A 53-year forcing data set for land surface models

T. Ngo-Duc, J. Polcher, and K. Laval

Laboratoire de Météorologie Dynamique du CNRS, Paris, France

[1]   As most variables describing the state of the surface are not directly observable, we have to use land surface models in order to reconstruct an estimate of their evolution. These large-scale land surface models often require high-quality forcing data with a subdiurnal sampling. Building these data sets is a major challenge but an essential step for estimating the land surface water budget, which is a crucial part of climate change prediction. To study the interannual variability of surface conditions over the last half century, we have built a 53-year forcing data set, named NCC. NCC has a 6-hourly time step from 1948 to 2000 and a spatial resolution of $1° \times 1°$. It is based on the National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis project and a number of independent in situ observations. In this study we show the adjustments which need to be applied to the reanalysis and how they impact the simulated continental water balance. The model outputs are validated with the observed discharges of the world's 10 largest rivers to estimate the combined errors of the forcing data and the land surface model. The seasonal and interannual variations of these discharges are used for this validation. Five numerical experiments have been carried out. They used the forcing data sets obtained after each step of data adjustment and the forcing of the Global Soil Wetness Project 2 as inputs for the Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) land surface model. The quality of forcing data is improved after each adjustment. The precipitation correction gives the most important improvement in the simulated river discharges, while the temperature correction has a significant effect only at high latitudes. The radiation correction also improves the forcing quality, especially in term of discharge amplitude. The NCC forcing data set can be used to study the water budget over many areas and catchment basins that have not been yet analyzed in this study. With its period of 53 years, NCC can also be used to evaluate the trends of terrestrial water storage in particular regions.

1. Introduction

[2]   Land surface models (LSMs) were initially developed for coupling to general circulation models (GCMs). In the last few years, the range of complexity among the land surface parameterizations has grown significantly. To improve our understanding and parameterizations of land surface processes, and to eliminate some biases that can be generated by the GCMs, the LSMs have been applied in an off-line mode. To be used in this stand-alone mode, a high quality prescribed atmospheric forcing data with a subdiurnal sampling is required. Building these data sets is a major challenge but an essential step for estimating the land surface water budget.

[3]   Up to present, there have been several attempts to produce atmospheric forcing data sets for use in the land surface modeling community. *Meeson et al.* [1995] produced the International Satellite Land surface Climatology Project (ISLSCP) Initiative I data set for 1987–1988. The ISLSCP Initiative I data were used as an upper boundary forcing for the LSMs in the pilot phase of the Global Soil Wetness Project (GSWP1) [*Dirmeyer et al.*, 1999]. Recently, in the framework of GSWP2 [*Dirmeyer et al.*, 2002], the Center for Ocean-Land Atmosphere Studies (COLA) has produced the near-surface data set for ISLSCP Initiative II [*Hall et al.*, 2003] from the NCEP/DOE reanalysis [*Kanamitsu et al.*, 2002]. This data set includes near-surface meteorology at a 3-hourly interval for the 10-year (1986–1995) and has been used to force LSMs in GSWP2. Another forcing data set can be mentioned here is VIC Retrospective Land Surface Data Set: 1950–2000 [*Maurer et al.*, 2002], which is for the conterminous United States and has a 3-hourly time step from 1950 to 2000 and a spatial resolution of 1/8 degree.

[4]   All these data sets are reanalysis estimates combined with gridded data sets from observations. Reanalysis or retrospective analysis are the production of longterm analysis using a frozen modeling and assimilation framework [*Bengtsson and Shukla*, 1988]. They are produced by

Table 1. Atmospheric Forcing Variables for LSMs

Name	Description	Units
Tair	near-surface air temperature at 2 m	K
Qair	near-surface specific humidity at 2 m	kg kg^{-1}
Wind	near-surface wind speed at 10 m	m s^{-1}
Psurf	surface pressure	Pa
SWdown	surface incident shortwave radiation	W m^{-2}
LWdown	surface incident longwave radiation	W m^{-2}
Rainf	rainfall rate	kg m^{-2} s^{-1}
Snowf	snowfall rate	kg m^{-2} s^{-1}

assimilating the most current set of atmospheric observations into a global circulation model of the atmosphere. These observations are obtained from surface station observations, radiosonde, aircraft and, in recent decades, satellite retrievals. Both the National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF) have produced global reanalyses spanning from 15 to more than 50 years, and other centers have produced additional reanalyses that are more limited in time or space. *Kalnay et al.* [1996] have shown that the reanalyses strongly reflect the biases and errors of the model used, particularly in the flux estimates (e.g., precipitation, radiation) and in the state variables (e.g., temperature, pressure) in regions that have little observational data input. The adjustment process of the reanalyses by global observationally based data sets is thus necessary in building a reliable meteorological data set to force LSMs.

[5] LSM simulations allow us to estimate the global land surface water and energy cycles. However, the period of actual available global forcing data sets (e.g., the period of ISLSCP-I and ISLSCP-II is 1987–1988 and 1986–1995, respectively) cannot be compared with remote sensing data, and it is too short for the detection and analysis of trends, such as those associated with global warming. A longer period will allow for a better study of interannual land surface climate variability, and also for application and further development of the methods of calibration, evaluation and validation of LSMs with in situ and remote sensing data.

[6] The following study takes place in this context and aims at building an atmospheric forcing data sets for LSMs. We began from the 53-year (1948–2000) pure reanalysis products of NCEP/NCAR [*Kistler et al.*, 2001] then used the observationally based data of the Climate Research Unit (CRU) from University of East Anglia [*New et al.*, 1999, 2000] and the Surface Radiation Budget (SRB) data produced at NASA Langley Research Center to correct the reanalysis products. The correction process is done with the same methods as those used to build GSWP2 near-surface meteorology data sets [*Dirmeyer et al.*, 2002; *Zhao and Dirmeyer*, 2003]. The new forcing data set is named NCC (NCEP/NCAR Corrected by CRU). NCC is a 6-hourly forcing data with a spatial resolution of 1° for the period of 1948–2000. The main reason for us to choose CRU data to correct reanalysis products is that CRU is the only available observationally based data over this long period.

[7] Our goal in this study is to construct and validate our new 53-year atmospheric forcing data sets NCC using the Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) land surface model [*Verant et al.*, 2004; *Krinner et al.*, 2005]. We show the importance of the

corrections applied to the reanalysis and how these corrections impact the simulated continental water budget.

[8] In section 2, the construction of the NCC data will be described. The ORCHIDEE LSM and experiments will be briefly presented in section 3. Section 4 will focus on the validation of NCC data. Section 5 is a comparison using ORCHIDEE of the NCC and GSWP2 forcing data sets over the common period. Finally, conclusions will be presented in section 6.

2. NCC Data Construction

[9] The input data for LSMs are generally divided into three categories: soils data (fixed in time), vegetation data (some fixed and some monthly varying) and meteorological data. The soil and vegetation data are parameter data sets that are used to specify characteristics of the land surface. The meteorological data provide the forcing at the upper boundary of the land surface.

[10] In this section, we aim to construct a new 53-year meteorological data for LSMs. The variables in the meteorological data (Table 1) are divided into two types: state variables (near-surface air temperature, specific humidity, wind speed and surface pressure) and flux fields (radiation and precipitation). The data construction involves two steps: interpolation of the NCEP/NCAR Reanalysis data to a grid of 1° × 1° and correction of reanalysis data with the observationally based data.

[11] To facilitate the exchange of forcing data for LSMs and the results produced by these models, the Global Energy and Water Cycle Experiment (GEWEX) Global Land Atmosphere System Study (GLASS) established the Assistance for Land surface Modeling activities (ALMA, http://www.lmd.jussieu.fr/ALMA/) convention for LSM input and output variables. The aim is to have a data exchange format which is stable but still general and flexible enough to evolve with the needs of LSMs. This should ensure that the implementation of procedures to exchange data only needs to be done once and that future intercomparisons of LSMs will be run more efficiently. NCC data were thus saved using the ALMA convention.

2.1. Interpolation of the NCEP/NCAR Reanalysis Data

[12] We began with the near-surface meteorological variables from NCEP/NCAR Global Reanalysis Products [*Kistler et al.*, 2001]. Reanalysis have T62 (Gaussian grid) resolution (192 × 94 grid boxes globally) and were saved 6-hourly for the period from 1948 to present. To keep the procedure close to the GSWP2 method, the reanalysis data was first regridded to 1° resolution. The method used here is the inverse distance weighted interpolation, sometimes called "Shepard's method" [*Shepard*, 1968]. Inverse distance weighted methods are based on the assumption that the interpolating surface should be influenced most by the nearby points and less by the more distant points. The reanalysis land-sea mask and the ISLSCP land-sea mask are used to ensure that only land points are transformed into land points.

[13] This method of interpolation is directly applied for near-surface wind speed at 10 m, surface incident radiation and precipitation rate. For near-surface air temperature, surface pressure and specific humidity, the procedure is a

little different because these variables must be corrected for differences in elevation between the reanalysis model topography and the NCC topography. To interpolate temperature and pressure, the key technique is to extrapolate first these variables to sea level height using the reanalysis topography, to regrid them to the $1°$ resolution and then to calculate their new values with the NCC topography. The vertical extrapolation needed for the two data sets uses

$$T_z = T_0 - \gamma z \qquad (1)$$

where T is temperature, z is height above mean sea level (T_0 is thus the temperature at the sea level), and $\gamma = -\partial T/\partial z$ is the environmental lapse rate. In this study, we use $\gamma = 0.65°C/100$ m.

$$p_z = p_0 \left(\frac{T_0 - \gamma z}{T_0} \right)^{g/\gamma R_a} \qquad (2)$$

where p is pressure and R_a is gas constant of air ($R_a = R/m_{air} = 287$ J kg^{-1} K^{-1}).

[14] The methodology for computing humidity at the $1° \times 1°$ resolution has been chosen so that the relative humidity remains constant with pressure and temperature changes caused by their interpolation. The actual vapor pressure and the saturated vapor pressure both change, but by the same factor. Thus, to interpolate specific humidity, the following steps are taken: (1) calculate specific humidity at saturation $q_s(T_{NCEP}, p_{NCEP})$, (2) calculate relative humidity $f_{NCEP} = q/q_s$, (3) interpolate f_{NCEP} to f_{NCC} in using Shepard's method, and (4) recalculate specific humidity in the new grid $q_{NCC} = f_{NCC} * q_s(T_{NCC}, p_{NCC})$.

2.2. Correction of the 53-year NCEP/ NCAR Data

[15] After the interpolation step, we have the new 53-year meteorological data set on the $1° \times 1°$ grid. This data set is a pure reanalysis product. As discussed above, it should be amended by combining with gridded observational data to remove systematic errors in the reanalysis fields. The observational data are thus required to have a high resolution in space and cover a long period. There are few data sets that satisfy this demand. Notable exceptions are the monthly time step Global Precipitation Climatology Project (GPCP) data set [*Xie and Arkin*, 1996; *Xie et al.*, 1996], the monthly 1900–1988, $2.5° \times 2.5°$ precipitation data set of *Dai et al.* [1997] (hereinafter referred to as Dai), and the $0.5° \times 0.5°$ daily time step data set being developed by *Piper and Stewart* [1996, henceforth PS]. These products either cover relatively short periods (1970s to present; GPCP, PS) or are coarse resolution (Dai). The precipitation and temperature of the CRU data set [*New et al.*, 1999, 2000] appear to be the best choice to correct NCEP/NCAR products. The CRU data set is a high-resolution ($0.5°$) monthly product over continents only which includes for the 1901–2000 period a gauge-only estimate of precipitation as well as other near-surface climatic variables.

[16] Because no observationally based radiation data for the period of 1948–present is available, the Surface Radiation Budget (SRB) data produced at NASA Langley Research Center for the period of 1983–1995 will be

applied here for a simple bias correction of the reanalysis product.

2.2.1. Precipitation

[17] The hybridization of the reanalysis with CRU precipitation has been done in using the process proposed in GSWP2 [*Dirmeyer et al.*, 2002; *Zhao and Dirmeyer*, 2003]. Monthly CRU data are first transformed from the original $0.5°$ resolution to the $1°$ resolution of NCC. Then, the reanalysis systematic errors are removed via a multiplicative scaling factor based on the ratio of observed monthly rainfall to reanalysis estimates:

$$[P]_{Y,M,D,T} = \frac{[P_{CRU}]_M}{[P_{NCEP}]_M} [P_{NCEP}]_{Y,M,D,T} \qquad (3)$$

[18] To adjust the precipitation data, the value at a grid box of the reanalysis precipitation at a given year (Y), month (M), day (D) and 6-hourly time interval (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 such as precipitation.

[19] The partition of snowfall and rainfall is calculated by a simple function of total precipitation and near-surface air temperature (Tair). When Tair is smaller than 273.15 K, precipitation is assumed to be all snowfall. When Tair is greater or equal to 273.15 K, precipitation is assumed to be rainfall.

2.2.2. Atmospheric State Variables

[20] To make a consistent adjustment of near-surface air temperature to the $1° \times 1°$ grid, the CRU temperatures were corrected for the altitude difference between the CRU grid and the NCC grid. The monthly CRU temperature data were aggregated from $0.5°$ to $1°$, and then used to hybridize the 6-hourly NCEP/NCAR reanalysis 2-m air temperature data by correcting the differences of monthly mean:

$$[T]_{Y,M,D,T} = [T_{NCEP}]_{Y,M,D,T} + [T_{CRU}]_M - [T_{NCEP}]_M \qquad (4)$$

[21] The correction of temperature will affect the surface pressure, which depends on the temperature (equation (2)), and will also affect the estimated saturation specific humidity. Thus it is necessary to adjust the estimates of surface pressure and specific humidity from the reanalysis. This is done by using the equation (2) with the temperature before and after the correction. In the new 53-year NCC atmospheric forcing data, the reanalysis wind products have not been corrected and are used as they are.

2.2.3. Radiation

[22] To our knowledge there are no radiation data sets over the period of 1948 to present available with a high enough resolution in space and time to be used here for the correction of the reanalysis. The SRB data produced at NASA Langley Research Center for the period of 1983–1995 which is available at a resolution of $1° \times 1°$ in space and 3 hours in time seems to be the best choice and is thus used here for a simple bias correction of the reanalysis product. Thus NCC uses the same observational data as GSWP2 for the hybridization of radiation.
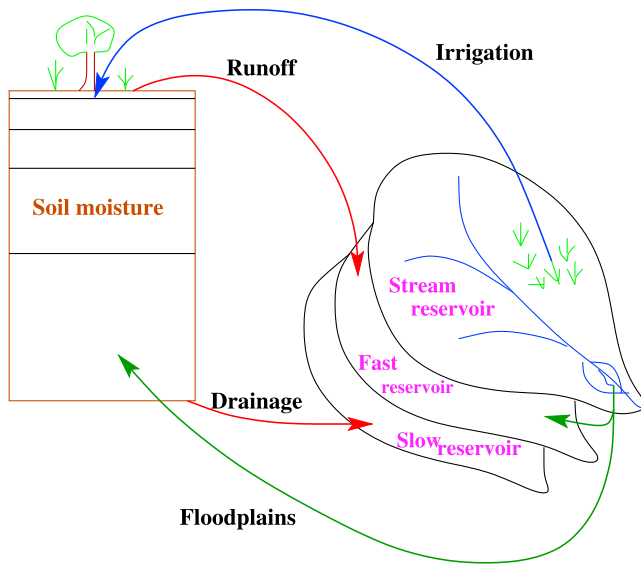
Figure 1. Principle of the river routing scheme.

[23] The equation for generating a bias corrected downward shortwave radiation is

$$[SW]_{Y,M,D,T} = \frac{[SW_{\mathrm{SRB}}]_{\overline{M}}}{[SW_{\mathrm{NCEP}}]_{\overline{M}}}[SW_{\mathrm{NCEP}}]_{Y,M,D,T} \qquad (5)$$

where SW is surface incident shortwave radiation. The index $\overline{M}$ means the mean monthly value of SW for the period of 1983–1995. The same calculation is applied for downward long-wave data.

3. Model Description and Experimental Design
3.1. Brief Description of ORCHIDEE LSM and the Runoff Routing Scheme

[24] The description of ORCHIDEE is given by *Verant et al* [2004] and *Krinner et al.* [2005], but the key items are briefly presented here. ORCHIDEE is the new LSM of the Pierre-Simon Laplace Institute (IPSL). It has been developed for regional studies either within a GCM or in a stand-alone mode. It is composed of (1) the previous LSM of the Laboratoire de Météorologie Dynamique, SECHIBA, which computes the physical processes at the interface between soil, vegetation and atmosphere, the water fluxes in the soil and the control of evaporation by soil moisture [*Ducoudré et al.*, 1993; *de Rosnay and Polcher*, 1998], (2) the carbon cycle model of the Laboratoire des Sciences du Climat et de l'Environnement, STOMATE, which simulates the biochemical processes at the surface [*Viovy*, 1996], and (3) a representation of the dynamical evolution of the vegetation and the carbon budget derived from the LPJ (Lund-Potsdam-Jena) model [*Sitch*, 2000]. The two last components of ORCHIDEE are not used in this study.

[25] The vegetation distribution in ORCHIDEE is based on the International Geosphere Biosphere Programme (IGBP) land cover map [*Belward et al.*, 1999]. The soil hydrology consists of two moisture layers with the upper one having a varying depth. The total soil depth is constant at 2 m and the soil has a maximum water content per unit of soil volume [*de Rosnay and Polcher*, 1998]. Runoff occurs when the soil is saturated and it is the only runoff mechanism in the model. A new development of the model is to include a routing scheme, which uses a map of the world basins built by combining the map built by *Vörösmarty et al.* [2000] and the one built by *Oki et al.* [1999]. At each time step, runoff and "drainage" fluxes are temporarily stored in three reservoirs which have different residence time constants (Figure 1). The water is progressively routed to the oceans, following the main slopes and taking into account the tortuous path of the river channels, through a cascade of linear reservoirs.

[26] The representation of vertical and horizontal water fluxes in the same model allow us to simulate the impact of floodplains and irrigation on the continental water cycle and to represent endorheic basins which are not connected to oceans by rivers. The water carried by rivers in endorheic basins will return to the soil moisture reservoir at the point of convergence. The floodplains outside of the high latitudes are treated by a simple parameterization which disperses the flood wave and returns waters to the soil moisture. This process is only activated in regions identified as wetlands in the vegetation map. The parameterization of the irrigation is described by *de Rosnay et al.* [2003].

3.2. Experimental Design

[27] In sections 4 and 5, we analyze the outputs of ORCHIDEE LSM forced by different meteorological data to validate our new NCC data. In the NCC data construction process, we have obtained several atmospheric forcing data sets for 53 years from 1948 to 2000 as described in Table 2. Each forcing data set contains eight variables (see Table 1) which may either be the pure reanalysis products (NCEP), the reanalysis corrected by CRU data (CRU) or the reanalysis corrected by the SRB data.

[28] These data sets are used to force the ORCHIDEE LSM, and the model output obtained with these respective data sets will also be called NCEP, NPRE, NCRU and NCC. In section 5, the GSWP2 forcing data are also used for an ORCHIDEE simulation and the obtained output for the period 1986–1995 will be called GSPW2.

4. Validation of NCC and ORCHIDEE LSM

[29] As most variables describing the state of the surface are not directly observable, to assess the quality of

Table 2. Description of the Atmospheric Forcing Data Sets Constructed to Evaluate the Added Value of NCC[a]

Forcing data sets	Rainf	Snowf	Tair	Qair	Psurf	SWdown	LWdown	Wind
NCEP	NCEP	NCEP	NCEP	NCEP	NCEP	NCEP	NCEP	NCEP
NPRE	CRU	CRU	NCEP	NCEP	NCEP	NCEP	NCEP	NCEP
NCRU	CRU	CRU	CRU	CRU	CRU	NCEP	NCEP	NCEP
NCC	CRU	CRU	CRU	CRU	CRU	SRB	SRB	NCEP

[a]NCEP, pure reanalysis products; CRU, reanalysis corrected by CRU data; and SRB, reanalysis corrected by SRB data.

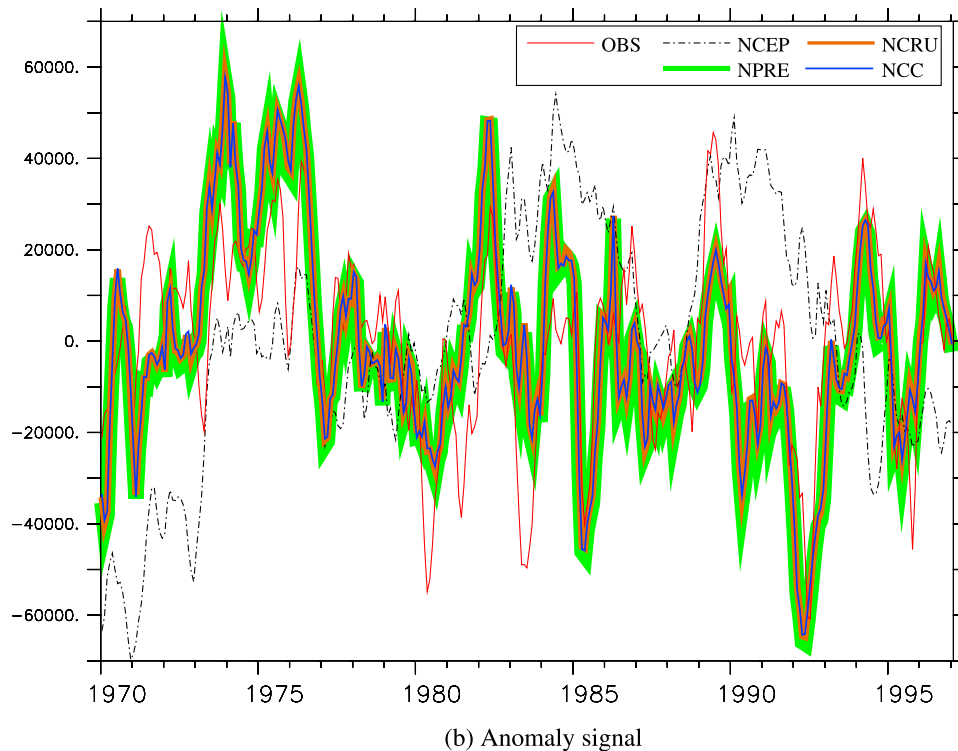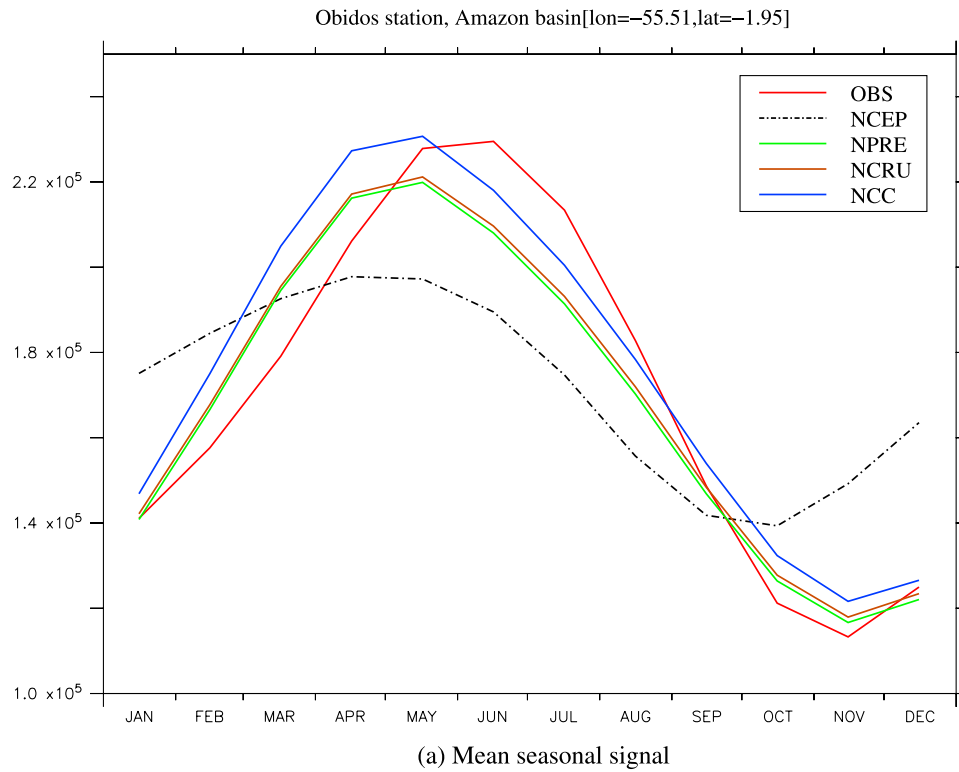(a) Mean seasonal signal

(b) Anomaly signal

Figure 2. Discharge from 1970 to 1997 simulated by NCEP, NPRE, NCRU, and NCC experiments compared to the observed river discharge at Obidos, Amazon: (a) mean seasonal signal and (b) anomaly signal. Because of the similarity of the NPRE, NCRU, and NCC anomaly curves, different thicknesses are used to distinguish them. Units are in m^3/s.

ORCHIDEE LSM, we take advantage of the integrated routing scheme which allows us to compare the simulated and the observed river discharge over the largest river basins. River discharge is an appropriate observable measure to validate the large-scale water balance. The river discharge measurements used here are the data set provided by the Global Runoff Data Center (GRDC) (http://www.grdc.sr.unh.edu/) and the data set provided by UCAR
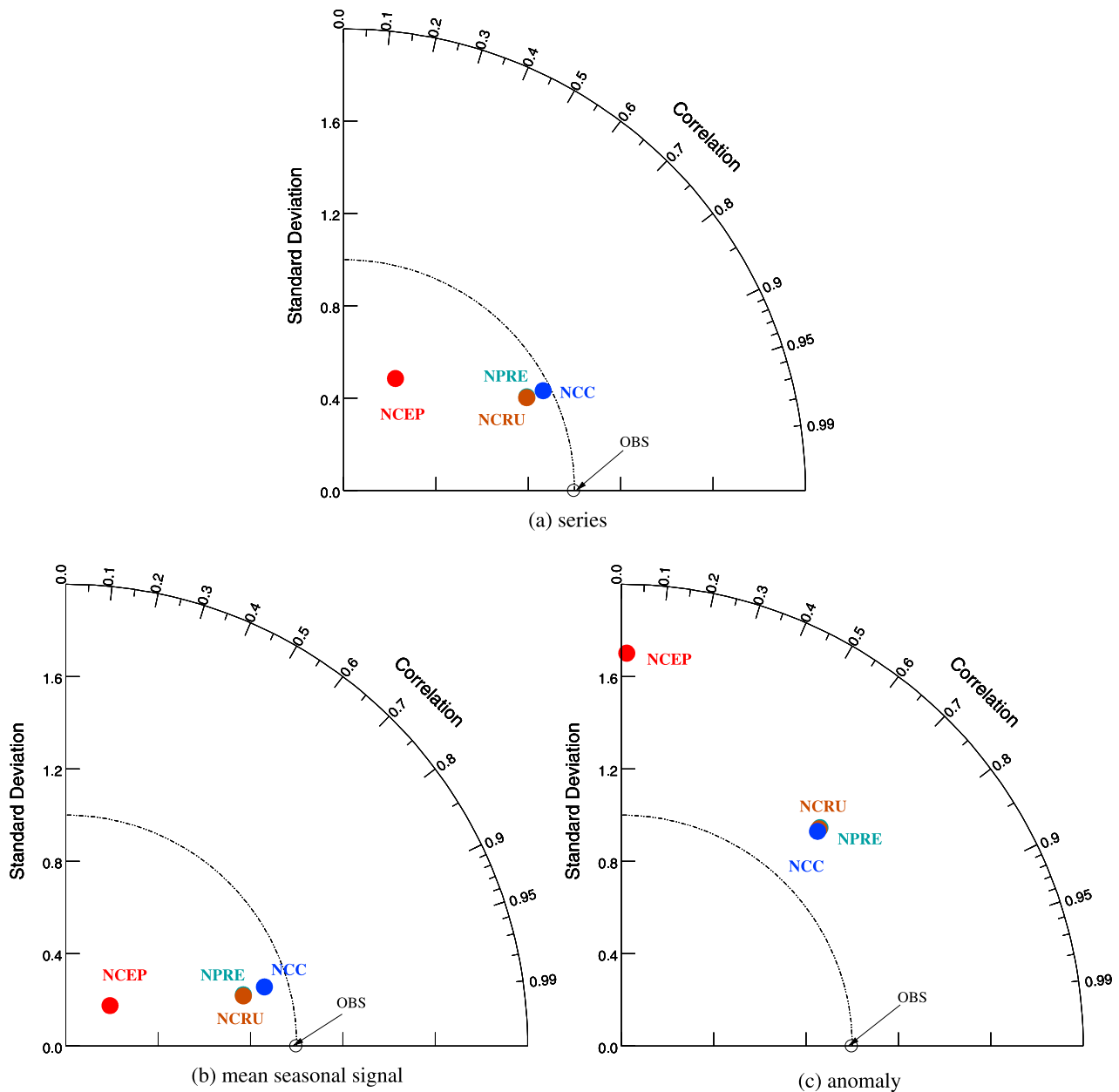
Figure 3. Taylor diagram illustrating the statistics of Amazon river discharge at Obidos simulated by NCEP, NPRE, NCRU, and NCC experiments compared against the observations from 1970 to 1997 for (a) full series, (b) mean seasonal signal, and (c) anomaly.

(http://dss.ucar.edu/data sets/ds552.1/). The last data set contains monthly river discharge rates for 4425 locations around the world except for the former Soviet Union. This data set has certain observations which do not exist in the GRDC data, for example, the discharge measurements at the station Timbues of Paraná, the measurements from 1983 to 1997 at the station Vicksburg of Mississippi, etc. In this study, the GRDC data were used for the stations in the former Soviet Union. When the simulated river discharges and the observations are compared, only the period overlap is used to compute statistics.

4.1. The Amazon Basin, a Test Case

[30] The objective of the present section is to show the importance of data correction as well as to prove the quality

of NCC data set over the Amazon basin, which is the world's largest basin with a total area of 6.14×10^6 km^2.

[31] The result obtained over the Amazon basin is illustrated in Figure 2. Figure 2 represents the mean seasonal (Figure 2a) and the anomaly signals (Figure 2b) of the river discharge observed and simulated at the station Obidos of the Amazon (1.95°S, 55.51°W) from 1970 to 1997. Figure 2a shows that the mean seasonal discharge simulated by NCEP experiment (black dotted curve) is lower than the observations (red curve) and reflects well the error in the precipitation of the reanalysis. After correcting the precipitation (NPRE experiment, green curve), the simulated river discharge gives more satisfactory results. Figure 2b shows that NPRE experiment describes well the interannual signal from 1970 to 1990 and even better in the 1990s.

Table 3. The 10 Stations Closest to the Ocean of the World's 10 Largest Rivers[a]

No	Station	River	Longitude	Latitude	Period
1	Obidos	Amazon	−55.51	−1.94	1970–1997
2	Kinshasa	Congo	15.30	−4.30	1950–1982
3	Puente Angostura	Orinoco	−63.60	8.15	1950–1989
4	Datong	Changjiang	117.62	30.77	1950–1988
5	Bahadurabad	Brahmaputra	89.67	25.18	1973–1975
6	Vicksburg	Mississippi	−90.90	32.31	1950–1997
7	Igarka	Yenisey	86.50	67.48	1950–1983
8	Timbues	Paraná	−60.71	−32.67	1950–1993
9	Kusur	Lena	127.65	70.70	1950–1983
10	Pakse	Mekong	105.80	15.12	1982–1984

[a]Arranged by the estimated river discharge. Period indicates the years within 1948–2000 where we have a continuously observed discharge.

[32] The correction of temperature (NCRU experiment, brown curve) changes slightly the Amazon discharge (maximum 1.5% in amplitude). This correction becomes more important in the regions of high latitude (about 25% in amplitude and 1 month of phase shift over the Lena basin (see section 5). The additional correction of radiation (NCC experiment, blue curve) increases the amplitude of the simulated discharge and provides NCC discharge a more realistic amplitude (Figure 2a). However, we should note that the high flow in the mean seasonal signal is too soon with a phase shift of about 1 month in the simulations. The temperature and radiation corrections have a very little effect on the anomaly signal (Figure 2b).
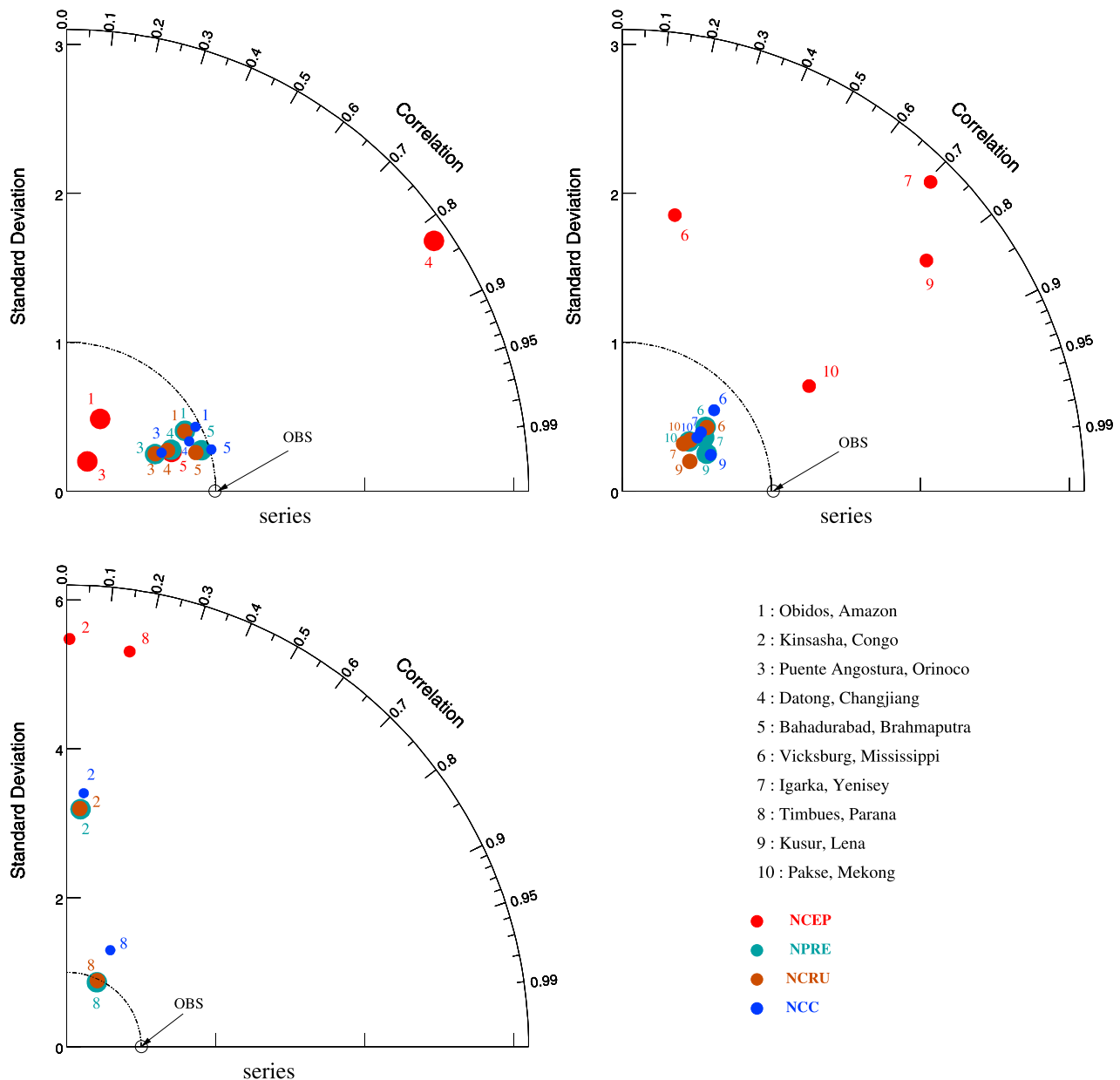

1 : Obidos, Amazon

2 : Kinsasha, Congo

3 : Puente Angostura, Orinoco

4 : Datong, Changjiang

5 : Bahadurabad, Brahmaputra

6 : Vicksburg, Mississippi

7 : Igarka, Yenisey

8 : Timbues, Parana

9 : Kusur, Lena

10 : Pakse, Mekong

● NCEP

● NPRE

● NCRU

● NCC

Figure 4. Taylor diagram illustrating the statistics of 10 largest rivers discharges simulated by NCEP, NPRE, NCRU, and NCC experiments compared against the observations for the whole common (observation and simulation) period. Note the different axis scales between the plot that represents two stations Kinsasha and Timbues and the two other plots.
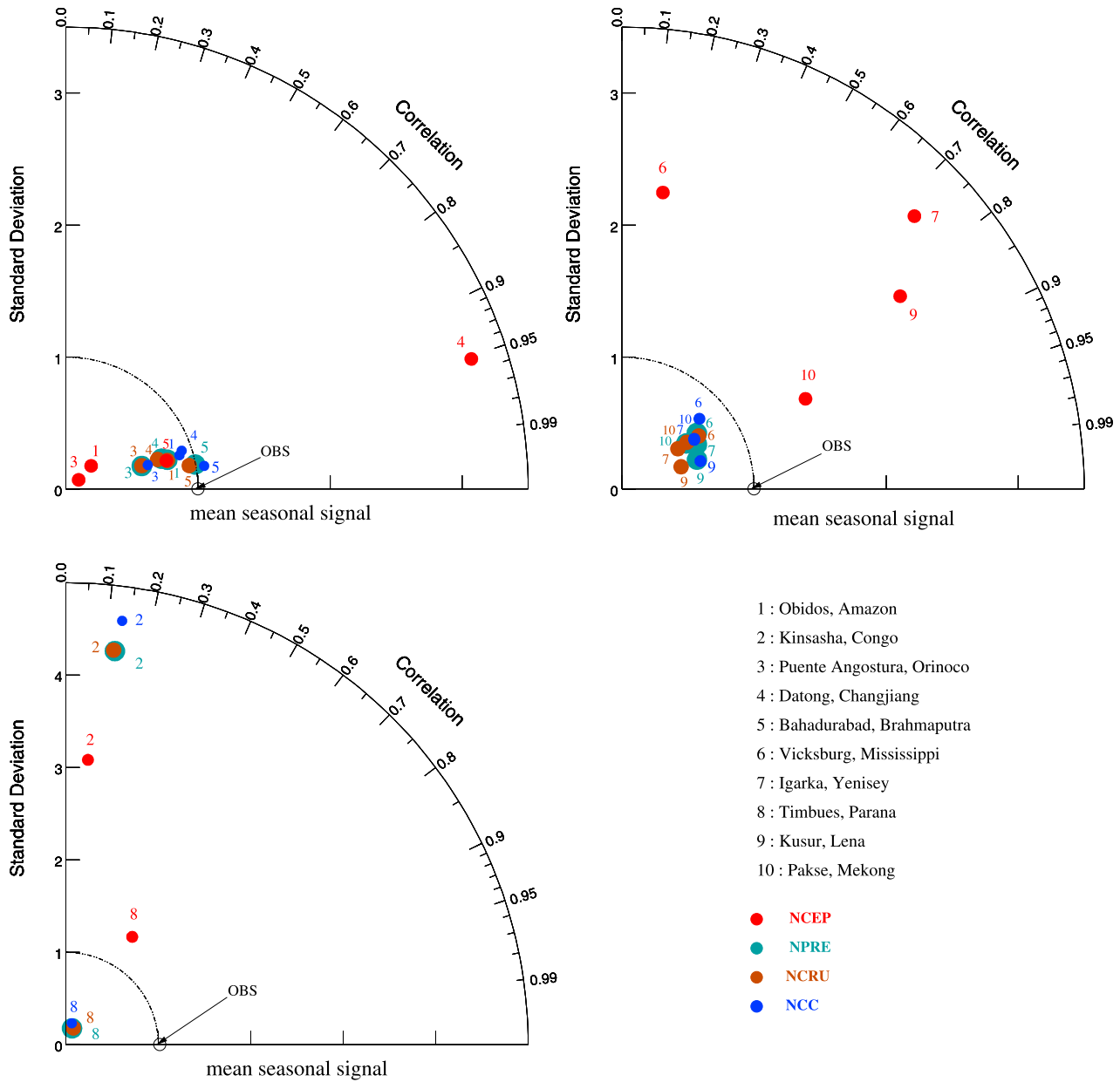
Figure 5. Same as Figure 4 but for mean seasonal signal.

[33] To provide a more condensed view of these results, Figure 3a displays a Taylor diagram [*Taylor*, 2001], which shows the error in the simulated discharge for the full time series. This error is then decomposed into the part present in the mean annual cycle and the part associated to the interannual variations and displayed using two other Taylor diagrams (Figures 3b and 3c). The Taylor diagrams provide the ratio of standard deviation as a radial distance and the correlation with observations as an angle in the polar plot for all four simulations. Consequently, the observed behavior of the Amazon's discharge at Obidos is represented by a point on the horizontal axis (zero correlation error) and at unit distance from the origin (no error in standard deviation). In this coordinate system, the linear distance between each experiment's point and the "observed" point is proportional to the root mean square model error. Figure 3a is a discharge comparison for the entire period 1970–1997 for which observations are available. It shows the qualities of the river discharges simulated by the 4 experiments. NCEP discharge is far from the observations, its simulated variance is underestimated at 50% of observed and its correlation with observations is only of 0.4. NPRE and NCRU are better simulations and very similar because of the small effect of temperature correction. The NCC simulation has the best simulated discharge. It closely matches the observed magnitude of variance and exhibit a correlation of about 0.9 with the observations.

[34] Figures 3b and 3c illustrate the statistics of the mean seasonal signal and the anomaly of the simulated Amazon discharge compared against the observations, respectively. They strengthen the results from Figure 3a: over the Amazon, the quality of simulated discharge depends principally on the quality of precipitation input; and the temperature correction has little effect on the simulated
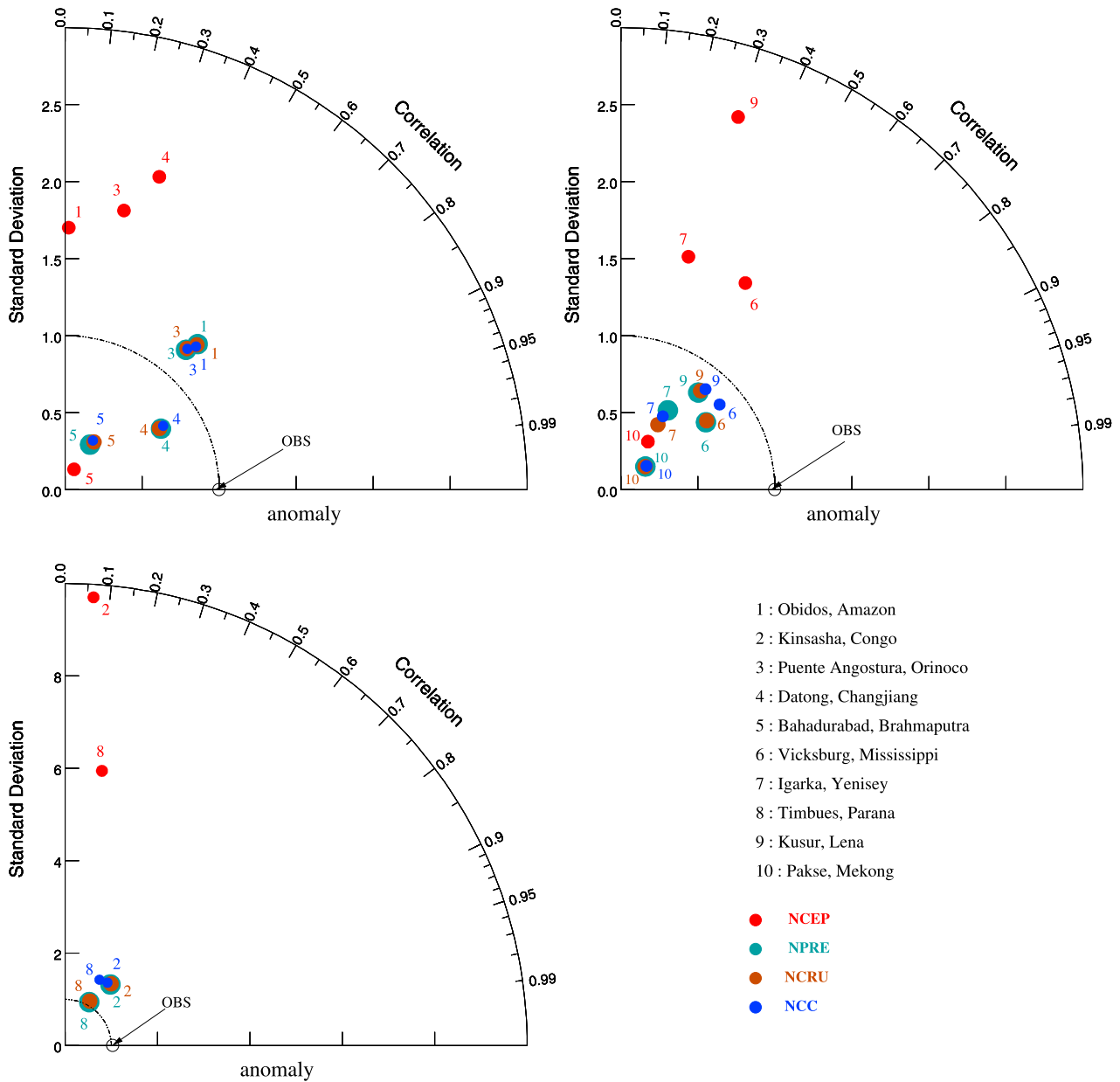
Figure 6. Same as Figure 4 but for anomaly signal.

discharge. For the mean seasonal signal, the NCC simulation shows the best amplitude but its correlation with the observations (about 0.9) is slightly smaller than the ones for NPRE and NCRU (about 0.95). In Figure 3c, the correlation is not as good as in Figures 3a and 3b (about 0.7 for the three experiments NPRE, NCRU, and NCC) and the variance is too strong and points to problems in the simulated interannual variability which will be discussed later.

4.2. The 10 Largest Rivers

[35] The objective of this section is to show the improvement of 53-year forcing data set after each step of data correction over the world's 10 largest rivers (by the estimated river mouth flow rate). Table 3 presents the 10 stations closest to the mouth of these rivers and the period in common for the observations and the numerical experiments. For most of these rivers a station could be

found with a long enough record to provide a meaningful evaluation. Notable exceptions are the Brahmaputra and the Mekong. This choice of basins and stations is biased toward the tropical basins which is favorable to our model as the treatment of soil and river freezing is still simplistic and an evaluation of the atmospheric forcing in the high latitudes with ORCHIDEE would not be very meaningful. Over the Congo basin, there are two discharge stations at nearly the same location, Kinshasa and Brazzaville. Kinshasa provides continuously observations from 1903 to 1983 while for Brazzaville only the period 1971 to 1989 is available. During their common period (1970–1983) the recorded discharges show discrepancies of about 6%. In this section, the discharge measured at Kinshasa station from 1950 to 1983 are used because of the longer common period with NCC. Later, in the comparison with GSWP2 the observations from Brazza-
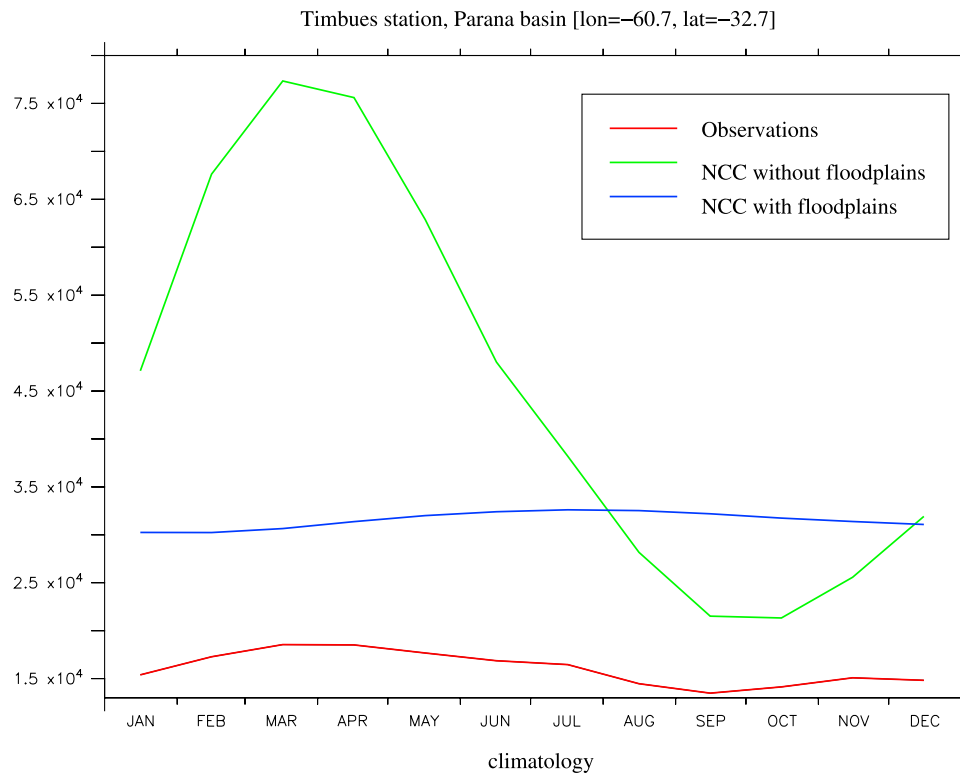
Timbues station, Parana basin [lon=−60.7, lat=−32.7]

Figure 7. Mean seasonal signals observed, simulated by NCC experiment with and without floodplains at Timbues station, Paraná. Units are in m^3/s.

ville are used because the overlap for the 1986–1996 period is larger.

[36] Figures 4, 5, and 6 illustrate the statistics of the simulated world's 10 largest river discharges compared to the observations for the whole observed period, the mean seasonal signal, and the anomaly signal, respectively. For almost all the basins, Figure 4 shows very clearly that the data have better quality after each correction. The simulated discharges of the Amazon, Changjiang, and Brahmaputra are quite realistic; their correlations with the observations vary from 0.9 to 0.95 and their normalized standard deviations are very close to 1. For all basins, discharges obtained for NPRE show a great improvement compared to the NCEP experiment, which illustrates that precipitation biases input are the main reason of nonrealistic simulated discharges.

[37] The temperature correction (NCRU) has a small effect on the river discharge at low and middle latitudes. At high latitudes (see Yenisey and Lena basin), temperature correction becomes more important because change in temperature will also affect the precipitation input (partition between rain and snow), which consequently changes the processes at the surface.

[38] The radiation correction (NCC) improves the forcing quality, especially in term of discharge amplitude. This correction has nearly no effect in term of phase as only a bias correction was performed in the reanalysis products.

[39] For the Kinshasa station (Congo basin), the simulated discharges is practically uncorrelated with the observations for all experiments. As discussed by *Oki et al.* [1999], the low rain gauge density [*New et al.*, 2000] leads to poor precipitation estimates over the basin which in turn degrades

considerably the simulated discharge. If one can use ORCHIDEE as an indirect evaluator of precipitation climatologies, one would have to conclude that the annual cycle of precipitation is better represented in the reanalysis than by the estimation based on rain gauges. However, before one would reach such a conclusion ORCHIDEE will need to be more thoroughly evaluated, especially the parameterization of floodplains, as they play an important role in the Congo basin.

[40] Except for poor forcing precipitation, this bias may also be explained by the fact that ORCHIDEE does not properly represent the natural dissipation of water from river channels to surrounding land, water used by the cities, irrigation and dam constructions.

[41] Figure 5 shows that there is a small variance ratio and a weak correlation in the mean seasonal cycle of the simulated discharge at Timbues (Paraná basin). The main cause of this deficiency is the existence of floodplains over the Paraná basin (see Figure 7) which flatten the annual cycle and which are very difficult to represent in a coarse land surface model. The inclusion of a representation of floodplains in the ORCHIDEE LSM improves the amplitude of simulated Paraná river discharge; however, large discrepancies with seasonal variations remain.

[42] In Figure 4 one may note that except for the Congo and Paraná, the simulated discharges tend to be lower than the observations. Too high temperature, too strong wind, too high incoming solar radiation, too low humidity in the forcing data or inadequate land surface parameters could be the causes of discharge underestimates. Among them, *Oki et al.* [1999] have considered two causes. The first one is related to the disaggregation into 6 hourly of the CRU
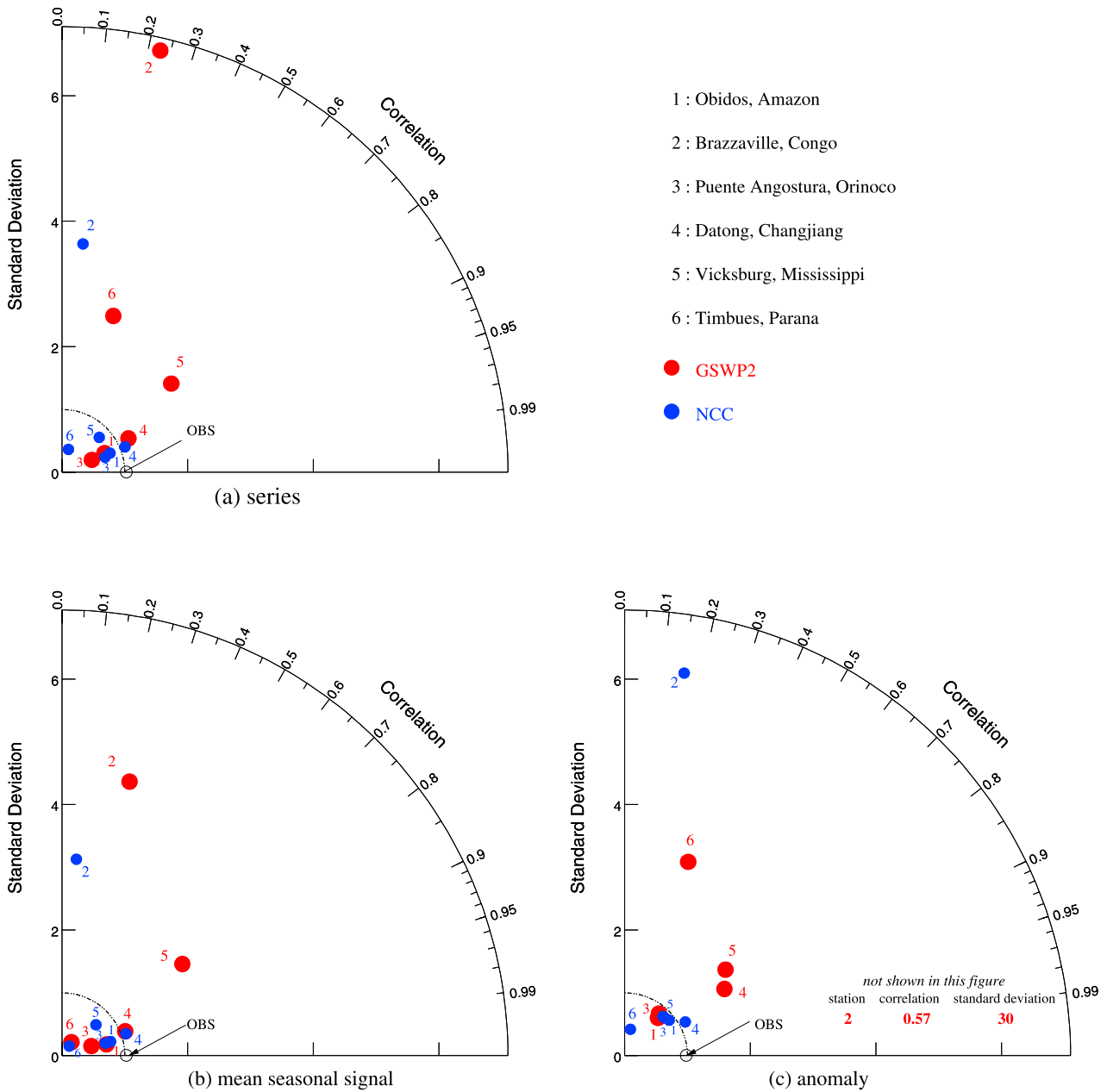
Figure 8. Taylor diagram illustrating the statistics of discharges simulated using GSWP2 and NCC compared to observations for (a) full series, (b) mean seasonal signal, and (c) anomaly.

monthly precipitation. The precipitation intensity is weaker and more continuous than reality so that the simulated evapotranspiration from intercepted water should be larger than reality and result in too low runoff. As a consequence discharges simulated by ORCHIDEE LSM should be lower than observed. Another cause of the underestimation could be due to the observational problems of rain gauges. Gauge measurements that tend to underestimate the true precipitation because of strong wind will reduce the capture ratio of a rain gauge. This effect is especially significant for snow precipitation (see the simulated discharges of Yenisey and Lena).

[43] In the mean seasonal signal (Figure 5), except for the Congo and Paraná, the correlations of simulated and observed discharges vary between 0.8 and 0.99, which are

quite satisfactory. On the other hand, for the interannual variances of simulated discharges the correlation is always lower than 0.8, although the amplitudes are well represented. This result could very well point to limitations in the routing scheme of ORCHIDEE as the three-reservoir approach might be unable to represent properly fluctuations over longer timescales. As the quality of the observed discharge over long time periods is not know, the severity of the model's deficiency cannot be evaluated.

5. Comparison Between NCC and GSWP2

[44] In section 4, we have shown that each step of correction makes the 53-year forcing data better. The NCC forcing allows us to simulate realistically the discharge of

Table 4. Stations and Period Used to Compare the Discharges Simulated by GSWP2 and NCC Experiments to the Observations[a]

	Station	River	Longitude	Latitude	Period
1	Obidos	Amazon	−55.51	−1.94	1986–1997
2	Brazzaville	Congo	15.28	−4.29	1986–1989
3	Puente Angostura	Orinoco	−63.60	8.15	1986–1989
4	Datong	Changjiang	117.62	30.77	1986–1988
5	Vicksburg	Mississippi	−90.91	32.31	1986–1995
6	Timbues	Paraná	−60.71	−32.67	1986–1993

[a]See Figure 8.

the world's largest rivers. This section is a comparison between the new NCC data and the atmospheric forcing data set of GSWP2 over the common period 1986–1995. Figure 8 shows Taylor diagrams which compare the discharges (Figure 8a), the mean seasonal cycle (Figure 8b) and the interannual anomalies (Figure 8c) for 6 of the 10 largest rivers of the world simulated to observations for ORCHIDEE forced by GSWP2 and NCC (see Table 4). Four rivers were left out here because of the lack of discharge observations during 1986–1995 period. Over the Congo basin, the discharge measured at Brazzaville was used because Kinshasa only provides the observations until 1983 (Brazzaville provides the data from 1971 to 1989).

[45] Figure 8a shows that the discharges for Brazzaville (Congo) and Timbues (Paraná) simulated with GSWP2 and NCC are far from the observations. The reasons have been discussed in section 4. The discharges at Obidos (Amazon) and Datong (Changjiang) are slightly better in NCC than in GSWP2. For the Orinoco and Mississippi, discharges simulated with NCC are better than those obtained with the GSWP2 forcing. Figures 8b and 8c confirm the better quality of NCC discharges compared to GSWP2, especially for the interannual variability (Figure 8c) where NCC provides a better simulation.

[46] At the stage of this study, we are unable to understand why the simulated river flows with NCC data seem in better agreement with observations than GSWP2. The fact that the precipitation data are not the same and are used with different time step (3-hourly and 6-hourly for GSWP2 and NCC, respectively) are probably the causes of these differences. This question would deserve another careful study in the near future.

[47] For the period in common, the results of ORCHIDEE forced by GSWP2 and NCC are similar with a small advantage for the longer forcing data set. The fact that for the construction of GSWP2 better observations were available and used gives us some confidence that NCC is reliable forcing data set for the period 1948 to 2000.

6. Conclusions

[48] This study was aimed at constructing a 53-year forcing data set for land surface models. We have used NCEP/NCAR reanalysis and have corrected them with a number of independent in situ observations. The resulting data set, called NCC, has been used as input for the ORCHIDEE LSM and the simulated river discharges were used to validate the simulations.

[49] It was shown that, by far, the correction of precipitation, with the observations collected by the CRU, gives the most important improvement in river discharge. However, the temperature and net radiation corrections also provide improvements which are not negligible. When the simulated river discharges are compared with the ones obtained using the GSWP2 forcing data set, for the same period, the similar discharges produced by ORCHIDEE gives some confidence that the 53-year forcing is reliable. It will be certainly very interesting to confirm these results with other simulations performed using different land surface models.

[50] The NCC data is an important step for understanding of the continental water cycle evolution over the last half century. It will allows us to validate the ability of our land surface models to respond to the interannual variability of the atmospheric forcing. A progressively increasing confidence in the output of our LSMs will enable the community to provide for instance oceanographers with estimates of the freshwater input into the oceans and its variability and trends. The simulated surface processes over the last 53 years will also allow us to refine studies of the impact of climate change on water such as the one proposed by *Milly et al.* [2002].

References

Belward, A., J. Estes, and K. Kline (1999), The IGBP-DIS global 1-km land-cover data set DISCover: A project overview, *Photogram. Eng. Remote Sens.*, 9, 1013–1020.

Bengtsson, L., and J. Shukla (1988), Integration of space and in situ observation to study climate change, *Bull. Am. Meteorol. Soc.*, 69, 1130–1143.

Dai, A., I. Fung, and D. G. Genio (1997), Surface observed global land precipitation variations during 1900–1988, *J. Clim.*, 10, 2943–2962.

de Rosnay, P., and J. Polcher (1998), Modelling root water uptake in a complex land scheme coupled to a gcm, *Hydrol. Earth Syst. Sci.*, 2, 239–255.

de Rosnay, P., J. Polcher, K. Laval, and M. Sabre (2003), Integrated parameterization of irrigation in the land surface model ORCHIDEE, Validation over Indian Peninsula, *Geophys. Res. Lett.*, 30(19), 1986, doi:10.1029/2003GL018024.

Dirmeyer, P. A., A. Dolman, and N. Sato (1999), The pilot phase of the Global Soil Wetness Project, *Bull. Am. Meteorol. Soc.*, 80, 851–878.

Dirmeyer, P. A., X. Gao, and T. Oki (2002), GSWP-2: The second Global Soil Wetness Project science and implementation plan, *IGPO Publ. Ser.*, 65 pp., Int. Global Energy and Water Cycle Exp. (GEWEX) Proj. Off., Silver Spring, Md.

Ducoudré, N. I., K. Laval, and A. Perrier (1993), A new set of parameterizations of the hydrologic exchanges and the land-atmosphere interface within the LMD atmospheric global circulation model, *J. Clim.*, 6, 248–273.

Hall, F. G., B. Meeson, S. Los, L. Steyaert, E. B. de Colstoun, and D. Landis (Eds.) (2003), ISLSCP Initiative II [CD-ROM], NASA, Greenbelt, Md.

Kalnay, E., et al. (1996), The NCEP-NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 77, 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, and S. K. Yang (2002), NCEP/DOE AMIP-II Reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, 83, 1631–1643.

Kistler, et al. (2001), The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, *Bull. Am. Meteorol. Soc.*, 82, 247–267.

Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Oge, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A

dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochem. Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.

Maurer, E., A. Wood, J. Adam, D. Lettenmaier, and B. Nijssen (2002), A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States, *J. Clim.*, *15*, 3237–3251.

Meeson, B. W., F. E. Corprew, J. M. P. McManus, D. M. Myers, J. W. Closs, K. J. Sun, D. J. Sunday, and P. Sellers (1995), ISLSCP Initiative I: Global data sets for land-atmosphere models, 1987–1988 [CD-ROM], NASA, Greenbelt, Md.

Milly, P., R. Wetherald, K. Dunne, and T. Delworth (2002), Increasing risk of great floods in a changing climate, *Nature*, *415*, 514–517.

New, M., M. Hulme, and P. Jones (1999), Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology, *J. Clim.*, *12*, 829–856.

New, M., M. Hulme, and P. Jones (2000), Representing twentieth-century space-time climate variability. Part II: Development of a 1901–90 mean monthly grids of terrestrial surface climate, *J. Clim.*, *13*, 2217–2238.

Oki, T., T. Nishimura, and P. Dirmeyer (1999), Assessment of annual runoff from land surface models using total runoff integrating pathways (TRIP), *J. Meteorol. Soc. Jpn.*, *77*, 235–255.

Piper, S. C., and E. F. Stewart (1996), A gridded global data set of daily temperature and precipitation for terrestrial biosphere modeling, *Global Biogeochem. Cycles*, *10*, 757–782.

Shepard, D. (1968), A two-dimensional interpolation function for irregularly-spaced data, *Proc. Natl. Conf. ACM, 23rd*, 517–524.

Sitch, S. (2000), The role of vegetation dynamics in the control of atmospheric CO_2 content, Ph.D. thesis, 213 pp., Univ. of Lund, Lund, Sweden.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*, 7183–7192.

Verant, S., K. Laval, J. Polcher, and M. Castro (2004), Sensitivity of the continental hydrological cycle to the spatial resolution over the Iberian Peninsula, *J. Hydrometeorol.*, *5*, 265–283.

Viovy, N. (1996), Interannuality and CO_2 sensitivity of the SECHIBA-BG coupled SVAT-BGC Model, *Phys. Chem. Earth*, *21*, 489–497.

Vörösmarty, C., B. Fekete, B. Meybeck, and R. Lammers (2000), Global system of rivers: Its role in organizing continental land mass and defining land-to-ocean linkages, *Global Biogeochem. Cycles*, *14*, 599–621.

Xie, P., and P. A. Arkin (1996), Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, *J. Clim.*, *9*, 840–858.

Xie, P., B. Rudolf, U. Schneider, and P. A. Arkin (1996), Gauge-based monthly analysis of global land precipitation from 1971 to 1994, *J. Geophys. Res.*, *101*, 19,023–19,034.

Zhao, M., and P. A. Dirmeyer (2003), Production and Analysis of GSWP-2 near-surface meteorology data sets, *COLA Tech. Rep. 159*, 38 pp., Cent. for Ocean-Land-Atmos. Stud., Calverton, Md.

―――――――――――

K. Laval, T. Ngo-Duc, and J. Polcher, LMD/CNRS, BP99, 4 place Jussieu, F-75252 Paris Cedex 5, France. (katia.laval@lmd.jussieu.fr; thanh.ngo-duc@lmd.jussieu.fr; jan.polcher@lmd.jussieu.fr)