12

Land, Water and Energy Data Assimilation

David L. Toll and Paul R. Houser

12.1 Introduction

12.1.1 Land–Water–Energy Systems

The land surface stores (temperature, soil moisture, snow, etc.) and modulates global energy and water fluxes that pass between the surface and the atmosphere. Hydrological cycle fluxes move energy as water. Energy used to evaporate water may be released hundreds of kilometres away during condensation, producing clouds and precipitation. Rainfall-runoff processes, weather and climate dynamics, and ecosystem changes all are highly dependent on land surface water and energy budgets.

As people alter the land surface, concern grows about the ensuing consequences for weather and climate, water supplies, crop production, biogeochemical cycles and ecological balances at various time scales. Therefore, it is crucial that any natural or human-induced water cycle changes in the land and atmosphere be assessed, monitored and predicted. For example, Gornitz *et al.* (1997) report that nearly 1% of the total, global annual stream flow is reduced by human activities such as irrigated agriculture contributing to a sea level lowering of 0.8 ± 0.4 mm per year. This offsets the predicted 1–2 mm/year sea level rise attributed to global warming. Accurately assessing land surface hydrology and energy flux spatial and temporal variation is essential for understanding and predicting biospheric and climatic responses. Data assimilation is a key means of improving our knowledge of these processes by optimally constraining model predictions with observational information.

Spatial Modelling of the Terrestrial Environment. Edited by R. Kelly, N. Drake, S. Barr.
 ISBN: 0-470-84348-9.

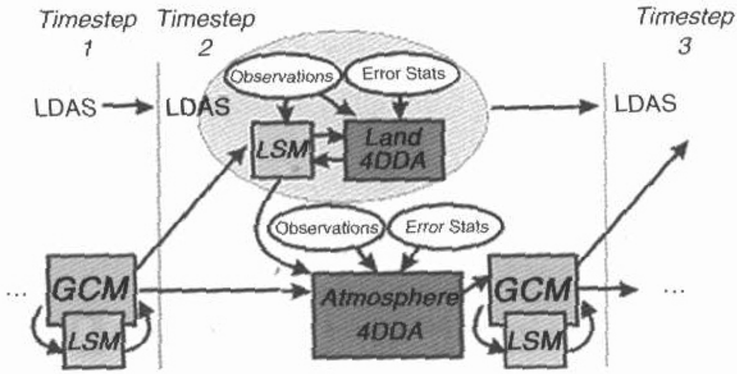

Figure 12.1 *Interaction of the Land Data Assimilation Scheme (LDAS) with an operational Numerical Weather Prediction (NWP) system. The atmospheric General Circulation Model (GCM) is coupled with the Land Surface Model (LSM), and both use a 4-Dimensional Data Assimilation (4DDA) process to integrate past forecasts with observations to improve performance*

12.1.2 Land Data Assimilation

Data assimilation is a numerical scientific tool that can improve land, water and energy budget model estimations through the incorporation of observational constraints. This leads not only to better overall predictions, but also helps to diagnose model weaknesses and can suggest where better parameterizations are needed. The fusion of operational model and observation data via data assimilation requires access to large near-real time surface atmospheric and hydrologic information volumes. Data processing, modelling and data assimilation require large computational and data storage resources that are becoming more achievable with evolving computer technology.

Additionally, new remotely sensed land surface observations are becoming available that will provide the additional information necessary to constrain land surface predictions at multiple time and space scales. These constraints can be imposed two ways (Figure 12.1). First, by forcing the land surface primarily by observations (e.g. precipitation and radiation), the often-severe atmospheric numerical weather prediction biases can be avoided. Second, by employing innovative land surface data assimilation techniques, land surface storage observations such as soil temperature and moisture can be used to constrain unrealistic simulated storages. Land data assimilation techniques also have the ability to maximize the utility of limited land surface observations by propagating their information throughout the land system to times and locations which lack observational data. The primary thrust of this chapter is to highlight land, energy and water data assimilation. Specifically we will do so through the 'Land Data Assimilation Systems' (LDAS) framework.

12.2 Land Data Assimilation Systems (LDAS)

Significant land-surface observation and modelling progress has been made at a wide range of spatial and temporal scales. Projects such as the International Satellite Land Surface Climatology Project (ISLSCP) (Hall *et al.*, 2002), the Global Soil Wetness Project

(GSWP) (Koster and Milly, 1997), and the Global Energy and Water Experiment (GEWEX) Continental-Scale International Project (GCIP), among others have paved the way for operational Land Data Assimilation System (LDAS) development. The LDAS development serves as an integrating linkage between a variety of Earth science disciplines and geographical locations. The LDAS are forced with real time output from numerical prediction models, satellite data, and radar precipitation measurements. Many model parameters are derived from high-resolution satellite-based vegetation coverage. But most importantly, LDAS integrates state-of-the-art modelling and observation on an operational basis to provide timely and consistent high quality land states to be used in real-time applications such as coupled land-atmosphere models and mesoscale climate models. A primary LDAS goal is to provide a broad range of information useful for applications, policy making and scientific research. We are currently associated with three LDAS projects: (1) the North American LDAS; (2) the Global LDAS; and (3) the Land Information System (LIS). The Global and North American LDAS provide real-time and selected retrospective simulations and the LIS team is currently developing the high-performance computation capability to perform a relatively high spatial resolution (1 km) global land prediction.

12.2.1 The North American LDAS

The North American LDAS (NLDAS) was initiated in 1998 primarily to derive land surface modelling with observation and model-based forcing fields (e.g. radiation and precipitation) to avoid biases from atmospheric models (Mitchell *et al.*, 1999). NLDAS consists of a number of land surface models that use remote sensing and *in situ* observations gridded to 1/8 degree. NASA, NOAA National Center for Environmental Prediction (NCEP), Princeton University, Rutgers University, the University of Maryland and the University of Washington implement NLDAS in near real time using existing Land Surface Models (LSM's). NLDAS has also been run for a 50-year retrospective period from 1950–2000 (Maurer *et al.*, 2002). NLDAS uses NCEP 'Eta' model analysis fields, along with observed precipitation and radiation fields to force several different land surface models in an uncoupled modelling system.

12.2.2 The Global LDAS

The Global Land Data Assimilation System (GLDAS) enlarges upon NLDAS to the global scale using many of the same algorithms, but necessarily requiring global forcing and parameter fields (Rodell *et al.*, 2002). Whereas a primary NLDAS emphasis is to improve local weather forecasts, a primary GLDAS emphasis is to provide initialization of global coupled weather and climate prediction models. In addition to improved weather prediction and climate modelling, both systems have a broad range of applications in studies of the terrestrial energy and water cycles. Examples include flood and drought assessment, water usage for crops, snowpack and snowmelt for water availability, water and energy data for ecosystem modelling and net primary productivity.

Both NLDAS and GLDAS are able to drive multiple land surface models under one system. A summary of the key system features is given in Table 12.1. Furthermore, input and output fields for NLDAS and GLDAS are summarized in Table 12.2. Outputs from the LDAS simulations are freely available (see http://ldas.gsfc.nasa.gov) and the LDAS

Table 12.1 *Typical programme execution options in LDAS (from Rodell et al., 2002)*

Attribute/Option	Description
Spatial resolution	0.125° (NLDAS); 0.25° to 2.5° (GLDAS)
Land surface model	MOSAIC; CLM; NOAH; catchment (in preparation)
Forcing	Various model and satellite-derived products
Initialization	None (constant value); restart file; forcing data
Subgrid variability	1–13 tiles per grid cell
Elevation adjustment	Temperature; pressure; humidity; long-wave radiation
Data assimilation	Surface temperature; snow cover; soil moisture
Soil classification	Look-up table; Reynolds *et al.* (1999)
Leaf area index	Look-up table; AVHRR/model derived
Inland water tiles	CLM lakes option

website includes a real-time image generator and data subsetting tool that permits viewing and acquisition of recent LDAS results.

12.2.3 Land Information System

Increases in GLDAS resolution to 1 km are planned and will improve land-atmosphere process understanding. However, to process a year of 1 km global data using conventional computers would require an unavailable amount of computer runtime. A new Land Information System (LIS), building on the same land surface modelling and observation fields as GLDAS, is being constructed (http://lis.gsfc.nasa.gov). The LIS uses a high performance, massively parallel computer including a 192-node Beuwolf computer cluster to support throughput demands of a near real-time global 1 km land prediction and data assimilation system. The system will have a web-based user interface designed to facilitate broad and efficient usage. Input and output are based on developing an Earth System Modelling Framework (ESMF, http://www.esmf.ucar.edu) that demonstrates the interoperability of disparate model components and enables the use of remotely sensed data in coupled Earth system models.

12.3 LDAS Components

Land data assimilation typically refers to the incorporation of observational (*in situ* and remote sensing) data into an LSM. Land surface modelling provides the spatial and temporal predictions, whereas observations are typically used as input data or 'correction' data for the numerical simulation. The three central land surface data assimilation system components are the following: (1) land surface simulation; (2) land surface observations; and (3) land surface data assimilation.

12.3.1 Land Surface Simulation

Recent advances in understanding soil-water dynamics, plant physiology, micrometeorology, and biosphere-atmosphere interactions have spurred LSM developments that seek to realistically represent mass, energy, and momentum transfer between a vegetated surface and the atmosphere (Dickinson *et al.*, 1993; Sellers *et al.*, 1986). LSM predictions are regular in time and space, but are influenced by model structure, input forcing and model parameter errors, and inadequate sub-grid-scale spatial variability treatment. Consequently, LSM hydrology and energy prediction will likely be much improved by using assimilation strategies to remove biases and better constrain boundary conditions.

There are many different approaches to land surface prediction, which has led to great diversity in LSMs. Four LSMs either used or soon to be used in LDAS are presented here. These are the: (1) Mosaic LSM of Koster and Suarez (1992); (2) the Catchment LSM of Koster *et al.* (1998); (3) the National Centers for Environmental Prediction (NCEP), Oregon State University (OSU), United States Air Force (USAF) and Office of Hydrology (OH), LSM, called the **NOAH** LSM (Mitchell, 2002); and (4) the recently emerging Community Land Model (CLM) (Bonan, 1998 and Dickinson *et al.*, 2000).

The Mosaic LSM addresses the sub-grid heterogeneity issue by subdividing each GCM grid cell into a user-specified *mosaic* of tiles (after Avissar and Pielke, 1989), with each tile having different vegetation characteristics and hence water and energy balance. Surface flux calculations for each tile are similar to those described by Sellers *et al.* (1986). Like the plethora of LSMs that have been developed over the past decade, tiles do not directly interact with each other, but influence each other indirectly by their collective influence on the coupled overlying atmosphere (Henderson-Sellers *et al.*, 1993). Vukovich *et al.* (1997) report using tiles reduced the average error for three test sites by 11% for sensible heat and 20% for latent heat fluxes. Toll *et al.* (2001) report much of the error is in geographic regions with contrasting cover such as forests, grasses, and crops when using only one land cover class per grid cell. Although the Mosaic LSM is well suited to modelling the vertical exchange of mass, energy and momentum between the land surface and the overlying atmosphere, Mosaic includes no lateral moisture movement representation, which can significantly impact soil water, surface energy fluxes and runoff variation at some scales.

Recognizing this weakness, Koster *et al.* (1998) developed a new, *catchment*-based LSM that includes a more realistic hydrological process representation, including soil water lateral transport through the subsurface. The catchment-based land surface model uses a topographic land atmosphere transfer scheme (TOPLATS) that relies heavily on the concepts originally put forth by Famiglietti and Wood (1994) and Peters-Lidard *et al.* (1997) (i.e., the TOPLATS model). It represents a major advance in LSMs for the following two reasons. First, TOPMODEL's topographically based framework (Beven and Kirkby, 1979) will result in improved runoff prediction, and consequently, a more realistic catchment-scale water balance. Second, the downslope moisture movement within the watershed will yield sub-catchment-scale surface and unsaturated-zone moisture variations, which will result in more realistic prediction of intra-catchment surface flux variations. Ultimately, improved runoff simulation will result in more realistic continental-scale stream flow estimates from the land to the oceans, and similarly, the intra-catchment surface flux variation will improve catchment-average exchanges with the atmosphere.

The NOAH LSM simulates soil moisture (both liquid and frozen), soil temperature, skin temperature, snowpack water equivalent, snowpack density, canopy water content, and the traditional energy flux and water flux terms of the surface energy and surface water balance (Mitchell, 2002). NOAHLSM uses global satellite-derived monthly climatological vegetation greenness fraction, albedo values for different surfaces, and accounts for the snow albedo. Recently, the NOAH has significantly improved both cold season process physics and bare soil evaporation. This model has been used in: (1) the NCEP-OH submission to the 'PILPS-2d' tests for the Valdai, Russia site; (2) the emerging, real-time, North American Land Data Assimilation System (NLDAS); (3) the coupled NCEP mesoscale model and the Eta model's companion 4-D Data Assimilation System (EDAS); and (4) the coupled NCEP global Medium-Range Forecast model (MRF) and its companion 4-D Global Data Assimilation System (GDAS).

The Community Land Model (CLM) is under development by a grass-roots collaboration of scientists who have an interest in making a general land model available for public use. The CLM development philosophy is to use only proven and well-tested physical parameterizations and numerical schemes. The current CLM version includes superior components from each of several contributing models (Bonan, 1998 and Dickinson *et al.*, 1993). The CLM code is managed in an open source style, in that updates from multiple groups will be included in future model versions. Also, the land model was run for a test case suite including many of the Projects for the Intercomparison of Land Parameterization Schemes (PILPS) case studies (Koster and Milly, 1997).

There are strong justifications for using an uncoupled LSM modelling system for LDAS. Although coupling the LSM to an atmospheric model permits the study of the interaction and feedbacks between the atmosphere and land surface, coupled modelling also imposes strong Numerical Weather Prediction land surface forcing biases on the LSM (Mitchell *et al.*, 1999). These precipitation and radiation biases can overwhelm the behaviour of LSM physics. In fact, several numerical weather prediction centres must 'correctively nudge' their LSM soil moisture estimates towards climatological values to eliminate soil moisture 'drift' (Mitchell *et al.*, 1999). By using an uncoupled LSM, LDAS users can better constrain land surface forcing via observations, use fewer computational resources, and still address all of the relevant LDAS goals. The physical understanding and modelling insights gained from implementing distributed, uncoupled land-surface schemes with observation-based forcing have been demonstrated in recent off-line land surface modelling projects such as the GEWEX Global Soil Wetness Project (Koster and Milly, 1997).

12.3.2 Observations and Observation-Based Data

Atmospheric model simulations and observation-based data are used as a baseline to provide forcing fields as input for LDAS and LIS (Table 12.2). Forcing fields for both NLDAS and GLDAS may be obtained through the LDAS website (http://ldas.gsfc.nasa.gov) and from NCEP (ftp://ftp.ncep.noaa.gov/pub/gcp/ldas/noaaoutput). The LDAS is forced with real-time output from numerical prediction models, satellite data and surface radar precipitation measurements. Many LDAS vegetation parameters are derived from high-resolution AVHRR and MODIS observations. See Mitchell *et al.* (1999) for a further description of NLDAS forcing and see Rodell *et al.* (2002) for GLDAS forcing.

Atmospheric Forcing Data. GLDAS provides a choice between three meteorological model inputs. First, the NASA/Goddard Earth Observing System (GEOS) data assimilation system supports Level-4 data products from the NASA TERRA satellite. The data

Table 12.2 *LDAS Model output summary using the mosaic LSM*

Atmospheric	Land Surface and Subsurface	
Net shortwave radiation (W m^{-2})	Surface runoff (Kg mv)	Top 1 m soil moisture (Kg m^{-2})
Net longwave radiation (W m^{-2})	Subsurface runoff (Kg mv)	Snowmelt (Kg m^{-2})
Downward solar radiation flux (W m^{-2})	Average surface temperature (K)	Layer 2 soil moisture (Kg m^{-2})
Downward longwave radiation flux (W m^{-2})	Surface Albedo paul (%)	Layer 3 soil moisture (Kg m^{-2})
Snowfall, frozen (Kg m^{-2}) precipitation (Kg m^{-2})	Total soil column wetness (%)	Snowpack water equivalent (Kg m^{-2})
Rainfall, unfrozen (Kg m^{-2})	Snow depth (m)	Root zone wetness (%)
Surface pressure (Pa)	Snow cover (%)	Canopy surface water evaporation (W m^{-2})
Air temperature, 2 m (K)	Plant canopy surface water storage (kg mv)	Canopy transpiration (W m^{-2})
Specific humidity, 2 m (Kg/Kg)	Deep soil temperature (K)	Aerodynamic conductance (m s^{-1})
U wind component (m s^{-2})	Total column soil moisture (Kg m^{-2})	Canopy conductance (m s^{-1})
V wind component (m s^{-2})	Root zone soil moisture (Kg m^{-2})	Sensible heat flux (W m^{-2})
Convective precipitation (Kg m^{-2})	Vegetation greenness (%)	Latent heat flux (W m^{-2})
	Leaf area index (dimensionless)	Ground heat flux (W m^{-2})

are produced on a 1-degree global grid in 3-hour assimilation datasets. GEOS incorporates Physical-Space Statistical Analysis System (PSAS) data assimilation techniques that combine current boundary conditions (e.g. sea surface temperature) with updated observations and error statistics. Second, the Global Data Assimilation System (GDAS) operational weather forecast model of National Centers for Environmental Prediction (NCEP) is available (Derber *et al.*, 1991). GDAS data at a native resolution of 0.7° is mapped to a 1-degree global grid in the GLDAS project. Third, the European Centre for Medium-Range Weather Forecasts (ECMWF) products are also available. ECMWF produces 6-hour forecasts at approximately 39-km spatial resolution. The analysis includes *in situ* conventional and satellite-derived data using a four dimensional multivariate data assimilation technique. Rodell *et al.* (2002) report the GEOS data provides the best comparison to observation data and was chosen as the primary forcing data for GLDAS.

When possible, observation-based data from satellites, precipitation gauges and Doppler radar is used. For example, in GLDAS and NLDAS observational data may be used to replace or update modelled data that is spatially and temporally contiguous. Observational data are typically used when validation and quality control efforts indicate satisfactory accuracy. With increases in computer power, observational data are now used more frequently in data assimilation (Cohn *et al.*, 1998; Mitchell, 2002).

Land Surface State Data. Operational and accurate global land surface hydrologic predictions require the assimilation of spatially distributed remote sensing-derived observations. Observations of interest include temperature, soil moisture (surface moisture content, surface saturation, total water storage), lake/river height and flow, snow areal extent and snow water equivalent, land cover, leaf area index and albedo.

Surface temperature remote sensing is a relatively mature technology. 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 3–5 and 8–14 μm wavelength ranges, making them the best windows for sensing land surface temperature. Some errors due to atmospheric absorption and improperly specified surface emissivity are possible, and the presence of clouds can obscure the signal. Generally, surface temperature remote sensing can be considered an operational technology (Ma *et al.*, 2002), with many spaceborne sensors making regular observations (i.e., Landsat Thematic Mapper, NOAA Advanced Very High Resolution Radiometer (AVHRR), Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) and Terra Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)). The land surface temperature evolution is linked to all other land surface processes through physically based relationships, which makes its assimilation possible.

Soil moisture remote sensing 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 in correcting for the vegetation and roughness effects. Soil moisture remote sensing has previously been limited to aircraft campaigns (e.g. Jackson, 1997a), or Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager (SSM/I) analysis (Jackson, 1997b) data. SSM/I data was also successfully employed to monitor surface saturation/inundation (Basist and Grody, 1997). The Advanced Microwave canning Radiometer (AMSR) instrument provides additional C-band microwave observations that may be useful for soil moisture determination. The Tropical Rainfall Measuring Mission's (TRMM) Microwave Imager (TMI), which is very similar to AMSR radiometrically, is much better suited to soil moisture measurement (because of its 10 Mhz channels) than SSM/I, and is also currently available. All of these sensors have adequate spatial resolution for land surface applications, but have a very limited quantitative measurement capacity, especially over dense vegetation. However, Sipple *et al.*, (1994) demonstrated that it is possible to detect saturated areas through dense vegetation using Scanning Multichannel Microwave Radiometer (SMMR), which can greatly aid land surface predictions. The SMMR has similar radiometric characteristics to AMSR. Because of the large remotely sensed microwave soil moisture observation error, there is a real need to maximize its information content by using algorithms, such as data assimilation, that can account for measurement error and extend satellite information in time and space.

There is a potential to monitor total water storage variations (ground water, soil water), surface waters (lakes, wetlands, rivers), water stored in vegetation, and snow and ice using time variable gravity field satellite observations. The Gravity Recovery and Climate Experiment (GRACE), an Earth System Science Pathfinder mission, will provide highly accurate terrestrial water storage change estimates in large watersheds. Wahr *et al.* (1998) note that GRACE will provide water storage variation estimates to within 5 mm on a monthly basis. Rodell and Famiglietti (1999) have demonstrated the potential utility of these data for hydrologic application is aimed more at large watersheds (>150,000 km^2). They further discuss the potential power of GRACE to constrain land surface modelled water storage when combined with surface soil moisture and altimetery observations. Birkett (1998) demonstrated the potential of satellite radar altimeters to monitor height variations over inland waters, including climatically sensitive lakes and large rivers and wetlands. Such altimeters are currently operational on the ERS-2, TOPEX/POSEIDON, ENVISAT and JASON-1 satellites.

Key snow variables of interest to land data assimilation include areal coverage and snow water equivalent. While snow water equivalent estimation by satellite is currently in research mode, snow areal extent can be routinely monitored by many operational platforms (Tait *et al.*, 2000), including AVHRR, GOES, MODIS and SSM/I. Recent algorithm developments even permit snow cover fraction determination within Landsat-TM pixels (Rosenthal and Dozier, 1996). Cline *et al.* (1998) describe an approach for retrieving snow water equivalent from the jointly using remote sensing and energy balance modelling.

Other key variables conducive to remote sensing are surface albedo (Toll *et al.*, 1997), land cover and leaf area index (Justice *et al.*, 1998). They are each key variables to global climate, ecology, hydrology and biogeochemical models that may help describe the energy, mass (e.g. water and CO_2) and momentum exchanges between the land and atmosphere. In addition, they are available as data products globally from satellite sensors such as the Terra and Aqua MODIS (Justice *et al.*, 1998).

12.3.3 Data Assimilation

The presence of model and observation error causes the study of highly interactive large-scale land hydrology and energy budgets to be a complex task. A combination of information, including LSM data, remote sensing observations and *in situ* surface data is used for study. There have been recent data assimilation theory advances that have provided quantitative methods for merging the various information types to provide more accurate estimates (Errico, 1999). Lorenc (1995) defines assimilation as the process of finding the model representation that is most consistent with the observations. 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. Most data assimilation techniques can be applied to almost any dynamic geoscience problem, but are often limited by computational feasibility. Earth scientists now face the challenge to apply true data assimilation techniques to all problems where the incorporation of observations can provide new insights. However, this is a difficult task due to the highly nonlinear nature of land surface processes, the problem size, and the lack of data and experience to determine error statistics accurately. Consequently, data assimilation implementation always requires trade-offs between resolution, complexity, computational effort, and data availability.

Data assimilation techniques are used extensively in meteorology (Daley, 1991) and oceanography (Wunsch, 1996). For example, in meteorology data assimilation is routinely used to improve weather forecasting. Currently most operational weather forecast centres use optimal interpolation type schemes (Daley, 1991). NASA's Data Assimilation Office has recently improved this technique by developing the Physical-Space Statistical Analysis

System (PSAS) (Cohn *et al.*, 1998). PSAS operates in the spatial domain and improves the complicated and time-dependent error covariance estimation.

Hydrologic data assimilation, especially at large scales, is in its early stages (Reichle, 2000; McLaughlin, 1995). One formidable problem is that many hydrological processes are non-linear. Currently, there are only a few studies that use distributed watershed models to assimilate field data. Reichle (2000) provides a short Earth sciences data assimilation review with a focus on hydrology. Surface hydrology data assimilation is based primarily on soil moisture information from surface observations or remote sensing. There are several soil moisture estimation data assimilation techniques that use a one-dimensional optimal estimation approach, including studies by Milly (1986), Katul *et al.* (1993), Parlange *et al.*, (1993), Entekhabi *et al.* (1994), Galantowicz *et al.* (1999), Calvet *et al.* (1998) and Castelli *et al.* (1999). Several additional studies have used low-level atmospheric observations to infer soil moisture using one-dimensional optimal variation assimilation approaches (Mahfouf, 1991; Bouttier *et al.*, 1993; Hu *et al.*, 1999; Callies *et al.*, 1998; Rhodin, *et al.*, 1999). In these approaches the calculation of soil moisture is a 'parametric approach' and not physically based. Reichle *et al.* (2002) used four-dimensional variational data assimilation with improved physics. They used large-scale soil moisture profiles along with other soil and vegetation parameters from passive microwave measurements in their data assimilation.

Unlike previous efforts to test assimilation of soil moisture using synthetic data, two recent studies assimilated soil moisture into a LSM with actual remote sensing data. Both Crow and Wood (2003) and Margulis *et al.* (2002) demonstrated the utility of an extended Kalman filter that assimilated airborne microwave brightness data into a LSM. Both studies used ESTAR (Electronically Scanned Thinned Array Radiometer) brightness temperature data (1.4 GHz frequency, L-band), from the 1997 Southern Great Plains Experiment (SGP97), sensitive to soil moisture variations to 5 cm. The data assimilation was reported to be a computationally efficient and more accurate approach than modelling or remote sensing alone. The extended Kalman filter data assimilation was able to derive spatial and temporal trends of soil moisture in the root zone, significantly below the sensitivity of L-band data (to 5 cm). A summary of recent data assimilation soil moisture papers related to LDAS is given next.

Houser *et al.* (1998) demonstrated the feasibility of synthesizing distributed soil moisture fields by the novel application of four-dimensional data assimilation in a hydrological model. Six Push Broom Microwave Radiometer (PBMR) images gathered over the USDA-ARS Walnut Gulch Experimental Watershed in southeast Arizona were assimilated into the TOPLATS hydrological model (Peters-Lidard *et al.*, 1997) using several alternative assimilation procedures. Modification of traditional assimilation methods was required to use the high-density PBMR observations. Information on surface soil moisture was assimilated into the subsurface using surface-subsurface correlation knowledge. Newtonian nudging assimilation procedures were found to be preferable to other techniques because generally they preserve the observed patterns within the sampled region, but also yield plausible patterns in unmeasured regions, and allow information to be advected in time.

Reichle *et al.* (2002) used the ensemble Kalman filter to estimate soil moisture by assimilating microwave L-band (1.4 GHz) brightness temperatures into a land surface model.

They concluded the ensemble Kalman filter specifically reduced soil moisture estimation errors in comparison to results without data assimilation. They also concluded that assimilation schemes that use 'static' forecast error covariances such as in statistical interpolation produce less accurate estimates than the ensemble Kalman filter. In addition, they found that the ensemble Kalman approach is very flexible and is applicable over a broad model error range.

Sun *et al.* (2002) used an extended Kalman filter to assimilate observed snow water equivalent, available through using satellite remote sensing for snow depth, snow temperature and snow water equivalent retrieval. They used a NASA Seasonal-to-Interannual Prediction Project (NSIPP) land surface model and derived 'true' snow states with European Centre for Medium-Range Weather Forecasting (ECMWF) atmospheric forcing data. The data were degraded and then assimilated for comparisons to the 'true' snow states. Because of snow's high albedo, thermal properties, feedback to the atmosphere and as medium-term water storage, improved snow state estimation has the potential to greatly increase the climatological and hydrological prediction accuracy.

Walker and Houser (2001) found the one-dimensional extended Kalman filter effective in assimilating near-surface soil moisture into a land surface model. They degraded a simulation by setting the initial soil moisture prognostic variables to arbitrarily wet values throughout North America. A 'true' land surface simulation was run using International Satellite Land Surface Climatology Project (ISLSCP) forcing data. The study showed assimilation of near-surface soil moisture observations from a remote sensing satellite reduced soil moisture storage error to 3% after a one-month assimilation and to 1% after a 12 month assimilation. They concluded that data assimilation of remotely sensed data may provide accurate initial conditions to GCMs and that these models need not rely on initial conditions from a spun-up LSM simulation.

12.4 LDAS Applications

Real-time North American LDAS and Global LDAS (see http://ldas.gsfc.nasa.gov) that use atmospheric forcing fields to initialize land surface models are currently in place. Figure 12.2 illustrates for July 1996 a continental U.S. 1/8° precipitation image (upper) from assimilated modelling and observational data that were input to LDAS to derive the average surface soil moisture (lower). The precipitation data were assimilated using a NCEP Eta model, NEXRAD Doppler radar and Higgins gauge precipitation data (Mitchell *et al.*, 1999) combination. The precipitation data along with radiation data are key input parameters (see Table 12.2) to LDAS. Inspection of Figure 12.2 helps illustrate the relationship between precipitation and soil moisture that is influenced by other LDAS output parameters (Table 12.2) such as soil properties, LAI and radiation supply.

Figure 12.3 shows precipitation data over North America available for use as forcing data for LDAS. Comparisons between the precipitation plots show the large variation in estimates. The DAO GEOS and NOAA NCEP-EDAS data are the primary baseline forcing fields and provide spatially and temporally contiguous data. In addition, there are other observational-based datasets that may offer an improved precipitation dataset. The U.S. Naval Research Laboratory (NRL) provided near-real time satellite precipitation data from

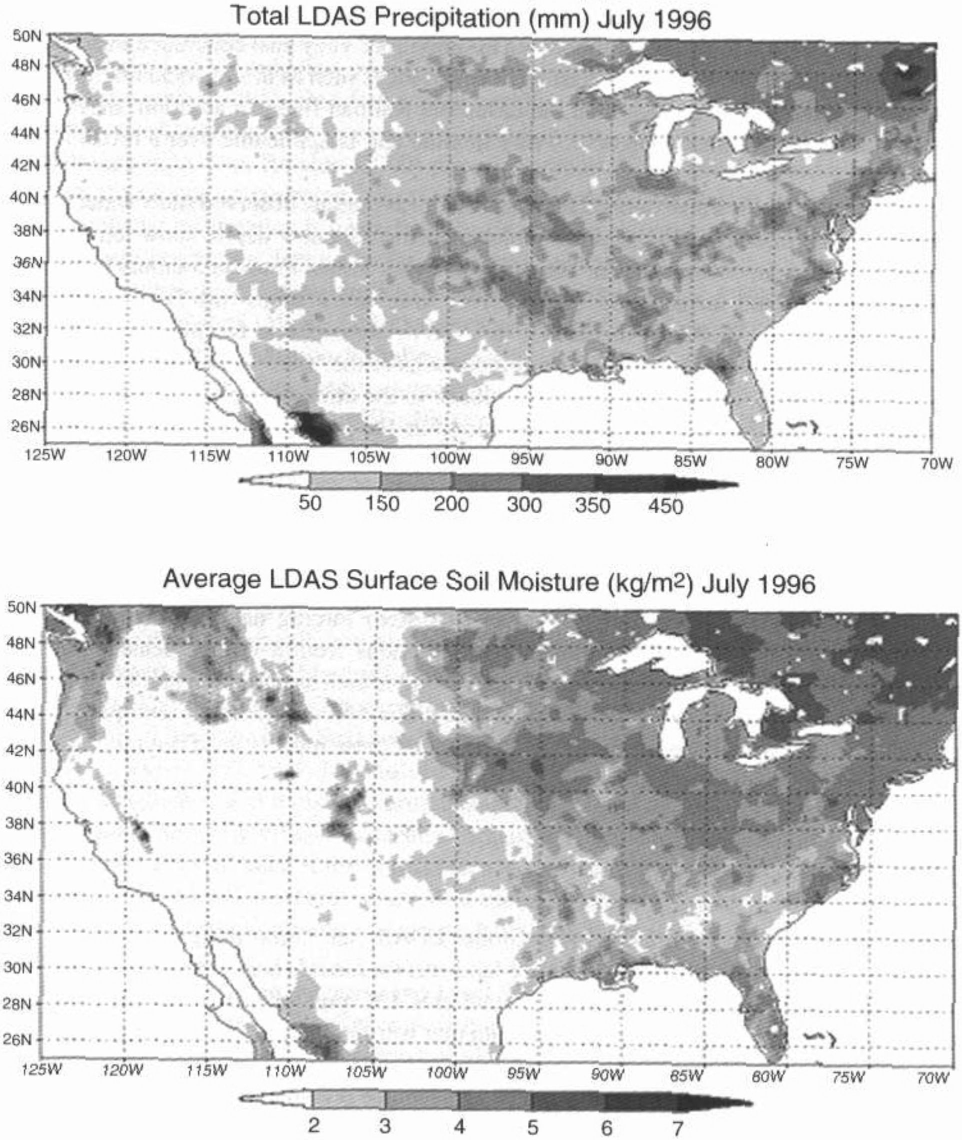

Figure 12.2 *A continental United States 1/8° LDAS merged Eta model, NEXRAD Doppler radar and gauge precipitation field (a) and soil moisture field (b) (courtesy B. Cosgrove)*

both infrared data and microwave sensors (Turk *et al.*, 2000). The microwave data are from TRMM and SSM/I satellites. In addition, Figure 12.3 shows other precipitation data sources which include the University of Arizona, Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Hsu *et al.*, 1997) and the Higgins interpolated gauge data from NOAA Cooperative Institute for Research in Environmental Sciences (CIRES) Climate Diagnostic Center (CDC). Since much of this

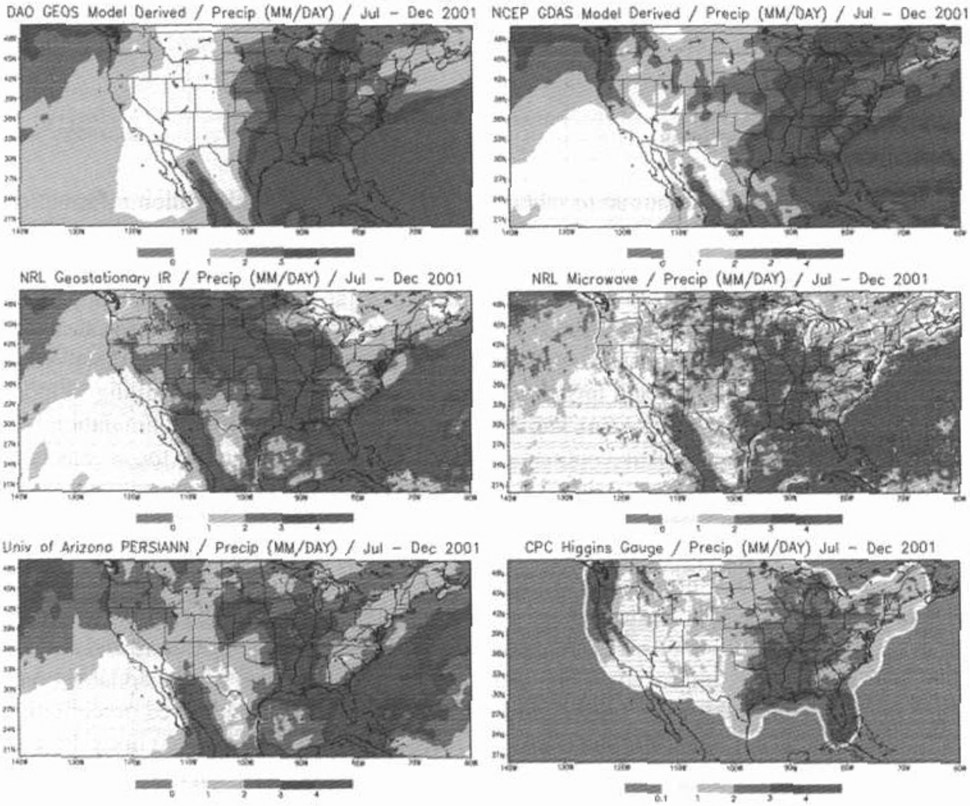

Figure 12.3 *Comparison of partial North American precipitation estimates for July–December 2001. Precipitation is a key forcing parameter for LDAS modelling and data assimilation (courtesy J. Gottschalck)*

data is spatially and temporally limited, LDAS also provides options to merge the sparse observational data with the baseline modelled data.

Plate 12 illustrates SMMR-derived surface soil moisture used to constrain a land surface model prediction using a one-dimensional extended Kalman filter (Walker and Houser, 2001). The SMMR has similar frequencies to the recently launched AMSR sensor on the EOS Aquas satellite. The plots in the top row show SMMR-derived soil moisture with LDAS forcing (input) data of snow depth and precipitation. The middle row shows LDAS-simulated soil moisture at the surface (left), root zone (centre) and over the soil layer profile (right). The assimilated surface moisture plots (bottom row) exhibit a soil moisture increase from the modelled data, especially in the northern Great Plains and south central US. Inspection of the Plate demonstrates how sparse observational data (upper-left) may be used in conjunction with more temporally and spatially covered model data. Approximately 15% of the area is observational data used in data assimilation with the contiguous modelling data. In the Kalman filtering technique the error covariances are used to provide an assimilated output that differs from both the observational and modelled data (Plate 12). This is the first known use of actual satellite-derived soil moisture within an assimilation framework.

12.5 Future Directions

The LDAS projects will continue to incorporate new model and observation information to improve land surface knowledge. The fourth LDAS model to be implemented soon is the Catchment Land Surface Model (CLSM) (Koster *et al.*, 1998). The catchment model signifies a major advance for the following two reasons. First, the modelling framework will result in improved runoff prediction, and consequently, more realistic catchment-scale water balance. Second, the downward slope movement of moisture within the watershed will yield sub-catchment-scale surface and unsaturated-zone moisture variations, resulting in more realistic intra-catchment surface flux variation prediction. Improved runoff simulation will ultimately yield a more realistic continental stream flow from the land to the oceans, and similarly, the within-catchment surface flux variations will result in more representative catchment-average exchanges with the atmosphere.

The significant increase in satellite observations is providing a global supply of atmosphere and land surface information available to improve land surface simulation and data assimilation. For example, higher spatial resolution data (to 1 km) will become available from MODIS and will include leaf area index (LAI), land surface temperature and surface albedo products. Also, improved global land cover datasets (to 1 km) will be updated every three months. New microwave sensors such as AMSR will permit improved precipitation estimation critical to LSM water and energy budgets. Moreover, the planned Global Precipitation Mission (GPM) will involve a satellite constellation that will enhance precipitation estimates to 3 hours temporally and to 4 km horizontal and 250 m vertical global cells. GPM should greatly improve land surface water and energy budget data, and the improved accuracy will benefit data assimilation.

Enhanced modelling and observational data with expanded computer power will improve data assimilation in hydrological sciences. Previous developments of data assimilation in hydrological applications have lagged behind in comparison to atmospheric and ocean applications. However, the wealth of emerging Earth science data coupled with hydrological land surface model physical improvements and computer capability will permit data assimilation to be more routinely implemented. Massively parallel computing techniques such as those developed by the Land Information System will support the near real-time data assimilation throughput demands. In this mode, users will have a web-based interface designed to facilitate broad and efficient information use for land, water and energy data assimilation.

Acknowledgements

Special thanks to Brian Cosgrove, Jon Gottschalck, Rolf Reichle, Jonathan Triggs and Matt Rodell for their inputs and paper review. We also thank Rolf Reichle, Chaojiao Sun and Jeffery Walker for inputs on data assimilation. We express gratitude to other members of the LDAS team for their support: Christa Peters-Lidard, Jared Entin, Jon Radakovich, Mike

Bosilovich, Urszula Jambor, Kristi Arsenault, Jesse Meng, Aaron Berg, Guiling Wang, Ken Mitchell, John Schaake, Eric Wood, Dennis Lettenmaier, Alan Robock and Jeff Basara.

References

Avissar, R. and Pielke, R., 1989, A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology, *Monthly Weather Review*, **117**, 2113–2136.

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

Beven, K. and Kirkby, M., 1979, A physically-based variable contributing area model of basin hydrology, *Hydrological Sciences Journal*, **24**, 43–69.

Birkett, C.M., 1998, Contribution of the TOPEX NASA radar altimeter to the global monitoring of large rivers and wetlands, *Water Resources Research*, **34**, 1223–1239.

Bonan, G.B., 1998, The land surface climatology of the NCAR land surface model coupled to the NCAR community climate mode, *Journal of Climate*, **11**, 1307–1326.

Bouttier, F., Mahfouf, J.-F. and Noilhan, J., 1993, Sequential assimilation of soil moisture from atmospheric low-level parameters. Part I: sensitivity and calibration studies, *Journal of Applied Meteorology*, **32**, 1335–1351.

Callies, U., Rhodin, A. and Eppel, D.P., 1998, A case study on variational soil moisture analysis from atmospheric observations, *Journal of Hydrology*, **212–213**, 95–108.

Calvet, J.-C., Noilhan, J. and Bessemoulin, P., 1998, Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: a feasibility study based on field measurements, *Journal of Applied Meteorology*, **37**, 371–386.

Castelli, F., Entekhabi, D. and Caporali, E., 1999, Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance, *Water Resources Research*, **35**, 3115–3125.

Cline, D.W., Bales, R.C. and Dozier, J., 1998, Estimating the spatial distribution of snow in mountain basins using remote sensing and energy balance modeling, *Water Resources Research*, **34**, 1275–1285.

Cohn, S.E., Da Silva, A., Guo, J., Sienkiewicz, M. and Lamich, D., 1998, Assessing the effects of data selection with the DAO physical-space statistical analysis system, *Monthly Weather Review*, **126**, 2913–2926.

Crow, W.T. and Wood, E.F., 2003, The assimilation of remotely sensed soil brightness temperature imagery into a land-surface model using Ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97, *Advances in Water Resources*, **26**, 137–159.

Daley, R., 1991, *Atmospheric Data Analysis* (New York: Cambridge University Press).

Derber, J.C., Parrish, D.F. and Lord, S.J., 1991, The new global operational analysis system at the National Meteorological Center, *Weather Forecasting*, **6**, 538–547.

Dickinson, R.E., Henderson-Sellers, A. and Kennedy, P.J., 1993, Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as coupled to the NCAR community climate model, *NCAR Technical Note*, NCAR/TN-387+STR, National Center for Atmospheric Research (NCAR).

Eley, J., 1992, Summary of Workshop, Soil Moisture Modeling, *Proceedings of the NHRC Workshop* held March 9–10, 1992, *NHRI Symposium No. 9 Proceedings, Canada.*

Entekhabi, D., Nakamura, H. and Njoku, E.G., 1994, Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations, *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 438–448.

Errico, R.M., 1999, Workshop on assimilation of satellite data, *Bulletin of the American Meteorological Society*, **80**, 463–471.

Famiglietti, J.S. and Wood, E.F., 1994, Multi-scale modeling of spatially-variable water and energy balance processes, *Water Resources Research*, **30**, 3061–3078.

Galantowicz, J.F., Entekhabi, D. and Njoku, E.G., 1999, Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness, *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 1860.

Gornitz, V., Rosenzweig, C. and Hillel, D., 1997, Effects of anthropogenic intervention in the land hydrologic cycle on global sea level rise, *Global Planetary Change*, **14**, 147–161.

Hall, F.G., Meeson, B., Los, S., Steyaert, L., de Colstoun, E. and Landis, D., 2002, ISLSCP Initiative II, DVD/CD-ROM, NASA/Goddard Space Flight Center.

Henderson-Sellers, A., Yang Z.-L. and Dickinson, R.E., 1993, The project for intercomparison of land-surface parameterization schemes, *Bulletin of the American Meteorological Society*, **74**, 1335–1349.

Houser, P., Gupta, H., Shuttleworth, W.J. and Famiglietti, J.S., 2001, Multiobjective calibration and sensitivity of a distributed land surface water and energy balance model, *Journal of Geophysical Research*, **106**, 421–433.

Hsu, K., Gao, X., Sorooshian, S. and Gupta, H.V., 1997, Precipitation estimation from remotely sensed information using artificial neural networks, *Journal of Applied Meteorology*, **36**, 1176–1190.

Hu, Y., Gao, X., Shuttleworth, W.J., Gupta, H., Mahfouf, J. and Viterbo, P., 1999, Soil-moisture nudging experiments with a single-column version of the ECMWF model, *Quarterly Journal of the Royal Meteorological Society*, **125**, 1879–1902.

Jackson, T.J., (1997a) Southern Great Plains, 1997, (SGP97) Hydrology Experiment Plan, United States Department of Agriculture, (http://hydrolab.arsusda.gov/sgp97).

Jackson, T.J., 1997b, Soil moisture estimation using special satellite microwave/imager satellite data over a grassland region, *Water Resources Research*, **33**, 1475–1484.

Justice, C., Vermote, E., Townshend, J.R.G., Defries, R., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J., Riggs, G., Strahler, A., Lucht, W., Myneni, R., Knjazihhin, Y., Running, S., Nemani, R., Wan, Z., Huete, A., van Leeuwen, W., Wolfe, R., Giglio, L., Muller, J-P., Lewis, P. and Barnsley, M., 1998, The moderate resolution imaging spectroradiometer (MODIS) land remote sensing for global change, *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 1228–1249.

Katul, G.G., Wendroth, O., Parlange, M.B., Puente, C.E., Folegatti, M.V. and Nielsen, D.R., 1993, Estimation of in situ hydraulic conductivity function from nonlinear filtering theory, *Water Resources Research*, **29**, 1063–1070.

Koster, R.D. and Milly, P.C.D., 1997, The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models, *Journal of Climate*, **10**, 1578–1591.

Koster, R.D. and Suarez, M.J., 1992, Modeling the land surface boundary in climate models as a composite of independent vegetation stands, *Journal of Geophysical Research*, **97**, 2697–2715.

Koster, R.D., Suarez, M.J., Ducharne, A., Kumar, P. and Stieglitz, M., 1998, A catchment-based approach to modeling land surface processes in a GCM. Part 1: model structure, *Journal of Geophysical Research*, **99**, 14,415–14,428.

Lorenc, A.C., 1995, *Atmospheric Data Assimilation*, Meteorological Office Forecasting Research Division Scientific Paper No. 34, UK Meteorological Office.

Ma, X.-L., Wan, Z., Moeller, C.C., Menzel, W.P. and Gumley, L.E., 2002, Simultaneous retrieval of atmospheric profiles, land-surface temperature and emissivity from moderate resolution imaging spectroradiometer thermal infrared data: extension of a two-step physical algorithm, *Applied Optics*, **41**, 909–924.

Mahfouf, J.-F., 1991, Analysis of soil moisture from near-surface parameters: a feasibility study, *Journal of Applied Meteorology*, **30**, 1534–1547.

Margulis, S.A., McLaughlin, D., Entekhabi, D. and Dunne, S., 2002, Land data assimilation and soil moisture estimation using measurements from the Southern Great Plains 1997 field experiment, *Water Resources Research* (in press).

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P. and Nussen, B., 2002, A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, *Journal of Climate*, **15**, 3237–3251.

McLaughlin, D., 1995, Recent developments in hydrologic data assimilation. *Reviews of Geophysics*, Supplement, 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, July 1986). IHAS Publ. no. 158, pp. 201–211.

Mitchell, K., 2002, The Community *NOAH Land-Surface Model (LSM) User's Guide*, Version 2.5.2. (ftp://ftp.ncep.noaa.gov/pub/gcp/ldas/noahlsm/ver_2.5.2/NOAH_LSM_USERGUIDE_2.5.2.doc).

Mitchell, K., Houser, P., Wood, E., Schaake, J., Tarply, D., Lettenmaier, D., Higgins, W., Marshall, C., Lohman, D., Ek, M., Cosgrove, B., Entin J., Duan, Q., Pinker, R., Robuck, A., Habelts, F. and Vinnikov, K., 1999, *GCIP Land Data Assimilation System (LDAS) project* now underway, GEWEX News, 9(4),3–6, GEWEX Project Office, Silver Spring, MD.

Parlange, M.B., Katul, G.G., Folegatti, M.V. and Nielsen, D.R., 1993, Evaporation and the field scale soil water diffusivity function, *Water Resources Research*, **29**, 1279–1286.

Peters-Lidard, C.O., Pan, F. and Wood, E.F., 2001, A re-examination of modeled and measured soil moisture spatial variability and its implications for land surface modeling, *Advances in Water Resources*, **24**, 1069–1083.

Peters-Lidard, C.O., Zion, M.S. and Wood, E.F., 1997, A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes, *Journal of Geophysical Research*, **102**, 4304–4324.

Reichle, R.H., 2000, Variational assimilation of remote sensing data for land surface hydrologic applications, Ph.D. dissertation, Massachusetts Institute of Technology.

Reichle, R.H., McLaughlin, D.B. and Entekhabi, D., 2002, Hydrologic data assimilation with the Ensemble Kalman filter, *Monthly Weather Review*, **130**, 103–114.

Reynolds, C.A., Jackson, T.J. and Rawls, W.J., 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 Transactions*, AGU, 80.

Rhodin, A., Kucharski, F., Callies, U., Eppel, D.P. and Wergen, W., 1999, Variational analysis of effective soil moisture from screen-level atmospheric parameters: application to a short-range weather forecast model, *Quarterly Journal of the Royal Meteorological Society*, **125**, 2427–2448.

Rodell, M. and Famiglietti, J.S., 1999, Detectability of variations in continental water storage from satellite observations of the time-variable gravity field, *Water Resources Research*, **35**, 2705–2723.

Rodell, M., Houser, P.R., Jambor, U., Gotschalck, J., Meng, C.-J., Arsenault, K., Cosgrove B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J., Toll, D. and Mitchell, K., 2004, The Global Land Data Assimilation System (GLDAS), *Bulletin of the American Meteorological Society*, in press.

Rosenthal, W. and Dozier, J., 1996, Automated mapping of montane snow cover at subpixel resolution from the Landsat Thematic Mapper, *Water Resources Research*, **32**, 115–130.

Sellers, P.J., Mintz, Y. and Dalcher, A., 1986, A simple biosphere model (SiB) for use within general circulation models, *Journal of Atmospheric Science*, **43**, 505–531.

Sipple, S.J., Hamilton, S.K., Melak, J.M. and Choudhury, B.J., 1994, Determination of inundation area in the Amazon river flood plain using SMMR 37 GHz polarization difference, *Remote Sensing Environment*, **48**, 70–76.

Sun, C., Walker, J.P. and Houser, P.R., 2002, Snow assimilation in a catchment-based land surface model using the extended Kalman filter. Extended abstract and poster. *Symposium on Observations, Data Assimilation and Probabilistic Prediction, AMS 82nd Annual Meeting, Orlando, Florida.*

Tait, A.B., Hall, D.K., Foster, J.L. and Armstrong, R.L., 2000, Utilizing multiple datasets for snow-cover mapping, *Remote Sensing Environment*, **72**, 111–126.

Toll, D.L., Entin, J. and Houser, P., 2001, Land surface heterogeneity on surface energy and water fluxes, in M. Owe and G. D'Urso (eds), *Remote Sensing for Agriculture, Ecosystems, and Hydrology III, Proceedings of SPIE, Toulouse France*, 17–19 September 2002, Volume 4542, 267–271.

Toll, D.L., Shirey, D. and Kimes, D.S., 1997, NOAA AVHRR surface albedo algorithm development, *International Journal of Remote Sensing*, **18**, 3761–3796.

Turk, F.J., Rohaly, G., Hawkins, J.D., Smith, E.A., Grose, A., Marzano, F.S., Mugnai, A. and Levizzani, V., 2000, Analysis and assimilation of rainfall from blended SSM/I, TRMM and geostationary satellite data, *AMS 10th Conference on Satellite Meteorology and Ocean., Long Beach*, CA, 9–14 January, 66–69.

Vukovich, F.M., Wayland, R. and Toll, D.L., 1997, The surface heat flux as a function of ground cover for climate models, *Monthly Weather Review*, **125**, 573–586.

Wahr, J., Molenaar, M. and Bryan, F., 1998, Time-variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE, *Journal of Geophysical Research*, **103**, 30,205–30,229.

Walker, J.P. and Houser, P., 2001, A method for initializing soil moisture in a global climate model: assimilation of near-surface soil moisture observations, *Journal of Geophysical Research*, **106**, 11,761–11,774.

Wunsch, C., 1996, *The Ocean Circulation Inverse Problem* (New York: Cambridge University Press).