An Integrated Infrastructure for land surface data assimilation using the NASA Land Information System (LIS): Description of Recent Enhancements

Sujay V. Kumar[a,b], Christa D. Peters-Lidard[b], Rolf H. Reichle[a,c]

[a]University of Maryland Baltimore County, Goddard Earth Sciences and Technology Center, Baltimore, MD

[b]Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD

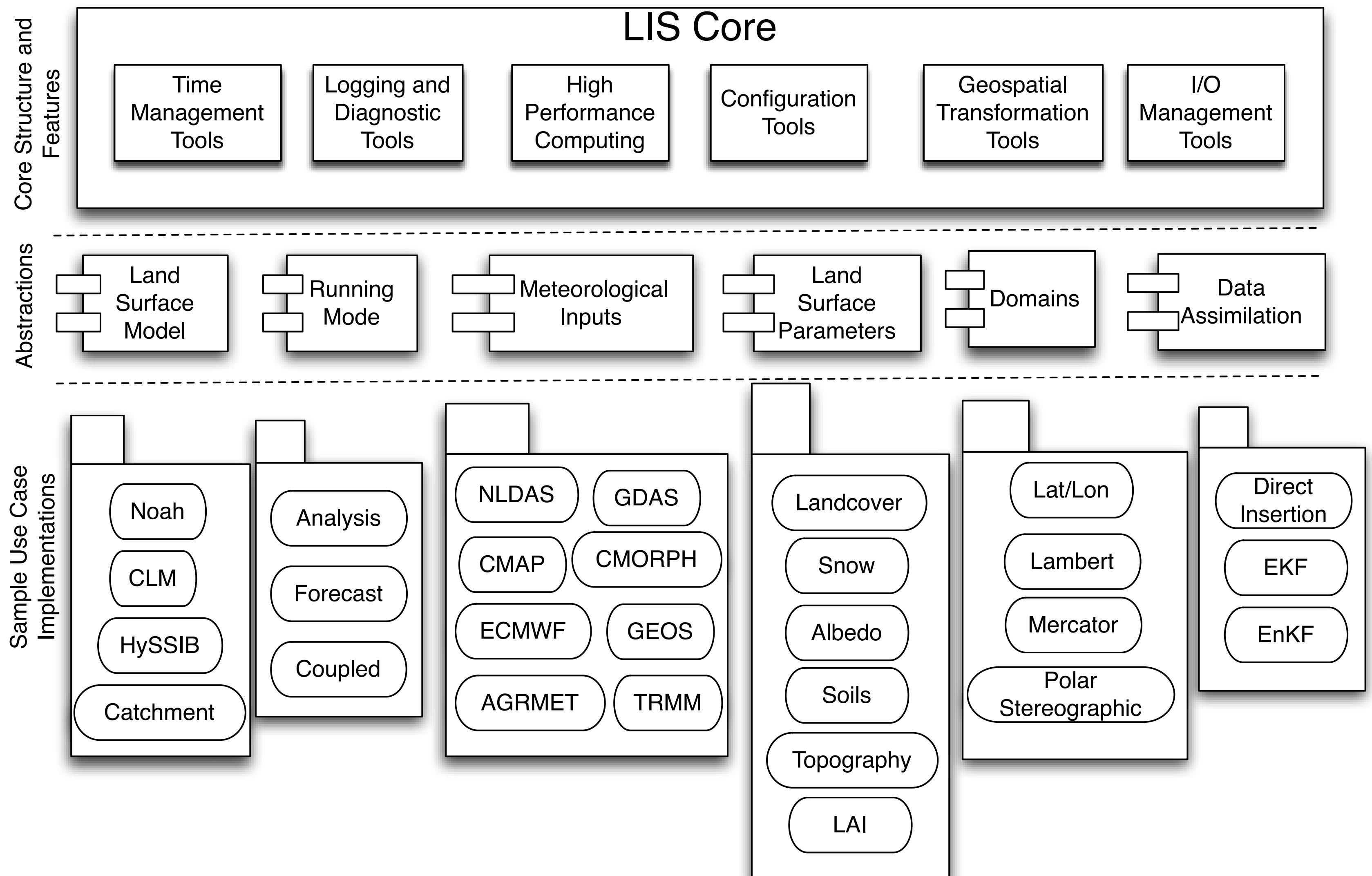
[c]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD

http://lis.gsfc.nasa.gov

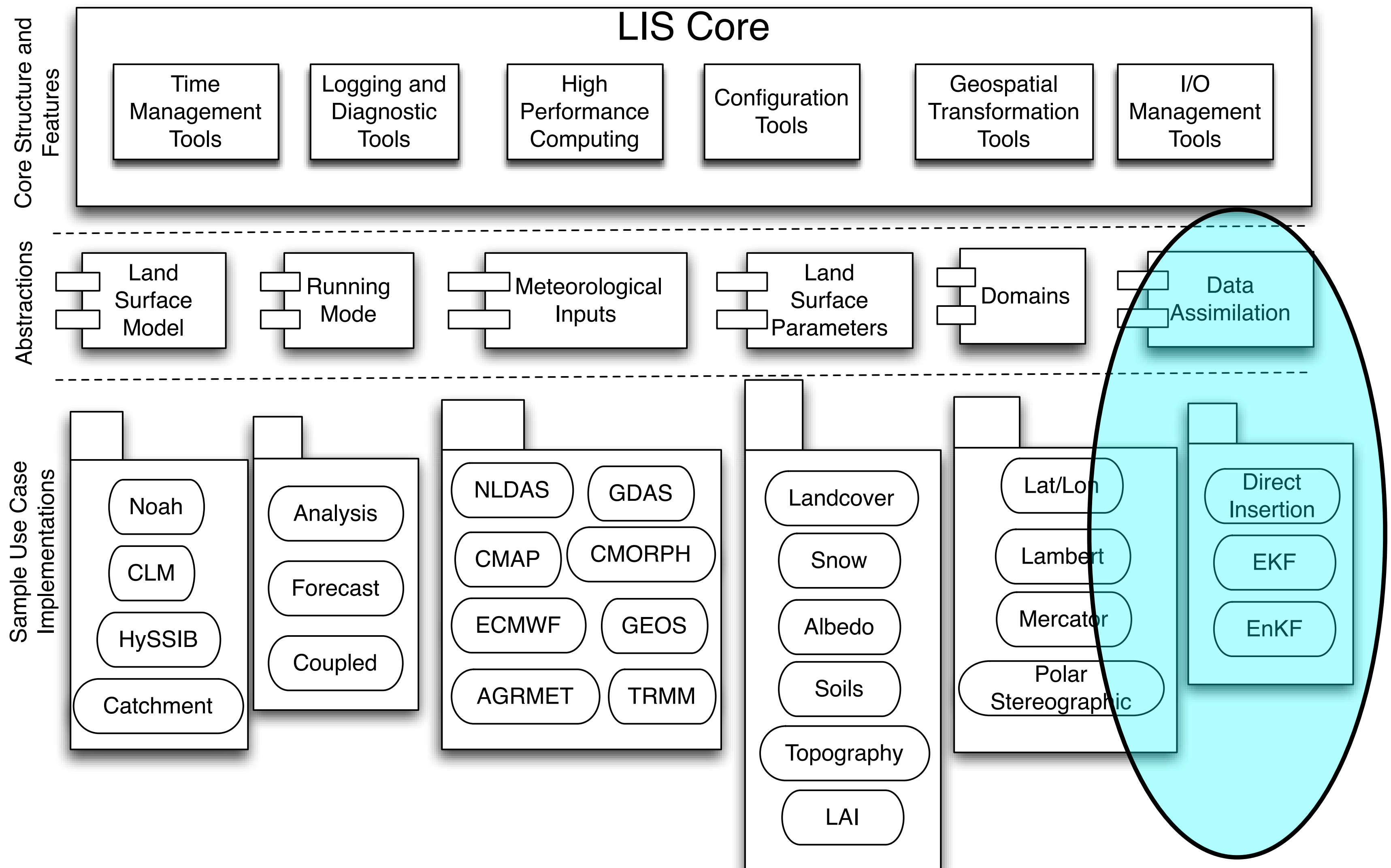
Outline

- LIS Data Assimilation - background and capabilities (Joint with GMAO, AFWA, USDA and NESDIS)
- Examples
  - Soil moisture assimilation
  - Snow assimilation
  - Skin Temperature assimilation

LIS Software Structure

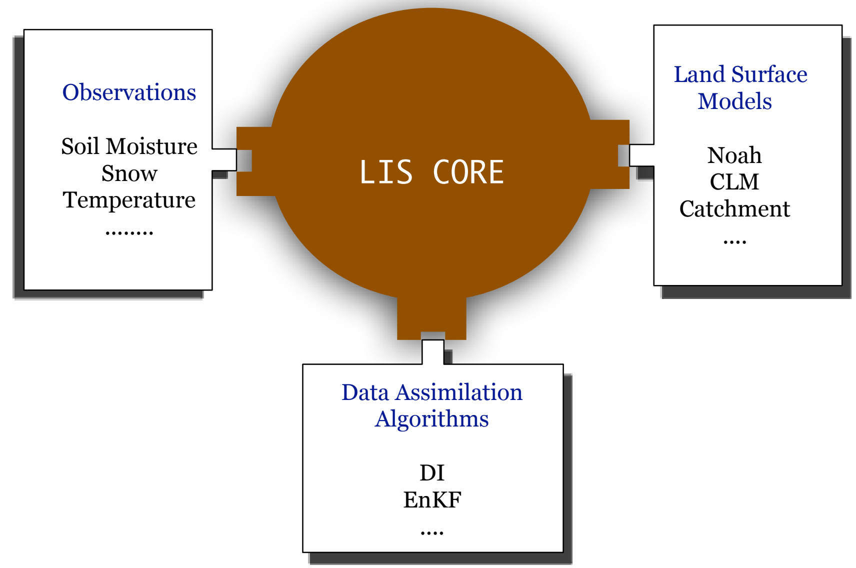


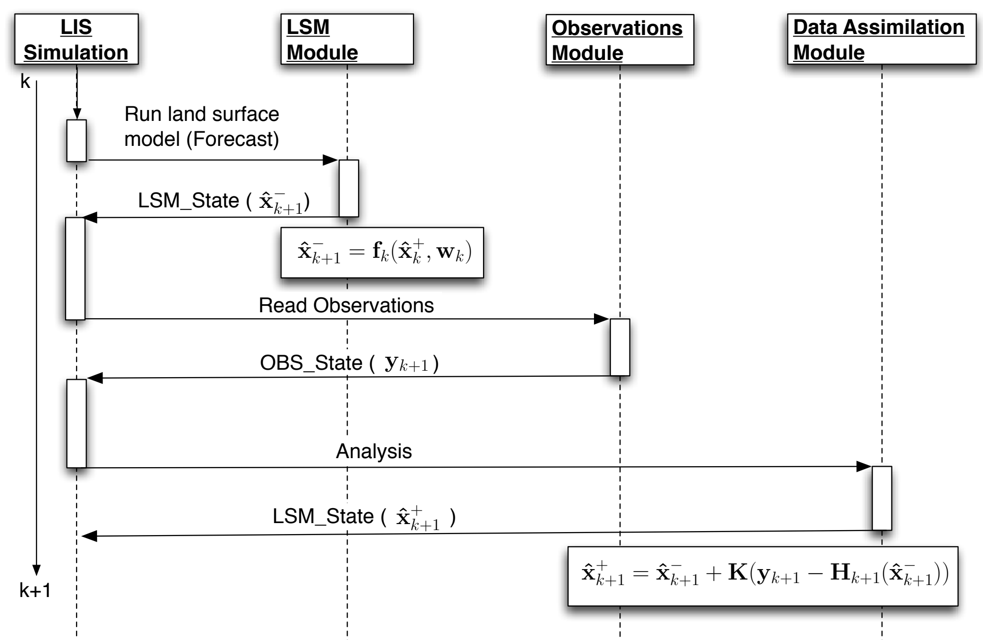
LIS Software Structure



Data Assimilation Support in LIS

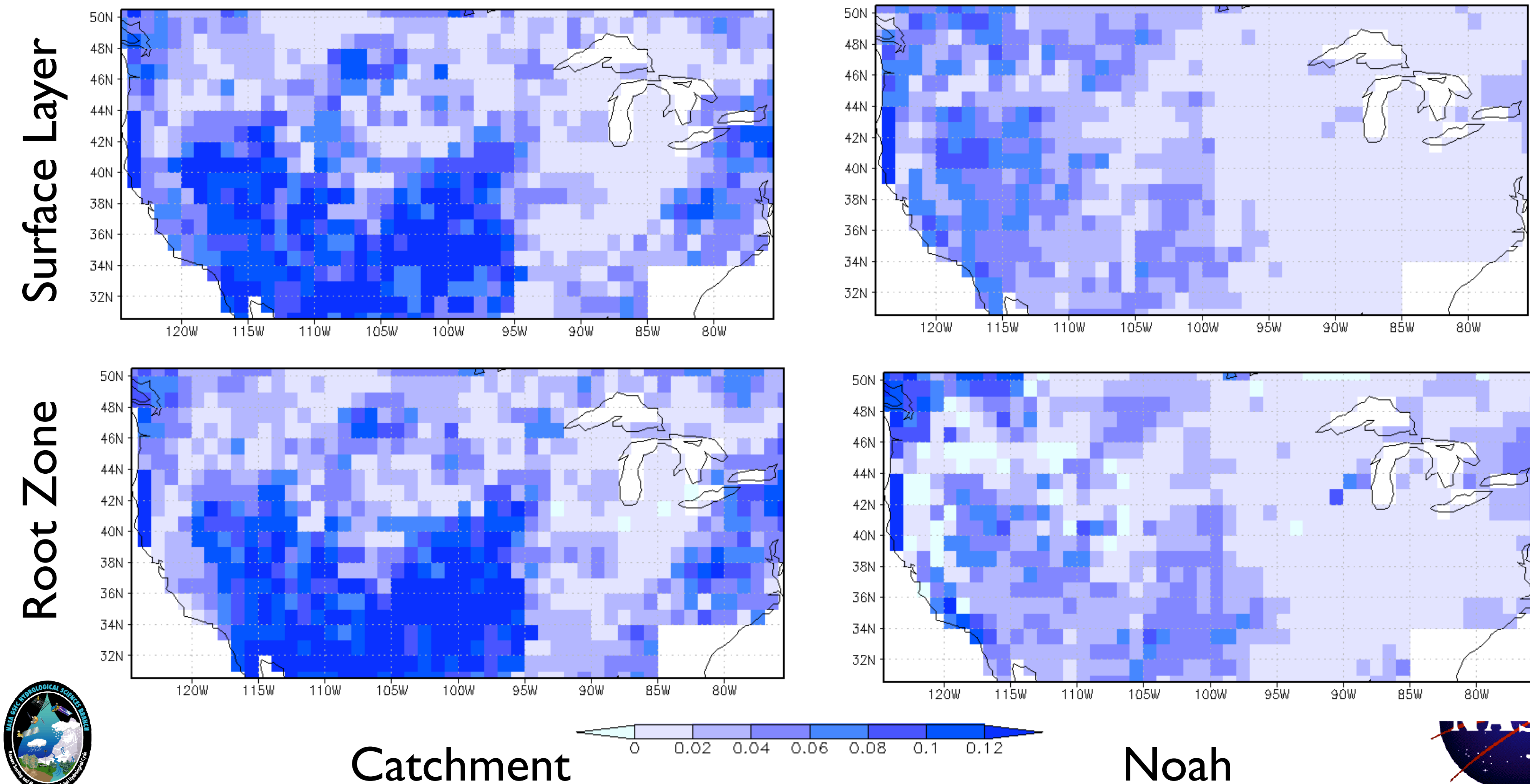
- Capabilities for sequential data assimilation
- Supports multiple LSMs, multiple observational types, multiple DA algorithms
- Computation supported by LIS high performance computing infrastructure





Surface Soil Moisture Assimilation

Improvement Metric (RMSE(OpenLoop) - RMSE(EnKF)) for soil moisture OSSEs



Impact of surface soil moisture assimilation on root zone soil moisture

Motivation

- Characterization of soil moisture stores is important for many real-world applications
  - Agricultural and water resources management, meteorological/climate studies, flood/drought forecasting
- Common approaches to soil moisture estimation
  - ground observations
  - land surface modeling
  - remote sensing observations
  - constraining model predictions with observations through data assimilation

Using remotely sensed soil moisture observations

- Remotely sensed observations (TRMM, AMSRE, SMOS, SMAP) are limited to observing a top thin portion of the soil column, down to depths between 1 cm and 5 cm.
- Large scale observations of root zone soil moisture do not exist
- Various computational methods have been used to retrieve the soil moisture profile from surface measurements
- Integrated use of data assimilation techniques and hydrologic models is an effective method (Reichle and Koster (2003, 2005), Walker et al (2001), Reichle (2007)).

Problem Statement

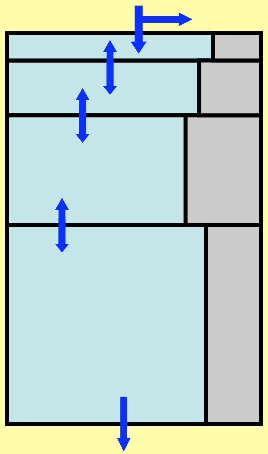
- How do the model representations impact the efficiency of soil moisture assimilation?
- How do the LSMs perform in a data assimilation system under various different representations of possible true land surface processes?

Approach

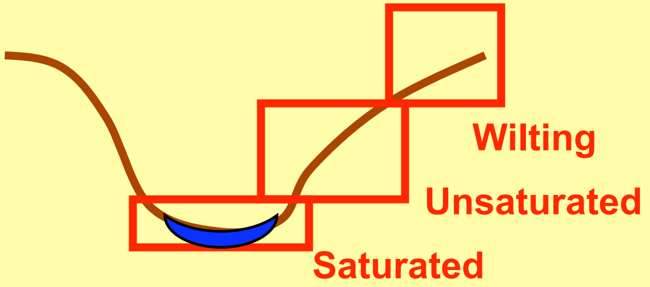
- We use four different LSMs to assimilate synthetic observations
  - NASA Catchment
  - Mosaic
  - Noah
  - CLM version 2.0
- Study is conducted using the recently developed Land Information System data assimilation system infrastructure
- LIS provides a uniform basis for this intercomparison

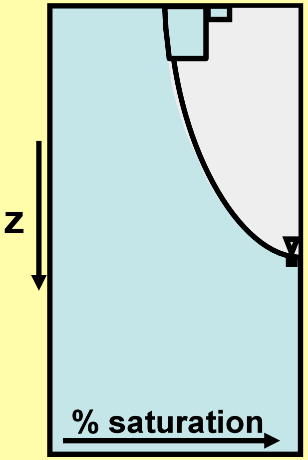
Land surface models

- Mosaic, Noah, and CLM
  - traditional, layer-based soil moisture dynamics
  - standard diffusion equation for unsaturated flow
  - free drainage
  - • Mosaic 3 layers (2, 148, 200 cm; Σ=3.5m)
  - • Noah 4 layers (10, 30, 60,100 cm; Σ=2.0m)
  - • CLM 10 layers (1.75, 3, 5, 8, 12, 34, 55, 91, 114 cm; Σ=3.2m)



- NASA Catchment LSM
  - catchments divided into sub-areas (saturated, unsaturated, and wilting)
  - soil moisture profile determined by deviations from equilibrium
  - dynamic water table
  - diagnose soil moisture content for
    - 2 cm surface layer
    - 100 cm root zone layer



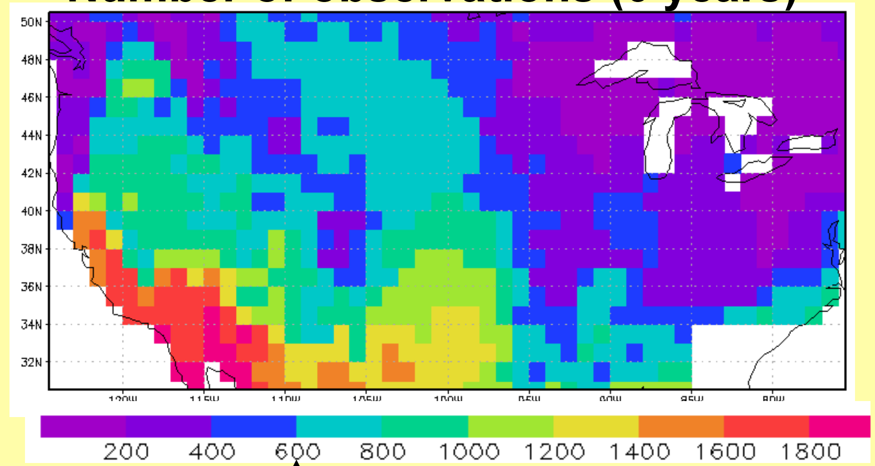


- For our analysis:
  - **surface** **≡ native (model-dependent) surface layer (1.75 – 10 cm)**
  - **root zone** **≡ top 100 cm of soil column (computed from available layers)**

Experiment setup

Domain: CONUS 1° by 1°
Period: Jan 1, 2001 to Jan 1, 2007
Forcing: GDAS

Number of observations (6 years)



Minimum N_obs

For each LSM:

1. **Generate ensemble integration**
   -- 12 members
   -- perturb soil moisture states *and* surface meteorological forcing inputs
   -- (model-specific) perturbations are correlated in time and in vertical dimension

2. **Pick one member as (synthetic) "truth" integration and generate synthetic observations**
   -- sample "truth" as with typical MW sensor (once daily; mask out dense vegetation, precipitation events, frozen soil, snow on the ground)
   -- add synthetic observation error

3. **Use ensemble mean as "Open Loop" (no assimilation) estimates**

Suite of Assimilation Experiments

- For each LSM, four assimilation integrations are conducted, using synthetic observations generated from each model

	TRUTH			
	CLSM	Mosaic	Noah	CLM
CLSM				
Mosaic				
Noah				
CLM				

- 16 assimilation experiments
- 4 Identical Twin Experiments and 12 Fraternal Twin experiments

Suite of Assimilation Experiments

- For each LSM, four assimilation integrations are conducted, using synthetic observations generated from each model

	TRUTH			
	CLSM	Mosaic	Noah	CLM
CLSM				
Mosaic				
Noah				
CLM				

- 16 assimilation experiments
- 4 Identical Twin Experiments and 12 Fraternal Twin experiments

Suite of Assimilation Experiments

- For each LSM, four assimilation integrations are conducted, using synthetic observations generated from each model

	TRUTH			
	CLSM	Mosaic	Noah	CLM
CLSM				
Mosaic				
Noah				
CLM				

- 16 assimilation experiments
- 4 Identical Twin Experiments and 12 Fraternal Twin experiments

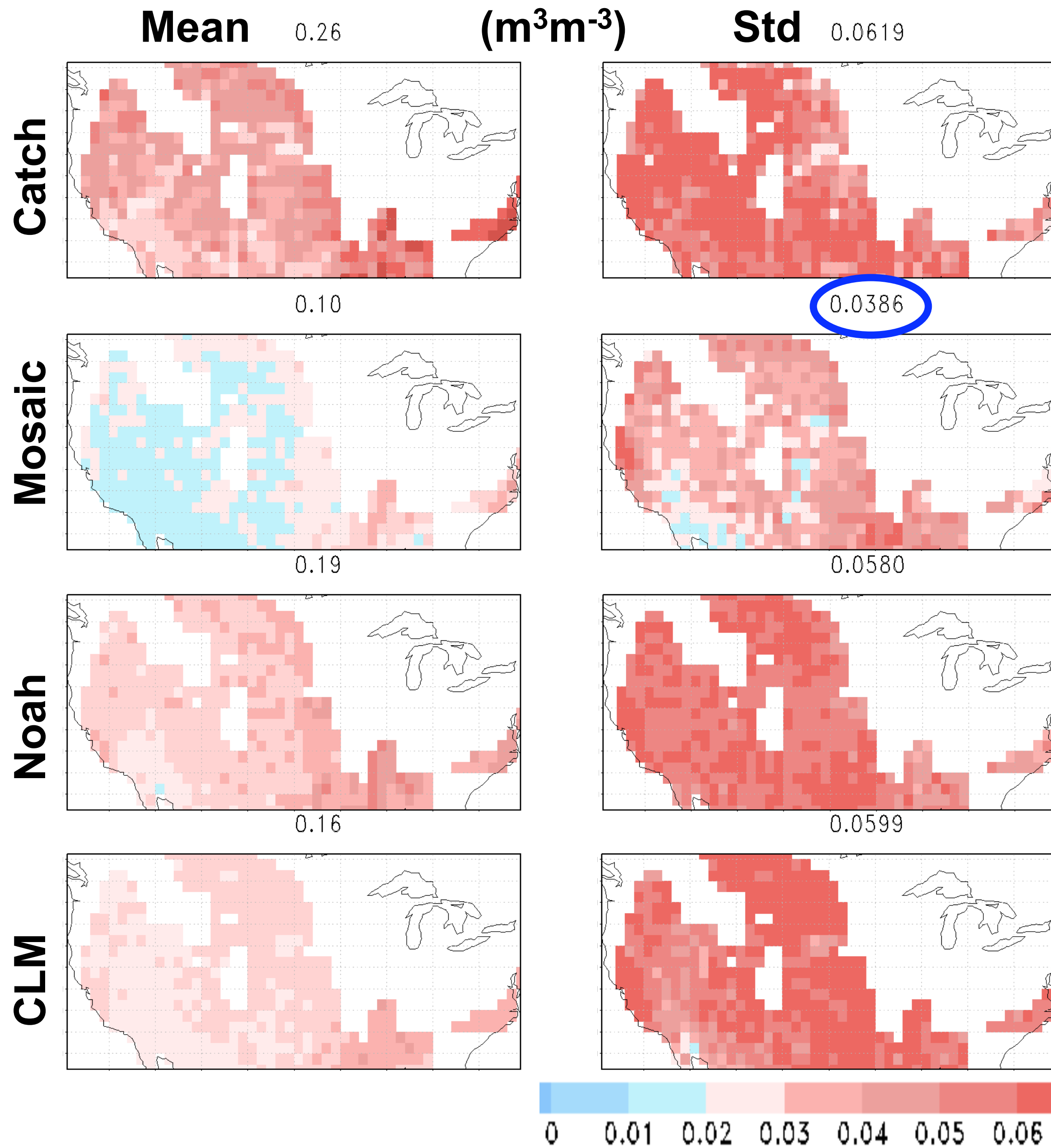
Surface soil moisture climatology

Climatology
≡ Mean and std of *raw* time series (masked to obs times and locations)

Soil moisture climatologies differ ➔

1.) Must match cdf's of synthetic obs and model prior to assimilation (*Reichle & Koster* 2004).

2.) Obs error std
Mosaic:
0.02 m^3m^{-3}
Cat/Noah/CLM: 0.03 m^3m^{-3}



Skill measure

RMSE not useful because of cdf matching.
Instead, remove seasonal cycle and compute

R = anomaly time series correlation coefficient

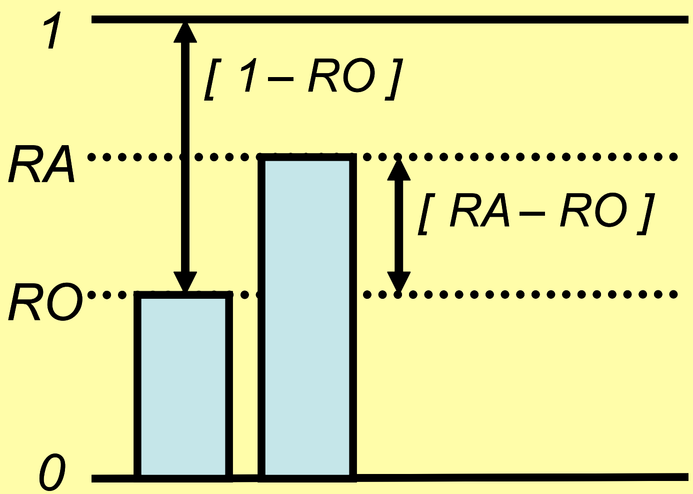
RO = skill (w.r.t. truth) of model ("open loop")
RA = skill (w.r.t. truth) of assimilation integration

RO depends on perturbation settings (error std, vertical corr)
RA depends on RO

Use "normalized information contribution"

$$NIC = [\, RA - RO \,] \,/\, [\, 1 - RO \,]$$

How much of the missing model skill is contributed by the assimilation?



Root zone soil moisture skill improvement from assimilation

NIC rzmc	Catch	Mos	Noah	CLM	Avg
Catch	0.71	0.54	0.36	0.38	0.50
Mosaic	0.55	0.69	0.31	0.33	0.47
Noah	0.43	0.43	0.36	0.26	0.37
CLM	0.11	0.21	0.10	0.45	0.22
	0.45	0.47	0.28	0.36	

1.) Average across rows (known truth physics):

Mosaic or Catchment "truth" is "easier" to estimate in data assimilation than Noah or CLM "truth".

2.) Average across columns (unknown truth physics):

Use of Catchment, Mosaic, and Noah in assimilation system is better than use of CLM.

Why???

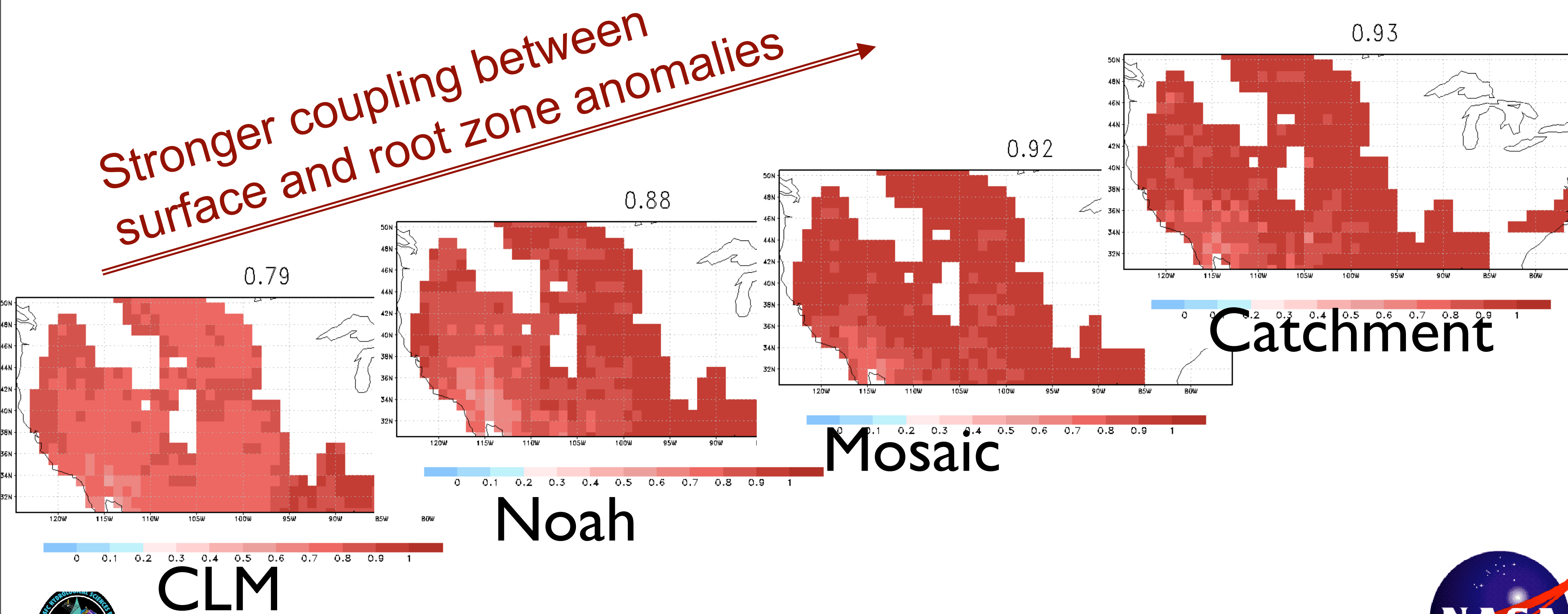
Vertical coupling strength (VCS)

For a given model integration (without assimilation), define

$$\text{VCS} \equiv R(\ sfmc,\ rzmc\ )$$

where R = anomaly time series correlation coefficient

Measures (time series) correlation between surface and root zone anomalies.



Gain correlation

In a given data assimilation integration, how much do surface obs contribute to root zone updates?

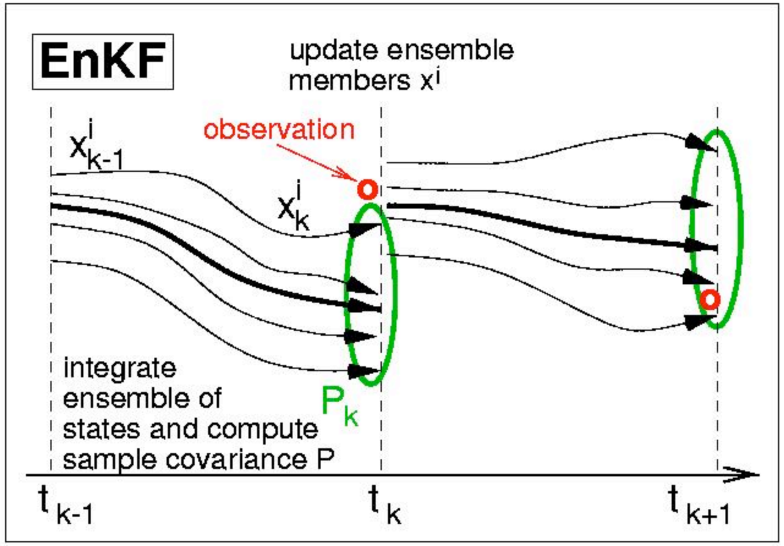
Increment = Gain * (Observation – Model forecast)

Root zone gain: *K_rz* ~ Cov(*sfmc*, *rzmc*)

"Gain correlation" = *K_rz* Std(*sfmc*) $^{-1}$ Std(*rzmc*) $^{-1}$

Note: **Cov = *ensemble* covariance**, Std = *ensemble* std

Depends on perturbations settings!

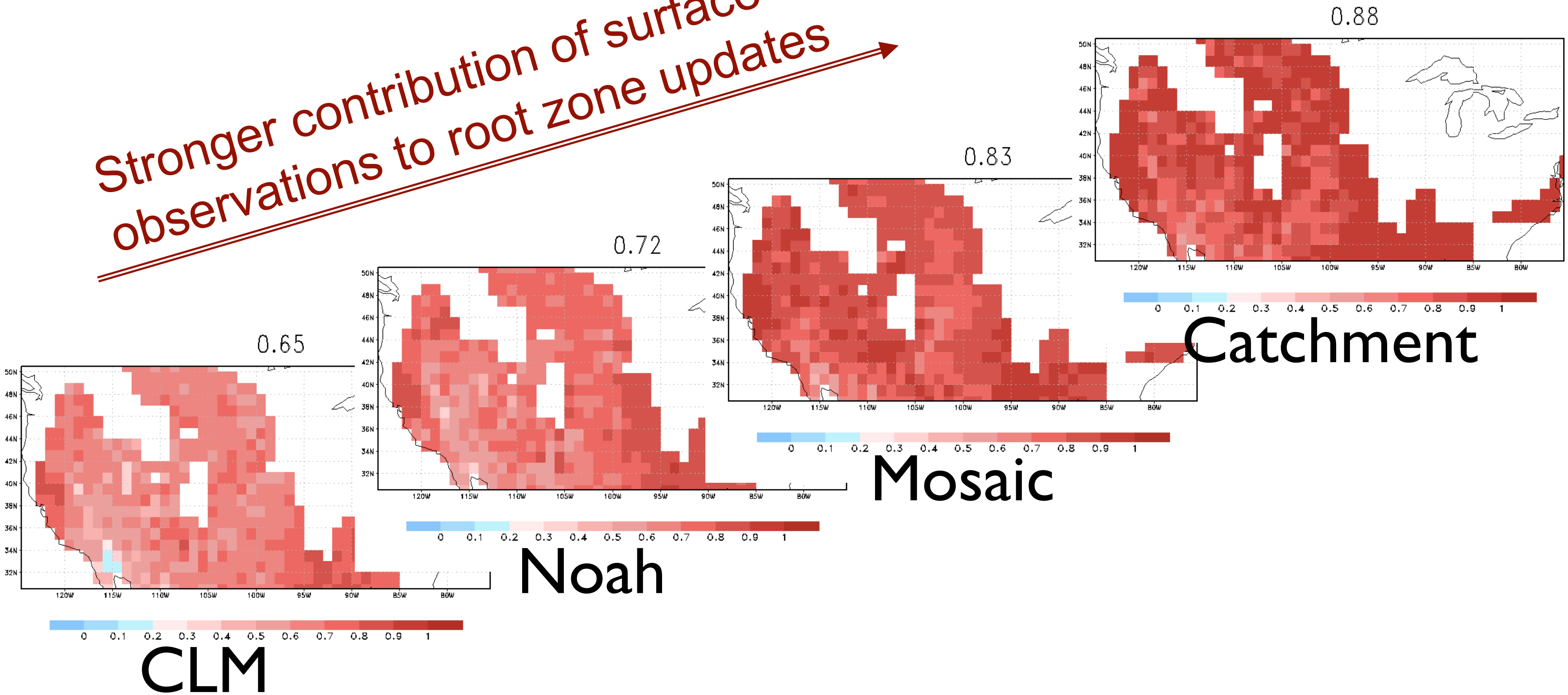


State vector	X = [*sfmc*, *rzmc*]^T
Obs operator	H = [1 0]^T
Model error cov	P = Cov(X)
Obs error cov	R
Kalman gain	$K = PH^T (HPH^T + R)^{-1}$

Gain correlation

How much do surface obs contribute to root zone updates?

Stronger contribution of surface observations to root zone updates



Same order as for vertical coupling strength.

Root zone soil moisture skill improvement from assimilation

NIC rzmc	Catch	Mos	Noah	CLM	Avg
Catch	0.71	0.54	0.36	0.38	0.50
Mosaic	0.55	0.69	0.31	0.33	0.47
Noah	0.43	0.43	0.36	0.26	0.37
CLM	0.11	0.21	0.10	0.45	0.22
	0.45	0.47	0.28	0.36	

2.) Average across columns (unknown truth physics):

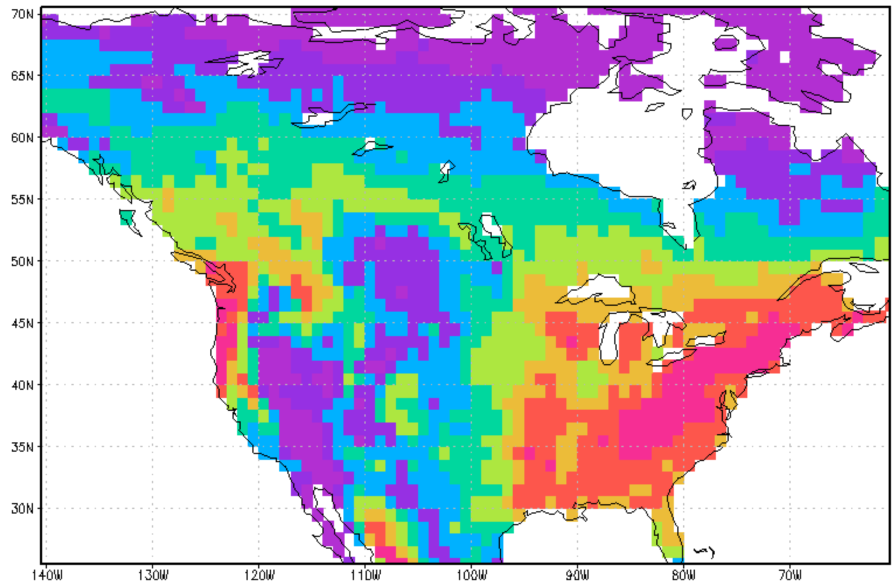
Use of Catchment, Mosaic, and Noah in assimilation system is better than use of CLM.

1.) Average across rows (known truth physics):

Mosaic or Catchment "truth" is "easier" to estimate in data assimilation than Noah or CLM "truth".

If coupling between surface and root zone is weak in truth, assimilation of surface observations is less efficient.

Data Assimilation Experiment Setup
Snow OSSEs

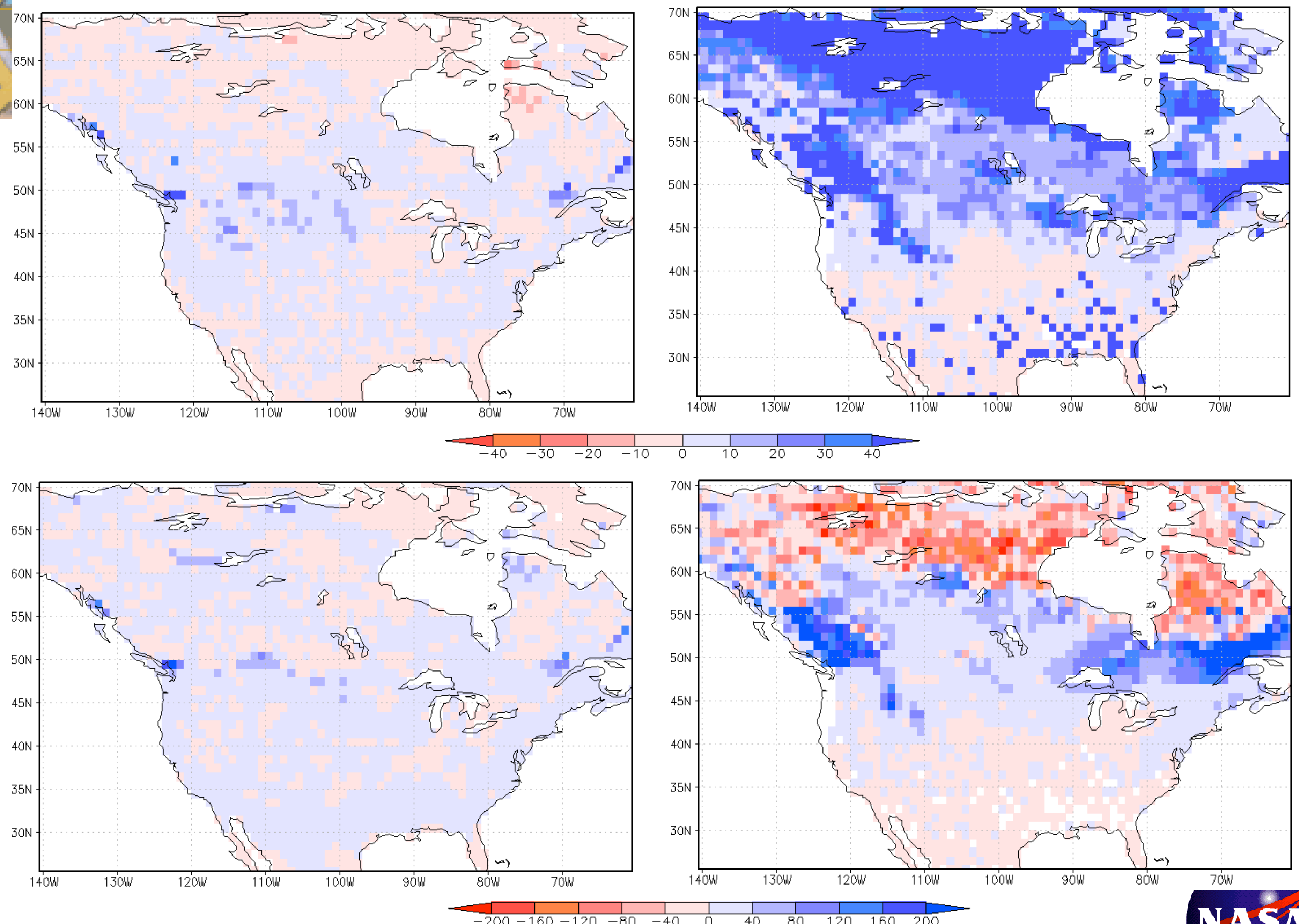


- Modeling domain: North America
- SWE Assimilation using EnKF and and SCA Assimilation using a rule based Direct Insertion (Rodell and Houser, 2004)
- October 1, 2003 to June 1, 2004
- Control/Truth runs using GDAS forcing (spun up from January 1, 2000) and Catchment LSM.
- OpenLoop runs using GEOS forcing and Noah LSM
- Synthetic SCA observations flagged using cloud cover masks from the MODIS Level 3 product (Hall et al, 2002)
- Synthetic SWE observations generated by
  - data masks for dense vegetation
  - random noise of 10mm error and 10mm minimum and 200mm maximum cutoffs
- Assimilation runs
  - SCA obs into the Open Loop run once a day at 12Z using the rule-based DI
  - SWE obs into the Open Loop run once a day at 12Z using the EnKF

Improvement Metric (RMSE(Assim) - RMSE (OL) for snow OSSEs

SWE

Snow Depth

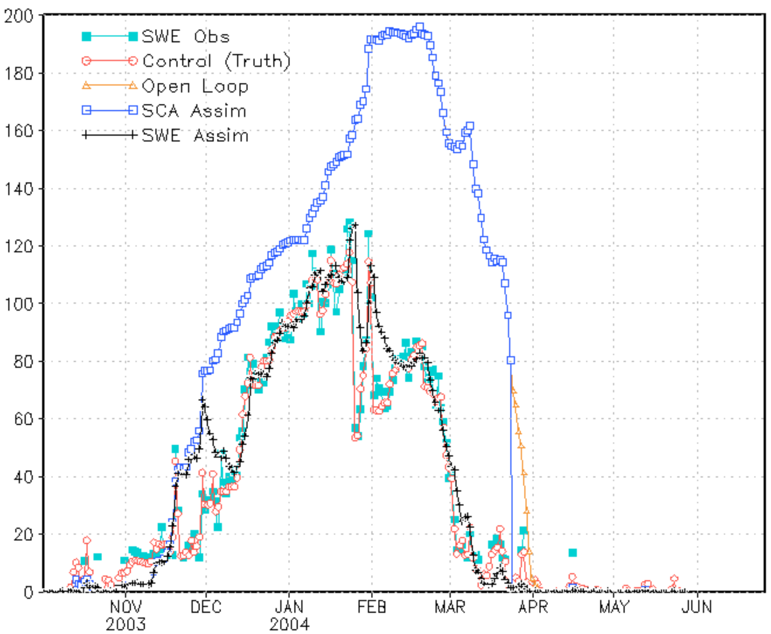


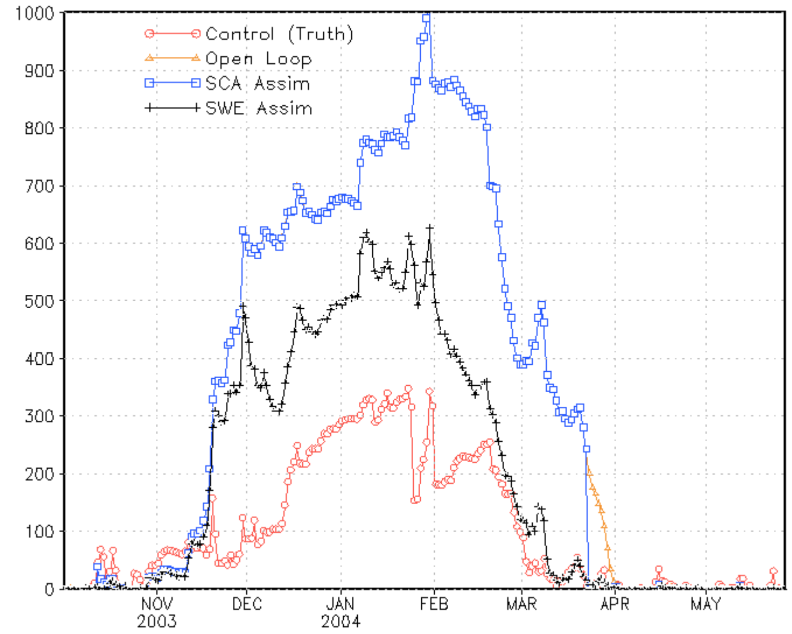
SCA Assimilation

SWE Assimilation

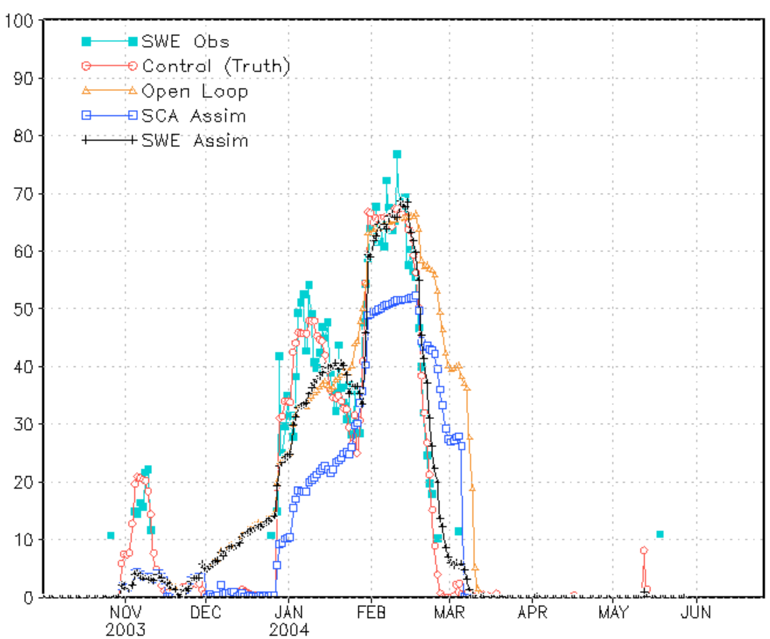
Time Series Comparisons of Snow fields

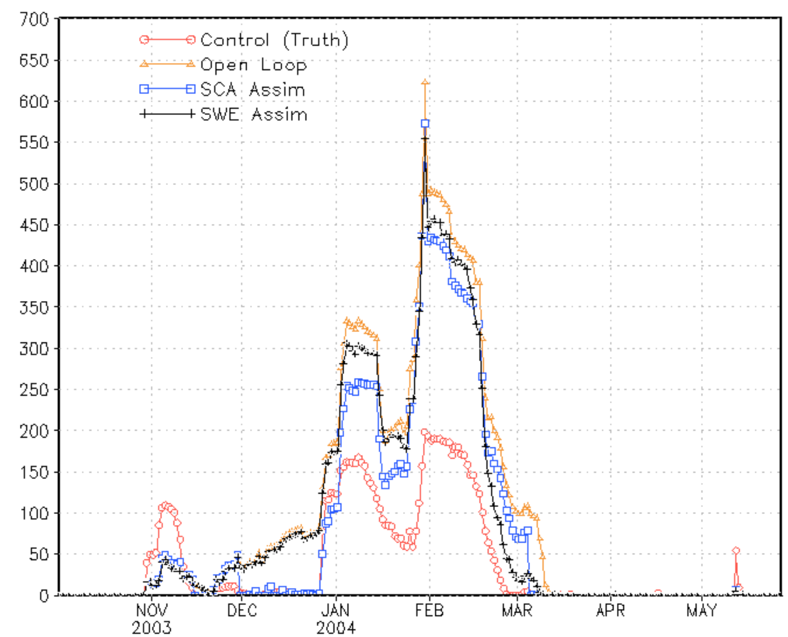
Plateau Mountain (50.2N, 116.5W)





Chinook, MT (48.6N, 109.2W)



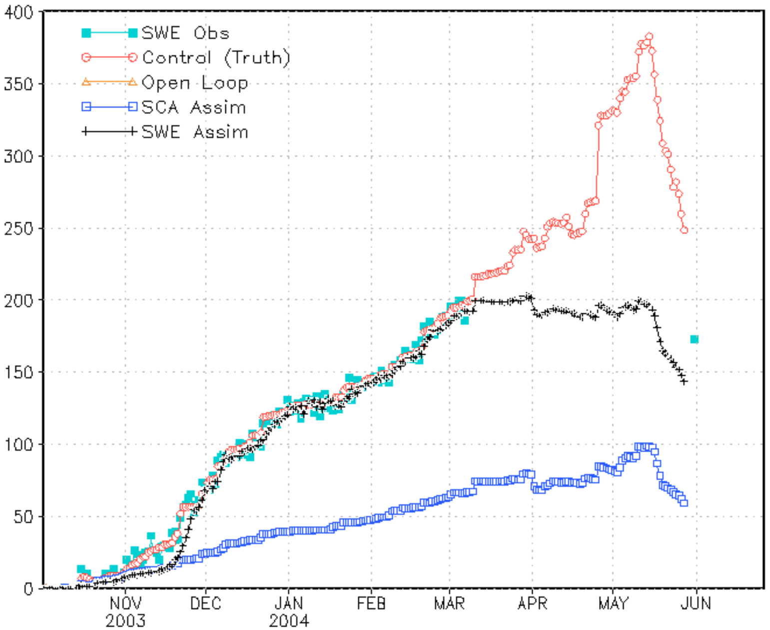


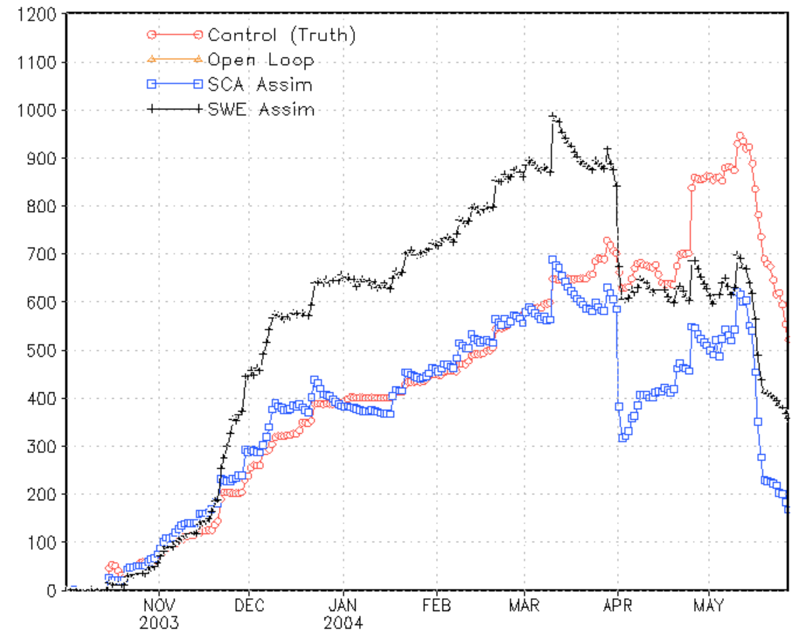
SWE

Snow Depth

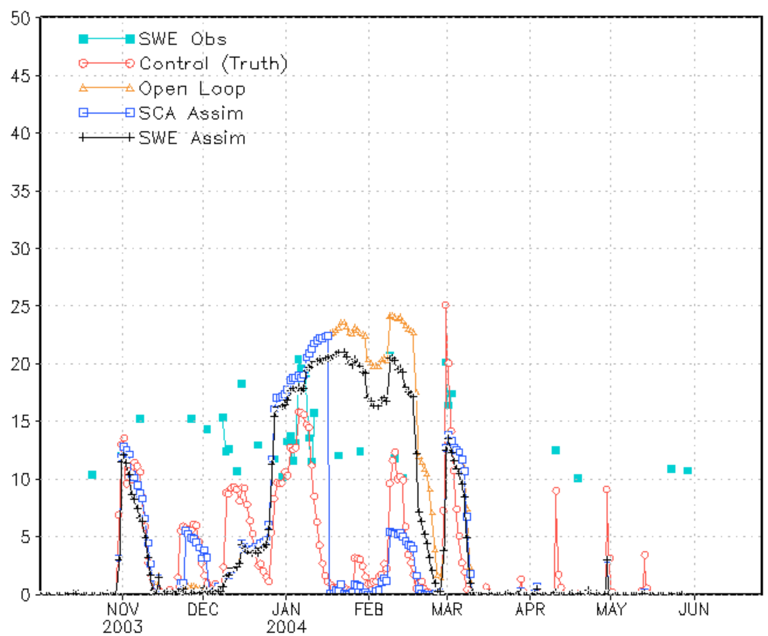
Time Series Comparisons of Snow fields

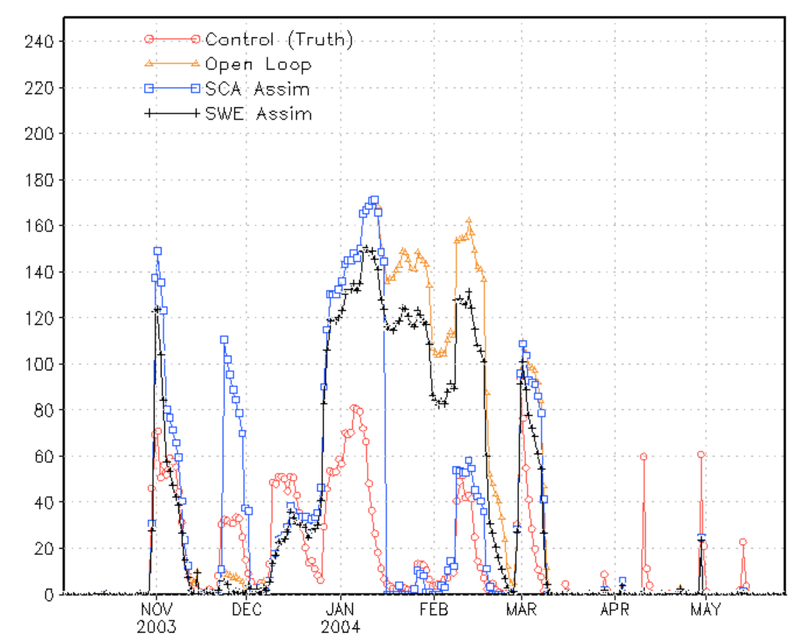
Churchill, RCT (58.8N, 94.5W)





Riverton, WY (43.0N, 108.4W)





SWE

Snow Depth

Skin Temperature Assimilation

- Need accurate assessment of LST because
  - LST is at the heart of the surface energy balance and impacts surface fluxes to the atmosphere (sensible, latent, and upward longwave)
  - LST is used as the lower boundary condition for retrieval of atmospheric profiles and cloud detection
  - LST has very little memory and as a result, the assimilation must apply a continuous correction, to prevent the drift back to the model climatology
  - Assimilation of LST must consider differences between satellite and model climatologies
  - Strategies
    - Scaling of LST obs to model climatology
    - Dynamic bias estimation

Bias estimation approach

1.) *Off-line (a priori) scaling between climatology of obs. and land model:*

+ No assumption whether model or observations are biased.
+ Easy to implement in pre-processing.
− Static (cannot adjust to changes in bias).

2.) *Dynamic model bias estimation:*

− Assume obs. climatology is correct and the model is biased.
+ Dynamic (adjusts to changes in bias).

Standard Kalman filter: $x^+ = x^- + K_x(y - Hx^-)$

$K_x = P_xH^T(HP_xH^T + R)^{-1}$

Bias estimation: $b^+ = b^- + K_b(y - Hb^-)$ (2nd Kalman filter)

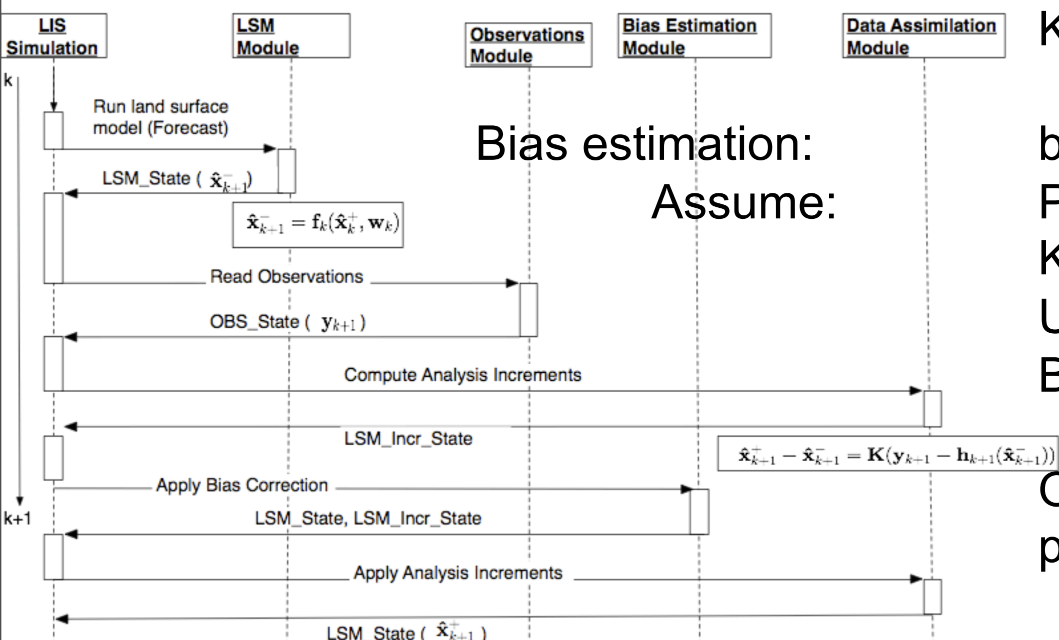
Assume: $P_b \sim P_x$

$K_b = \text{function}(K_x)$

Use KF increments to update bias.

Bias estimate is effectively time average of increments.

Options for diurnal and semi-diurnal bias parameterization.



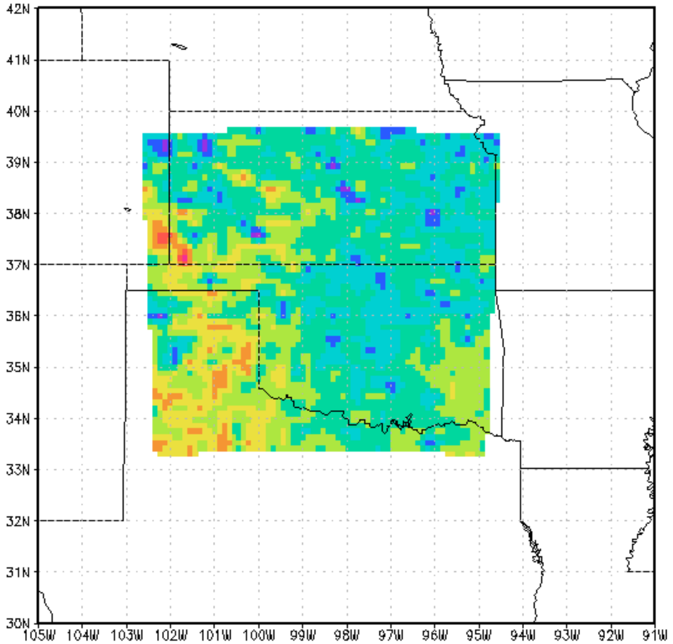
Synthetic Experiment Setup

Modeling domain centered around IHOP'02

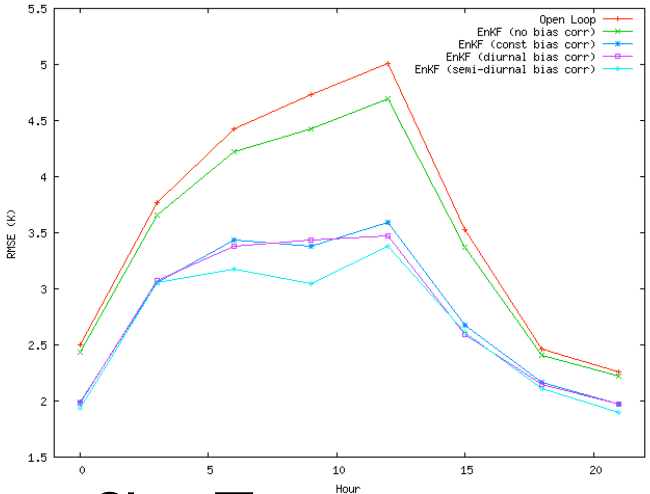
GSWP forcing at 1degree spatial resolution

Control run and synthetic observations produced using the Catchment LSM

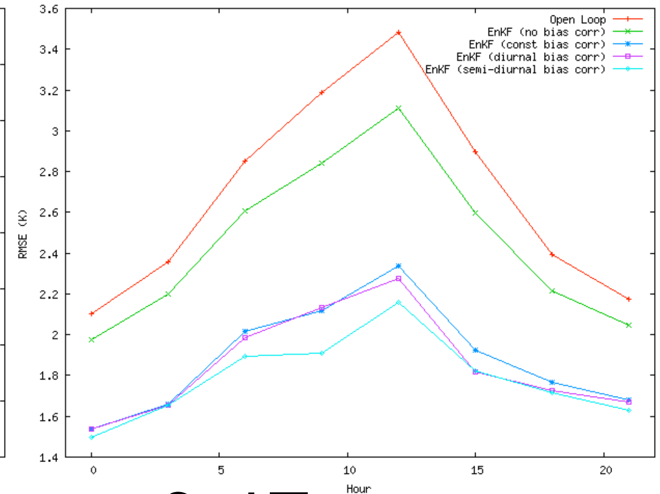
Open loop and assimilation runs using Catchment LSM



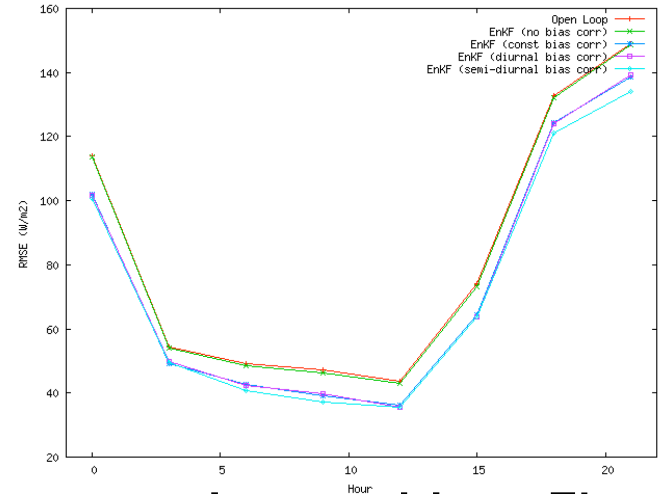
Results



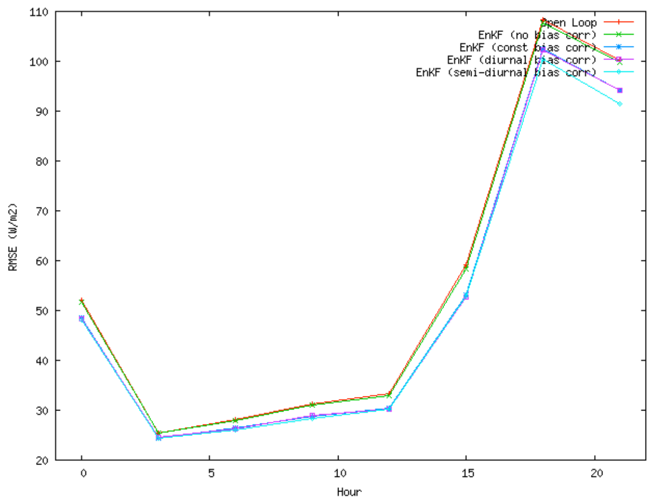
Skin Temperature



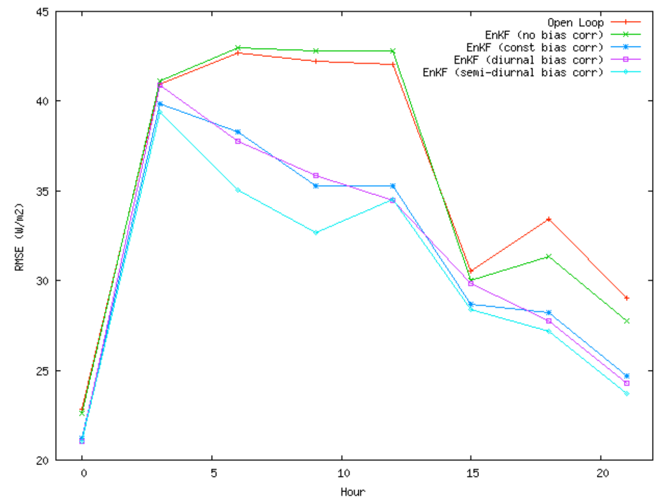
Soil Temperature



Latent Heat Flux



Sensible Heat Flux



Ground Heat Flux

Summary and Future Work

- A flexible, reusable, extensible framework for land surface data assimilation
- Use of multiple observations, support for variational, smoothing algorithms, 3d algorithms, radiance based assimilation (through CRTM)
- Support for parameter estimation, calibration and generic optimization requirements - towards a SODA (Simultaneous Optimization and Data Assimilation) framework