Use of inverse modeling tools for improved utilization of earth science data in land surface modeling and data assimilation

Sujay Kumar
SAIC/NASA GSFC
Hydrological Sciences Branch
Greenbelt, MD

Christa Peters-Lidard
NASA GSFC
Hydrological Sciences Branch
Greenbelt, MD

Ken Harrison
ESSIC/NASA GSFC
Hydrological Sciences Branch
Greenbelt, MD

Soni Yatheendradas
ESSIC/NASA GSFC
Hydrological Sciences Branch
Greenbelt, MD

Joe Santanello
NASA GSFC
Hydrological Sciences Branch
Greenbelt, MD

Abstract—**In this work, the development and application of optimization and uncertainty estimation tools within the NASA Land Information System (LIS) is presented. These tools are employed in the context of soil moisture simulation. The results demonstrate the application of these tools for improved exploitation of Earth science data relevant for hydrology.**

I. Introduction

The NASA Land Information System (LIS [1], [2]) is a high-resolution, high-performance, land surface modeling and data assimilation system to support a wide range of land surface research and applications. LIS integrates various community land surface models, ground and satellite-based observations, and ensemble-based data assimilation tools to enable assessment and prediction of hydrologic conditions at various spatial and temporal scales of interest. More recently, the LIS infrastructure has been enhanced with the development of a suite of inverse modeling tools (optimization and uncertainty estimation subsystems) that help in the improved use of Earth science data for hydrologic modeling. The optimization infrastructure in LIS includes a suite of both deterministic and stochastic algorithms and provides a new capability to estimate and improve global soil or vegetation parameters from satellite data. The uncertainty estimation algorithms, which include Bayesian approaches such as Markov Chain Monte Carlo (MCMC), acknowledge the uncertainty in modeling land surface processes and use the observational information to enable probabilistic predictions. These important capabilities represent significant advancements over current practice and further exploit the information content of the Earth science data.

II. Approach

To demonstrate the new optimization infrastructure in LIS, three optimization algorithms were implemented: (1) Levenberg-Marquardt (LM), (2) Genetic Algorithm (GA) and (3) Shuffled Complex Evolution from University of Arizona (SCE-UA). The three algorithms represent a range of search strategies suitable to different problem types. LM exploits local information (numerically evaluated derivatives) to efficiently identify the nearest local optimum in problems in which a single local optimum exists. In contrast, GAs use a random search strategy that helps avoid convergence to local, inferior optima. SCE-UA combines local and random search strategies.

Two test cases involving the estimation of soil parameters were formulated to demonstrate the implementation of the algorithms and the optimization infrastructure. Test case I was based on the prior work ([3]) that involved determining soil properties given estimates of near surface soil moisture derived from passive (L-band) microwave remote sensing over the Walnut Gulch watershed in Southeastern Arizona. Errors in the simulated versus observed soil moisture were minimized by adjusting the soil texture, which in turn controls the hydraulic properties through the use of pedotransfer functions (PTF). To further test the algorithms, test case II was formulated by modifying test case I to directly estimate the soil properties. Test case II, therefore presents a more challenging optimization problem, since it involves (a) the estimation of more parameters compared to test case I, and (b) a decision space solution characterized by a higher degree of freedom due to the absence of the PTF constraints.

In addition to the optimization infrastructure, a comprehensive uncertainty estimation subsystem was developed in LIS that enables the estimation of uncertainties in parameters and prediction. The design accommodates uncertainty estimation methods based on Bayesian analysis. The aim of such algorithms is to probabilistically infer the values of parameters given an initial description of the parameter uncertainties and given the observations and any associated modeling or observational errors. A random walk MCMC algorithm was implemented in the uncertainty estimation subsystem and was applied for test case II. The MCMC algorithm, not unlike the parameter estimation algorithms, involves exploration of

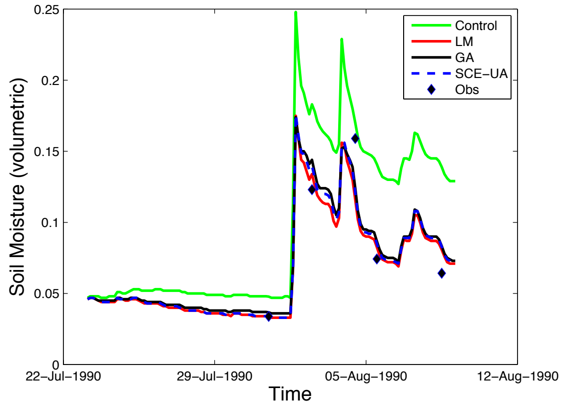

Fig. 1. Comparison of soil moisture simulations at Lucky Hills. The control represents the model simulation using the default parameters and LM, GA, and SCE-UA represents the model simulations using the optimized parameters which were estimated by these algorithms, respectively. The PBMR soil moisture observations are also shown.

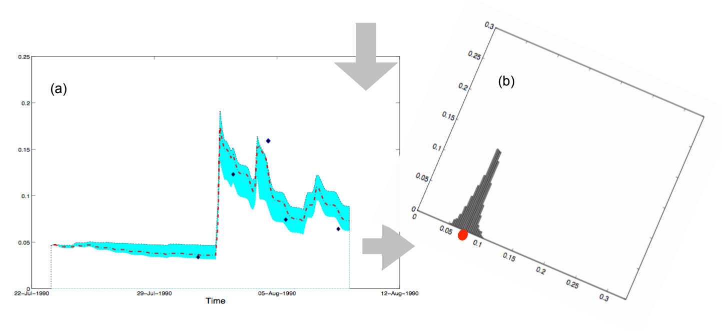

Fig. 3. The uncertainty estimation capability yields multiple sets of parameter values, and in turn, multiple sets of outputs from which confidence intervals can be computed. Here the range between 5th and 95th percentiles is plotted and envelopes the parameter estimation solution, including the GA solution (dashed). Also probabilistic forecasts can be generated that reflect the uncertainty. Above, a probabilistic forecast is given for the final timestep in the model run. The spread in soil moisture is approximately 15% of the dynamic range.

the parameter space. Unlike those algorithms, however, the method by which new parameter values are proposed is carefully constructed such that the points visited represent a random sample from the probability distribution over the parameter set and that are conditioned on the observations.

III. Results

The soil moisture simulations are conducted over two sites in the Walnut Gulch watershed: (1) Lucky Hills and (2) Kendall. The Noah land surface model in LIS is employed at 40m spatial resolution, forced with observed meteorology at these locations. Figure 1 shows the results from test case II, where the soil moisture simulations from using the default model parameters (control) is compared against the estimates generated by the model using the optimized parameters, using the three optimization algorithms (LM, GA, SCE-UA). It can be noticed that the optimized parameters improves the ability of the model to match the simulated soil moisture much more closely than when compared to the control that relied on mappings of hydraulic parameters from soil texture classifications.

Figure 2 illustrate the results from the use of MCMC algorithm for testcase II. The learning of parameter values can be inferred from examination of points visited by MCMC, as shown in Figure 2. Prior to the soil moisture observations, all combinations of parameter values within the scatterplot boxes were considered equally likely. After the observations, the probability coalesces around a localized region of the parameter space, though considerable uncertainty remains.

In addition to providing information on model parameter values, uncertainty estimation enables probabilistic estimates and forecasts of model variables such as soil moisture. LIS was run for each set of parameter values represented by the sample points shown in Figure 2. Time series were generated for soil moisture for each parameter set and at each timestep the 5th and 95th percentile values were computed. The range of soil moisture values between these percentiles is shown in Figure 3. In addition, a probabilistic forecast of the soil moisture is provided in the histogram inset in Figure 3. For comparison, the time series of GA parameter estimation solution (dashed line) is also shown.

Thus, these above experiments demonstrate the use of optimization and uncertainty algorithms towards the increased exploitation of the information content of remotely sensed observations.

Acknowledgment

We gratefully acknowledge the financial support from the NASA Earth Science Technology Office (ESTO). Resources supporting this work were provided by the NASA High-End Computing (HEC) program through the NASA Center for Computational Sciences (NCCS) at Goddard Space Flight Center.

References

[1] Kumar, S., C. Peters-Lidard, T. Tian, P. Houser, J. Geiger, S. Olden, L. Lighty, J. Eastman, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. Wood, and J. Sheffield, 2006: Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modeling and Software*, **21**, 1402–1415.

[2] Peters-Lidard, C.D., 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, and J. Sheffield, 2007: High Performance Earth system modeling with NASA/GSFC's Land Information System. *Innovations in Systems and Software Engineering*, **3**(3), 157–165.

[3] Santanello, J.A., C.D. Peters-Lidard, M. Garcia, D. Mocko, M. Tischler, M.S. Moran, and D.P. Thoma, 2007: Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed, *Remote Sensing of Environment*, **110**(1), 79–97.

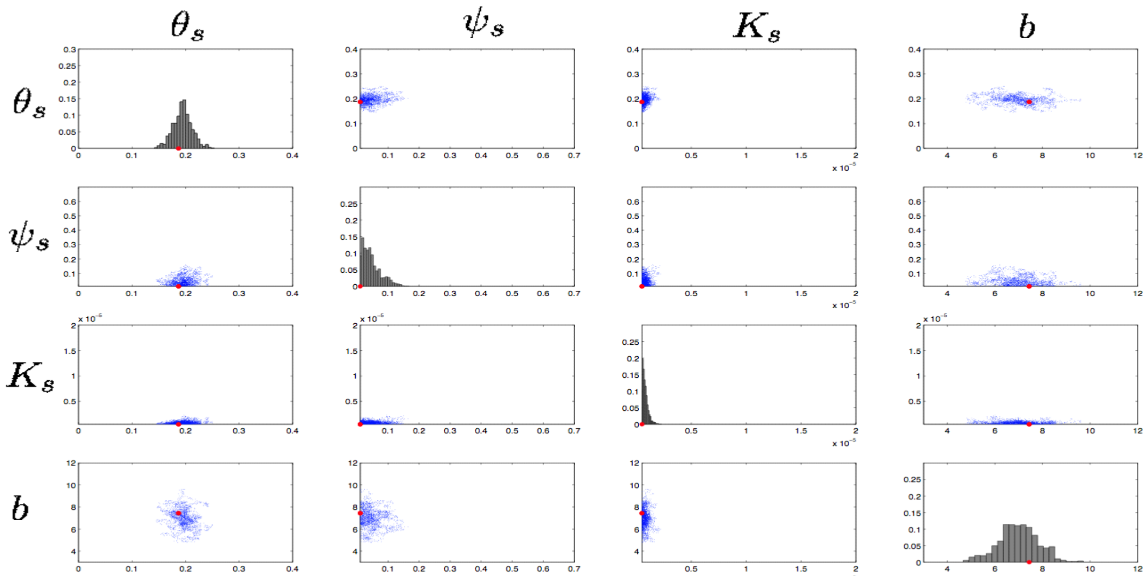

Fig. 2. The "learning" of the model parameter values is apparent with the examination of the points (blue dots) visited by the MCMC algorithm, shown here in 2D space for each pair of the four uncertain parameters. Initially all points within the boxes are equally likely. The points visited (after an initial "burn-in" period) are a sample of the probability distribution conditioned on the observations. The move to a localized region within the parameter space shows how the uncertainty, initially spread out over the entire space shown, narrows considerably. To contrast uncertainty estimation with the parameter estimation, the GA solution (red dot) is also shown.