

UCRL-TR-217474



LAWRENCE  
LIVERMORE  
NATIONAL  
LABORATORY

# Large-scale Validation of AMIP II Land-surface Simulations: Preliminary Results for Ten Models

T. J. Phillips, A. Henderson-Sellers, P. Irannejad,  
K. McGuffie, H. Zhang

December 1, 2005

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

## Large-scale Validation of AMIP II Land-surface Simulations: Preliminary Results for Ten Models

T.J. Phillips<sup>1</sup>, A. Henderson-Sellers<sup>2</sup>, P. Irannejad<sup>2</sup>, K. McGuffie<sup>3</sup>, H. Zhang<sup>4</sup>

<sup>1</sup> Program for Climate Model Diagnosis & Intercomparison, Livermore, California, USA

<sup>2</sup> Environment, Australian Nuclear Science and Technology Organisation, Menai NSW, Australia

<sup>3</sup> University of Technology, Sydney, Australia

<sup>4</sup> Bureau of Meteorology, Melbourne, Australia

February, 2002

### Abstract

This report summarizes initial findings of a large-scale validation of the land-surface simulations of ten atmospheric general circulation models that are entries in phase II of the Atmospheric Model Intercomparison Project (AMIP II). This validation is conducted by AMIP Diagnostic Subproject 12 on Land-surface Processes and Parameterizations, which is focussing on putative relationships between the continental climate simulations and the associated models' land-surface schemes. The selected models typify the diversity of representations of land-surface climate that are currently implemented by the global modeling community.

The current dearth of global-scale terrestrial observations makes exacting validation of AMIP II continental simulations impractical. Thus, selected land-surface processes of the models are compared with several alternative validation data sets, which include merged in-situ/satellite products, climate reanalyses, and off-line simulations of land-surface schemes that are driven by observed forcings. The aggregated spatio-temporal differences between each simulated process and a chosen reference data set then are quantified by means of root-mean-square error statistics; the differences among alternative validation data sets are similarly quantified as an estimate of the current observational uncertainty in the selected land-surface process. Examples of these metrics are displayed for land-surface air temperature, precipitation, and the latent and sensible heat fluxes.

It is found that the simulations of surface air temperature, when aggregated over all land and seasons, agree most closely with the chosen reference data, while the simulations of precipitation agree least. In the latter case, there also is considerable inter-model scatter in the error statistics, with the reanalyses' estimates of precipitation resembling the AMIP II simulations more than to the chosen reference data. In aggregate, the simulations of land-surface latent and sensible heat fluxes appear to occupy intermediate positions between these extremes, but the existing large observational uncertainties in these processes make this a provisional assessment. In all selected processes as well, the error statistics are found to be sensitive to season and latitude sector, confirming the need for finer-scale analyses which also are in progress.

## 1. Introduction

The Atmospheric Model Intercomparison Project ([AMIP](#)) has promoted a widely adopted experimental protocol in which ocean and sea ice boundary conditions as well as certain radiative forcings are prescribed in common, while the state of the land surface is determined by model-specific schemes. Hence, there is especial interest in analyzing differences among the AMIP continental simulations that may be attributable to variations in the detailed parameterizations of the associated land-surface schemes (LSSs). This is the focus of AMIP Diagnostic Subproject 12 (DSP 12) on Land-Surface Processes and Parameterizations, an initiative of the Project for Intercomparison of Land-surface Processes and Parameterizations ([PILPS](#)).

DSP 12's work in the first phase of the AMIP (aka AMIP I) is summarized by [Henderson-Sellers \(1999\)](#). We now are implementing our diagnostic strategy ([Phillips et al. 2001](#)) for the current-generation AMIP II models whose output data are being archived by the Program for Climate Model Diagnosis and Intercomparison ([PCMDI](#)). In addition to analyzing the coupled land-atmosphere interactions at regional scales (e.g. [Irannejad et al. 2000](#) and [Irannejad et al. 2001](#)), we are validating the land-surface

simulations at large (continental-to-global) scales.

This report describes preliminary results of such a large-scale validation of a subset of the participating AMIP II models ([Table 1](#)) whose output data were first made available by PCMDI. These ten models typify the variety of representations of atmosphere and land surface that are in current use (e.g. only models **ugamp-98a** and **ukmo-98a** share many common characteristics). When considering the selected models' LSSs, for example, it is seen that **ccsr-98a** adopts a "bucket" representation of soil hydrology that only implicitly incorporates the effects of vegetation canopy and roots (but modified from the classic bucket scheme of [Manabe et al. 1965](#) insofar as a minimum stomatal resistance for land-surface evaporation is prescribed). In contrast, the [other models](#) include diverse soil-vegetation-atmosphere transfer schemes (SVATs) that more explicitly represent the details of continental processes, albeit at varying levels of complexity. This widespread usage of SVATs in the selected AMIP II models contrasts with usual practice in the earlier-generation AMIP I models, which mostly included variations of the bucket scheme ([Phillips 2001](#)).

Table 1: Selected AMIP II modeling groups (partial listing of participants), information on the horizontal/ vertical resolution of the corresponding AGCMs, and the name, type, and relevant reference for each embedded land-surface schemes (LSS).

| Modeling Group (Location)   | Model Version (Abbreviation) | Horiz./Vert. Resolution | LSS Name, Type, Reference                              |
|---|------------------------------|-------------------------|--|
| Center for Climate System Research (Tokyo, Japan)                     | <b>ccsr-98a (ccsr)</b>       | spectral T42/L18        | CCSR modif. bucket, <a href="#">Manabe et al. 1965</a> |
| Centre National de Recherches Meteorologiques (Toulouse, France)      | <b>cnrm-00a (cnrm)</b>       | spectral T63/L45        | ISBA SVAT, <a href="#">Mahfouf et al. 1995</a>         |
| Department of Numerical Mathematics (Moscow, Russia)                  | <b>dnm-98a (dnm)</b>         | 4x5 degs/L21            | DNM SVAT, <a href="#">Volodin&amp;Lykossov 1998</a>    |
| European Centre for Medium- range Weather Forecasts (Reading, UK)     | <b>ecmwf-98a (ecm)</b>       | spectral T63/L50        | ECMWF SVAT, <a href="#">Viterbo&amp;Beljaars 1995</a>  |
| Japan Meteorological Agency (Tokyo, Japan)                            | <b>jma-98a (jma)</b>         | spectral T63/L30        | modified SiB SVAT, <a href="#">Sato et al. 1989</a>    |
| National Center for Atmospheric Research (Boulder, Colorado, USA)     | <b>ncar-98a (ncar)</b>       | spectral T4 /L18        | LSM SVAT, <a href="#">Bonan 1996</a>                   |
| National Centers for Environmental Prediction (Washington, D.C., USA) | <b>ncep-99a* (ncep)</b>      | spectral T62/L28        | NCEP SVAT, <a href="#">Pan&amp;Mahrt 1987</a>          |
| Pacific Northwest National Laboratory (Richland, Washington, USA)     | <b>pnnl-97a (pnnl)</b>       | spectral T42/L18        | BATS1e SVAT, <a href="#">Dickinson et al. 1993</a>     |
| UK Universities' Global Atmospheric Modelling Programme (Reading, UK) | <b>ugamp-98a (ugam)</b>      | 2.5x3.75 degs/ L58      | MOSES SVAT, <a href="#">Cox et al. 1999</a>            |
| United Kingdom Meteorological Office (Bracknell, UK)                  | <b>ukmo-98a (ukmo)</b>       | 2.5x3.75 degs/ L19      | MOSES SVAT, <a href="#">Cox et al. 1999</a>            |

\* now designated **ncep-99b**

In the remainder of this report, we discuss our validation approach ([Section 2](#)) and choice of reference data ([Section 3](#)) for evaluating the ten AMIP II models with respect to four land-surface processes: surface air temperature, precipitation, and the latent

and sensible heat fluxes. In [Sections 4](#) and [5](#), we explain our validation metrics and show a number of examples. Finally, in [Section 6](#) we summarize the main findings of the large-scale land-surface validation thus far.

## 2. Validation Approach

We acknowledge that the present dearth of global terrestrial data sets renders large-scale validation of the AMIP II continental simulations a problematical endeavor. Nonetheless, these data deficiencies are gradually being ameliorated as initiatives such as the Global Precipitation Climatology Project ([GPCP](#)) and the International Satellite Land-Surface Climatology Project ([ISLSCP](#)) construct consistent global data sets by merging in-situ observations with satellite-derived products. However, such merged global data sets are not yet available for many land-surface processes of interest (e.g. the latent and sensible heat fluxes). In such cases, the only available global validation references are model-derived products such as climate reanalyses or off-line simulations of particular LSSs that are driven by estimated observational forcings.

We therefore want to compare the AMIP II land-surface simulations against a variety of available validation data sets (cited in [Section 3](#)), since current observational uncertainties are large, and likely to remain so for some time. That is, we wish to assess whether or not the "average simulation" (i.e. the centroid of the ensemble of AMIP simulations) lies within the envelope of current observational estimates of a land-surface process. Thus, we seek to quantify the goodness-of-fit of the average AMIP land-surface simulation with the chosen reference data sets by means of appropriate statistical metrics (see [Section 4](#)). The amount of inter-model scatter about this average simulation also is pertinent, since it implies how sensitive the simulations of the continental process are to model representations of atmospheric forcings and/or land-surface responses.

In such coupled land-atmosphere simulations, of course, we cannot cleanly separate "forcings" from "responses", and so we are not able to validate the LSSs *per se*. Nonetheless, there is reason to expect that the "signature" of a land-surface scheme should be detectable in such coupled simulations (e.g. [Gedney et al. 2000](#)). In the longer run, therefore, we want to evaluate the performance of an individual AMIP II model *insofar as this may be attributable to the complexities of its LSS parameterizations*. Hence, our ultimate goal is to formulate hypotheses on the impacts on model performance of including LSS parameterizations of different complexities in AGCMs. These hypotheses then could be tested in future sensitivity experiments where the continental state is subject to more control than is practical in the AMIP intercomparison.

## 3. Selected Processes and Reference Data

For this preliminary model validation, we focus on four land-surface processes: precipitation, surface (two-meter) air temperature, and the terrestrial latent and sensible heat fluxes. The first two of these are atmospheric prognostics that influence the land-surface moisture and energy budgets. Likewise, the turbulent fluxes exemplify key land-atmosphere interactions that are presumed to be especially sensitive to LSS parameterizations. For each of these processes, we choose a validation reference, but also consider several alternative validation data sets for comparison in Table 2.

Table 2: For selected land-surface processes, the chosen validation reference is listed, as well as alternative validation data sets. These monthly mean data all overlap for the period 1979-1993.

| Land-surface Process (AMIP acronym) | Reference Data Set   | Alternative Validation Data   |
|-------------------------------------|--|---|
| Sfc. air temperature ( <b>tas</b> ) | <a href="#">Jones et al. 1999</a>                                | <a href="#">ERA15</a> , <a href="#">NCEP/NCAR</a> , <a href="#">NCEP/DOE</a>  |
| Precipitation rate ( <b>pr</b> )    | Global Precipitation Climatology Project( <a href="#">GPCP</a> ) | Climate Prediction Center Merged Analysis of Precipitation ( <a href="#">CMAP</a> )<br><a href="#">VIC</a> precipitation forcing,<br><a href="#">ERA15</a> , <a href="#">NCEP/NCAR</a> , <a href="#">NCEP/DOE</a> |

|                                    |   |  |
|------------------------------------|---|--|
| Latent heat flux ( <b>hfls</b> )   | <a href="#">VIC</a> off-line simulation | <a href="#">ERA15</a> , <a href="#">NCEP/NCAR</a> , <a href="#">NCEP/DOE</a> |
| Sensible heat flux ( <b>hfss</b> ) | <a href="#">ERA15</a> reanalysis        | <a href="#">NCEP/NCAR</a> , <a href="#">NCEP/DOE</a>                         |

The choices of validation reference data reflect our preference for in-situ terrestrial observations (e.g. [Jones et al. 1999](#)) or merged in situ/satellite observations (e.g. [GPCP](#) precipitation). Because comparable merged data sets are presently unavailable for the turbulent fluxes, we have resorted to using model-derived products as validation references. For example, we have selected a 1979-1993 global simulation by the Variable Infiltration Capacity (VIC) scheme ([Nijssen et al. 2001](#)) as reference data for land-surface latent heat flux. Observed precipitation station data (interpolated to a 2x2-degree grid) were used for off-line forcing of the VIC scheme, and its routed climatological runoff was calibrated to be in rough agreement with continental-scale observations. Thus, there is reason to prefer this estimate of land-surface latent heat flux over those of the reanalyses, which rely only on simulated precipitation forcings and reflect other biases, for example those introduced by soil moisture nudging.

No alternatives to the reanalyses' estimates of land-surface sensible heat flux were available, however. The [VIC](#) off-line simulation, for example, computed the latent heat flux from the Penman-Monteith equation, and did not provide a time series of sensible heat flux. Hence, we arbitrarily chose the [ERA15](#) sensible heat flux as the validation reference in this case. (In future validation work, however, we plan to compare the AMIP II simulations against other estimates of global turbulent fluxes that are becoming more available. The latter include, for example, a 10+ year data set to be produced by the [ISLSCP](#) satellite initiative mentioned by [Hall et al. 2001](#), as well as other synthetic global data sets generated from selected land-surface schemes that are driven off-line by observational forcings, such as those described by [Dirmeyer et al. 1999](#), [Dirmeyer and Tan 2001](#), and [Houser et al. 2001](#).)

The reanalyses also supply many of the alternative validation data sets of [Table 2](#) for estimating current observational uncertainties. In the case of precipitation, though, other alternatives besides the reanalyses are available. These include the VIC precipitation forcings mentioned above and the Climate Prediction Center Merged Analysis of Precipitation ([CMAP](#)) data set which, like the [GPCP](#), consists of merged gauge and satellite estimates (see [Gruber et al. 2000](#) for a detailed comparison).

The data sets of [Table 2](#) consist of time series of monthly mean data that overlap for the years 1979-1993, a substantial portion of the AMIP II simulation period (1979-1995). In all cases, evaluations of the selected AMIP II models' simulations of these land-surface processes were limited to this 1979-1993 period and were interpolated to the same grids as the validation reference. For consistency, the same temporal/spatial constraints also were imposed on the alternative validation data sets, since we wish to use their goodness-of-fit to the reference data as estimates of current observational uncertainties in the selected land-surface processes.

#### 4. Validation Metrics and Displays: Examples

In this section, we elaborate on our chosen validation metrics and associated displays, and illustrate how these are applied to the land-surface temperature fields simulated by the AMIP II models of [Table 1](#).

##### 4.1 RMS Error and Components

Large-scale model validation involves, in essence, an estimation of the average spatio-temporal "distance" between a given simulation of a spatio-temporal surface process  $P_s(i,j,k)$  and that of the reference surface process  $P_r(i,j,k)$ . This is often expressed as a root-mean-square statistic

$$E = \{ \sum_w [P_s(i,j,k) - P_r(i,j,k)]^2 \}^{1/2}$$

Here  $(i,j,k)$  denotes, respectively, the longitude, latitude, and time coordinates of each data point, and  $\sum_w$  signifies a summation

over the product  $N = QT$  of the total number of space samples  $Q$  and time samples  $T$ , with suitable weighting by fractional area and time slice, for example

$$w \sim \cos(j) / \{N \cdot \sum_j [\cos(j)]\}$$

for data points that are uniformly distributed in time. We will commonly refer to  $E$  as the *total RMS error* in a model simulation of land-surface process  $P$ ; however, given the large uncertainties in the currently available validation data ([Table 2](#)),  $E$  perhaps instead should be regarded as simply the *total RMS difference* between the model simulation and the reference data.

For model validation, it is also useful to distinguish the components of  $E$  associated with differences in the spatio-temporal means ("bias error") versus spatio-temporal structures ("pattern error"). From this perspective, the simulated and reference processes can be decomposed into their respective spatio-temporal means and associated departures:

$$P_s(i,j,k) = \langle P_s \rangle + P'_s(i,j,k)$$

and

$$P_r(i,j,k) = \langle P_r \rangle + P'_r(i,j,k)$$

Then the bias error is defined as

$$\langle E \rangle = \langle P_s \rangle - \langle P_r \rangle$$

and the pattern error as

$$E' = \{ \sum_w [P'_s(i,j,k) - P'_r(i,j,k)]^2 \}^{1/2}$$

It follows that the total mean-square error  $E^2$  is expressed by the quadratic sum of these bias and pattern errors:

$$E^2 = \langle E^2 \rangle + E'^2$$

To allow consistent comparison of RMS errors across many models and/or processes that exhibit different natural variabilities, it is also advantageous to derive a normalized RMS error  $E_n$  by dividing the mean-square error by the spatio-temporal variance  $\sigma_r^2$  of the reference process:

$$E_n = [E^2 / \sigma_r^2]^{1/2}$$

where

$$\sigma_r^2 = \sum_w [P'_r(i,j,k)]^2$$

This, in turn, defines *normalized* bias and pattern errors  $\langle E_n \rangle$  and  $E'_n$ :

$$E_n = \{ \langle E \rangle^2 / \sigma_r^2 + E'^2 / \sigma_r^2 \}^{1/2} = \{ \langle E_n \rangle^2 + E'_n{}^2 \}^{1/2}$$

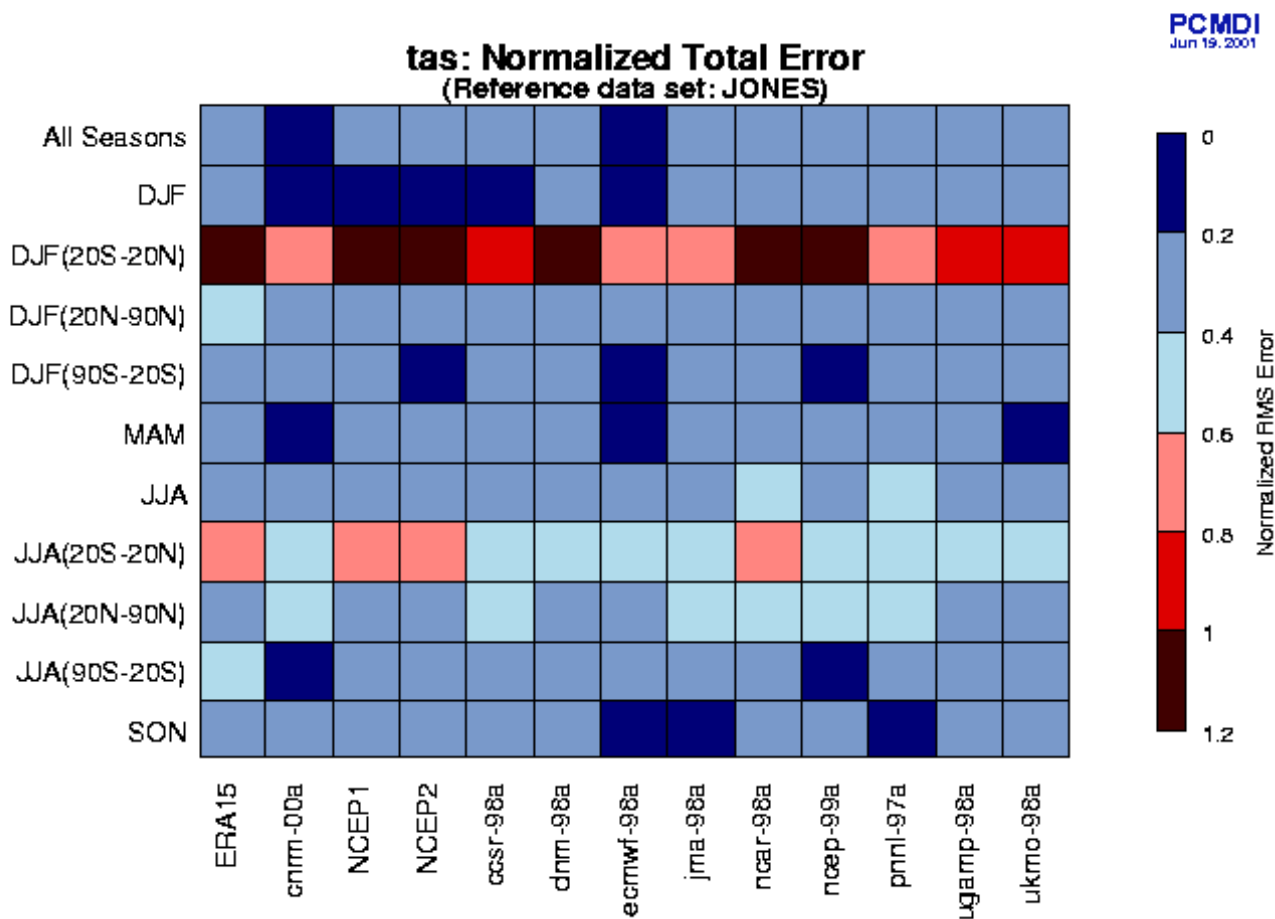
#### 4.2 Portrait Diagrams

Our model evaluation tasks include not only the calculation of appropriate large-scale validation metrics, but also their concise depiction for the many simulations and processes of interest. Both these needs are addressed by software developed by [PCMDI](#)

[staff members](#), which DSP 12 has previously applied in a more limited way to the AMIP I continental simulations ([Phillips et al. 2000](#)).

For example, a simulation's performance relative to chosen validation reference data can be conveyed by color-coded illustrations of the total normalized RMS error  $E_n$ . Consider [Figure 1](#), a "portrait diagram" of total RMS error in two-meter land-surface air temperature for the selected AMIP II simulations and for the [ERA15](#), NCEP1 (shorthand for [NCEP/NCAR](#)), and NCEP2 (shorthand for [NCEP/DOE](#)) reanalyses, where the reference is the [Jones et al. \(1999\)](#) data set. The total RMS errors are generally small (as indicated by the preponderance of bluish squares) when computed over all land surfaces and seasons (i.e. the top row of this portrait diagram), but are considerably larger (as indicated by the reddish squares) when computed over only tropical sectors in DJF, and to a lesser extent in JJA.

Figure 1: Portrait of total normalized RMS error  $E_n$  in two-meter land-surface air temperature **tas** for selected AMIP II models and for [ERA15](#), [NCEP/NCAR](#) (denoted 'NCEP1'), and [NCEP/DOE](#) (denoted 'NCEP2') reanalyses relative to [Jones et al. 1999](#) reference data for the period 1979-1993. (Note, the display software has placed results for the **cnrm-00a** model in the second column, between those for the ERA15 and NCEP1 reanalyses, and apart from those of the other nine models.) The magnitudes of the  $E_n$  are indicated by color coding, with bluish squares indicating relatively small RMS errors and reddish squares relatively large ones. Results are shown for aggregation over all land surfaces and seasons (first row of the portrait diagram), when disaggregated by season (December- January- February: DJF, March - April - May: MAM, June - July - August: JJA, and September - October - November: SON), and when, for the DJF and JJA seasons, disaggregated spatially in tropical (20 S to 20 N) and extratropical (20 N to 90 N and 20 S to 90 S) sectors.





gn r r α dr ecm jr nr nc pr ugar ukr

The bias and pattern components of this normalized total error are similarly illustrated by the portrait diagrams of [Figure 2](#) and [Figure 3](#), respectively. These error components are relatively small when they are computed over all land surfaces and seasons (top row of each portrait diagram), but are considerably larger for particular seasons and/or sectors, such as for tropical latitudes in DJF (as is the case also for the total normalized RMS error in [Figure 1](#)). It is also apparent, however, that the generally large total errors in the DJF tropical sector are more commonly due to pattern errors ([Figure 3](#)) than to bias errors ([Figure 2](#)), although the latter are also substantial in a number of the simulations and reanalyses.

Figure 2: As in [Figure 1](#), except for the bias component  $\langle E_n \rangle$  of the total normalized RMS error  $E_n$ .

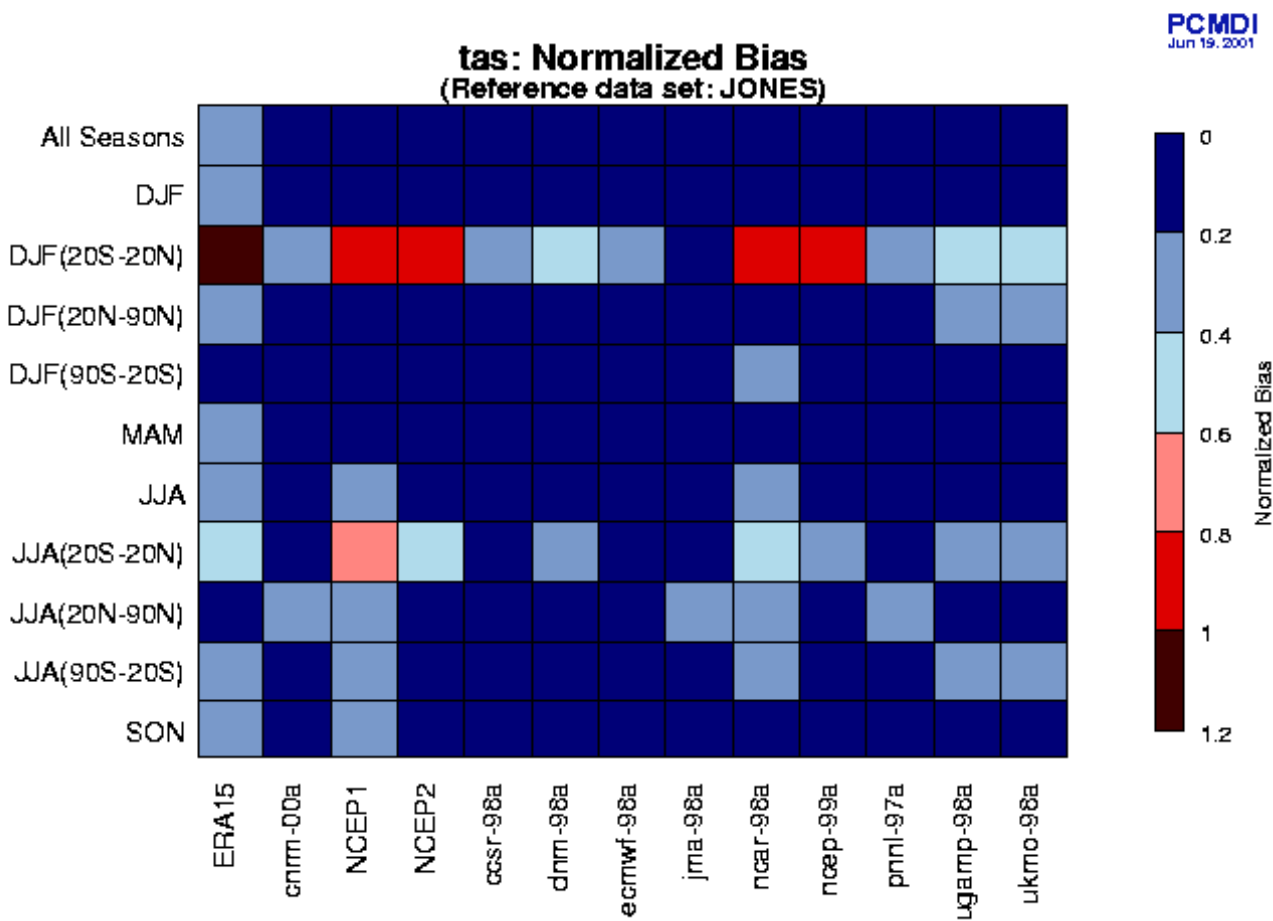
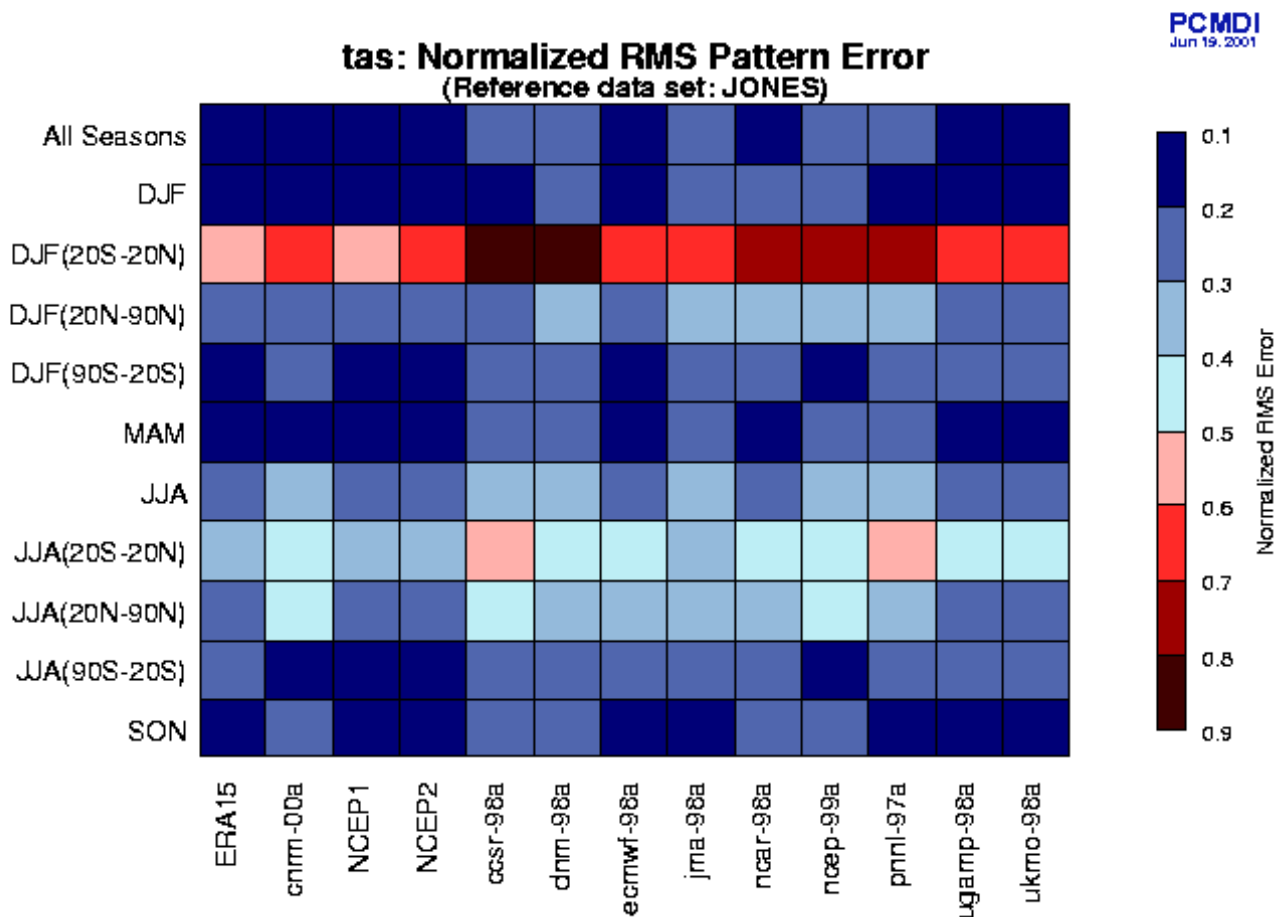


Figure 3: As in [Figure 1](#), except for the pattern component  $E_n'$  of the total normalized RMS error  $E_n$ .



### 4.3 Taylor Diagrams

Another useful validation metric is the spatio-temporal correlation  $R$  between simulated and reference data, defined as

$$R = \frac{\sum_w [P_s' P_r']}{[\sigma_s \sigma_r]}$$

where  $\sigma_s$  and  $\sigma_r$  are spatio-temporal standard deviations of the simulated and reference data given by

$$\sigma_s = \{ \sum_w [P_s'(i,j,k)]^2 \}^{1/2}$$

and

$$\sigma_r = \{\sum_w [P_r'(i,j,k)]^2\}^{1/2}$$

[Taylor \(2001\)](#) has shown that the mean-square pattern error  $E'^2$  is related to the quantities  $R$ ,  $\sigma_s$ , and  $\sigma_r$  by

$$E'^2 = \sigma_r^2 + \sigma_s^2 - 2\sigma_r\sigma_s R$$

This corresponds to a law-of-cosines relationship

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma)$$

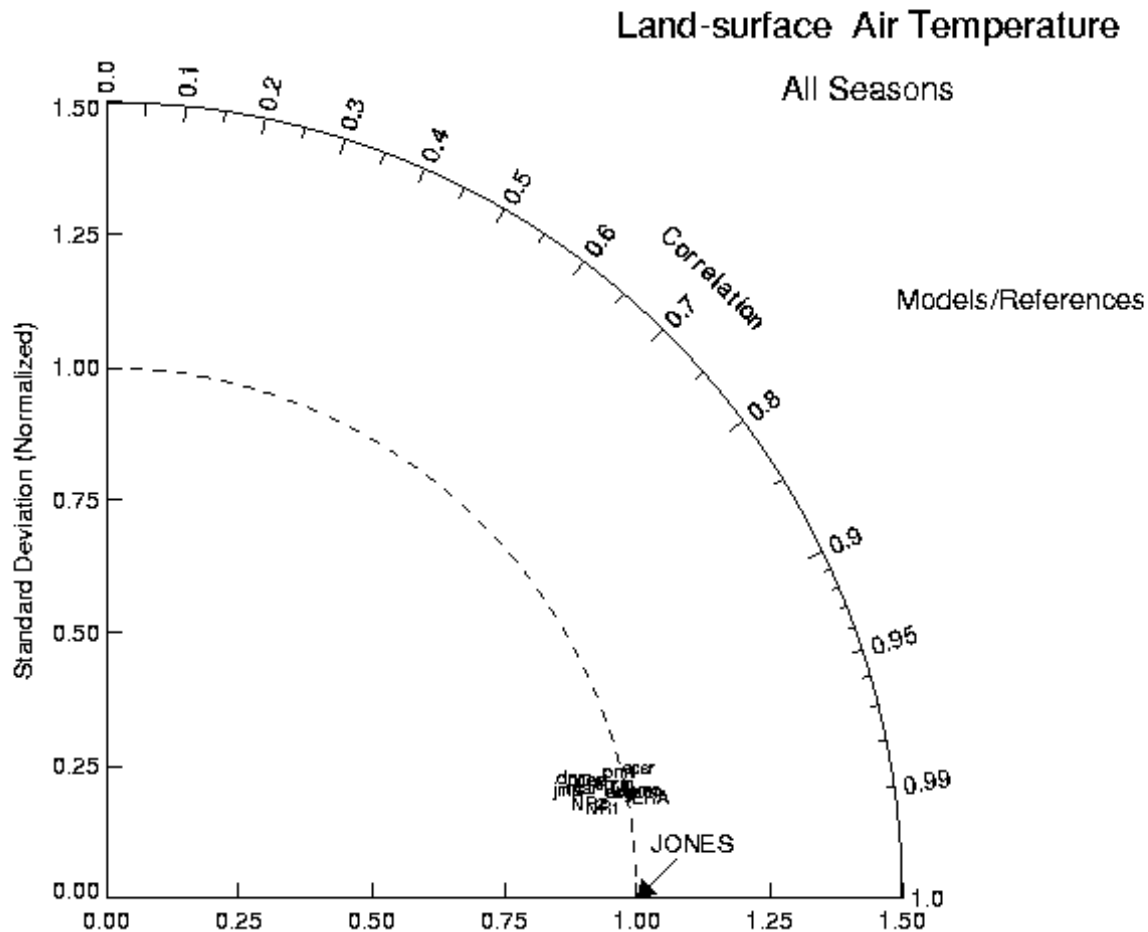
where  $a = \sigma_r$  and  $b = \sigma_s$  are the sides of a triangle that include angle  $\gamma = \cos^{-1}(R)$ , while  $c = E'$  is the length of the side opposite angle  $\gamma$ . This relationship also can be expressed in terms of the normalized mean-square pattern error  $E_n'^2 = E'^2/\sigma_r^2$ :

$$E_n'^2 = 1 + (\sigma_s^2/\sigma_r^2) - 2(\sigma_s/\sigma_r)R$$

Taylor depicts the corresponding geometry in a polar plot that decomposes the normalized RMS pattern error  $E_n'$  into amplitude and phase errors. An example of such a Taylor diagram is shown in [Figure 4](#) for the land-surface air temperature simulated by each AMIP II model or estimated by each reanalysis, relative to the [Jones et al. \(1999\)](#) reference data (labeled 'JONES'). Here, the respective  $E_n'$  statistics are aggregated over all land areas and seasons. The magnitude of  $E_n'$  for a given simulation/reanalysis is conveyed by its *linear displacement* from the 'JONES' reference point in [Figure 4](#). That part of  $E_n'$  associated with amplitude error, expressed as  $(\sigma_s - \sigma_r)/\sigma_r$ , is indicated in this polar plot by the *radial displacement* from the dashed-line quarter circle. That part of  $E_n'$  associated with phase error is indicated by the *azimuthal displacement* from the 'JONES' reference point, which is scaled proportional to the cosine of the spatio-temporal correlation  $R$ . (The equivalence of this azimuthal displacement to phase error is only strictly true, however, when both the model and reference data exhibit a single simple harmonic pattern, but with relative offsets in phase.)

From [Figure 4](#), it is apparent there are scant differences among the globally and seasonally aggregated values of  $E_n'$  for the simulations or reanalyses: they are all highly correlated ( $R \sim 0.98$ ) and their variability amplitudes also closely approximate that of the reference data (i.e. they all fall near the dashed-line quarter circle that passes through the 'JONES' reference point).

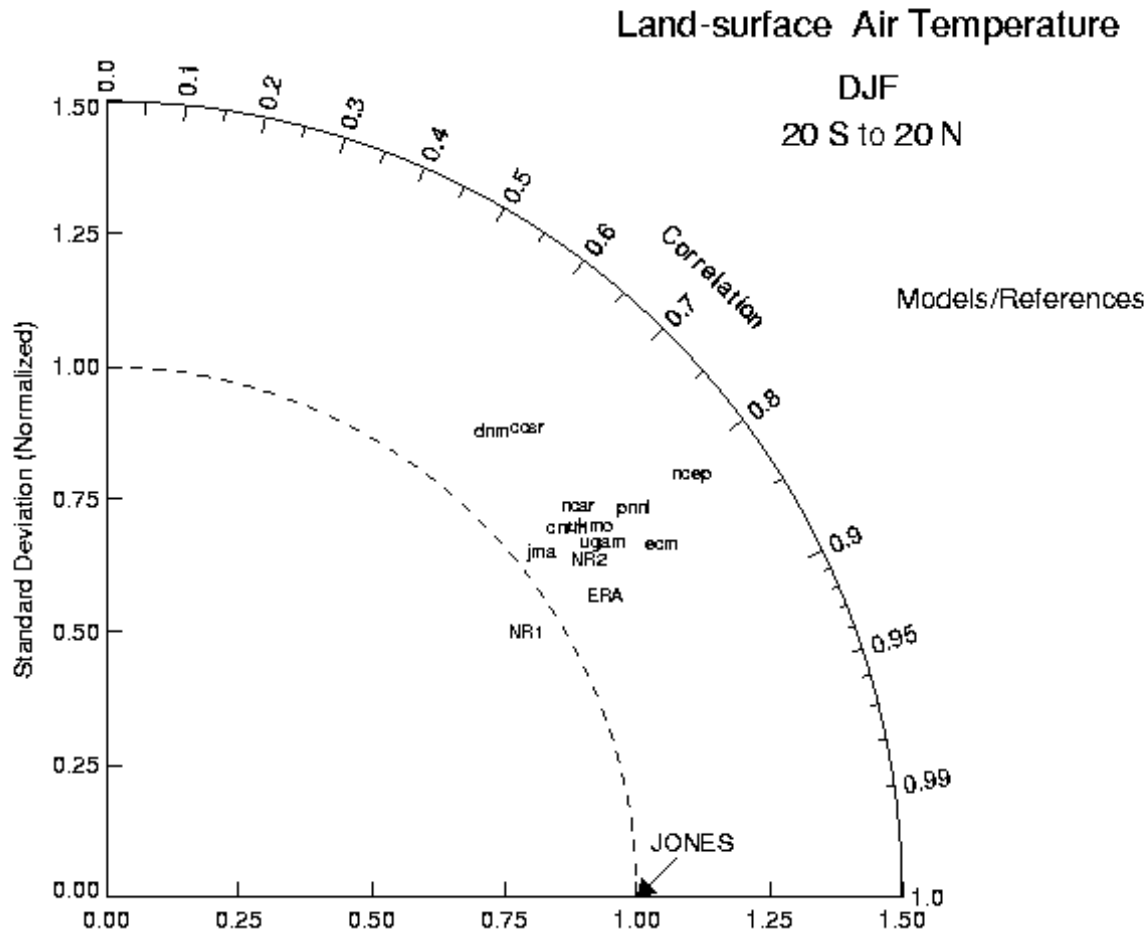
Figure 4: Taylor diagram of normalized RMS pattern errors  $E_n'$  in two-meter land-surface air temperature for selected AMIP II models (labeled according to the model version acronyms of Table 1) and for [ERA15](#), [NCEP/NCAR](#), and [NCEP/DOE](#) reanalyses relative to [Jones et al. 1999](#) reference data for the period 1979-1993. (The errors are aggregated over all land surfaces and seasons.) For each simulation/reanalysis, the magnitude of  $E_n'$  is indicated by the linear displacement from the 'JONES' reference point, and  $E_n'$  is decomposed as well into amplitude and phase errors. That part of  $E_n'$  due to amplitude error is indicated by the radial displacement from the dashed-line quarter circle with radius 1, where points interior to the quarter circle denote amplitudes of spatio-temporal variability smaller than that of the reference data, while exterior points denote larger amplitudes. That part of  $E_n'$  due to pattern error is indicated by the azimuthal displacement from the 'JONES' reference point, which is scaled proportional to the cosine of the spatio-temporal correlation  $R$  of each simulation/reanalysis with the reference data.



The structure of spatio-temporal subcomponents of the normalized RMS pattern error  $E_n'$  of surface air temperature can be displayed in similar fashion (Figure 5), e.g. for the problematical DJF tropical case in the portrait diagram of Figure 3. Compared with Figure 4, there are considerably larger normalized pattern errors in the model simulations and reanalyses (the latter being labeled as 'ERA' for ERA15, 'NR1' for NCEP/NCAR, and 'NR2' for NCEP/DOE reanalyses), as manifested by their greater distances from the 'JONES' reference point. These larger pattern errors, in turn, are associated with substantially larger amplitude and phase errors than are seen in Figure 4.

In Figure 5, the reanalyses show marginally lower pattern errors than do the model simulations. However, except for the NCEP/NCAR (NR1) reanalysis, the amplitudes of spatio-temporal variability all exceed the reference value (i.e. they lie outside the dashed-line quarter circle). The scatter in magnitudes of pattern errors among the simulations/reanalyses thus are due mainly to differences in their respective phase errors.

Figure 5: As in Figure 4, except displaying the structure of the normalized pattern error  $E_n'$  computed only over the DJF season and only in the tropical latitude sector 20 S to 20 N.

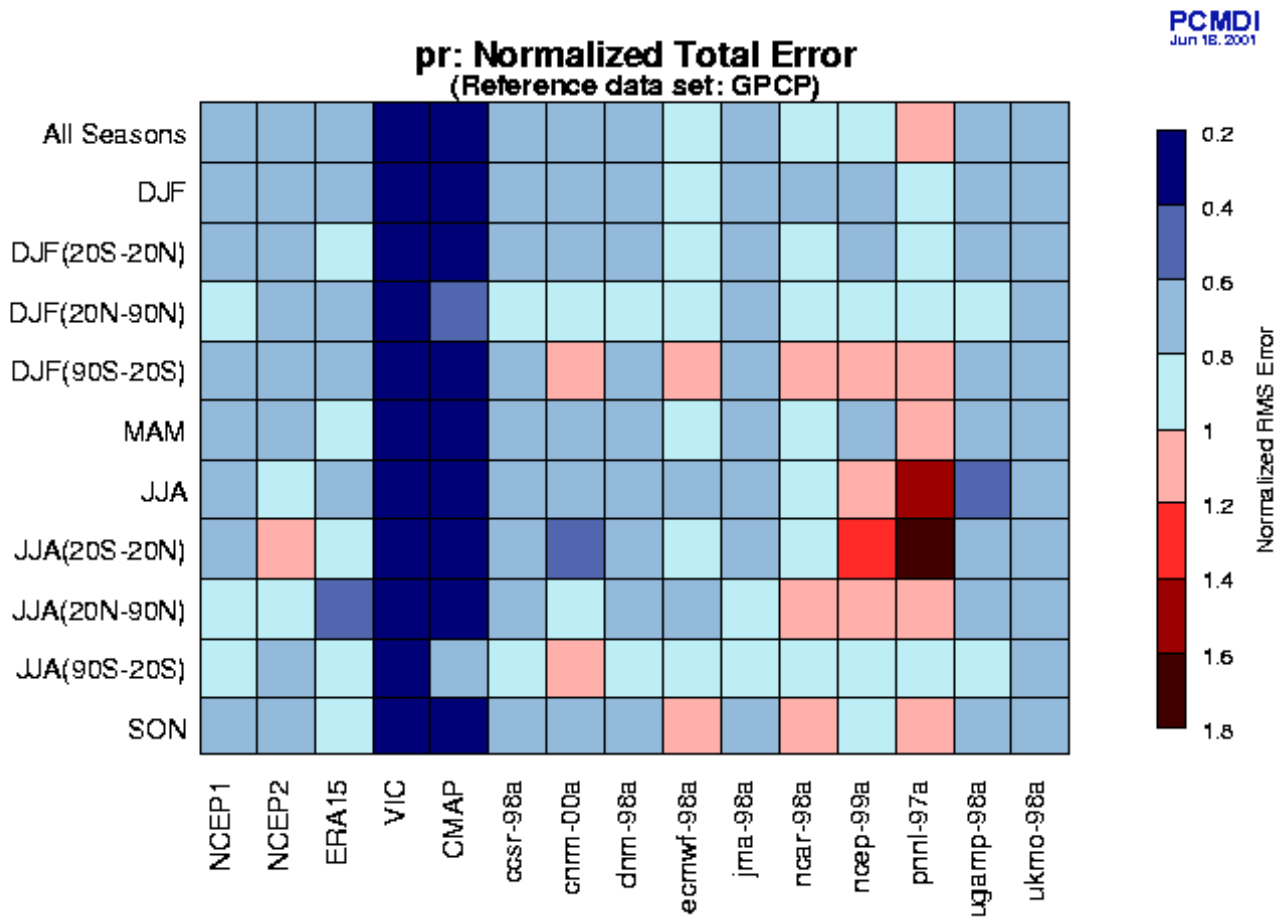


## 5. Additional Validation Results

### 5.1 Land-surface Precipitation

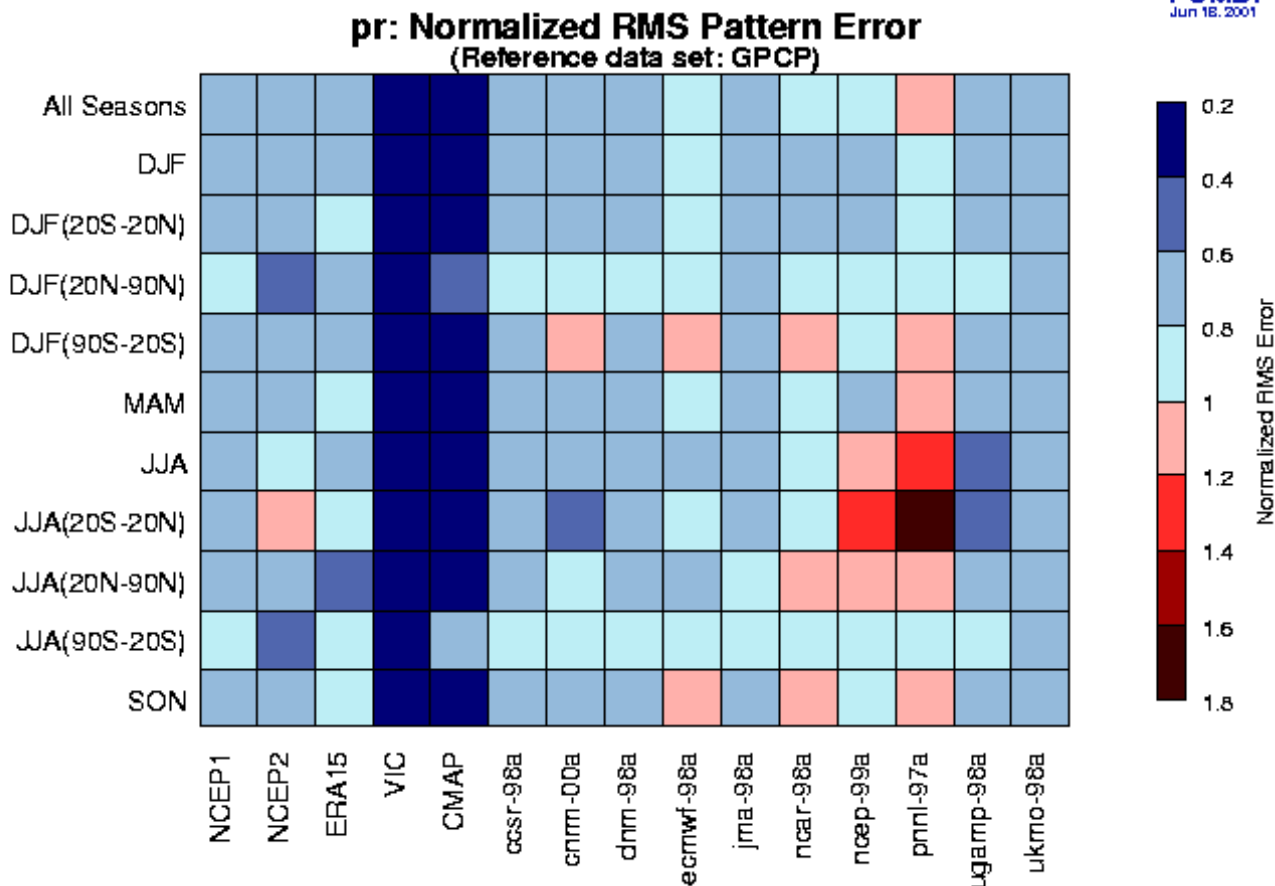
The normalized total RMS errors in land-surface precipitation relative to the [GPCP reference data](#) are illustrated by the portrait diagram of [Figure 6](#). The globally aggregated spatio-temporal variability of the CMAP and VIC alternative precipitation data sets are seen to agree closely with the reference, as manifested by low normalized total RMS error  $E_n$  when aggregated over all seasons/latitude sectors. The total errors in the reanalyses' precipitation are substantially larger, and are greater still in the model simulations, especially in selected seasons/latitudes (e.g. in tropical and Northern extratropical latitudes in JJA). Considerable inter-model differences in performance are also evident, however.

Figure 6: As in [Figure 1](#), except for land-surface precipitation **pr** relative to the [GPCP reference data](#).



Overall, these normalized total errors in precipitation are due mostly to pattern errors, which are exhibited to varying degrees by individual simulations/reanalyses, and depend as well on season and latitude sector ([Figure 7](#)). For example, the pattern errors in some model simulations are greatest in JJA, especially in the Northern Hemisphere.

Figure 7: As in [Figure 3](#), except for land-surface precipitation **pr** relative to the [GPCP reference data](#).

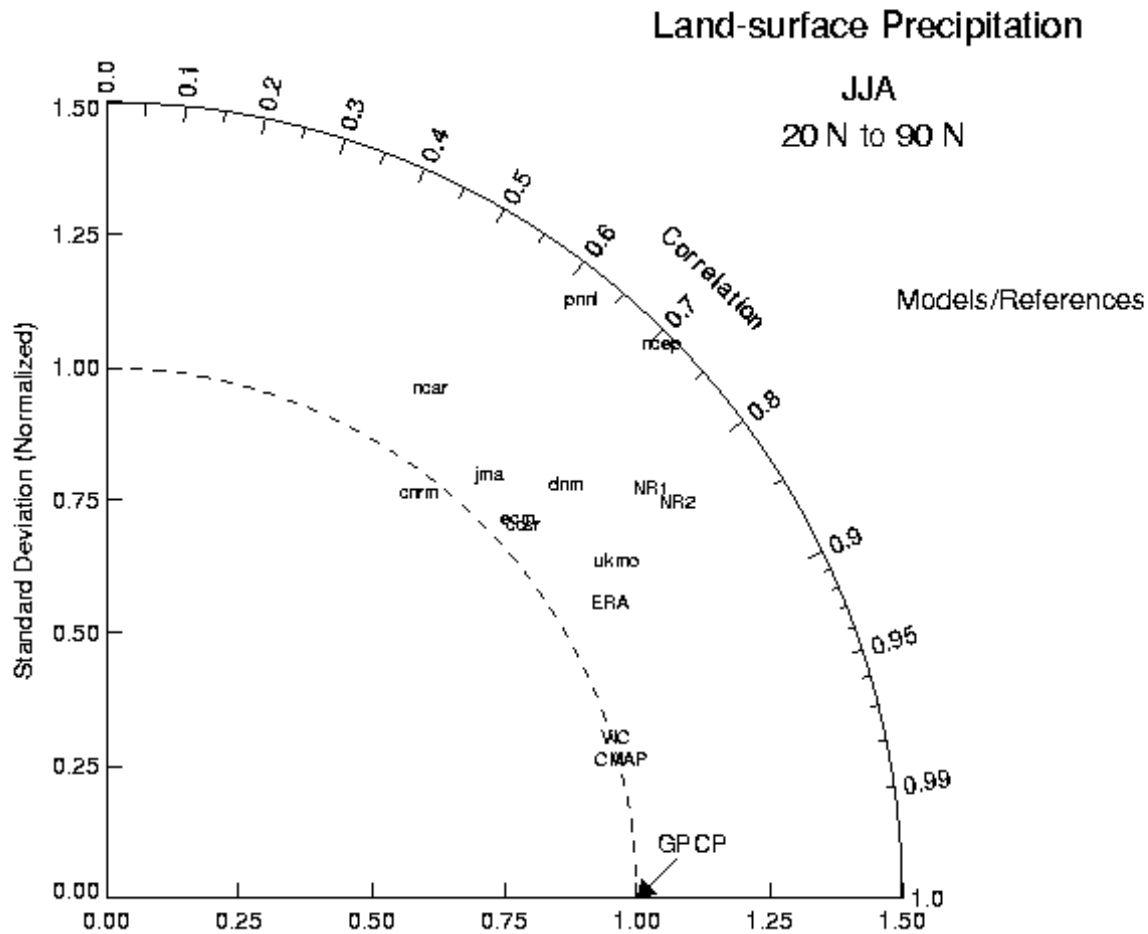


The Taylor diagram of [Figure 8](#) shows the amplitude and phase components of these pattern errors, when aggregated over all seasons and latitude sectors (as in the top row of the portrait diagram of [Figure 7](#)). Both the amplitude and phase errors in land-surface precipitation are generally much greater than the comparable errors in land-surface air temperature ([Figure 4](#)). Note also that the precipitation errors in the reanalyses, as well as in many of the model simulations, are due to excessive amplitudes of variability relative to the [GPCP reference](#); however, the phase errors in the NCEP reanalyses are marginally less than in the majority of the models. On the other hand, the amplitude and phase errors of the [VIC](#) and [CMAP](#) alternative validation data relative to the [GPCP reference](#) are qualitatively less than any of the reanalyses.

Figure 8: As in [Figure 4](#), except for land-surface precipitation **pr** relative to the [GPCP reference data](#).



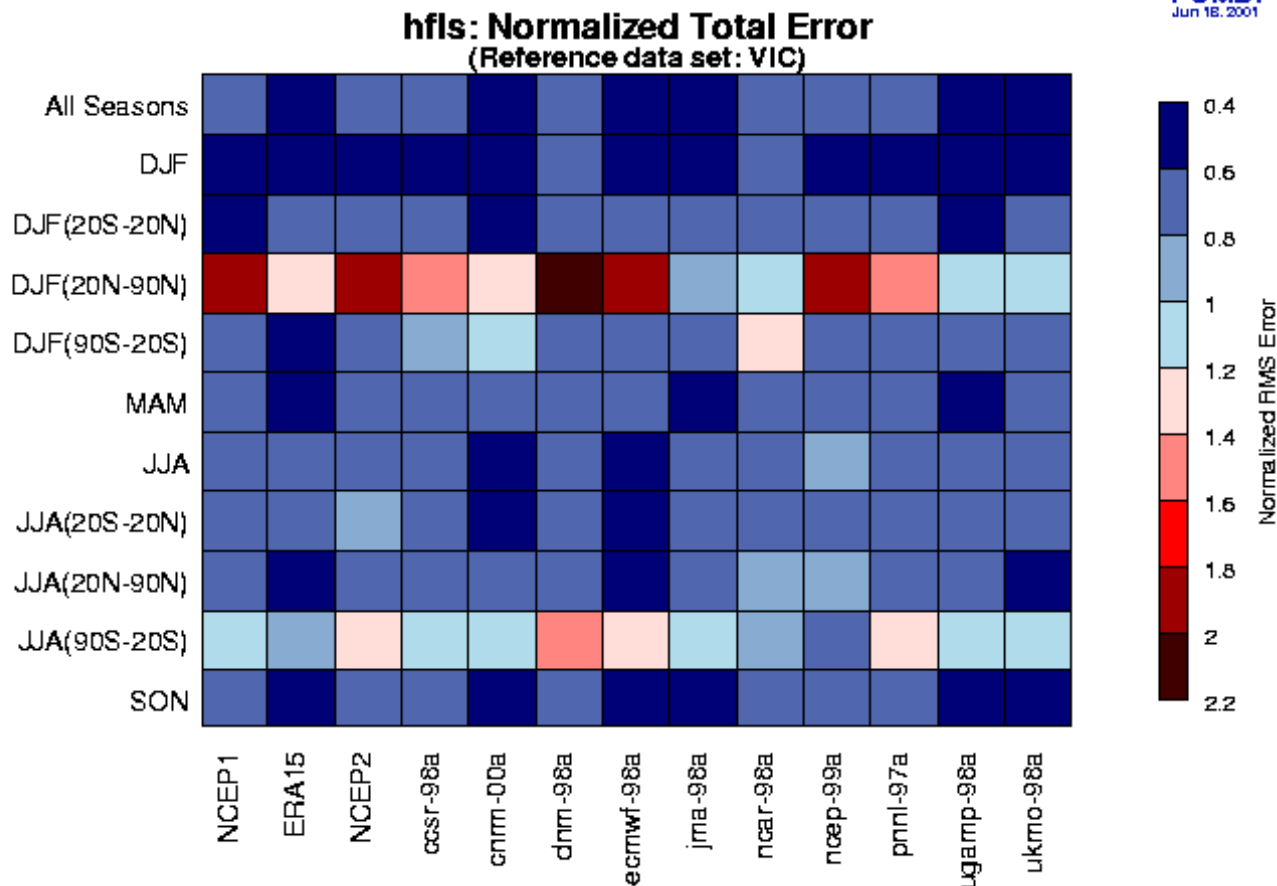




### 5.2 Land-surface Latent Heat Flux

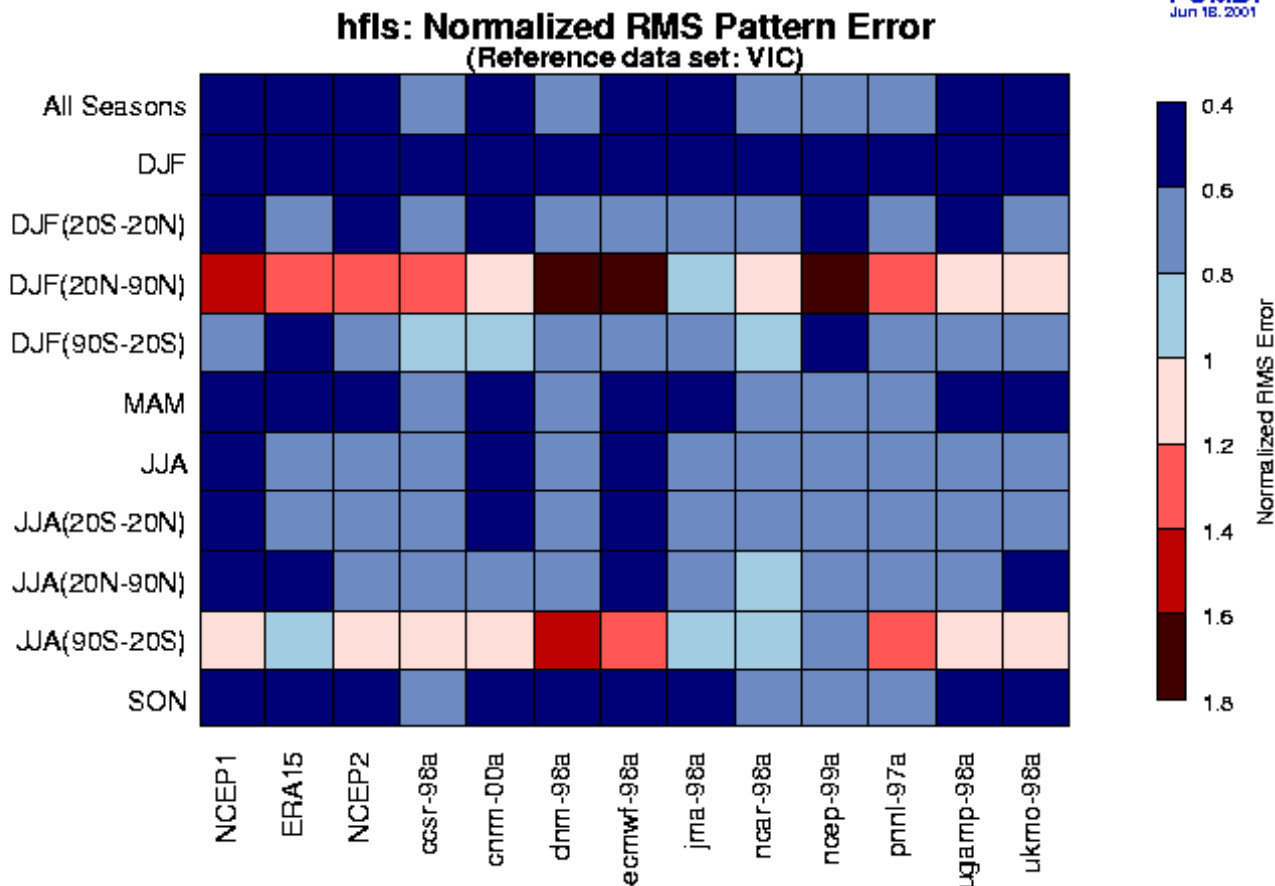
Relative to the [VIC reference data](#), the normalized total RMS errors in the surface latent heat flux are sizeable (~0.4-0.8) when aggregated over all land and seasons (see top row of portrait diagram in [Figure 10](#)), but they are much larger (up to ~ 2) in the extratropics of the winter hemisphere, especially for DJF in the Northern Hemisphere. The latter latitude/seasonal-specific characteristics are seen in the total error of the reanalyses in addition to most of the AMIP simulations.

Figure 10: As in [Figure 1](#), except for land-surface latent heat **hfls** relative to the [VIC reference data](#).



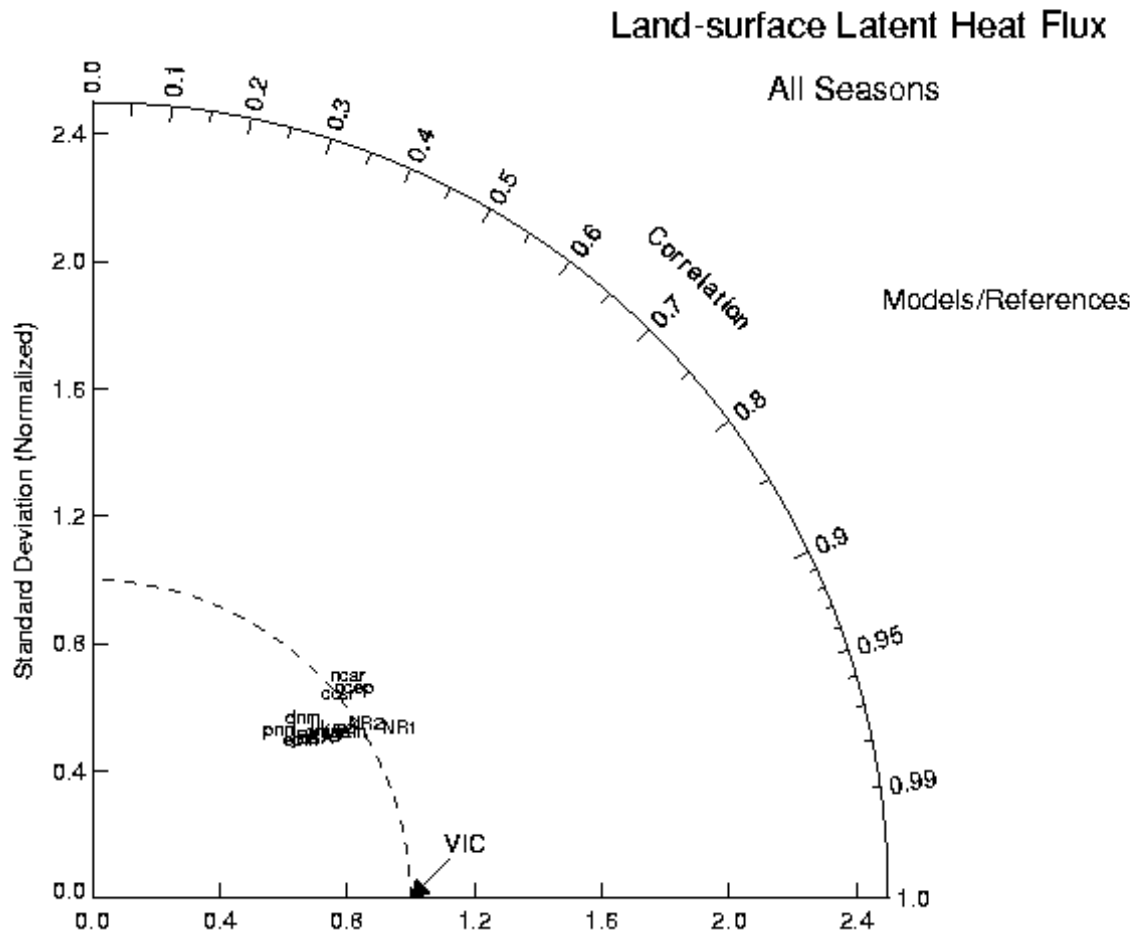
In these seasons/latitudes, pattern errors and bias errors are both sizeable, but the former are somewhat larger (up to ~1.8--see [Figure 11](#)) for both the reanalyses and the majority of model simulations.

Figure 11: As in [Figure 3](#), except for land-surface latent heat **hfls** relative to the [VIC reference data](#).



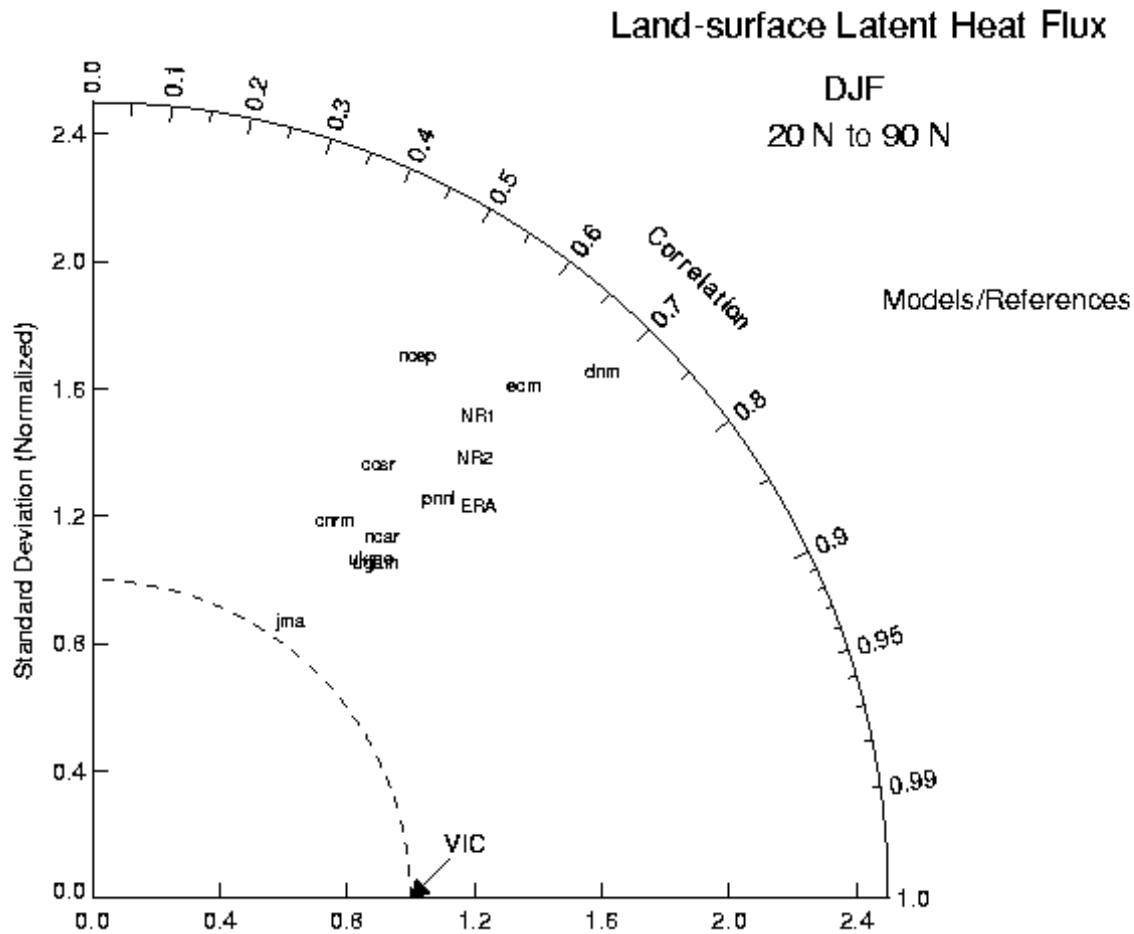
The Taylor diagram decomposition of pattern errors aggregated over all land/seasons ([Figure 12](#)) indicates that the average simulation has about the same amplitude as the [VIC reference](#), with relatively high correlation  $R \sim 0.8-0.9$ .

Figure 12: As in [Figure 4](#), except for land-surface latent heat flux relative to the [VIC reference data](#). (Note, the radial scale of the plot is extended farther than necessary for consistency with that of [Figure 13](#).)



However, the Taylor diagram of pattern errors in DJF Northern extratropics ([Figure 13](#)) indicates general overprediction of amplitude with very large inter-model scatter, and generally large phase errors (correlations reduced to ~ 0.5-0.7); moreover, in both these respects, the reanalyses yield results similar to the model simulations. From comparison of individual model/reanalysis maps with the [VIC reference](#) in DJF Northern extratropics (not shown), these large pattern errors appear to be related mainly to differences in observed (as in the VIC forcing) vs. simulated snow cover.

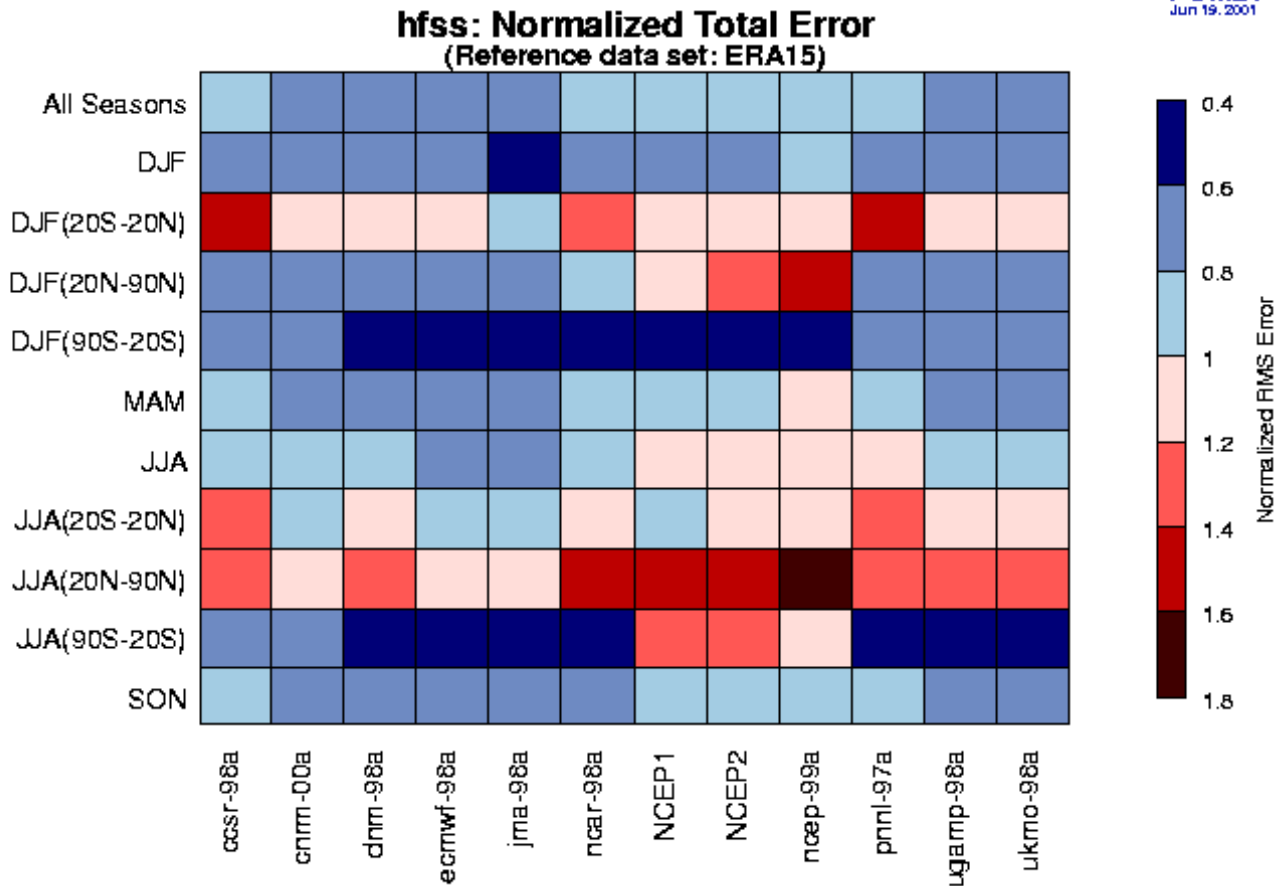
Figure 13: As in [Figure 12](#), except displaying the structure of the normalized pattern error  $E_n'$  computed only over the DJF season and only in the Northern extratropical latitude sector 20 N to 90 N.



### 5.3 Land-surface Sensible Heat Flux

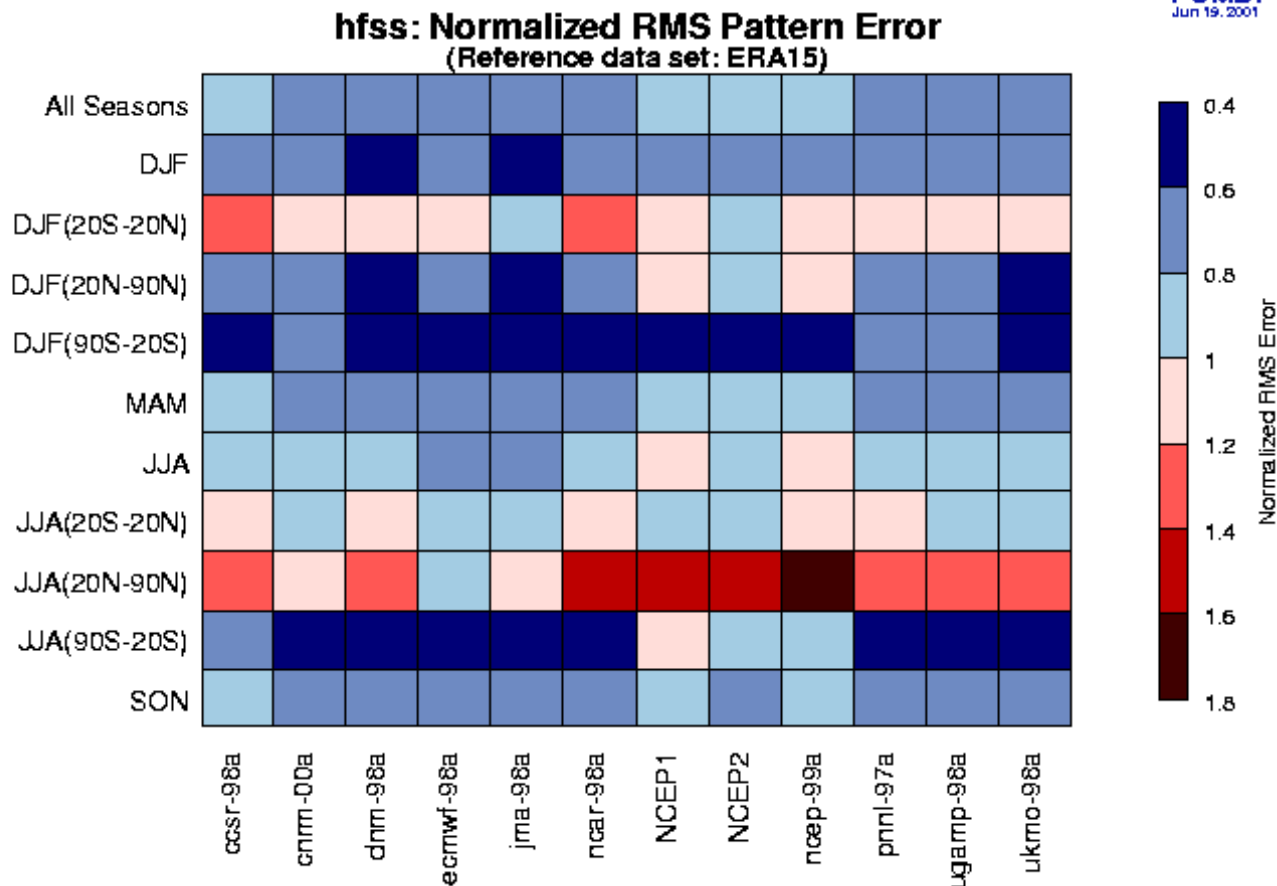
In the case of land-surface sensible heat flux, we had no alternative validation data to the reanalyses (see [Table 2](#) and related comments), so the [ERA15](#) data set was arbitrarily chosen as the reference. Relative to the ERA15 reference, the model simulations of land-surface sensible heat flux show sizable ( $\sim 0.6$ - $1.0$ ) normalized total RMS errors, even when aggregated over all land/seasons (see top row of the portrait diagram in [Figure 14](#)). These errors are generally greatest ( $\sim 1.2$ - $1.8$ ) in the Northern extratropics during JJA, but are also substantial in tropical latitudes during both JJA and DJF. Both NCEP reanalyses also show comparably large total errors relative to the ERA15 reference.

Figure 14: As in [Figure 1](#), except for land-surface sensible heat flux **hfss** relative to the [ERA15 reanalysis](#). Note that results for the alternative NCEP1 and NCEP2 reanalyses are interspersed with those of the ten models.



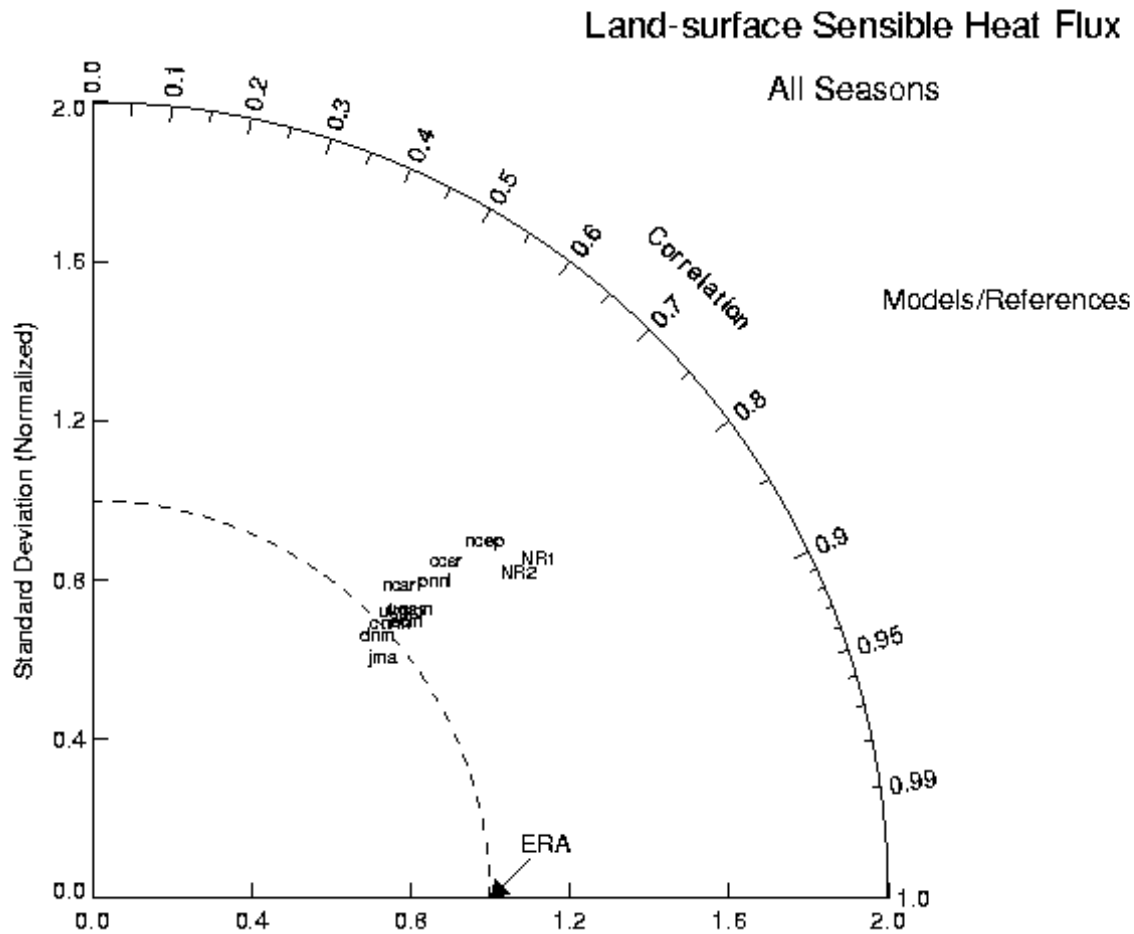
These normalized total errors are found to be due more to pattern errors (Figure 15) than to bias errors (not shown).

Figure 15: As in Figure 3, except for land-surface sensible heat flux hfss relative to the ERA15 reanalysis.



The Taylor diagram decomposition of these pattern errors, when aggregated over all land areas/seasons ([Figure 16](#)), indicates a greater range in amplitude errors (range of amplitude ratios ~ 0.9-1.3) than phase errors (correlations ~ 0.7-0.8 in all instances). In particular, both NCEP reanalyses show comparatively large variability amplitudes relative to both the ERA15 reference and to most of the model simulations; however, the NCEP reanalyses correlate marginally better with the ERA15 reference than do the majority of the model simulations.

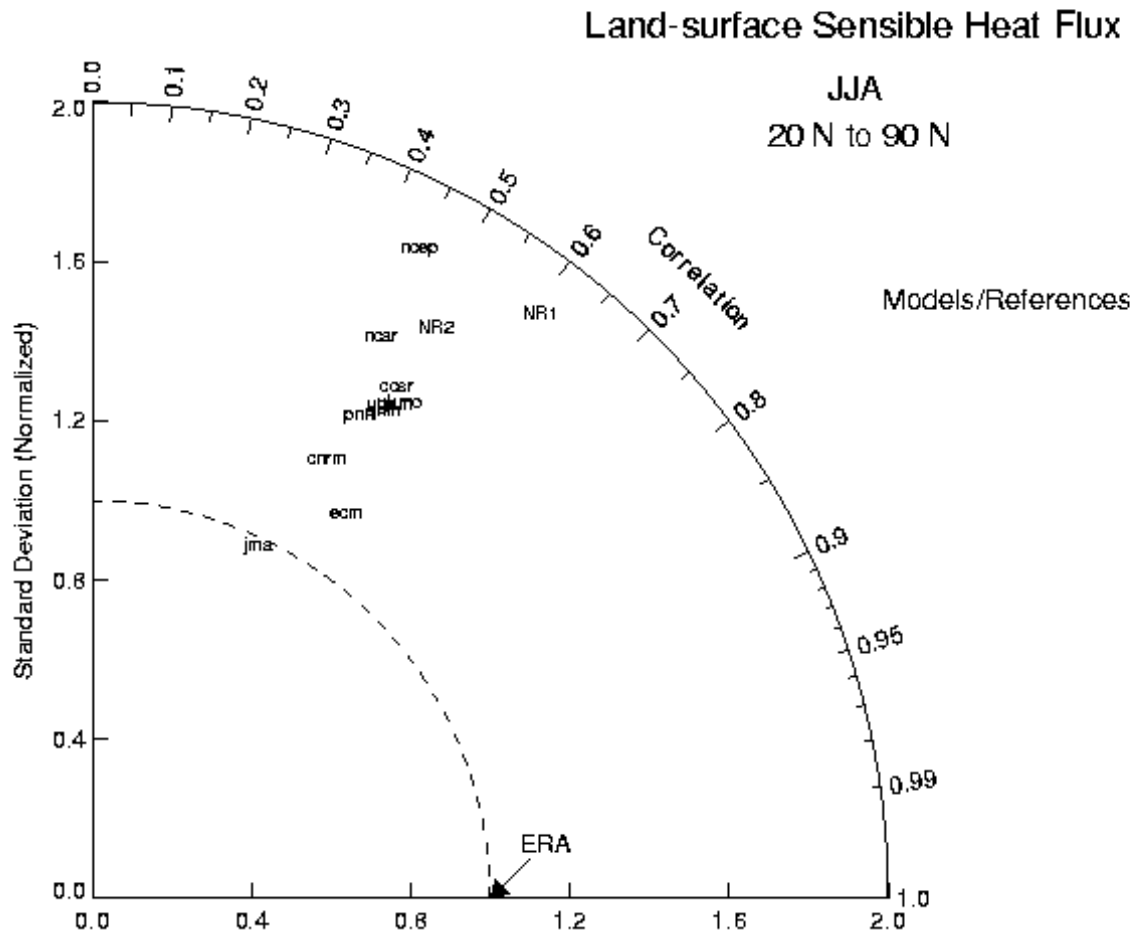
Figure 16: As in [Figure 4](#), except for land-surface sensible heat flux relative to the [ERA15 reanalysis](#).



The Taylor diagram of pattern errors, when limited to the problematical simulations of the JJA Northern extratropics (Figure 17), indicates considerably larger phase errors (correlations reduced to ~ 0.4-0.6) and even greater amplitude errors (range of amplitude ratios ~ 0.9-1.8) than the all-land/all-seasons results (Figure 16). In Figure 17, both of the NCEP reanalyses also again show relatively large amplitude errors relative to the ERA15 reference, with correlations that are not substantially different than those of the model simulations.

Figure 17: As in Figure 16, except displaying the structure of the normalized pattern error  $E_n'$  computed only over the JJA season and only in the Northern extratropical latitude sector 20 N to 90 N.





## 6. Summary

Salient features of this preliminary large-scale model validation are summarized as follows:

- The AMIP II simulations of land-surface air temperature are in generally good agreement with the [Jones et al.\(1999\)](#) reference data--perhaps to be expected when the same estimates of observed SSTs are prescribed, as in this AMIP II model intercomparison. When calculations of error statistics are restricted to selected seasons/latitude sectors, however, the simulated results fall somewhat outside the observational uncertainties bracketed by the reference data and various reanalyses.
- Relative to the [GPCP reference data](#) and similar merged in-situ/satellite estimates of precipitation, the selected AMIP II models collectively show substantial errors in both amplitude and phase, even for the greatest degree of aggregation. There is, however, considerable inter-model scatter in these statistics. Also, while the variability structures of the reanalyses' precipitation estimates are marginally in better agreement with that of the reference data, they resemble more the variability of the average AMIP II simulation.
- Relative to the [VIC-derived estimates](#) of land-surface latent heat flux (from observational forcings), the selected AMIP II

simulations do not show a very large range in amplitude or phase errors when calculations are aggregated over all seasons and land areas. However, the inter-model scatter in simulation errors is much greater (especially in amplitude) when attention is restricted to the winter extratropics, probably due mainly to differences in simulated snow cover. Again, the variability structures of the reanalyses' estimates of latent heat flux resemble that of the average AMIP II simulation more than the reference data.

- The AMIP II simulations show substantial RMS errors in land-surface sensible heat flux relative to the [ERA15](#) reference, with amplitude errors predominating. Nevertheless, the model simulations appear to fall within large existing uncertainties in land-surface sensible heat flux, as estimated from RMS differences among the ERA15 and the two NCEP reanalyses. Thus, it is important to include alternative estimates of turbulent fluxes that now are becoming more available.

The tendency for inter-model differences in the large-scale simulation of various land-surface processes to be more sharply highlighted when attention is restricted to particular seasons and/or latitude sectors confirms the need for detailed analysis of the land-surface simulations at finer scales, work which is also in progress (e.g. see [Irannejad et al. 2000](#), [2001](#)).

We also note that, thus far, the large-scale validation statistics do not seem to stratify in an obvious way according to models which employ a bucket scheme versus those with more complex SVAT schemes. However, because only a single example of a bucket scheme, is presently represented in our analysis (i.e. only model **ccsr-98a**), this provisional result will bear watching as we incorporate more AMIP II models with a still broader spectrum of LSS complexities.

**Access to other in-progress work by DSP 12:** <http://www.pcmdi.llnl.gov/pilps3/inprogress/index.html>

### Acknowledgments

In this report, we have applied validation and graphical display software developed by Karl Taylor, Charles Doutriaux, and Krishna AchutaRao of [PCMDI](#), who also assisted us with many implementation details. We gratefully acknowledge, as well, the provision of output data by the [AMIP II modeling groups](#), without which such a study would be impossible.

### References

Bonan, G.B., 1996: A Land Surface Model (LSM Version 1.0) for Ecological, Hydrological, and Atmospheric Studies: Technical Description and User's Guide. NCAR Technical Note NCAR/TN-417+STR, National Center for Atmospheric Research, Boulder, Colorado, 150 pp.

Cox, P.M., R.A. Betts, C.B. Bunton, R.L.H. Essery, P.R. Rowntree, and J. Smith, 1999: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *ClimateDyn.*, **15**, 183-203.

Dickinson, R.E., A. Henderson-Sellers, and P.J. Kennedy, 1993: Biosphere-Atmosphere Transfer Scheme Version 1e as Coupled to the NCAR Community Climate Model. NCAR Technical Note NCAR/TN-387+STR, National Center for Atmospheric Research, Boulder, Colorado.

Dirmeyer, P., 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 Liqin Tan, 2001: A multi-decadal global land-surface data set of state variables and fluxes. COLA Technical Report 102, Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland, USA, 43 pp.

- Gedney, N., P.M. Cox, H. Douville, J. Polcher, and P.J. Valdes, 2000: Characterizing GCM land-surface schemes to understand their responses to climate change. *J. Climate*, **13**, 3066-3079.
- Gruber, A., X. Su, M. Kanamitsu, and J. Schemm, 2000: The comparison of two merged rain gauge-satellite precipitation datasets. *Bull. Amer. Meteor. Soc.*, **81**, 2631-2644.
- Hall, F.G., B. Meeson, S.O. Los, and E. Brown de Colstoun, 2001: ISLSCP II: New 10+ year multi-parameter near-surface global data set--including greater carbon and vegetation data--nears completion. BAHC-GEWEX News (joint issue: BAHC News No. 9 and Gewex News Vol. 11, No. 2), November, 2001, pp. 13-15.
- Henderson-Sellers, A., 1999: Atmospheric global climate models' representation of the land surface: From AMIP I to AMIP II. URL <http://www-pcmdi.llnl.gov/pilps3/amip1review/amip1review.pdf>.
- Houser, P.R., M. Rodell, U. Jambor, J. Gottschalck, B. Cosgrove, J. Radakovich, K. Arsenault, M. Bosilovich, J.K. Entin, J.P. Walker, K. Mitchell, H.L. Pan, and C.-J. Meng, 2001: The Global Land Data Assimilation System (GLDAS), BAHC-GEWEX News (joint issue: BAHC News No. 9 and Gewex News Vol. 11, No. 2), November, 2001, pp. 11-13.
- Irannejad, P., A. Henderson-Sellers, T.J. Phillips, and K. McGuffie, 2000: Soil moisture and evaporative fluxes from selected AMIP II global models: Analyzing land-surface complexity by climate stratification. Poster presentation at the GEWEX/BAHC International Workshop on Soil Moisture Monitoring, Analysis, and Prediction for Hydrometeorological and Hydroclimatological Applications, University of Oklahoma at Norman, May 2000. See also <http://www-pcmdi.llnl.gov/pilps3/inprogress/poster2000/poster.html>
- Irannejad, P., A. Henderson-Sellers, T.J. Phillips, and K. McGuffie, 2001: Sensitivity of climate simulations to land-surface complexity: Beginning AMIP II Diagnostic Subproject No. 12. Proceedings of the 12th Symposium on Global Change and Climate Variations, American Meteorological Society, Albuquerque, New Mexico, pp. 193-196. See also <http://www-pcmdi.llnl.gov/pilps3/ams2001/>
- Jones, P.D., M. New, D.E. Parker, S. Martin, and I.G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, **37**, 173-199.
- Mahfouf, J.-F., A.O. Manzi, J. Noilhan, H. Giordani, and M. Deque, 1995: The land surface scheme ISBA within the Meteo-France climate model ARPEGE. Part 1: Implementation and preliminary results. *J. Climate*, **8**, 2039-2057.
- Manabe, S., J. Smagorinsky, and R.F. Strickler, 1965: Simulated climatology of a general circulation model with a hydrologic cycle. *Mon. Wea. Rev.*, **93**, 769-798. *Symp Soc. Exptl. Biol.*, **19**, 205-234.
- Nijssen, B., R. Schnur, and D.P. Lettenmaier, 2001: Global retrospective estimation of soil moisture using the Variable Infiltration Capacity land surface model, 1980-1993. *J. Climate*, **14**, 1790-1808.
- Pan, H-L., and L. Mahrt, 1987: Interaction between soil hydrology and boundary layer developments. *Bound. Layer Meteor.*, **38**, 185-202.
- Phillips, T.J., 2001: Summary documentation of models for intercomparison experiments: Models in the Atmospheric Model Intercomparison Project (AMIP I and AMIP II). URL <http://www-pcmdi.llnl.gov/modeldoc/>.
- Phillips, T.J., A. Henderson-Sellers, P. Irannejad, K. McGuffie, Huqiang Zhang, and the AMIP I Modeling Groups, 2000: Validation of land-surface processes in AMIP models: A pilot study. PCMDI Report No. 63, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California, USA, 20 pp. See also <http://www-pcmdi.llnl.gov/pcmdi/pubs/pdf/rpt63.pdf>
- Phillips, T.J., A. Henderson-Sellers, P. Irannejad, and K. McGuffie, 2001: AMIP II Diagnostic Subproject 12: Land-surface

processes and parameterizations (a joint AMIP/PILPS project). URL <http://www-pcmdi.llnl.gov/pilps3/proposal/>.

Sato, N., P.J. Sellers, D.A. Randall, E.K. Schneider, J. Shukla, J.L. Kinter III, Y-T. Hou, and E. Albertazzi, 1989: Implementing the Simple Biosphere Model in a general circulation model. *J. Atmos.Sci.*, **46**, 2757-2782.

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

Viterbo, P., and A.C.M. Beljaars 1995: An improved land surface parameterization scheme in the ECMWF model and its validation. *J. Climate*, **8**, 2716-2748.

Volodin, E.M., and V.N. Lykossov, 1998: Parameterization of heat and moisture transfer in the soil-vegetation system for use in atmospheric general circulation models: 1. Formulation and simulations based on local observational data. *Izvestiya Atmospheric and Oceanic Physics*, **34**, 405-416.

---

Last update 28 February, 2002. Please direct comments and questions to [Tom Phillips](#).