

Continental-Global Scale Validation of Land Surface Models

Paul Dirmeyer, Zhichang Guo and Xiang Gao

*Center for Ocean-Land-Atmosphere Studies
4041 Powder Mill Road, Suite 302, Calverton, Maryland 20705-3106 USA
dirmeyer@cola.iges.org*

Abstract

We provide a number of examples from recent international, multi-model experiments of the performance and application of land surface schemes in both offline and coupled simulations. Science-driven investigations show a wide range of skill and behavior among models that all purport to represent the same quantities (e.g., near-surface soil moisture). Combining estimates from multiple models can produce an analysis which is of high quality, and usually superior to any individual model. If there is sufficient observational data for calibration, a more sophisticated regression can be formulated to improve skill further. However, it appears that calibrations over data rich regions may not be transferable to other locations. The quality of forcing data, or the coupled atmospheric model, can play a large role in the quality of the resulting land model estimates. Land models can also produce estimates of brightness temperatures as they would be observed by satellite, opening new opportunities for calibration, validation and assimilation. Finally, the behavior of coupled land-atmosphere model systems is beginning to be validated to ascertain whether the feedbacks are being simulated at the appropriate strength. In all cases, research would greatly benefit from a greater amount of observational data.

1. Introduction

For weather and climate forecasting with numerical models, we need land surface initial and boundary conditions that have four basic qualities: 1) They are consistent with the land surface scheme (LSS) in the forecast model; 2) They are consistent with the spatial and temporal resolution(s) of the forecast model; 3) They are consistent with the precursory observed atmospheric conditions; 4) They are complete through the soil column. We then need the LSSs to perform adequately, representing realistically the physical processes which govern the conservation and transfer of water and energy at the land surface, and the exchanges between land and atmosphere.

Development of LSSs is typically performed at the laboratory level, which encourages scientific inquiry and innovation. However, the tasks of calibration and validation are tedious, require large amounts of observational data from disparate sources, and often necessitate drawing on resources greater than individual labs can provide.

Recent activities within the Global Energy and Water Experiment (GEWEX) Modeling and Prediction Panel (GMPP) have shown the value of participation in focused multi-model process investigations to improve both our understanding of the natural world, and to accelerate model development and the improvement in performance of participating schemes (Holtslag and Randall 2001, Randall et al. 2003). Within GMPP, and particularly the GEWEX Global Land-Atmosphere System Study (GLASS; Polcher et al. 2000), a number of actions are underway specifically targeting the application of land surface models and coupled land-atmosphere models towards understanding key problems in earth system science.

Offline testing of process models includes stand-alone testing of the LSS at individual locations where observed forcing and validation data are complete, as well as integration in a distributed framework akin to the scheme's application in weather and climate applications. There exists through the Second Global Soil Wetness Project (GSWP-2; IGPO 2002) a ready framework for global model evaluation and comparison for the 10-year period 1986-1995. Regional and local evaluations can take place in the framework of various

phases of the Project to Intercompare Land-surface Parameterization Schemes (PILPS; Pitman et al. 1998). Both of these efforts use a common data formatting standard that simplifies participation in multiple tests. Testing of the fully coupled land-atmosphere system in GLASS takes place in the framework of the Global Land-Atmosphere Coupling Experiment (GLACE; Koster et al. 2004). A new effort is beginning to bring together the land, boundary layer, and cloud modelers to understand land-atmosphere interactions at the process level. All of the GLASS activities, and related studies that consolidate and share large validation data sets, are of tremendous use to large-scale validation of individual LSSs. Recent efforts in these areas form the basis of this overview of continental to global scale validation of land models.

2. Observational data and validation

Observational data are necessary for assimilation, initialization, model validation and calibration. The lack of global observational networks is a hindrance for weather and climate modeling (Leese et al. 2001). What exists is a smattering of highly localized and spatially irregular *in situ* measurements in national or state networks, sparse international networks dedicated to research aims only, and one-of-a-kind field experiments which are not coordinated nor distributing data in real time (Dirmeyer et al. 2004a). The global network of rain and streamflow gauges has been deteriorating for decades over large regions of the world, particularly in developing countries (Vörösmarty et al. 2001).

Satellites can provide measurements of surface skin temperature and other surface properties that are related to brightness temperatures at various electromagnetic wavelengths. Remote sensing holds great promise to provide spatially distributed measurements on a large scale for immediate use, but many key aspects of the land surface cannot or will not be measured from space in the foreseeable future. Surface soil wetness and snow coverage can be measured by satellite where the signal is not degraded by excessive overlying vegetation, but the amount of water in the snowpack or below the surface is only measured in an insufficiently small number of observing stations around the world. Therefore, one must rely on numerical models to fill the gaps and serve as the basis for global analyses that are continuous in space and time, for purposes of monitoring as well as the initialization of weather and climate forecasts.

There are second-order ways to infer the skill of LSSs in simulating land surface state variables and fluxes. For instance, there have been numerous studies which show that changes in the formulation, initialization, or prescription of the land surface in coupled land-atmosphere models improves the simulation of the atmospheric component of climate (e.g., Fennessy and Shukla 1999, Dirmeyer 2000, Douville et al. 2001, Douville 2002, Schlosser and Milly 2002, Koster and Suarez 2001, 2003). In these cases, the coupled land-atmosphere system is validated against distributed near-surface meteorological measurements on large spatial scales.

State variables such as surface soil moisture and surface temperature are observable to some degree from space, which can provide global-scale coverage for validation of LSSs. Soil moisture is also measured in the soil column in a number of agricultural regions, either gravimetrically or electronically. Snow coverage can also be measured from space, but the amount of water in the snowpack is difficult to measure except *in situ*.

Radiative fluxes may also be measured remotely, but fluxes of sensible heat and water (evapotranspiration and runoff) cannot be measured directly. All measurements of heat fluxes from the surface must rely on some assumptions about conduction, convection, gradients and turbulence. Instrumented towers are used to estimate heat fluxes that can be compared to LSS simulations. Runoff can be measured integratively through streamflow, but model grids generally do not conform to the boundaries of catchment basins. In all cases, calculated variables from a gridded LSS will not compare directly to the spatial and temporal scales sampled by instruments. Some compromising assumptions must be made in any validation exercise

Beyond the choice of variables, there is also the question of what statistics of the model to validate. For ease of comparison on large spatial and temporal scales, one typically tries to distill the comparisons to a single number. For measurement of accuracy, root mean square error is most often used over a spatial or temporal domain, although there are arguments against its use in favor of mean absolute error or mean bias error (C. J. Willmott, pers. comm.).

If one is more concerned with the ability of a LSS to simulate the pattern of variations in space or time, some form of correlation may be more appropriate. Correlation is a particularly important metric when one is interested transferability of a quantity such as soil moisture from observations to a model (e.g., for initialization of a forecast) or between models. Because different models (and observations) have different climatologies and “operating ranges” for soil moisture (Koster and Milly 1997), values cannot be conveyed without some sort of transformation. Transformations can be constructed so that the correlation with observations is not changed (e.g., rescaling so that the mean and variance are preserved; Dirmeyer et al. 2004b).

3. Validation in uncoupled LSSs

Large-scale validation in GSWP-2 is being conducted in several ways, including overall skill compared to all stations in the Global Soil Moisture Data Bank (Robock et al. 2000). Figure 1 shows the median correlation with all stations where at least 25% of monthly readings for the period 1986-1995 are available. Results from the COLA version of the Simplified Simple Biosphere model (SSiB; Xue et al 1991; 1996; Dirmeyer and Zeng 1999) are shown for a number of GSWP-2 sensitivity studies, where substitutions are made to elements of the forcing data or boundary conditions compared to the baseline simulation (the black bar, with characteristics given at the bottom of the plot). The shades for each bar denote which aspect of the input data has been altered (*Hyb* stands for a hybrid of reanalysis and observational data sets). It is clear that the quality of the simulation is dependent on the quality of the forcing data. A change to even one field (e.g., precipitation or vegetation) can have a profound effect on the results. Likewise, the result is not always intuitive. For example, substitution of the ERA40 model downward surface radiation for SRB leads to the best simulation of soil wetness, even though SRB (and ISCCP) are derived directly from remote sensing observations.

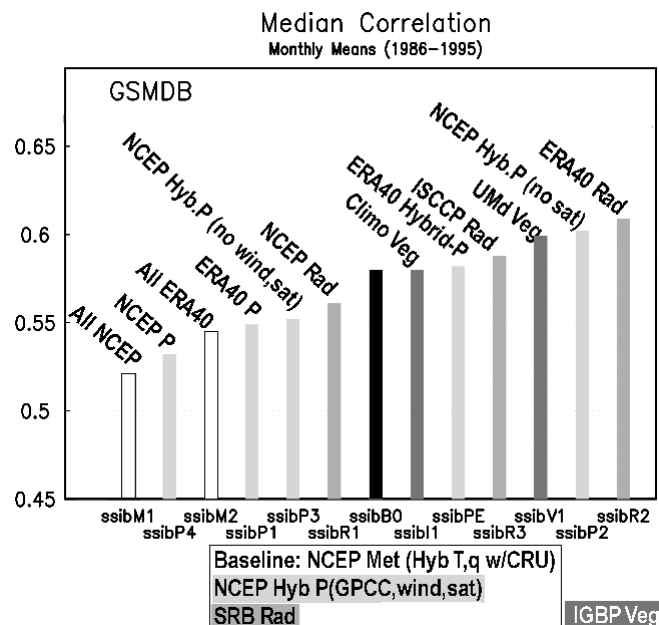


Figure 1: Median correlation of 120-month time series of 1m soil wetness from SSiB simulations in GSWP-2 sensitivity studies and station data from GSMDB..

Comparison of longer-term offline model products (1980-1999; Dirmeyer et al. 2004b) are shown as the first, second and fifth sets of bars in the panels of Fig 2. Clearly there are products which tend to perform better than others on the whole, but no model is always best or worst, and every model ranks first or last for some stations. This fact helps to motivate our attempts at multi-model analyses, described in section 6.

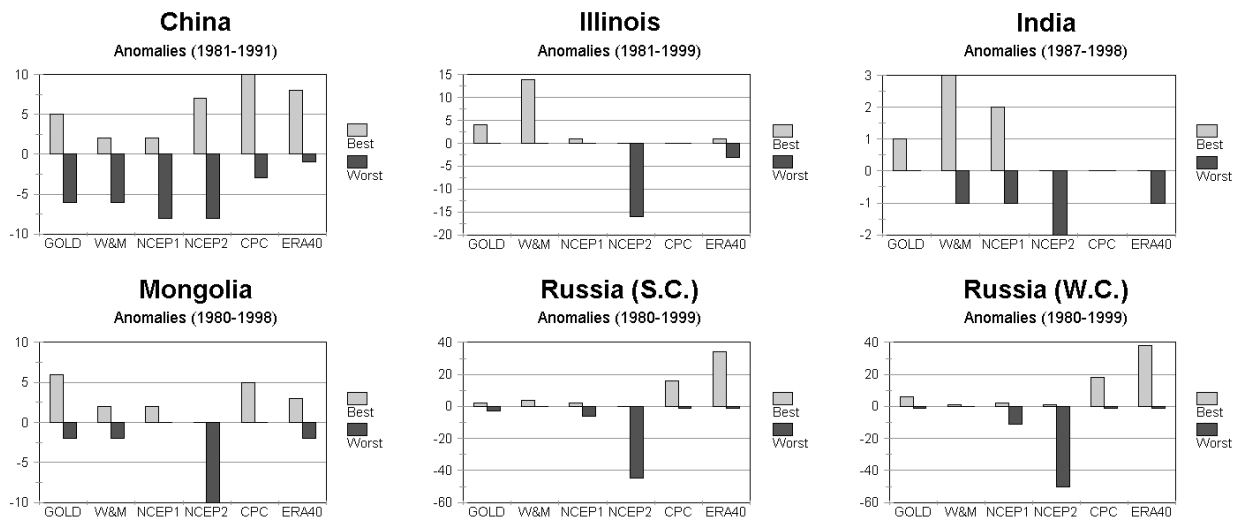


Figure 2: Number of stations for which each model gave the best or worse correlation with the observed 20-year time series of monthly anomalies, where at least one model showed a statistically significant correlation at the 95% level.

Forward brightness temperature modeling has also been applied to model output in GSWP-2. This provides another means to expand the validation and assimilation capabilities of current LSSs beyond those few areas where in situ data are readily available. An application for surface soil moisture validation is to couple the LSSs with a validated L-band microwave emission model – MEB (Pellarin et al. 2003) to simulate forward brightness temperature as would be observed from microwave radiometers. Several participating LSSs have been examined for their performance in simulated brightness temperature, which is assessed by comparisons with airborne measurements from several large-scale campaigns held during the GSWP2 period. Such an assessment constitutes a useful prototype of LSSs validation on a global scale when future satellite-based L-band radiometry data are available, such as the NASA Hydrosphere State (HYDROS) mission and the European Space Agency (ESA) Soil Moisture Ocean Salinity (SMOS) mission.

4. Validation in coupled LSSs

Figure 2 also shows the performance of soil moisture estimates from coupled land-atmosphere analyses, namely the reanalysis products of operational weather centers in Europe and the United States. There appears to be a greater range of skill in the reanalysis products, with some performing very poorly in this metric (it should be noted that while the NCEP-DOE reanalysis (Kanamitsu et al. 2002) often has very poor time series correlations compared to other products, including the original NCEP-NCAR reanalysis (Kalnay et al. 1996), it has smaller mean errors than the original reanalysis (Lu et al. 2004).

One drawback of surface water and energy balance estimates from reanalyses is that the assimilation of state variables applied to reduce the errors in the analyses necessarily introduce increments into the predictive equations such that the local water and energy budget does not close. This can be a serious drawback when one is studying the water and energy balances themselves. Of course, local observations of all terms of these budgets never perfectly balance either, due to measurement errors and limits on instrument accuracy.

The models participating in GLACE have been subject to an experiment which quantifies the degree of coupling from land to atmosphere within the hydrologic cycle (Koster et al. 2004). A great deal of variability is found both in the spatial patterns of feedback strength over the globe, and in the relative strengths of the feedbacks among models. Table 1 shows the rankings of the twelve models' globally averaged coupling strength, and the separate components arising within the associated LSS and AGCM (Guo et al. 2004).

Model	L – A.	Land	Atmosphere
GFDL	1	1	4
NSIPP	2	4	2
NCAR	3	8	1
CCC	4	2	6
CSIRO	5	6	3
UCLA	6	7	7
CCSR	7	9	8
COLA	8	11	5
GEOS	9	3	10
BMRC	10	10	9
Hadley	11	5	11
NCEP	12	12	12

Table 1. Ranks of GLACE GCMs in land-atmosphere coupling strength, and the contributions from the LSS and AGCM separately.

Of course, the pressing question is whether any of these models exhibit a realistic degree of land-atmosphere feedback. The quantity used to estimate coupling strength is the change in the coherence in the time series of precipitation among 16 ensemble members across a 3-month period between a control ensemble with each member initialized with the soil moisture from 1 June of a different year of an AMIP-2 simulation, and a test ensemble where soil moisture is prescribed at each model time step to exactly match a randomly chosen member of the control ensemble. The real world progresses as a single realization, not as an ensemble, so the same statistical manipulation cannot be performed on observational data. However relationships among components of the surface energy and water budgets can be plotted and compared with observations at selected sites where data records are sufficiently long.

Figure 3 shows the relationship between evaporative fraction (here given as the ratio of daily mean evapotranspiration to daily net surface radiation) and soil moisture at each model's grid box containing the AmeriFlux site at Little Washita, Kansas during June-August (note: only ten of the models reported all of the necessary variables to perform this calculation). Also shown in Fig 4 is the corresponding plot from the observational site for 1996-1998 (Falge et al. 2001). The models show a range of coherence in the relationship between soil wetness and evaporation. The models with the strongest coupling strengths generally show the tightest relationships between soil wetness and evaporative fraction. Several models have a very limited range of soil wetness for this location. None show the same distribution of soil wetness as observations, although some models (e.g., Hadley Centre and NSIPP) span a similar range to the observations. And none of the models except the Hadley Centre show the weak relationship between soil wetness and evaporative fraction over a range of soil wetnesses that appears in the observations at this site. It should be noted that the Little Washita site is in a transitional zone between arid and humid regimes, where the multi-model GLACE analysis suggested one of the strongest coupling strengths in the world (Koster et al. 2004). On the other hand, a tighter relationship between available energy and evaporation is found in the observations than the models at this site (not shown). Initial investigations at more humid locations show a similarly weak relationship in the observations between soil moisture and evaporative fraction, and a strong relationship between net radiation and evapotranspiration, with the models showing somewhat better

agreement with observations. Over more arid regions, one would expect evaporation to be more moisture limited than energy limited, but observational data from more arid sites has yet to be analyzed. Nevertheless, these initial comparisons suggest that the model-based conclusions of Koster et al. (2004) may need to be revisited, as the implied regions of strong land-atmosphere coupling strength for the real world may be either weaker, or displaced farther into the arid zones than the models suggest.

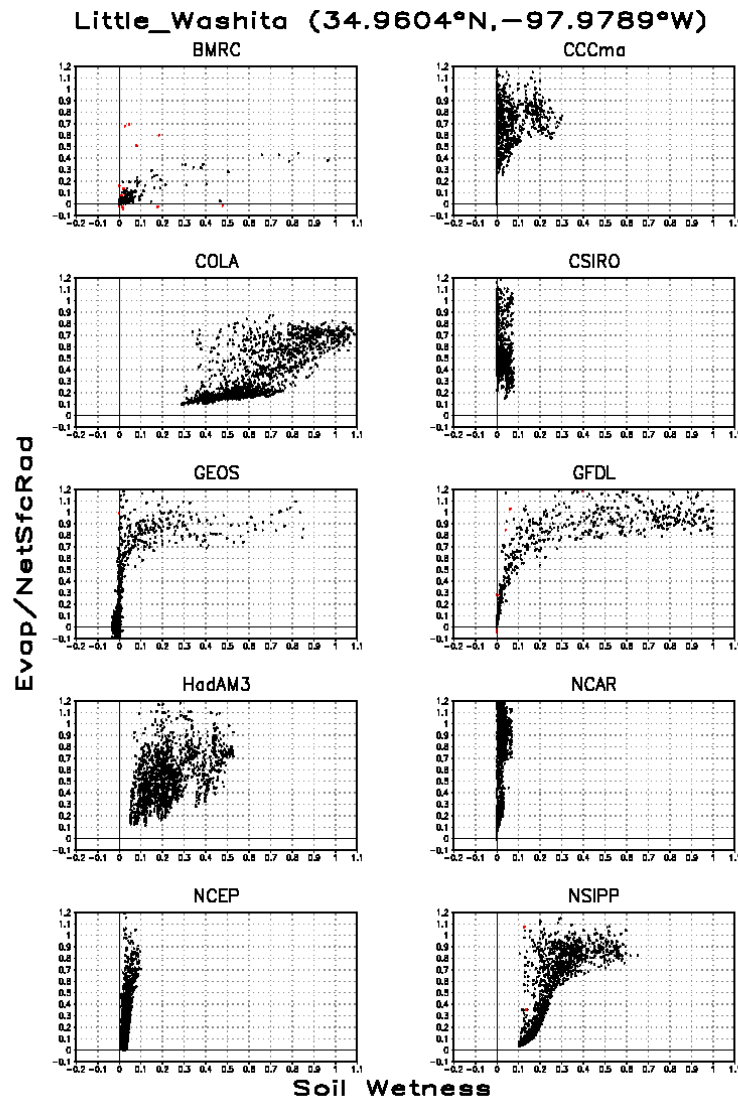


Figure 3 Plot of daily mean evaporative fraction versus soil wetness for model grid box containing the Little Washita field site.

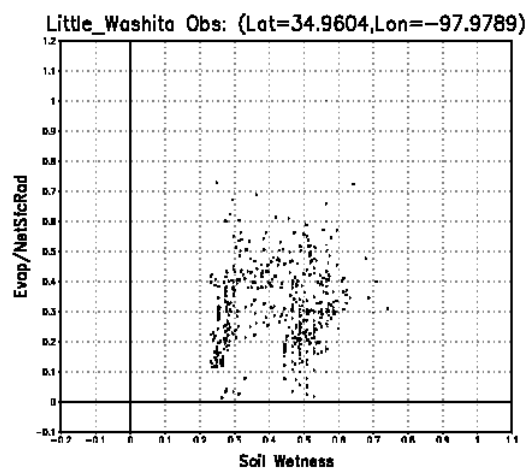


Figure 4 As in Fig 3 for observations.

5. Multi-model analyses

One way to easily improve upon the performance of individual land surface models is to statistically combine the simulations of multiple models. It can be shown analytically that given a large enough sample, an ensemble mean should always outperform individual members in predicting nonlinear systems. For atmosphere and ocean models, the skill is gained by reduction of noise, and thus enhancing the ratio of signal to noise in the simulations. For land models, noise is not a major issue, since they behave in a more deterministic fashion than fluid dynamical systems. The gain in skill is realized by the reduction of systematic errors and pathological behaviors in individual LSSs, which arise from poor calibration or an incomplete representation of physical processes.

One caveat for the construction of multi-model analyses of the land surface is to carefully consider how to deal with inconsistencies between models (their formulation, resolution, representation or omission of certain state variables or fluxes in their output streams). For example, rarely will two unique models have the same vertical structure of soil levels. Thus an analysis of soil moisture at 10cm intervals will require interpolation. And if some models represent ice in the soil and others do not, how should one represent the multi-model analysis of frozen soil?

Gao et al. (2004) showed that when sufficient validation data are available, one can use a seasonally-dependent linear regression to improve the correlation of any individual model simulation of soil wetness time series with observations. A regression on the multi-model mean performs even better, and the best skill can be had when each model has its own calibrated weight in the multi-model mean. However, in the absence of regression, the multi-model mean generally does as well or better than the best single model at any location.

This approach is being used in GSWP-2 to create a multi-model analysis of the land surface for the period 1986-1995 from the baseline simulations of all participating LSSs. This analysis is scheduled to be completed and released in early 2005.

In a very real sense, the conclusion of Koster et al. (2004) is the result of a multi-model analysis of coupled land-atmosphere models. Because the pattern of land-atmosphere feedbacks is highly model dependent, an average across twelve models is means to produce an assessment that is not heavily dependent on the vagaries of any single model. It remains to be seen if the multi-model estimate of the strength and distribution of land atmosphere coupling is in better agreement with observational evidence than any of the individual model products.

6. Conclusions

The lack of observational data makes large-scale validation of land models quite difficult. We need global *in situ* observational networks for land surface state variables (Leese et al., 2001). Remote sensing holds promise for providing the global coverage needed, especially if combined with land surface models in data assimilation systems. However, ground truth capabilities are still needed, so that remote sensing can be used to fill the gaps in the network and contribute to continuous spatial coverage.

Numerous ways of testing offline and coupled land-atmosphere models are being developed and applied. Some examples have been given here, both in the area of validation of LSSs, and improving upon basic simulations. Land surface modeling (and its relationship to weather and climate forecasting, climate change and applications to society) is still a relatively immature field of study, and it will be a grand research area during the next decade.

References

- Dirmeyer, P. A., and F. J. Zeng, 1999: An update to the distribution and treatment of vegetation and soil properties in SSiB. COLA Technical Report 78 [Available from the Center for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705 USA], 25 pp.
- Dirmeyer, P. A., 2000: Using a global soil wetness data set to improve seasonal climate simulation. *J. Climate*, **13**, 2900-2922.
- Dirmeyer, P. A. (ed.), 2004a: Part C: The value of land surface data consolidation, *Vegetation, Water, Humans and the Climate: A New Perspective on an Interactive System*. Springer Verlag, Berlin, 245-296.
- Dirmeyer, P. A., Z. Guo and X. Gao, 2004b: Comparison, validation and transferability of eight multi-year global soil wetness products, *J. Hydrometeor.*, (submitted).
- Douville, H., F. Chauvin, and H. Broqua, 2001: Influence of soil moisture on the Asian and African monsoons. Part I: Mean monsoon and daily precipitation. *J. Climate*, **14**, 2381-2403.
- Douville, H., 2002: Influence of soil moisture on the Asian and African monsoons. Part II: Interannual variability. *J. Climate*, **15**, 701-720.
- Falge, E., and 33 co-authors, 2001: Gap filling strategies for longterm energy flux data sets. *Agr. Forest Meteor.*, **107**, 71-77.
- Fennessy, M. J., and J. Shukla, 1999: Impact of initial soil wetness on seasonal atmospheric prediction. *J. Climate*, **12**, 3167-3180.
- Gao, X., P. A. Dirmeyer, Z. Guo, 2004: Multi-model analysis, validation, and transferability study for global soil wetness products. (in preparation).
- Guo., Z., R. D. Koster, P. A. Dirmeyer, G. Bonan, E. Chan, P. Cox, H. Davies, C. T. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C.-H. Lu, S. Malyshev, B. McAvaney, J. L. McGregor, K. Mitchell, D. Mocko, T. Oki, K. Oleson, A. Pitman, Y. C. Sud, C. M. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada, 2004: GLACE: The Global Land-Atmosphere Coupling Experiment. 2. Model characteristics and comparison. *J. Hydrometeor.*, (submitted).
- Holtslag, B., and D. Randall, 2001: New boundary layer study. *GEWEX News*, **11**, No. 1, 6.
- International GEWEX Project Office, 2002: The Second Global Soil Wetness Project Science and Implementation Plan. *IGPO Publication Series No. 37*, 69 pp.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, & D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631-1648.
- Koster, R. D., and P. C. D. Milly, 1997: The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *J. Climate*, **10**, 1578-1591.
- Koster, R. D., and M. J. Suarez, 2001: Soil moisture memory in climate models, *J. Hydrometeor.*, **2**, 558-570.
- Koster, R. D., and M. J. Suarez, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeor.*, **4**, 408-423.

- Koster, R. D., P. A. Dirmeyer, Z. Guo, G. Bonan, E. Chan, P. Cox, C. T. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C.-H. Lu, S. Malyshev, B. McAvaney, K. Mitchell, D. Mocko, T. Oki, K. Oleson, A. Pitman, Y. C. Sud, C. M. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada, 2004: Regions of coupling between soil moisture and precipitation. *Science*, **305**, 1138-1140.
- Leese, J., T. Jackson, A. Pitman, and P. Dirmeyer, 2001: GEWEX/BAHC international workshop on soil moisture monitoring, analysis, and prediction for hydrometeorological and hydroclimatological applications. *Bull. Amer. Meteor. Soc.*, **82**, 1423-1430.
- Lu, C.-H., M. Kanamitsu, J. O. Roads, W. Ebisuzaki, and K. E. Mitchell, 2004: Evaluation of soil moisture in the NCEP-NCAR and NCEP-DOE global reanalyses. *J. Hydrometeorol.*, (submitted).
- Pellarin, T., J.-P. Wigneron, J. C. Calvet, M. Berger, H. Douville, P. Ferrazzoli, Y. H. Kerr, E. Lopez-Baeza, J. Pulliainen, L. P. Simmonds, and P. Waldteufel, 2003: Two-year global simulation of L-band brightness temperature over land. *IEEE Trans. Geosci. Remote Sens.*, **41**, 2135-2139.
- Pitman, A. J., and A. Henderson-Sellers, 1998: Recent progress and results from the project for the intercomparison of landsurface parameterizations schemes. *J. Hydrol.*, **212-213**, 128-135.
- Polcher, J., P. Cox, P. Dirmeyer, H. Dolman, H. Gupta, A. Henderson-Sellers, P. Houser, R. Koster, T. Oki, A. Pitman, and P. Viterbo, 2000: GLASS: Global Land-Atmosphere System Study. *GEWEX News*, **10**, No. 2, 3-5.
- Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.*, **84**, 1547-1564.
- Robock, A., K. Ya. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai, 2000: The global soil moisture data bank. *Bull. Amer. Meteor. Soc.*, **81**, 1281-1299.
- Schlosser, C. A., and P. C. D. Milly, 2002: A model-based investigation of soil moisture predictability and associated climate predictability. *J. Hydrometeorol.*, **3**, 483-501.
- Vörösmarty, C. J., A. Askew, W. Grabs, R. G. Barry, C. Birkett, P. Döll, B. Goodison, A. Hall, R. Jenne, L. Kitaev, J. Landwehr, M. Keeler, G. Leavesey, J. Schaake, K. Strzepek, S. S. Sundarvel, K. Takeuchi, and F. Webster, 2001: Global water data: A new endangered species. *Eos, Transactions, American Geophysical Union*, **82**, 54, 56, 58.
- Xue, Y., P. J. Sellers, J. L. Kinter and J. Shukla, 1991: A simplified biosphere model for global climate studies. *J. Climate*, **4**, 345-364.
- Xue, Y., F. J. Zeng, and C. A. Schlosser, 1996: SSiB and its sensitivity to soil properties - a case study using HAPEX-Mobilhy data. *Global and Planetary Change*, **13**, 183-194.