Innovations Syst Softw Eng (2007) 3:157–165
DOI 10.1007/s11334-007-0028-x

ORIGINAL PAPER

High-performance Earth system modeling with NASA/GSFC's Land Information System

C. D. Peters-Lidard · P. R. Houser · Y. Tian · S. V. Kumar · J. Geiger · S. Olden · L. Lighty · B. Doty · P. Dirmeyer · J. Adams · K. Mitchell · E. F. Wood · J. Sheffield

Received: 22 September 2006 / Accepted: 20 February 2007 / Published online: 24 July 2007

Abstract The Land Information System software (LIS; http://lis.gsfc.nasa.gov/, 2006) has been developed to support high-performance land surface modeling and data assimilation. LIS integrates parallel and distributed computing technologies with modern land surface modeling capabilities, and establishes a framework for easy interchange of subcomponents, such as land surface physics, input/output conventions, and data assimilation routines. The software includes multiple land surface models that can be run as a multi-model ensemble on global or regional domains with horizontal resolutions ranging from 2.5° to 1 km. The software may execute serially or in parallel on various high-performance computing platforms. In addition, the software has well-defined, standard-conforming interfaces and data structures to interface and interoperate with other Earth system models. Developed with the support of an Earth science technology office (ESTO) computational technologies project round 3 cooperative agreement, LIS has helped advance NASA's Earth–Sun division's software engineering principles and practices, while promoting portability, interoperability, and scalability for Earth system modeling. LIS was selected as a co-winner of NASA's 2005 software of the year award.

Keywords Land surface modeling · Earth system modeling · High-performance computing · Information systems · Object-oriented frameworks · Weather · Climate · Ensemble · Interoperability

C. D. Peters-Lidard (✉) · Y. Tian · S. V. Kumar
Goddard Space Flight Center, Hydrological Sciences Branch, NASA, Code 614.3, Greenbelt, MD 20771, USA
e-mail: christa.peters@nasa.gov
URL: http://lis.gsfc.nasa.gov

P. R. Houser
George Mason University and Center for Research in Environment and Water, Calverton, MD, USA

Y. Tian · S. V. Kumar
Goddard Earth Sciences Technology Center, University of Maryland at Baltimore County, Baltimore, MD, USA

J. Geiger · S. Olden · L. Lighty
Goddard Space Flight Center, Information Systems Division, NASA, Code 580, Greenbelt, MD, USA

B. Doty · P. Dirmeyer · J. Adams
Center for Ocean–Land–Atmosphere Studies, Calverton, MD, USA

K. Mitchell
NCEP Environmental Modeling Center, NOAA/NWS, Camp Springs, MD, USA

E. F. Wood · J. Sheffield
Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

1 Introduction

Observation-driven Earth system modeling supports NASA's mission "to understand and protect our home planet" by leveraging capabilities in software engineering, computational technologies, and remote sensing to maximize the benefits of NASA science and technology for society. Earth system modeling uses quantitative methods to simulate the interactions of the atmosphere, oceans, land, and ice to study the dynamics of the weather and climate system. The Land Information System (LIS; http://lis.gsfc.nasa.gov; [1]) has been developed as an observation-driven land surface component of an Earth system model to advance our understanding of terrestrial water, energy and carbon cycles by incorporating unique NASA Earth observing system

Table 1 Computational resource requirements for LIS at different horizontal spatial resolutions

Horizontal resolution	1/4°	5 km	1 km
Number of land grid points	2.43E+05	5.73E+06	1.44E+08
Output disk space/day (Gb)	1	28	694
Memory (Gb)	3	62	1,561

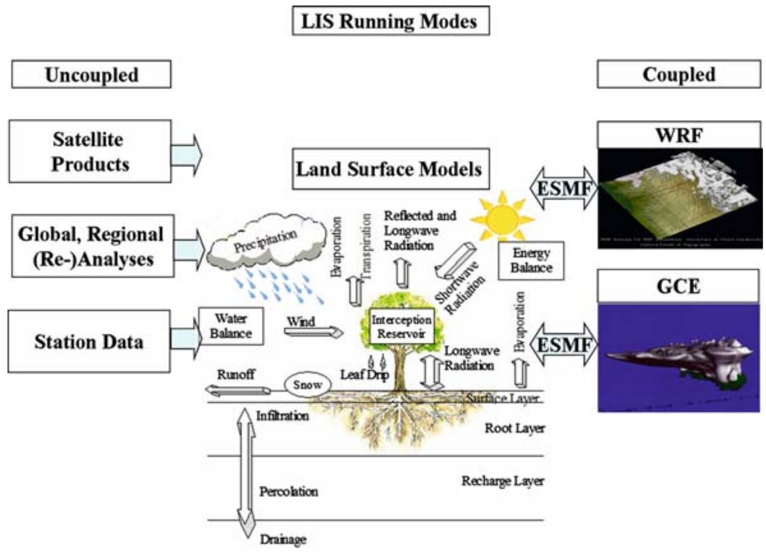

Fig. 1 The two LIS execution modes: uncoupled and coupled. In the uncoupled mode, input meteorological data is obtained from various sources, including in situ weather station data, global and regional model-based analyses and satellite products. In coupled mode, LIS has been coupled using ESMF and ALMA to two atmospheric model components: WRF and GCE

(EOS)-era observations into a near-real-time modeling capability at the 1 km scale of the observations [i.e., Moderate-Resolution Imaging Spectroradiometer (MODIS) instrument-based]—a challenge that was impossible prior to the development of LIS.

Land surface models predict terrestrial water, energy, momentum, and in some cases, bio-geochemical exchange processes by solving the governing equations of the soil–vegetation–snowpack medium. By constraining land surface models with observations, land surface data assimilation improves our ability to understand and predict these processes. The ability to predict terrestrial water, energy and bio-geochemical processes is critical for applications in weather and climate prediction, agricultural forecasting, water resources management, hazard mitigation, and mobility assessment. However, until recently, global land surface modeling at scales down to 1 km was infeasible due to limits in computational and observational resources (Table 1).

LIS supports regional to global land surface modeling, with a focus on land–atmosphere interactions to enable its coupling with, and incorporation into Earth system models. LIS is a high-performance Land Data Assimilation System (LDAS), that unifies and extends the capabilities of its precursors, the global (GLDAS; [2]) and North American (NLDAS; [3]) LDAS. As shown in Fig. 1, LIS incorporates several land surface models, which may be executed in two running modes: (1) uncoupled (or offline) retrospective analysis using observationally based precipitation, radiation and meteorological inputs, and (2) coupled to an atmospheric model component, such as the Goddard Cumulus Ensemble (GCE; [4]) model or the Weather Research and Forecasting (WRF; [5]) model. The coupled mode is used for weather forecasts, as shown in Fig. 1.

This paper describes the LIS software, and details how the LIS software design promotes an innovative approach to high-performance observation-driven Earth system modeling. We show how the LIS computational performance and interoperability considerations enable its use in a variety of high-impact science and application projects. A summary and discussion of ongoing and future work concludes the paper.

2 Software design

The LIS software design is discussed in detail in [1], so only a brief summary will be given here. As shown in Fig. 2, the major components of the LIS include: (1) the LIS core, which is the software that coordinates the use of land surface models (LSMs), various data sources, high-performance computing and execution domains; (2) a number of community LSMs; (3) data servers to provide a common interface to heterogeneous data and handle access requests; and (4) visualization tools to provide interactive access to the LIS products.

LIS unifies and extends the capabilities of the 1/4° GLDAS [2] and the 1/8° NLDAS [3] in a common software

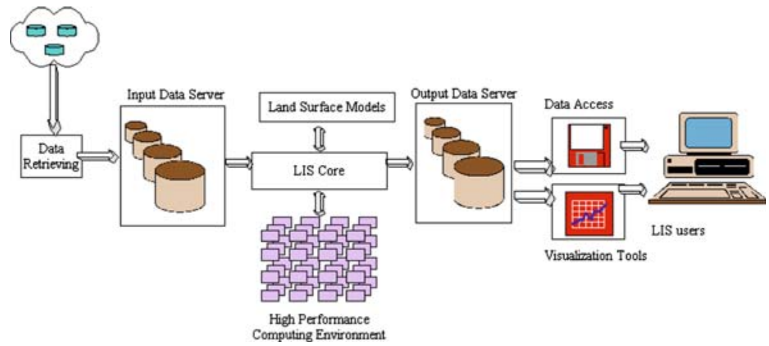

Fig. 2 The LIS system architecture, illustrating the use of input and output data servers as well as the high-performance computing environment

framework capable of ensemble land surface modeling on points, regions or the globe at spatial resolutions from 2° down to 1 km or finer. Similar to other LDASs, the LIS framework acts as a driver for typically one-dimensional (1D) land surface models that predict energy and water states (e.g., snowpack and soil moisture and temperature profiles) and fluxes (e.g., evaporation, sensible heat flux and runoff). The 1 km capability of LIS allows it to take advantage of the latest NASA EOS-era products, such as the MODIS land cover, leaf area index, albedo, emissivity and surface temperature, at their full resolution.

Central to the LIS design is its object-oriented software engineering framework [1] and integrated high-performance computing and communications technologies that enable high-resolution ensemble land surface modeling. This object-oriented design also makes LIS the ideal testbed for ensemble studies of land–atmosphere interactions including the development of techniques for representing land surface heterogeneities, in the spirit of the world climate research programme's global energy and water experiment (GEWEX) activities such as the Project for Intercomparison of Land Surface Parameterization Schemes (PILPS; [6]) and the Global Soil Wetness Project (GSWP; [7,8]). In addition to providing an infrastructure to support land surface research and applications activities, LIS incorporates Earth system modeling standards and conventions that facilitate its use as the land surface component of coupled Earth system models. These standards include the Earth System Modeling Framework (ESMF; [9]) and Assistance for Land Modeling Activities (ALMA; [10]). ESMF is a framework that provides a flexible software infrastructure to foster interoperability, portability and code reuse in climate, numerical weather prediction, data assimilation and other Earth science applications. ALMA was originally developed by the GEWEX Land Atmosphere System Study (GLASS) to support model comparison efforts, and has now been adopted by the European programme for integrated Earth system modeling (PRISM; [11]), an infrastructure project for climate research in Europe.

LIS includes a highly modular system for high-performance modeling functions, whose central component (the LIS core) controls the execution of land surface models, defines the parallel-computing configurations and manages data input/output. The component-based design allows plug-and-play flexibility, which enables scientists to experiment with different input datasets, parameters, and land surface models without modifying the LIS core code.

LIS retrieves relevant atmospheric observational data from various data repositories on the Internet as its input, and creates an archive for future reuse. This function is carried out by the data retrieval component, which includes a collection of File Transfer Protocol (FTP)/Hypertext Transfer Protocol (HTTP) data downloading scripts and data reformatting programs. The HTTP-based Graphical Analysis and Display System (GrADS) data server (GDS; http://www.grads.iges.org/grads/gds/) is a key component for managing transparent and on-demand LIS input and output data.

LIS has been designed and built to support most parallel computing platforms, ranging from commodity Linux clusters to supercomputers at NASA, NOAA, and DoD. LIS has software components to adapt its parallel computing modes to the underlying computing platforms. In addition, LIS is flexible enough to run on commodity computers such as desktops or laptops for small-scale simulations.

A web-based data visualization system, the Land Explorer, has also been developed for users to interactively access and visualize the data. The Land Explorer is integrated with the GrADS-DODS data management server, and provides a unified interface for users to access LIS input, output and parameter data saved in different remote and local locations and formats.

2.1 LIS process representations

As described in detail in the online documentation and software design documents available at http://lis.gsfc.nasa.gov, the LIS software executes 1D land surface models on land cover-based tiles, which are employed to simulate sub-grid variability according to the mosaic approach of [12] and [13] when the grid resolution is coarser than the resolution of the supporting land cover data (globally 1 km, regionally 30 m). The land surface models currently incorporated in LIS include the community Noah LSM (Noah; [14]), the community land model version 2.0 (CLM; [15,16]), the variable infiltration capacity (VIC) model [17], Mosaic [18], and HySSIB [19], with more model plug-ins on the way. In keeping with the 1D physics of these land surface models, LIS was designed so that grids or tiles do not communicate horizontally, although extensions to LIS that support multidimensional communication are currently underway.

2.2 LIS parameters

LIS parameter datasets include topography, land cover, vegetation properties such as greenness fraction and/or leaf area index, and soil texture and/or sand and clay percentages. The global terrain maps of elevation, slope, aspect and curvature used in LIS are derived from the GTOPO30 global 30 arc second dataset [20]). A static, 1 km resolution, global vegetation classification dataset produced at the University of Maryland [21] provides a land cover classification to which each model's intrinsic land cover parameter tables are mapped. Monthly climatologies (i.e., multiyear averages) of vegetation properties such as Leaf Area Index (LAI), Stem Area Index (SAI) and greenness fraction are derived from two major sources: Advanced Very High Resolution Radiometer (AVHRR) and the EOS-Terra or EOS-Aqua MODIS. LIS currently supports the use of two LAI products: the first is an AVHRR-based 8 km product [22], and the second is a MODIS-based 1 km product [23]. The Noah LSM [14] requires monthly Green Vegetation Fraction (GVF) and quarterly snow-free albedo. The GVF fields in LIS are currently sampled from the AVHRR global, 0.144° (approx. 15 km), monthly 5 year climatology derived by [24], although
a 12-year climatology and near-real-time product are now available from NOAA. Other custom Noah datasets are based on the NLDAS [5] and GLDAS [6] databases.

Soil hydraulic properties are derived either from pedotransfer functions (e.g., CLM; [15]) or from texture-based lookup-tables (e.g., Noah; [14]) mapped using vertically uniform sand, silt and clay percentages from the 5 min global soils dataset of [25], which is based largely on that of the Food and Agriculture Organization (FAO). A finer, 1 km resolution State Soil Geographic Database (STATSGO) soils database from [26], which carries 16 texture classes by layer over 11 layers to 2 m depth, may be used to provide soil hydraulic properties in LIS for the continental United States.

2.3 LIS coupled and uncoupled running modes

Meteorological inputs or forcings in LIS are provided by a combination of global or regional model analyses superposed with satellite-, radar- and gauge-based observations of precipitation and by satellite-based observations of radiation, as discussed by [27] and [2]. Over the US, the highest resolution routinely available precipitation and radiation products have 4 km spatial resolution and an hourly temporal resolution ([28,29]). Therefore, an active area of current LIS research is to improve upon our precipitation and radiation downscaling using a combination of topographic (e.g., [30]) and statistical/dynamical (e.g., [31]) approaches. Other meteorological inputs such as near-surface air temperature, humidity, pressure, winds and downward longwave radiation are obtained by topographically downscaling atmospheric analyses from NASA and/or NOAA using lapse-rate and hypsometric adjustments to the 1 km GTOPO30 elevation data in LIS as discussed by [27].

By explicitly modeling 1 km heterogeneity in topography, land cover and soils, in addition to including downscaling strategies for forcings, LIS is an ideal testbed for studies of the impact and representation of mesoscale (>1 km) land surface heterogeneity in weather and climate models.

3 Software performance

LIS is designed to perform local, regional or global land surface simulations at horizontal spatial resolutions down to 1 km in near-real-time, and has been used at the meter scale on smaller domains. As shown in Table 1, the computational requirements for global 1 km horizontal resolution land surface modeling are almost three orders of magnitude larger than the current 1/4° scale of other global land surface modeling efforts, such as GLDAS mentioned previously. By successfully achieving such an unprecedented scale and intensity, LIS represents a major contribution to Earth science modeling capabilities.

At the Goddard Space Flight Center (GSFC), the LIS team designed and built a custom Linux cluster using commodity components as our primary development platform and testbed, because at the time it was infeasible to find dedicated resources on NASA's supercomputer systems to accommodate a problem of such scale. More importantly, the LIS Linux cluster gives us ultimate flexibility to test innovations and unconventional approaches. The LIS Beowulf cluster design includes the most economical commodity components ($100 k in 2002 dollars), including eight I/O nodes and 192 compute nodes. Overall, the LIS cluster has 208 AMD XP processors of 1.53 GHz and above, 160 GB of memory, 24 TB of disk space, 192 fast Ethernet connections, and 10 Gb Ethernet connections.

LIS has two built-in parallelization schemes. The 1D representation of water and energy transfer processes in the current generation of LSMs yields results in a coarse-grained parallelization problem. In the absence of data assimilation with horizontal correlations, lateral subsurface (groundwater) flow, and/or re-infiltration of surface runoff, the global land surface can be partitioned into several pieces and each piece can be processed independently without communication or synchronization with other pieces. LIS has components to employ the pool-of-tasks paradigm to deal with this mode of execution, allowing us to accomplish our performance objectives using the previously described Linux cluster, with near linear scalability.

The second parallelization scheme implemented in LIS is a conventional message passing interface (MPI)-based scheme. This allows us to handle the fine-grained computational problems involving horizontal coupling, and allows LIS to take advantage of shared-memory systems or distributed systems with fast inter-connections, as available on supercomputing platforms. The advantage of this scheme is that job management is not handled at the code level; instead, the queuing systems on most supercomputers perform this task.

For either parallelization scheme, input/output (I/O) has been identified as the primary LIS performance bottleneck, as indicated by the huge data volume in Table 1. On the LIS Linux cluster, three unique and unconventional approaches help overcome the I/O bottleneck: (1) parallel on-demand serving of input data; (2) distributed storage of output data on local disks; and (3) peer-to-peer data replication, as described below.

3.1 Parallel on-demand serving of input data

Eight input/output servers running the Graphical Analysis and Display System (GrADS)-Distributed Oceanographic Data System (DODS) servers provide a unified interface for LIS code to access the large collection of input datasets saved in various formats. Instead of the conventional scattering/broadcasting data serving paradigm from a master node, each compute node in the LIS system individually contacts one of the GrADS-DODS servers and requests the required subset. A load-balancing mechanism keeps the eight servers equally loaded, and the compute nodes do not need to connect and request the data from the servers in synchronization; in fact, they have natural timing differences when running uncoupled, which further spreads the load in time. These approaches greatly increase the input data serving throughput. This client–server mode of operation also includes a mechanism for fault tolerance in case any of the compute nodes fail during the computation. A time-out mechanism is used to determine the failure of a node and the subsequent redistribution of the job.

3.2 Distributed storage of output data on local disks

Distributed storage of LIS output data on local disks eliminates the need for time-consuming gathering and post-processing operations. For the 1 km global resolution, the standard practice of reassembling the data onto a central server is infeasible in near-real-time, because the computational intensity of those operations is in the same order as LIS producing the output data. Instead, compute nodes save the output data on their local disks, and a high-performance data serving system performs on-demand data gathering directly from the compute nodes' local disks. In addition, this system is interfaced with the GrADS-DODS system to replace the local file reading interface, thus enabling GrADS-DODS serve LIS output data directly off the distributed data sources to users. An additional advantage of this system is the increased data bandwidth, as the aggregated data transfer rate from multiple compute nodes with multithreaded data retrieving is 2–3 times larger than reading from one central server.

3.3 Peer-to-peer data replication

The combination of peer-to-peer data replication with the LIS distributed storage approach greatly reduces the time and complexity needed to stage 1 km global LIS runs. At 1 km resolution, in addition to the large volume of input and output data, the static parameter data volume is also significant. From the performance perspective, it is much better to pre-stage a copy of the parameter data on each compute node, instead of serving them repeatedly over the network in real time. However, the task to stage $\sim$100 GB of parameter data on 200 compute nodes rapidly is a nontrivial problem, but is well-suited to a peer-to-peer (P2P) data distribution system—in this case BitTorrent (http://www.bittorrent.com/), which is traditionally used to share large video files. This is an entirely different paradigm from the standard central client–server architecture, in which the central server becomes overloaded when the number of clients is large. BitTorrent eliminates the need for a central data server. Starting with a single copy of the data on a particular node, other nodes get a random piece of the data and then they share these pieces between themselves, and shortly all the nodes will have a complete copy of the data after each reassembles the pieces obtained from other nodes. This makes fast data replication possible, and estimated throughput from this P2P distribution is at least 1-times higher than the conventional central-server approach.

The LIS software exceeded our performance goal of near-real-time simulation by running one day of global simulations at 1 km resolution in less than one day. LIS also exceeded our performance, portability and interoperability goals to enable near-real-time simulations on a variety of platforms (e.g., distributed-memory Linux cluster, shared-memory SGI supercomputer, hybrid-memory Compaq cluster). As shown in Fig. 3, LIS could successfully execute up to about six days per day at the full 1 km resolution represented by EOS-era level 3 products, while generating 600 Gb of level 4 product output per day simulated. As shown in this figure, the performance varied somewhat by input dataset and by land surface model. For example, the performance was consistently worse for input meteorological data from the NASA Goddard Earth

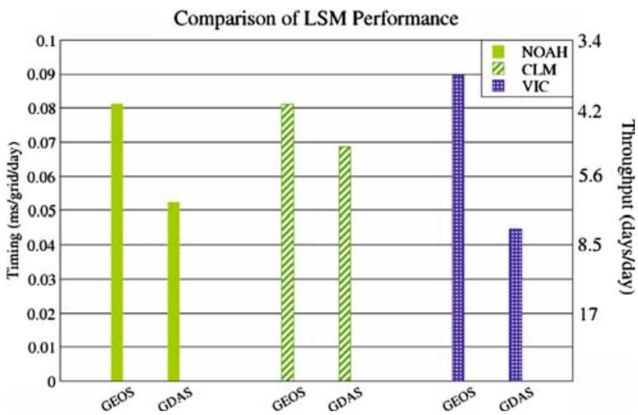

Fig. 3 Comparison of LIS land surface model (LSM) performance for global executions at 1 km spatial resolution showing that the throughput is better than three days per day for all cases. The three land surface models (as described in the text) include the Noah, CLM, and VIC models. The two *bars* for each model indicate the performance for different types of input meteorological data, namely the GEOS and the NOAA National Centers for Environmental Prediction's Global Data Assimilation System (GDAS) analysis systems

Observing System (GEOS) model compared to input data from the NOAA National Centers for Environmental Prediction's Global Data Assimilation System (GDAS). This was due to more physically based interpolations being performed in order to transform the data to the 1 km scale. As also seen in Fig. 3, there was some interplay between the land surface model complexity and the input data, so that the CLM model performance was less sensitive to input dataset, while the VIC model was much more sensitive. The high LIS throughput, wrapped with the ESMF interface, makes LIS a powerful Earth system modeling component with broader applications.

4 Applications and impact

LIS has been used extensively in four major areas: (1) satellite mission support; (2) technology development; (3) scientific research; and (4) end-use applications.

In the mission support area, LIS is being used on the science teams for the currently orbiting Gravity Recovery and Climate Experiment (GRACE; [32]), the Tropical Rainfall Measurement Mission (TRMM; Fig. 4) and the EOS-Aqua Advanced Microwave Scanning Radiometer-EOS (AMSR-E). Furthermore, LIS has helped to refine future snowpack and soil moisture mission requirements via observing system simulation experiments (OSSEs; [33]).

LIS is advancing Earth system modeling technologies in the ESTO-supported project "coupling high-resolution Earth system models using advanced computational technologies", in which LIS has been coupled to two different atmospheric modeling components—the Goddard Cumulus Ensemble (GCE; [4]) and Weather Research and Forecasting (WRF; [5]). This coupling has greatly enhanced our ability to study coupled energy and water cycles, and to further understand the requirements to better predict interactions between clouds, radiation and the land surface ([34]).

In the scientific research area, LIS is currently being used in several projects related to global water and energy cycle modeling. For example, recent work conducted by the GLDAS team [35] used the LIS code to investigate model

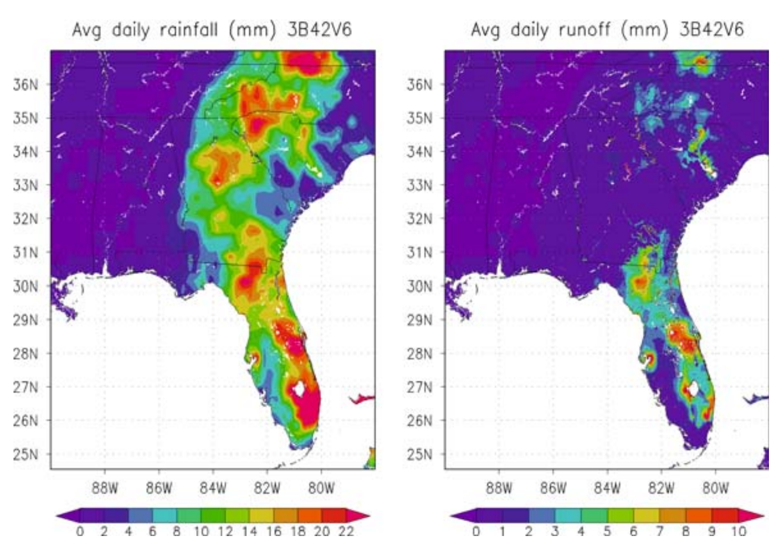

Fig. 4 Averaged daily rainfall (*left*) and surface runoff (*right*) predicted by LIS with the Noah LSM during hurricane Jeanne (25–30 September 2004), using a TRMM-based precipitation product as input. The input precipitation product is the NASA/TRMM and other satellites version 6 analysis, which contains rain gauge bias adjustments (3B42V6). As discussed in the text, the simulations use high-resolution (1 km) land cover data from MODIS/Terra-based satellite remote-sensing observations, MOD12Q1, collection 3, originally distributed by NASA, and soils data (STATSGO) from the US department of agriculture (USDA)

performance at reference sites for the World Meteorological Organization's coordinated enhanced observing period (CEOP; [36,37]). In addition, the LIS team submitted several results for the world climate research program's global energy and water cycle experiment (GEWEX) second global soil wetness project [8,38] baseline experiments.

LIS is being used in several applications projects in the areas of water resources, weather prediction, air quality, and homeland security with our partners at the Air Force Weather Agency (AFWA), the US Bureau of Reclamation, NOAA's National Centers for Environmental Prediction, the US Environmental Protection Agency (EPA), and the US Army Corps of Engineers/Engineering Research and Development Center (USACE). LIS's unique water and energy cycle prediction capabilities are essential for understanding the vulnerability of life on our planet to weather and climate extremes, including such high-impact events as flooding, droughts, tsunamis and hurricanes. LIS is being applied by our interagency users to help improve decision support in the following ways: (1) to provide better guidance on water resource management and planning (with the Bureau of Reclamation); (2) to provide energy and soil moisture inputs to improve weather and climate forecasting (with NOAA's National Centers for Environmental Prediction); and (3) to improve forecasts of agricultural production, flood/drought prediction, pollutant dispersion, military mobility/trafficability (e.g., with the US Army Corps of Engineers and Air Force Weather Agency).

The high spatial resolution of LIS makes it capable of resolving urban areas, as shown in Fig. 5. The impact of high-resolution input data (as previously discussed) is clearly seen as the small, medium, and large urban areas are more fully resolved and seen to contribute a net heat flux that is less than that modeled at the $1/4°$ scale. A key area of ongoing work is updating land surface models to represent the impact of engineered surfaces (buildings, roads, parking lots, sidewalks, etc.) on mesoscale land–atmosphere interactions, including water, energy, and momentum fluxes. In addition, due to its coupling with two atmospheric model components, LIS is being applied extensively in NASA's Energy and Water Study (NEWS) program to understand the nature of land–atmosphere coupling and help improve prediction of weather and climate.

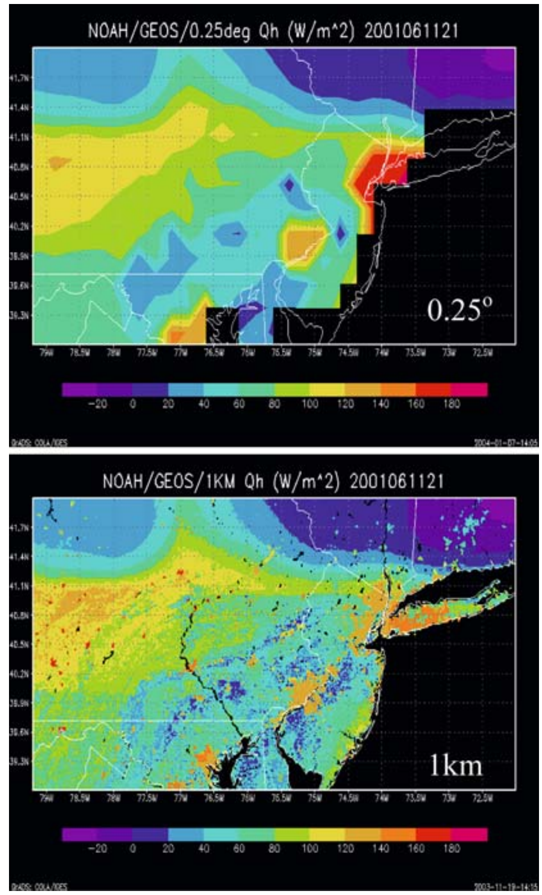

Fig. 5 Sensible heat flux (Qh) estimated by LIS/Noah for the mid-Atlantic region on 11 June 2001, 2100 GMT at (**a**) $1/4°$ (**b**) 1 km horizontal spatial resolution

5 Discussion and conclusions

LIS is a high-performance modeling and data-assimilation system that can perform local, regional or global land surface simulations. The LIS software was the co-winner of NASA's 2005 software of the year award, and is directly relevant to NASA's mission "to understand and protect our home planet". LIS enables accurate global water and energy cycle prediction, and provides useful information for water-resource management, weather prediction, air quality, and military operations. The LIS software met a high-performance standard to enable near-real-time simulations at the full global 1 km resolution represented by EOS-era level 3 products, while generating 600 Gb of level 4 products per simulated model day. Such an unprecedented scale and intensity is a major contribution to Earth science modeling capabilities.

The LIS software includes a fault-tolerant job management system run on Linux clusters or other parallel systems, high-performance, high-availability parallel I/O based on GDS servers with dynamic load balancing and distributed data storage, and highly scalable data replication with peer-to-peer technology. This design, in addition to complying with ESMF and ALMA, is critical in making LIS a high-performance Earth system modeling component with broader applications.

In the management and development of LIS, we have followed standard project management and software engineering procedures and practices, and have released LIS to the public and the scientific community as an open-source software package with extensive documentation and sample datasets. LIS has attracted over 300 registered accounts from academic institutions, government agencies, international organizations as well as individual users all over the world. With over 150,000 lines of C and Fortran90 code, and over 250,000 hits on our website over the past 12 months, NASA's initial investment in LIS development is now poised to benefit the nation through ongoing operational implementations with our partners at NOAA's National Centers for Environmental Prediction and the Air Force Weather Agency, as well as in education and research with our academic partners and the broader Earth system modeling community. Please visit the LIS website at http://lis.gsfc.nasa.gov for more information and to download the software package.

Acknowledgments We gratefully acknowledge support from NASA's Earth science technology office (ESTO) computational technologies (CT) project, for which this work has been performed under task agreement GSFC-CT-2 resulting from the round-3 CAN-00-OES-01 (Co-PI's Houser and Peters-Lidard). We also gratefully acknowledge the additional support of NASA ESTO advanced information systems technologies (AIST) NRA-02-OES-04 (PI Peters-Lidard), and more-recent support from our partners at the Air Force and the US Army Corps of Engineers. In addition, the LIS team appreciates the support and assistance from members of the GLDAS and NLDAS teams.

References

1. Kumar SV, Peters-Lidard CD, Tian Y, Geiger J, Houser PR, Olden S, Lighty L, Eastman JL, Dirmeyer P, Doty B, Adams J, Wood E, Sheffield J (2006) LIS—an interoperable framework for high resolution land surface modeling. Environ Model Softw 21:1402–1415
2. Rodell M, Houser PR, Jambor U, Gottschalck J, Mitchell K, Meng C-J, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Lohmann D, Toll D (2004) The global land data assimilation system. Bull Am Meteorol Soc 85(3):381–394
3. Mitchell KE, et al (2004) The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. J Geophys Res 109:D07S90. doi:10.1029/2003JD003823
4. Tao W-K, Simpson J, Baker D, Braun S, Chou M-D, Ferrier B, Johnson D, Khain A, Lang S, Lynn B, Shie C-L, Starr D, Sui C-H, Wang Y, Wetzel P (2003) Microphysics, radiation and surface processes in the Goddard Cumulus Ensemble (GCE) model. Meteorol Atm Phys 82:1–4, 97–137. doi:10.1007/s00703-001-0594-7
5. Skamarock WC,Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2005) A description of the advanced research WRF Version 2, NCAR/TN–468+STR, p 100
6. Henderson-Sellers A, Irannejad P, McGuffie K, Pitman A (2003) Predicting land-surface climates-better skill or moving targets? Geophys Res Lett 30:14, 1777. doi:10.1029/2003GL017387
7. Dirmeyer PA, Dolman AJ, Sato N (1999) The Global Soil Wetness Project: a pilot project for global land surface modeling and validation. Bull Am Meteorol Soc 80:851–878
8. Dirmeyer PA, Gao X, Zhao M, Guo Z, Oki T, Hanasaki N (2006) The Second Global Soil Wetness Project (GSWP-2): multi-model analysis and implications for our perception of the land surface. Bull Am Meteorol Soc (in press)
9. Hill C, DeLuca C, Balaji V, Suarez M, da Silva A, and The ESMF Joint Specification (2004) The architecture of the Earth system modeling framework. Comput Sci Eng 6:18–28
10. Polcher J, McAvaney B, Viterbo P, Gaertner M-A, Hahmann A, Mahfouf J-F, Noilhan J, Phillips T, Pitman A, Schlosser CA, Schulz J-P, Timbal B, Verseghy D, Xue Y (1998) A proposal for a general interface between land-surface schemes and general circulation models. Glob Planet Change 19:261–276
11. Guilyardi E, Budich R, Valcke S (2003) PRISM and ENES: a European approach to Earth System Modelling. In: Zwieflhofer W, Kreitz N (eds) Realizing teracomputing, Proceedings of the tenth ECMWF workshop on the use of high performance computing in meteorology. World Scientific, pp 146–164 (ISBN 981-238-376-X)
12. Avissar R, Pielke RA (1989) A parameterization of heterogeneous land surfaces for atmospheric models and impact on regional meteorology. Mon Weather Rev 117(10):2113–2136
13. Koster RD, Suarez MJ (1992) Modeling the land surface boundary in climate models as a composite of independent vegetation stands. J Geophys Res 97(D3):2697–2715
14. Ek MB, Mitchell KE, Lin Y, Rogers E, Grunmann P, Koren V, Gayno G, Tarpley JD (2003) Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model. J Geophys Res 108(D22):8851. doi:10.1029/2002JD003296
15. Bonan GB, Oleson KW, Vertenstein M, Levis S, Zeng X, Dai Y, Dickinson RE, Yang Z-L (2002) The land surface climatology of the community land model coupled to the NCAR community climate model. J Clim 15(22):3123–3149
16. Dai Y, Zeng X, Dickinson RE, Baker I, Bonan GB, Bosilovich MG, Denning AS, Dirmeyer PA, Houser PR, Niu G-Y, Oleson KW, Schlosser CA, Yang Z-L (2003) The common land model (CLM). Bull Am Meteorol Soc 84(8):1013–1024
17. Liang X, Wood EF, Lettenmaier D (1996) Surface and soil moisture parameterization of the VIC-2L model: evaluation and modifications. Glob Planet Change 13:195–206
18. Koster RD, Suarez MJ (1996) Energy and water balance calculations in the Mosaic LSM. NASA Tech. Memo. 104606, vol. 9, p 59 (Available from NASA/Center for Aerospace Information, 800 Elkridge Landing Road, Linthicum Heights, MD 21090)
19. Mocko, David M, Sud YC (2001) Refinements to SSiB with an emphasis on snow physics: evaluation and validation using GSWP and valdai data, earth interactions 5(1):1–31. doi:10.1175/1087-3562(2001)005<0001:RTSWAE>2.0.CO;2
20. Gesch DB, Verdin KL, Greenlee SK (1999) New land surface digital elevation model covers the Earth. EOS Trans Am Geophys Union 80(6):69–70
21. Hansen MC, DeFries RS, Townsend JRG, Sohlberg R (2000) Global land cover classification at 1 km spatial resolution using a classification tree approach. Int J Remote Sens 21:1331–1364
22. Myneni RB, Nemani RR, Running SW (1997) Algorithm for the estimation of global land cover. IEEE Trans Geosci Remote Sens 35:1380–1393
23. Myneni RB, Hoffman S, Knyazikhin Y, Privette JL, Glassy J, Tian Y, Wang Y, Song X, Zhang Y, Smith GR, Lotsch A, Friedl M, Morisette JT, Votava P, Nemani RR, Running SW (2002) Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens Environ 83:214–231
24. Gutman G, Ignatov A (1998) Derivation of green vegetation fraction from NOAA/AVHRR for use in numerical weather prediction models. Int J Remote Sens 19:1533–1543
25. Reynolds CA, Jackson TJ, Rawls WJ (2000) Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions. Water Resour Res 36:3653–3662

26. Miller DA, White RA (1998) A conterminous United States multi-layer soil characteristics dataset for regional climate and hydrology modeling. Earth Interact 2:1–26
27. Cosgrove BA et al (2003) Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. J Geophys Res 108(D22):8842. doi:10.1029/2002JD003118
28. Baldwin M, Mitchell KE (1997) The NCEP hourly multi-sensor US precipitation analysis for operations and GCIP research. In: Preprints, 13th AMS conference on hydrology. Am Meteorol Soc. Boston, Mass, pp 54–55
29. Pinker RT et al (2003) Surface radiation budgets in support of the GEWEX Continental-Scale International Project (GCIP) and the GEWEX Americas Prediction Project (GAPP), including the North American Land Data Assimilation System (NLDAS) project. J Geophys Res 108(D22):8844. doi:10.1029/2002JD003301
30. Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J Appl Meteorol 33:140–158
31. Nykanen DK, Foufoula-Georgiou E, Lapenta WM (2001) Impact of small-scale rainfall variability on larger-scale spatial organization of land-atmosphere fluxes. J Hydrometeorol 2(2):105–121
32. Rodell M, Chao BF, Au AY, Kimball J, McDonald K (2005) Global biomass variation and its geodynamic effects, 1982–1998. Earth Interact 9(2):1–19
33. Zhan X, Houser PR, Walker JP, Crow WT (2006) A method for retrieving high-resolution surface soil moisture from hydros L-band radiometer and radar observations. IEEE Trans Geosci Remote Sens (in press) doi:10.1109/TGRS.2005.863319
34. Eastman J, Peters-Lidard C, Tao W, Kumar S, Tian Y, Lang SE, Zeng X, Houser P (2006) A meteorological model's dependence on radiation update frequency, Submitted to Geophys Res Lett
35. Kato H, Rodell M, Beyrich F, Cleugh H, van Gorsel E, Liu H, Meyers TP (2006) Sensitivity of land surface simulations to model physics, parameters, and forcings, at four CEOP sites, Submitted to J Meteor Soc Jpn
36. Koike T, The coordinated enhanced observing period—an initial step for integrated water cycle observation. WMO Bull 53(2):115–121
37. Lawford R, Bosilovich M, Eden S, Benedict S, Brown C, Gruber A, Houser P, Meyers T, Mitchell K, Peters-Lidard C, Roads J, Rodell M, Tarpley D, Williams S (2006) US contributions to the Coordinated Enhanced Observing Period (CEOP). Submitted to the Bull Am Meteorol Soc
38. Dirmeyer PA, Guo Z, Gao X (2004) Comparison, validation and transferability of eight multi-year global soil wetness products. J Hydrometeorol 5:1011–1033