The Land Information System

P. Houser, C. Peters-Lidard, S. Kumar, and Y. Tian, J. Geiger, S. Olden, L. Lighty, J. Sheffield, E. Wood, K. Mitchell, P. Dirmeyer, B. Doty, J. Adams

***Abstract*—A global Land Information System (LIS; http://lis.gsfc.nasa.gov) has been developed that is capable of modeling land-atmosphere interactions at spatial resolutions down to 1km. LIS consists of an ensemble of land surface models (e.g., CLM, Noah, VIC) run offline using satellite-based precipitation, radiation and surface parameters, in addition to model-derived surface meteorology. Satellite-based surface parameters include AVHRR-based or MODIS-based land cover and Leaf Area Index (LAI), and MODIS-based albedo and emissivity. The high spatial resolution of LIS, enabled by the use of high performance computing and communications technologies, is capable of resolving mesoscale features, including urban areas, lakes, and agricultural fields. We will present results demonstrating LIS applied at ¼ degree, 5km and 1km resolutions. Several demonstration case studies conducted with LIS demonstrate that using LIS has significant advantages for predicting land surface temperatures and surface water and energy fluxes.**

***Index Terms*—Land, Modeling, Software, Remote Sensing**

I. INTRODUCTION

LAND surface water, energy, and carbon conditions have profound influences on the overall behavior of the Earth's climate system. A better understanding of these conditions enables improved use of natural resources, mitigation of environmental hazards, and knowledge of climate change scenerios. To this end, much research has been devoted to the integration of land surface simulation, observation, and analysis methods to accurately determine land surface energy and moisture states. Examples of such systems include the 1/8 degree North American Land Data Assimilation System [1] and the 1/4 degree Global Land Data Assimilation System (GLDAS) [2]. Computational limitations in hardware and software have impeded the development and application of such systems at higher spatial resolutions. The Land Information System (LIS) is a software system that takes advantage of the technological improvements in computing and environmental monitoring tools to enable a global high resolution land modeling. LIS is also designed to directly ingest the vast array of high resolution observations such as those available from the next generation NASA earth science instruments. The ability to operate at the same fine spatial scales of the atmospheric boundary layer and cloud models also helps in improving water and energy cycle prediction capabilities. In addition to providing a land surface modeling infrastructure, the portable, interoperable design of LIS enables it to be a valuable research tool for land surface researchers and other interdisciplinary scientists.

The LIS infrastructure consists of several land surface models (LSMs), forced with a combination of observationally-based precipitation and radiation with downscaled model based meteorological inputs. LIS makes use of the state-of-the-art scalable high performance computing technologies to achieve ever-higher spatial resolutions and temporal durations. To provide efficient management, storage, and high throughput data access in simulations, LIS also employs a number of generic tools that manage this data. LIS also provides intuitive web-based interfaces to enable access for a wide variety of users.

Many existing earth science applications lack the ability to interoperate with other earth system applications. As a result, the cost of adding new functionalities and adapting the existing systems to function with other applications may be prohibitively high. LIS attempts to achieve the desired interoperability by applying advanced software engineering

Manuscript received May 23, 2005. This work was supported by the NASA Computational Technologies Program.

P. Houser is with the George Mason University, Fairfax, VA 22030 USA (phone: 301-595-7000; fax: 301-595-9793; e-mail: phouser@gmu.edu).

C. Peters-Lidard is with the National Aeronautics and Space Administration, Greenbelt, MD 20771 USA. (e-mail: Christa.Peters @nasa.gov).

S. Kumar is with the University of Maryland, Baltimore County, Baltimore MD 21250 USA. (e-mail: sujay@hsb.gsfc.nasa.gov).

Y. Tian is with the University of Maryland, Baltimore County, Baltimore MD 21250 USA. (e-mail: yudong@hsb.gsfc.nasa.gov).

J. Geiger is with the National Aeronautics and Space Administration, Greenbelt, MD 20771 USA. (e-mail: James.V.Geiger.1@gsfc.nasa.gov).

S. Olden is with the National Aeronautics and Space Administration, Greenbelt, MD 20771 USA. (e-mail: Susan.P.Olden.1@gsfc.nasa.gov).

L. Lighty is with the National Aeronautics and Space Administration, Greenbelt, MD 20771 USA. (e-mail: Luther.Lighty.1@gsfc.nasa.gov).

J. Sheffield is with the Princeton University, Princeton, New Jersey 08544 USA. (e-mail: justin@princeton.edu).

E. Wood is with the Princeton University, Princeton, New Jersey 08544 USA. (e-mail: efwood@princeton.edu).

K. Mitchell is with the National Oceanic and Atmospheric Administration, Camp Springs, MD 20746 USA. (e-mail: Kenneth.Mitchell@noaa.gov).

P. Dirmeyer is with the Center for Ocean-Land-Atmosphere Studies, Calverton, MD 20705 USA. (e-mail: dirmeyer@cola.iges.org).

B. Doty is with the Center for Ocean-Land-Atmosphere Studies, Calverton, MD 20705 USA. (e-mail: doty@cola.iges.org).

J. Adams is with the Center for Ocean-Land-Atmosphere Studies, Calverton, MD 20705 USA. (e-mail: jma@cola.iges.org).

concepts. The system is designed as an object oriented framework that can be shared and reused by scientists and practitioners. The use of object oriented principles help in designing LIS to be flexible and extensible, enabling rapid prototyping of new applications. In addition to providing an infrastructure to support land surface research and applications activities, LIS has also adopted other earth system modeling standards and conventions, such as the Earth System Modeling Framework (ESMF) [3] and Assistance for Land Modeling Activities (ALMA) [4].

The following sections describe the land observation, modeling and assimilation tools in LIS, the design of the LIS system, and some demonstrations of LIS in simulating land surface processes.

II. BACKGROUND

A. *Land Surface Observations*

The use of observation-driven land models and data assimilation is a fundamental principle of the LIS framework. As such, we have identified the current and future observational systems that may guide the development of improved parameters, forcing, validation, and data assimilation constraints for LIS. A huge volume of land surface observations may be operationally sensed from space, including surface temperatures, vegetation conditions, snow, albedo, longwave and solar radiation, precipitation, surface moisture, freeze/thaw state, runoff, total water storage, and elevation, among others. LIS must be able to compile and assimilate those highly dynamic remotely-sensed observations of the land surface that previous research suggests will provide information critical for land surface research synthesis. Many of the available remotely-sensed land surface observations are summarized in Table 1.

B. *Land Surface Observations*

Recent advances in understanding of soil-water dynamics, plant physiology, micrometeorology, and the hydrology that control biosphere-atmosphere interactions have spurred the development of Land Surface Models (LSMs), whose aim is to represent simply, yet realistically, the transfer of mass, energy, and momentum between a vegetated surface and the atmosphere [5]. LSM predictions are regular in time and space, but these predictions are influenced by errors in model structure, input variables, parameters, and inadequate treatment of sub-grid scale spatial variability. Consequently, LSM predictions are significantly improved through observation constraints. LIS has adopted a “ensemble physics” land surface modeling philosophy to enable straightforward collaboration with operational weather, climate, and decision support partners. It should be noted that LIS uses model-independent software frameworks (e.g., ESMF, ALMA, etc.) and data assimilation algorithms to enable its application in a wide range of end-uses.

TABLE I

LIST OF LAND REMOTELY-SENSED SURFACE PROPERTIES

Class	Observation	Example Platform	Temporal	Spatial
Land Parameters	Leaf area and greenness	AVHRR, MODIS, NPOESS	weekly	1km
	Albedo	MODIS, NPOESS	weekly	1km
	Emissivity	MODIS, NPOESS	weekly	1km
	Vegetation structure	ICESAT, ESSP lidar mission	weekly-monthly	100m
	Topography	GTOPO30, SRTM	episodic	30m–1km
Land Forcings	Wind profile			
	Air Humidity and temperature	TOVS, GOES, AVHRR, MODIS, AMSR	hourly-weekly	5 km
	Near- surface radiation	GOES, MODIS, CERES, ERBS, etc.	hourly-weekly	1km
	Precipitation	TRMM, GPM, SSMI, GEO-IR, etc.	hourly-monthly	10km
Land States	Temperature	IR-GEO, MODIS, AVHRR, TOVS	hourly-monthly	10m-4km
	Thermal anomalies	AVHRR, MODIS, TRMM	daily-weekly	250m–1km
	Snow cover and water	SSMI, TM, MODIS, AMSR, AVHRR, etc.	weekly-monthly	1km
	Freeze/thaw	Quickscat, HYDROS, IceSAT, CryoSAT	weekly	3km
	Total water storage	GRACE	monthly	1000km
	Soil moisture	SSMI, AMSR, HYDROS, SMOS, etc.	3-30 day	10-100 km
Land Fluxes	Evapotranspiration	MODIS, GOES	hourly-weekly	10m-4km
	Solar radiation	MODIS, GOES, CERES, ERBS	hourly-monthly	
	Longwave radiation	MODIS, GOES	hourly-monthly	10m-4km
	Sensible heat flux	MODIS, ASTER, GOES	hourly-monthly	10m-4km

C. *Land Surface Data Assimilation*

Charney [6] 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 between the fields. This concept has evolved into a family of techniques known as data assimilation. In essence, hydrologic data assimilation aims to utilize both our hydrologic process knowledge, as embodied in a hydrologic model, and information that can be gained from observations. Both model predictions and observations are imperfect, and we wish to use both synergistically to obtain a more accurate result. Moreover, both contain different kinds of information that, when used together, provide an accuracy level that cannot be obtained individually. The past few years have seen a sharp increase in the volume of scientific literature devoted to land surface data assimilation techniques, with the most promising techniques being the Extended Kalman filter (EKF) and the Ensemble Kalman filter (EnKF). We have implemented both the EKF and EnKF algorithms in LIS as routine and systematic procedures for assimilating continuously available land surface data products.

III. LIS COMPONENTS

The LIS software system consists of a four components:

(1) The LIS core that integrates the use of LSMs, high performance computing, use of various sources of data, and the domains of execution;

(2) Several community Land Surface Models;

(3) Data servers that provide an interface to heterogeneous data and,

(4) Visualization tools that provide interactive access to the LIS products.

The LIS core is primarily an infrastructure that operates multiple one-dimensional LSMs providing it the appropriate inputs. These models are typically run in an uncoupled manner, where the boundary conditions for the atmosphere are provided either from meteorological forecast model outputs or from various satellite and ground-based observational systems. The input data, including the initial conditions and model parameters derived from topography, vegetation, and soil coverage describing the land surface states are processed and supplied to the LSMs. The models in turn produce optimal output fields of land surface states and fluxes.

The LSMs in LIS use and produce numerous data for analysis and modeling purposes in different data formats and resolutions. LIS provides a number of generic data management utilities to ensure a seamless, efficient access and use of data. The heterogeneity of diastases are encapsulated by the use of data servers based on Grid Analysis and Display System (GrADS) [7])-DODS (Distributed Oceanographic Data System) (GDS [8]). The client-server model of data serving provided by the GDS server and GrADS clients allows distributed data sets in various formats to be accessed dynamically and transparently. LIS also provides tools to retrieve input data from various sources, interpolate, reproject, and subset them to the required domain and resolution. The LIS datasets include satellite and remote-sensing land surface data, and products such as temperature, Normalized Difference Vegetation Index (DVI) greenness, Leaf Area Index (LAI), surface albedo, and emissivity from Advanced Very High Resolution Radiometer (AVHRR), Geostationary Operational Environmental Satellite (GOES), Moderate Resolution Imaging Spectroradiometer (MODIS) and various airborne sensors.

The ability to spatially resolve 1km resolutions enables the direct use of the high resolution data produced by satellite technologies such as the MODIS instrument on Terra and Aqua satellites in LIS. Land surface modeling at high spatial resolutions such as 1km presents considerable computational challenges. Typically, the land surface is modeled by dividing it into two -dimensional regions or gridcells (for example, cells of size lkm × 1km globally would lead to approximately 5×108 gridcells). Assuming approximately 15 milliseconds for each day of land surface model execution on a single gridcell with 15 minute timesteps, it can be estimated that to conduct a day's simulation at 1km on a single processor would require approximately 3 months. Further, as the number of grid points increases with resolution, the memory and disk storage requirements also increase significantly. The output requirements for the global 1km LIS simulation, is on the order of 1 terabyte per day of simulation.

Due to the significant computational requirements, the use of scalable computing technologies is critically important for LIS. Land surface processes have rather weak horizontal coupling on short time and large space scales. LIS exploits this inherent parallelism to achieve highly efficient scaling across massively parallel computational resources.

To adequately address the computational requirements at different resolutions and domains, LIS provides a number of high performance operating modes. For resolutions and domains where the available memory is not a limiting constraint, LIS employs a simple master slave paradigm, with a master processor performing the initializations and domain decompositions. Subsequently, the slave nodes perform computations on the decomposed domain. The temporal synchronizations and output aggregations are conducted by the master processor during a simulation at specified intervals. However, this mode becomes intractable for large domains at high spatial resolution. For such cases, LIS makes use of the GDS data servers to handle I/O. GDS provides capabilities for a client to dynamically retrieve subsets of data on a global domain. The domain decomposition is done as before by a master processor, and the initializations and subsequent computations are performed by the slave nodes. These slave nodes request the required subset of data from the GDS data server as the computations proceed. The data is retrieved from the server by each slave node using a GrADS client. The GDS server per-forms output aggregations subsequent to the completion of computations by the slave nodes.

The LIS user interface allows the interactive and flexible use of the LIS products by end users. The visualization capabilities in LIS are built based on a multi-tier client-server system architecture. GDS data servers are employed to handle various types of client requests. This enables the user to access the LIS user interface from virtually any program. LIS uses the capabilities of the GDS server to handle DODS client requests, and a web server to handle others. When a DODS client is used, the user can also perform data manipulations such as subsetting and dynamic generation of images. LIS also provides an alternate method to visualize the LIS data using the Live Access Server (LAS [9]). The LAS server allows users to search a data catalog, visualize data interactively, request subsets, view metadata, and a multitude of other functions.

IV. LIS DEMONSTRATION

Here we summarize a few examples of the scientific studies enabled by LIS. LIS's interoperable features enable a system with a growing suite of LSMs, forcing schemes, different types of parameter data, and domains. The extensible interfaces for these different components help in rapid prototyping and inclusion of new components into LIS. The

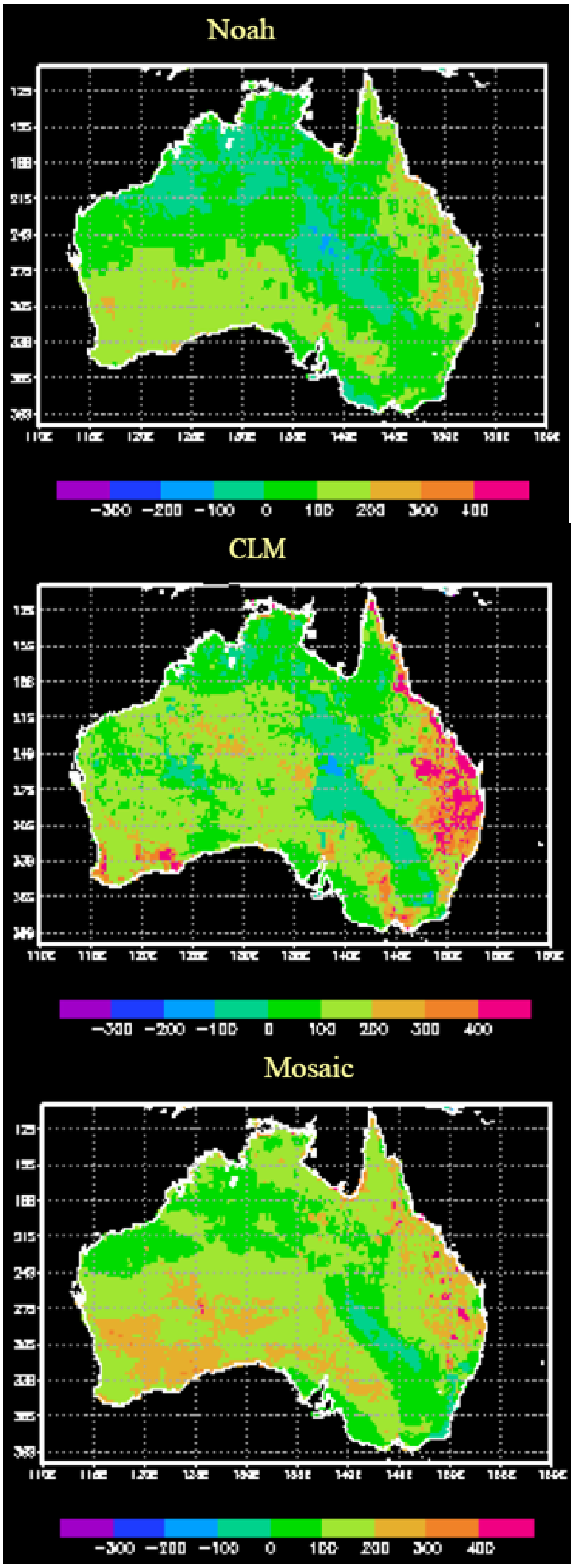


Fig. 1. Sensible heat fluxes (W/m2) produced by Noah, CLM, and Mosaic over Australia, at 4GMT for October 09, 2001.

configurable features in LIS also allow the user to select different components to build the application instead of having to use a monolithic system. For example, although LIS includes a number of land surface models and forcing schemes, the user can build the LIS application with the only required LSM and data scheme. These configurable features improve technology transfer and continued innovative efforts in modeling.

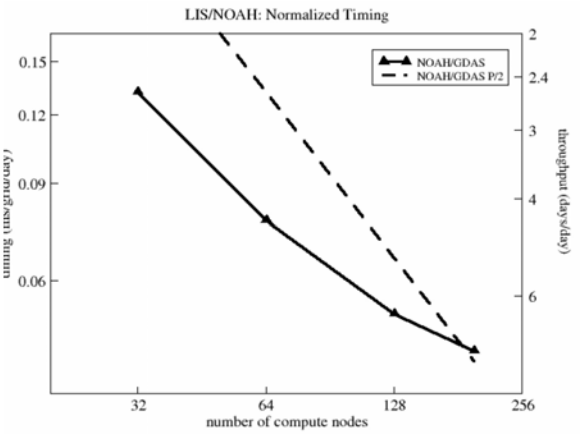
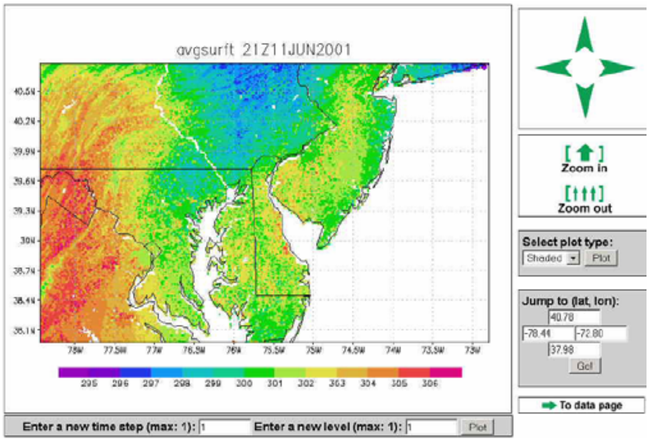


Fig. 2. Normalized timing and simulation throughput for a global 1km LIS simulation with Noah using the GDAS forcing compared with the theoretical P/2 estimate.

There have been numerous studies for intercomparing LSMs, such as the Global Soil Wetness Project (GSWP [10]) and Project for Intercomparison of Land-surface Parameterization Schemes (PILPS [11]). One of the main challenges to performing such intercomparions is the configuration of each model for the specific set of data and domain. LIS provides an ideal platform to perform a multi-model land surface analysis and sensitivity studies of models to different parameters. As a demonstration, land surface simulations were carried out over Australia using three different models (Noah, CLM, and Mosaic). Sensible heat fluxes on October 9, 2001 at 4GMT predicted by the three models are shown in Figure 1. It can be observed that there are significant differences between the model predictions. Mosaic model in this case predicts higher values of sensible heat fluxes compared to that of CLM and Noah.

One of the unique capabilities of LIS is the infrastructure it provides to support global land-atmosphere interactions at spatial resolutions down to 1km and finer. The high spatial resolution of LIS makes it capable of resolving spatial features such as urban areas that could not be resolved at coarser resolutions. Our work has shown that the higher-resolution LIS predictions are more representative of the station data than coarser predictions.

The high performance operating modes in LIS not only provides the ability to perform massive simulations, but also the environment to provide high throughput for the simulations. To demonstrate the improved performance with an increase in the number of processors, a global 1km simulation was performed on a custom built, 200 node Linux cluster at GSFC. The simulation was carried out using the Noah LSM and the GDAS forcing scheme. To manage the huge data throughput at 1km, the use of GDS servers to dynamically subset the data was used. Figure 2 shows the scaling of performance with the number of processors. The performance is compared with a theoretical P/2 estimate,



Select plot type: Shaded

Jump to (lat, lon): 40.78 -78.44 -72.80 37.98

Enter a new time step (max: 1): 1 Enter a new level (max: 1): 1

Fig. 3. A snapshot of the surface temperature datafield for a region around Washington D.C. retrieved dynamically by the Land Explorer from a global 1km output of a Noah LSM simulation.

which is based on the assumption that the performance improves by a factor of P/2 when P processors are employed. It can be seen that the LIS performance scales better than the P/2 estimate.

The visualization tools in LIS allow for interactive use of both input and output data. The Land Explorer, which is based on the GDS server, allows user to interactively visualize and explore data at all resolutions. The LAS interface allows more advanced features such as performing data analysis, interactive subsetting etc. A snapshot of surface temperature for a region around Washington D.C. retrieved dynamically using the Land Explorer from a simulation using Noah model is shown in Figure 3.

REFERENCES

[1] Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F. Schaake, J. C., Robock, A., Cosgrove, B., Sheffield, J., Duan, Q., Luo, L., Higgins, W. R., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, 27 A. A. (2004). The Multi-institution North American Land Data Assimilation system (NLDAS): Utilization of multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, 109.

[2] Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng., C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85(3):381–394.

[3] Hill, C., DeLuca, C., Balaji, V., Suarez, M., and da Silva, A. (2004). The architecture of the earth system modeling framework. *Computing in Science and Engineering*, 6(1).

[4] ALMA (2002). Assistance for land modeling activities, Version 3. http://www.lmd.jussieu.fr/ALMA/.

[5] 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. 28.

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

[7] Doty, B. and Kinter, J. L. (1993). The grid analysis and display system (GrADS): A desktop tool for earth science visualization. In American Geophysical Union 1993 Fall Meeting, pages 6–10, San Francisco, CA..

[8] Wielgosz, J., Doty, B., Gallagher, J., and Holloway, D. (2001). GrADS and DODS. In Seventeenth International Conference on Interactive Information and Processing, Albuquerque, NM.

[9] Hankin, S. C., Callahan, J., and Sirott, J. (2001). The Live Access Server and DODS. In 17th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, Albuquerque, NM..

[10] Dirmeyer, P., Dolman, A., and Sato, N. (1999). The global soil wetness project: A pilot project for global land surface modeling and validation. Bulletin of the American Meteorological Society, 80:851–878..

[11] Henderson-Sellers, A., Irannejad, P., McGuffie, K., and Pitman, A. (2003). Predicting land-surface climates-better skill or moving targets? *Geophysical Research Letters*, 30(14).