LIS - An Interoperable Framework for Global, High Resolution Land Surface Modeling

Sujay Kumar¹, Paul Houser², and Christa Peters-Lidard²

Abstract. Knowledge of land surface water, energy, and carbon conditions are of critical importance due to their impact on many real world applications such as agricultural production, water resource management, and flood, weather, and climate prediction. The need to predict accurate, optimal fields of land surface states and fluxes globally has led to the development of the Land Information System (LIS), a near-real-time, high resolution system that integrates satellite and ground-based observational data along with advanced land modeling and data assimilation models. LIS employs the use of scalable, high performance computing technologies to meet the increased computational and memory requirements. LIS is adopting various Earth system modeling standards such as the Earth System Modeling Framework (ESMF) and Assistance for Land Modeling Activities (ALMA) for future coupling with other earth system models. In addition, by developing a system with well defined interfaces that enable the users to reuse the underlying computing infrastructure and land modeling tools, LIS itself is evolving into a framework for land surface modeling community.

1 Introduction

Land surface modeling seeks to predict the terrestrial water, energy, and biogeochemical processes by solving the governing equations of soil-vegetation-snowpack medium. The land surface and atmosphere are coupled to each other over a variety of time scales through the exchanges of water, energy, and carbon. An accurate representation of land surface processes is critical for improving models of the boundary layer and land-atmosphere coupling at all spatial and temporal scales and over heterogeneous domains. Long term descriptions of land use and fluxes also enable in the accurate assessments of climate characteristics. In addition to the impact on the atmosphere, predicting land surface processes is also critical for ecosystem modeling and water resources prediction and management.

The need for accurate predictions of land surface states has led to the development of a Land Data Assimilation System (LDAS) [4] that consists of a number of uncoupled land models using remotely-sensed and in-situ observations within a land data assimilation framework. LDAS has been successfully used for simulations upto 1/8 degree resolution in both real-time and long-term (50 years) retrospective simulations. However, to improve the understanding of land surface processes and their interaction with atmospheric processes, such a system needs to be implemented globally at a high resolution such as 1km. The

¹ Goddard Earth Science Technology Center, University of Maryland at Baltimore County, Greenbelt, MD, 20771
² Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD, 20771
motivation behind LIS is to extend the capabilities of LDAS and to provide a computing environment for global, high resolution land surface modeling.

Many existing Earth science applications, though highly scalable and computationally capable, lack the ability to interoperate with other Earth system applications. The cost to adapt such applications to function with other Earth system applications may be prohibitively high. LIS attempts to achieve code interoperability by applying advanced software engineering concepts in its design. The system is designed as an object oriented framework that can be shared and reused by scientists and practitioners in the land modeling community. One of the design goals of LIS is to allow use of a complete, usable, and integrated set of high level facilities without the necessary knowledge of underlying computer hardware or software. The use of object oriented principles help in designing LIS to be flexible and extensible, enabling rapid prototyping of new applications into LIS.

In addition to providing an infrastructure to support land surface research and applications activities, LIS is adopting other Earth system modeling standards and conventions, such as ESMF [3] and ALMA [1]. ESMF is a system 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 is a land-atmosphere coupling standard that is being developed by the broad land-atmosphere research community. By conforming to the standards laid out by ESMF and ALMA, LIS can provide capabilities to interact with other Earth system models.

The following sections describe the computing infrastructure, the interoperable features, and the adoption of ESMF and ALMA standards, tools and interfaces in LIS.

2 High Performance Computing in LIS

LIS includes a model control and input/output system that drives multiple offline one-dimensional land surface models (LSMs) to facilitate global modeling. The one-dimensional LSMs apply the governing equations of the physical processes of the soil-vegetation-snowpack medium to characterize the transfer of mass, energy, and momentum between a vegetated surface and the atmosphere.

The global land surface is modeled by dividing it into two-dimensional regions or gridcells (for example cells of size 1km × 1km would lead to approximately $5 \times 10^8$ gridcells). Assuming approximately 15ms for each day of land surface model execution on a particular gridcell, it can be estimated that to conduct a day’s simulation at 1km using 15 minute timesteps would require approximately 3 months. It is clear that global, high resolution land surface modeling presents a significant computational challenge.

In addition to the computational challenges, high resolution land surface modeling also presents significant I/O bottlenecks. The definition of data structures associated with each gridcell and the variables associated with defining sub-grid variability at the global resolution causes the memory requirements to scale significantly with the domain resolution. From the experiments conducted at low resolutions using LDAS, it can be estimated that memory in the order of terabytes would be required for global land surface modeling at 1km.

Due to the computational challenges described above, the use of scalable computing technologies is critically important and relevant for LIS. Land surface processes have rather weak horizontal coupling on short time and large space scales. LIS is taking advantage of
this inherent parallelism that enables highly efficient scaling across massively parallel computational resources. LIS employs a unique custom-built 200 node Linux cluster consisting of 192 computing nodes and 8 I/O nodes, specifically to handle the huge data management requirements of high resolution land surface modeling.

Simulations of LIS at 5km global resolution were conducted using two different LSMs: 1) Community Land Model (CLM) [2] and 2) the National Oceanic and Atmospheric Administration’s NOAH [5]. The experiments were conducted on an SGI Origin 3000 (Lomax) at the NASA Ames Research Center. There are approximately 21 million gridcells at 5km resolution. Currently LIS employs a simple master-slave paradigm for parallel processing. A master processor initializes the global grid and the associated parameters. Since land surface modeling is carried out only on land points, the global 2-dimensional grid is reduced to a vectorized representation, eliminating the ocean points. The master processor carries out the domain decomposition and the slave processors carry out the land surface model executions and other computations on the subdomains. The master processes collects variables required for output at every output time interval.

Figure 1 shows the scaling curves at 5km for a day of simulated run. The sequential executions of NOAH and CLM required approximately 12 and 18 hours, respectively. It can be observed that the parallelization scheme in LIS significantly reduces the computation times to the order to 2 hours for a day’s simulation.

![Fig. 1. LIS execution times on SGI Origin 3000 at 5km](image)

3 Object Oriented Principles in LIS

LIS system is designed using object oriented design principles, providing a number of well-defined interfaces or “hook points” for enabling development of new features and applications into LIS. As described earlier, LIS incorporates a number of one-dimensional LSMs. These LSMs typically require three types of inputs: 1) initial conditions, which describe the initial state of land surface; 2) boundary conditions, which describe both the atmospheric states
also known as forcings and the soil states; and 3) parameters, which are functions of soil, vegetation, topography, etc. and are used to solve the governing equations. LIS makes use of data from various satellite and ground-based observations as well as output from numerical prediction models to force the LSMs in LIS. A common interface to different LSMs enables the reuse of this broad set of data and other tools. Further, forcing different LSMs with the same input allows researchers to perform intercomparisons of output from different LSMs. The need for such a flexible interface is addressed in LIS by a component-based design that provides a number of well defined interfaces for the incorporation of new land surface schemes. The implementation of these interfaces allows the user to employ underlying computing and land modeling infrastructure when a new LSM is incorporated.

4 Interoperability in LIS

As mentioned earlier, LIS will comply with ALMA and ESMF to include interoperability in its design. The use of ALMA standards enables LIS to interoperate with other land modeling systems and the use of ESMF allows for future coupling with earth system models such as atmospheric models.

4.1 LIS and ALMA

ALMA is a flexible data exchange convention to facilitate the exchange of forcing data for LSMs and the results produced by them. The output data variables and formats, and the variables passed between LIS and the land models follow the ALMA specification. By implementing the ALMA convention, LIS can exchange data with other land modeling systems that are also ALMA compliant. Further, ALMA compliance enables LIS to be used for intercomparisons of land surface models.

4.2 LIS and ESMF

The purpose of ESMF is to develop a framework that provides a structured collection of building blocks that can be customized to develop model components. ESMF can be broadly viewed as consisting of an infrastructure of utilities and data structures for building model components and a superstructure for coupling and running them.

ESMF provides a utility layer that presents a uniform interface for common system functions such as time manager, basic communications, error handler, diagnostics, etc. LIS currently uses the ESMF time manager and the logging and error diagnostics tools. The time management utility provides useful functions for time and data calculations and higher level functions that control model time stepping and alarms. The log utility organizes diagnostic output and allows for searches and filters to be constructed. The error handler provides both uniform handling of errors and a way for users to select how the errors will be handled.

ESMF also defines a number of guidelines for applications that are intended to be coupled with other Earth system models. ESMF provides definitions of a Gridded Component class for user-supplied components discretized on grids and a Coupler Component class for the software that is used to couple them together. LIS will implement the interfaces required to be a Gridded Component and will use ESMF_State class to exchange information with
other models and systems. By implementing the ESMF interfaces at the top of the LIS control system, each LSM in LIS will not need to be customized to comply with the ESMF standards. The implementation of these interfaces will enable LIS to couple with other Earth system models. Figure 2 shows a schematic view of the interaction between LIS and ESMF.

![ESMF LIS Interaction Diagram](image)

**Fig. 2.** LIS and ESMF

## 5 Summary and Future Directions

LIS is an evolving framework for global high resolution land surface modeling. Some of the computational challenges faced by LIS are addressed by the use of scalable parallel computing architectures as described in the results above. However, the I/O requirements still pose a significant challenge and will be the focus of the development efforts in the future. LIS is also actively participating in the development of Earth system modeling standards by complying with the ESMF guidelines and standards. ESMF is a project in development and LIS will adopt the utilities and interfaces as they become available. In addition to adopting the ESMF for future coupling with other earth system models, LIS itself provides an interoperable framework for the land modeling community by adopting land surface modeling standards such as ALMA and by defining extensible interfaces in its design.

## References