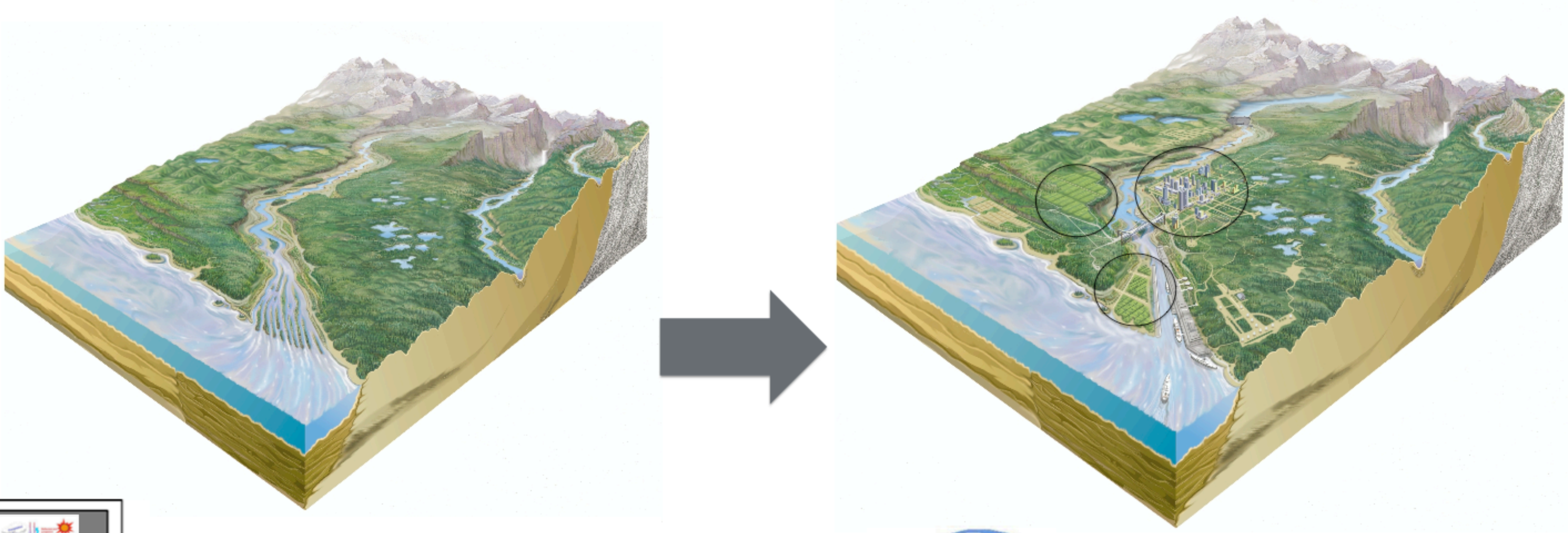
Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes

Sujay Kumar[1], Christa Peters-Lidard[1], Joseph Santanello[1], Rolf H. Reichle[2], Clara Draper[3,2], Randal Koster[2],Grey Nearing[4,1], Michael Jasinski[1]

(1) Hydrological Sciences Laboratory, NASA GSFC, Greenbelt, MD
(2) Global Modeling and Assimilation Office, NASA GSFC, Greenbelt, MD
(3) Universities Space Research Association, NASA GSFC, Greenbelt, MD
(4) Science Applications International Corporation, McLean, VA

Human impacts from expansion of agriculture and infrastructure have significantly (>50%) transformed the natural features of the land surface



Land surface models : fairly utopian; hard to realistically represent subjective practices

Remote sensing: practical method to observe these 'unmodeled' features

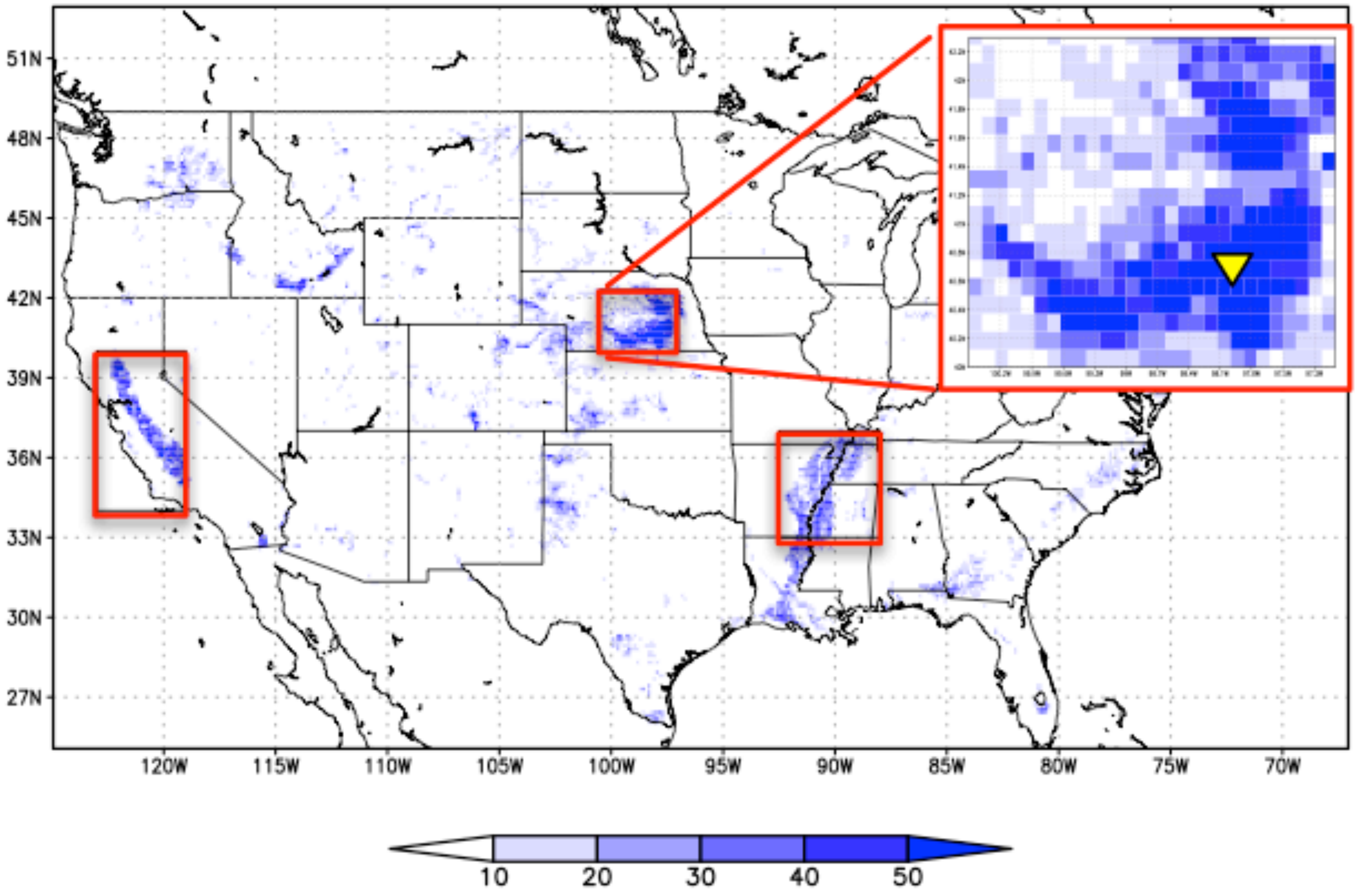
Using irrigation as an example of a human engineered, often unmodeled process ….

1. Can modern soil moisture remote sensing datasets detect such features?

2. Are the DA methods effective in incorporating such signals into the models?

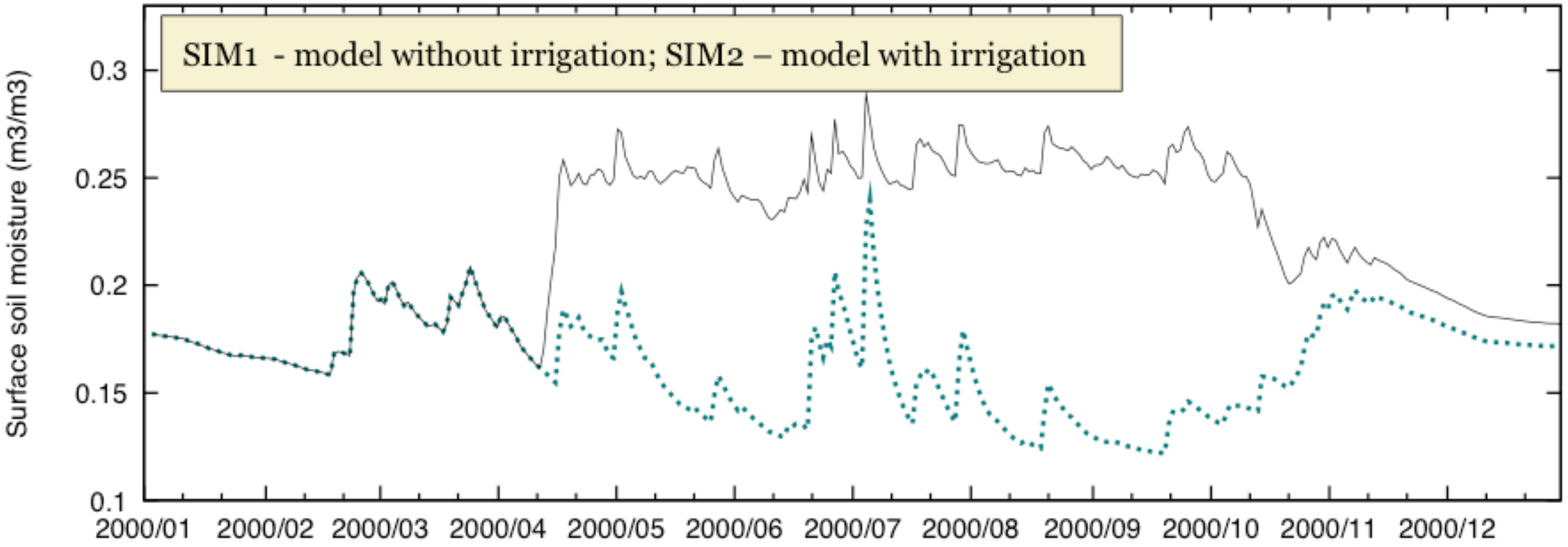
Kumar, S.V. et al. (2015), Evaluating the utility of modern soil moisture remote sensing retrievals over irrigated areas and the suitability of data assimilation methods for incorporating unmodeled artifacts, *HESS,* 19, 4463–4478.

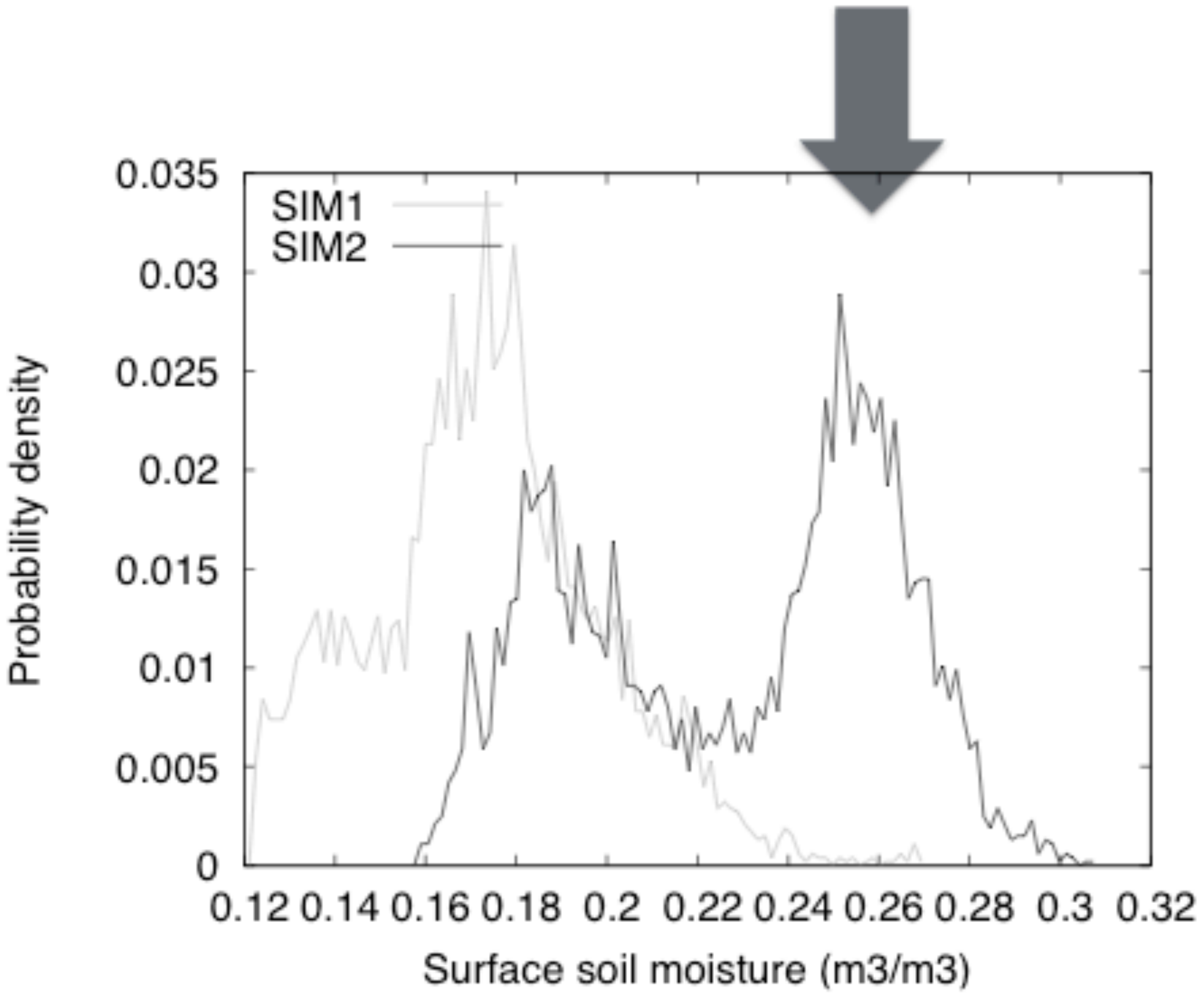
Irrigation “hot-spots” in the U.S



MODIS-based irrigation intensity map (Ozdogan & Gutman, 2008)

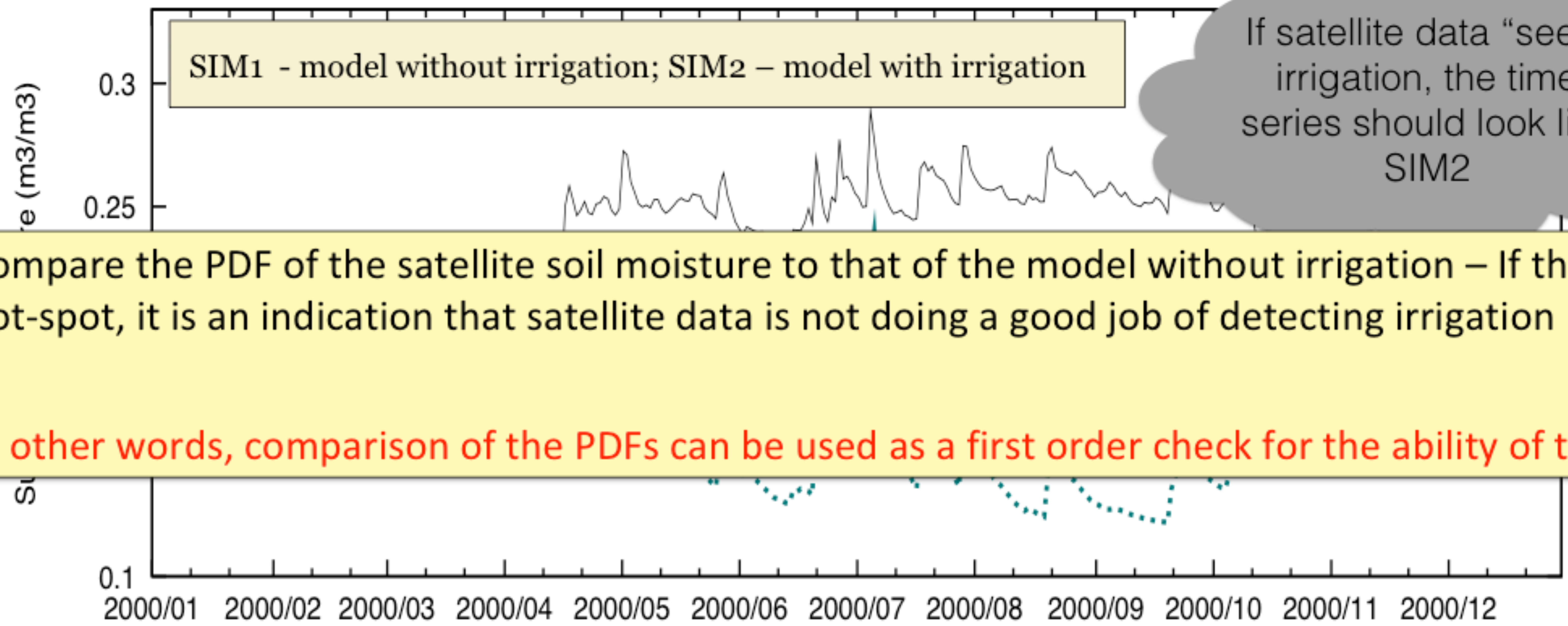
An idealized simulation of irrigation in the LSM





Significant changes in soil moisture distribution due to irrigation with changes in mean, location and symmetry

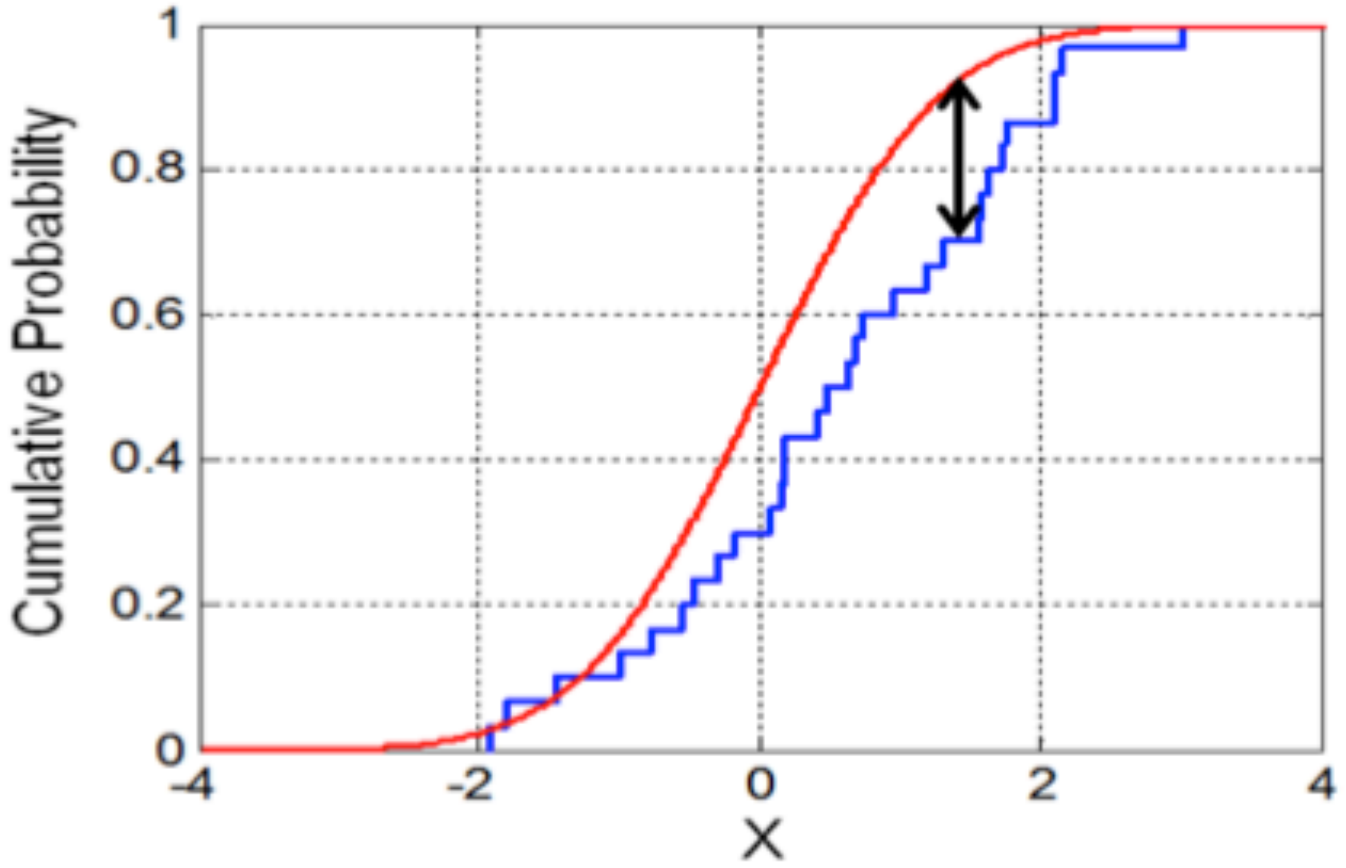
Bimodality due to the seasonal effect of irrigation



If satellite data "sees" irrigation, the time series should look like SIM2

Compare the PDF of the satellite soil moisture to that of the model without irrigation – If the PDFs are similar at an irrigation hot-spot, it is an indication that satellite data is not doing a good job of detecting irrigation

In other words, comparison of the PDFs can be used as a first order check for the ability of these datasets to detect irrigation



$$D_{m,n} > c(\alpha)\sqrt{\frac{m+n}{mn}}$$

$$D_{m,n} = \max_x |F(x) - G(x)|$$

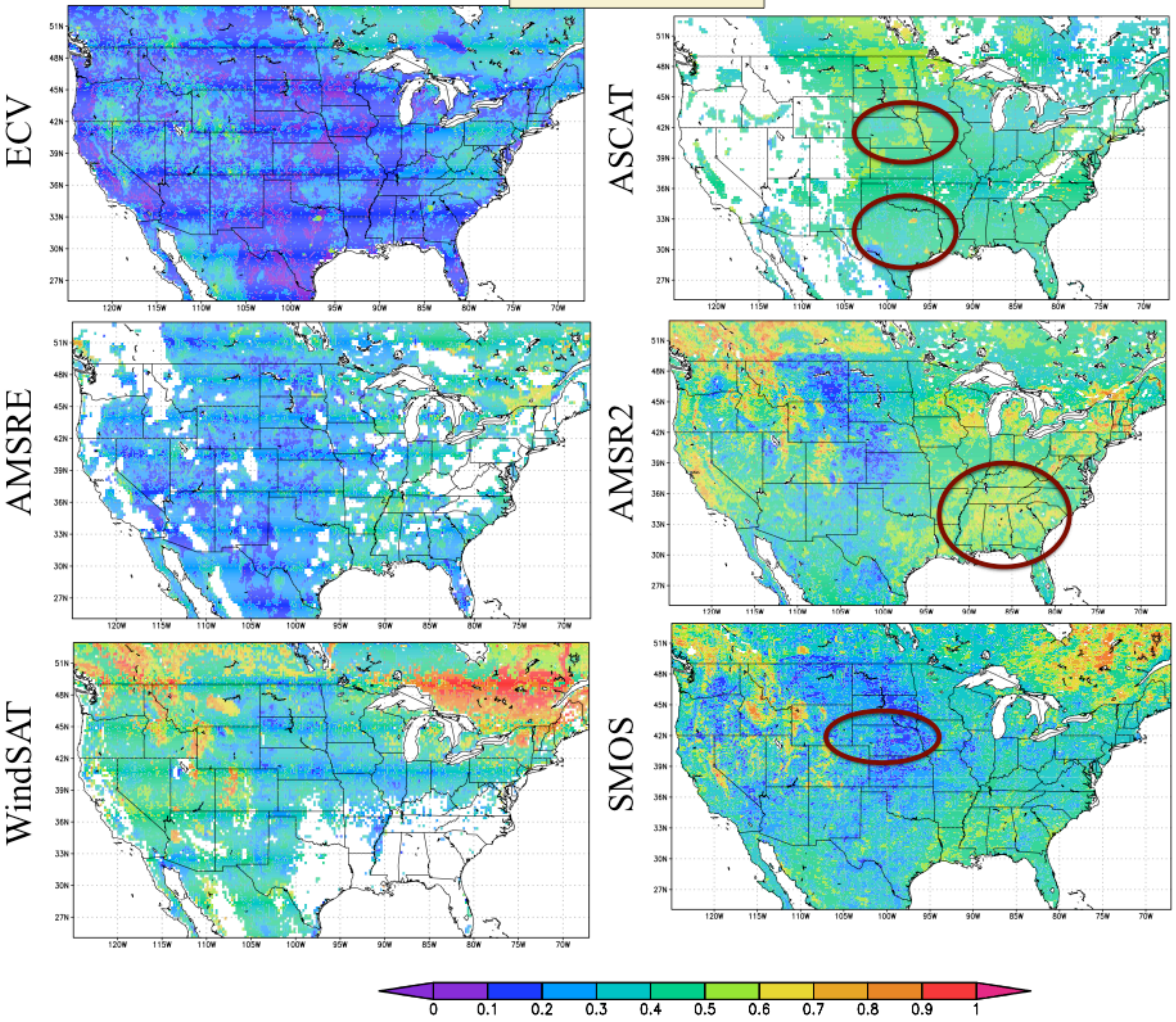
A two-sample Kolmogorov – Smirnov (K-S) test is used to quantitatively compare the PDFs

K-S distance (D) measures the distance between the PDFs. Values closer to zero indicate that the distributions are similar; larger values of D indicate locations where the PDFs differ

K-S distance (D)

ECV shows the lowest D values (possibly because ECV was generated by CDF matching soil moisture estimates from different sensors to GLDAS Noah)

Larger differences seen with other sensors – a mix of biases from instrument noise, retrieval algorithm errors and **unmodeled observational processes**



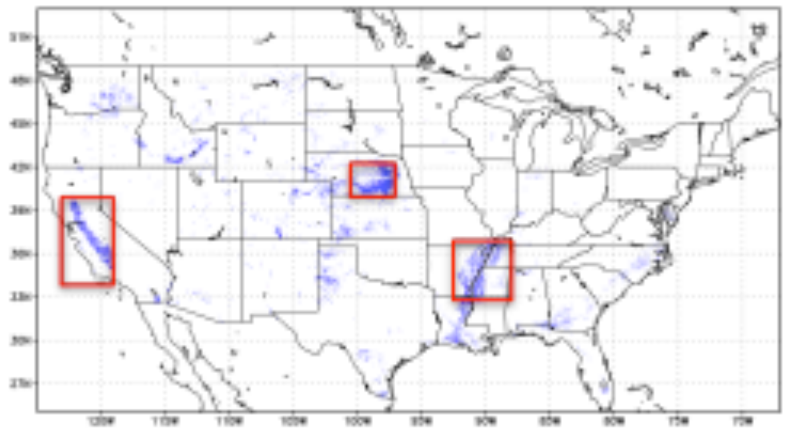
Larger D values in eastern Nebraska – possible signal of irrigation

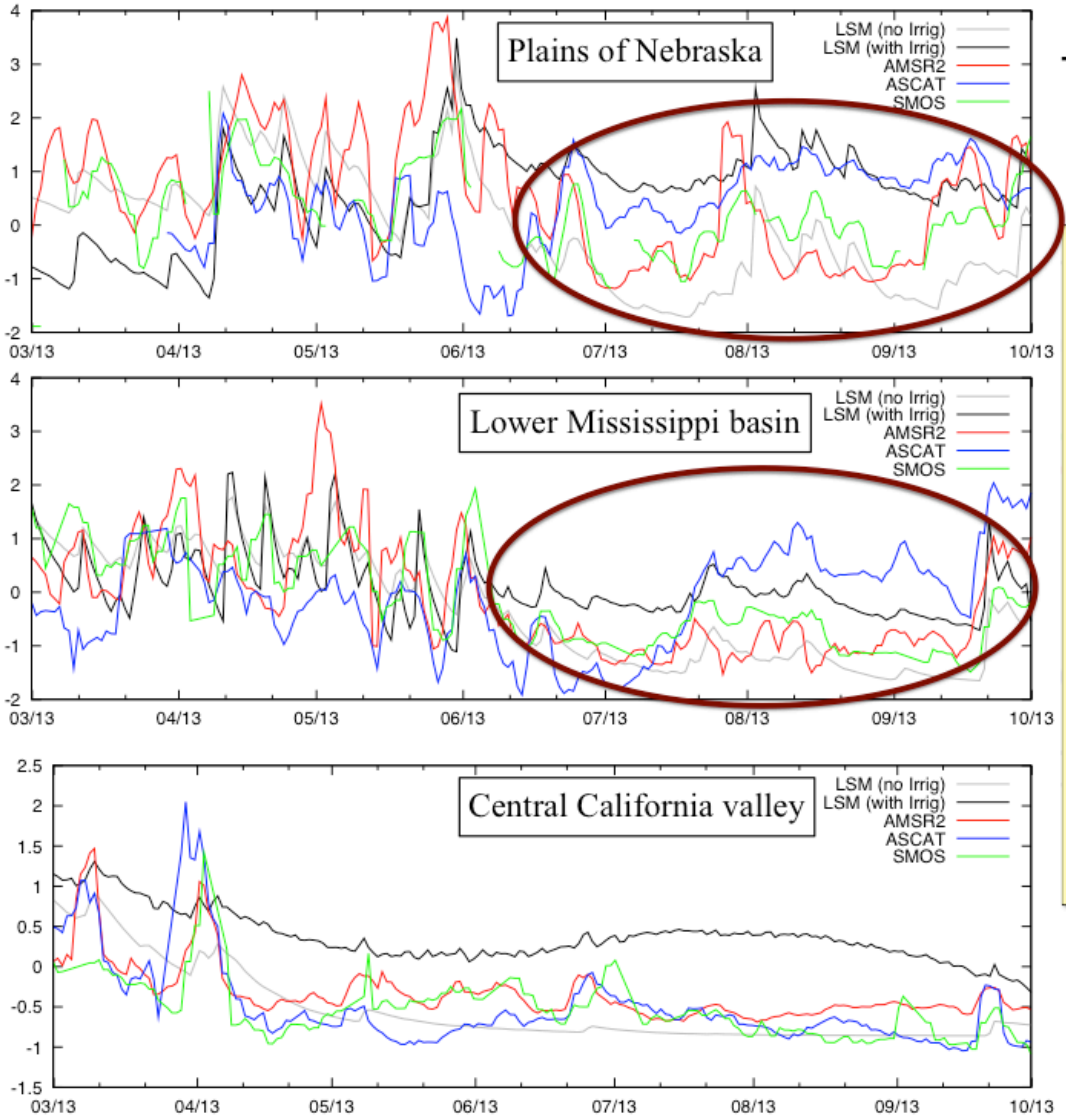
Signal of urban areas – Dallas, Houston

Strong signal of vegetation density in the eastern U.S.

Low D values in the SMOS comparison over Nebraska – an indication that SMOS retrievals are not doing a good job at detecting irrigation

Normalized soil moisture time series averaged over the irrigation hot-spots

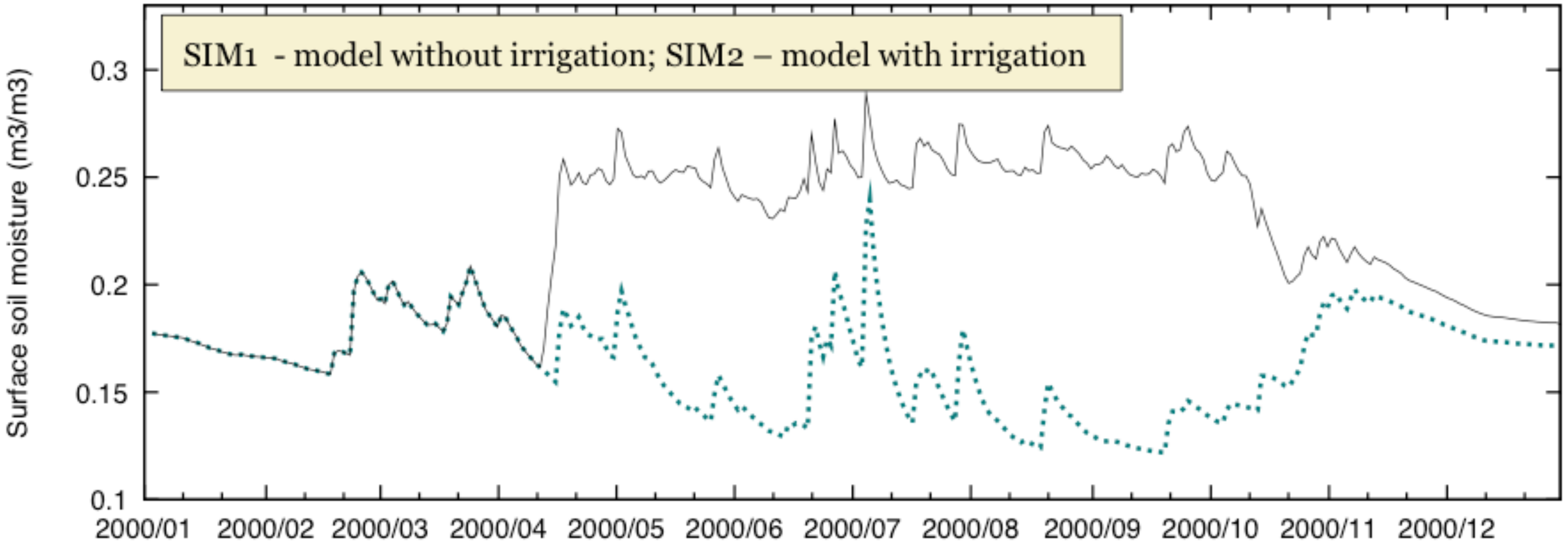




ASCAT time series shows better agreement with the LSM with irrigation time series in the summer and fall months (Nebraska, Lower Mississippi); No such distinct contrast in the central California valley.

AMSR2 and SMOS time series show better agreement with LSM without irrigation

The skill of AMSR2 and SMOS retrievals are low in detecting irrigation whereas ASCAT retrievals are somewhat effective in detecting these features



If unmodeled processes such as irrigation are present in satellite retrievals, can we represent them through data assimilation?

Data assimilation methods are primarily designed to work with random errors. Proper treatment of biases is important for the success of data assimilation.

If unmodeled processes (irrigation) are the major source of biases, are the typical bias mitigation strategies in DA systems appropriate?

$$\underbrace{\mathbf{x}_k^{i+} - \mathbf{x}_k^{i-}}_{\text{Analysis increment}} = \mathbf{K}_k \underbrace{[\mathbf{y}_k^{i} - \mathbf{H}_k \mathbf{x}_k^{i-}]}_{\text{Innovation}}$$

x_k Model state vector

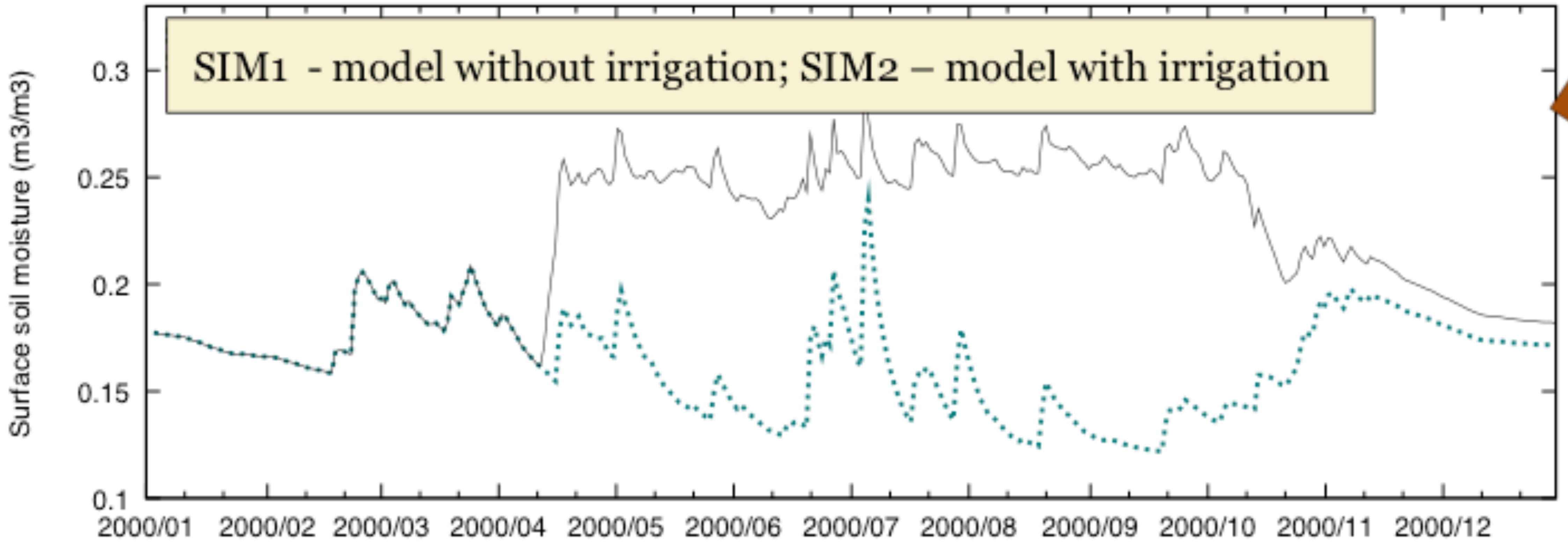
y_k Observation vector

In bias-blind DA systems, observations (y_k) and model forecasts ($H_k x_k^{i-}$) are expected to be unbiased relative to each other.

There are two choices for bias correction:

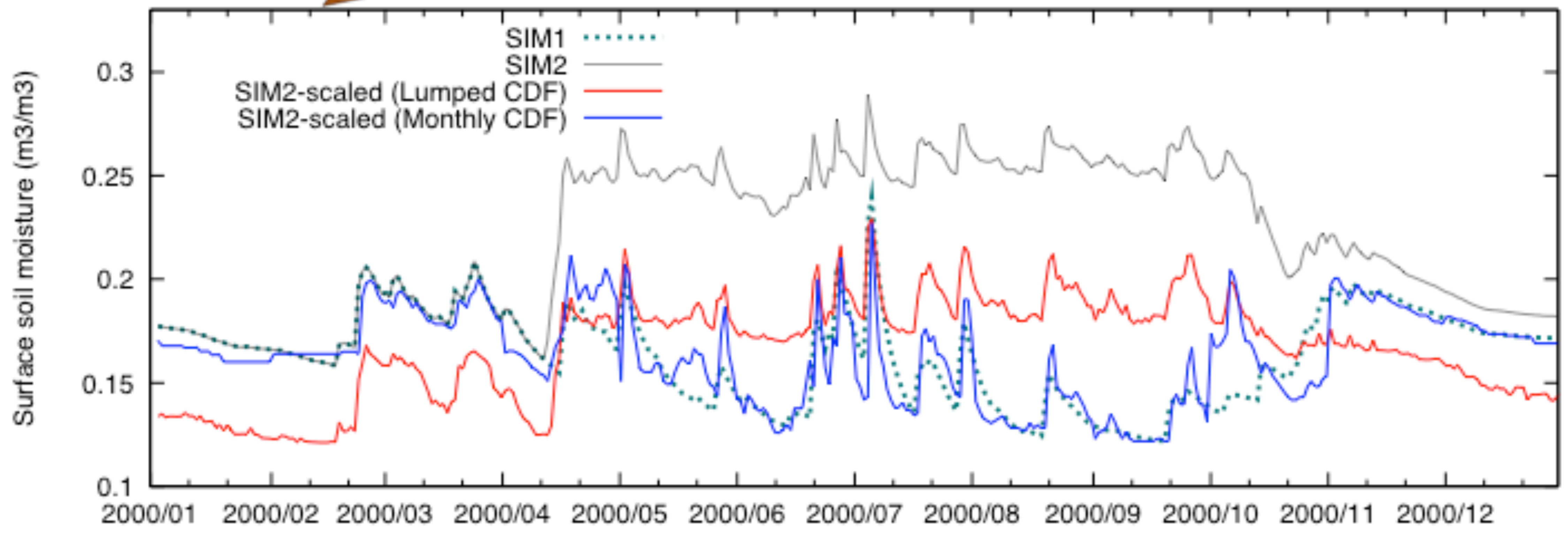
- Rescale observations into the model climatology so that the innovations are computed in the climatology of $H_k x_k^{i-}$
  - Standard normal deviate scaling, CDF matching
- Compute the innovations in the observation space by having an operator (H_k) that translates the model states into the observation space.
  - Trained forward models (RTMs, ANNs)

"Bias-blind" DA strategies

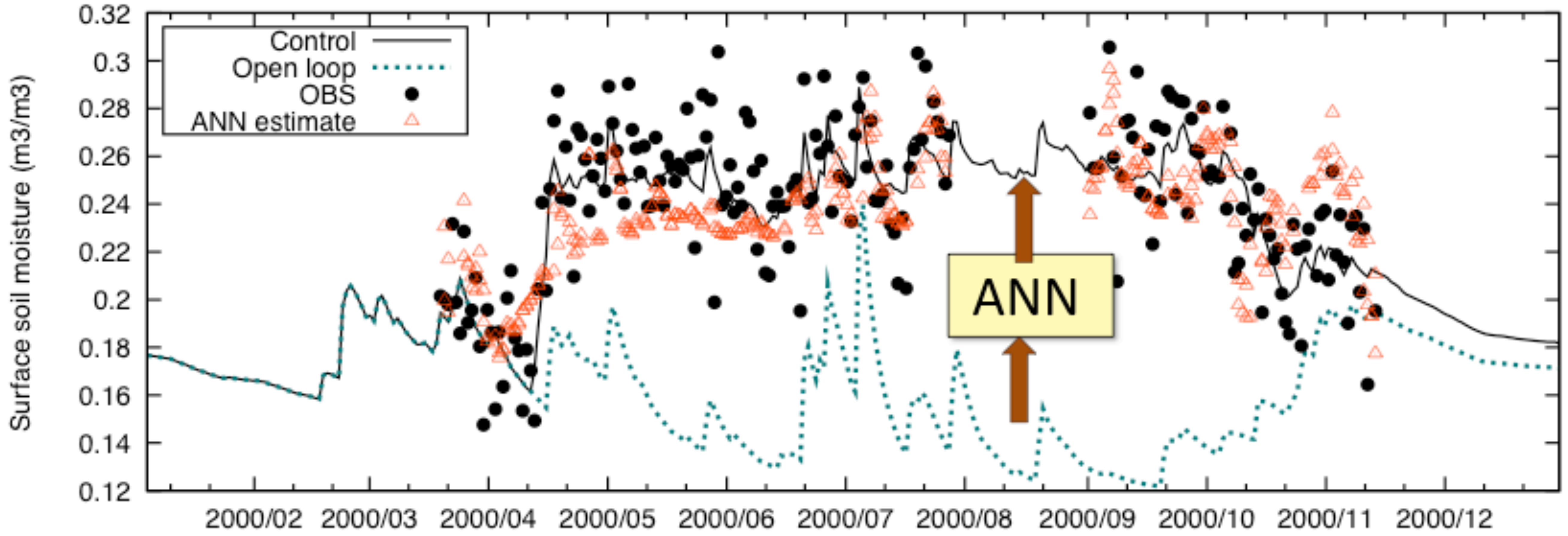


Lumped CDF matching (CDFs computed using all years and all seasons lumped together)

Monthly CDF matching (CDFs computed separately for each month)



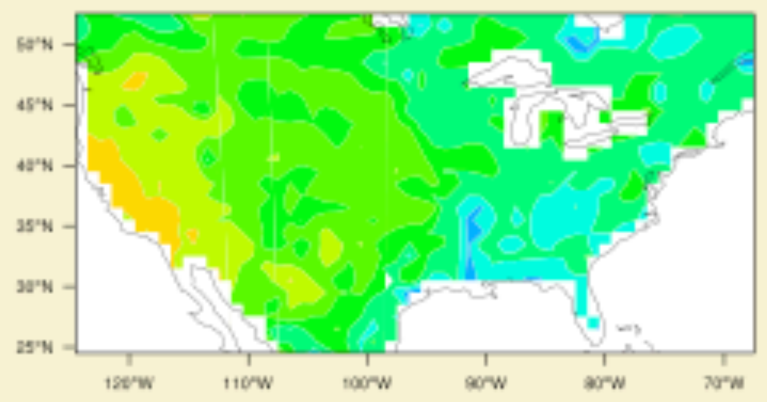
Spurious statistical artifacts introduced in the lumped CDF matching approach, which are reduced when CDF matching is more finely temporally resolved



Trained ANN simulates the anomalous wet signals of irrigation

Do these bias correction strategies work when the model background is the major source of biases?

Synthetic DA experiment



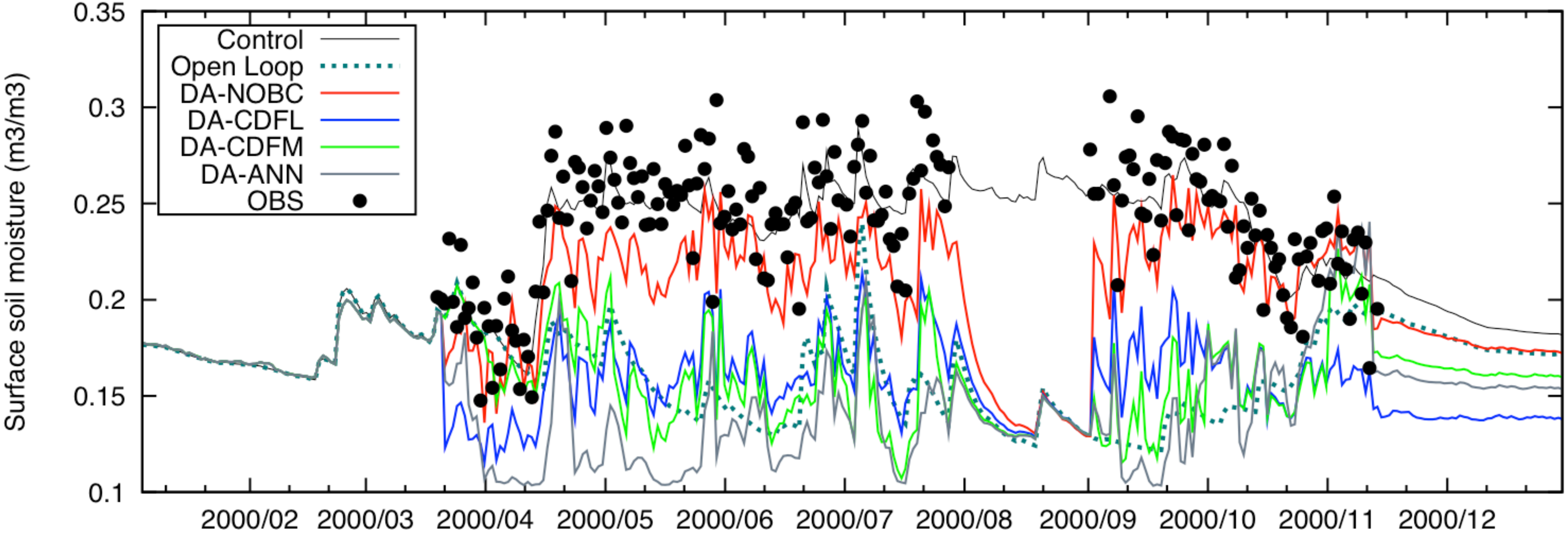
Simulate the "Control" run (Noah LSM forced with NLDAS2 and irrigation simulated)

Generate "observations" 0.02 m3/m3 gaussian noise Masking for dense vegetation, rain/snow events

Assimilate observations using different bias correction strategies into an OL configuration

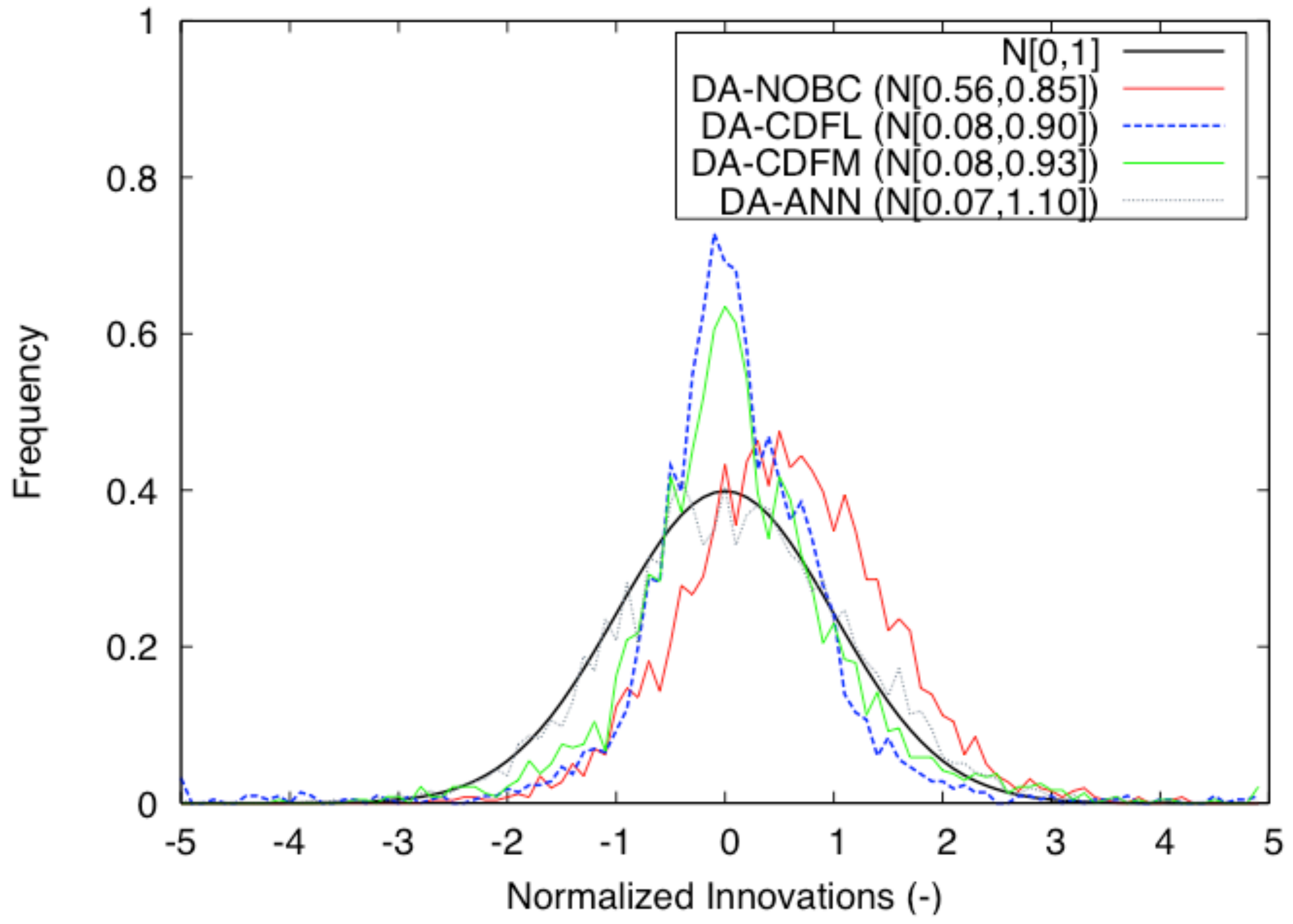
DA-NOBC – assimilate observations directly without any bias correction
DA-CDFL – assimilate by rescaling observations using lumped CDF matching
DA-CDFM – assimilate by rescaling observations using monthly CDF matching
DA-ANN –assimilate raw observations using trained ANN as a forward operator

Compare the DA run against the Control run



- DA runs with bias correction do not incorporate the wet signal of irrigation!
- The size of the innovations remain small in these runs as the anomalous wet signal is treated as a bias artifact and removed in the DA system
- The standard bias correction strategies make no distinction of the source of the biases. The signals from unmodeled processes are therefore, excluded

- The deviations from the standard normal N(0,1) of the normalized innovations are usually used to infer the optimality of the DA configuration.



- DA-NOBC shows the largest deviation from N(0,1), indicating the presence of bias
- Other DA integrations show close to optimal behavior
- Reliance on these diagnostics could be misleading if unmodeled processes are the source of the bias.

Summary and Conclusions

- Simulating subjective practices such as irrigation is inherently hard to do in conceptual models. Remote sensing and data assimilation are practical ways to incorporate these unmodeled features.
- The skill of the soil moisture retrievals, however, must be improved to effectively monitor such human engineered processes.
- LDA methods must adapt if we are to represent unmodeled processes on the land surface through data assimilation.
- Challenge in LDA systems is to understand the source of the biases. Focus on understanding and separating systematic errors from unmodeled processes and other sources such as retrieval algorithm errors, instrument errors is needed.