4.2 LAND SURFACE DATA ASSIMILATION

Paul R. Houser
NASA Goddard Space Flight Center, Greenbelt, MD

1. INTRODUCTION

Accurate initialization of land surface moisture and energy stores is critical in weather and climate prediction because of their regulation of surface water and energy fluxes between the surface and atmosphere over a variety of time scales. Since these are integrated states, errors in land surface forcing and parameterization accumulate in land stores, leading to incorrect surface water and energy partitioning. However, many new land surface observations are becoming available that may provide additional information necessary to constrain the initialization of land surface states critical for weather and climate prediction. These constraints can be imposed in two ways. Firstly, by forcing the land surface primarily by observations (such as precipitation and radiation), the often severe atmospheric numerical weather prediction land surface forcing biases can be avoided. Secondly, by employing land surface data assimilation techniques, observations of land surface storages (soil temperature, soil moisture, and snow depth/cover) can be used to constrain unrealistic simulated storages.

Therefore, high-resolution continental and global Land Data Assimilation System that uses relevant remotely-sensed and in-situ observations within a land data assimilation framework has been developed. This development will greatly increase our skill in land surface, weather, and climate prediction, as well as provide high-quality, global land surface *assimilated data fields* that are useful for subsequent research and applications. Analysis of the constant confrontation of model predictions with observations at various time and space scales provides an opportunity to improve our understanding and assessment of the space-time structure of land-atmosphere interaction, the relationship between model estimates and observations of land surface conditions, and the role of the land surface in regulating hydrologic and climatic variability.

2. METHODOLOGY

2.1 Land Surface Modeling

Recent advances in understanding soil-water dynamics, plant physiology, micrometeorology, and hydrology, all of which control biosphere-atmosphere interactions, have spurred the development of land surface models (Figure 1). The primary goal of a land surface model is to represent in a simple, yet realistic way, the transfer of mass, energy, and momentum between a vegetated surface and the atmosphere [Dickinson et al., 1993; Sellers et al., 1986]. Land surface model predictions are regular in time and space, but these predictions are influenced by model structure, errors in input variables and model parameters, and inadequate treatment of sub-grid scale spatial variability. Consequently, land surface model predictions of land surface hydrology and land surface states are much improved by the assimilation of land surface observations.

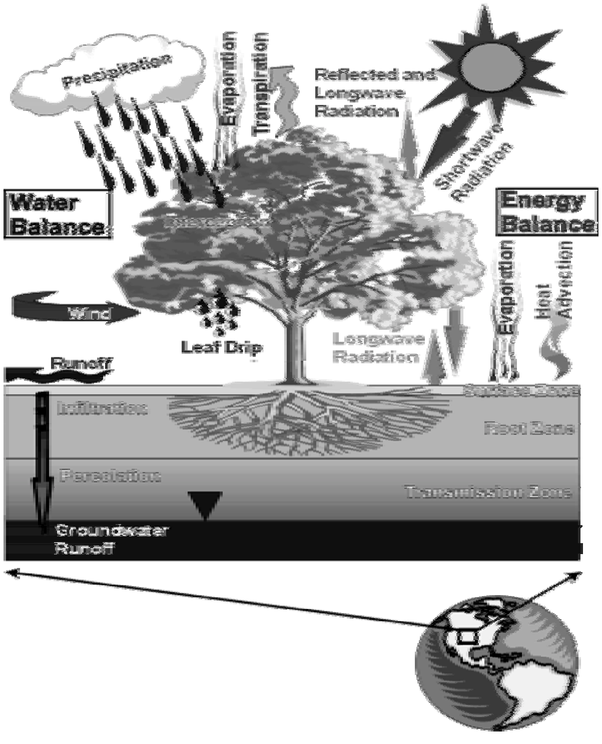

Figure 1: Land surface modeled processes.

---

[1] *Corresponding Author Address*: Paul R. Houser, NASA Goddard Space Flight Center, Greenbelt, MD 20771, Paul.Houser@nasa.gov

2.2 Land Surface Remote Sensing

The emphasis of land surface data assimilation research is to assimilate remotely-sensed observations of the land surface that previous research suggests will provide memory to the land-atmosphere interaction. Remote observations of interest include: (1) temperature, (2) soil moisture (surface moisture content, surface saturation, total water storage), (3) other surface water bodies (lakes, wetlands, and large rivers) and (4) snow (areal extent, snow water equivalent).

The land surface emits thermal infrared radiation at an intensity directly related to its emissivity and temperature. The absorption of this radiation by atmospheric constituents is smallest in the 3 to 5 and 8 to 14 micrometer wavelength ranges, making them the best atmospheric windows for sensing land surface temperature. Generally, surface temperature remote sensing can be considered an operational technology, with many spaceborne sensors making regular observations (for example, the Landsat Thematic Mapper, Advance Very High Resolution Radiometer, the Moderate Resolution Imaging Spectroradiometer, and the Advanced Spaceborne Thermal Emission and Reflection Radiometer [Lillesand and Kiefer, 1994]). The evolution of land surface temperature is linked to all other land surface processes through physical relationships.

Remote sensing of soil moisture content is a developing technology, although the theory and methods are well established [Eley, 1992]. Long-wave passive microwave remote sensing is ideal for soil moisture observation, but there are technical challenges involved in correcting for the effects of vegetation and roughness. Soil moisture remote sensing has previously been limited to aircraft campaigns [e.g. Jackson, 1997a], or analysis of the Defense Meteorological Satellite Program Special Sensor Microwave Imager [Engman, 1995; Jackson, 1997b]. The Special Sensor Microwave Imager has also been successfully employed to monitor surface saturation/inundation [Achutuni and Scofield, 1997; Basist and Grody, 1997]. The Earth Observing System Advanced Microwave Sounding Unit will provide additional C-band microwave observations that may be useful for soil moisture determination.

An important and emerging technology with respect to land surface observation is the potential to monitor variations in total water storage (ground water, soil water, surface waters (lakes, wetlands, rivers), water stored in vegetation, snow and ice) using satellite observations of the time variable gravity field. The Gravity Recovery and Climate Experiment, an Earth System Science Pathfinder mission launched in 2002, will provide highly accurate estimates of changes in terrestrial water storage in large. Wahr et al. [1998] note that the Gravity Recovery and Climate Experiment will provide estimates of variations in water storage to within 5 millimeters on a monthly basis.

Key snow variables of interest to land surface understanding include area coverage and snow water equivalent. While the estimation of snow water equivalent by satellite is currently in research mode, snow areal extent can be routinely monitored by many operational platforms, including The Advanced Very High Resolution Radiometer, the Geostationary Operational Environmental Satellite and the Special Sensor Microwave Imager. Recent algorithm developments even permit the determination of the fraction of snow cover within Landsat Thematic Mapper pixels [Rosenthal and Dozier, 1996]. Cline et al. [1998], describe an approach to retrieve snow water equivalent from the joint use of remote sensing and energy balance modeling.

2.3 Land Surface Data Assimilation

Charney et al. [1969] first suggested combining current and past data in an explicit dynamical model, using the model's prognostic equations to provide time continuity and dynamic coupling amongst the fields (Figure 2). This concept has evolved into a family of techniques known as four-dimensional data assimilation. “Assimilation is the process of finding the model representation which is most consistent with the observations” [Lorenc, 1995]. In essence, data assimilation merges a range of diverse data fields with a model prediction to provide that model with the best estimate of the current state of the natural environment so that it can then make more accurate predictions. The application of data assimilation in hydrology has been limited to a few one-dimensional, largely theoretical studies [i.e. Entekhabi et al., 1994; Milly, 1986], primarily due to the lack of sufficient spatially-distributed hydrologic observations [McLaughlin, 1995]. However, the feasibility of synthesizing distributed fields of soil moisture by the novel application of four-dimensional data assimilation applied in a hydrological model was demonstrated by Houser et al. [1998]. Six Push Broom Microwave Radiometer images gathered over the United

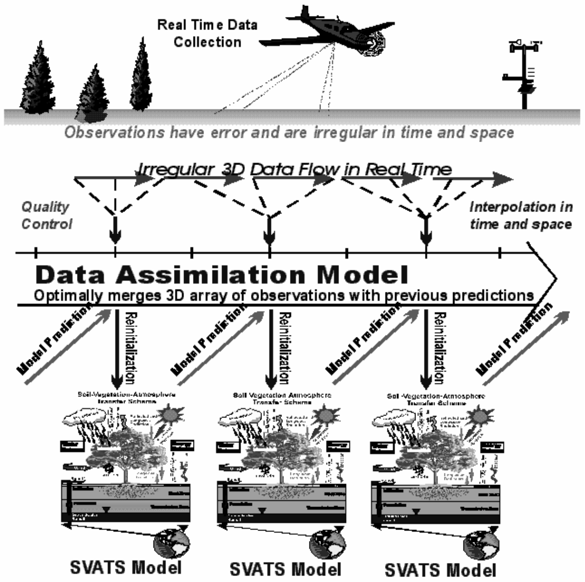

Figure 2: The land surface data assimilation process.

States Department of Agriculture, Agricultural Research Service Walnut Gulch Experimental Watershed in southeast Arizona were assimilated into a land surface model using several alternative assimilation procedures. Modification of traditional assimilation methods was required to use these high-density Push Broom Microwave Radiometer observations. The images were found to contain horizontal correlations with length scales of several tens of kilometers, thus allowing information to be advected beyond the area of the image. Information on surface soil moisture was also assimilated into the subsurface using knowledge of the surface-subsurface correlation. Newtonian nudging assimilation procedures were found to be preferable to other techniques because they nearly preserve the observed patterns within the sampled region, but also yield plausible patterns in unmeasured regions, and allow information to be advected in time.

2.4 Land Data Assimilation Systems

The Global Land Data Assimilation System has its basis in the North American Land Data Assimilation System project [Mitchell et al. 1999]. The North American Land Data Assimilation System was initiated in 1998 with the goal of modeling land surface states and fluxes, while relying as much as possible on observation-based parameter and forcing fields in order to avoid biases that are known to exist in forcing fields produced by atmospheric models. The study region for the North American Land Data Assimilation System encompasses the conterminous United States and parts of Mexico and Canada. The land surface models implemented in the North American Land Data Assimilation System are run at 1/8th degree latitude by 1/8th degree longitude resolution. Separate versions of the system have been developed at the National Aeronautics and Space Administration's Goddard Space Flight Center, the National Oceanic and Atmospheric Administration's National Centers for Environmental Prediction, Princeton University, the University of Washington, and National Oceanic and Atmospheric Administration's Office of Hydrology. Each group runs their land surface models both in real time, retrospectively using the same high quality parameter and forcing fields, thus enabling unambiguous intercomparison of the land surface model simulations. Results are being validated by researchers at Rutgers University, using time series of observed variables, including soil moisture and temperature, to validate the strengths and weaknesses of each model. Much of the Global Land Data Assimilation System program code was derived from Goddard Space Flight Center's North American Land Data Assimilation System program code, and many of the project specifications are identical.

One of the primary objectives of the Global Land Data Assimilation System was to develop a system that would allow users to run multiple land surface models without specific knowledge of the models' architectures or physics. Currently, program code for three land surface models has been installed. Designing a Global Land Data Assimilation System simulation only requires modification of a single, simple interface file, which includes switches and variables for many run time options (summarized in Table 2). The Global Land Data Assimilation System program code interprets the forcing data to the individual input requirements of each respective land surface model, so that the same data can be used to force multiple land surface models. Thus, the influence of discrepancies in forcing data can be eliminated when comparing land surface fields simulated by different land surface models.

As a standard, all Global Land Data Assimilation System models run on a common 0.25 degree longitude by 0.25 degree latitude grid which is nearly global, covering all of the land north of latitude 60 degrees South. The Global Land Data Assimilation System also is able to run on 0.5 degree longitude by 0.5 degree latitude, a 1.0 degree longitude by 1.0 degree latitude, and a 2.5 degree longitude by 2.5 degree latitude global

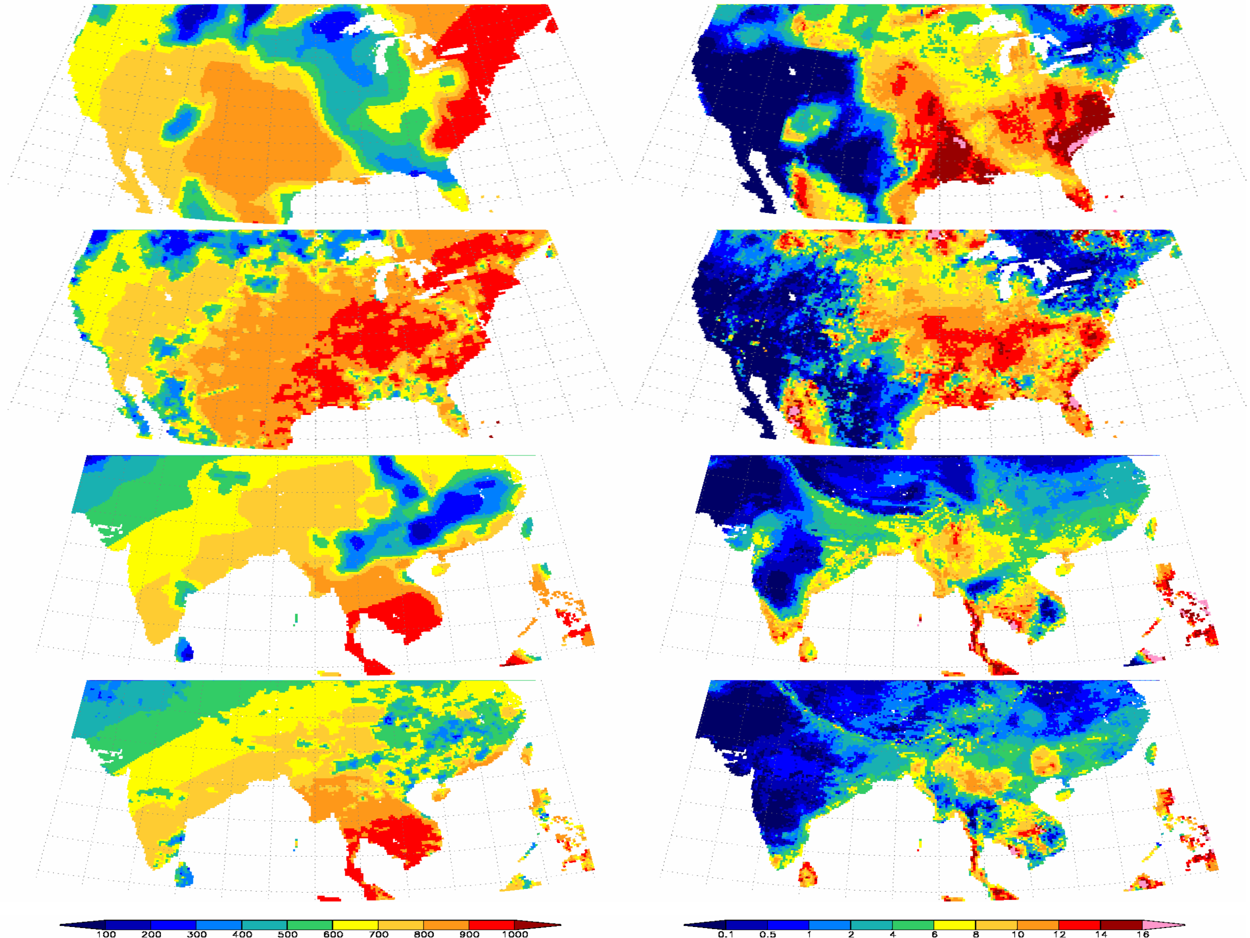

Figure7: Downward shortwave radiation forcing (W/m^2; left) and output total evapotranspiration rate (mm/day; right). From top to bottom: Control Run 18-21Z 31 July 2001; Derived Forcing Run 18-21Z 31 July 2001; Control Run 6-9Z 31 January 2002; Derived Forcing Run 6-9Z 31 January 2002. Top four: central North America; bottom four: southeast Asia.

grid. Subgrid variability is simulated using a vegetation-based tiling approach, as described in the next section. The model time step is user-defined (15 minutes is standard). Forcing data is typically available on 0.25 degree longitude by 0.25 degree latitude to 1.0 degree longitude by 1.0 degree latitude grids with three or six hourly resolution. The Global Land Data Assimilation System includes spatial and temporal interpolation routines based on commonly accepted algorithms. The Global Land Data Assimilation System uses a static, 1 kilometer resolution, global vegetation classification dataset produced by the University of Maryland [Hansen et al., 2000] from the Advanced Very High Resolution Radiometer data. The Global Land Data Assimilation System also employs a satellite observation-based, 1 kilometer resolution climatology and, when available, a time series of leaf area index. The soil parameter maps used in the Global Land Data Assimilation System were derived from the global soils dataset of Reynolds et al. [1999]. That dataset includes 5 minute resolution global maps of porosity and the percentages of sand, silt, and clay, which are based on the United Nations Food and Agriculture Organization Soil Map of the World (FAO 1990) linked to a global database of over 1300 soil pedons. The Global Land Data Assimilation System uses the Global 30 Arc-Second Elevation Data Set [Verdin and Greenlee, 1996] as its standard. The Global Land Data Assimilation System corrects the modeled temperature, pressure, humidity, and longwave radiation forcing fields based on the difference between the Global Land Data Assimilation System elevation definition and the elevation definition of the model that created the forcing data. Because some land surface models, including the Mosaic land model, ingest surface or bedrock slope as a parameter, geographic information systems software was used to assess the slope at each Global 30 Arc-Second Elevation Data Set pixel, and from those values the mean slope within each Global Land Data Assimilation System grid cell was computed.

3. RESULTS

The Global Land Data Assimilation System runs daily in an operational mode. The Mosaic land model is the current operational model, but parallel simulations with the Community Land Model and the Noah land model are also used. The Goddard Earth Observing System Data Assimilation System is currently the baseline forcing source. The Goddard Earth Observing System Data Assimilation System precipitation and radiation fields are overwritten with the observation-based fields, when and where these are available. The model spatial resolution is 0.25 degrees longitude by 0.25 degrees latitude, and 10% is the minimum tile area allowed. The model time step is 15 minutes, and output is 3-hourly. Typically the daily near real time runs are complete within 36 to 48 hours of real time.

Results are presented from two simulations (Figure 7) : a Control Run and a Derived Forcing Run. Each run started on 1 January 2001. The Mosaic land model was used for both runs with the operational settings defined in the previous paragraph, except that the combination of forcing fields varied. The forcing data initialization option was used so that the Goddard Earth Observing System Data Assimilation System provided the initial surface energy and water storage states. Evidence suggests that this allowed the model to spin up and achieve reasonable stability in about three months. The Control Run relied on the Goddard Earth Observing System Data Assimilation System forcing exclusively. The Derived Forcing Run used the United States Naval Research Laboratory observation-based precipitation fields and the observation-based downward shortwave and longwave radiation fields.

The greater fine scale variability of the observation-based precipitation is reflected in the fine scale patterns of soil moisture in the Derived Forcing Run. Because rainfall tends to be spatially heterogeneous at local to regional scales and soil moisture shows a high degree of variability at all scales (e.g., Famiglietti et al., [1999]), the fine scale soil moisture variability evident in the Derived Forcing Run results may be preferable to the Control Run results. However, the exact locations of the fine scale features are unlikely to be reliable due to the imprecision of precipitation maps derived from satellite infrared observations of cloud top temperatures.

One of the most ambitious activities of the Global Land Data Assimilation System project has been the assemblage of an archive of global, operational weather forecast model output and observation-based data fields for parameterizing and forcing land surface models. Most of the time series begin around January 2001 and continue up to present. The most recent fields are downloaded daily from forecast centers and groups that process satellite data.

Output fields of land surface states and fluxes from the Global Land Data Assimilation System model simulations are also freely available to the public (http://ldas.gsfc.nasa.gov). The Global Land Data Assimilation System website includes a real time image generator which allows users to view the most recent output fields. Time series are available by request, subject to manpower limitations. It is the intention of the Global Land Data Assimilation System project to encourage broad use of Global Land Data Assimilation System results: for education, policy making, and social, agricultural, and natural hazards planning, as well as scientific research.

4. SUMMARY

Land surface data assimilation is in its infancy, with many open areas of research. Development of land surface data assimilation theory and methods is needed to: (i) better quantify and use model and observation errors, (ii) optimize data assimilation computational efficiency for use in large operational applications, (iii) use radiative transfer forward models to enable the assimilation of brightness temperatures directly, (iv) link model calibration and data assimilation to optimally use available observation information, (v) create multivariate land surface assimilation methods to use multiple observations with complementary information, and (vi) quantify the potential of data assimilation downscaling. Further, the regular provision of remotely-sensed land surface variables with improved knowledge of observation errors in time and space are essential to advance land surface data assimilation. Land surface models must also be improved to: (i) provide more observable land model states, parameters, and fluxes, (ii) include advanced processes such as river runoff and routing, vegetation and carbon dynamics, and groundwater interaction to enable the assimilation of emerging observations, (iii) have valid and easily updated adjoint models, and (iv) have knowledge of prediction errors in time and space. The assimilation of new types of land surface observations, such as streamflow, vegetation dynamics, evapotranspiration, and groundwater or total water storage must be developed. Finally, we must understand the impact of land surface assimilation feedbacks on earth system predictions, and optimize the complexity of model, observation, and assimilation for practical real-world land surface problem solving.

5. REFERENCES

Achutuni, R. and R. A. Scofield (1997) The spatial and temporal variability of the DMSP SSM/I global soil wetness index, *AMS Annual Meeting, Proceedings of the13th Conference on Hydrology*, 188-189.

Basist, A. and N. Grody (1997) Surface wetness and snow cover, *AMS Annual Meeting, Proceedings of the13th Conference on Hydrology*, 190-193.

Charney, J. G., M. Halem, and R. Jastrow (1969) Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.* **26**, 1160-1163.

Cline, D. W., R. C. Bales and J. Dozier, Estimating the spatial distribution of snow in mountain basins using remote sensing and energy balance modeling, *Wat. Resour. Res.* **34**(5), 1275-1285.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, D. Lamich (1998) Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. *Mon. Weather Rev.* **126**, 2913-2926.

Dickinson, R. E., A. Henderson-Sellers, and P. J. Kennedy,(1993) Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model. *NCAR Technical Note* 387+STR.

Eley, J. (1992) Summary of Workshop, Soil Moisture Modeling. Proceedings of the NHRC Workshop held March 9-10, 1992, *NHRI Symposium Proceedings* **9**.

Engman, E. T. (1995) Recent Advances in Remote Sensing in Hydrology. *Reviews of Geophysics*, Supplement, 967-975.

Entekhabi, D., H. Nakamura and E. G. Njoku (1994) Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Trans. Geosci. Remote Sensing* **32,** 438-448.

Hansen, M.C., R.S. DeFries, J. R. G. Townshend, and R. Sohlberg (2000) Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing* **21,** 1331-1364.

Houser, P. R., W. J. Shuttleworth, H. V. Gupta, J. S. Famiglietti, K. H. Syed, and D. C. Goodrich (1998) Integration of Soil Moisture Remote Sensing and Hydrologic Modeling using Data Assimilation. *Wat. Resour. Res.* **34**(12), 3405-3420.

Jackson, T. J. (1997a) Southern Great Plains 1997 (SGP97) Hydrology Experiment Plan, http://hydrolab.arsusda.gov/~tjackson.

Jackson, T. J. (1997b) Soil moisture estimation using special satellite microwave/imager satellite data over a grassland region, *Wat. Resour. Res.* **33**(6) 1475-1484.

Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.* **82,** 35-45.

Koster, R. D., and M. J. Suarez (1992) Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *J. Geophys. Res.* **97,** 2697-2715.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000) A catchment-based approach to modeling land surface processes in a GCM, Part 1, Model Structure, *J. Geophys. Res.* **105**, 24809-24822.

Lorenc, A. C. (1995) Atmospheric Data Assimilation. *Meteorological Office Forecasting Research Div*. **34**, The Met Office, UK.

McLaughlin, D. (1995) Recent developments in hydrologic data assimilation, *Reviews of Geophysics,* 977-984.

Milly, P. C. D. (1986) Integrated remote sensing modeling of soil moisture: sampling frequency, response time, and accuracy of estimates. *Integrated Design of Hydrological Networks - Proceedings of the Budapest Symposium* **158,** 201-211.

Mitchell, K., P. Houser, E. Wood, J. Schaake, D. Tarpley, D. Lettenmaier, W. Higgins, C. Marshall, D. Lohmann, M. Ek, B. Cosgrove, J. Entin, Q. Duan, R. Pinker, A. Robock, F. Habets, and K. Vinnikov (1999) GCIP Land Data Assimilation System (LDAS) project now underway, *GEWEX News* **9**(4), 3-6.

Reynolds, C. A., T. J. Jackson, and W. J. Rawls (1999) Estimating available water content by linking the FAO Soil Map of the World with global soil profile databases and pedo-transfer functions, American Geophysical Union, Fall Meeting, *Eos Trans. AGU,* **80.**

Rosenthal, C. W. and Dozier, J. (1996,) Automated mapping of montane snow cover at subpixel resolution from the Landsat Thematic Mapper, *Wat. Resour. Res.* **32** (1), 115-130.

Sellers, P. J., B. W. Meeson, J. Closs, J. Collatz, F. Corprew, D. Dazlich, F. G. Hall, Y. Kerr, R. Koster, S. Los, K. Mitchell, J. McManus, D. Myers, K.-J. Sun, and P. Try (1996) The ISLSCP Initiative I global datasets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies. *Bull. Amer. Meteor. Soc.* **77**, 1987-2005.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986) A simple biosphere model (SiB) for use with general circulation models. *J. Atmos. Sci.* **43**, 505-531.

Verdin, K. L., and S. K. Greenlee (1996) Development of continental scale digital elevation models and extraction of hydrographic features. *Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling,* Santa Fe, NM, January 21-26, National Center for Geographic Information and Analysis, Santa Barbara, CA.

Wahr, J.; Molenaar, M.; Bryan, F. (1998) Time variablity of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.* **103** (B12), 30205-30229.