Uncertainty Analysis in the Decadal Survey Era: A Hydrologic Application using the Land Information System (LIS)

Ken Harrison¹,², Sujay Kumar¹,³, Christa Peters-Lidard¹, Joe Santanello¹

¹ – Hydrological Sciences Branch, NASA/GSFC
² – Earth System Science Interdisciplinary Center (ESSIC), U. of Maryland
³ – Science Applications International Corporation (SAIC)

Earth Science Technology Forum 2011
Pasadena, CA

Acknowledgements: We gratefully acknowledge the financial support from the NASA Earth Science Technology Office (ESTO). Computing was supported by the resources at the NASA Center for Climate Simulation.
Uncertainty analysis: Scientists and decision-makers need a better understanding of the uncertainty in data products.
LIS Architecture
LIS Architecture

Core Structure and Features:
- Time Management Tools
- Logging and Diagnostic Tools
- High Performance Computing
- Configuration Tools

Abstractions:
- Land Surface Model
- Running Mode
- Meteorological Inputs
- Land Surface Parameters
- Domains
- RTM

Sample Use Case Implementations:
- Noah
- CLM
- Mosaic Catchment
- SiB2
- Hyssib
- Sacramento SNOW17
- Analysis Forecast Coupled RTM Forward mode
- NLDAS GDAS CMAP CMORPH ECMWF AGRMET GEOS TRMM
- Landcover (UMD, USGS, MODIS)
- Soils (FAO, STATSGO)
- Topography (USGS)
- LAI (AVHRR/MODIS)
- Greenness (AVHRR/MODIS)
- Albedo (AVHRR/MODIS)
- Lat/Ion Gaussian Lambert Conformal Mercator Polar Stereographic UTM
- CRTM CRTM2E CMEM HUT
- Direct
LIS Architecture
LIS Architecture

---

Sample Use Case Implementations:
- Noah
- CLM
- MOSAIC
- Eta
- SM2
- MesoNH
- Snowpack
- GEOS-5
- NLDAS
- GEWEX
- CMORPH
- ECWMF
- AGRMET
- GEOS-TRMM
- Landcover (LMDI, USGS, MODIS)
- Soil (FAO, STATSGO)
- Topography (USGS)
- LAI (AVHRR/MODIS)
- Greenness (AVHRR/MODIS)
- Albedo (AVHRR/MODIS)
- LAI: Gaussian
- Lambert Conformal Mercator
- Polar Stereographic
- UTM
- CRIM
- CRIM2E
- CPDM
- MCM
- Direct Insertion
- EnKF
- LM
- GEA
- SCE-UA
- MC-SIM
- RW1
- MCMC
- DEMC
- Landslide models
- Crop models
- NLDAS
- router
- CREST
Uncertainty Estimation (UE)

\[ \theta = (\theta_1, \theta_2, \ldots, \theta_N) \]

\[ X_t \]

\[ U_t \]

\[ Y_t \]

\[ M(X_t, U_t, \theta) \]
Decadal Survey Era goals w/ regards to uncertainty?

• Capture important sources
  – Stochastic behavior (forcings, errors)
  – Physical model time-invariants (properties)
  – Statistical parameters of stochastic model

• Update with observations ("learning")
Current-Decadal Observations

- Vegetation/Carbon (Landsat, AVHRR, MODIS, DESDynI, ICESat-II, HySpIRI, LIST, ASCENDS)
- Surface soil moisture (SMMR, TRMM, AMSR-E, SMOS, Aquarius, SMAP)
- Snow water equivalent (AMSR-E, SSM/I, SCLP, GCOM-W, MIS)
- Snow cover fraction (MODIS, VIIRS, MIS)
- Water surface elevation (Jason-2, SWOT)
- Land surface temperature (MODIS, AVHRR, GOES, ...)
- Terrestrial water storage (GRACE, GRACEII)
- Precipitation (TRMM, GPM)
- Radiation (CERES, CLARREO)
- LDAS
Decadal Survey Era goals w/ regards to uncertainty?...

- Capture important sources
  - Stochastic behavior (forcings, errors)
  - Physical model time-invariants (properties)
  - Statistical parameters of stochastic model
- Update with observations (“learning”)
- Estimate worth of data
- Estimate worth of data yet to be collected (mission simulation experiment)
Uncertainty Estimation (UE) Case Study

- \( \theta_s \): porosity
- \( K_s \): saturated hydraulic conductivity
- \( \psi_s \): saturated matric potential
- \( b \): pore size distribution index

### Parameters
- \([\theta_s, \psi_s, K_s, b]\)

### Walnut Gulch Station Data

### Noah Land Surface Model

### Observations
- Water and Energy Fluxes, Soil Moisture and Temperature profiles, Land surface states
- Observations (Fluxes, soil moisture, snow, skin temperature)

### Noah soil moisture

### PBMR soil moisture observations
Cosby et al. (1984)

TABLE 3. Means and Standard Deviations for the Four Hydraulic Parameters in Each Textural Class

<table>
<thead>
<tr>
<th>Class</th>
<th>n</th>
<th>Mean</th>
<th>S.D.</th>
<th>Mean</th>
<th>S.D.</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sandy loam</td>
<td>124</td>
<td>4.74</td>
<td>1.40</td>
<td>1.15</td>
<td>0.73</td>
<td>-0.13</td>
<td>0.67</td>
</tr>
<tr>
<td>Sand</td>
<td>14</td>
<td>2.79</td>
<td>1.38</td>
<td>0.84</td>
<td>0.56</td>
<td>0.82</td>
<td>0.39</td>
</tr>
<tr>
<td>Loamy sand</td>
<td>30</td>
<td>4.26</td>
<td>1.95</td>
<td>0.56</td>
<td>0.73</td>
<td>0.30</td>
<td>0.51</td>
</tr>
<tr>
<td>Loam</td>
<td>103</td>
<td>5.25</td>
<td>1.66</td>
<td>1.55</td>
<td>0.66</td>
<td>-0.32</td>
<td>0.63</td>
</tr>
<tr>
<td>Silt loam</td>
<td>394</td>
<td>5.33</td>
<td>1.72</td>
<td>1.88</td>
<td>0.38</td>
<td>-0.40</td>
<td>0.55</td>
</tr>
<tr>
<td>Sandy clay loam</td>
<td>104</td>
<td>6.77</td>
<td>3.39</td>
<td>1.13</td>
<td>1.04</td>
<td>-0.20</td>
<td>0.54</td>
</tr>
<tr>
<td>Clay loam</td>
<td>147</td>
<td>8.17</td>
<td>3.74</td>
<td>1.42</td>
<td>0.72</td>
<td>-0.46</td>
<td>0.59</td>
</tr>
<tr>
<td>Silty clay loam</td>
<td>325</td>
<td>8.72</td>
<td>4.33</td>
<td>2.17</td>
<td>0.58</td>
<td>-0.54</td>
<td>0.61</td>
</tr>
<tr>
<td>Sandy clay</td>
<td>16</td>
<td>10.73</td>
<td>1.54</td>
<td>0.99</td>
<td>0.56</td>
<td>0.01</td>
<td>0.33</td>
</tr>
<tr>
<td>Silty clay</td>
<td>43</td>
<td>10.39</td>
<td>2.47</td>
<td>1.51</td>
<td>0.84</td>
<td>-0.72</td>
<td>0.69</td>
</tr>
<tr>
<td>Light clay</td>
<td>148</td>
<td>11.55</td>
<td>3.93</td>
<td>1.67</td>
<td>0.59</td>
<td>-0.86</td>
<td>0.62</td>
</tr>
<tr>
<td>All classes</td>
<td>1448</td>
<td>7.22</td>
<td>3.86</td>
<td>1.59</td>
<td>0.70</td>
<td>-0.42</td>
<td>0.64</td>
</tr>
</tbody>
</table>

$\theta_s$: porosity  
$K_s$: saturated hydraulic conductivity (m/s)  
$\psi_s$: saturated matric potential (m)  
$b$: pore size distribution index

$\theta = \theta_s \psi_s K_s b$

Passive (L-band) microwave remote sensing (using NASA's push broom microwave radiometer - PBMR) (Monsoon '90 experiment in SE Arizona)

Observation error described by normal independent distribution with std. dev.
Uncertainty estimation (UE)

In conducting uncertainty analysis, we acknowledge that many model fits have probability.

How do we correctly generate an ensemble of such solutions (An ensemble that reflects the unknown posterior distribution)?
Bayesian inference involves using observations to update/infer the probability that hypothesis (set of parameters) is true.

\[
p(\theta_i | y) = \frac{p(y | \theta_i)p(\theta_i)}{\int_\theta p(y | \theta)p(\theta)}
\]

- \(\theta\) Array of unobservable, uncertain model parameters
- \(\theta_i\) A particular model “fit” of \(\theta\)
- \(y\) Observations

Exceptionally computationally expensive to evaluate with standard integration methods
How is the likelihood evaluated?

- Common likelihood model: independent, normally distributed residuals: Zero mean ($\mu$) and constant standard deviation ($\sigma$)

\[
p(y|\theta) = \prod_{i=1}^{N} p(y_i|\theta)
\]

\[
p(y_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(M - y)^2}{2\sigma^2}}
\]

- This likelihood model is implicitly assumed in least squares PE:
  - "Results in maximum likelihood solution" (given many observations)
  - $\sigma$ falls out of analysis
- In contrast, here we consider $\sigma$ to be uncertain (along with hydraulic parameters)
Uncertainty estimation algorithms - MCMC

Markov Chain Monte Carlo (MCMC) method is a stochastic simulation that is based on constructing a Markov chain that has the desired distribution as its target (posterior).

Run sufficiently long, the frequency of the states of the chain is an estimate of the target distribution.

The main challenge with MCMC is in reducing the number of iterations required to converge to an equilibrium distribution.

Different implementations of the algorithm exist: Metropolis-Hastings, Gibbs, DE-MC, DREAM, etc.

Can be illustrated with simple random walk (RW) Metropolis algorithm...
MCMC - Generating candidate solutions

\[ p(\theta | y) \]

Current solution \( \theta_j \)

\( \theta \)

\( \theta^*_j + 1 \)

Proposed/candidate Solution
MCMC - Acceptance/Rejection Test

\[ R = \frac{p(\theta_j^* + 1 | y)}{p(\theta_j | y)} \]

if \( R > 1 \) then accept the proposed solution with 100\% probability, i.e., \( \theta_j + 1 = \theta_j^* + 1 \)
MCMC - Acceptance/Rejection Test
(building the Markov Chain)

\[ R = \frac{p(\theta^*_j | y)}{p(\theta_{j+1} | y)} \]

if \( R < 1 \)
Accept with probability \( R \)
else
Reject the solution
Uncertainty estimation MCMC algorithms

- **Random Walk (RW)** (Just demonstrated)
  - Most straightforward implementation of MCMC
  - Slow to converge

- **Differential Evolution Monte Carlo (DE-MC)** (Ter Braak et al., 2006)
  - Example of a population-based, adaptive algorithm
  - Borrows concepts from evolutionary search strategies (e.g., genetic algorithms)
  - We have developed a parallel implementation of the algorithm
Proposal distribution for the step is pre-specified and fixed.

Proposal distribution for the step determined by spacing of population → scale continually adjusts as posterior distribution is “learned”.
RW-MCMC and DEMC were implemented to take advantage of the efficiency of LIS ensemble runs:

- RW: 50 parallel, independent chains
- DE-MC: 50 population size. Less amenable to parallelization: 1 DE-MC iteration = cost of 2.5 MCMC iterations

Runs initialized with GA-found maximum probability solution to avoid long burn-in time

Convergence

DEMC converges in fewer LIS evaluations than RW MCMC.

...but with larger computational cost per iteration, MCMC performs marginally better than DEMC in wall clock time
**TABLE 3.** Means and Standard Deviations for the Four Hydraulic Parameters in Each Textural Class

<table>
<thead>
<tr>
<th>Class</th>
<th>n</th>
<th>Mean</th>
<th>S.D.</th>
<th>Mean</th>
<th>S.D.</th>
<th>Mean</th>
<th>S.D.</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sandy loam</td>
<td>124</td>
<td>4.74</td>
<td>1.40</td>
<td>1.15</td>
<td>0.73</td>
<td>-0.13</td>
<td>0.67</td>
<td>43.4</td>
<td>8.8</td>
</tr>
<tr>
<td>Sand</td>
<td>14</td>
<td>2.79</td>
<td>1.38</td>
<td>0.84</td>
<td>0.56</td>
<td>0.82</td>
<td>0.39</td>
<td>33.9</td>
<td>7.3</td>
</tr>
<tr>
<td>Loamy sand</td>
<td>30</td>
<td>4.26</td>
<td>1.95</td>
<td>0.56</td>
<td>0.73</td>
<td>0.30</td>
<td>0.51</td>
<td>42.1</td>
<td>7.2</td>
</tr>
<tr>
<td>Loam</td>
<td>103</td>
<td>5.25</td>
<td>1.66</td>
<td>1.55</td>
<td>0.66</td>
<td>-0.32</td>
<td>0.63</td>
<td>43.9</td>
<td>7.4</td>
</tr>
<tr>
<td>Silt loam</td>
<td>394</td>
<td>5.33</td>
<td>1.72</td>
<td>1.88</td>
<td>0.38</td>
<td>-0.40</td>
<td>0.55</td>
<td>47.6</td>
<td>5.4</td>
</tr>
<tr>
<td>Sandy clay loam</td>
<td>104</td>
<td>6.77</td>
<td>3.39</td>
<td>1.13</td>
<td>1.04</td>
<td>-0.20</td>
<td>0.54</td>
<td>40.4</td>
<td>4.8</td>
</tr>
<tr>
<td>Clay loam</td>
<td>147</td>
<td>8.17</td>
<td>3.74</td>
<td>1.42</td>
<td>0.72</td>
<td>-0.46</td>
<td>0.59</td>
<td>46.5</td>
<td>5.4</td>
</tr>
<tr>
<td>Silty clay loam</td>
<td>325</td>
<td>8.72</td>
<td>4.33</td>
<td>1.79</td>
<td>0.58</td>
<td>-0.54</td>
<td>0.61</td>
<td>46.4</td>
<td>4.6</td>
</tr>
<tr>
<td>Sandy clay</td>
<td>16</td>
<td>10.73</td>
<td>1.54</td>
<td>0.99</td>
<td>0.56</td>
<td>0.01</td>
<td>0.33</td>
<td>40.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Silty clay</td>
<td>43</td>
<td>10.39</td>
<td>4.27</td>
<td>1.51</td>
<td>0.84</td>
<td>-0.72</td>
<td>0.69</td>
<td>46.8</td>
<td>6.2</td>
</tr>
<tr>
<td>Light clay</td>
<td>148</td>
<td>11.55</td>
<td>3.93</td>
<td>1.67</td>
<td>0.59</td>
<td>-0.86</td>
<td>0.62</td>
<td>46.8</td>
<td>3.5</td>
</tr>
<tr>
<td>All classes</td>
<td>1448</td>
<td>7.22</td>
<td>3.86</td>
<td>1.59</td>
<td>0.70</td>
<td>-0.42</td>
<td>0.64</td>
<td>45.7</td>
<td>6.1</td>
</tr>
</tbody>
</table>

\( \theta_s \) : porosity  
\( K_s \) : saturated hydraulic conductivity (m/s)  
\( \psi_s \) : saturated matric potential (m)  
\( b \) : pore size distribution index

\[ \theta = \theta_s \psi_s K_s b \]

**Passive (L-band) microwave remote sensing (using NASA's push broom microwave radiometer - PBMR) (Monsoon '90 experiment in SE Arizona)**

**Observation error described by normal independent distribution with std. dev.**
Posterior distribution $p(\theta | y)$ (marginal distributions)

- Large shift in probability mass away from prior
- Correlations between parameters
- Lack of influence of the prior; small, flat in region of posterior
- Much uncertainty remains
- "Best" fits ignore uncertainty
- Uncertainty remains in error ($\sigma$)
Probabilistic simulation

- Many model fits are consistent with obs
- Value of remote sensing = reduction in uncertainty (and bias)
- Needed input to decision-makers
- Important within mission simulation experiment (OSSE) framework

a. The (time, soil moisture) points associated with all sample fits ($\theta$); curve is the median across fits.

b. The probabilistic forecast of soil moisture for the final time step

prior=blue  posterior=red
Conclusions

• Demonstrated new Uncertainty Estimation capability in LIS (now extending to land microwave emissivity modeling, applications (landslides))
• The uncertainty reduction is a measure of worth of remote sensing observations
• For decision-making (science and applications) need to understand the uncertainty
• Estimated models ignore uncertainty
• Ability to simulate multiple truths is important in mission simulation experiments (OSSEs)
Challenges

• Better priors (uncertainty reporting in soil texture–property relationships is terrible)
• Need more realistic forms of error models
• Continued advancements in algorithms & computing
• Integration with Data Assimil. systems
Conclusion: Capitalizing on algorithmic and computing advances, uncertainty analysis via Bayesian methods can provide needed input for scientists and decision-makers.