also includes a parameterization that

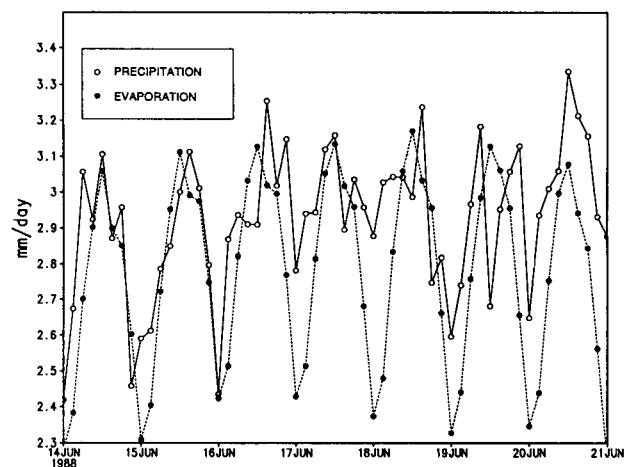


FIG. 2. Globally averaged precipitation and evaporation from the GEOS-1 assimilation plotted every three hours during a one-week period in June of 1988. Both quantities show a pronounced diurnal cycle.

models the evaporation of falling convective rain as described in Sud and Molod (1988). Negative values of specific humidity produced by the finite-differenced advection are filled by borrowing from below.

The planetary boundary layer (PBL) is explicitly resolved in a two- to four-layer region. Wind, temperature, and humidity profiles in an "extended" surface layer (which can be up to 150 m thick) and the turbulent fluxes of heat, moisture, and momentum at the surface are obtained from Monin–Obukov similarity theory by selecting similarity functions that approach the convective limit for unstable profiles and that agree with observations for very stable profiles. Surface roughness lengths are taken as functions of vegetation type over land and as a function of surface stress over water. Turbulent fluxes above the extended surface layer are computed using the second-order closure model of Helfand and Labraga (1988). In this scheme, the turbulent kinetic energy is a prognostic variable, and the remaining second-order moments are diagnosed from it and the atmospheric sounding.

The standard resolution of GEOS-1 is 2° latitude by 2.5° longitude with 20 sigma levels. The sigma levels are distributed to provide enhanced resolution in the planetary boundary layer and at upper levels (see

Fig. 3). The topography used in GEOS-1 was prepared from the 10-min topography obtained from the Navy Fleet Numerical Oceanography Center in Monterey, California. The 2° latitude by 2.5° longitude elevation values were obtained by averaging the high-resolution values (areas with more than 60% water were considered water points) and then applying a Lanczos (1966) filter. The Lanczos filter was designed to remove small-scale structure (it completely removes  $2\Delta X$  waves) while minimizing the Gibbs phenomenon.

This version of the AGCM is run without a land surface model. For the assimilation described here, soil moisture is computed off-line based on a simple bucket model using monthly mean observed surface air temperature and precipitation (Schemm et al. 1992). The snow line and surface albedo are prescribed and vary with the season. The sea surface temperature is updated according to the observed monthly mean values provided by the Climate Analysis Center at NMC and the Center for Ocean, Land and Atmosphere (COLA) at the University of Maryland. Long-term plans call for the incorporation of both land surface and ocean models.

### 3. The assimilated data

The assimilated dataset will cover the period March 1985 through February 1990. The assimilation actually starts 1 January 1985 from an analysis produced by ECMWF, but we allow two months for our system to adjust. The dataset may be extended beyond this time period depending on user demand and computing resources. The assimilation system requires 1 h of Cray YMP CPU time to assimilate 1 day. We are currently in production and assimilating (on average) 4 days per day; this will increase to about 20 assimilated days per day in September of 1993 when we will have access to a Cray C90. We anticipate that the full five years will be completed by the summer of 1994.

In order to understand the data products, it is important to distinguish between the *analysis* and the *assimilation*. As described in section 2, the analysis is performed in the conventional way every 6 h on the standard pressure levels to produce the *analysis increments* (as a function of the difference between the first guess from the AGCM and the observations). The IAU procedure (see section 2a) uses a small fraction of the analysis increments as a forcing (centered on the analysis times and updated every 6 h) in the AGCM to produce the assimilated fields. The assimilated data are thus essentially time continuous.

We have chosen to save the assimilated data on the model's 20 sigma levels to avoid any loss of information due to vertical interpolation. The assimilated

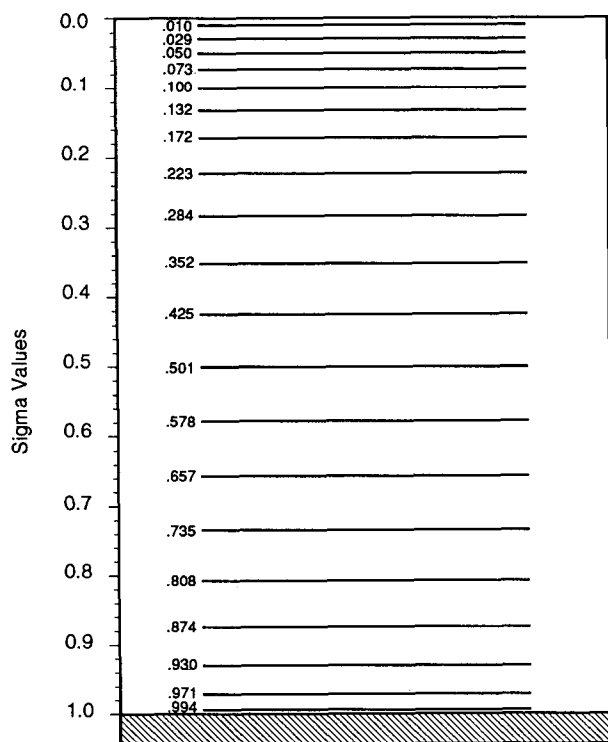


FIG. 3. Sigma levels in the GEOS-1 general circulation model, where sigma is defined as  $(p - p_s) / (p_s - p_s) - 1$ . Here  $p_s$  is the surface pressure and  $p_s = 10$  mb.

lated data are divided into five groups. The first consists of the assimilated model prognostic variables. These are instantaneous fields sampled every 6 h. The second and third sets consist of various diagnostic fields saved as 6-h averages centered on the output times. The second group contains what we consider the "primary" diagnostic fields—that is, those terms needed to compute heat, moisture, and momentum budgets. The third group of "secondary" diagnostics consists of such quantities as clear-sky radiative heating rates and cloud mass flux. The fourth group consists of various single-level, surface-integrated, and vertically integrated fields. These are saved every 3 h accumulated over the previous 3 h. The fifth group consists of the gridded analysis increments used to force the model during the assimilation as part of the IAU procedure described in the previous section. The increments are constant over the 6-h period centered on the output times. These increments should prove valuable for researchers interested in studying the influence of the observations and model bias. Our choice of providing most of the diagnostic quantities as 6-h averages is consistent with the IAU procedure that operates over the same 6-h time period with a fixed analysis increment. However, we have chosen to provide the prognostic variables as instantaneous quantities to allow these to be used consistently as initial conditions for an AGCM. The Appendix includes a complete list of all the quantities saved.

A subset of the aforementioned assimilated data consisting of groups 1, 2, and 4 (see above) and the estimated analysis errors (see below) saved on pressure levels in a time series format. Here each variable is stored by month in a single file. This is different from the original output that groups many variables together in daily files. The size of the entire time series dataset is 150 GB, or about 117 MB per month for each upper-level field. We anticipate that, initially, the pressure level data time series data will be in highest demand; these will be made available through the Goddard Distributed Active Archive Center (DAAC) as quickly as possible. Issues concerning the method of shipping the data, the media, and whether the data should be packed are still being worked out with the Goddard DAAC.

The variables that are stored were chosen to assure that all components of the energy and moisture budget could be calculated. These are fundamental

### Forecast error 165° W - 105°W, 60°S - 36°S

E7328 GLA retrievals 22 Jul - 20 Aug 1979

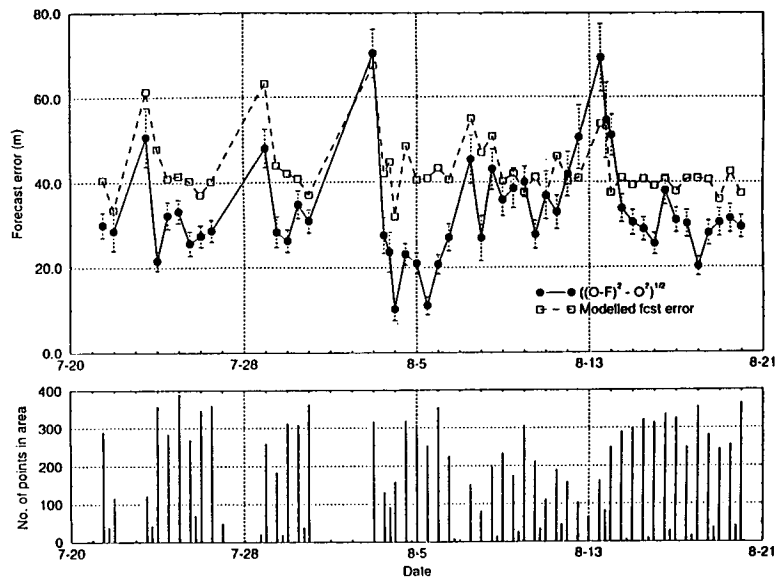


FIG. 4. Upper panel: the area-averaged first-guess 500-mb height error estimates for a test case in a region in the South Pacific for 1979. The solid curve shows the errors computed directly from the satellite observations and forecasts. The dashed curve shows the estimate of the errors using the analysis error-growth model. Bottom panel: a bar graph of the number of observations in this region during each time period.

measures of our understanding of the atmospheric system, and the assimilation process allows all components to be estimated. Surface quantities are stored eight times a day in order to resolve diurnally varying quantities such as precipitation. Surface wind data are also needed at this frequency to drive ocean models. We are also trying to save adequate information to perform tropospheric transport experiments; however, the details of representing transport by cumulus convection are not well known.

An important quantity is the OI estimate of the analysis error variance. While this is a rather crude estimate of the true analysis error, it is valuable for assessing the impact of poor data coverage. Maps of these error estimates provide a detailed picture of the regions that have a recent history of poor data coverage. This is illustrated in Fig. 4 (upper panel), which shows the area-averaged first-guess 500-mb height error estimates for a test case in a region in the South Pacific for 1979. The solid curve shows the errors computed directly from the satellite observations and forecasts. The dashed curve shows the estimate of the errors using the analysis error-growth model. The similarity of the two curves gives us confidence that the empirical model is providing qualitative estimates of the errors. The bottom panel in Fig. 4 shows a bar

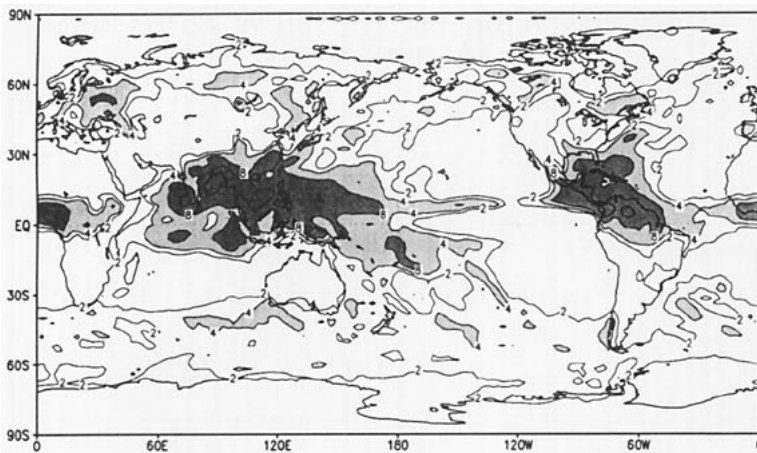


FIG. 5. Average June 1988 precipitation rate from the GEOS-1 assimilation. Contour intervals are 2, 4, 8, 16, and 32 mm day<sup>-1</sup>. Values greater than 4 are shaded.

graph of the number of observations in this region during each time period. The largest peaks in the error curves correspond well with the time periods with few observations.

In the following, we present selected results from a test case assimilation for May/June 1988. Figure 5 shows the June precipitation. The field looks quite reasonable, exhibiting the well-known tropical maxima over South and Central America, Indonesia, and Africa, which are north of the equator during this time of the year. Also evident are the unusually low precipitation amounts over much of the United States associated with the North American drought during this time period. Figures 6 and 7 show some comparisons with observations. Figure 6 compares the precipitation for 0000 UTC 1 June 1988 with an infrared (IR) image from a NOAA geostationary satellite. The correspondence between the high clouds in the IR image and the precipitation from the assimilation is quite good. Of particular note are the precipitation band extending from the tropics into the North Atlantic, another band extending from South America into the South Atlantic, and the precipitation over the central United States. A more quantitative comparison is shown in Fig. 7, which compares the precipitation over the United States for two weeks during May and June 1988 from the assimilation with the corresponding fields from station reports (NOAA 1988). The assimilation appears to do a very credible job of reproducing the major wet and dry regions throughout the May/June 1988 drought period. The aforementioned results give us some confidence that 1) the GCM's convective parameterization is reasonable (most of the precipitation shown is convective), 2) the observations provide information to the precipitation estimates, and 3) the

estimates are not dominated by adverse reactions (shocks) resulting from the introduction of the observations.

Figure 8 shows the diurnal cycle in the wind speed at a point near Fort Worth, Texas, averaged over May and June 1988. Figure 8a is from a May/June simulation with the GEOS-1 AGCM, and Fig. 8b is from the May/June 1988 assimilation with the GEOS-1 system. The profiles show the development of the well-known Great Plains nocturnal low-level jet (e.g., Bonner 1968). During May/June 1988, the jet (Fig. 8b) is clearly evident but considerably weaker than the simulated jet (Fig. 8a). The results of the simulation, analyzed by Helfand and Schubert (1993), highlight the importance of the low-level jet for the moisture budget over the United States. It is a key climatological feature that must be resolved in order to obtain accurate estimates of the moisture fluxes in this region. The fact that the low-level jet is resolved in the assimilation (but has a different signature from the simulation) gives us confidence that 1) the planetary boundary layer parameterization is performing well, 2) the resolution of our system is adequate, and 3) the observations are providing information about the nature of the jet. The behavior of the low-level jet during the 1988 North American drought is currently under investigation.

In order to obtain some measure of the accuracy of the assimilated fields it is useful to compare the DAO fields with those of other assimilation systems. An example of such a comparison is shown in Figs. 9 and 10, which compare the 1988 June 500-mb height field assimilated by the GEOS-1 system with the corresponding ECMWF product generated by their system in 1988. The mean squared height ( $h$ ) differences may be written as

$$\overline{(h_D - h_E)^2} = \overline{(h'_D - h'_E)^2} + (\overline{h_D} - \overline{h_E})^2, \quad (1)$$

where the bar denotes a time average and the primes denote a deviation from the time average. The subscript  $E$  denotes ECMWF and the subscript  $D$  denotes DAO. The first term on the right-hand-side (rhs) of (1) is the mean squared random difference, and the second term on the rhs of (1) is the squared systematic difference. The random difference is a measure of the uncertainty in the assimilated fields. Nonzero systematic differences suggest bias (with respect to the true state) in one or both of the fields. Note that the lack of a systematic difference does not imply an absence of

systematic error in the two products since both fields may have the same bias.

Figure 9 shows the DAO mean height field and the systematic difference between the DAO and ECMWF height fields. The differences are generally less than 20 m everywhere except in the high latitudes of the Southern Hemisphere. In data-dense regions, the values are less than 10 m. The largest differences occur over Antarctica; this may be partly due to vertical interpolation differences (note also the differences over Greenland and the Himalayas). There also appears to be a general “background” difference with the DAO heights tending to be about 10 m lower than the ECMWF heights (see Fig. 9b). Figure 10 shows the standard deviation of the DAO 500-mb height field (Fig. 10a) and the root-mean-square random differences [the square root of the first term on the rhs of (1), Fig. 10b]. The random differences are (like the systematic differences) largest in the Southern Hemisphere high latitudes. In the Northern Hemisphere (NH), the differences are generally less than 10 m over the landmasses and between 10 and 20 m over the oceans. The magnitude of the NH random errors is not inconsistent with what is generally accepted as the magnitude of typical analysis errors. On the other hand, the differences in the Southern Hemisphere high latitudes exhibit magnitudes that are a substantial fraction of the variability in that region (see Fig. 10a). Since both systems incorporate satellite observations, it is likely that the lack of adequate surface pressure information needed to convert the satellite thickness values to geopotential heights is contributing to these large differences.

#### 4. Other ongoing and future work

The DAO is pursuing long-range data assimilation goals using estimation theory based on Kalman filtering (KF). Estimation theory provides a framework in which to develop algorithms that are suboptimal compared to the KF yet are improvements over OI and are economically affordable (Cohn 1993; Todling and Cohn 1993). Such an approach provides an environment in which to incrementally develop systems with a maximum amount of internal physical and chemical consistency. One of the strengths of the DAO effort, and retrospective analyses in general, is that there is no operational constraint to only being able to use observations available up to the analysis time. We can wait for data that are reported late (often from important remote locations) as well as use data collected long after the analysis time. In addition, observations that define the boundary conditions can be obtained. Using data collected after the analysis time more

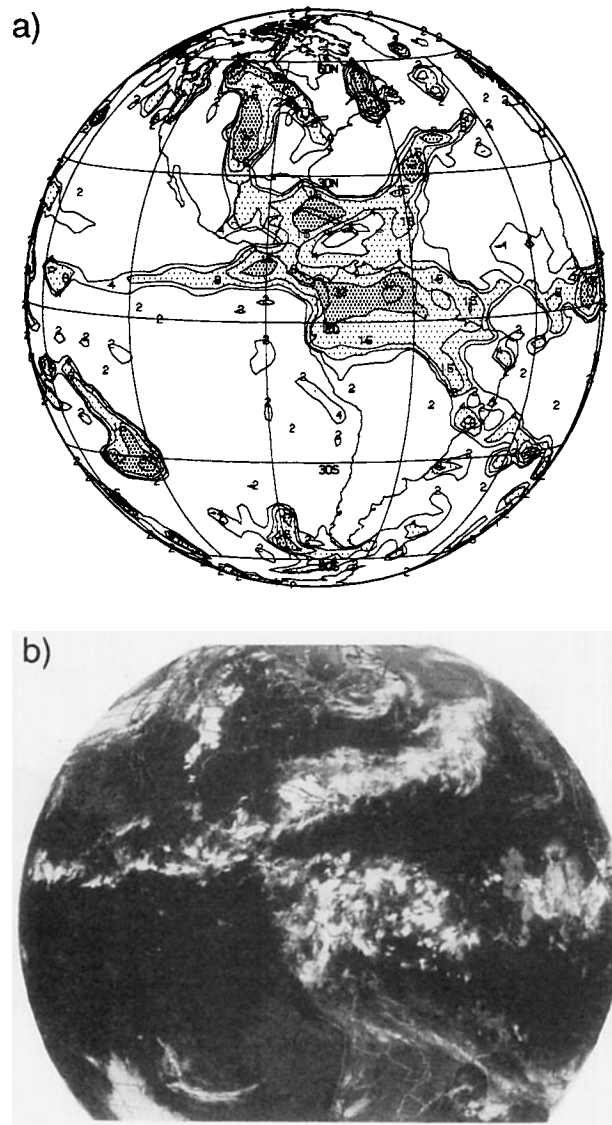


FIG. 6. (a) Precipitation rate at 0000 UTC on 1 June 1988 from the GEOS-1 assimilation. Contour intervals are 2, 4, 8, 16, and 32 mm day<sup>-1</sup>. Values greater than 4 are shaded. (b) Infrared image at 0000 UTC on 1 June 1988 from a NOAA geostationary satellite.

uniformly brackets the analysis time with observations and in theory should improve the quality of the analysis. Estimation theory provides an ideal framework for developing these retrospective analysis techniques. We also intend to exploit advanced numerical techniques in both model and analysis development to reduce the numerical artifacts of the system.

To guide the development effort we intend to produce datasets and apply them to generalized earth science problems. To gain the needed diversity, we intend to distribute the datasets to a wide range of users. An important component of this effort is DAO involvement in field campaigns to study particular earth system processes. Therefore, in addition to the

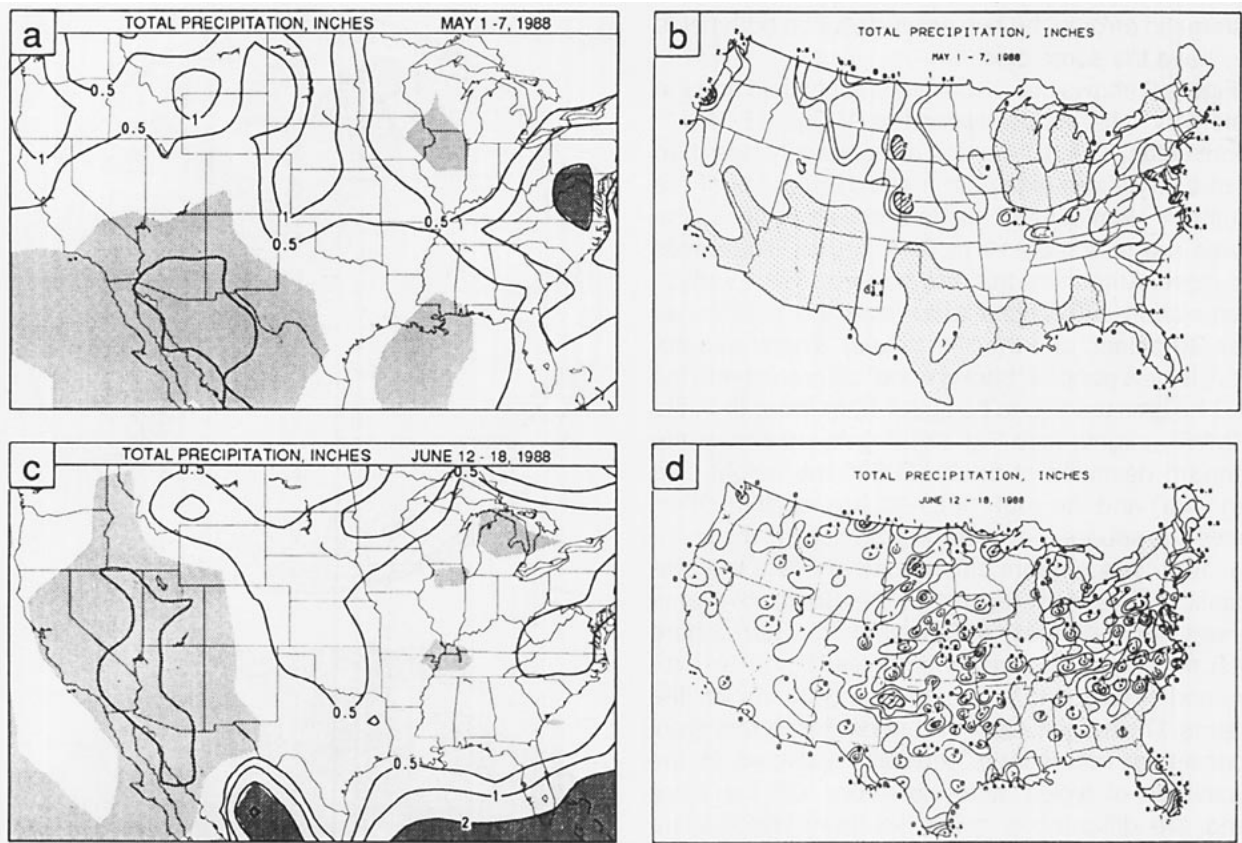


FIG. 7. (a) The precipitation amounts over the United States during the first week of May 1988 from the GEOS-1 assimilation. Light shading indicates values less than 0.1 inches. Heavy shading indicates values greater than 2 inches. (b) The precipitation amounts over the United States during the first week of May 1988 from station reports (NOAA/USDA Joint Agricultural Weather Facility) published in the *Weekly Weather and Crop Bulletin*. Stippled areas indicate no rain. Hatching indicates values greater than 2 inches, and cross-hatching indicates values greater than 4 inches. (c) Same as (a) except for 12–18 June. (d) Same as (b) except for 12–18 June.

baseline assimilation described above, we plan to carry out shorter assimilations supporting several special observing periods and testing the sensitivity of the results to the analysis system formulation and observational data. These include experiments employing the latest HIRS2 physical retrievals produced in-house (Susskind et al. 1984) and assimilations in support of the Tropical Ocean Global Atmosphere subproject Coupled Ocean–Atmosphere Response Experiment, and subprojects of the Global Energy and Water Cycle Experiment.

An important activity in the DAO is stratospheric assimilation. There are currently two configurations to the GEOS-1 data assimilation system. One is the tropospheric system described above. The other is a stratospheric system that employs a 46-level (top pressure of 0.1 mb) version of the GEOS model. The analysis is carried out at 19 pressure levels extending to 0.4 mb. Due to the computational constraints, this system is typically run at a horizontal resolution of  $4^\circ$  latitude by  $5^\circ$  longitude. This system, which has been

published under the name STRATAN, has been used in numerous stratospheric chemistry and meteorological studies (e.g., Rood et al. 1992). Recently, forecasts and analyses from STRATAN have been supplied operationally to the Stratospheric Photochemistry, Aerosols, and Dynamics Expedition. The DAO products were used for scientific flight planning and will be used in a wide range of interpretive studies. Previously, an older version of STRATAN was used in the Airborne Arctic Stratosphere Expedition in 1989. The participation in these missions has pushed the quality of the stratospheric analysis forward tremendously, and it is this experience that motivates DAO participation in a wider range of applications. The DAO also plans to assimilate data from the Upper Atmosphere Research Satellite (UARS), which has temperature, wind, and constituent measurements. In many ways UARS provides a prototype for the EOS effort, making UARS a very important mission to study prior to the launch of EOS satellites.

Various improvements are planned for both the

assimilating AGCM and the analysis scheme. In the short term, the model improvements include a more accurate moisture advection scheme, further improvements to the PBL, convection, and radiation parameterizations, including the introduction of a land surface model and a cloud liquid water scheme. The OI scheme is currently being modified to use a variational approach to solve the OI analysis equations; the method solves the system globally, thus eliminating the need to perform data selection. Longer-term developments include the introduction of semi-Lagrangian dynamics (Bates et al. 1993), a coupled ocean model, and a simplified Kalman filtering scheme.

*Acknowledgments.* We wish to thank Ms. L. Rumburg for her help in drafting the figures. This work is supported by NASA's Earth Observing Systems (EOS) projects on four-dimensional data assimilation and computing, and the Earth System Modeling and Global Analysis Branch at NASA Headquarters.

## Appendix: Output files

The following is a list of all the quantities being saved during the assimilation. For convenience the data have been divided into the five groups listed below. The first group (dataset 1) consists primarily of the assimilated prognostic variables on model sigma levels. These are instantaneous fields sampled every 6 h. The second and third groups (datasets 2 and 3) consists of various diagnostic fields from the assimilation on the model sigma levels. These are 6-h averages centered on the output times (0000, 0600, 1200 UTC, etc.). The fourth group (dataset 4) consists of various single-level, surface-integrated or vertically integrated diagnostic fields. These are saved eight times daily at 0000, 0300, 0600 UTC, etc., accumulated over the previous 3 h. The fifth group (dataset 5) consists of the analysis increments (four times daily) on the model sigma levels.

While all the data are available to the community, we believe that a subset of the aforementioned sigma-level quantities consisting of datasets 1, 2, and 4 and the OI estimates of the analysis errors, saved on pressure levels (1000, 950, 900, 850, 800, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20 mb), and in a time series format, will be the preferred datasets for many users. In the time series format each file contains one month of a single (four times daily) upper-level quantity. The single-level quantities (dataset 4) are also subdivided into

several smaller groups. This should make it easier to analyze longer records by keeping the file sizes more manageable. These data will be made available through the Goddard DAAC as quickly as possible. For more information and for data requests, please contact Head, Data Assimilation Office, Code 910.3, NASA/GSFC, Greenbelt, MD 20771; (301) 286-8203; email: data@dao.gsfc.nasa.gov.

### Dataset 1: Four times daily/instantaneous

1. Surface geopotential heights ( $m s^{-2}$ )
2. Surface albedo (0–1)
3. Ground wetness (0–1)
4. Surface pressure–pressure at top (mb)
5. Ground temperature (K)
6. Sea level pressure (mb)
7. Water (1), land (2), permanent ice (3), and sea ice (4) flags

### Upper-air quantities (on the model sigma levels)

1. Zonal wind ( $u$ ) ( $m s^{-1}$ )
2. Meridional wind ( $v$ ) ( $m s^{-1}$ )
3. Perturbation geopotential heights ( $m s^{-1}$ )<sup>2</sup>
4. Temperature ( $T$ ) (K)
5. Specific humidity ( $q$ ) ( $g kg^{-1}$ )
6. Turbulent kinetic energy ( $m s^{-1}$ )<sup>2</sup>

### Dataset 2: Four times daily/6-h average

1. Surface geopotential heights ( $m s^{-1}$ )<sup>2</sup>
2. Surface pressure–pressure at top (mb)

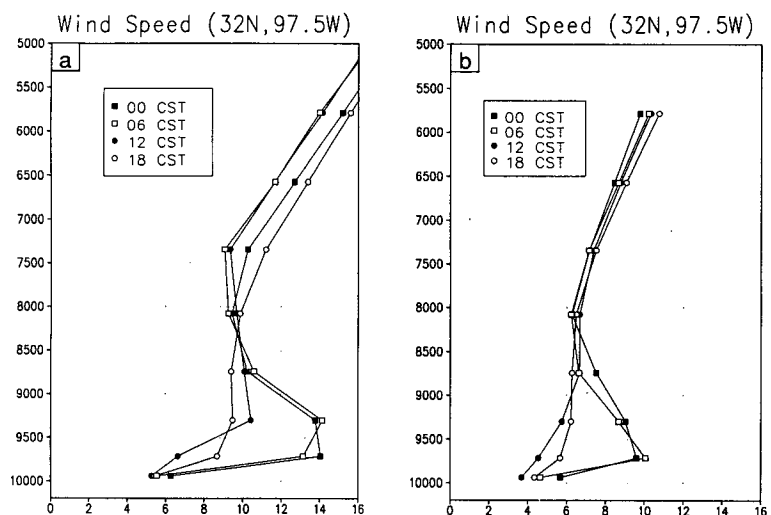


FIG. 8. (a) Diurnal cycle in wind speed at a point near Fort Worth, Texas, from a 2-month simulation with the GEOS-1 GCM. Values are the averages over the 2-month (May/June) period. (b) Same as (a) except for the GEOS-1 assimilation for May and June 1988. Units are meters per second. Vertical axis shows approximate pressure levels ( $\times 10$ ) computed from the sigma levels assuming a surface pressure of 1000 mb.

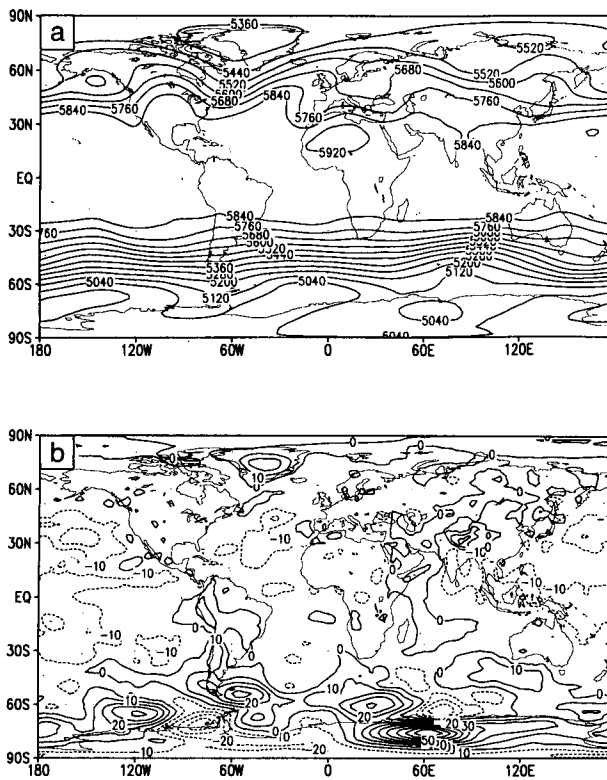


Fig. 9. (a) Time-mean height of the 500-mb pressure surface for 1–25 June 1988 from the GEOS-1 assimilation. (b) The time-mean (1–25 June 1988) difference (DAO–ECMWF) between the heights computed with the GEOS-1 assimilation and those computed with the ECMWF system. Units are in meters.

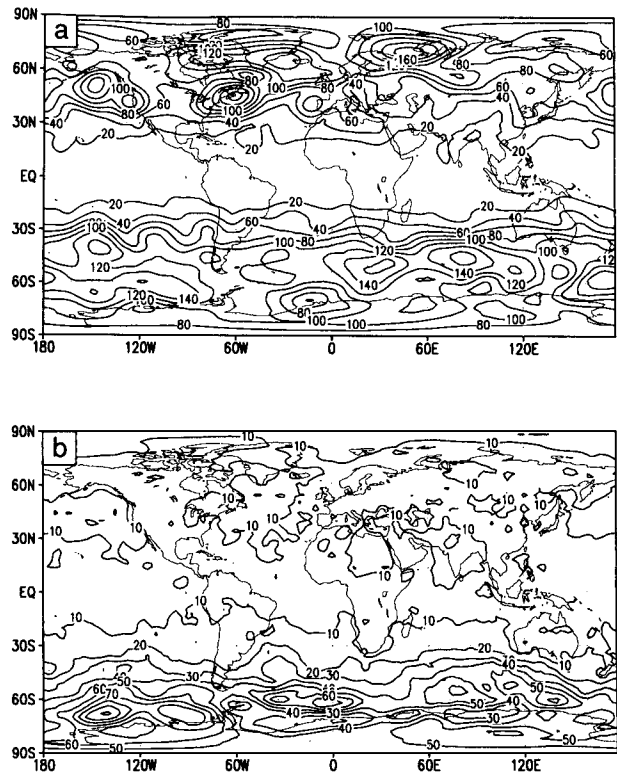


Fig. 10. (a) The standard deviation of the 500-mb height field for 1–25 June 1988. The heights are from the GEOS-1 data assimilation during this time period. (b) The root-mean-square difference between the GEOS-1 heights and the ECMWF heights (after removing their respect time means) for 1–25 June 1988. See text for details. Units are in meters.

3. Water (1), land (2), permanent ice (3), and sea ice (4) flags

*Upper-air quantities* (on the model sigma levels)

1. Vertical velocity ( $dp/dt$ ) ( $\text{mb day}^{-1}$ )
2. Temperature changes due to turbulence ( $\text{deg day}^{-1}$ )
3. Temperature changes due to moist processes ( $\text{deg day}^{-1}$ )
4. Longwave heating rates ( $\text{deg day}^{-1}$ )
5. Shortwave heating rates ( $\text{deg day}^{-1}$ )
6.  $u$ -momentum changes due to turbulence ( $\text{m s}^{-1} \text{day}^{-1}$ )
7.  $v$ -momentum changes due to turbulence ( $\text{m s}^{-1} \text{day}^{-1}$ )
8. Moisture changes due to turbulence ( $\text{g kg}^{-1} \text{day}^{-1}$ )
9. Moisture changes due to moist processes ( $\text{g kg}^{-1} \text{day}^{-1}$ )

**Dataset 3:** *Four times daily/6-h average*

1. Surface geopotential heights ( $\text{m s}^{-1}$ )<sup>2</sup>
2. Water (1), land (2), permanent ice (3), and sea ice (4) flags

*Upper-air quantities* (on the model sigma levels)

1. Clear-sky longwave heating rates ( $\text{deg day}^{-1}$ )
2. Clear-sky shortwave heating rates ( $\text{deg day}^{-1}$ )
3. Maximum overlap (convective) cloud fraction (0–1)
4. Random overlap (large-scale) cloud fraction (0–1)
5. Cloud mass flux ( $\text{kg m s}^{-1}$ )
6. Cloud detrainment ( $\text{kg m s}^{-1}$ )

**Dataset 4:** *Eight times daily/3-h average*

1. Surface geopotential heights ( $\text{m s}^{-1}$ )<sup>2</sup>
2. Water (1), land (2), permanent ice (3), and sea ice (4) flags
3. Ground wetness (0–1)
4.  $u$ -momentum surface stress ( $\text{N m}^{-2}$ )
5.  $v$ -momentum surface stress ( $\text{N m}^{-2}$ )
6. Surface flux of sensible heat ( $\text{W m}^{-2}$ )
7. Evaporation ( $\text{mm day}^{-1}$ )
8. Total precipitation ( $\text{mm day}^{-1}$ )
9. Convective precipitation ( $\text{mm day}^{-1}$ )
10. Outgoing longwave radiation ( $\text{W m}^{-2}$ )
11. Clear-sky outgoing longwave radiation ( $\text{W m}^{-2}$ )

12. Net upward longwave radiation at the ground ( $W m^{-2}$ )
13. Clear-sky longwave radiation at the ground ( $W m^{-2}$ )
14. Outgoing shortwave radiation ( $W m^{-2}$ )
15. Clear-sky outgoing shortwave radiation ( $W m^{-2}$ )
16. Net downward shortwave radiation at the ground ( $W m^{-2}$ )
17. Clear-sky shortwave radiation at the ground ( $W m^{-2}$ )
18. Overlap cloud fraction (0–1)
19. Surface drag coefficient for  $T$  and  $q$  ( $m s^{-1}$ )
20. Surface drag coefficient for  $u$  and  $v$  ( $m s^{-1}$ )
21. Surface pressure tendency ( $mb day^{-1}$ )
22. Ustar ( $u^*$ ) ( $m s^{-1}$ )
23. Surface roughness ( $z_0$ ) (m)
24. PBL depth (mb)
25. Ground temperature (deg)
26. Surface (skin) air temperature (deg)
27. Saturation surface specific humidity ( $g kg^{-1}$ )
28. Surface wind speed ( $m s^{-1}$ )
29. Temperature at 2 m (deg)
30. Specific humidity at 2 m ( $g kg^{-1}$ )
31.  $u$  wind at 2 m ( $m s^{-1}$ )
32.  $v$  wind at 2 m ( $m s^{-1}$ )
33. Temperature at 10 m (deg)
34. Specific humidity at 10 m ( $g kg^{-1}$ )
35.  $u$  wind at 10 m ( $m s^{-1}$ )
36.  $v$  wind at 10 m ( $m s^{-1}$ )
37. Vertically integrated specific humidity ( $g kg^{-1}$ )
38. Vertically integrated  $uq$
39. Vertically integrated  $vq$
40. Vertically integrated  $uT$
41. Vertically integrated  $vT$

#### Dataset 5: Four times daily

1. Analysis increments in surface pressure ( $mb s^{-1}$ )
2. Analysis increments in  $u$  ( $m s^{-1} s^{-1}$ )
3. Analysis increments in  $v$  ( $m s^{-1} s^{-1}$ )
4. Analysis increments in  $T$  ( $deg s^{-1}$ )
5. Analysis increments in  $q$  ( $g kg^{-1} s^{-1}$ )

## References

- Arakawa, A., and W. Schubert, 1974: Interaction of a cumulus ensemble with the large-scale environment. Part I. *J. Atmos. Sci.*, **31**, 674–701.
- , and M. J. Suarez, 1983: Vertical differencing of the primitive equations in sigma coordinates. *Mon. Wea. Rev.*, **111**, 34–45.
- Asselin, R., 1972: Frequency filter for time integrations. *Mon. Wea. Rev.*, **100**, 487–490.
- Baker, W. E., S. C. Bloom, J. S. Woollen, M. S. Nestler, E. Brin, T. W. Schlatter, and G. W. Branstator, 1987: Experiments with a three-dimensional statistical objective analysis scheme using FGGE data. *Mon. Wea. Rev.*, **115**, 272–296.
- Bates, J. R., S. Moorthi, and R. W. Higgins, 1993: A global multilevel atmospheric model using a vector semi-Lagrangian finite-difference scheme. Part I: Adiabatic formulation. *Mon. Wea. Rev.*, **121**, 244–263.
- 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.
- Bloom, S. C., L. L. Takacs, and E. Brin, 1991: A scheme to incorporate analysis increments gradually in the GLA assimilation system. *Ninth Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 110–112.
- Bonner, W. D., 1968: Climatology of the low level jet. *Mon. Wea. Rev.*, **96**, 833–850.
- Burridge, D. M., and J. Haseler 1977: A model for medium range weather forecasting—Adiabatic formulation. Tech. Rep. No. 4, European Centre for Medium-Range Weather Forecasts, Bracknell, Berkshire, United Kingdom, 46 pp.
- Chou, M. D., 1984: Broadband water vapor transmission functions for atmospheric IR flux computations. *J. Atmos. Sci.*, **41**, 1775–1778.
- , and L. Peng, 1983: A parameterization of the absorption in the 15  $m CO_2$  spectral region with application to climate sensitivity studies. *J. Atmos. Sci.*, **40**, 2183–2192.
- Cohn, S. E., 1993: Dynamics of short-term univariate forecast error covariances. *Mon. Wea. Rev.*, **121**, 3123–3149.
- Davies, R., 1982: Documentation of the solar radiation parameterization in the GLAS climate model. NASA Tech. Memo. 83961, Goddard Space Flight Center, 57 pp.
- Harshvardhan, R. Davies, D. A. Randall, and T. G. Corsetti, 1987: A fast radiation parameterization for atmospheric circulation models. *J. Geophys. Res.*, **92**, 1009–1016.
- Helfand, H. M., and J. C. Labraga, 1988: Design of a non-singular level 2.5 second-order closure model for the prediction of atmospheric turbulence. *J. Atmos. Sci.*, **45**, 113–132.
- , and S. D. Schubert, 1993: Contribution of the Great Plains low-level jet to the simulated continental moisture budget of the United States. *J. Climate*, submitted.
- Joseph, J. H., W. J. Wiscombe, and J. E. Weinman, 1976: The delta-Eddington approximation for radiative flux transfer. *J. Atmos. Sci.*, **33**, 2452–2459.
- Kalnay, E., and R. Jenne, 1991: Summary of the NMC/NCAR reanalysis workshop of April 1991. *Bull. Amer. Meteor. Soc.*, **72**, 1897–1904.
- King, M. D., and Harshvardhan, 1986: Comparative accuracy of selected multiple scattering approximations. *J. Atmos. Sci.*, **43**, 784–801.
- Lacis, A. A., and J. E. Hansen, 1974: A parameterization for the absorption of solar radiation in the Earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133.
- Lanczos, C., 1966: *Discourse on Fourier Series*. Hafner Publishing, 255 pp.
- Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa–Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120**, 978–1002.
- National Research Council, 1991: Four-dimensional model assimilation of data: A strategy for earth system sciences. *Report from the Panel on Model-Assimilated Data Sets for Atmospheric and Oceanic Research*, National Academy Press, 78 pp.
- NOAA, 1988: *Weekly Weather and Crop Bulletin*. 10 May and 21 June 1988 issues (28 pp. and 24 pp.), NOAA/USDA Joint Agricultural Weather Facility.
- Rodgers, C. D., 1968: Some extensions and applications of the new random model for molecular band transmission. *Quart. J. Roy. Meteor. Soc.*, **94**, 99–102.
- Rood, R. B., J. E. Nielsen, R. S. Stolarski, A. R. Douglass, J. A. Kaye, and D. J. Allen, 1992: Episodic total ozone minima and associ-

ated effects on heterogeneous chemistry and lower stratospheric transport. *J. Geophys. Res.*, **97**, 7979–7996.

Rosenfield, J. E., M. R. Schoeberl, and M. A. Geller, 1987: A computation of the stratospheric diabatic circulation using an accurate radiative transfer model. *J. Atmos. Sci.*, **44**, 859–876.

Sadourney, R., 1975: The dynamics of finite difference models of the shallow water equations. *J. Atmos. Sci.*, **32**, 680–689.

Schemm, J.-K., S. Schubert, J. Terry, and S. Bloom, 1992: Estimates of monthly mean soil moisture for 1979–89. NASA Tech. Memo. 104571, 252 pp.

Suarez, M. J., and L. Takacs, 1993: Documentation of the ARIES/GEOS dynamical core version 2. [Source and documentation

available from M. Suarez, NASA/GSFC, Code 913, Greenbelt, MD 20771 (suarez@maxs.gsfc.nasa.gov).]

Sud, Y., and A. Molod, 1988: The roles of dry convection, cloud-radiation feedback processes, and the influence of recent improvements in the parameterization of convection in the GLA AGCM. *Mon. Wea. Rev.*, **116**, 2366–2387.

Susskind, J., J. Rosenfield, D. Reuter, and M. T. Chahine, 1984: Remote sensing of weather and climate parameters from HIRS2/MSU on TIROS-N. *J. Geophys. Res.*, **89**, 4677–4697.

Todling, R., and S. E. Cohn, 1993: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.*, submitted.



## AMS Ties

The American Meteorological Society is offering its weather-symbol ties at a special holiday price, making them perfect gifts for your favorite atmospheric scientist or enthusiast. The ties are available in three color combinations: maroon with gold symbols, navy with gold symbols, or navy with silver symbols.

\$15 each, or two for \$25 (includes shipping and handling). Holiday offer expires January 15, 1994. Please send prepaid orders to: American Meteorological Society, 45 Beacon St., Boston, MA 02108-3693.

# Holiday Special

