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

Ken Harrison[1,2], Sujay Kumar[1,3], Christa Peters-Lidard[1], Joe Santanello[1]

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.

Land Data Assimilation: Maximizing Mission Science and Applications Return

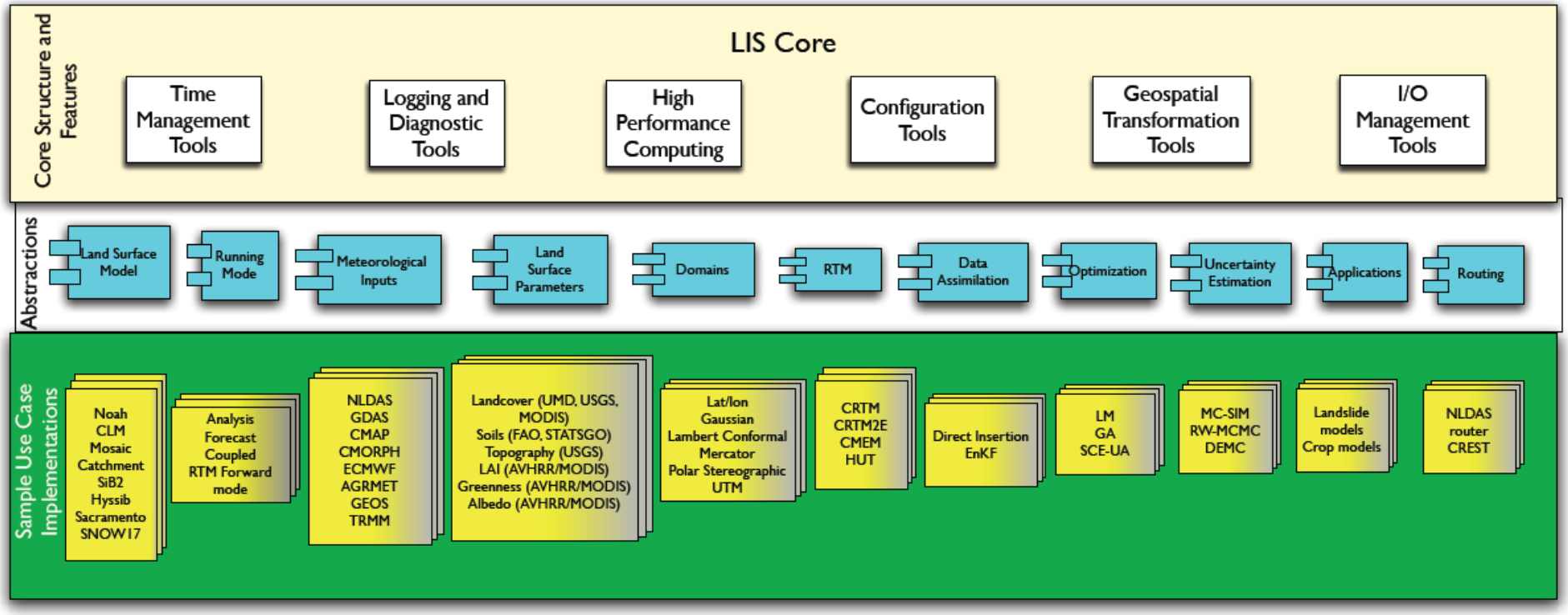
1. Observations

Uncertainty analysis: Scientists and decision-makers need a better understanding of the uncertainty in data products

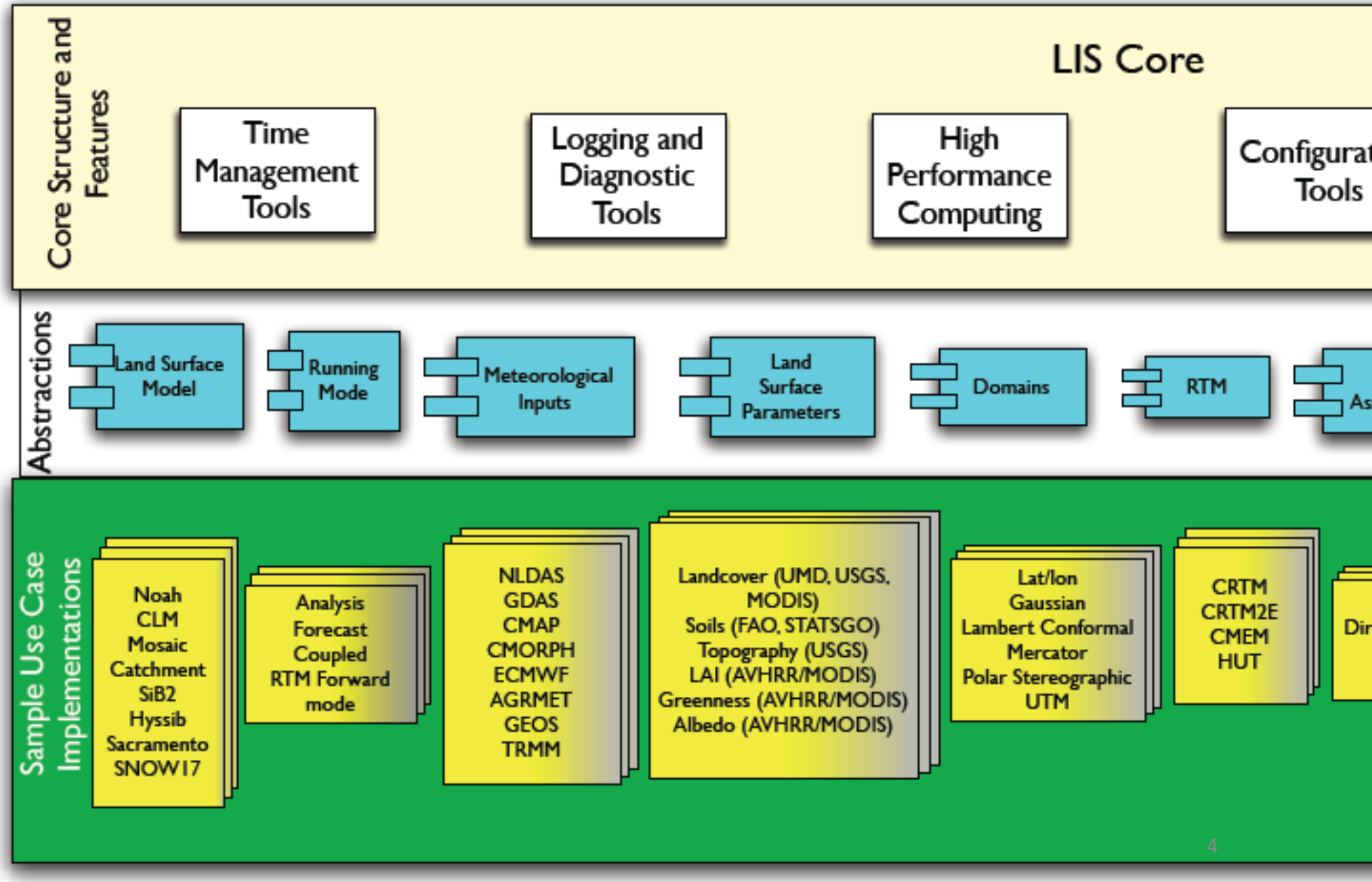
2. Modeling and Data Assimilation

3. Applications & Decisions

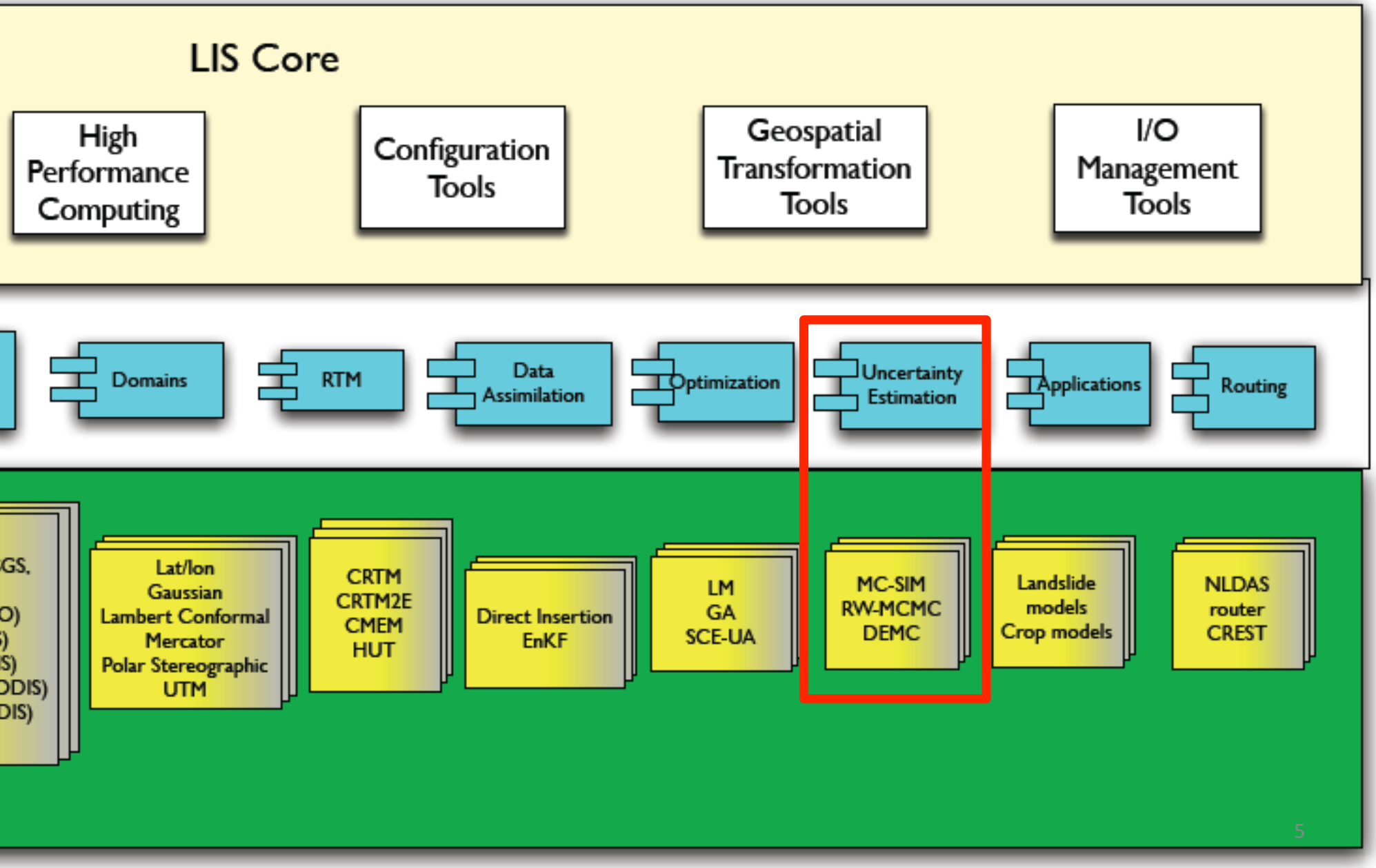
LIS Architecture



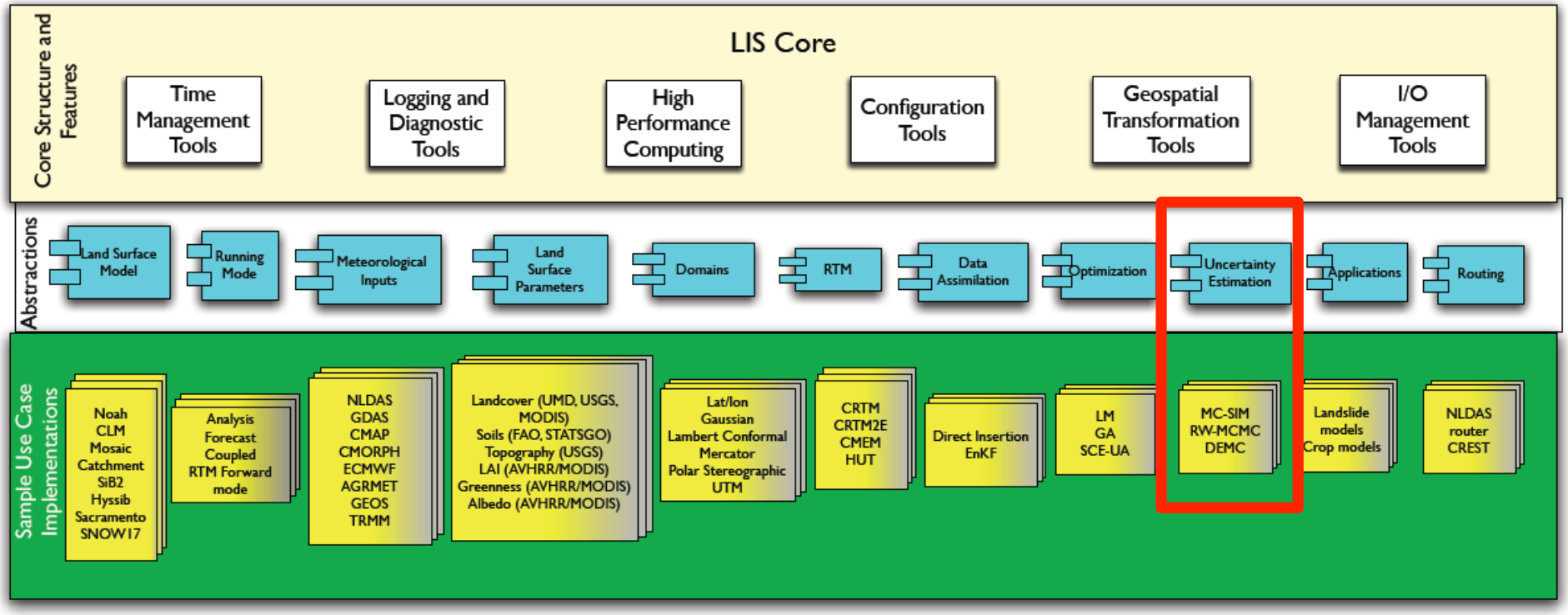
LIS Architecture



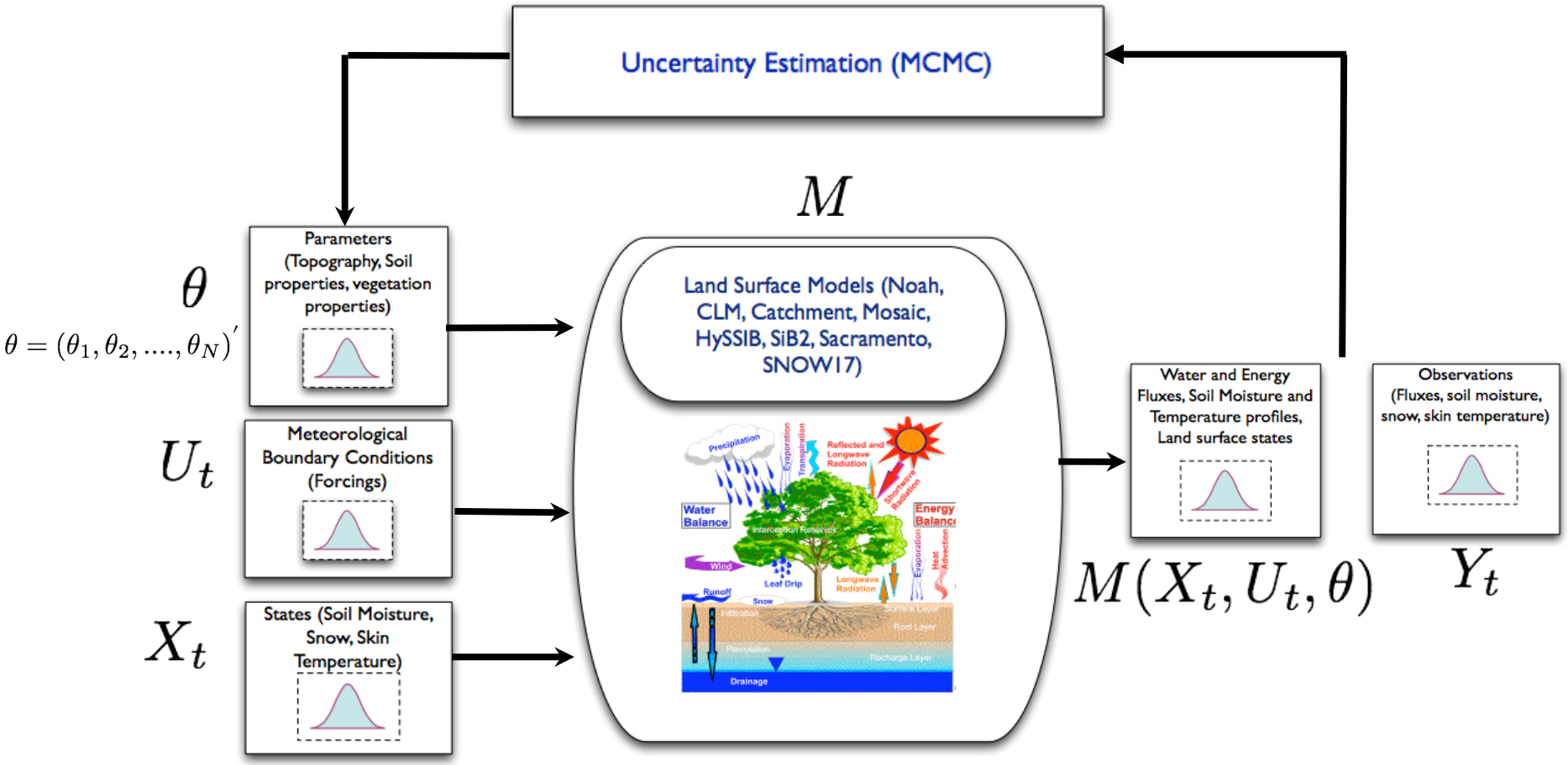
LIS Architecture



LIS Architecture



Uncertainty Estimation (UE)

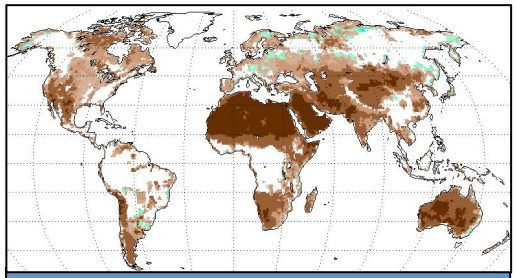


Decadal Survey Era goals w/ regards to uncertainty?

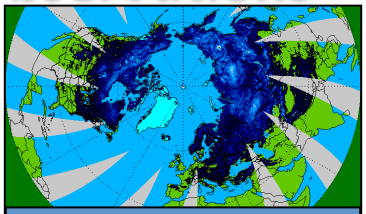
- Capture important sources
  - Stochastic behavior (forcings, errors)
  - Physical model time-invariants (properties)
  - Statistical parameters of stochastic model
- Update with observations ("learning")

Current-Decadal Observations

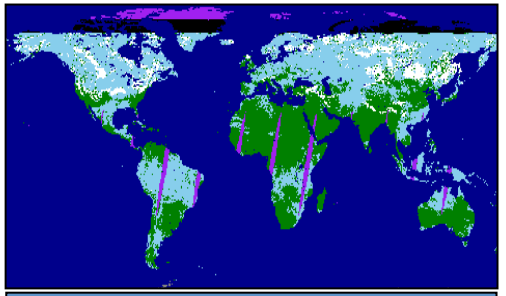
Vegetation/Carbon
(Landsat, AVHRR, MODIS, *DESDynI, ICESat-II, HyspIRI, LIST, ASCENDS*)



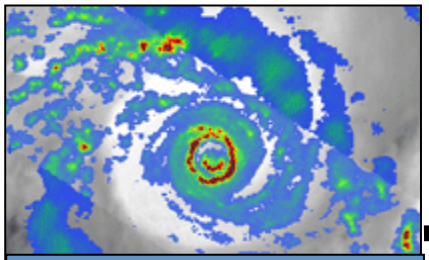
Surface soil moisture
(SMMR, TRMM, AMSR-E, SMOS, Aquarius, *SMAP*)



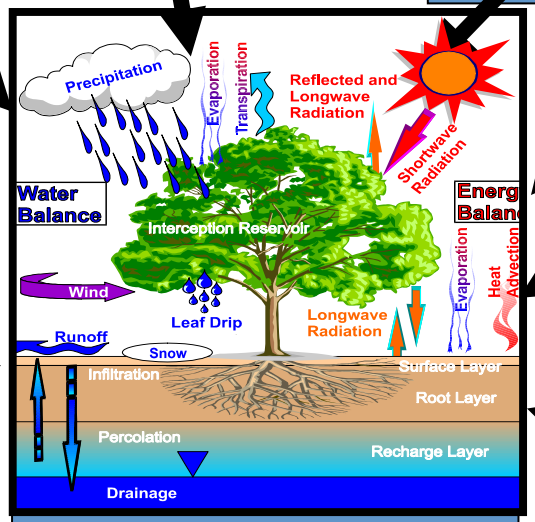
Snow water equivalent
(AMSR-E, SSM/I, *SCLP, GCOM-W, MIS*)



Snow cover fraction
(MODIS, *VIIRS, MIS*)

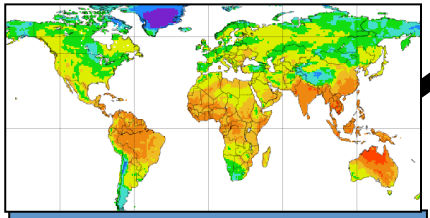


Precipitation
(TRMM, *GPM*)

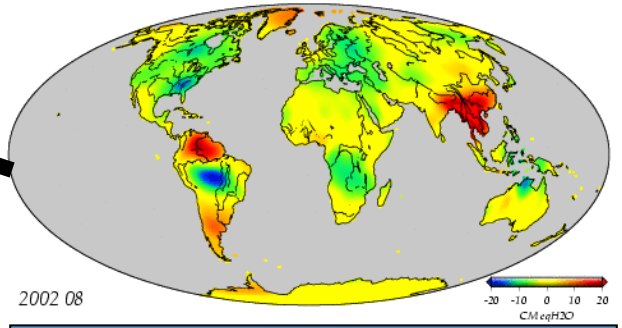


LDAS

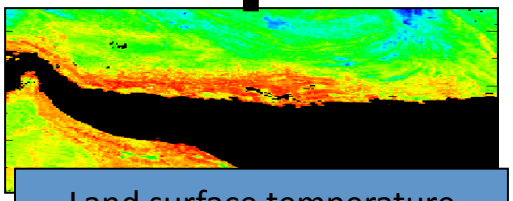
Water surface elevation
(Jason-2, *SWOT*)



Radiation
(CERES, *CLARREO*)



Terrestrial water storage
(GRACE,*GRACEII*)



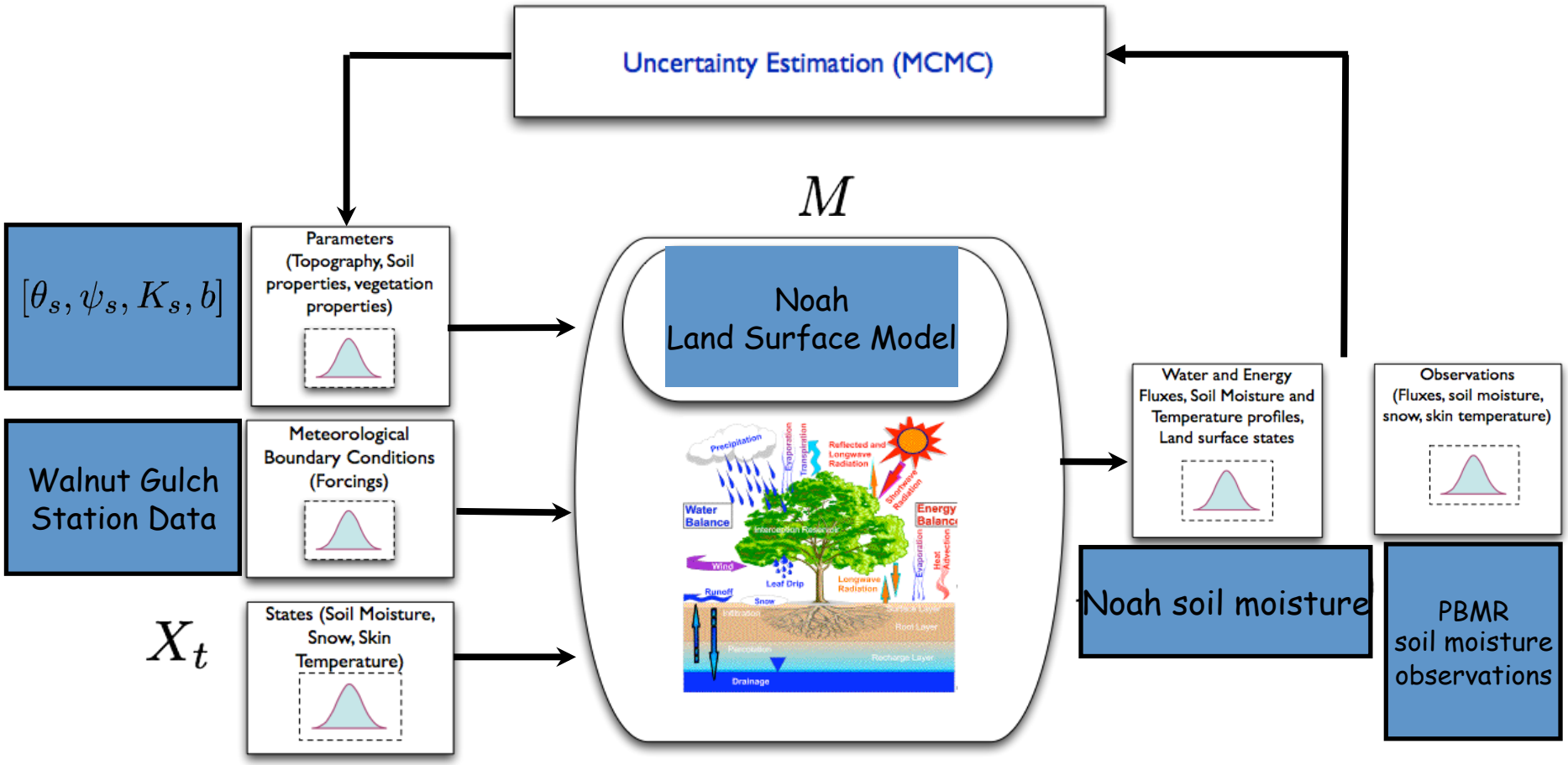
Land surface temperature
(MODIS, AVHRR,GOES,…)

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

θ_s : porosity K_s: saturated hydraulic conductivity
ψ_s: saturated matric potential b :pore size distribution index



Cosby et al. (1984)

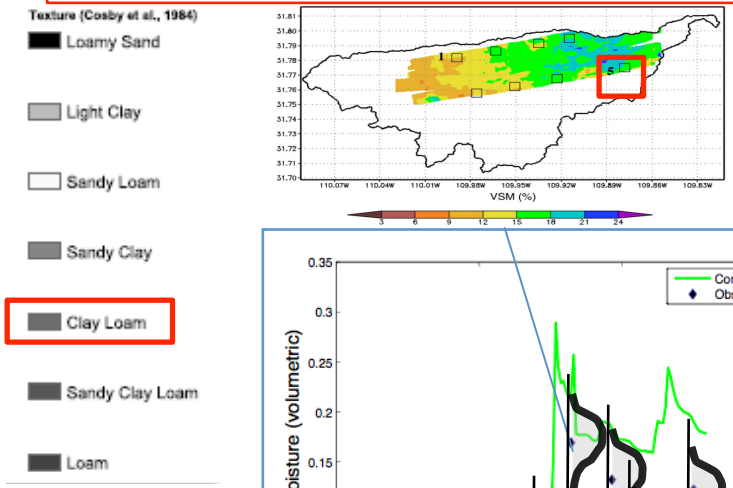
TABLE 3. Means and Standard Deviations for the Four Hydraulic Parameters in Each Textural Class

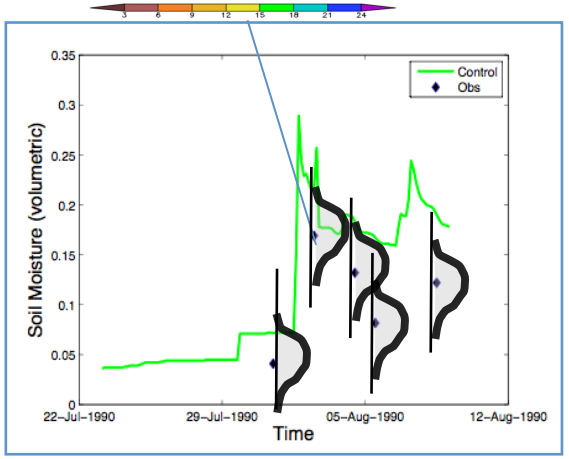
Class	n	b Mean	b S.D.	$\log \Psi_s$ Mean	$\log \Psi_s$ S.D.	$\log K_s$ Mean	$\log K_s$ S.D.	Θ_s Mean	Θ_s S.D.
Sandy loam	124	4.74	1.40	1.15	0.73	−0.13	0.67	43.4	8.8
Sand	14	2.79	1.38	0.84	0.56	0.82	0.39	33.9	7.3
Loamy sand	30	4.26	1.95	0.56	0.73	0.30	0.51	42.1	7.2
Loam	103	5.25	1.66	1.55	0.66	−0.32	0.63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

θ_s: porosity K_s: saturated hydraulic conductivity (m/s)

ψ_s: saturated matric potential (m) b :pore size distribution index

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





M

$\theta = \theta_s\ \psi_s\ K_s\ b$

θ

Parameters (Topography, Soil properties, vegetation properties)

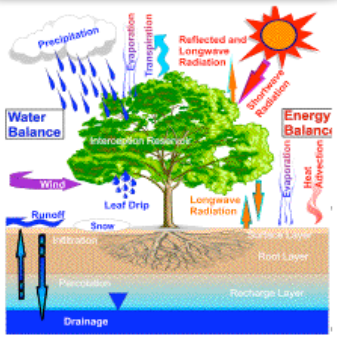
U_t

Meteorological Boundary Conditions (Forcings)

X_t

States (Soil Moisture, Snow, Skin Temperature)

Land Surface Models (Noah, CLM, Catchment, Mosaic, HySSIB, SiB2, Sacramento, SNOW17)



Water and Energy Fluxes, Soil Moisture and Temperature profiles, Land surface states

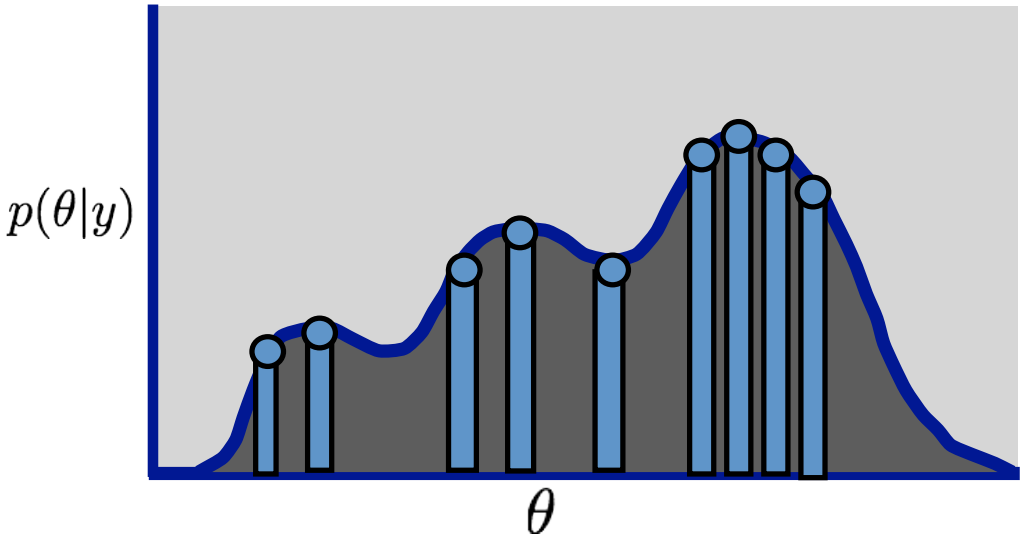
Observations (Fluxes, soil moisture, snow, skin temperature)

$M(X_t, U_t, \theta)$ Y_t

Observation error described by normal independent distribution with std. dev.

Uncertainty estimation (UE)

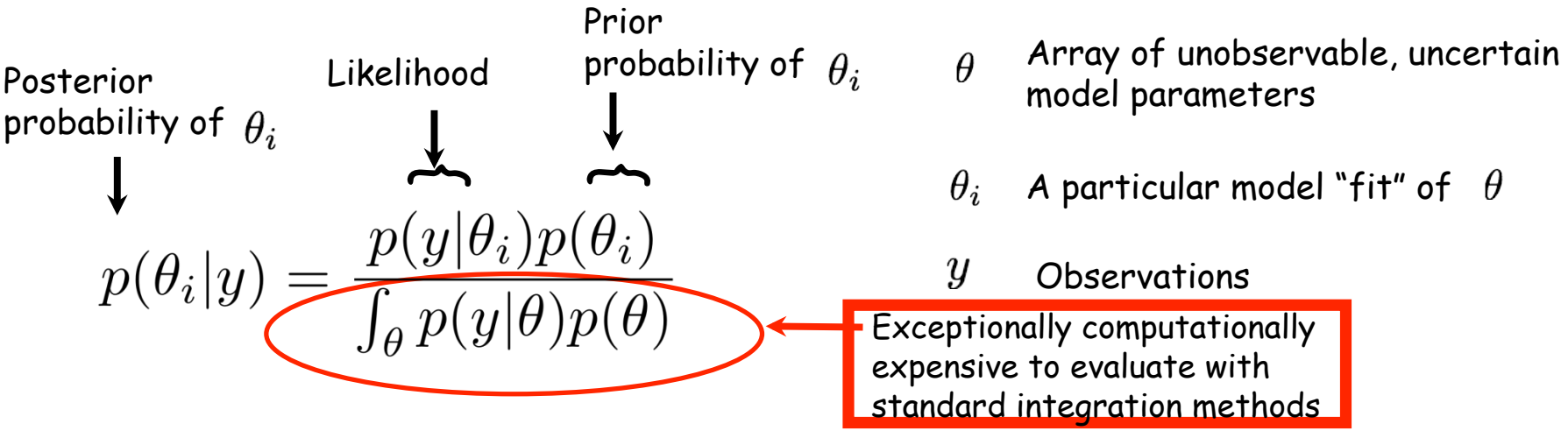
- In conducting uncertainty analysis, we acknowledge that many model fits have probability.
- How do we correctly generate an ensemble of such solutions (An ensemble that reflects the unknown posterior distribution)?



Bayesian Analysis

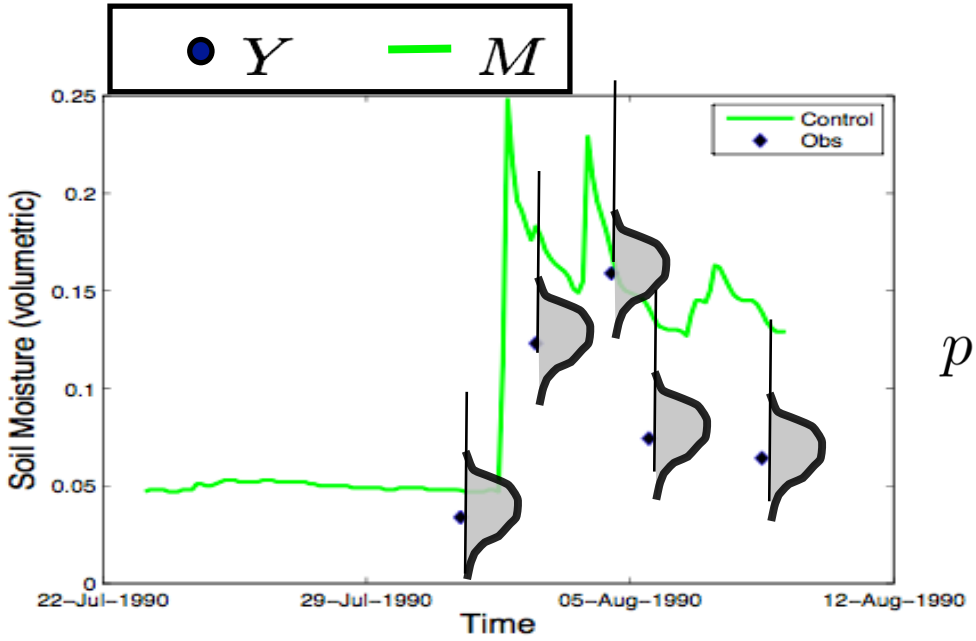
Bayes' rule

- Bayesian inference involves using observations to update/infer the probability that hypothesis (set of parameters) is true.



How is the likelihood evaluated?

•Common likelihood model: independent, normally distributed residuals: Zero mean ( ) and constant standard deviation (σ)



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

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

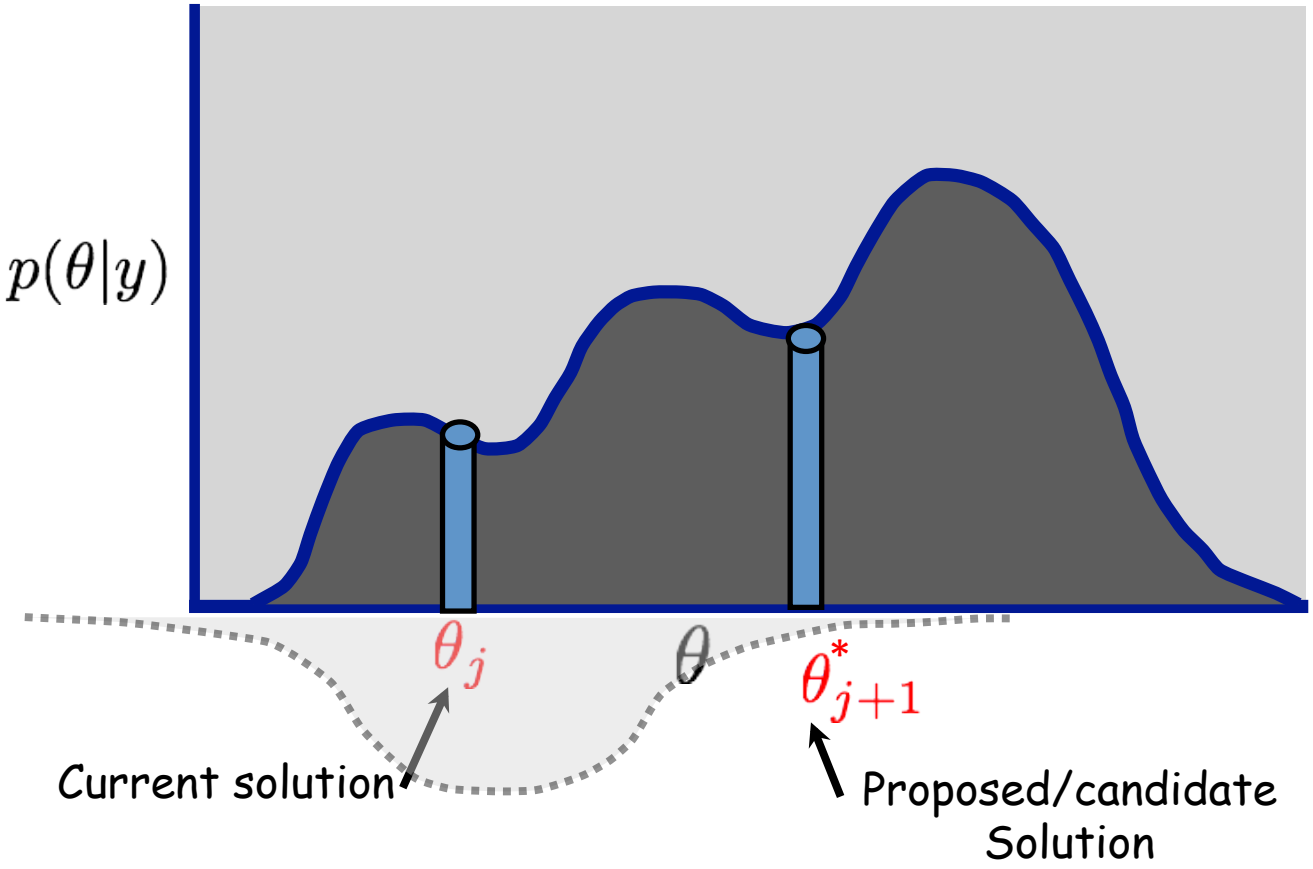
•This likelihood model is implicitly assumed in least squares PE:
- •"Results in maximum likelihood solution" (given many observations)
- •σ falls out of analysis

•In contrast, here we consider σ to be uncertain (along with hydraulic parameters)

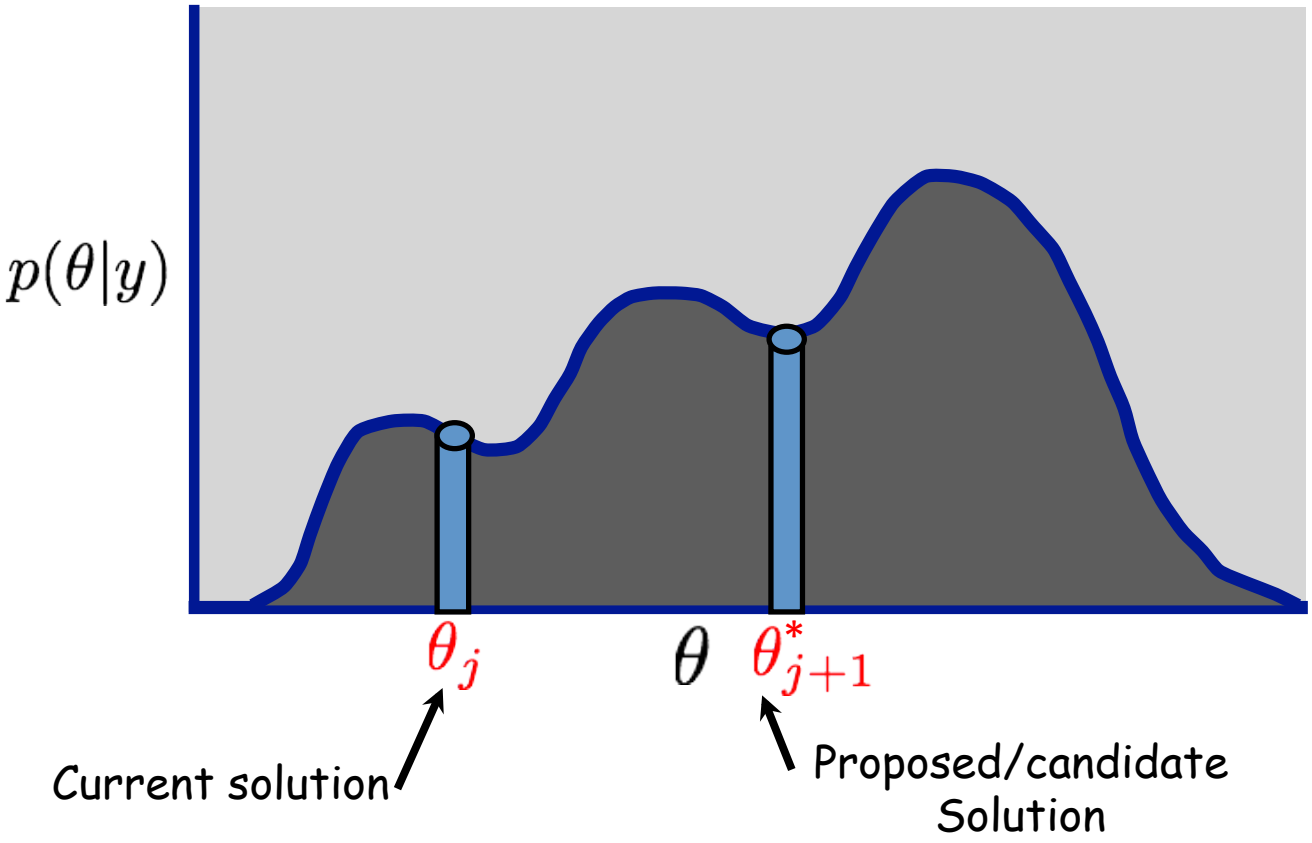
Uncertainty estimation algorithms - MCMC

- 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



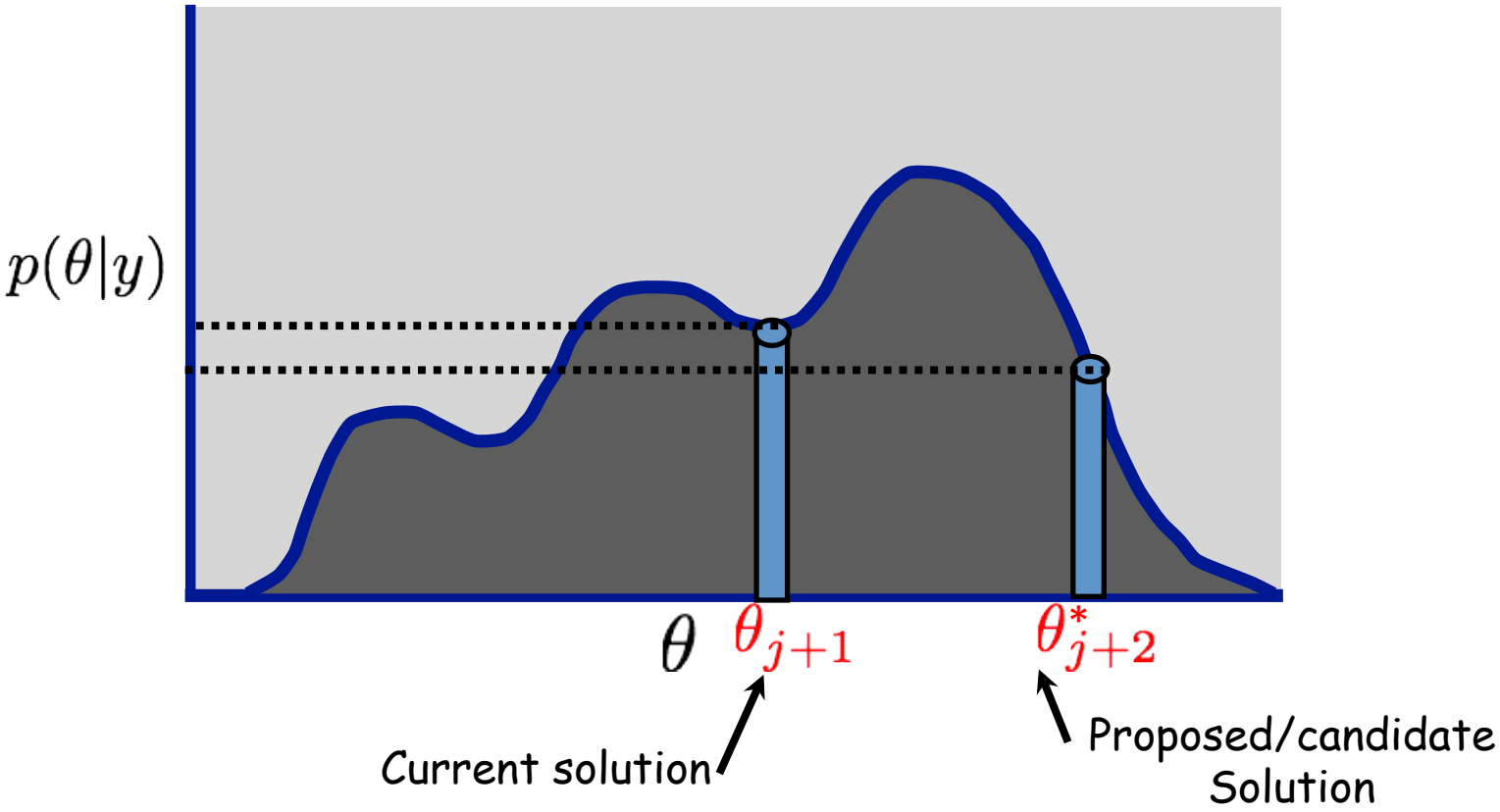
MCMC - Acceptance/Rejection Test



$$R = \frac{p(\theta^*_{j+1}|y)}{p(\theta_j|y)}$$

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)



$$R = \frac{p(\theta^*_{j+2}|y)}{p(\theta_{j+1}|y)}$$

if (R < 1)

Accept with probability R

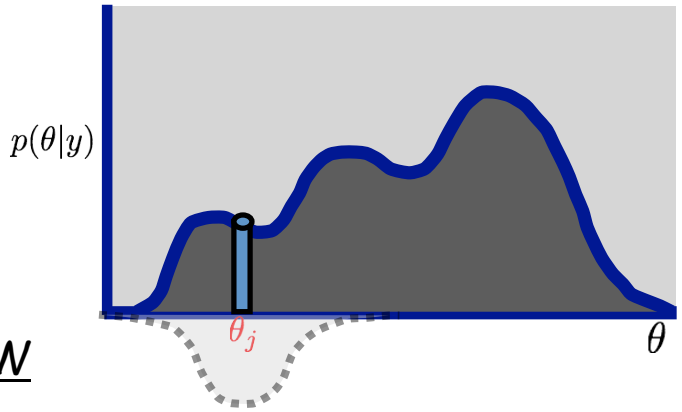
else

Reject the solution

Uncertainty estimation MCMC algorithms

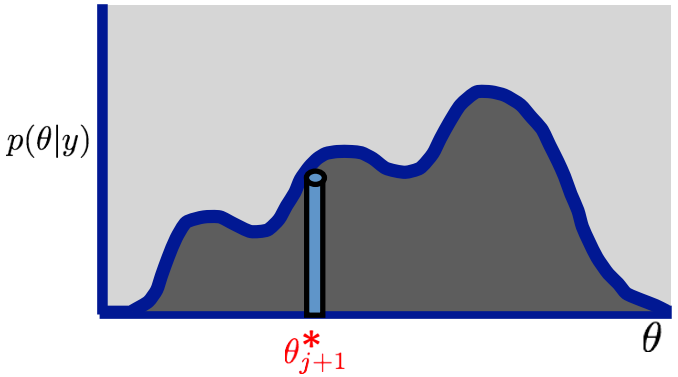
- Random Walk (RW) (Just demonstrated)
  - Most straightforward implementation of MCMC
  - Slow to converge
- Differential Evolution Monte Carlo (DE-MC) (Ter Braak et al., 2006)
  - Example of a population-based, adaptive algorithm
  - Borrows concepts from evolutionary search strategies (e.g., genetic algorithms)
  - We have developed a parallel implementation of the algorithm

RW MCMC

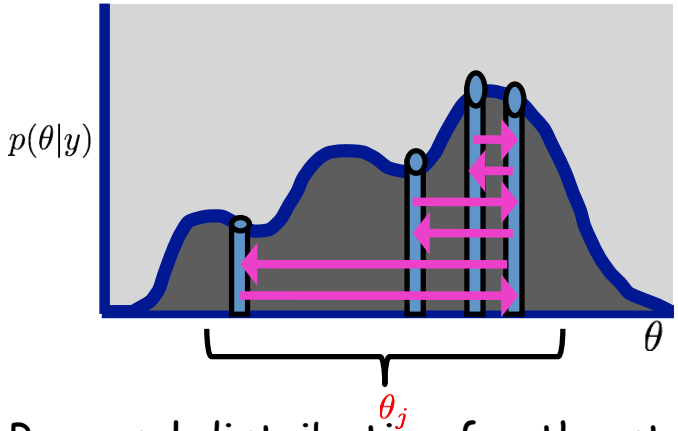


RW

Proposal distribution for the step is pre-specified and fixed

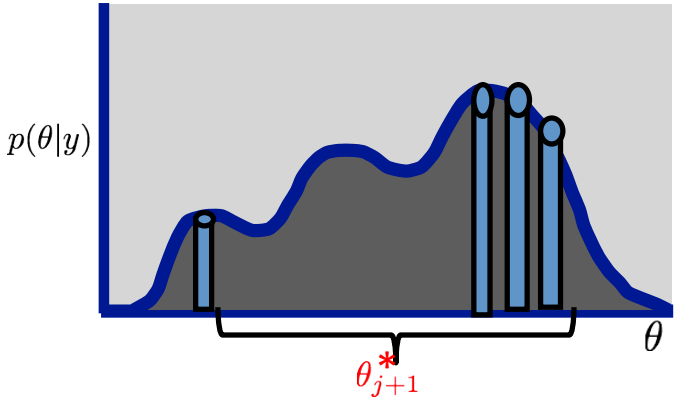


DE-MC



DE-MC

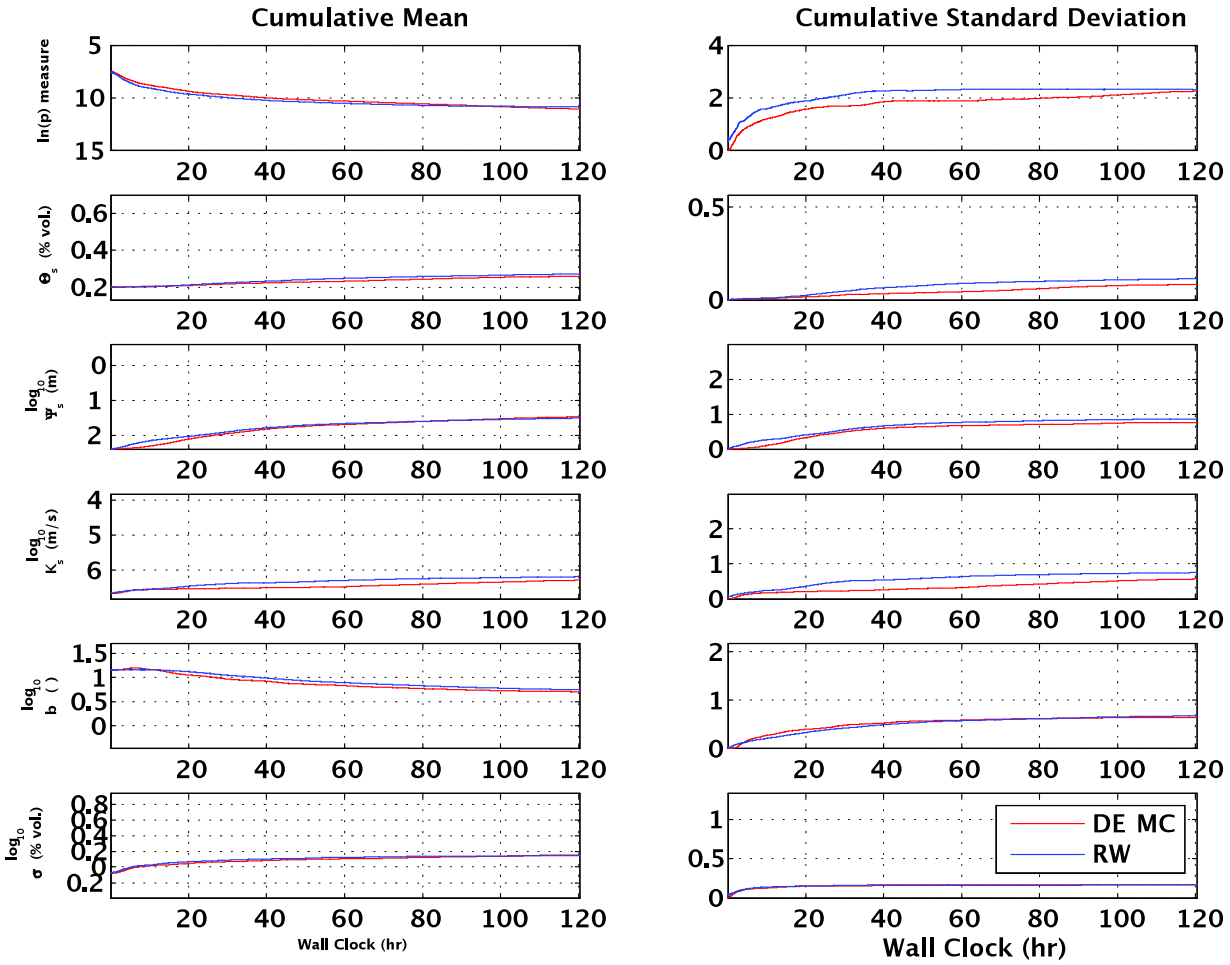
Proposal distribution for the step determined by spacing of population → scale continually adjusts as posterior distribution is "learned"



MCMC algorithm implementation

- 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

DEMC (red), RW (blue)



Cosby et al. (1984)

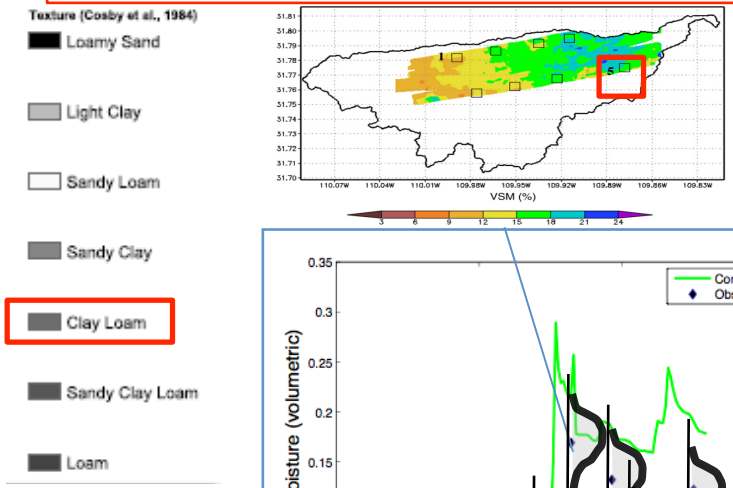
TABLE 3. Means and Standard Deviations for the Four Hydraulic Parameters in Each Textural Class

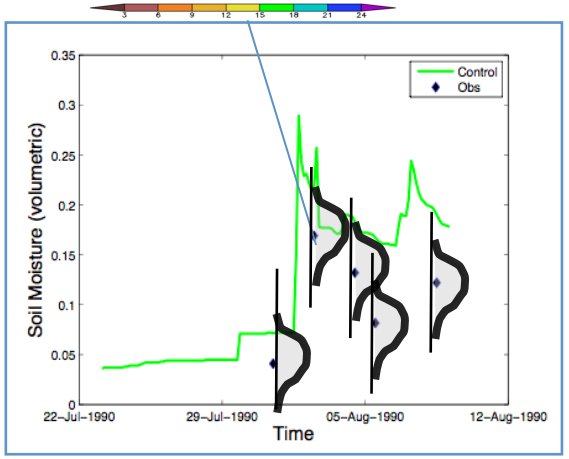
Class	n	b Mean	b S.D.	$\log \Psi_s$ Mean	$\log \Psi_s$ S.D.	$\log K_s$ Mean	$\log K_s$ S.D.	Θ_s Mean	Θ_s S.D.
Sandy loam	124	4.74	1.40	1.15	0.73	−0.13	0.67	43.4	8.8
Sand	14	2.79	1.38	0.84	0.56	0.82	0.39	33.9	7.3
Loamy sand	30	4.26	1.95	0.56	0.73	0.30	0.51	42.1	7.2
Loam	103	5.25	1.66	1.55	0.66	−0.32	0.63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

θ_s: porosity K_s: saturated hydraulic conductivity (m/s)

ψ_s: saturated matric potential (m) b :pore size distribution index

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





M

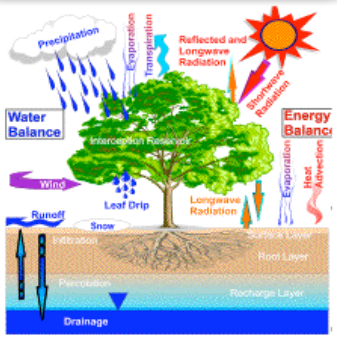
$\theta = \theta_s\ \psi_s\ K_s\ b$

Parameters (Topography, Soil properties, vegetation properties)

U_t Meteorological Boundary Conditions (Forcings)

X_t States (Soil Moisture, Snow, Skin Temperature)

Land Surface Models (Noah, CLM, Catchment, Mosaic, HySSIB, SiB2, Sacramento, SNOW17)



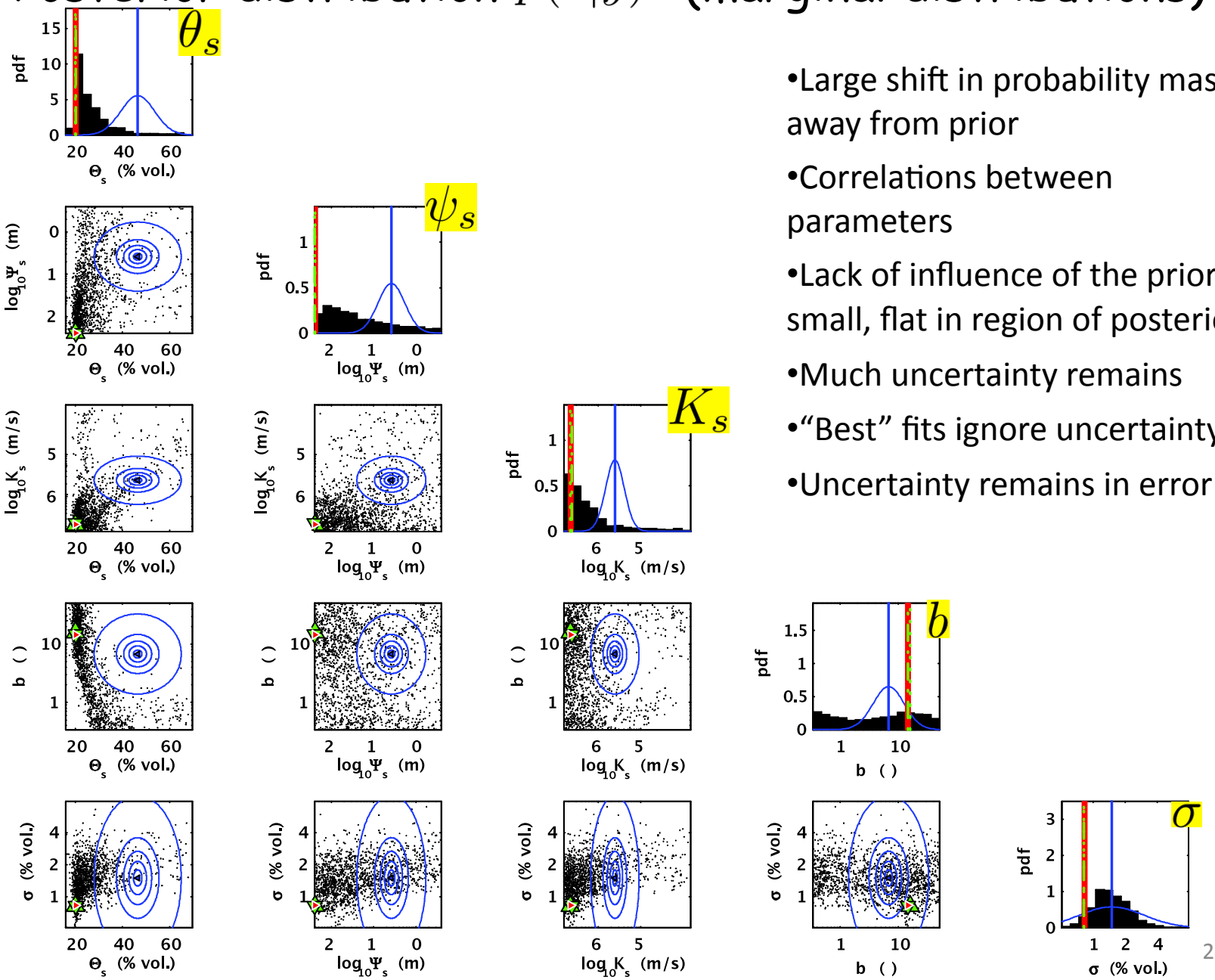
Water and Energy Fluxes, Soil Moisture and Temperature profiles, Land surface states

Observations (Fluxes, soil moisture, snow, skin temperature)

$M(X_t, U_t, \theta)$ Y_t

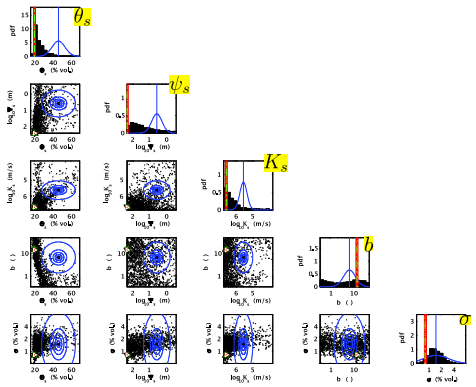
Observation error described by normal independent distribution with std. dev.

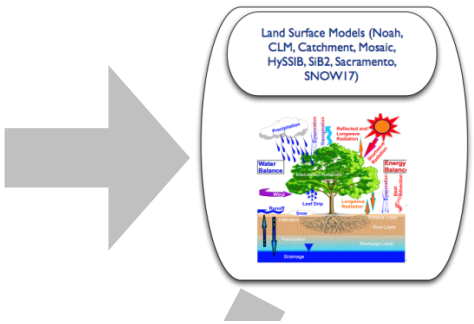
Posterior distribution $p(\theta|y)$ (marginal distributions)



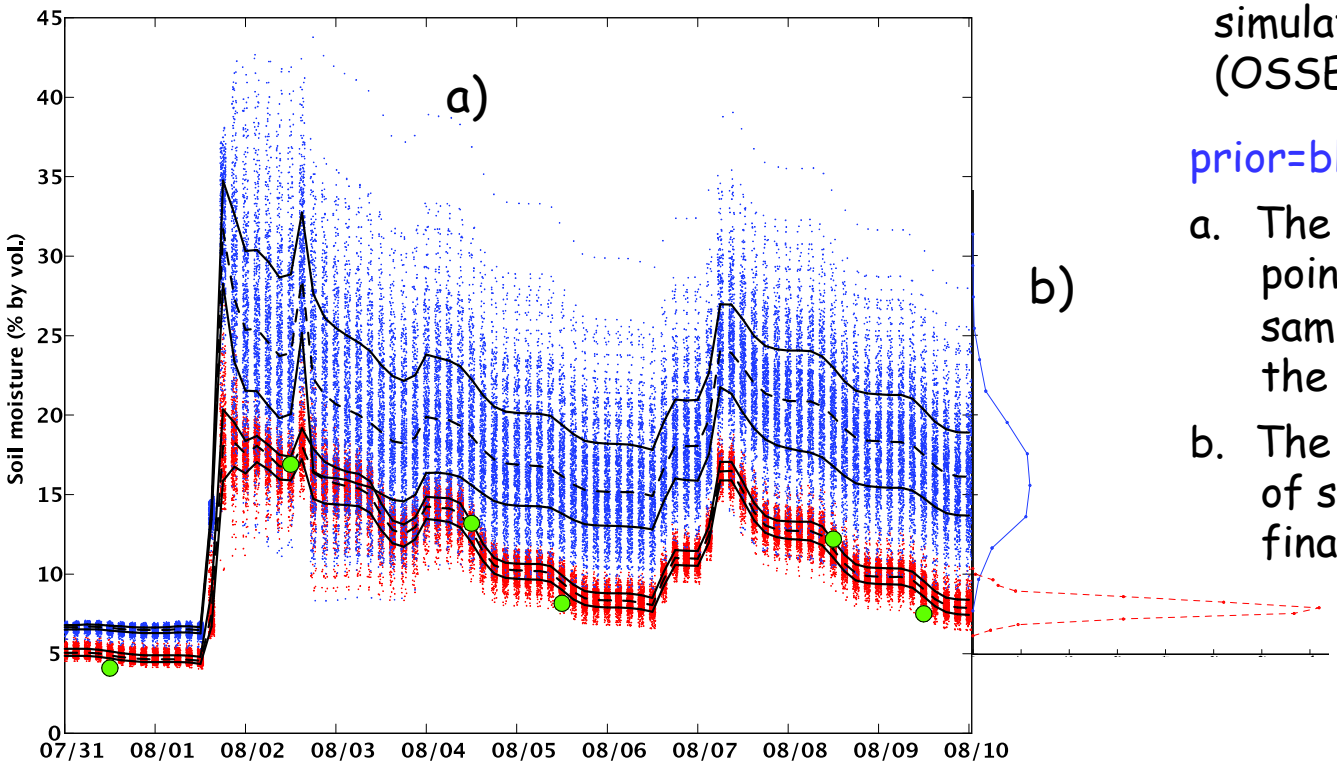
- Large shift in probability mass away from prior
- Correlations between parameters
- Lack of influence of the prior; small, flat in region of posterior
- Much uncertainty remains
- "Best" fits ignore uncertainty
- Uncertainty remains in error (σ)

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



prior=blue posterior=red

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

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

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)

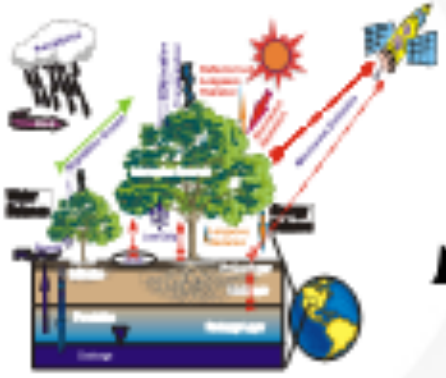
Challenges

- 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

Land Data Assimilation: Maximizing Mission Science and Applications Return

1. Observations

2. Modeling and Data Assimilation



3. Applications & Decisions

Conclusion: Capitalizing on algorithmic and computing advances, uncertainty analysis via Bayesian methods can provide needed input for scientists and decision-makers