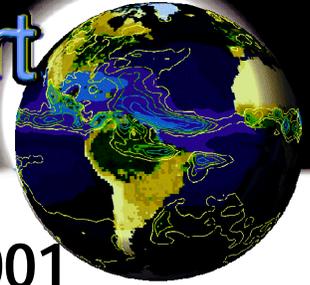


CTR
102

COLA Technical Report



August 2001

A Multi-Decadal Global Land-Surface Data Set of State Variables and Fluxes

Paul A. Dirmeyer, and Liqin Tan

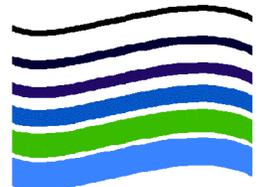
email: dirmeyer@cola.iges.org

4041 Powder Mill Road, Suite 302
Calverton, Maryland 20705-3106 USA

PHONE +1-301-595-7000
FAX +1-301-595-9793

www.iges.org

Center for
Ocean-Land-Atmosphere
Studies



COLA is a Center of the Institute of Global Environment and Society

Abstract

A spatially and temporally continuous global offline land-surface data set (GOLD) has been produced for the period 1979-1999. Its intent is to be used for specifying initial and/or boundary conditions for global climate models, climate variability analyses over land, and applications in hydrology, ecology or biogeochemistry. Hybrid sets of forcing data have been produced by combining the 6-hourly reanalyses with observed monthly data sets to reduce systematic errors in the reanalysis. These data are then used to drive a land surface scheme, producing a data set of surface state variables and fluxes. Both forcing and land surface data sets have been produced at several different resolutions. A brief analysis is presented of the grand mean, mean annual cycle, and seasonal-interannual variability of key surface energy and water balance terms. Calculations have been performed over a number of sub-continental scale regions to examine the climatology of the data set. Comparisons are also made between this 21-year period, and the shorter periods of ISLSCP Initiatives 1 and 2 to see if the ISLSCP periods are statistically representative of the longer term.

1. Introduction

One of the great ironies of the environmental sciences is that there is such a paucity of climatically useful data at the earth's land surface, where virtually all of the human population lives, compared with the sea surface and the atmosphere. There are several reasons for this. The sea surface, and some aspects of the atmosphere, are easily observed by satellite, providing continuous global coverage. Over land, most of the truly useful quantities for climate research and prediction (e.g., soil moisture profiles, snow liquid water content, vegetation density, surface roughness, and evapotranspiration) cannot be well estimated by current satellites. Only surface color (albedos, and by inference some vegetation coverage statistics), radiative skin temperature, and in some cases near surface soil moisture can be observed from space (Sellers et al. 1990). *In situ* measurements for the atmosphere are routinely collected as part of a global observing network (WMO 1996). Stationary buoys, research vessels and ships of opportunity collect and report ocean measurements for regular global analyses. But there are few routine observations of the land surface state, such as soil moisture, temperature, vegetative state and snowpack, and the ones that do exist tend to be highly regionalized or nationalized, and not fed into a global analysis. Regular measurements of surface heat and moisture fluxes are exceedingly difficult to make, and are thus virtually non-existent.

Because there are routine analyses of the near surface atmospheric state, made global through data assimilation into a dynamical atmospheric model, the option exists to calculate a global land surface analysis using an appropriate model of the land surface. Today there exist many so-called "second generation" land surface schemes (LSS) that incorporate at a suitable level of complexity the physical processes that control surface heat and moisture fluxes (Polcher et al. 2000). One or more of these LSSs, given a satisfactorily accurate global data set of land surface parameters (soil and vegetation properties), can be integrated over time, driven by the best near-surface atmospheric analyses, to create a self-consistent global data set of the evolving land surface state. In addition, model estimates of surface fluxes are also produced. They may

not be perfectly accurate, but they will be consistent and satisfy the basic water and energy balances at the surface. Such balance is not usually attained even over small spatial and time scales where intense surface measurements are made (e.g., Betts and Ball 1998).

The Global Land Atmosphere System Study (GLASS) of the Global Energy and Water Cycle Experiment (GEWEX) has identified the need for improved representation of surface and near surface fields and to initialize correctly the slowly varying components of the land surface for seasonal predictions (Polcher et al. 2000). There is little hope that *in situ* observations can provide all of the necessary information for the construction of global data sets, because of the large spatial heterogeneity of land-surfaces and the economic and political barriers to implementing and maintaining such a network. Remote sensing can provide global coverage, but the set of observable parameters is limited to the vegetation cover and the first few centimeters of the soil. In order to obtain the global state of the slower components, the *in situ* and remotely-sensed data will need to be assimilated in one or more LSS.

In recent years there have been several global atmospheric reanalysis projects (the National Centers for Environmental Prediction, Kalnay et al. 1996; the European Centre for Medium-range Weather Forecasts, Gibson et al. 1997; the NASA Data Assimilation Office, Schubert et al. 1993) with the aim of assimilating the most complete and accurate observational data into state-of-the-art atmospheric models to produce consistent multi-year analyses of the state of the atmosphere at sub-diurnal time intervals. Such data sets are ideal for driving a LSS to produce land-surface analyses over the same multi-year periods.

Of course, each reanalysis model is also coupled to a LSS to produce the analyses. Why not simply use those land-surface products? The reason is that the atmospheric models are imperfect, and they

introduce errors into the reanalysis, particularly in regions where there is little or no observational data to assimilate. There exist global observational data sets, often at much lower temporal resolution (typically monthly means), which are of superior quality to the reanalysis averaged over those same time intervals. If these data sets overlap the reanalysis time period, they can be used to scale the reanalysis data so that the monthly means agree (e.g., Mitchell and Lin 1994). Such a hybrid retains the high temporal resolution and global coverage of the reanalysis, but with a better climatology.

Often the observational data sets cover fewer years, or different years than the reanalyses. If the errors in the reanalysis are largely systematic (i.e., they do not vary much from year to year), the scaling of the reanalysis data can be extrapolated to years outside the observed data set, providing at least some degree of improvement over the original reanalysis.

These processes of producing hybrid near-surface atmospheric analyses are applied to create a data set for driving a LSS to produce a multi-decadal global analysis of the land surface. This data set is produced in an uncoupled mode — that is, the LSS is integrated offline without feedbacks to an atmospheric general circulation model. Thus we use the term Global Offline Land-surface Dataset (GOLD) to describe the result of this procedure. Our primary motivation is to produce a data set of land surface state variables for initializing climate model simulations. Since we are using the same LSS as in our climate model, the data are quite consistent for this application, minimizing land surface spin-up issues. The data set itself represents a multi-decadal climatology of the land surface for studying the seasonal and interannual variability of land surface climate. The data set may also be useful as an input for other models requiring continuous spatial and temporal coverage, such as hydrologic or biogeochemical models.

This paper describes the production of the hybrid forcing data set, as well as the resulting land surface climatological data set (GOLD). Section 2 describes the various observational and analysis data sets as well as the hybridization process. Section 3 describes the LSS and the process of integrating it in this context. Section 4 presents a description of the global land-surface data set. A summary is presented in section 5.

2. Atmospheric forcing data

A global offline integration of an LSS requires continuous near surface meteorological data at a high temporal resolution (resolving the diurnal cycle). The best sources of such continuous data are the operational meteorological reanalyses. For this project, we are using the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996; Kistler et al. 2001).

Eight fields are required to drive the typical LSS: near surface air temperature, humidity, wind speed, surface pressure, downward hemispheric shortwave and longwave radiation, total precipitation and the convective component of precipitation. The precipitation variables in the reanalysis are reported as 6-hour totals. The other fields are 6-hour averages. These fields are all part of the standard reanalysis data set.

The reanalysis is very much a model product, and is subject to errors that arise as a result of unavoidable shortcomings due to the model resolution, parameterizations, and inherent physical assumptions. This is particularly true in data-sparse regions, where there are few meteorological observations to assimilate. A number of gridded observational data sets at lower temporal resolution are used to help constrain the reanalysis. These observational data sets help to reduce systematic errors on

the monthly time scale in the reanalysis. However, they do not help us to improve upon the synoptic and diurnal variability depicted in the reanalysis.

A number of regional precipitation data sets are used to provide a regional correction to the monthly mean rainfall of the reanalysis. Over the United States, the rainfall data of Higgins (1996) are used. These data are at 2.5E spatial resolution, and cover the entire time period. There is recent evidence using an expanded observational network that the Higgins data may have a dry bias (Yarosh 2001). No attempt is made to correct for that bias, as that assessment was made with data from more recent years that do not overlap with most of our time period.

Over South America, the gridded data of Webber and Willmott (1998a) are used. These data are at a relatively high 1E spatial resolution, but do not cover the time period after 1990. Over India, the station data of Singh et al. (1992) are used. These data are gridded to the resolution of the offline LSS, using a Cressman (1959) analysis. These data cover the monsoon seasons of 1971 - 1990.

The global Climate Monitoring, Analysis and Prediction (CMAP; Xie and Arkin, 1997) precipitation data set is used to provide a correction to the monthly mean rainfall of the reanalysis over the regions mentioned above during the periods outside the range of those data, and over areas where we do not have regional precipitation data sets. The data set spans the period of 1979-1999. In fact, the span of the CMAP precipitation data was a major factor in defining the time period of the land surface data set. The data are at 2.5E spatial resolution.

For precipitation, a global mask is used so as to apply corrections based on the regional data sets, when available, in preference to the global CMAP data (Figure 1). The errors are removed by

multiplication by a scaling factor based on the ratio of observed monthly rainfall to NCEP/NCAR estimates, rather than by subtraction of the error:

$$[P]_{Y,M,D,T} = \frac{[P_{OBS}]_M}{[P_{NCEP}]_M} [P_{NCEP}]_{Y,M,D,T} \quad (1)$$

To get the corrected forcing data for precipitation, the value at a grid box of one of the reanalysis precipitation terms (total or convective) at a given year, month, day and 6-hour time interval $[R_{NCEP}]_{Y,M,D,T}$ is scaled by the ratio of the monthly mean observed precipitation to the corresponding mean value from the reanalysis for that month. This approach avoids problems of negative values in positive definite quantities with frequent zeroes, such as downward shortwave radiation. It provides the best attainable improvement in the NCEP/NCAR estimates given the lack a long-term sub-monthly observationally-based data set.

No attempt is made to adjust the monthly storm frequency (Liston et al. 1993), as was done for the 6-hourly precipitation estimates in the ISLSCP Initiative I data set (Mitchell and Lin 1994). Nor is any attempt made to adjust the diurnal cycle, which is known to be in error over some regions. For example, Fig. 2 shows two complete diurnal cycles of precipitation averaged over June-Sept 1986-1995 at 35EN from Higgins (1996) hourly data, and the NCEP/NCAR reanalysis. The percentage of daily precipitation falling during the ensuing 6-hour period is shown — large contrasts along the vertical indicate strong diurnal cycles peaking where the shading is darkest. There is a strong diurnal variation in the west, moderate in the east, and very little variation over the Mississippi Valley (near 95EW). The reanalysis shows generally little variation across North America, except for a slightly stronger diurnal cycle in the west. Also, the reanalysis fails to pick up the eastward propagation of storms forming in the afternoon near the front range of the Rockies. In the reanalysis, rainfall is strongly locked to the radiative maximum near local noon. The

only assurance that can be made in the adjustments we apply is that the monthly mean precipitation should agree with the observation data, with some small differences introduced as a result of spatial interpolation.

There is another caveat for the precipitation data. Because some of the regional observational data sets do not cover the entire 21-year period, there may be shifts in the statistics of monthly precipitation at the transitions between global and regional data sets. Similarly, there may be inconsistencies near the lateral boundaries of the regional observed data sets. We have chosen to sacrifice complete consistency in favor of providing the best possible estimates of precipitation at any particular place and time. We consider any gridded, or relatively dense gauge data set that covers at least one third of the 21-year period to be worthy of inclusion in the hybrid precipitation data set.

The surface temperature data of the Climate Anomaly Monitoring System (CAMS; Ropelewski et al. 1985) are used to correct systematic errors in the monthly reanalysis near surface air temperature. The CAMS data have been gridded to a $2.8E \times 1.8E$ grid, with corrections for variations in station altitudes (M. Fennessy, personal communication). Over South America, the gridded temperature data of Webber and Willmott (1998b) are used for the period 1979-1990.

To adjust downward radiation, the diurnal cycle is particularly important. The International Satellite Land Surface Climatology Project (ISLSCP) Initiative I data set (Meeson et al. 1995) contains a monthly-three-hourly global gridded data set of surface downwave and shortwave radiation derived from ISCCP (Rossow and Schiffer 1991). The data are for the 24 months of 1987-1988. Investigation of the errors in the monthly mean downward radiation at the surface from the NCEP/NCAR reanalysis, compared against the eight-year Surface Radiation Budget (SRB; Stackhouse et al. 2000), show that they are very systematic from year to year. Figure 3 shows an example for the month of July. The systematic nature

of the errors is shown by comparing the top and bottom panels — areas where mean errors are large also have a large ratio of root-mean-square error to interannual standard deviation. The SRB data compare very similarly to the radiation data of ISLSCP Initiative I during 1987-1988 (Figure 4), but lack the information on the diurnal cycle. Thus, the choice was made to use the shorter ISLSCP Initiative I data set. We create a hybrid radiation forcing data set for shortwave and longwave by removing the monthly mean diurnal cycle of systematic error from the NCEP/NCAR reanalysis estimates.

As with precipitation, a multiplicative scaling is used to adjust the NCEP/NCAR reanalysis estimates:

$$[R]_{Y,M,D,T} = \frac{[R_{ISLSCP,87;88}]_{M,T}}{[R_{NCEP,87;88}]_{M,T}} [R_{NCEP}]_{Y,M,D,T} \quad (2)$$

This is similar to the process used to produce hybrid data sets for the other quantities where suitable observational data sets are available. However, for temperature and precipitation, suitably long data sets exist, such that we do not have to assume interannual stationarity in the systematic errors.

The near surface temperature estimates from the NCEP/NCAR reanalyses have been adjusted by subtracting the monthly mean errors as estimated from the gridded CAMS station data, or the Webber and Willmott (1998b) temperature data over South America through 1990. Both data sets have been corrected for variations in elevation between the observing stations and the grid mean elevation. These adjustments to temperature also affect the estimated saturation specific humidity. Thus, it is also necessary to adjust the estimates of near surface specific humidity from the NCEP/NCAR reanalysis to avoid incidents of supersaturation. This is done by assuming the same relative humidity before and after the temperature correction, and then adjusting the specific humidity accordingly to agree with the adjusted temperature.

Ideally we would use an observed data set of near surface humidity consistent with the observed temperature data, but no such data set exists.

The only fields of the NCEP/NCAR reanalysis which are not adjusted are the wind speed and the surface pressure. Surface pressure is one of the most reliable quantities in the NCEP/NCAR reanalysis (Kalnay et al. 1996), and although it is a component in the calculation of some surface moisture terms in the LSS, it is not a critically important component. As with humidity, we do not have a suitable observational data set over land to improve upon the near surface winds of the NCEP/NCAR reanalysis.

Any of the near surface meteorological data sets can be easily recreated with the inclusion of newer or better observational data as they become available. With GOLD, we intend to use version numbers and associated updates to documentation of the data set to indicate such changes.

It is possible to extend this hybrid correction of the NCEP/NCAR analysis data set further back in time. Gridded observed near surface monthly temperature data cover the entire period of the NCEP/NCAR reanalysis. However, a similar observed precipitation product does not exist. Provided the biases of the reanalysis can be assumed to be stable, one could extend the correction procedure for precipitation just as it has been done for radiation. This may be a major condition, particularly for a quantity as important as precipitation. Therefore, while the forcing data set could be extended back to the late 1940s, an additional caveat would have to be made for any the period prior to 1979.

The hybridization process could easily be applied to any other high temporal resolution analysis, such as the NCEP/DOE reanalysis spanning only the AMIP-2 period, the ECMWF ERA 15 or ERA 40 reanalyses, or the products of the NASA Data Assimilation Office.

The resulting hybrid 6-hour reanalysis-based forcing data set is a valuable data set itself. It represents a first-order improvement over the original reanalysis, and may be useful to applications in land-surface, ecological or hydrologic modeling.

3. Model Integration

An updated version of the 2-dimensional implementation of the Simplified Simple Biosphere (SSiB; Xue et al. 1991; 1996, Dirmeyer and Zeng 1997; 1999) model with a new core/driver structure has been used to produce the land surface climatological data set.

a. SSiB model history

The Simple Biosphere model (SiB), one of the most successful implementations of a complete LSS, was originally developed by Sellers et al. (1986), and has served as a standard in the field ever since. The treatment of moisture and heat fluxes between soil, vegetation and atmosphere in SiB is based on the so-called “big leaf” philosophy, a circuit of rheostats.

Due to the SiB model’s complexity and some limitations for its application to NWP and climate simulations, Xue et al. (1991) developed a simplified version of SiB (SSiB) by analyzing the SiB equations, identifying the dominant parameters and simplifying the parameterization and structure of SiB. In this version of SSiB there are ten prognostic variables including temperature and soil moisture for each of the three soil layers, and a temperature and moisture store for the vegetation canopy and land surface.

SSiB has been used in a number of different applications. Used as a point model, SSiB can be driven by the observations from a single meteorological station or instrumented field site. SSiB can also be coupled with an atmospheric column model in one vertical dimension. The version of SSiB which is fully

coupled into an numerical atmospheric model is referred to as the three-dimensional implementation. In order to perform the testing and calibration of parameters and parameterizations in a global framework, a two-dimensional offline implementation of SSiB (2DSiB) model was developed by Dirmeyer and Zeng (1997). It allows global simulation of hydrologic and energy flux terms at the land surface driven by grided meteorological data. It has been used as a tool for land surface data assimilation and model development, such as the Global Soil Wetness Project (GSWP; Dirmeyer et al. 1999).

b. New core/driver structure & spatial independence

To make the SSiB model more flexible to different applications and to facilitate model updating, we have adopted and implemented a new core/driver model structure. This new model structure is coded in FORTRAN 90, and contains three parts: an updated version of the unified point/grid SSiB core model, a driver (which can be a general 1D or 2D offline land model driver, or an atmospheric model), and a standard interface between them. (Figure 5).

The new standard SSiB core model's function is to advance land states at a single point and a single time-step into the future. The intention of severing the core from the driver is to separate the coding tasks of model development for model operation. Thus, no user modification needs to be done on the core model to use it for a new application. In other words, the core model is independent from its driver. It communicates with the driving part (offline driver or GCM) through the standard interface. Dynamic memory allocation and data modules are used for data passing between the core and the driver. Any changes to the parameterizations of the core model will not affect the driver, and this makes both the model application and model updating more convenient. The new SSiB core model has been verified in single point/grid simulations with data sets from the Project for Intercomparison of Land Surface Parameterization

Schemes (PILPS; Henderson-Sellers et al. 1995), as well as in global scale simulations in the framework of GSWP (IGPO 1995).

A relatively general and flexible new 2D driver (following the GSWP data format convention) has been implemented based on the 2DSiB model and applied to this land data sets generation exercise. In this driver, the model integration time interval, forcing data time interval and model output time interval are all adjustable, and the model resolution is also easily adjusted for different applications by simply editing a control file and the driver's data model. The new 2-D driver has been tested in GSWP simulations. It can also be used for driving other core LSSs with minor modification.

c. Initialization and spin-up process

Since integration of the LSS requires initial conditions which include soil wetness and temperature, but observed global soil wetness and temperature data are nonexistent, we decided to create a realistic soil wetness initial condition by spinning the model up for a continuous 24-year period (1976-1999) instead of recursively reintegrating one year, as was done in GSWP. The final state for 0000UTC 1 January 2000 is used as the initial condition for the 1979-1999 integration, which generates the expected land surface climatological data set. This extended spin-up insures that the deep soil will be in reasonable equilibrium at the start of the integration. Of course, any unique anomalies induced by the conditions leading up to 1979 will be lost. However, it was found that in the absence of CMAP precipitation data to adjust the reanalysis precipitation prior to 1979, the statistics of the rainfall for 1976-1978 were significantly different from the latter period. The differences were large enough to adversely effect the simulation of runoff in SSiB. Thus, it was decided that starting from 1 January 2000 conditions was the less objectionable option.

d. Resolution, data intervals, flexibility (Fig 6 of resolution overlays)

Four different resolutions of the GOLD data sets have been produced, corresponding to the four principal Gaussian grids employed by the spectral COLA GCM:

- T31 (96 longitude \times 48 latitude) with 1034 ice-free land grid points
- T42 (128 longitude \times 64 latitude) with 1811 ice-free land grid points
- R40 (128 longitude \times 102 latitude) with 2902 ice-free land grid points
- T63 (192 longitude \times 96 latitude) with 4106 ice-free land grid points

These GCM resolutions can be contrasted with the GSWP experiment at 1E resolution, which had 14,637 ice-free land grid points. Figure 6 shows an example overlay of these various resolutions, as well as the grid of the NCEP/NCAR reanalysis and the original 1E ISLSCP Initiative I data set. Regional or global simulations can also be produced at any arbitrary resolution, as long as a land-sea mask is provided at that resolution. Vegetation and soil properties are interpolated from the ISLSCP Initiative I data set.

For each of the four resolutions listed above, a 30-minute integration time interval is used. The forcing data time interval is 6 hours, and model output time interval is once a month. All these time intervals are adjustable and can be controlled by the user. The forcing data are interpolated from the 6-hour interval to the LSS time step. Linear interpolation is used for most fields. Shortwave radiation is adjusted so that the local sunrise and sunset times are preserved. The convective component of precipitation is redistributed within the 6-hour interval according to a pre-defined probability density function to better represent the episodic nature of convective rainfall, and improve the partitioning of that precipitation between infiltration and runoff.

4. Land Surface Data Set

There are three basic types of data set that are associated with the LSS, one for input and two for output. First are the time-invariant fields that define characteristics of the landscape important to the LSS. Table 1 shows a list of these fields. They predominantly include soil parameters regarding the moisture-holding capacity. Also included is the vegetation type, which is categorized in Table 2. No attempt has been made to account for changing land use over the period of integration. At the spatial scales of GCM grids, the changes would be marginal over two decades. If integrated on regional scales at higher resolution, there could be a real impact to using a time-varying data set of vegetation cover.

There are two categories of time-varying output fields from the LSS, following the structure of GSWP (IGPO 1995): time-mean fields and instantaneous values. The baseline data sets have been produced with a one-month output interval, but as mentioned before, grids can be output at virtually any regular interval. The time mean fields are meant to provide a climatological data set of land surface fluxes and state variables. The instantaneous fields are particularly useful for providing initial conditions to coupled land-atmosphere model integrations. They can also be used for spot validation with remotely sensed data within a narrow time window. Tables 3 and 4 list the variables output at these time intervals and their units. Because of the legacy of GSWP, there are some redundancies and peculiarities in the list. For example, in the list of instantaneous fields (Table 4), surface soil wetness appears twice. There are also some special soil moisture variables. Soil wetness index (SWI) is a normalized linear measure of soil wetness that ranges from 0 (wilting point) to 1 (field capacity). SWI could conceivably be greater than 1, or negative in certain instances for this LSS. Normalized Available Water in the surface and rooting layers is similar to SWI in definition, and ranges from 0 (wilting point) to 1 (saturation). Available water in the top meter of soil is calculated for easy comparison to gravimetric measurements. It is interpolated depending on the soil profile

at a given grid point in the LSS. The assumption in this calculation is that soil moisture is evenly distributed throughout each layer, so that a linear interpolation can be used to estimate available water in the top meter of soil.

In the remainder of this section, we examine the climatology of the 21-year global offline land surface data set. The T63 resolution data set (1.875E) is shown here exclusively, although the data sets produced at lower resolutions are very comparable in their overall features. We present an examination of the mean state of the land surface, the mean annual cycle, interannual variability, and specific comparisons to the existing GSWP simulations during the 1987-1988 period, and an examination of characteristics of land surface variability during the ten-year period of the ISLSCP Initiative 2 (1986-1995).

a. Time mean

Figure 7 shows the annual mean climatological values for root zone soil wetness, surface latent and sensible heat flux, and runoff averaged over the 21-year period. The LSS reports high values of soil wetness over the tropics and at high latitudes, with the deserts well defined as dry. The zonal gradient across the central United States is evident, as is the sharp meridional gradient in the vicinity of the Sahel of Africa. Latent heat fluxes are highest in the tropics, but globally are generally lower than the sensible heat fluxes. This is a particular characteristic of the SSiB LSS (Dirmeyer et al. 2000). The largest sensible heat fluxes are over the interior of Australia and coastal regions around the Arabian Sea. Some of the lowest values are over Europe (recall that no calculation is performed over permanent ice over Greenland or Antarctica). The pattern of runoff resembles that of soil wetness, with the greatest values over Indonesia, Bangladesh, and the Amazon Basin.

Overall, the patterns of the time-mean surface energy and water balance terms appear reasonable and consistent with the forcing data. The LSS locally conserves energy and water when run in this uncoupled configuration.

b. The annual cycle

The zonally-averaged mean annual cycle of the same four quantities over land is shown in Fig 8. Soil wetness and runoff reflect the oscillation of the low-latitude Hadley circulation, and the cross-equatorial travels of the terrestrial inter-tropical convergence zone (ITCZ). The summer monsoons are evident in soil wetness around 20E-30E in both hemispheres. At higher latitudes, the springtime maximum in soil wetness is also present, with associated peaks in runoff about one month later that are driven in part by snowmelt. The annual cycles of latent and sensible heat fluxes follow the oscillation of available solar energy. The highest values of zonal mean latent heat fluxes occur off the equator in low latitudes, while sensible heat fluxes peak predominantly in the subtropics.

The annual cycle of surface water and energy budget terms have also been calculated over a number of regions that correspond with selected major river basins, continental-scale GEWEX research areas and other regions with particularly interesting seasonality. Figure 9 outlines one dozen areas where area averages have been calculated.

The terms of the water balance are shown in Fig 10. All regions chosen have a range of monthly mean precipitation that span at least a factor of two during the year. Recall that precipitation is a specified forcing, and the other terms shown are all the product of the LSS used here. Most areas show a synchronization of evapotranspiration with the precipitation cycle. In the subtropics and mid-latitudes, this

correspondence is due to the limitations on water availability in the soil during the dry season. At high latitudes, the correspondence has more to do with the synchronization of increased available radiant energy during the summer, and a greater capacity of the summertime atmosphere to hold water vapor and deliver larger precipitation events. Water is not usually a limiting factor for evaporation at high latitudes (Koster and Suarez 1999). Over the Amazon, model evapotranspiration is nearly uniform throughout the year.

As was evident in Fig. 8, runoff peaks during spring at mid- and high latitudes of the Northern Hemisphere. In the monsoonal regimes, runoff is synchronized with the rainy season. The LSS simulates a late autumn maximum in runoff over the La Plata basin of South America, largely as a response to the wet soils and shutdown of evapotranspiration during that season. Changes in storage reflect fluctuations in both soil moisture in the active column (which may range from 1 to 6 meters depending on location), and in snowpack. The loss in water storage during spring at high northern latitudes is the signature of seasonal snow melt. Drier areas respond to the wet season predominantly by recharging soil moisture, while wet regions typically have a peak in runoff.

Figure 11 gives the corresponding picture for the surface energy balance. The bars denote incoming radiant energy (positive downward), while the lines show responding heat fluxes (positive upward). The magnitude and phase of the cycle of net solar radiation is a clear indication of the latitude of the region. Net longwave radiation responds not only to the surface temperature, but to the effective atmospheric temperature, which can be strongly modulated by clouds. This can reduce the net longwave fluxes during the wet seasons at low latitudes. Latent and sensible heat fluxes often vary in tandem, except in areas where there is a stark contrast between wet and dry seasons, such as the Sahel and India. In those areas, high sensible heat fluxes give way to increasing latent heat fluxes with the onset of the rainy season.

At high northern latitudes there are strong negative ground heat fluxes during winter. This is a reflection of the definition of ground heat flux in this LSS, which is measured by the change in soil temperature. In this case, there is an exchange of heat from relatively warm soil to cold snowpack. The snow then dispels this energy by sensible heat fluxes and longwave cooling to the atmosphere.

c. Interannual variability

Figure 12 shows the 21-year time series of monthly precipitation anomalies (relative to the same 21-year period), and cumulative anomalies in the surface water balance terms. No special significance should be assigned to the fact that all of the cumulative anomalies return to zero at the end of the series, because that is an artifact of the choice of the reference period. The relative trends and changes in the cumulative anomalies over time are more meaningful than their absolute values. The trace of cumulative changes in soil wetness is in fact a proxy for the fluctuations in column soil moisture.

In the tropical basins, cumulative runoff anomalies track the cumulative rainfall anomalies very closely. The La Plata basin, though predominantly in the subtropics, also exhibits this behavior. Evapotranspiration and soil moisture changes are fairly consistent throughout the period. In fact, due to the finite holding capacity of the active soil column defined in the LSS and the lack of a parameterization of recharge from below, the cumulative change in soil wetness is necessarily limited. Presumably, the generally humid conditions at low latitudes necessarily drives fluctuations in rainfall to affect runoff, since evaporation is rarely stressed.

In the subtropical areas, a somewhat different behavior is found. Over the Sahel, there are similar responses of both runoff and evapotranspiration to fluctuations in rainfall, with evapotranspiration having

a somewhat stronger response. The strong decline in precipitation during the 1980s is evident in the concomitant drop in runoff and evapotranspiration. A moderate rebound is seen in the late 1990s. Over India, there is a curious trend from the mid-1980s to the mid-1990s when the runoff rate appears to have exceeded the precipitation rate for nearly a decade, at the expense of evapotranspiration. The impact of the record wet monsoon of 1988 is evident in the sharp increases in cumulative rainfall and runoff, offset later by the poor monsoons of 1991 and 1992. Over southern Africa, the response to fluctuations in rainfall appears to be almost wholly in cumulative evapotranspiration, with very little variation in runoff. In fact, soil wetness varies more than runoff.

At higher latitudes, there still exists some correlation between cumulative runoff and precipitation, but there also exists a great deal of variation in the other terms. The Mississippi-Missouri basin shows a fairly strong anticorrelation between runoff and evapotranspiration. Over this basin, the 1988 drought is very evident in the sharp decrease in cumulative precipitation and runoff. It is followed by an extended period of recovery culminating in the floods of 1993. In this model, there is not a profound response of evapotranspiration to the 1993 flood (again suggesting a largely unstressed background state for evapotranspiration), but there is a strong response during 1999. Basins such as the Lena and Yangtze show pronounced low-frequency cycles, while basins like the Baltic Sea drainage area are dominated by very short-term variations.

d. Comparison to other data sets

It is interesting to note the differences for 1987 and 1988 between this data set, and the GSWP simulations at 1E resolution performed with essentially the same LSS (Dirmeyer and Zeng 1999). There are many differences between the two simulations. They use different forcing data sets, although the hybrid

fields described in section 2, especially the precipitation, may bear more resemblance than the dynamical fields. Also, this data set is presumably well “spun up” by 1987, whereas the GSWP simulation began from an equilibrium 1987 state. Also, the models have different resolutions, and somewhat different surface boundary conditions (e.g., vegetation maps) as a result, although this should not have a great impact on basin-scale results.

One of the goals of ISLSCP is to compile and co-register data over a sufficiently long period that meaningful studies of climate means and variability can be performed. This is more an issue for ISLSCP Initiative 2 (II2), which covers the ten-year period from 1986-1995. ISLSCP Initiative 1 (II1) covered a two-year period (1987-1988) that was chosen in part because it exemplified hydrologically interesting years over many regions of the globe. Given the 21-year data set compiled here, encompassing the ISLSCP periods, we can compare to see if the shorter ISLSCP periods are statistically consistent with this longer period.

Table 5 shows results of this comparison for both the mean and monthly variance (mean annual cycle removed) of precipitation and temperature. Monthly means have been used in Student’s t-tests and Fisher’s F-tests for means and variances respectively. We see that, as might be expected from sample size, the very short II1 period shows more situations where it likely to be unrepresentative of the 21-year period, especially in terms of precipitation. Over seven of the basins there is at least an 80% probability that the 24 monthly means of precipitation for that two-year period are significantly different than the 21-year mean (Mississippi-Missouri, 0.40s less; Northern Amazon, 0.33s more; Southern Africa, 0.30s more; Mackenzie, 0.34s more; Lena, 0.36s more; La Plata, 0.38s less; Congo, 0.44s more). The only potentially significant anomalies in the II2 series are over the southern Amazon (0.18s cooler) and southern Africa (0.15s warmer) than the 21-year period of this data set.

There are some interesting aberrations in the variances too. The Yangtze and Congo basins show potentially excessive variance in monthly temperatures during the II1 interval — both having about 60% more variance than over the 21-year period. Over southern Africa during the II1 period there is about 34% less rainfall variability, suggesting that the wet anomaly was not caused by sporadic heavy rain events, but by a fairly consistent increase in rainfall during the period. The precipitation variance over the La Plata and Congo basins during the II2 interval was higher than in the 21-year period by 26-28% in each case. The identification of some of these precipitation variance changes as significant may be an artifact of our application of the F-test to a quantity which is not normally distributed on short time scales. This is less of an issue for monthly means than they would be for, say, daily rainfall, but certainly more of an issue than it would be for annual precipitation. Nevertheless, 25-30% increases in monthly variance during 10-year periods are probably noteworthy.

Overall, the short II1 period does show more statistical aberrations than the longer II2 period, with hydrologic anomalies being particularly prominent. The II2 window of 1986-1995 appears to be a reasonable proxy for the 1979-1999 period in most respects for precipitation and temperature.

5. Summary

A global offline land-surface data set (GOLD) has been produced for the period 1979-1999. Its intent is to be used for specifying initial and/or boundary conditions for global climate models, climate variability analyses over land, and applications in hydrology, ecology or biogeochemistry. We have chosen the SSiB LSS to produce the spatially and temporally continuous data set at several resolutions, corresponding to the GCM resolutions commonly used at COLA, but the data sets can be produced at any resolution. Hybrid data sets of forcings (precipitation, downward radiation and near surface

meteorological state variables) have also been produced at these resolutions by combining the NCEP/NCAR 6-hourly reanalysis data with observed monthly data sets when and where available to reduce systematic errors in the reanalysis. These data are then further downscaled in time and used to drive the LSS, producing a data set of surface state variables and fluxes. These data are being made available to the research community.

A brief analysis is presented of the grand mean, mean annual cycle, and seasonal-interannual variability of key surface energy and water balance terms. Calculations have been performed over a number of sub-continental scale regions to examine the climatology of the data set. The model produces a realistic climatology, and shows significant interannual variability over most of the regions examined. There is a great deal of inter-region variation in the response of surface fluxes to anomalies in climate.

Comparisons are also made between this 21-year period, and the shorter periods of ISLSCP Initiatives 1 and 2 to see if the ISLSCP periods are statistically representative of the longer term. By and large, they are found to be similar, except that the 2-year ISLSCP Initiative I period has a great deal of aberration in regional rainfall compared with the 21-year GOLD period.

Acknowledgments: This work was supported by NSF grant ATM 9814265, NOAA grant NA96GP0056 and NASA grant NAG5-8202.

References

- Betts, A. K., and J. H. Ball, 1998: FIFE surface climate and site-average dataset 1987-1989. *J. Atmos. Sci.*, **55**, 1091-1108.
- Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367- 374.
- Dirmeyer, P. A., A. J. Dolman, and N. Sato, 1999: The Global Soil Wetness Project: A pilot project for global land surface modeling and validation. *Bull. Amer. Meteor. Soc.*, **80**, 851-878.
- Dirmeyer, P. A., and F. J. Zeng, 1997: A two-dimensional implementation of the Simple Biosphere (SiB) model. *COLA Technical Report 48* [Available from the Center for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705 USA], 30 pp.
- Dirmeyer, P. A., and F. J. Zeng, 1999: An update to the distribution and treatment of vegetation and soil properties in SSiB. *COLA Technical Report 78* [Available from the Center for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705 USA], 25 pp.
- Gibson, R. K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA Description. *ERA Technical Report No. 1*, ECMWF, Reading, UK.
- Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad, and T. H. Chen, 1995: The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.*, **76**, 489-503.
- International GEWEX Project Office, 1995: *Global Soil Wetness Project*, 47 pp.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, & D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.
- Kistler, R., W. Collins, S. Saha, G. White, J. Woollen, E. Kalnay, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino, 2001: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247-268.
- Koster, R. D., and M. J. Suarez, 1999: A simple framework for examining the interannual variability of land surface moisture fluxes. *J. Climate*, **12**, 1911-1917.

- Liston, G. E., Y. C. Sud, and G. K. Walker, 1993: Design of a global soil moisture initialization procedure for the Simple Biosphere model. *NASA Tech. Memo. 104590*, Goddard Space Flight Center, Greenbelt, Maryland, 138 pp.
- Meeson, B. W., F. E. Corprew, J. M. P. McManus, D. M. Myers, J. W. Closs, K. J. Sun, D. J. Sunday, and P. J. Sellers, 1995: *ISLSCP Initiative I - Global data sets for land-atmosphere models, 1987-1988*. Volumes 1-5, Published on CD-ROM by NASA (USA_NASA_GDAAC_ISLSCP_001 - USA_NASA_GDAAC_ISLSCP_005).
- Mitchell, K. E., and Y. Lin, 1994: Production of 6-hourly continental precipitation data sets for 1987 and 1988 for ISLSCP Initiative I. Development Division, National Meteorological Center, Washington, DC.
- Polcher, J., P. Cox, P. Dirmeyer, H. Dolman, H. Gupta, A. Henderson-Sellers, P. Houser, R. Koster, T. Oki, A. Pitman, and P. Viterbo, 2000: GLASS: Global Land-Atmosphere System Study. *GEWEX News*, **10**, No. 2, 3-5.
- Ropelewski, C. F., J. E. Janowiak, and M. F. Halpert, 1985: The analysis and display of real time surface climate data. *Mon. Wea. Rev.*, **113**, 1101-1107.
- Rossow, W., and R. Schiffer, 1991: ISCCP cloud data products. *Bull. Amer. Meteor. Soc.*, **72**, 2-20.
- Schubert, S. D., R. B. Rood, and J. Pfaendtner, 1993: An assimilated dataset for earth science applications. *Bull. Amer. Meteor. Soc.*, **74**, 2331-2342.
- Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher, 1986: A simple biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.*, **43**, 505-531.
- Sellers, P. J., S. I. Rasool, and H.-J. Bolle, 1990: A review of satellite data algorithms for studies of the land surface. *Bull. Amer. Meteor. Soc.*, **71**, 1429-1447.
- Singh, S. V., R. H. Kripalani, and D. R. Sikka, 1992: Interannual variability of the Madden-Julian oscillations in Indian summer monsoon rainfall. *J. Climate*, **5**, 973-978.
- Stackhouse, P. W., Jr., S. K. Gupta, S. J. Cox, M. Chiacchio, and J. C. Mikovitz, 2000: The SRB Project Release 2 Data Set: An Update. *WCRP GEWEX News*, **10** (3), 4.
- Webber, S. R., and C. J. Willmott, 1998: *South American precipitation: 1960-1990 gridded monthly time series (Version 1.02)*. Center for Climatic Research, Dept. of Geography, Univ. of Delaware.

- Webber, S. R., and C. J. Willmott, 1998: *South American air temperature: 1960-1990 gridded monthly time series (Version 1.01)*. Center for Climatic Research, Dept. of Geography, Univ. of Delaware.
- World Meteorological Organization, 1996: *Guide to Meteorological Instruments and Methods of Observation*, WMO Publication 8, Sixth Edition [Available from the WMO, P.O.Box 2300, CH-1211 Geneva 2, Switzerland].
- Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539-2558.
- Xue, Y., F. J. Zeng, and C. A. Schlosser, 1996: SSiB and its sensitivity to soil properties - a case study using HAPEX-Mobilhy data. *Global and Planetary Change*, **13**, 183-194.
- Xue, Y., P. J. Sellers, J. L. Kinter and J. Shukla, 1991: A simplified biosphere model for global climate studies. *J. Climate*, **4**, 345-364.
- Yarosh, E. S., and C. F. Ropelewski, 2001: Sensitivity of the observed U.S. water budget to two different precipitation estimates. *Proc. 12ty Symp. Climate Variations*, American Meteorological Society.

Variable	Description
Vegetation type	Based on types from ISLSCP I-1 CD-ROM (see Table 2)
Soil type	Based on classes from ISLSCP I-1 CD-ROM
Depth of surface soil layer	5 cm globally
Depth of root layer	Function of vegetation type, total soil depth
Depth of recharge layer	Function of total soil depth from ISLSCP I-1 CD-ROM
Wilting point of surface + root layers	Function of soil type and vegetation type
Wilting point of total soil column	Function of soil type and vegetation type
Field capacity of surface + root layers	Function of soil type
Field capacity of total soil column	Function of soil type
Total capacity of surface + root layers	Function of soil porosity
Total capacity of total soil column	Function of soil porosity

Table 1. List of time-invariant fields.

Type	Definition	Seasonal Cycle
1	Broadleaf evergreen trees	No
2	Broadleaf deciduous trees	Yes
3	Broadleaf and needle leaf trees	Yes
4	Needle leaf evergreen trees	Yes
5	Needle leaf deciduous trees	Yes
6	Broadleaf trees with groundcover	Yes
7	Groundcover only	Yes
8	Not used	
9	Broadleaf shrubs with bare soil	Yes
10	Broadleaf shrubs with ground cover (Tundra)	Yes
11	Bare soil	No
12	Winter wheat and broadleaf deciduous trees	Yes

Table 2. List of SSiB vegetation types.

Variable	Units
Soil moisture in the surface layer	mm
Soil moisture in the surface and root layers	mm
Soil moisture in the total soil column	mm
Soil wetness index of the surface layer	—
Soil wetness index of the surface and root layers	—
Surface snow (liquid water equivalent)	mm
Effective skin temperature	K
Precipitation rate	mm day ¹
Surface runoff rate	mm day ¹
Total runoff rate	mm day ¹
Snowmelt rate	mm day ¹
Transpiration rate	mm day ¹
Bare soil evaporation rate	mm day ¹
Direct snow evaporation rate	mm day ¹
Interception loss	mm day ¹
Total evapotranspiration rate	mm day ¹
Downward shortwave radiation at the surface	W m ⁻²
Downward longwave radiation at the surface	W m ⁻²
Upward shortwave radiation at the surface	W m ⁻²
Upward longwave radiation at the surface	W m ⁻²
Sensible heat flux	W m ⁻²
Latent heat flux	W m ⁻²
Ground heat flux	W m ⁻²

Table 3. List of time mean fields.

Variable	Units
Soil moisture in the surface layer	mm
Soil moisture in the surface and root layers	mm
Soil moisture in the total soil column	mm
Soil wetness in the surface layer	—
Soil wetness in the surface and root layers	—
Temperature of vegetation canopy	K
Effective skin temperature	K
Temperature of surface layer	K
Temperature of root layer	K
Temperature of recharge layer	K
Total canopy water storage	mm
Surface storage (snow)	mm
Soil wetness in the surface layer	—
Soil wetness in the root layer	—
Soil wetness in the recharge layer	—
Soil wetness index	—
Available water in the top meter of soil	mm
Normalized available water	—

Table 4. List of instantaneous fields.

			Mississippi-Missouri	Southern Amazon	India	Northern Amazon	Sahel	Southern Africa	Baltic Sea	Mackenzie	Yangtze	Lena	La Plata	Congo
P	II1	Prob(t)	93%	14%	39%	88%	55%	83%	1%	89%	1%	91%	92%	96%
		Prob(F)	58%	1%	22%	47%	69%	87%	24%	18%	49%	26%	58%	21%
	II2	Prob(t)	8%	34%	60%	0%	5%	48%	12%	13%	2%	19%	73%	62%
		Prob(F)	22%	10%	71%	30%	72%	8%	2%	16%	37%	11%	85%	87%
T	II1	Prob(t)	52%	4%	62%	60%	65%	42%	54%	11%	51%	57%	14%	69%
		Prob(F)	7%	26%	47%	72%	39%	32%	12%	18%	81%	33%	52%	82%
	II2	Prob(t)	34%	90%	76%	63%	2%	84%	42%	21%	73%	14%	19%	76%
		Prob(F)	30%	23%	26%	59%	17%	64%	8%	44%	57%	14%	4%	52%

Table 5. Probabilities that the ISLSCP Initiative 1 (II1) and Initiative 2 (II2) periods are not representative of the 21 period of this data set in terms of anomalies of monthly mean (Prob(t)) and variance (Prob(F)) for precipitation (P) and surface temperature (T), for various regions. Probabilities of greater than 80% are shaded. Probabilities of greater than 90% are also bold.

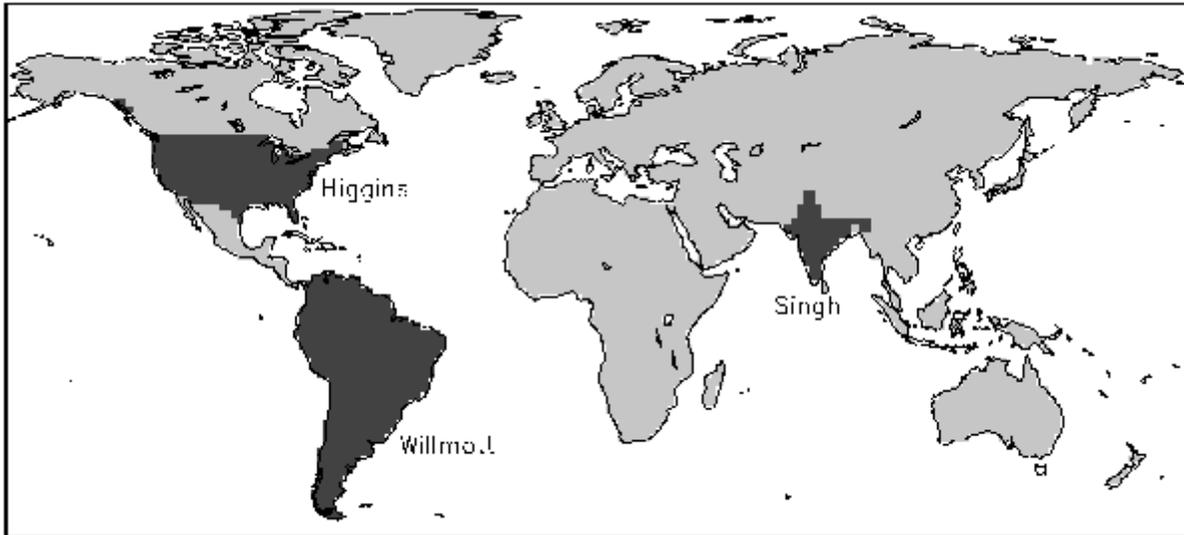


Figure 1 Areas of coverage for regional observational precipitation data sets.

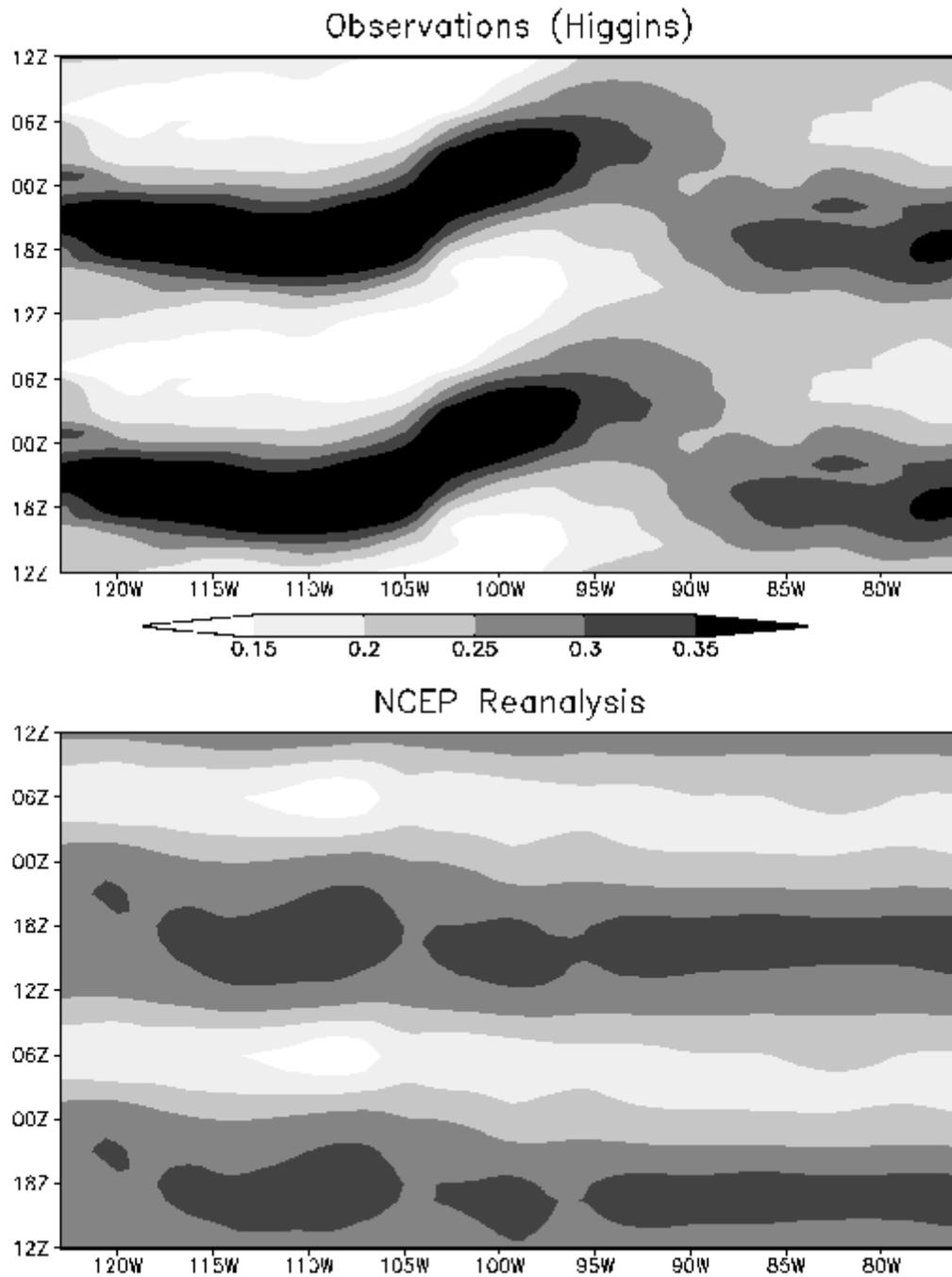


Figure 2 Mean June diurnal cycle of precipitation (repeated for 2 cycles) at 35 N from observations (top) and reanalysis (bottom). Units are percentage of daily precipitation falling during the ensuing 6-hour period.

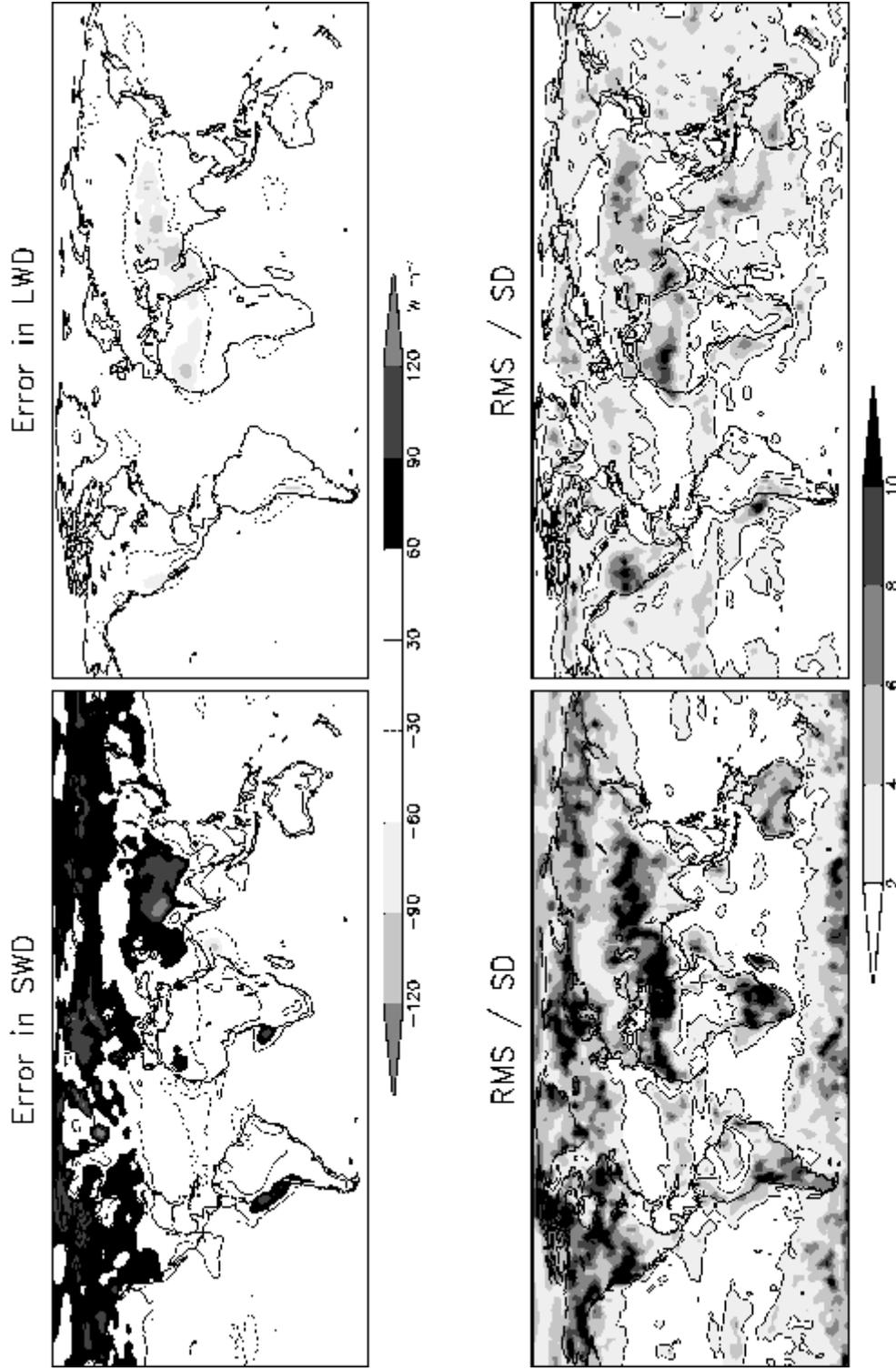
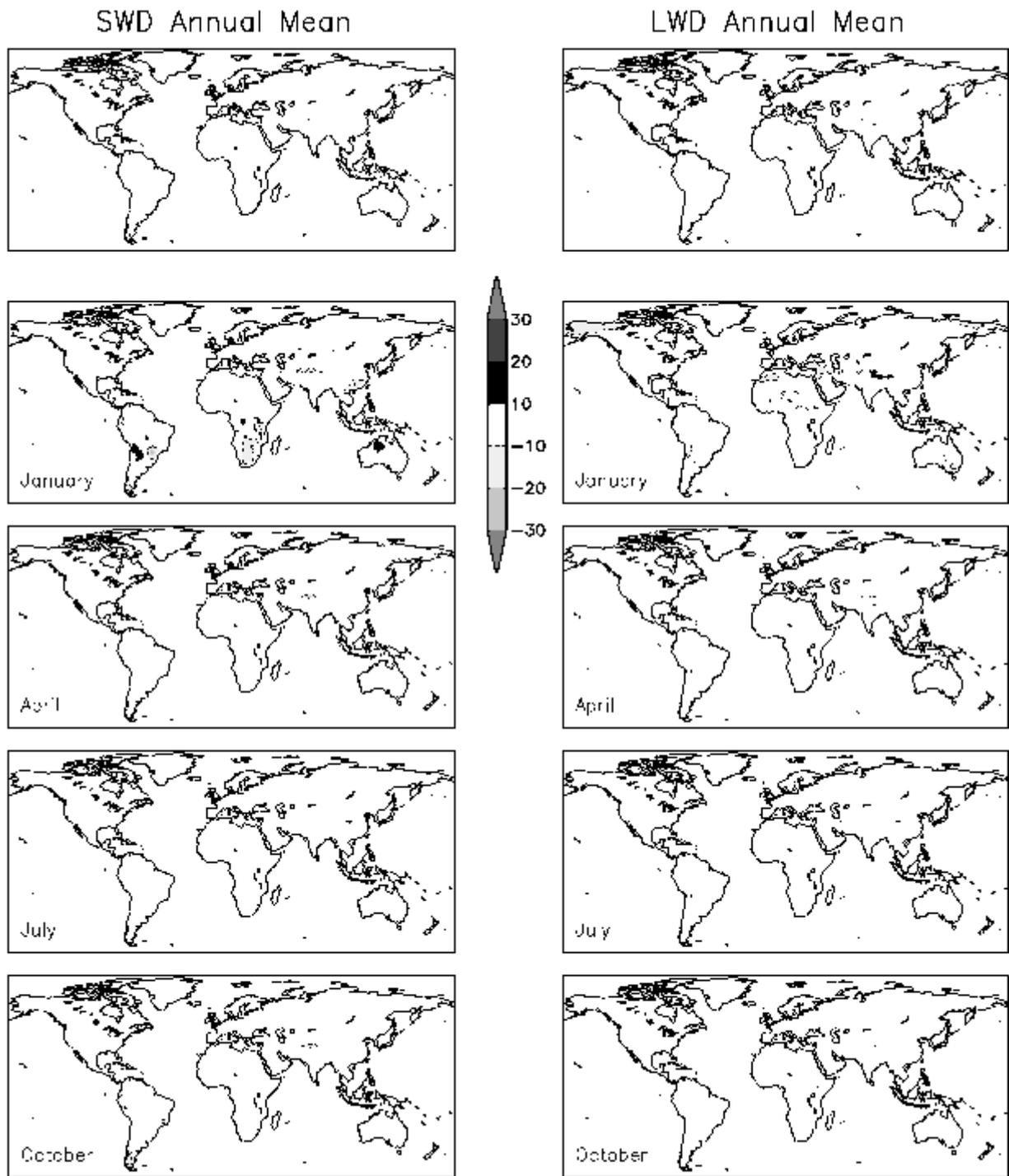


Figure 3 Errors in July NCEP reanalysis estimates of downward shortwave and longwave radiation at the earth's surface, averaged over 1983-1990 (top); ratio of root-mean-square error to interannual standard deviation of July surface radiances (bottom). Units for top panels are $W m^{-2}$.



SRB vs. ISCCP 2-Year Statistics (1987-1988)

Figure 4 Differences between SRB and ISCCP mean downward shortwave (left) and longwave (right) radiation during 1987-1988 for the annual mean (top) and selected months. Units are $W m^{-2}$.

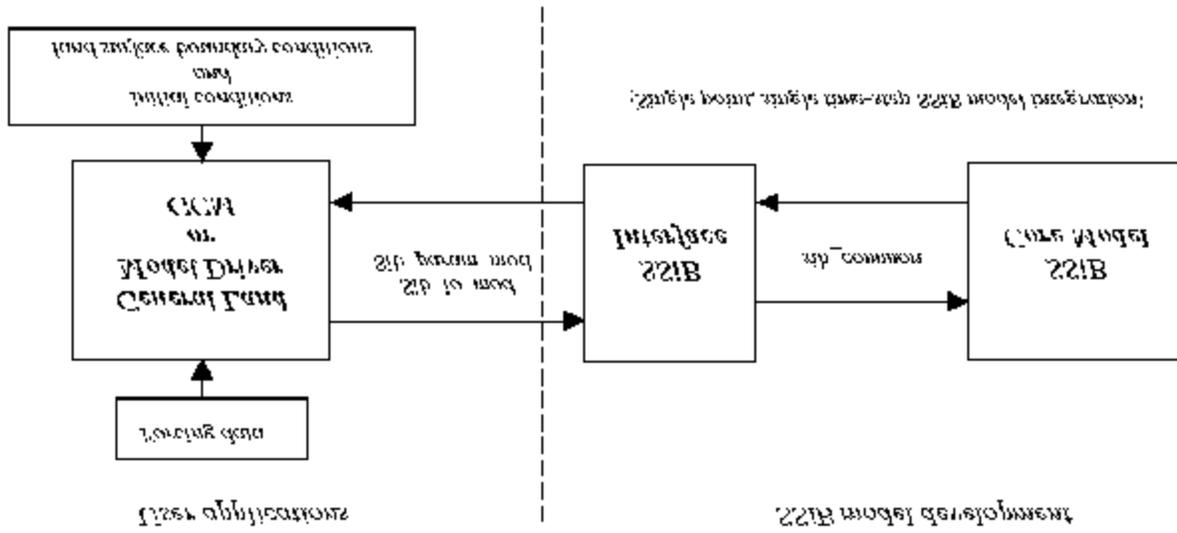


Figure 5 Schematic of the LSS structure.



Figure 6 Example of the grids of data sets used in this paper. Shading indicates land points on the ISLSCP Initiative 1 1E grid. Boxes outline the grid of the NCEP global reanalysis. Symbols show the centers of land grid boxes for four different resolutions of GOLD; open circles for T31, open squares for T42, pluses for R40, triangles for T63.

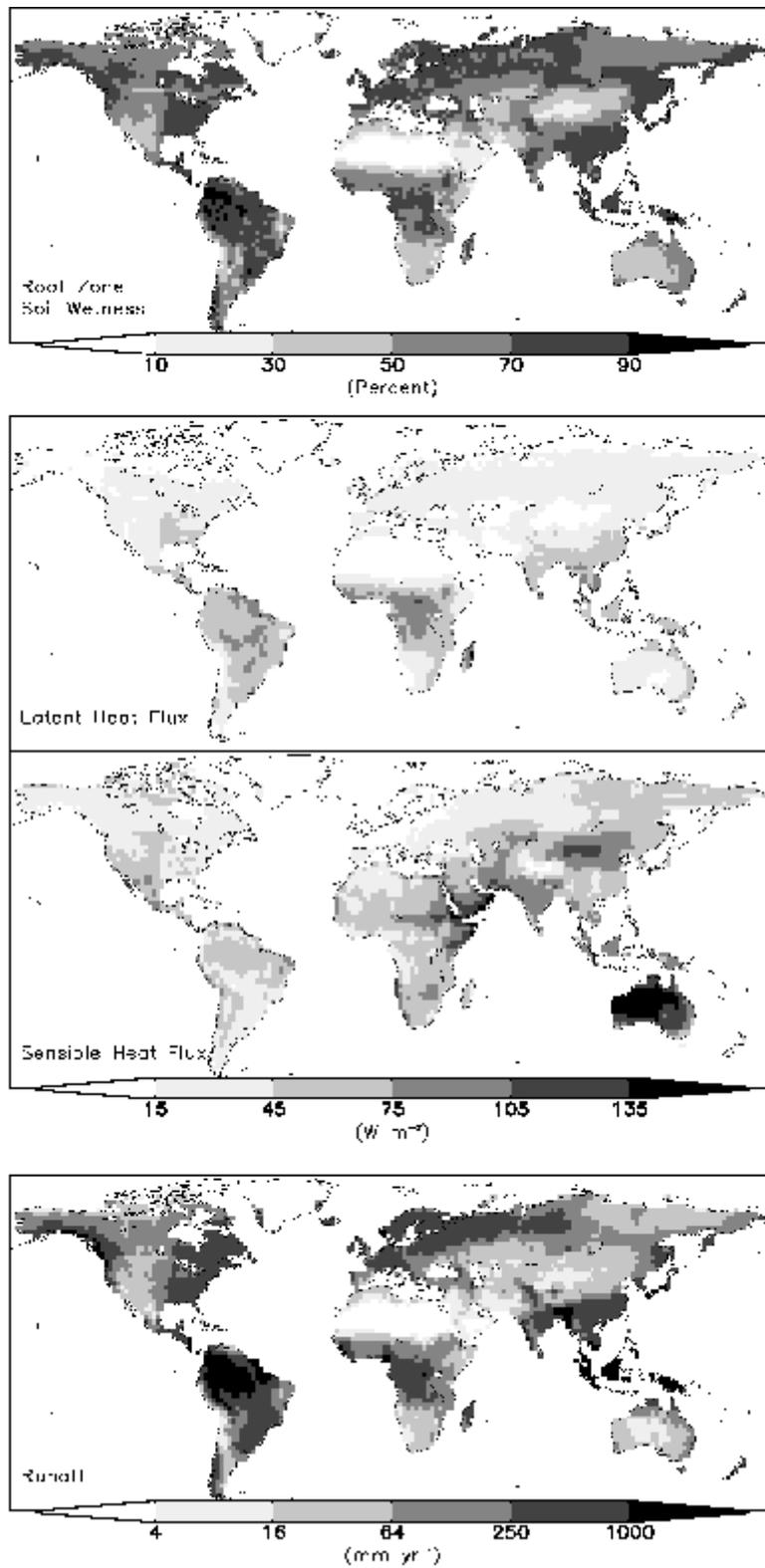


Figure 7 21-year mean of selected GOLD quantities.

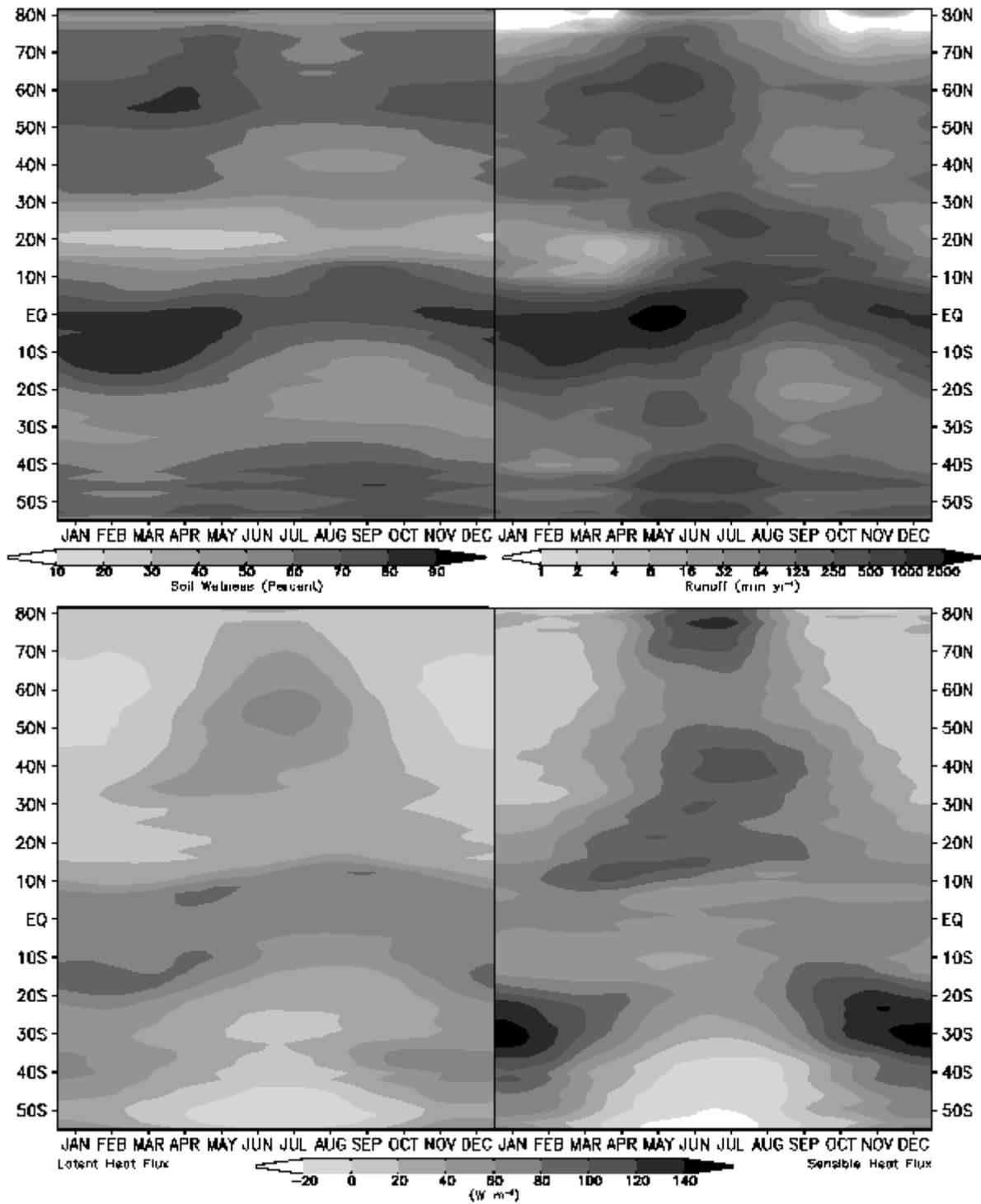


Figure 8 21-year averaged zonal mean annual cycle of selected GOLD quantities.

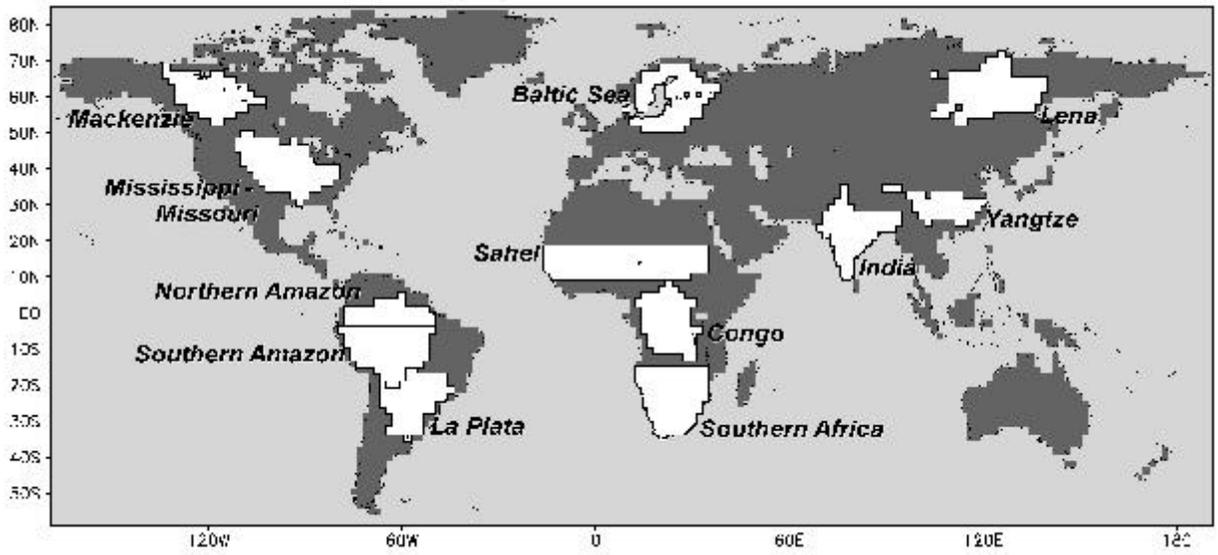


Figure 9 Selected areas for regional water and energy balance calculations.

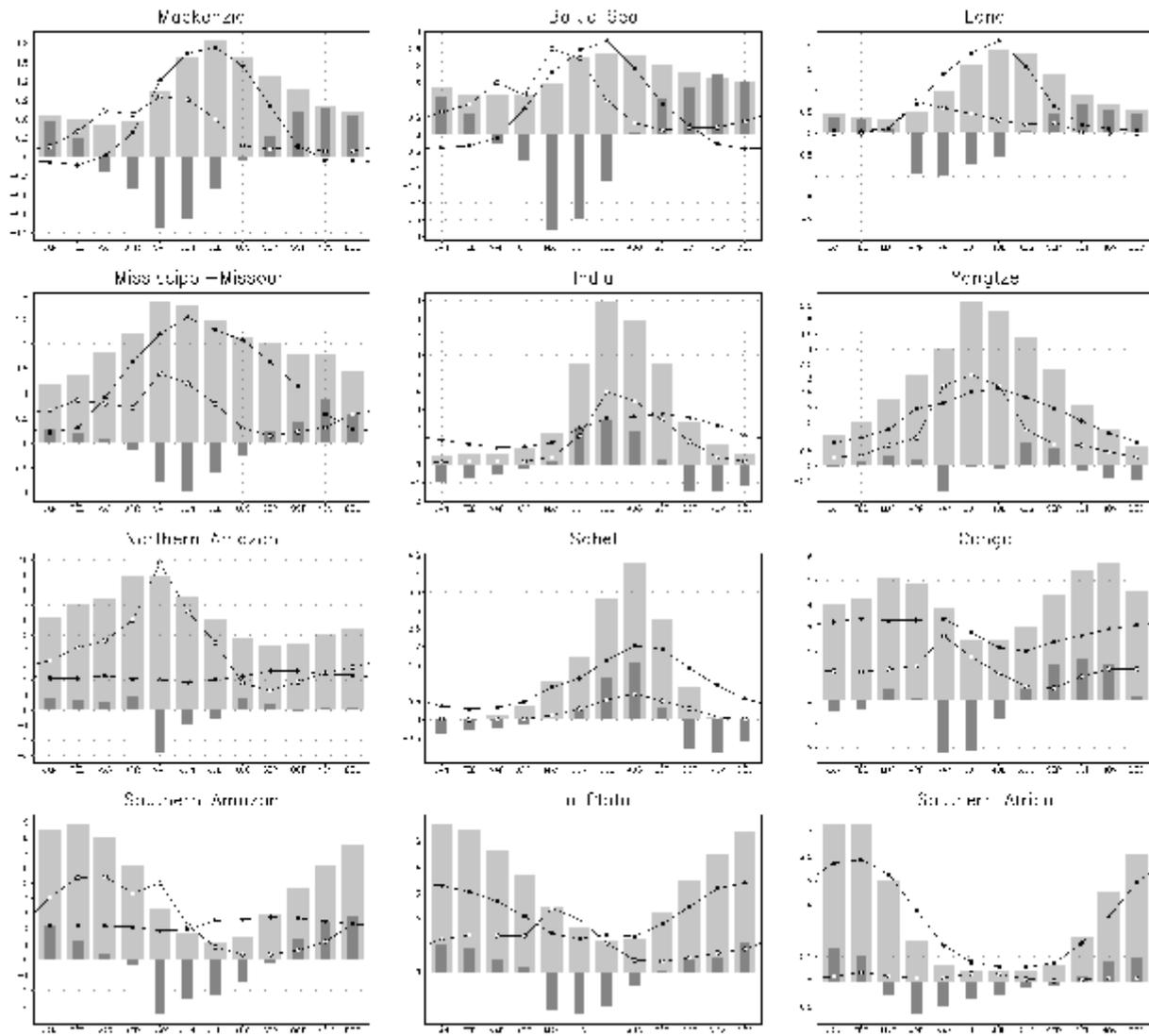


Figure 10 21-year mean annual cycles of surface water budget terms over the areas shown in Fig 9; wide bar = precipitation, narrow bar = change in storage (soil moisture + snow), filled circles = evapotranspiration, open circles = runoff. Units are mm d^{-1} .

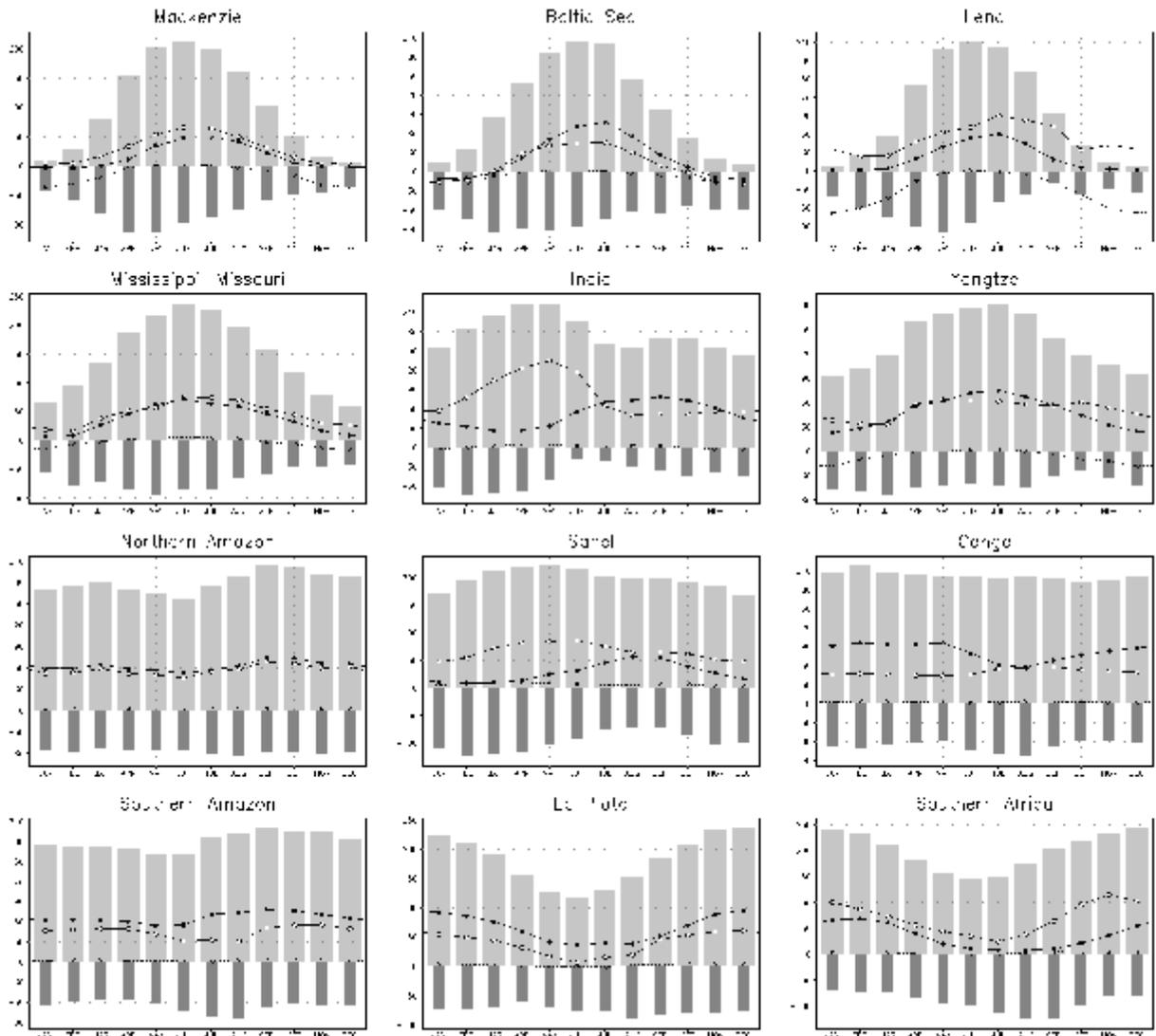


Figure 11 As in Fig 10 for the surface energy balance terms; wide bar = net shortwave radiation, narrow bar = net longwave radiation (both positive downward); filled circle = latent heat flux, open circle = sensible heat flux, cross with dotted line = ground heat flux (positive upward). Units are $W m^{-2}$.

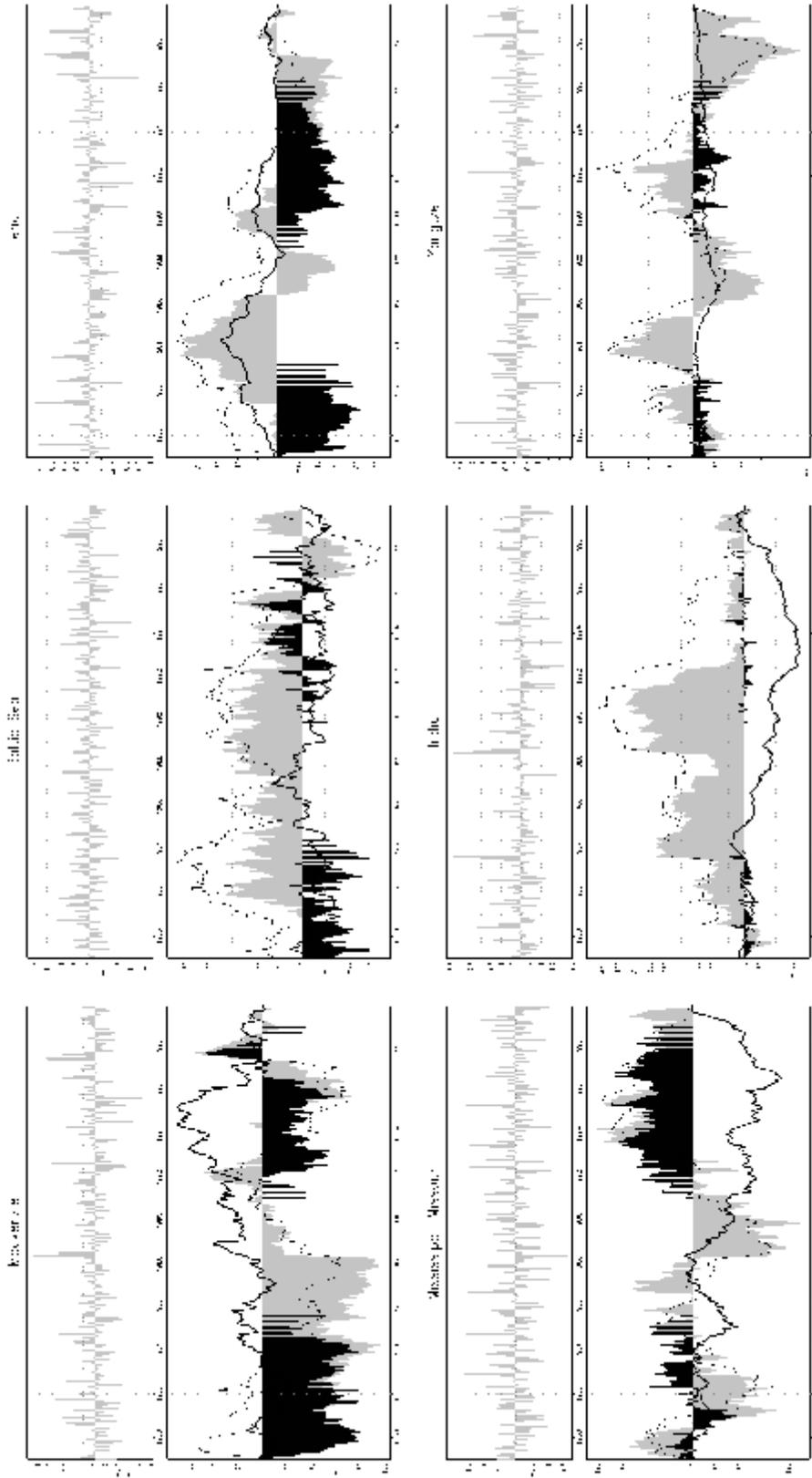


Figure 12 Monthly precipitation anomalies (top) and cumulative anomalies in water balance terms relative to the 21-year mean annual cycle over the areas shown in Fig 9. For the cumulative anomalies; grey shading = precipitation, black bars = change in storage (soil moisture + snow), bold line = evapotranspiration, thin line = runoff. Units are mm d^{-1} (top); mm (bottom).

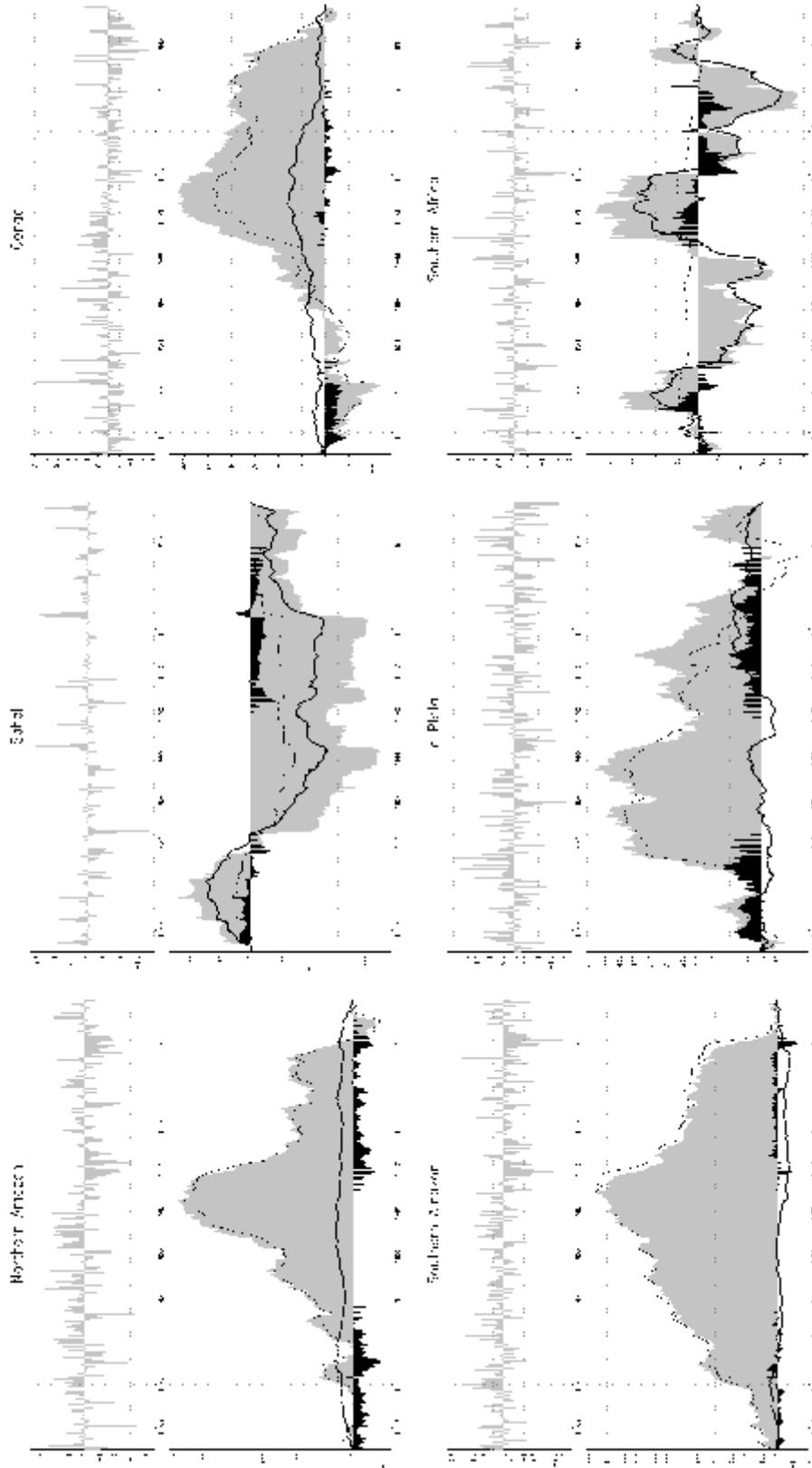


Figure 12 Cont.