

SMAP: A hydrologist goes crazy with a new high-quality dataset

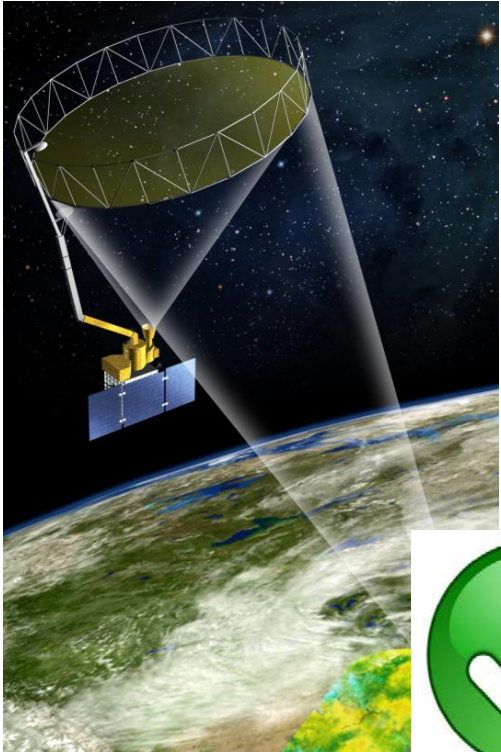
Randal Koster*

Global Modeling and Assimilation Office, NASA/GSFC

randal.d.Koster@nasa.gov

*with substantial help from Wade Crow, Rolf Reichle, and Sarith Mahanama

Clarification: this talk is about the NASA soil moisture mission, not about the Japanese boy band.



SMAP provides a capability for global mapping of soil moisture with unprecedented accuracy, resolution, and coverage.

Original specs:

Spatial resolution : 3-km, 9-km, and 36-km

Temporal resolution : every 3 days (at least)

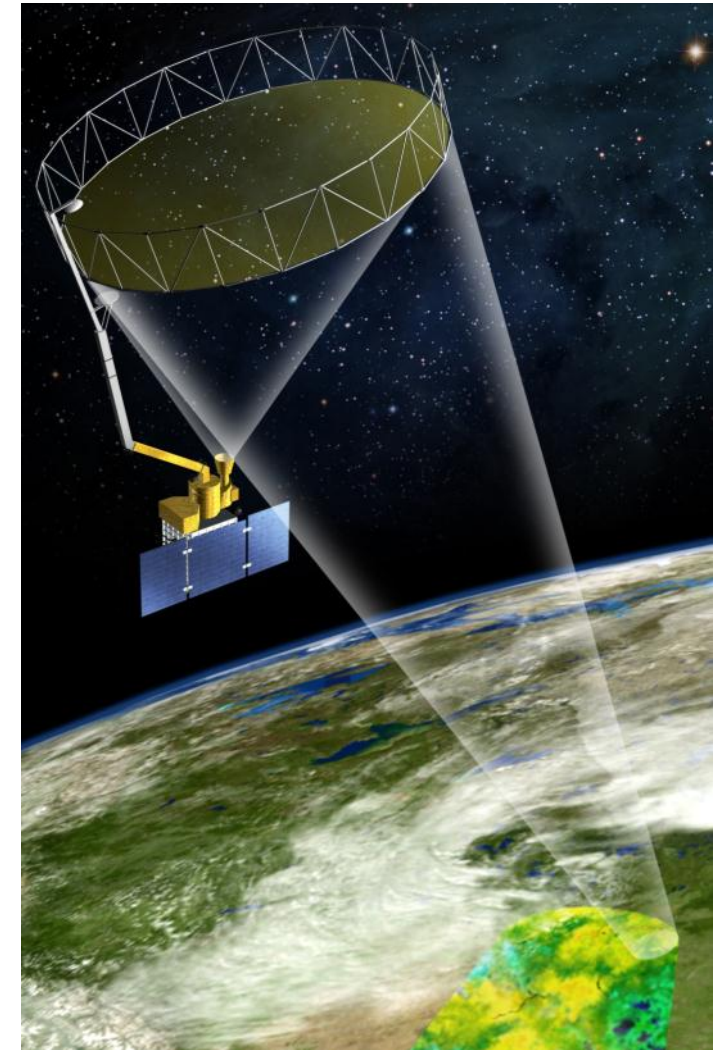
How deep into soil: several cm (Level 1-3)
1 m (Level 4)

Accuracy: RMSE < 4 volumetric percent

Latency: short! (hrs-days, depending on product)

Baseline mission duration: 3 years
(launched in 2015; mission just extended)

Orbit specs: Sun synchronous, 6 AM/PM,
8-day exact repeat with 2-3 day revisit



<http://smap.jpl.nasa.gov/>

SMAP provides a capability for global mapping of soil moisture with unprecedented accuracy, resolution, and coverage.

Original specs:

Spatial resolution ~~3-km~~ ~~9-km~~ and 36-km

Temporal resolution: every 3 days (at least)

How deep into soil: several cm (Level 1-3)

1 m (Level 4)

Accuracy: volumetric percent

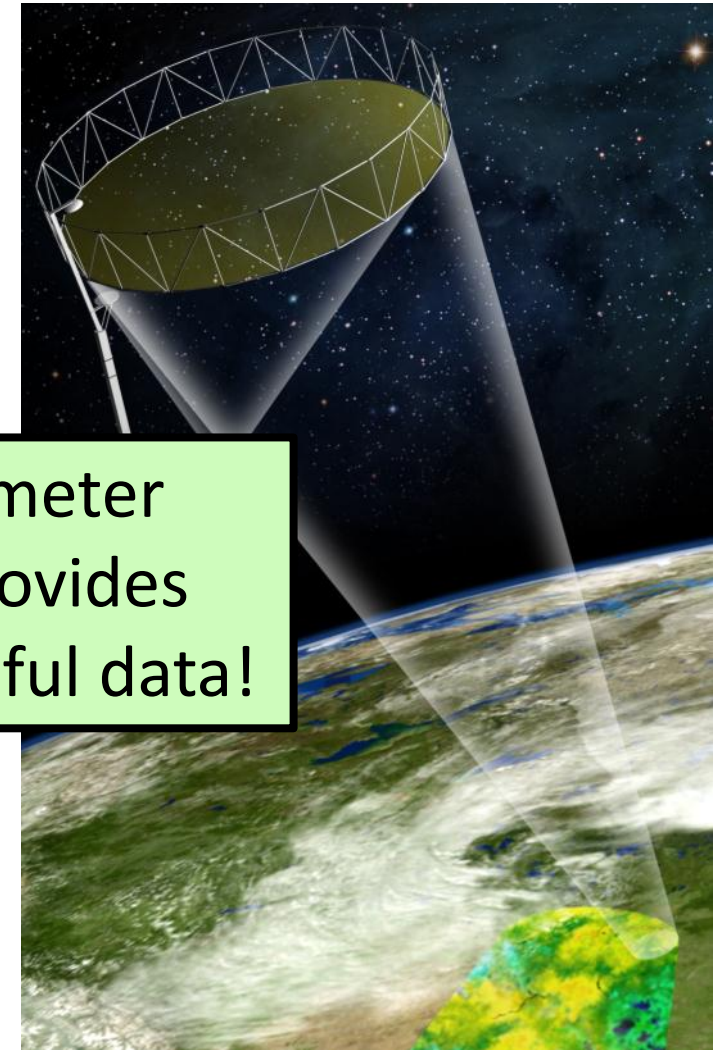
Latency: days, depending on

Baseline mission duration: 3 years
(just extended)

Orbit specs: Sun synchronous, 6 AM/PM,
8-day exact repeat with 2-3 day revisit

Radar
instrument
broke!

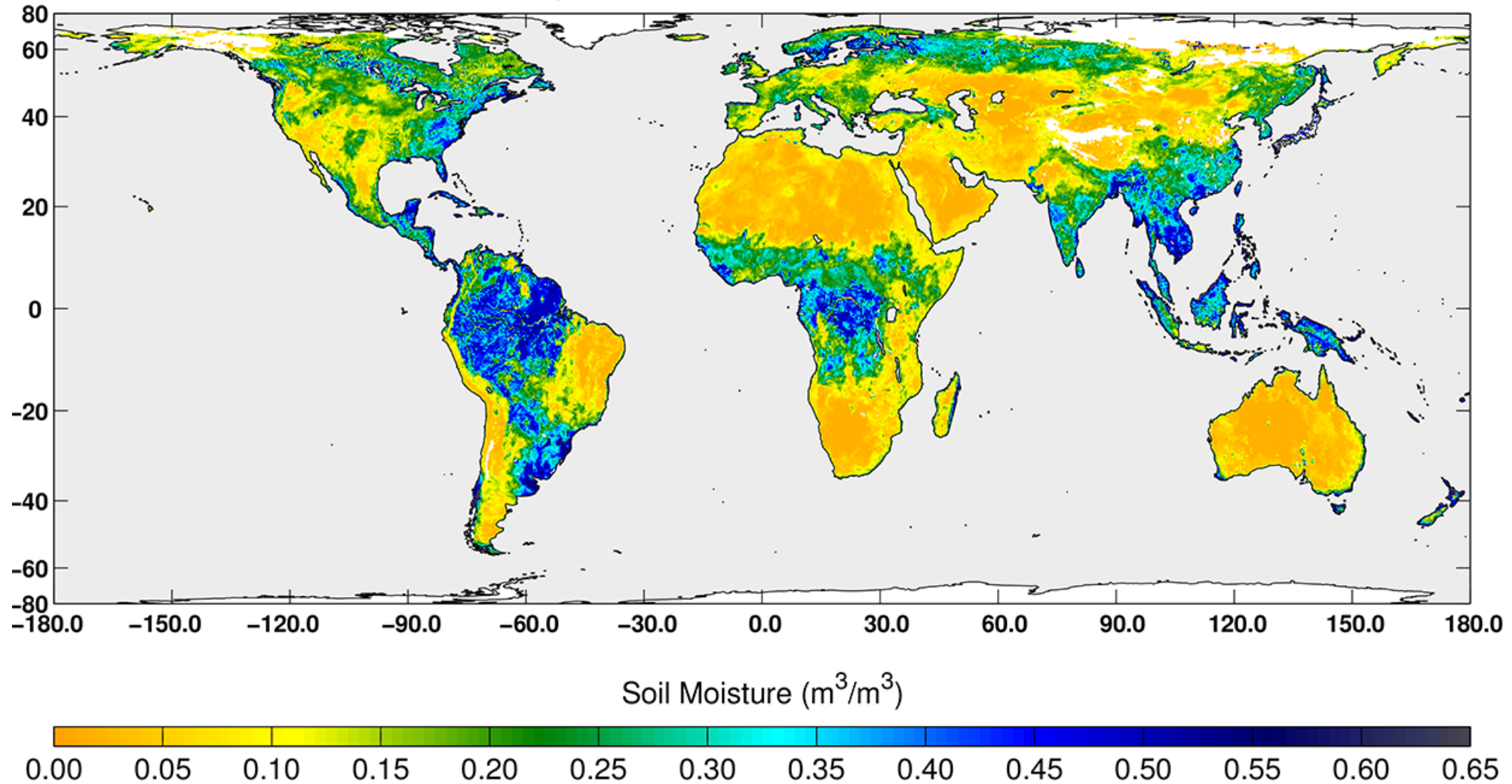
Radiometer
still provides
beautiful data!



<http://smap.jpl.nasa.gov/>

Example:

SMAP radiometer-only soil moisture between Oct 3, 2015 and Oct 5, 2015



So.... What can we do with all these data?

Three topics:

1. “Loss functions”: the characterization of soil moisture dynamics.
2. Data assimilation and calibration: do they access the same information?
3. Estimating rainfall and streamflow with SMAP data.

Three topics:

1. “Loss functions”: the characterization of soil moisture dynamics.

2. Data assimilation: how do they access the same information?

Published paper: Koster et al., J.
Hydromet., 18, 837-843, 2017;
DOI: 10.1175/JHM-D-16-0285.1

3. Estimating rainfall and streamflow with SMAP data.

What is a loss function?

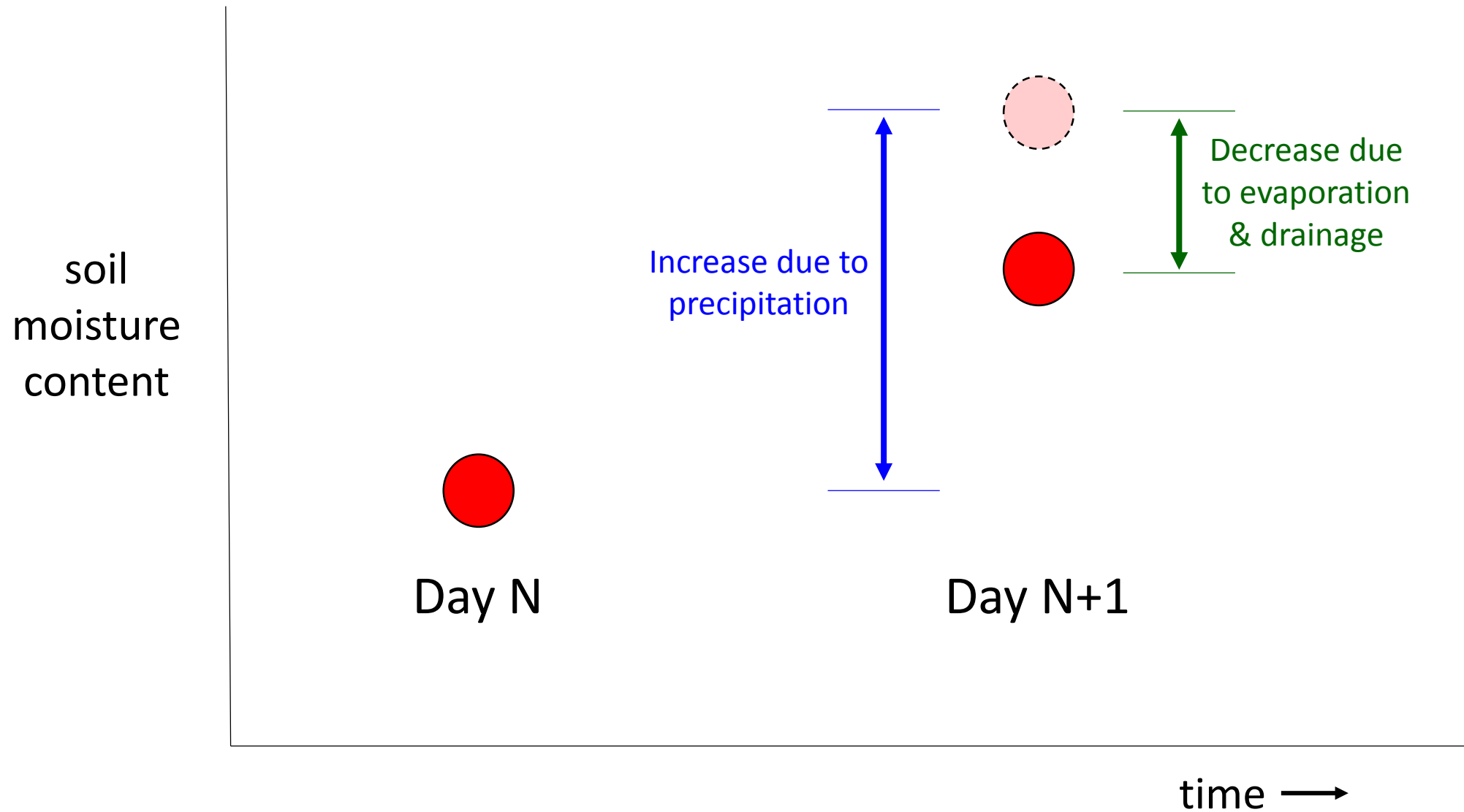
Consider two retrievals separated by 1 day:

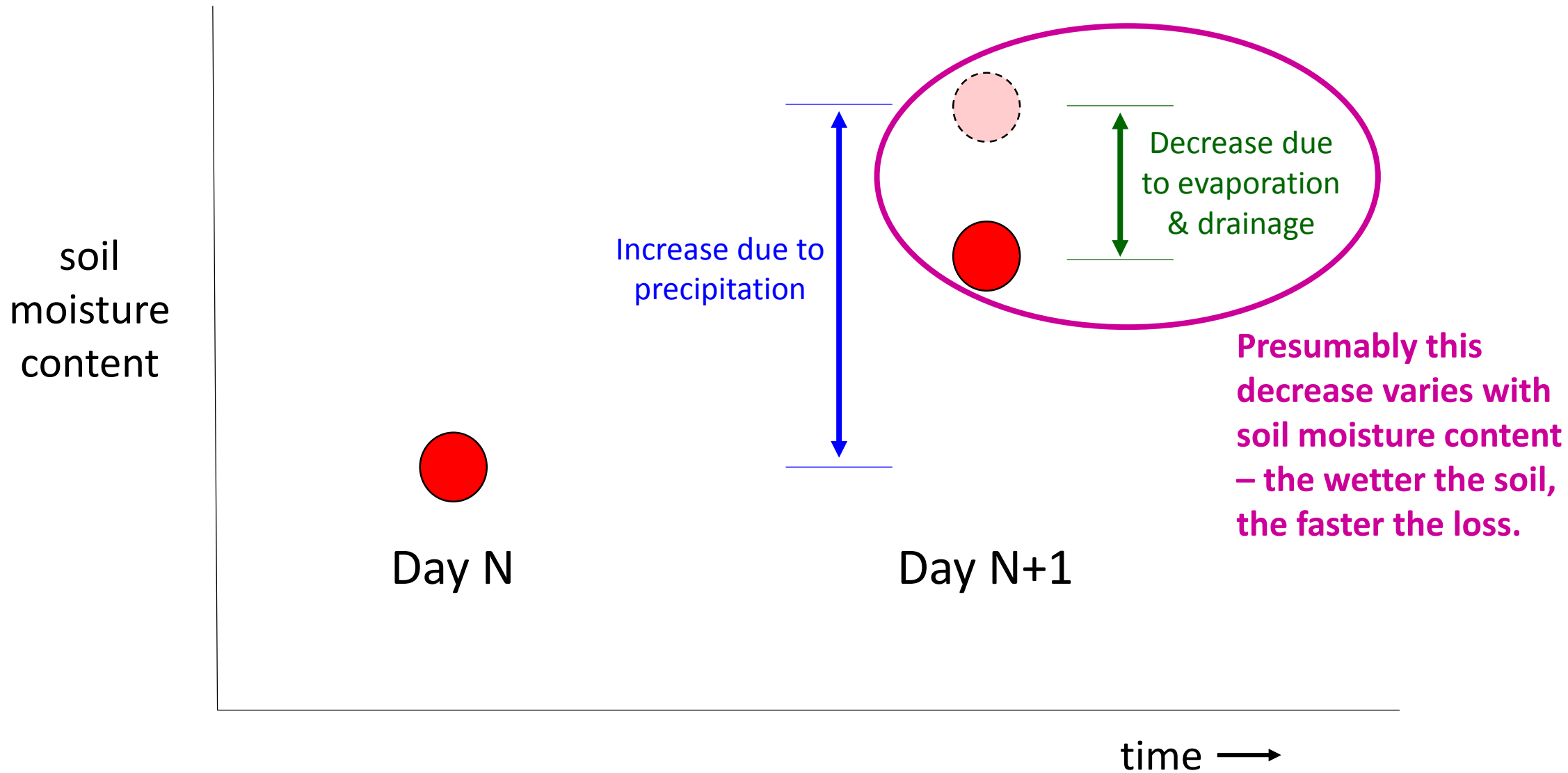
soil
moisture
content



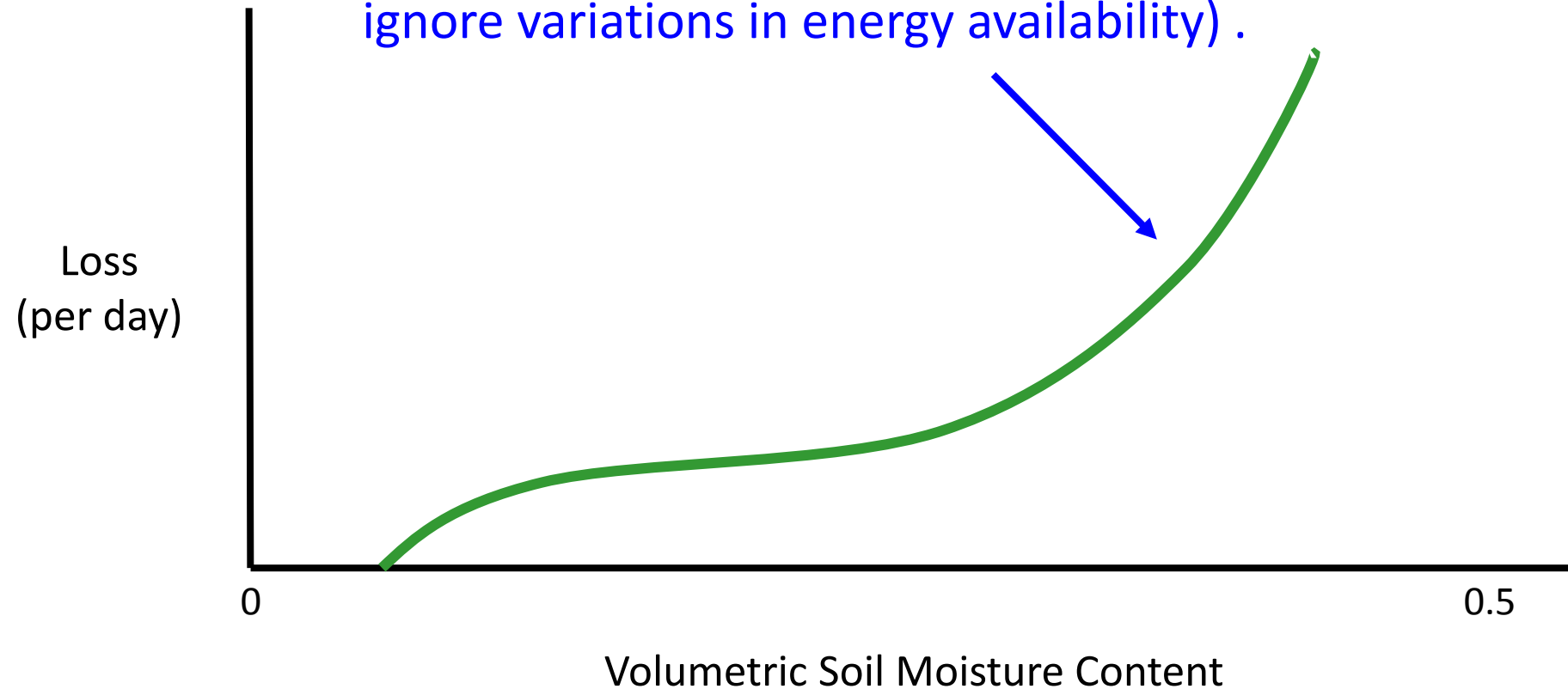
Apparently some rain
occurred between the
measurements. Of
course, there's more to
it than that...

time →





Assume that some monotonic loss function operates on the soil moisture whether or not precipitation falls during an interval. (We focus on the warm season and thus ignore variations in energy availability) .



The idea of the “loss function” is not new...

mainly on porability slow W_K , thereupon a amount cipated water eventually runs off by various routes. After adopting these accretion relations verified by the authors mentioned above and after taking into consideration the effect of evaporation, the rate of the change of soil moisture can be expressed by the following equations.

vs:

if $W = W_{FC}$ and $R_A > E_0$, $\partial W / \partial t = 0$ and $r_f = R_A - E_0$

(18) and

if $W < W_{FC}$, $\partial W / t = R_A - E$

(21)

where R_A is the rate of rainfall and r_f is that of runoff.

(19)

The field capacity of soil varies widely and depends very much upon the kind of soil (for example, the U.S. Department of Agriculture, 1955). It was decided, how-

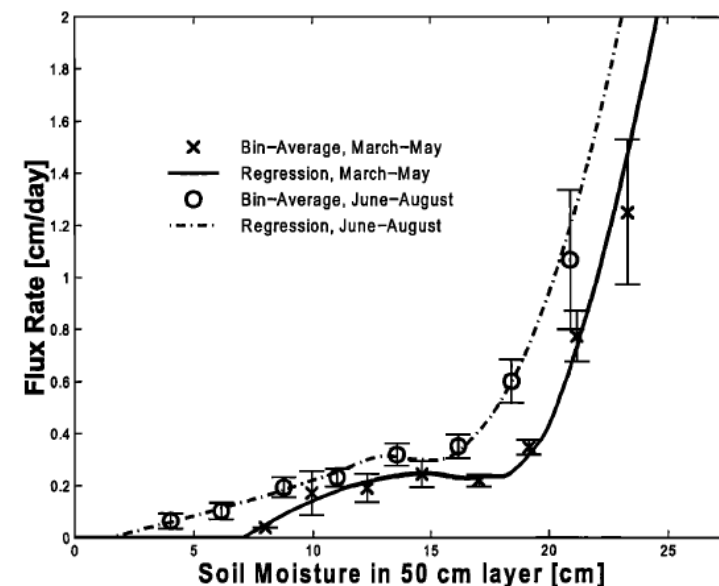


Figure 3. Estimation of moisture-dependent water loss from conditional mean precipitation for sites in Illinois. Note the plateau, presumably corresponding to potential evapotranspiration, and the sharp rise near saturation, presumably corresponding to percolation.

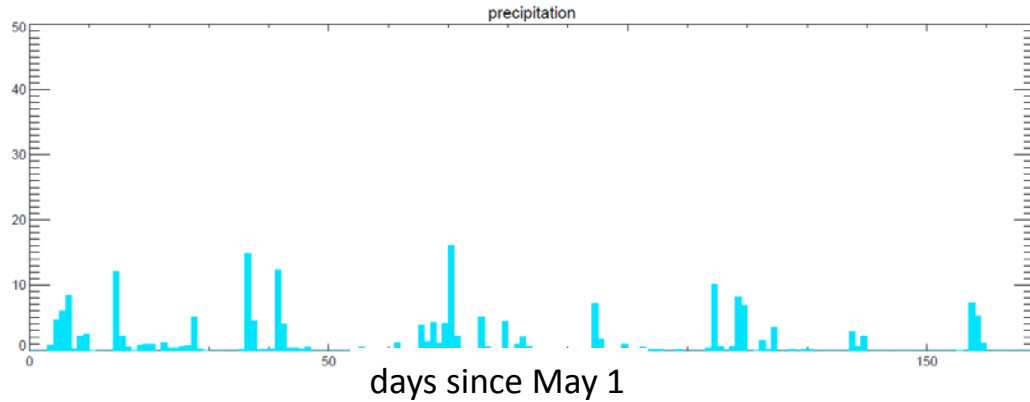
Manabe, Monthly Weath. Rev., 97,739-774, 1969

Salvucci, Water Resour. Res., 37,1357-1365, 2001

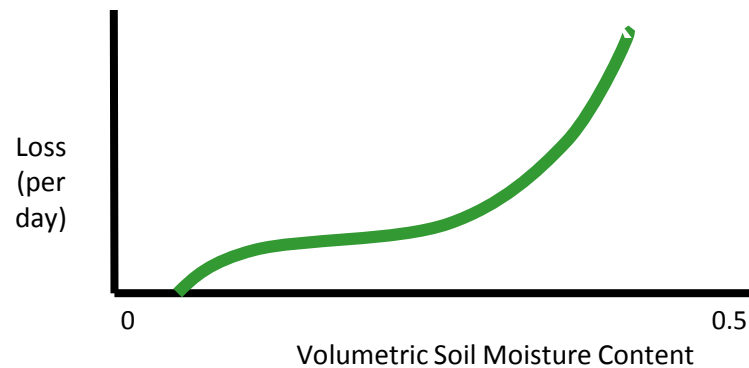
... and its determination has been a chief scientific motivation for the SMAP mission.

What is a simple way of determining a loss function from SMAP Level 2 soil moisture retrievals?

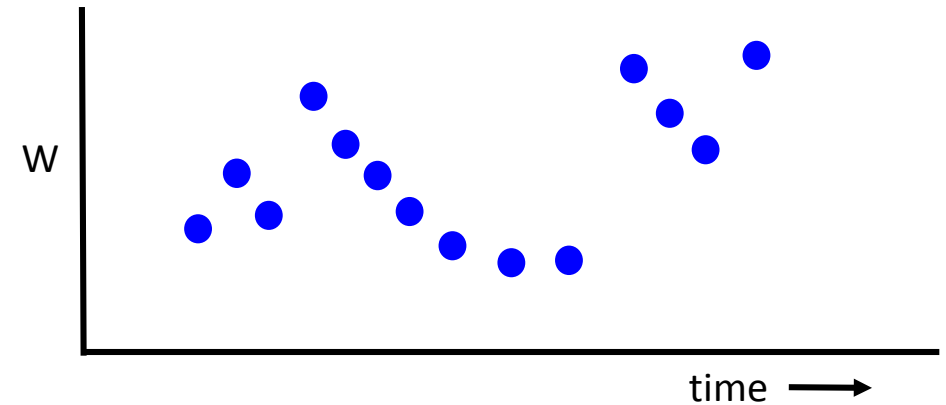
For a given time series of precipitation...



...and an assumed loss function...

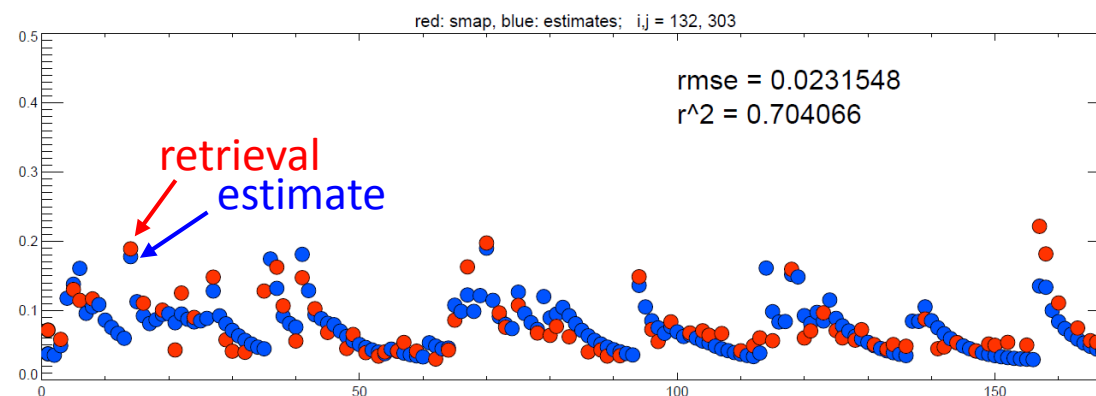
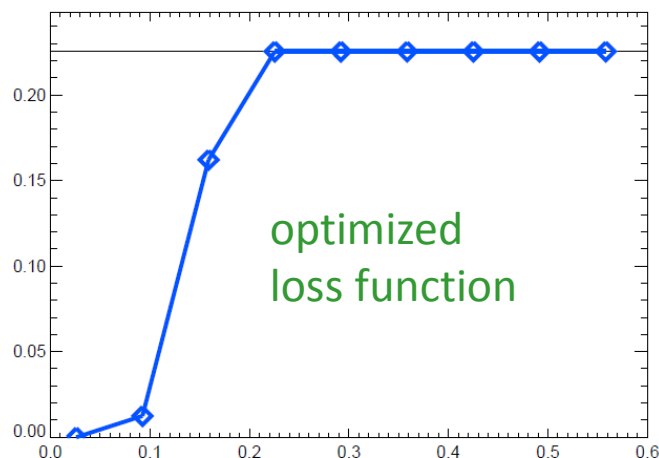
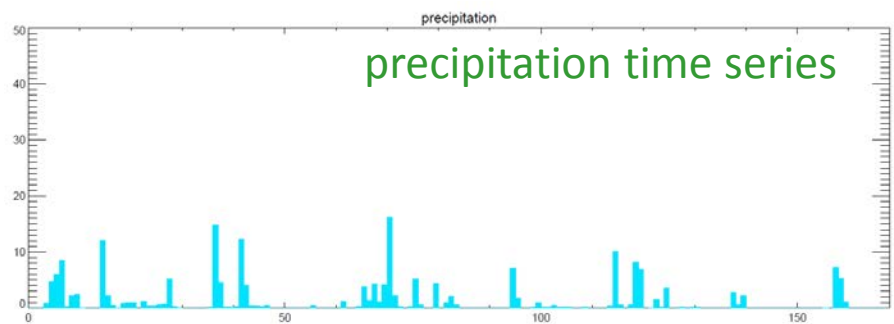


... we can generate a time series of soil moisture contents:

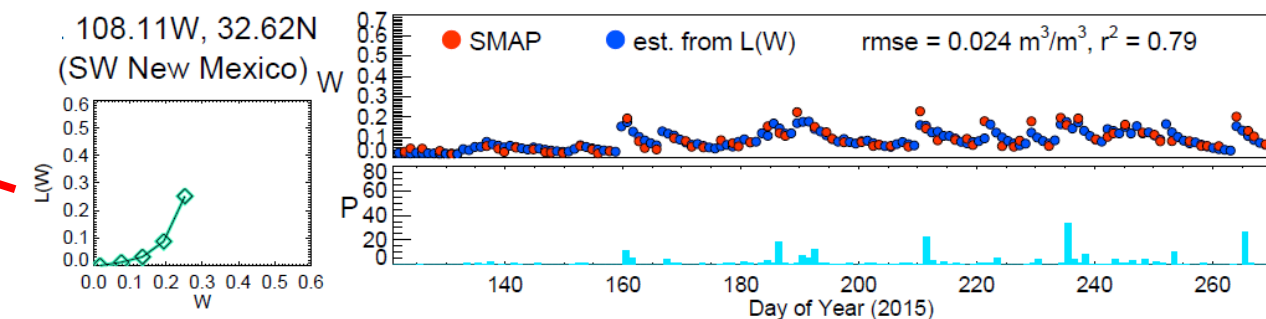
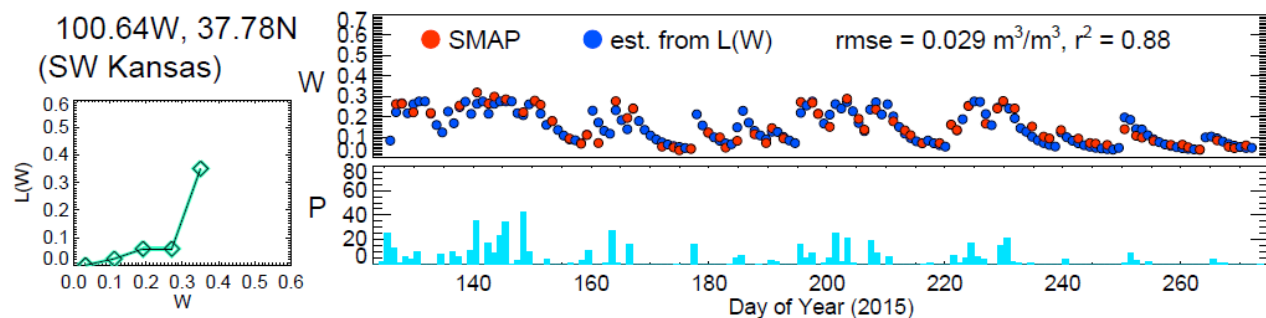
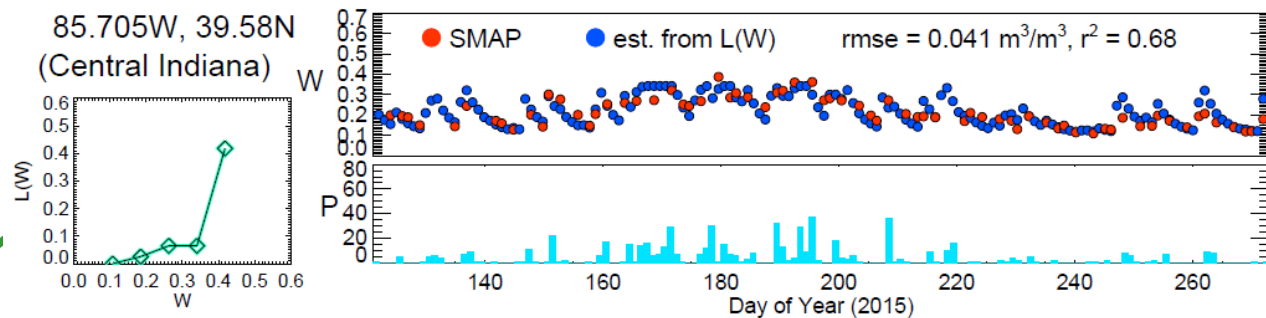
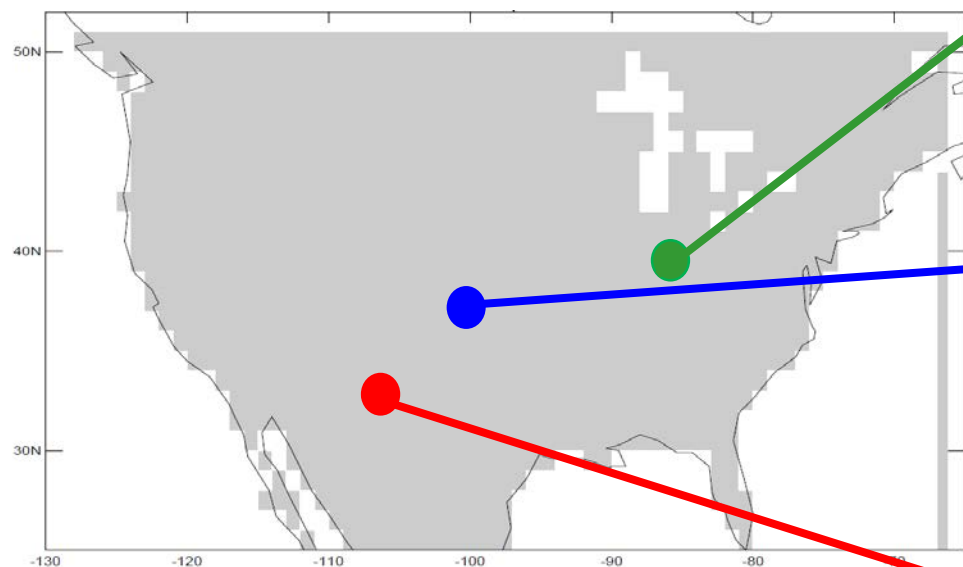


(generated with P data and assumed function alone – no retrievals used)

Procedure: through brute force, we find the loss function that produces the best reproduction (in terms of RMSE) of the SMAP Level 2 soil moisture time series.



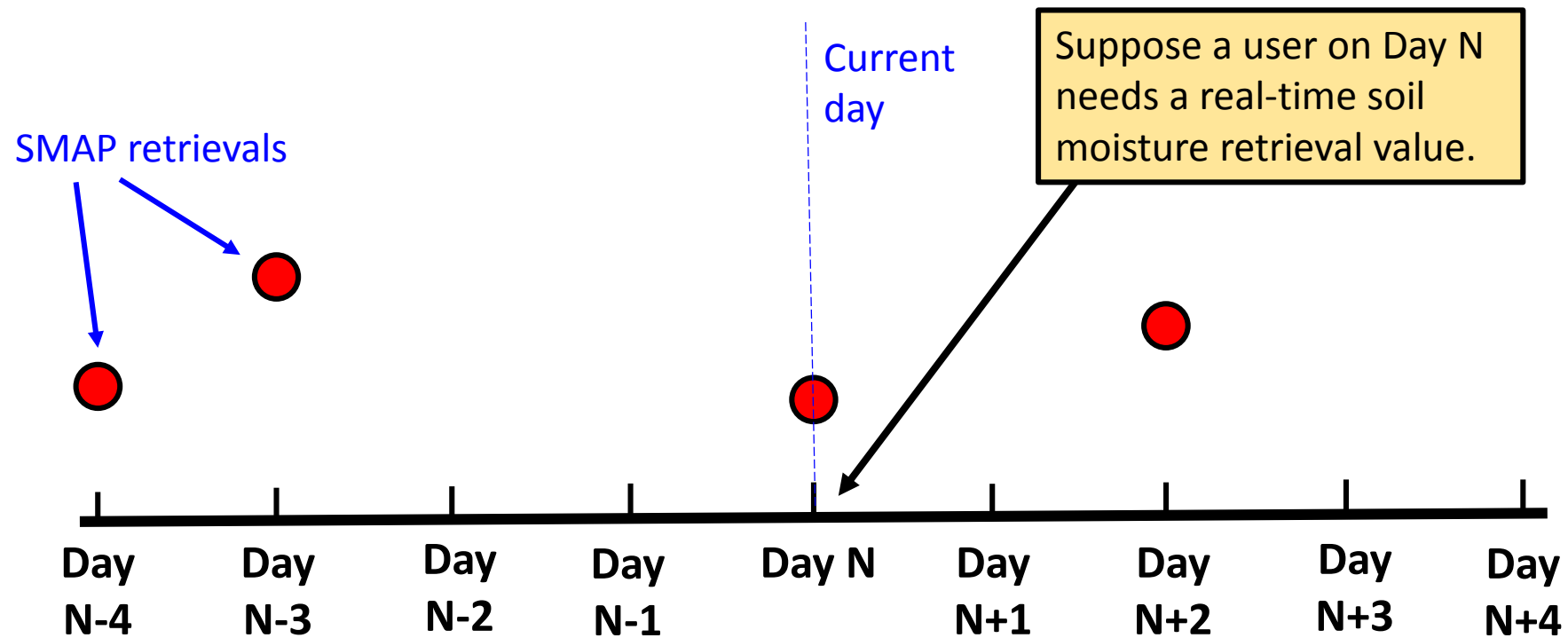
Examples of Loss Functions

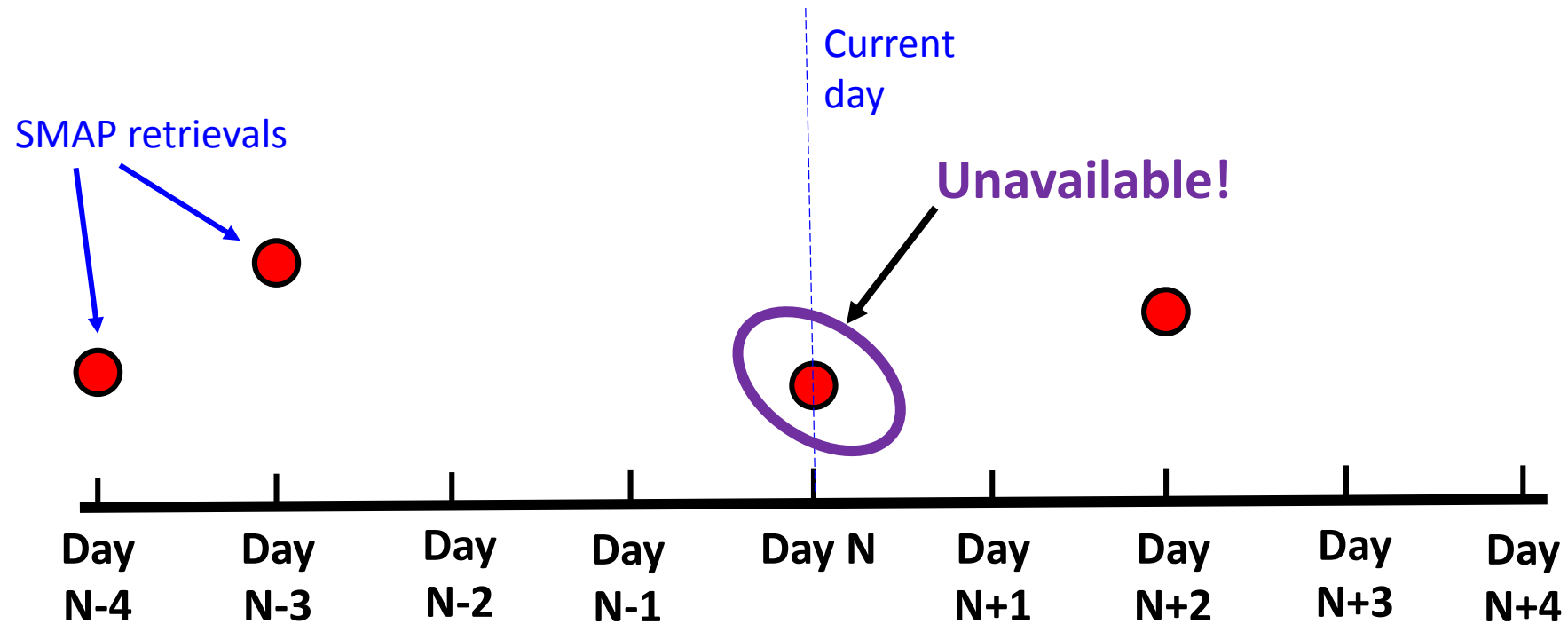


What can we do with these loss functions?

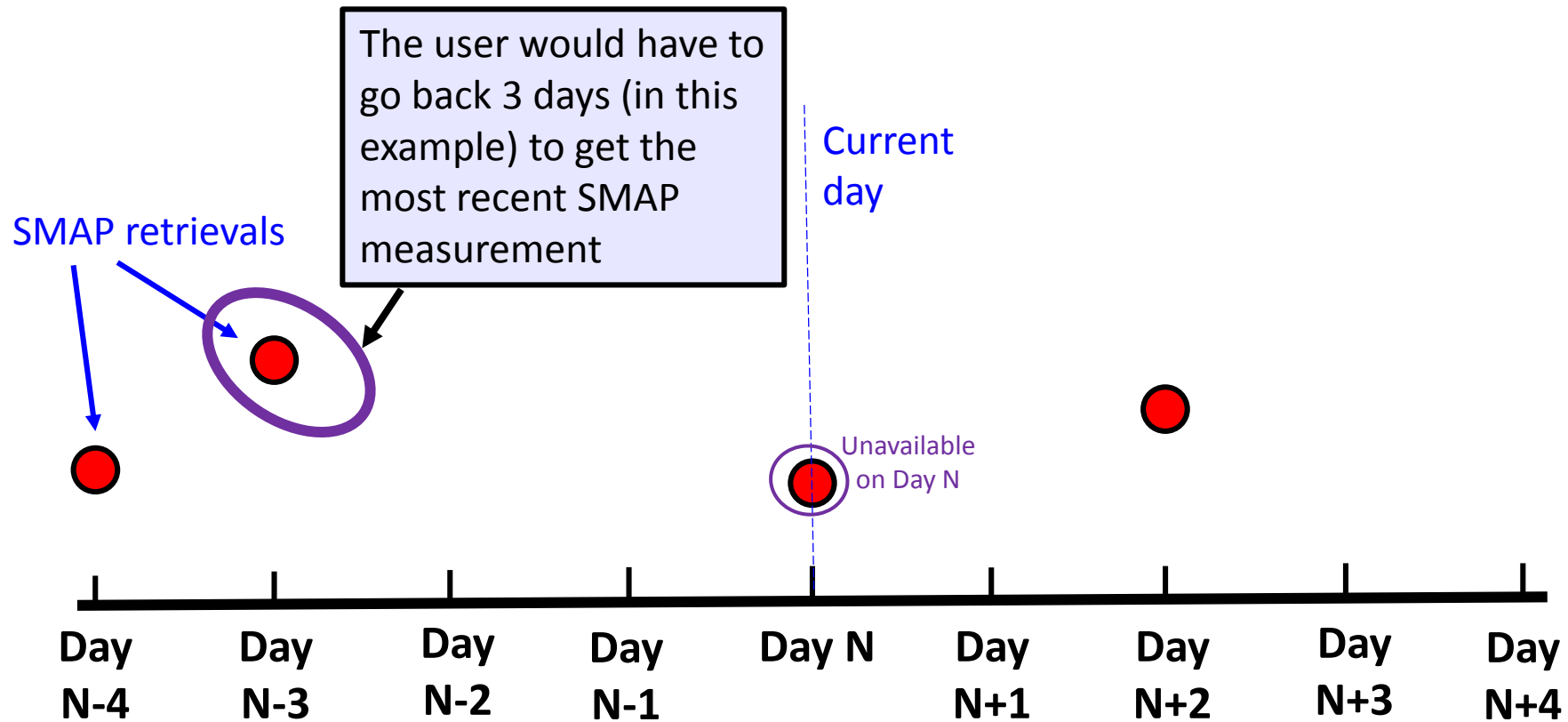
Answer: Many things. Discussed here:

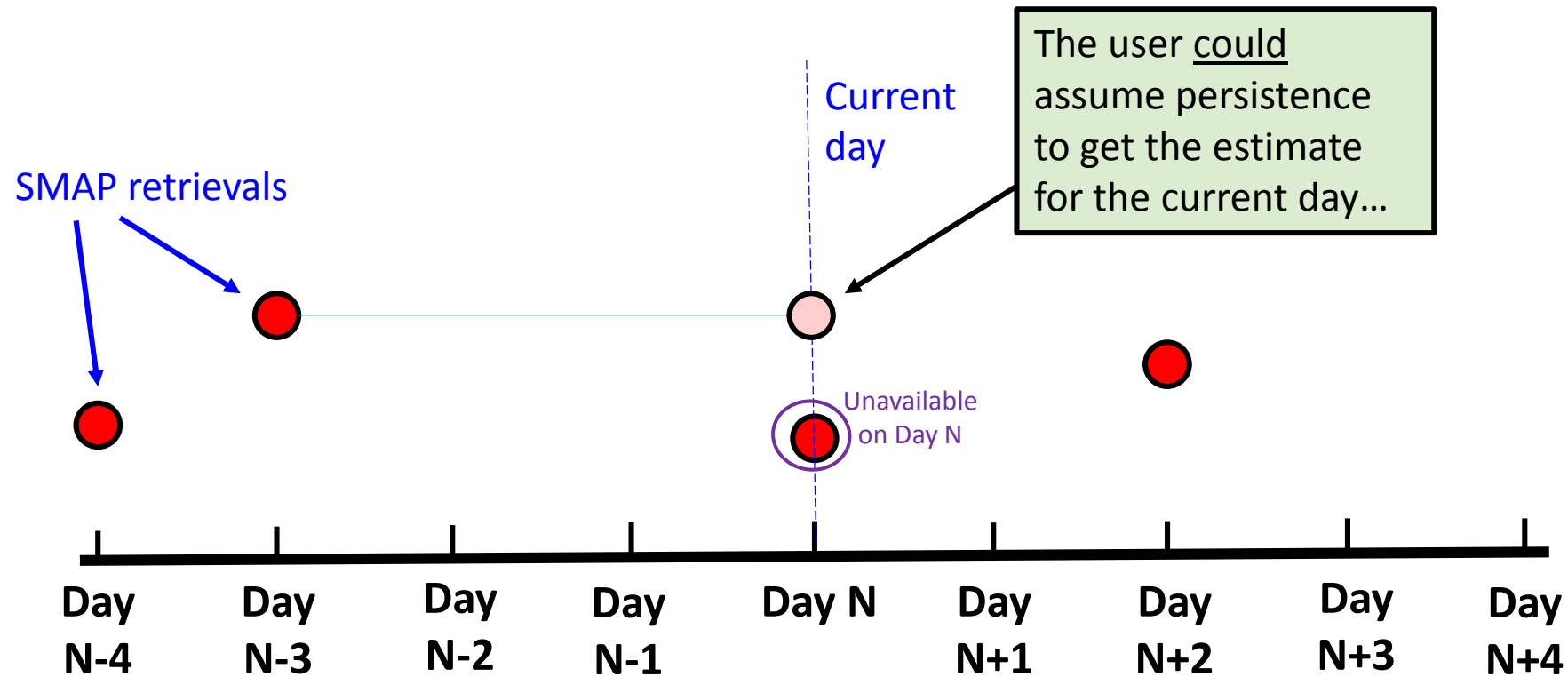
- *Decreasing the latency of SMAP retrieval information*
- *Providing soil moisture forecasts*

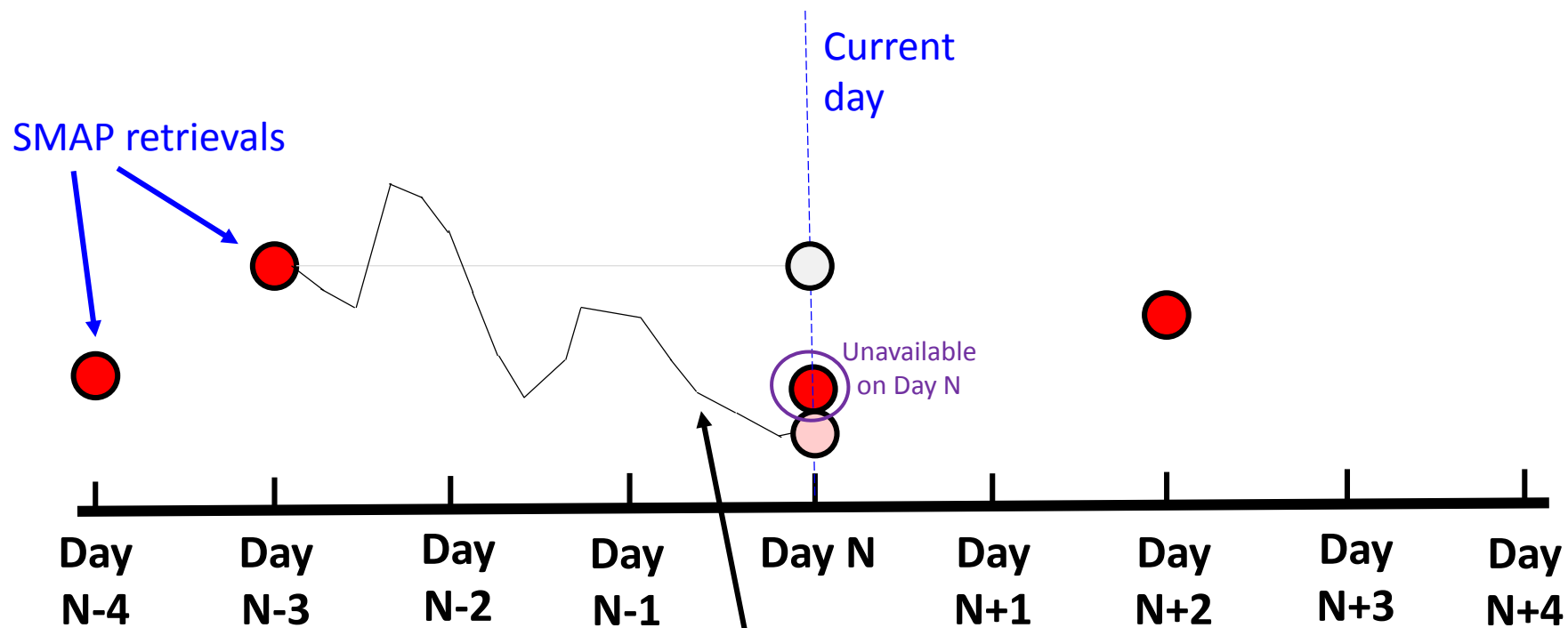


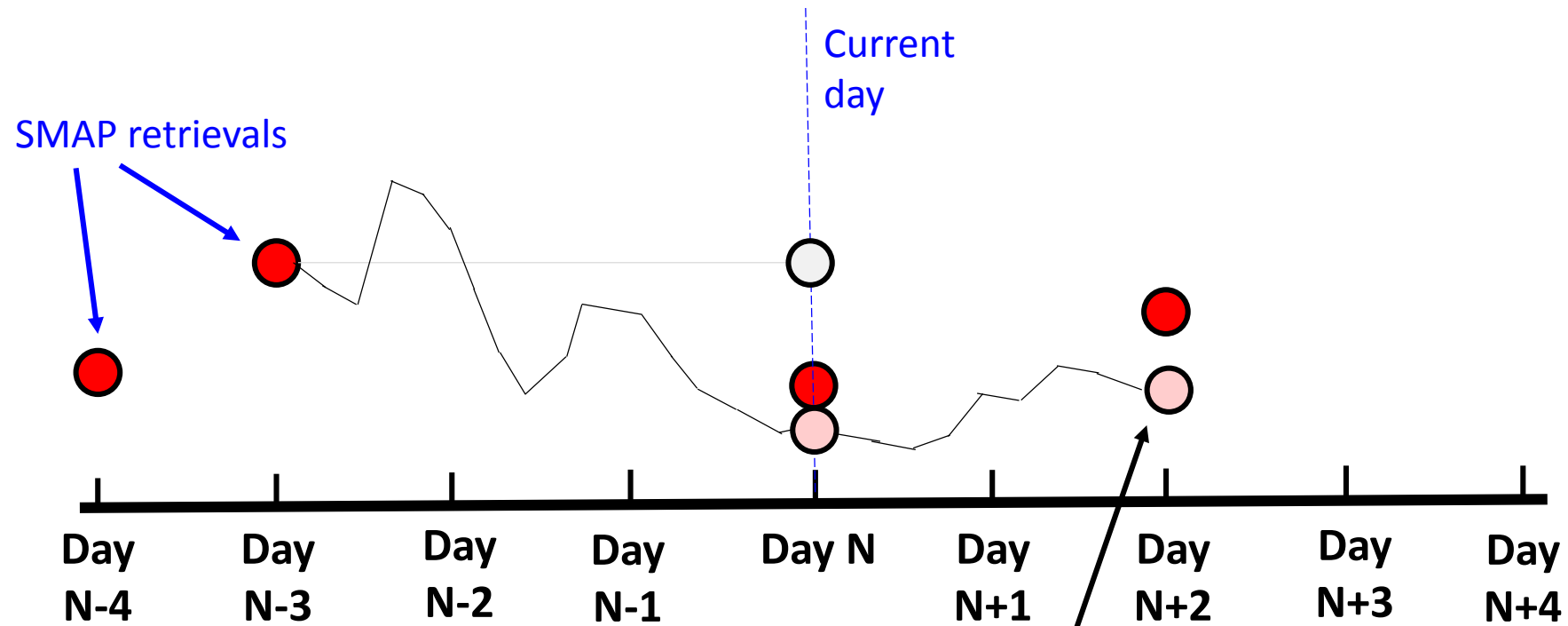


Given product latency, this retrieval value would not be available until Day N+1

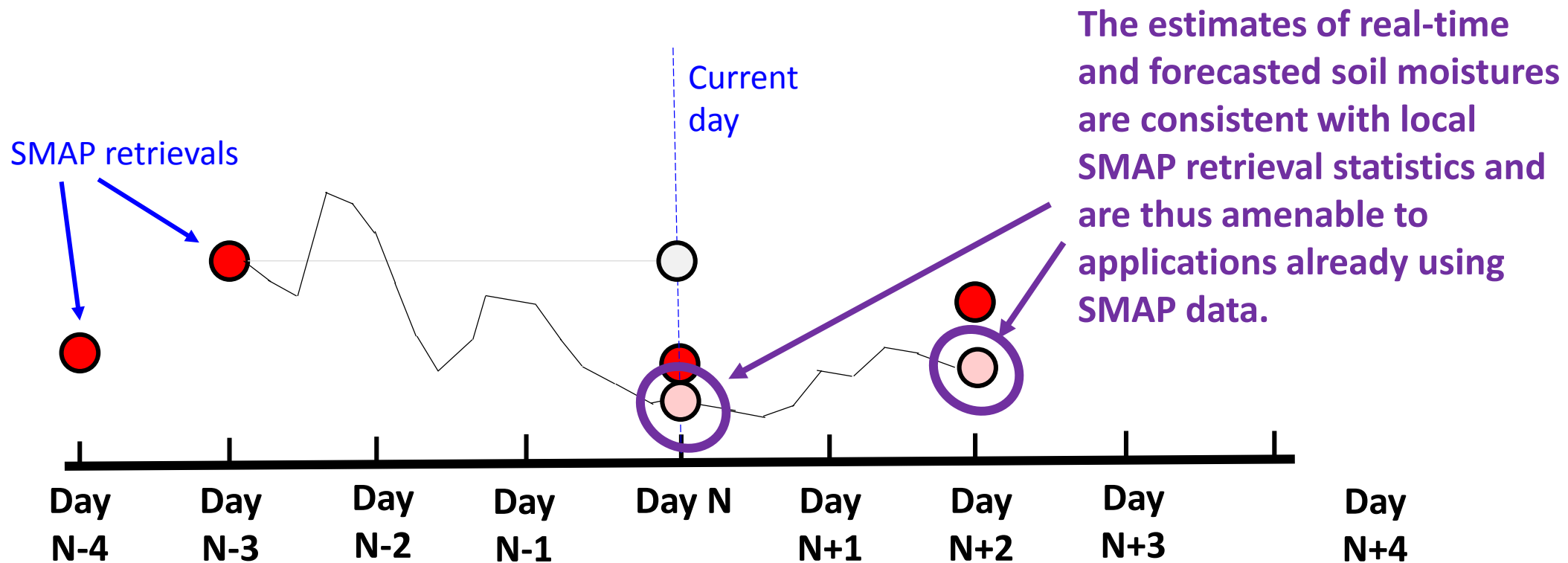








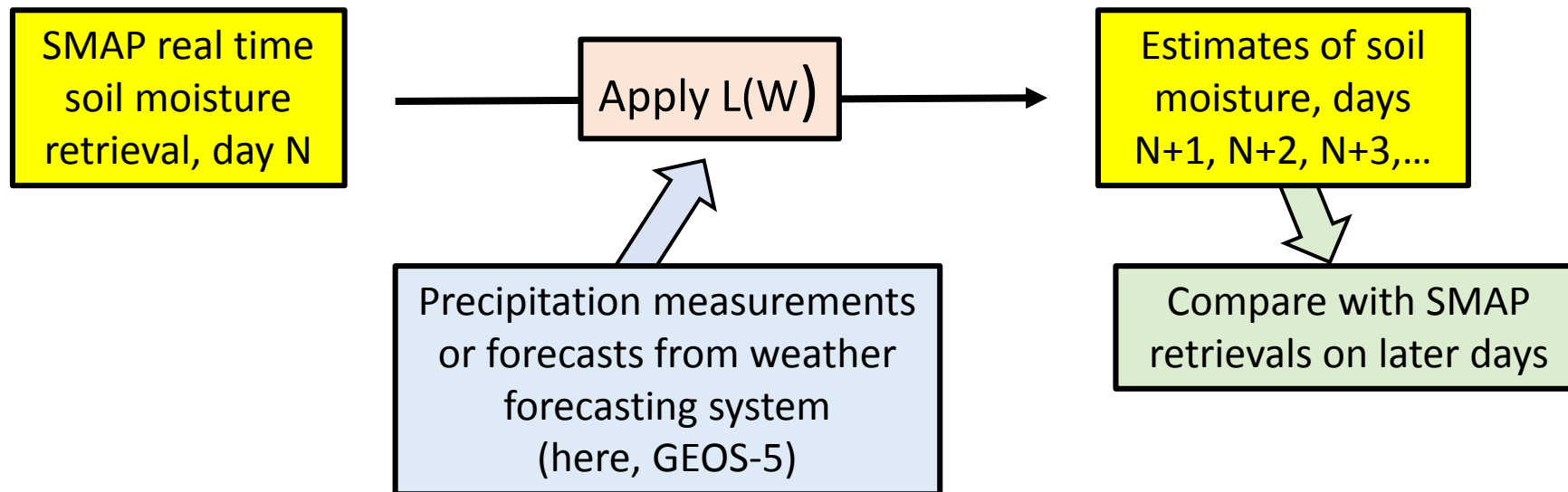
With precipitation forecasts, we could even produce soil moisture forecasts!

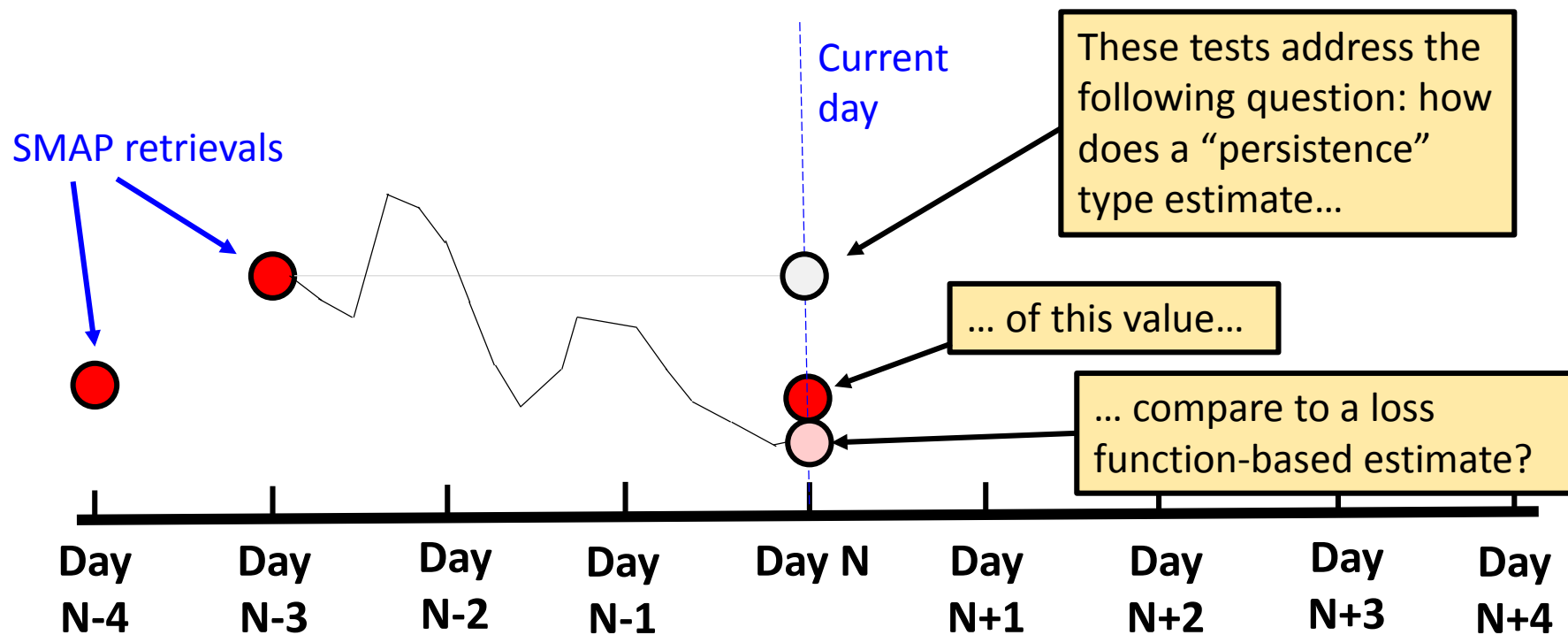


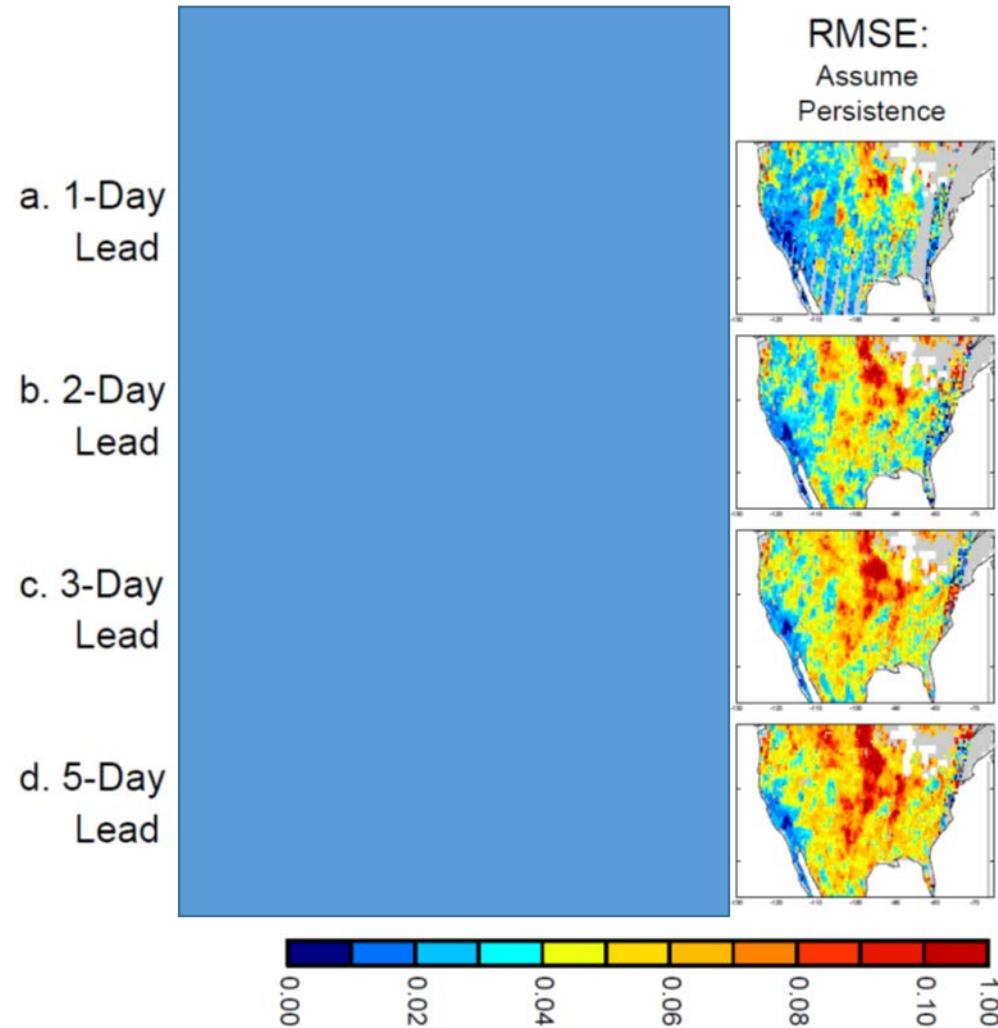
Overall strategy to test such soil moisture estimation:

STEP 1: Derive loss function, $L(W)$, from 2015 precipitation and SMAP data (warm season, May-Sept) over the US.

STEP 2: Utilize this function to predict 2016 soil moisture (again, May-Sept.):



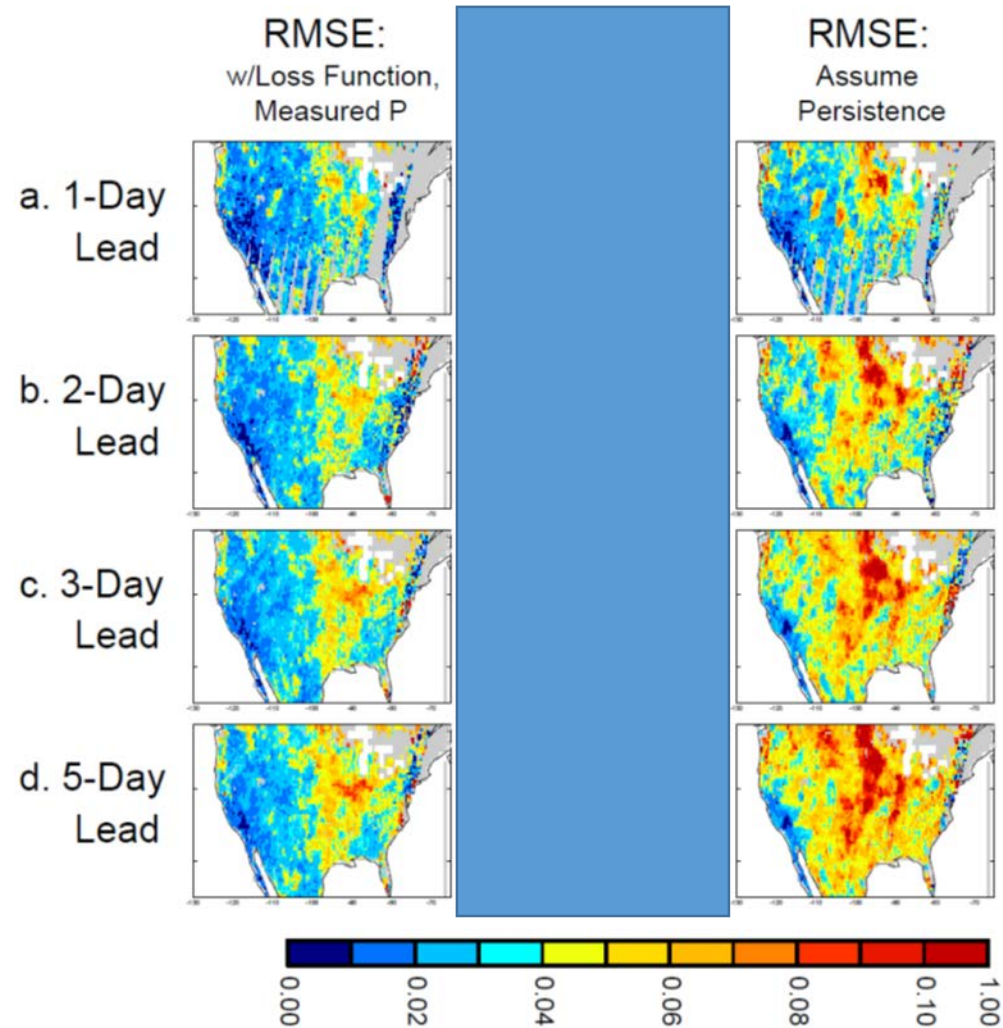




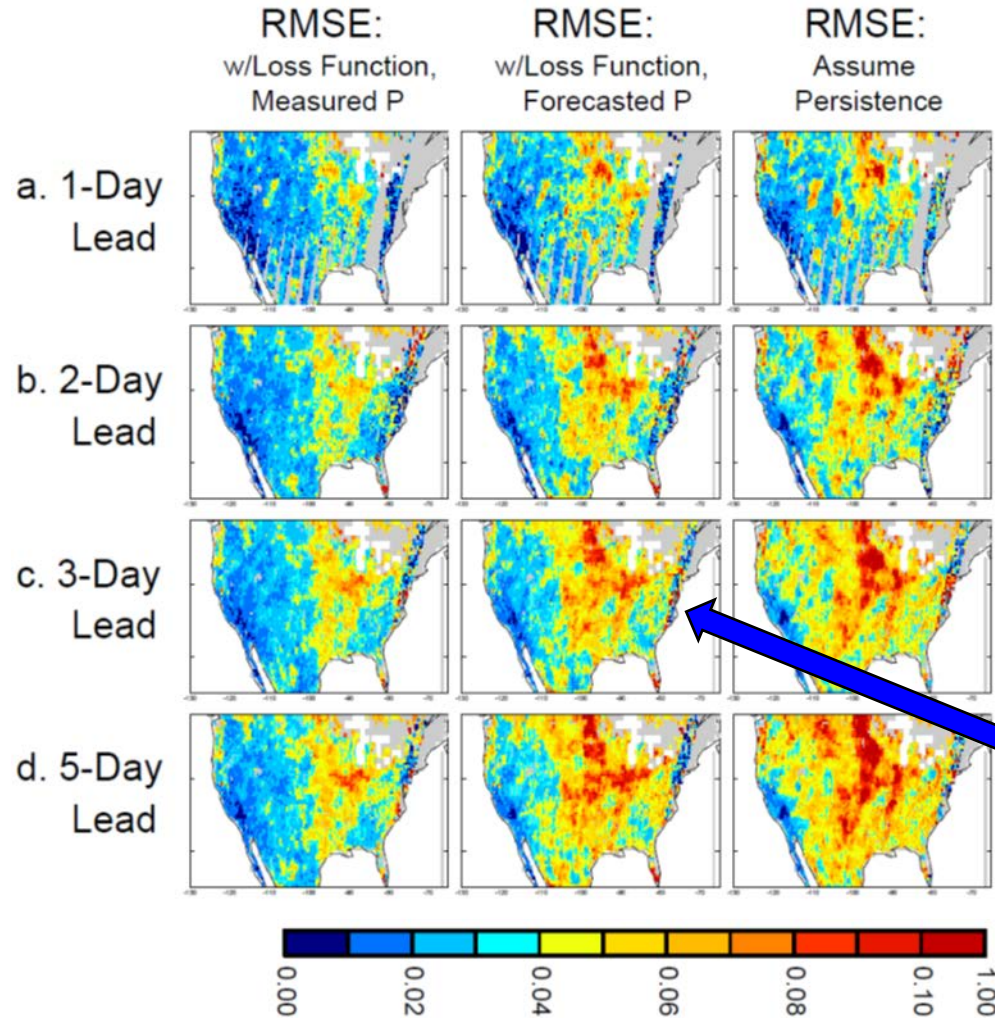
Persistence: If you assign the most recent SMAP retrieval value to the current day, these are the errors you obtain.

Units: m^3/m^3
(volumetric soil moisture)

Loss function: The errors go way down if you use loss functions in conjunction with precipitation information (for real-time estimates)....



Units: m^3/m^3
(volumetric soil moisture)



They also go down, though not as much, if you use loss functions in conjunction with precipitation forecasts (for soil moisture forecasts)....

Overall findings of loss function analysis:

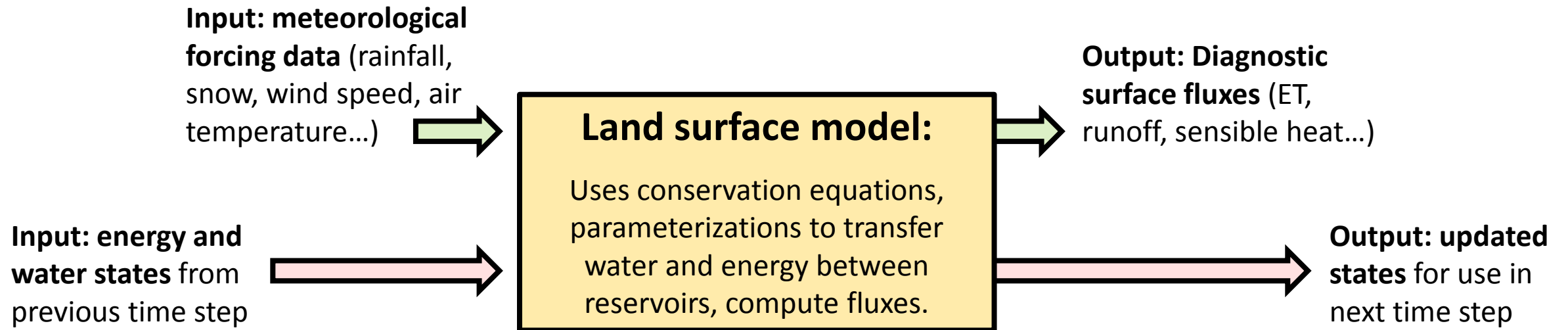
- o Loss functions (descriptions of how soil moisture decreases with evaporation and drainage) can be derived from joint analysis of SMAP data and precipitation data.
- o Using these functions along with precipitation measurements and/or precipitation forecasts, we can produce skillful soil moisture estimates with 0-day latency and even negative latency (soil moisture forecasts).
 - A potentially high impact on applications!

Three topics:

1. “Loss functions”: the characterization of soil moisture dynamics.
2. Data assimilation and calibration: do they access the same information?
3. Estimation of soil moisture with SMAP data.

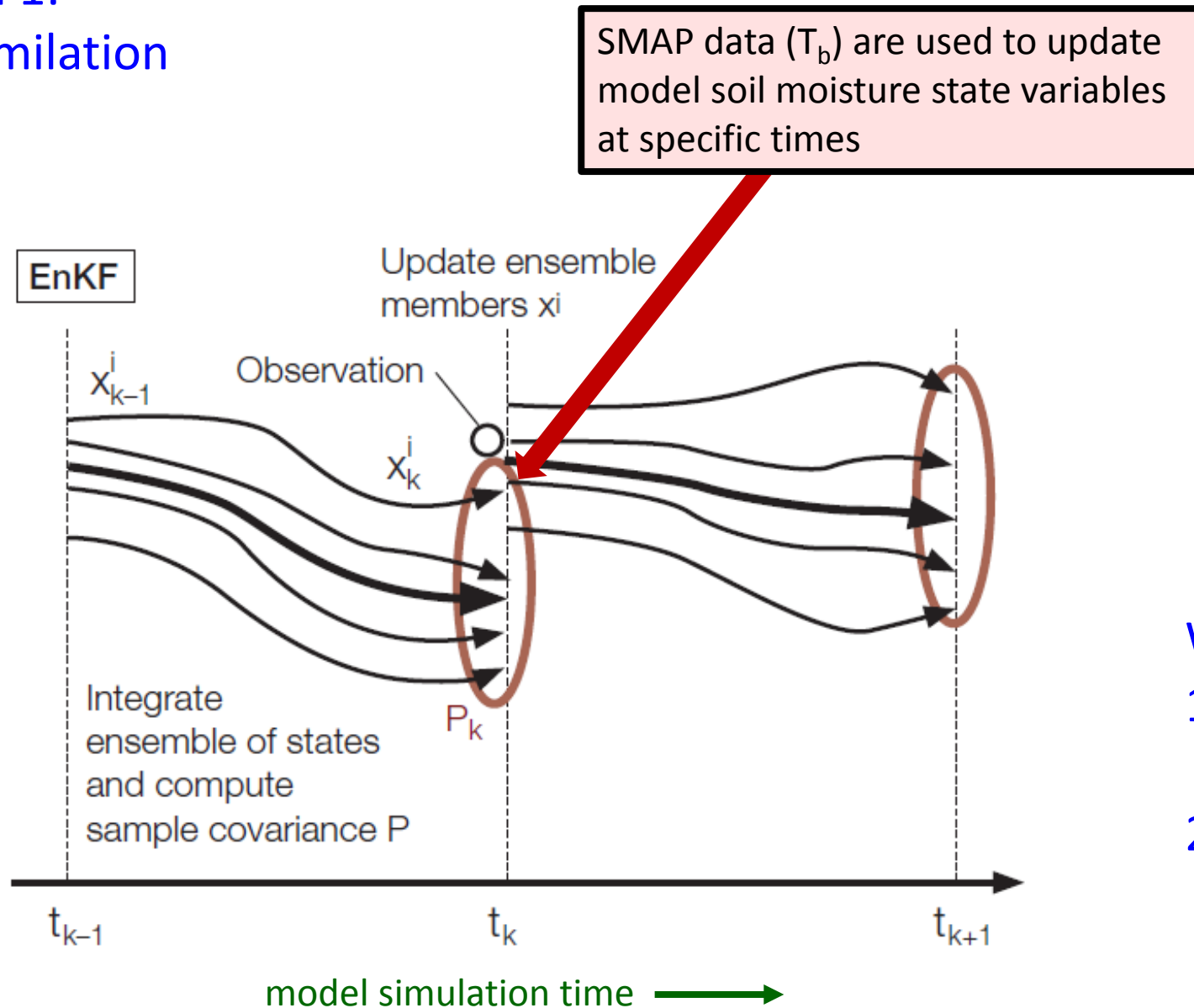
Paper in press:
<https://journals.ametsoc.org/doi/pdf/10.1175/JHM-D-17-0228.1>

Standard hydrological simulation (no satellite data!)



SMAP data can be used to improve such hydrological simulation in at least two ways...

Approach 1: Data assimilation

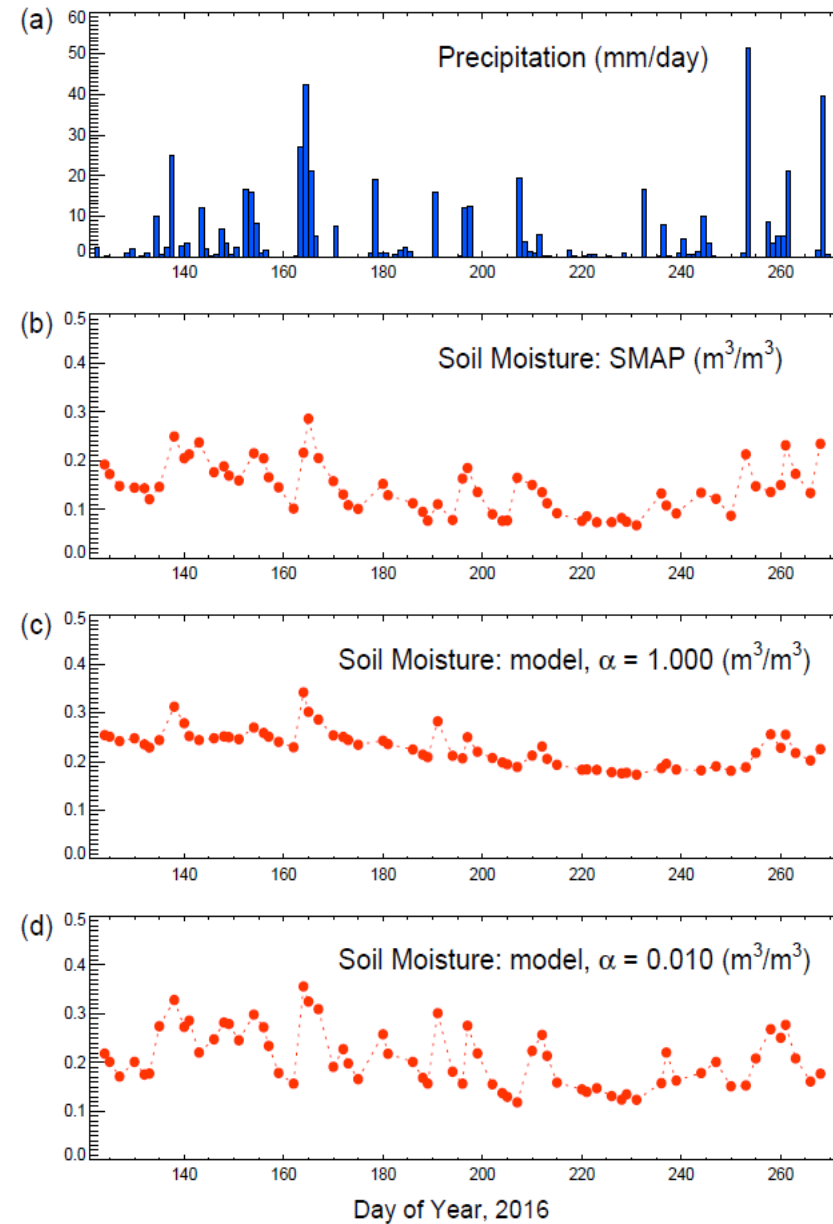


- With data assimilation:
- 1) the land model is used “as is”.
 - 2) Only the prognostic states are modified.

Approach 2: Model calibration

With model calibration:

- 1) the land model itself is changed – values of model parameter(s) are optimized.
- 2) SMAP data contribute to the parameter calibration but not to the updating of the prognostic states during a simulation.



The present study focuses on the calibration of a certain recharge parameter.

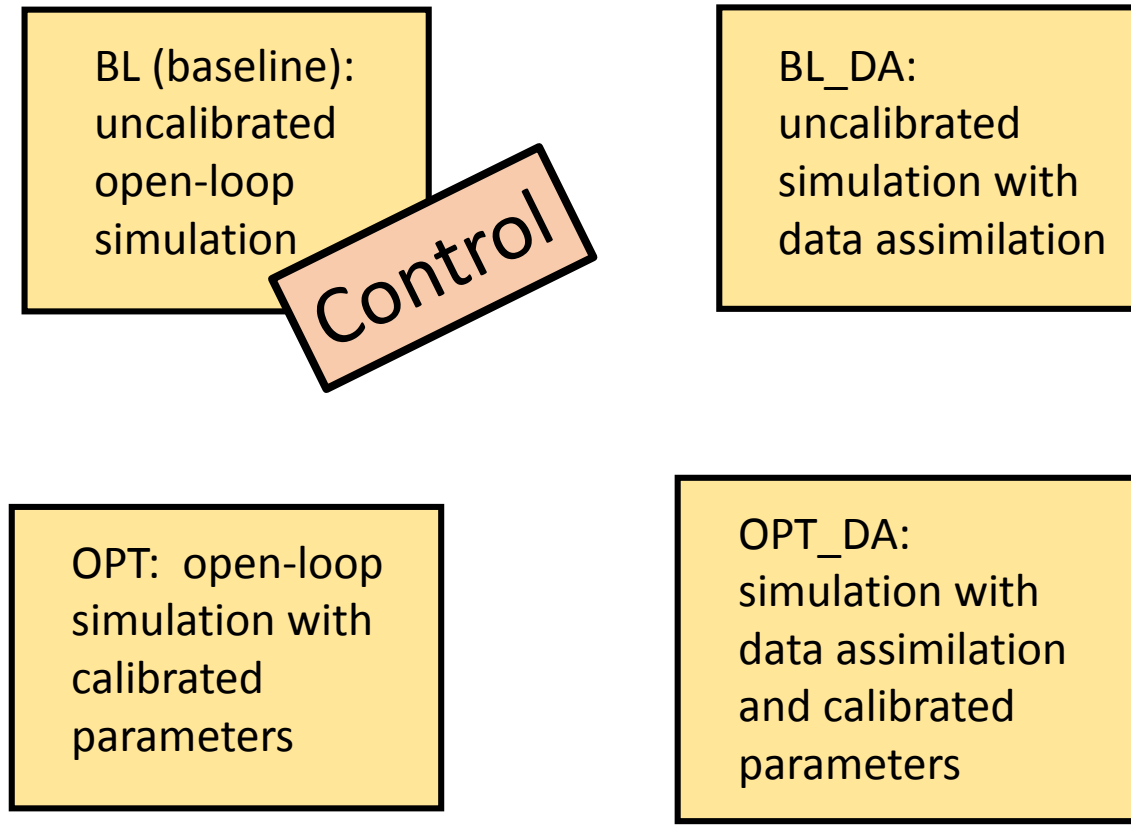
The value used in the default model gives soil moisture recessions that are too slow.

Calibrating the parameter allows more realistic recessions.

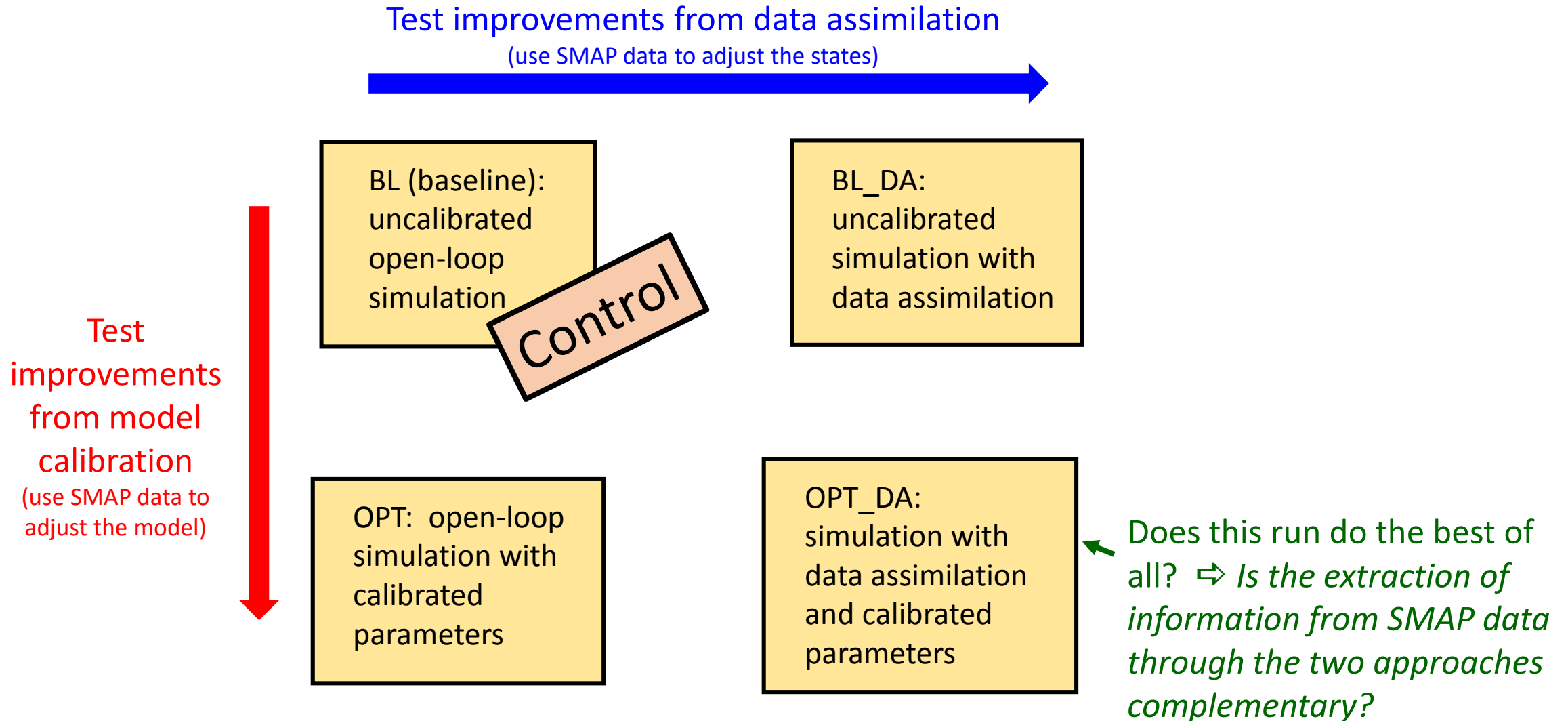
Again, the main question here:

Do these two approaches to using SMAP data extract complementary information?

This study: Perform four offline hydrological simulations with Catchment LSM, covering April 2015 – March 2017. Evaluate simulated soil moisture and streamflow against observations.

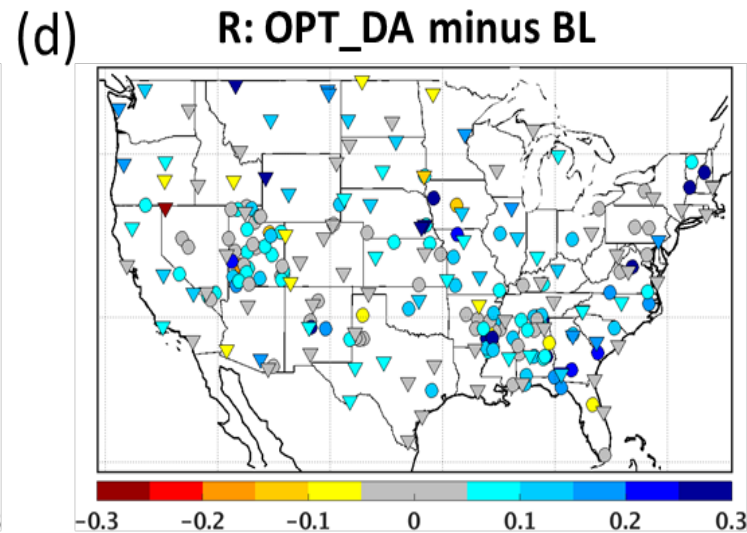
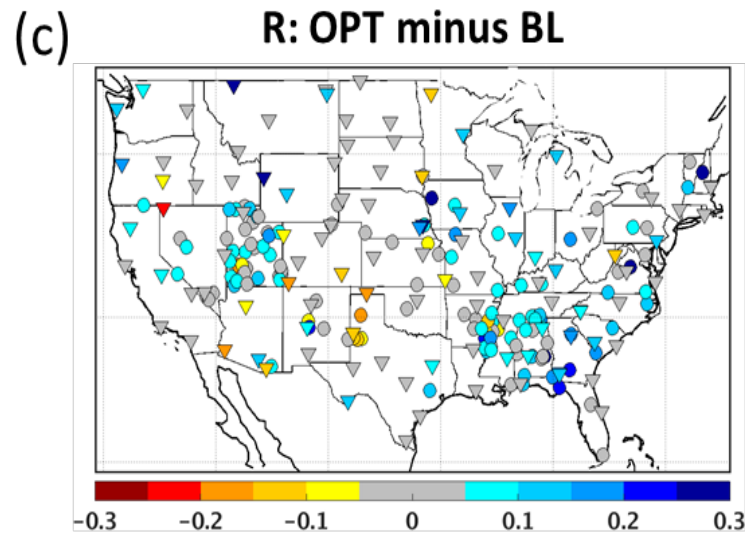
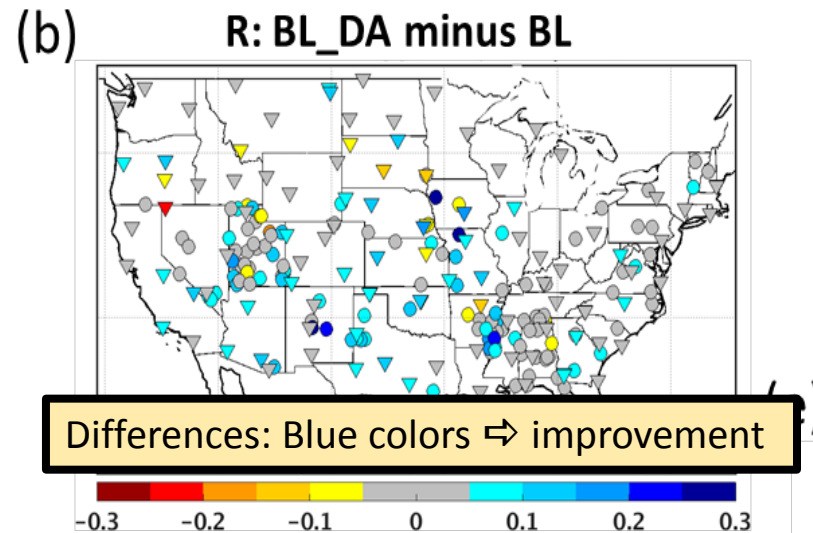
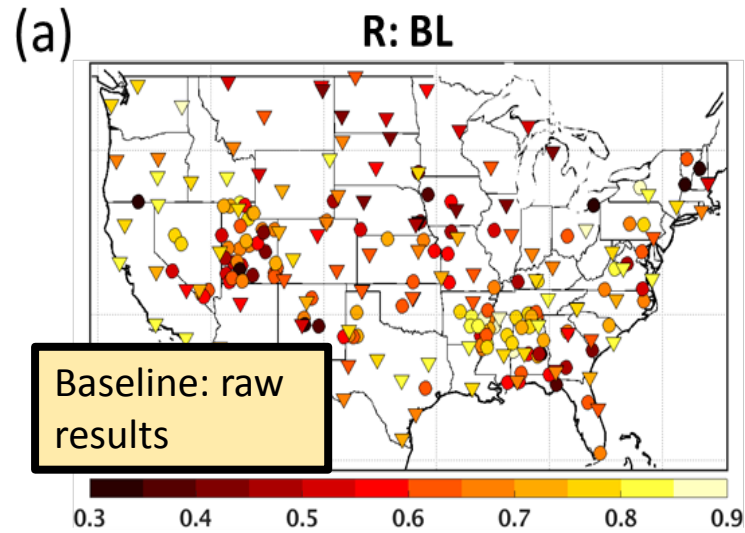


This study: Perform offline four hydrological simulations with Catchment LSM, covering April 2015 – March 2017. Evaluate simulated soil moisture and streamflow against observations.

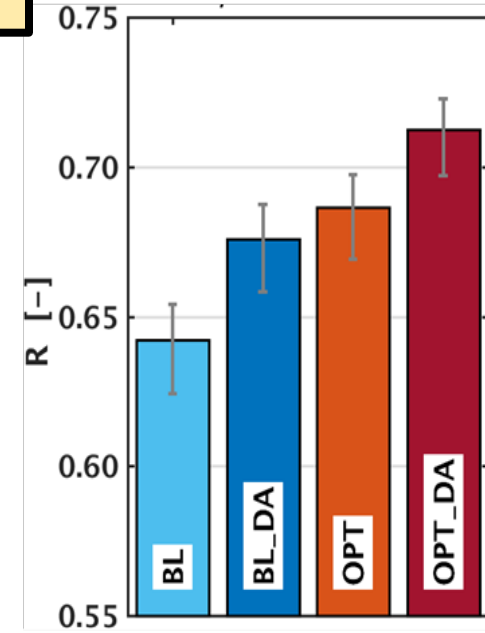


Test improvements from data assimilation

Test improvements from model calibration



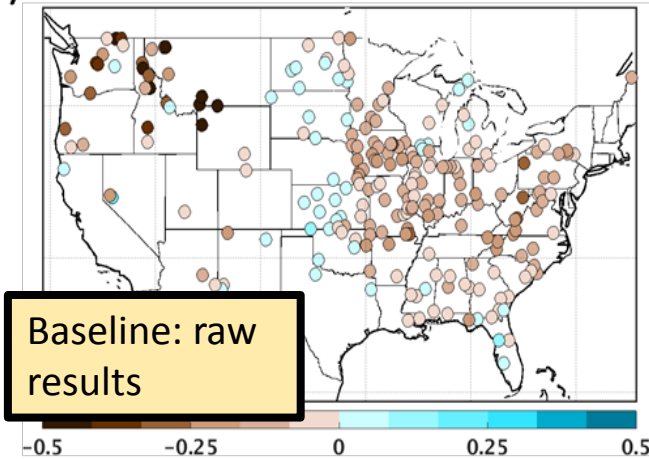
Timing of simulated surface soil moisture (r vs obs at USCRN, SCAN sites)



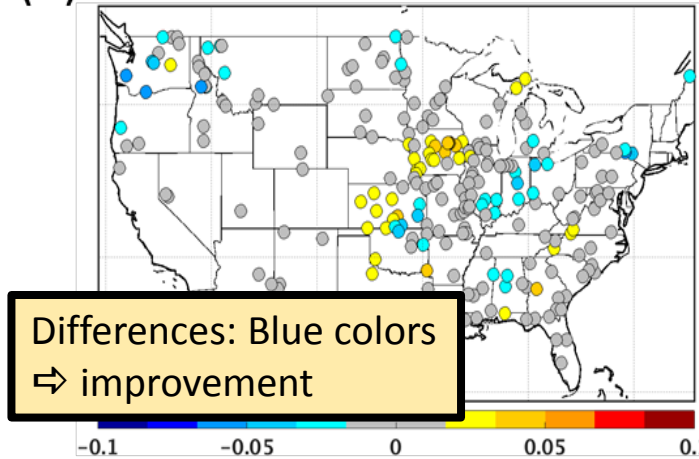
Test improvements from data assimilation

Test improvements from model calibration

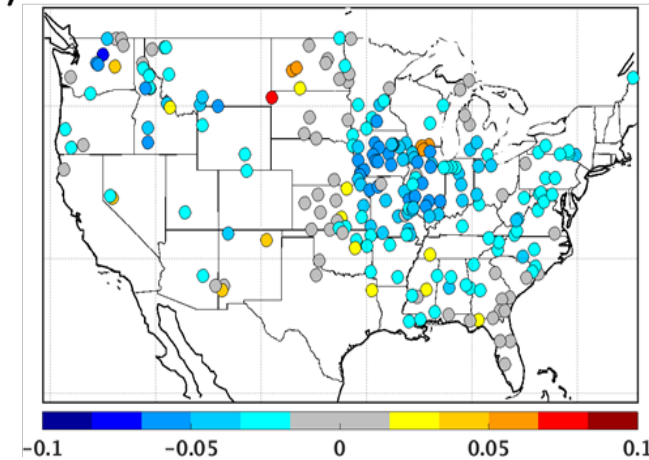
(a) Q/P error: BL



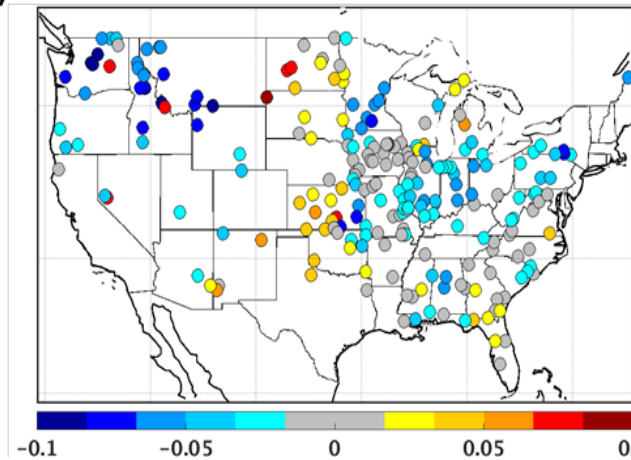
(b) Abs(Q/P error): BL_DA minus BL



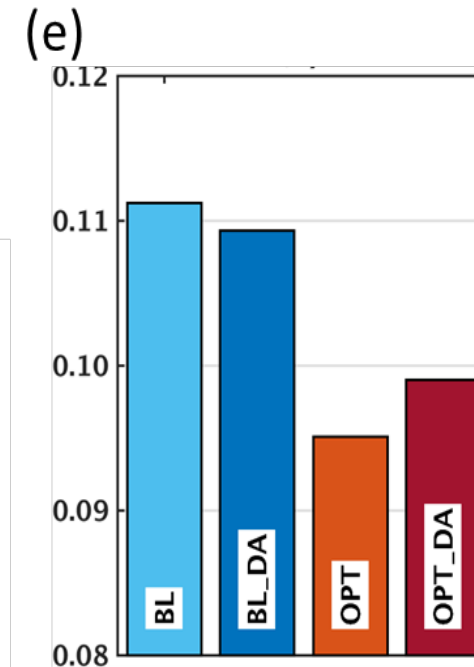
(c) Abs(Q/P error): OPT minus BL



(d) Abs(Q/P error): OPT_DA minus BL



**Error (bias) of
simulated
runoff ratio**
(vs results at
unregulated USGS
basins)



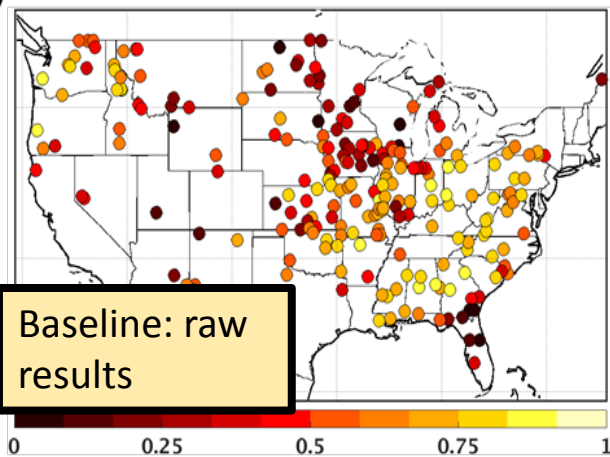
What's their favorite dessert?



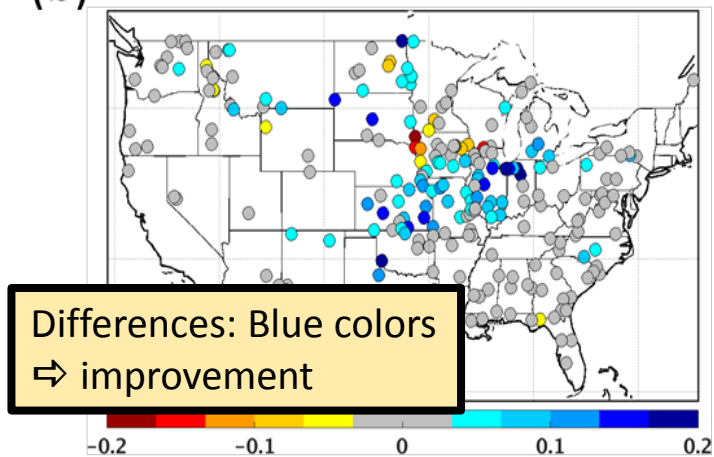
Test improvements from data assimilation

Test improvements from model calibration

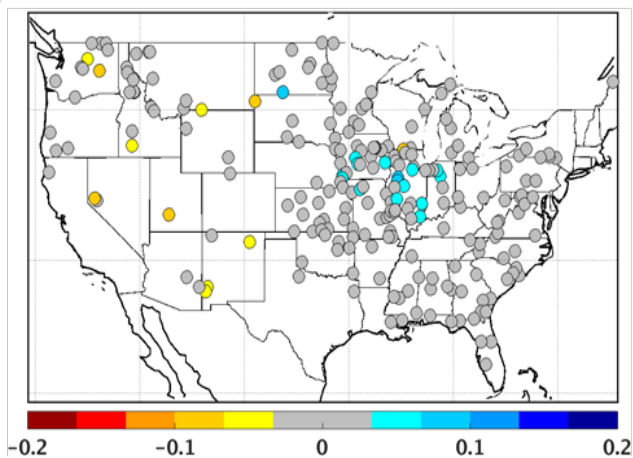
(a) R(streamflow): BL



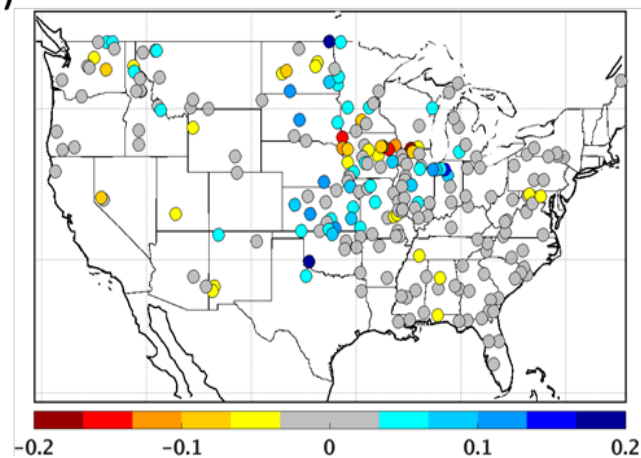
(b) R(streamflow): BL_DA minus BL



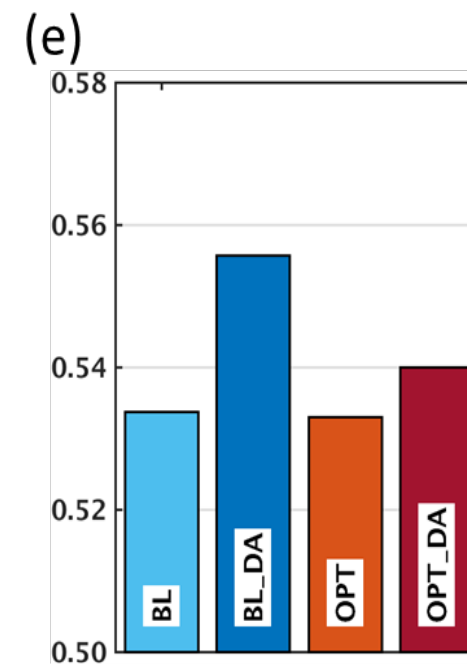
(c) R(streamflow): OPT minus BL



(d) R(streamflow): OPT_DA minus BL



Timing of
simulated
runoff (r vs obs)



To summarize the impacts:

Soil moisture

		Data assimilation	Model calibration	
Soil moisture	ubRMSE	Improvement	No change	(see paper)
	Abs(bias)	No change	Improvement	(see paper)
	timing (r)	Improvement	Improvement	

Streamflow

Streamflow	runoff ratio	No change	Improvement
	Abs(bias)		
Streamflow	timing (r)	Improvement	No change

As it turns out, we can explain at least some of this complementarity...

		Data assimilation	Model calibration
Streamflow	runoff ratio Abs(bias)	No change	Improvement
	timing (r)	Improvement	No change

In data assimilation, T_B is transformed to a value consistent with the land model's climatology prior to assimilation $\Rightarrow$ little impact on the climatology of the model.

With model calibration, the nature of the model itself is changed $\Rightarrow$ so is the climatology of the simulated fluxes.

Streamflow

	Data assimilation	Model calibration
runoff ratio Abs(bias)	No change	Improvement
timing (r)	Improvement	No change

Streamflow

	Data assimilation	Model calibration
runoff ratio Abs(bias)	No change	Improvement
timing (r)	Improvement	No change

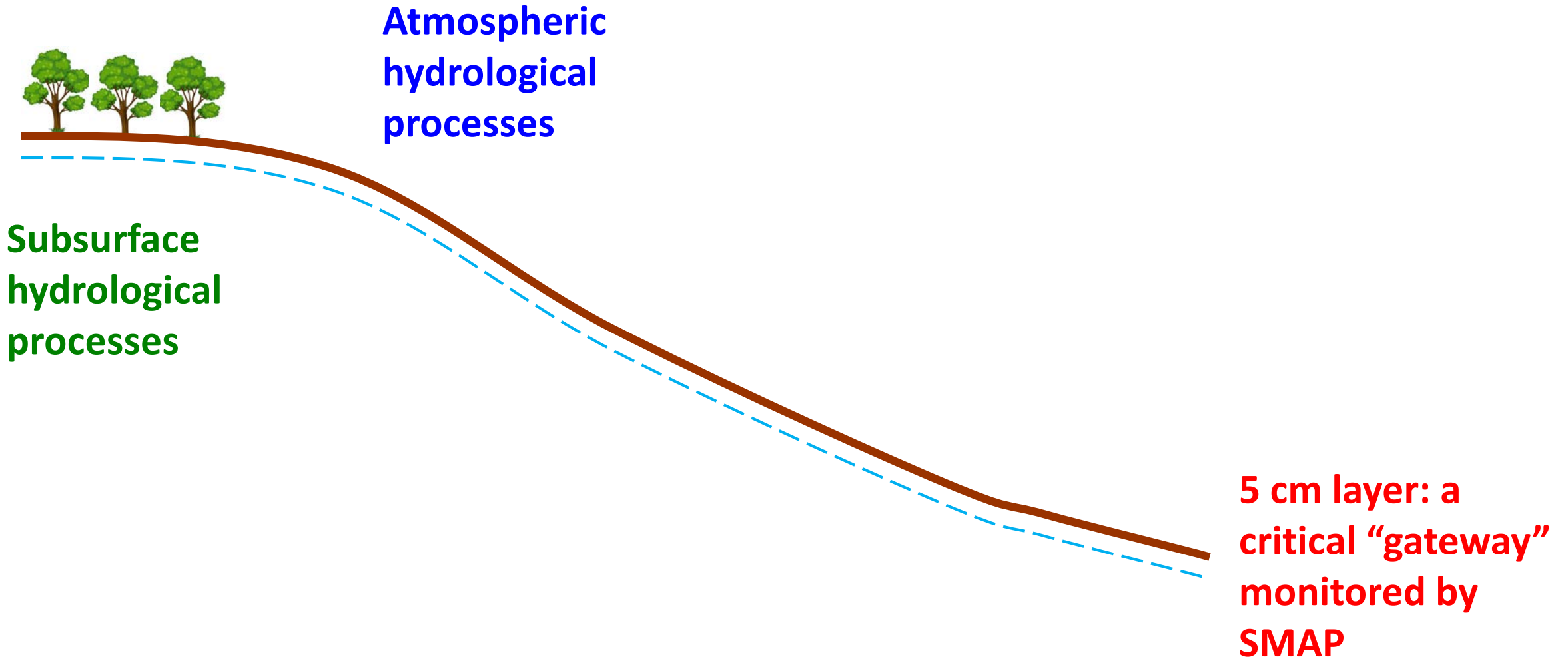
Data assimilation, however, does correct for errors in meteorological forcing, which is critical for streamflow timing.

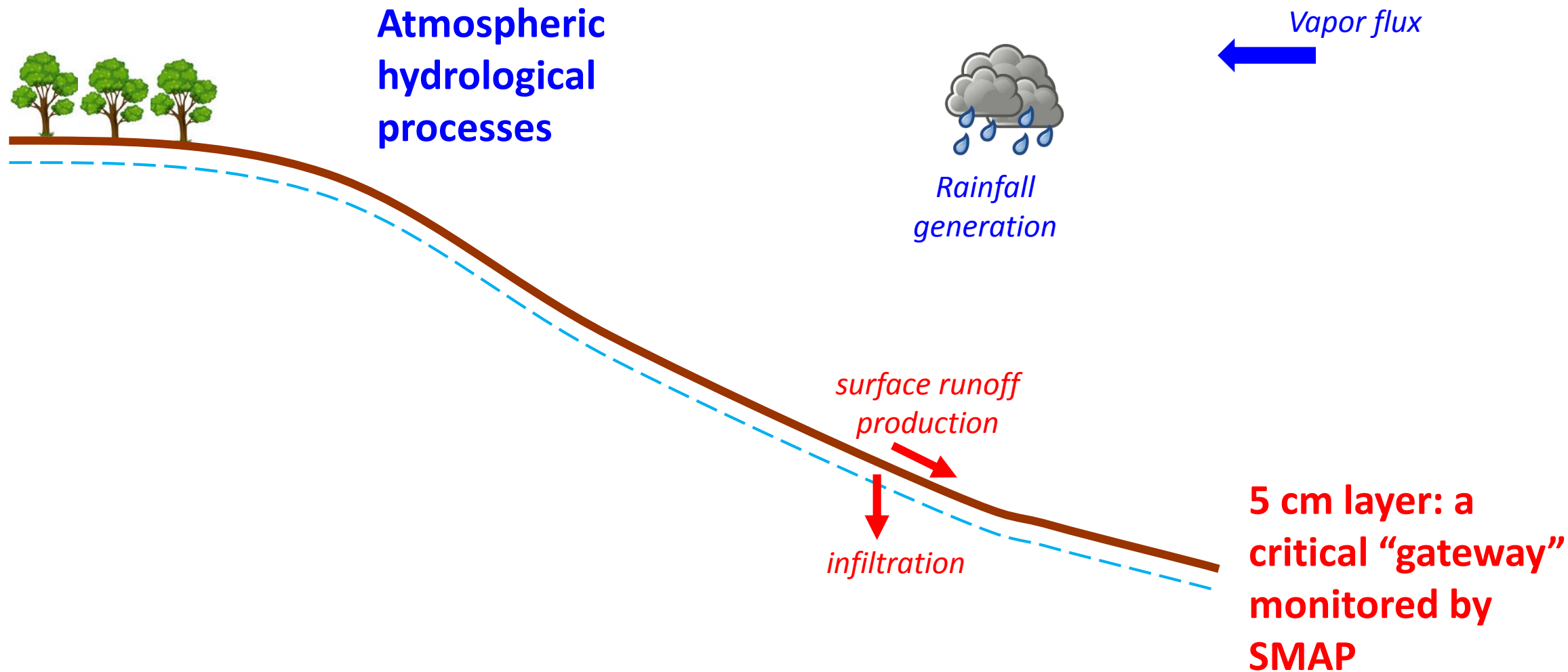
Calibration, which does not correct for errors in meteorological forcing, gains no such benefit.

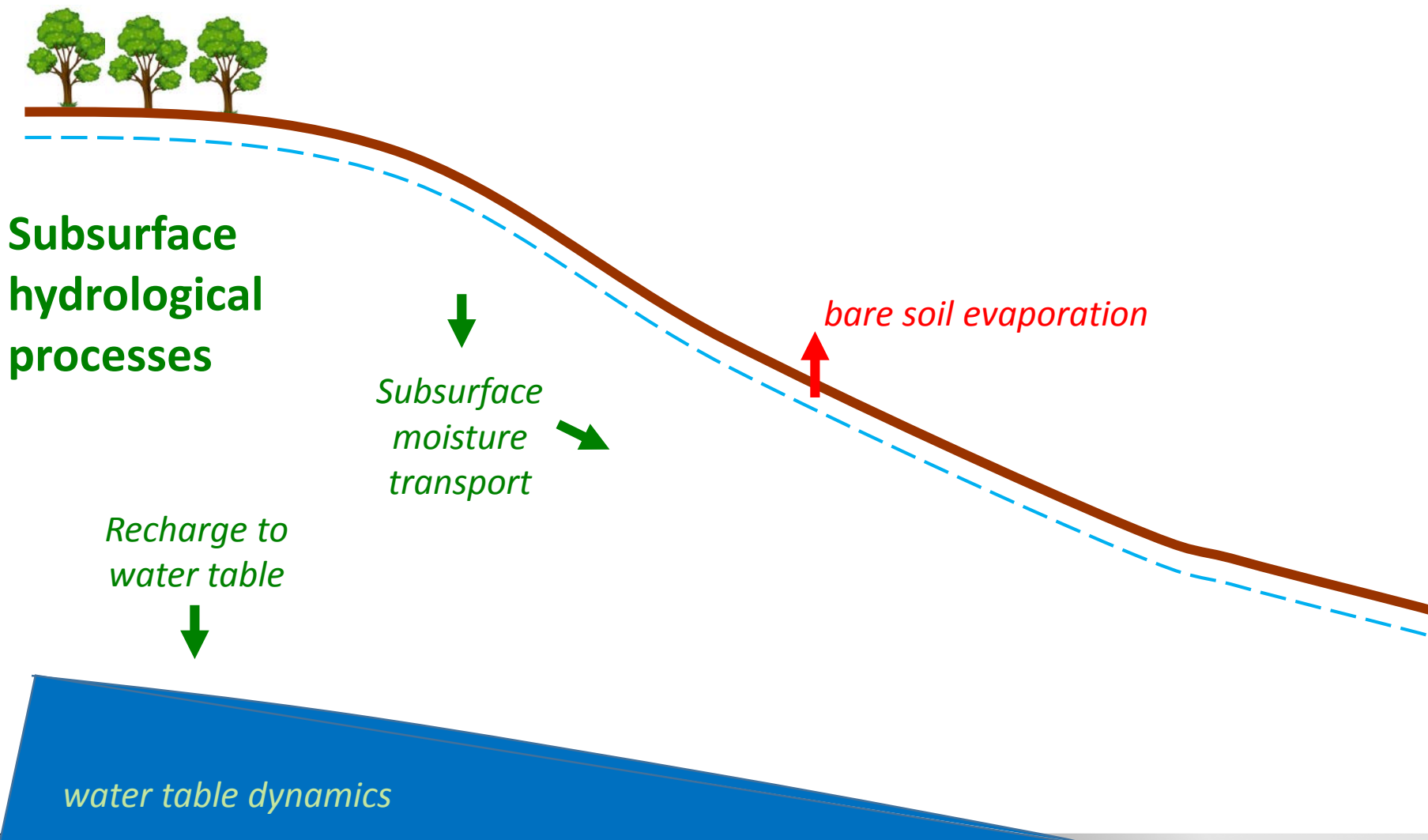
Three topics:

1. “Loss functions”: the characterization of soil moisture dynamics.
2. Data assimilation and calibration: do they access the same information?
3. Estimating rainfall and streamflow with SMAP data.

Paper under review in Water
Resources Research







Subsurface
hydrological
processes

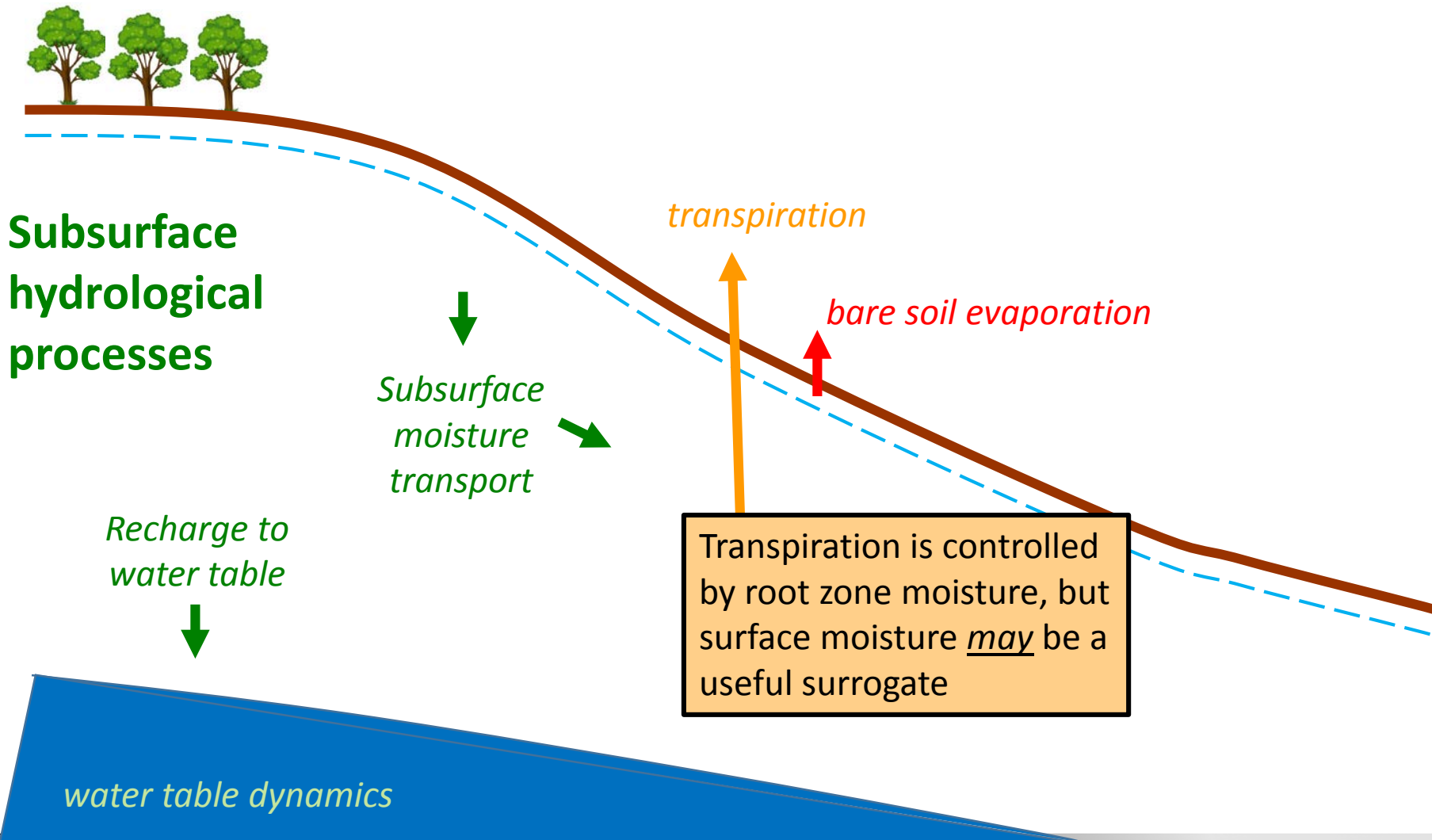
bare soil evaporation

*Subsurface
moisture
transport*

*Recharge to
water table*

water table dynamics

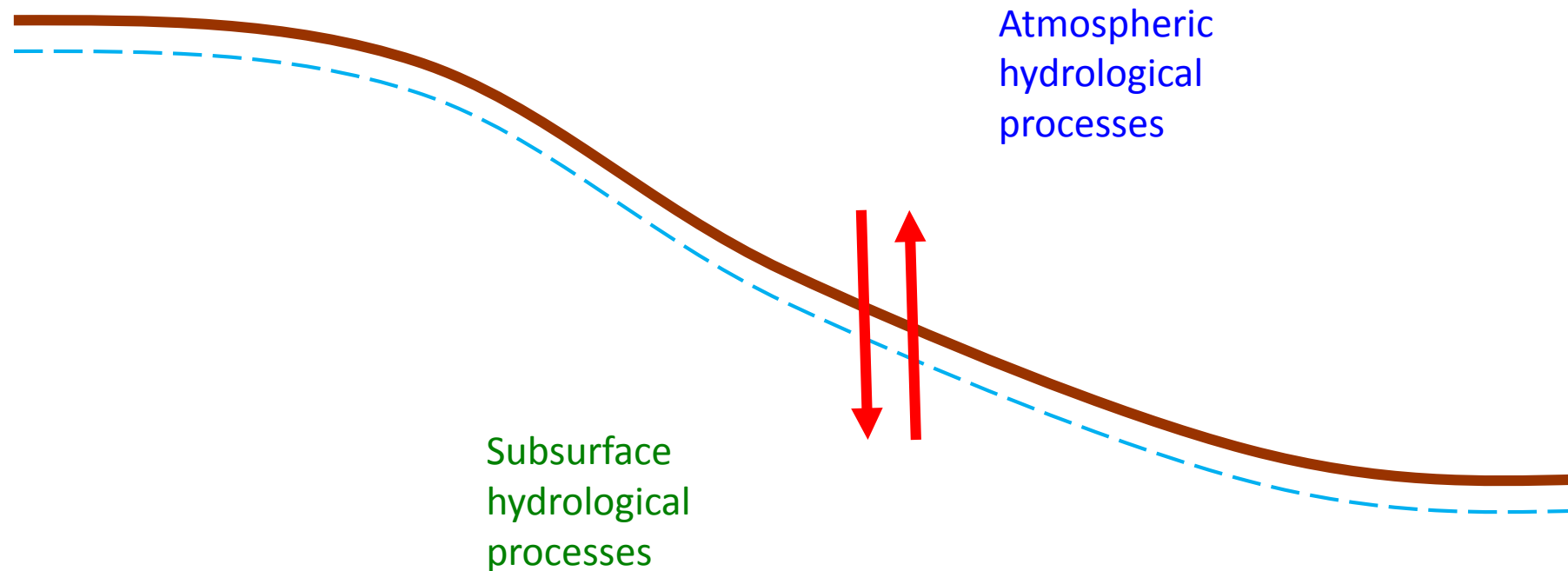
**5 cm layer: a
critical “gateway”
monitored by
SMAP**



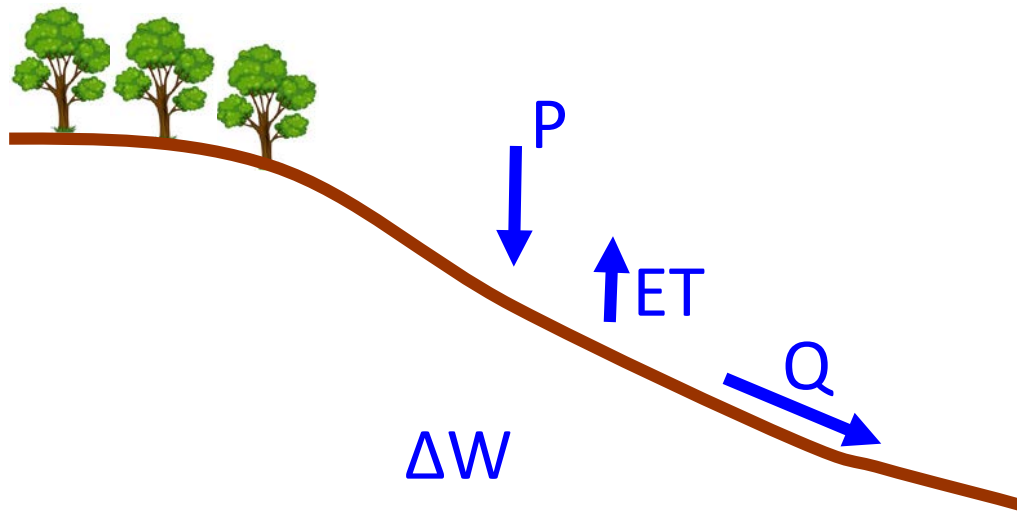
5 cm layer: a critical “gateway” monitored by SMAP

Key point: Because SMAP data monitor the “critical gateway”, they contain information on many important hydrological processes.

⇒ *In essence, SMAP does more than just measure soil moisture...*



Sample Hydrological Application: Estimating terms in a basin's water balance.



$$P = ET + Q + \Delta W$$

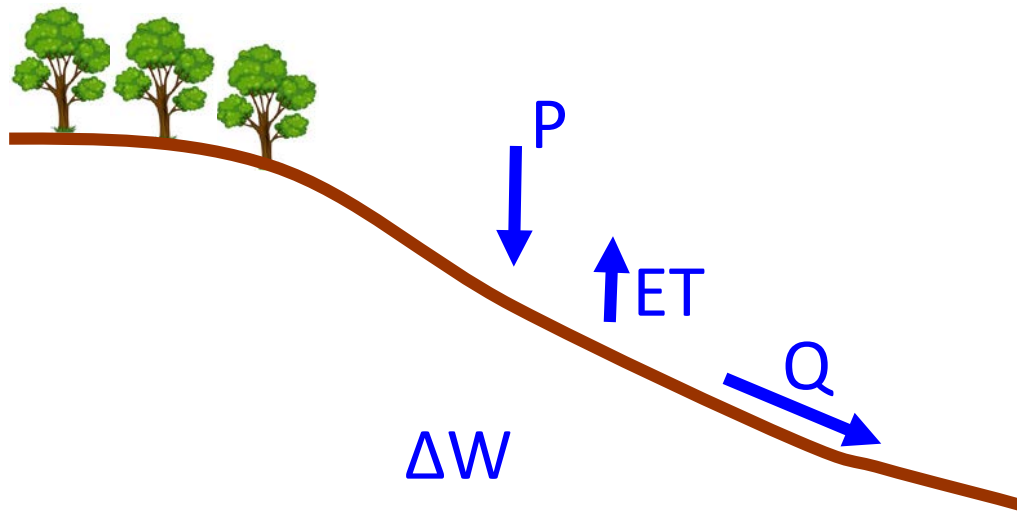
precipitation

runoff (streamflow)

evapotranspiration

change in storage

Sample Hydrological Application: Estimating terms in a basin's water balance.

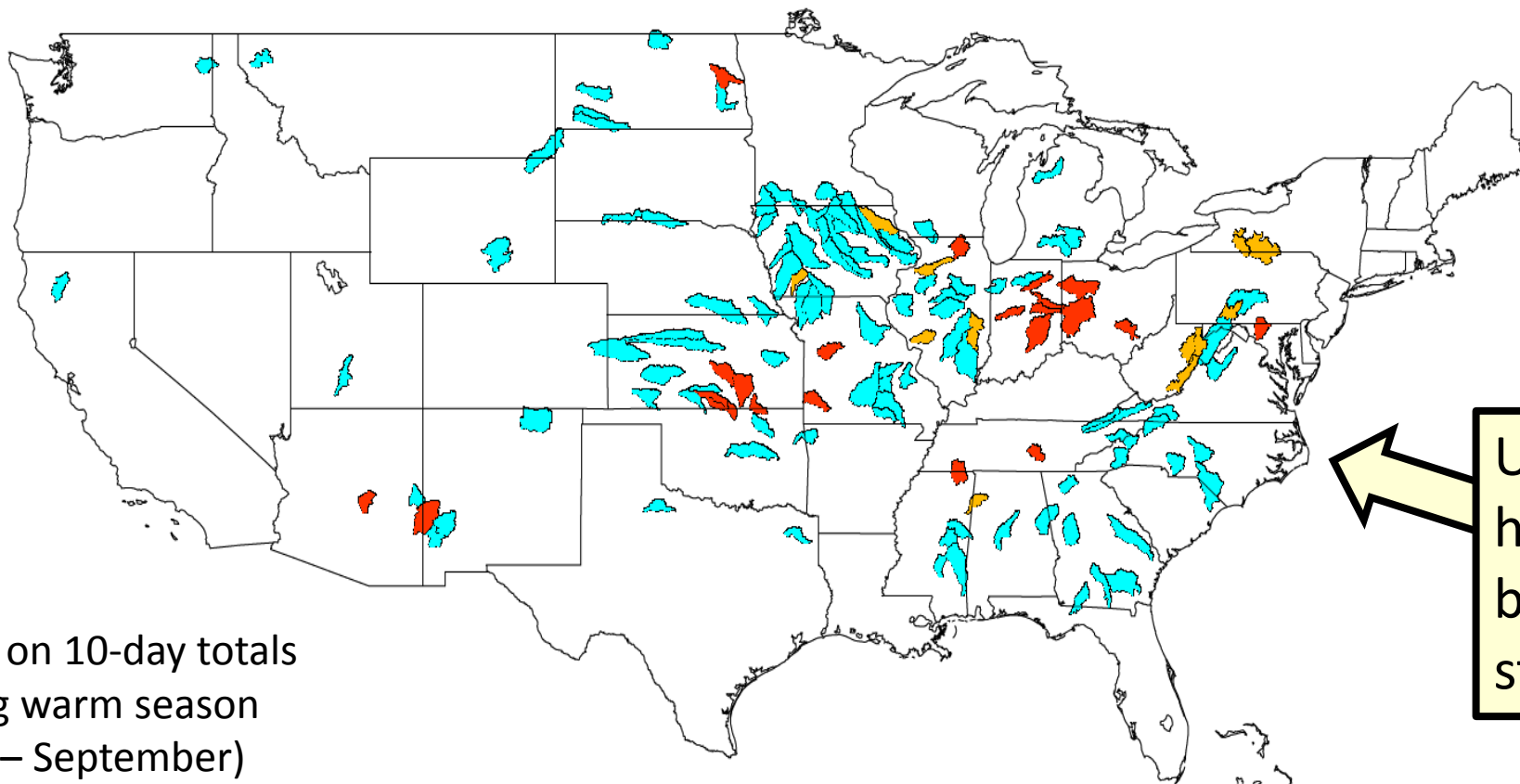


$$\text{P} = \text{ET} + \text{Q} + \Delta W$$

Today: a demonstration of how these terms can be estimated with SMAP soil moisture data
(not perfectly, but with some skill)

Basin level analysis (to allow for
joint calculation of P and Q)

$$\textcircled{P} = ET + \textcircled{Q} + \Delta W$$



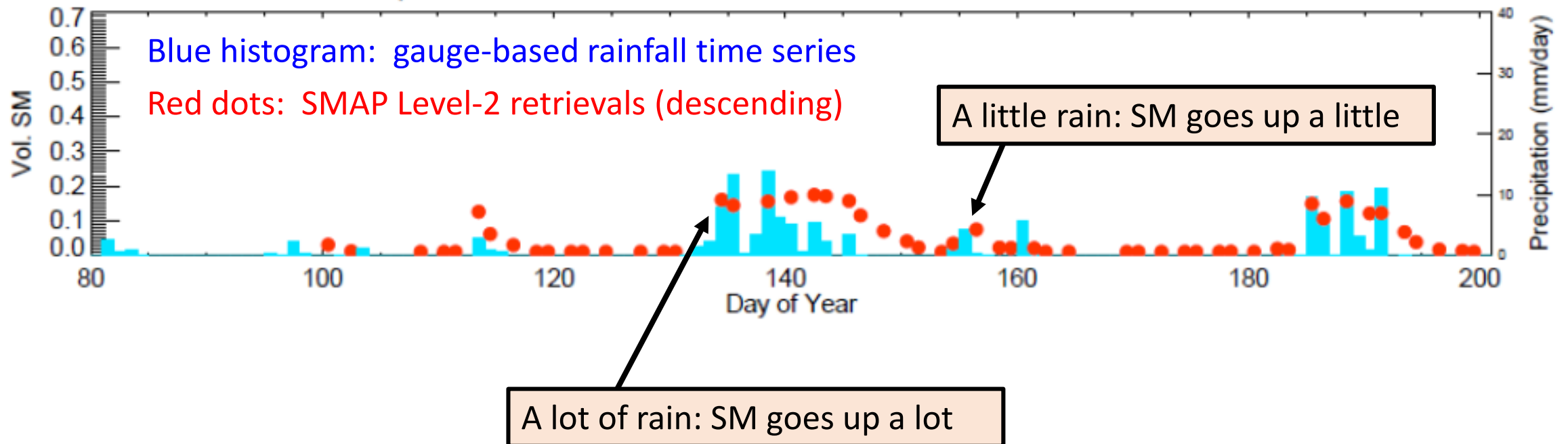
Focus on 10-day totals
during warm season
(June – September)

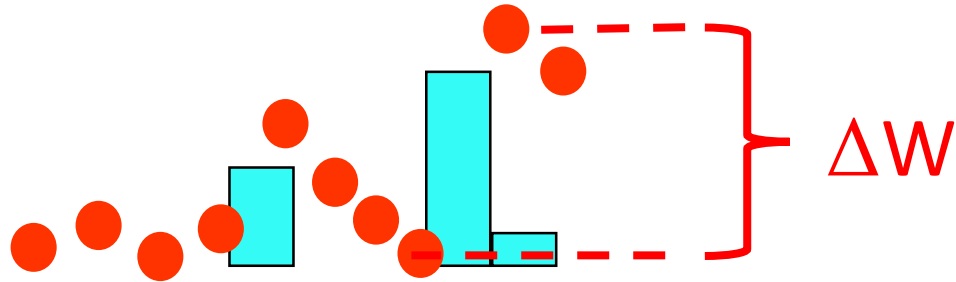
Unregulated
hydrological
basins with USGS
stream gauges

Step 1: Precipitation Estimation

Notice: Rain gauge data and SMAP radiometer data generally look nicely consistent.

a. Location A: 119.3W, 41.80N (Western U.S.)

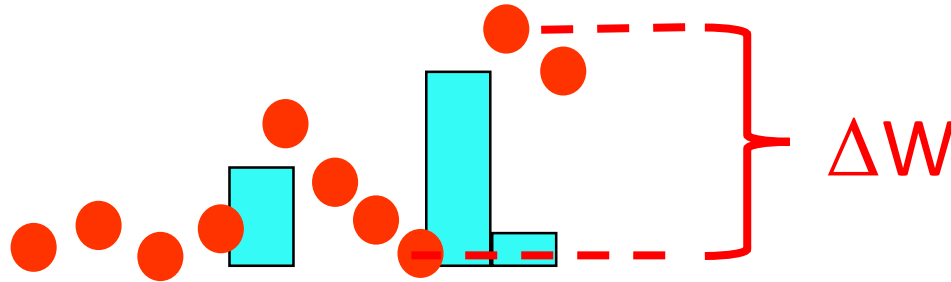




Precipitation estimation approach from Brocca et al. (2013):

$$P \sim \text{Max} (\Delta W - a W_{\text{ave}}^b , 0.)$$

W_{ave} = average of the two consecutive retrievals



Updated approach:
This term replaced with
calibrated loss function

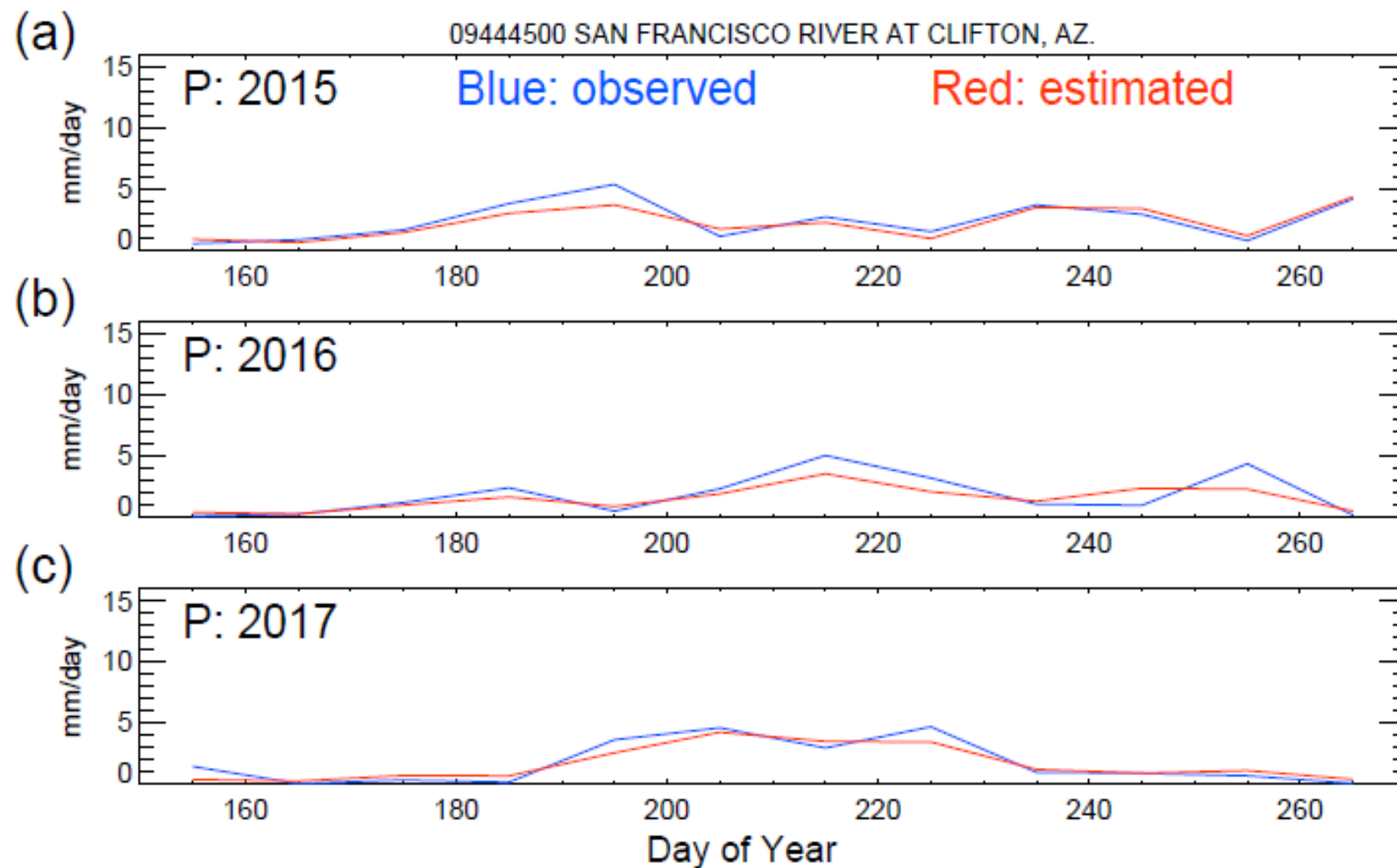
Precipitation estimation approach from Brocca et al. (2013):

$$P \sim \text{Max} (\Delta W - a W_{\text{ave}}^b, 0.)$$

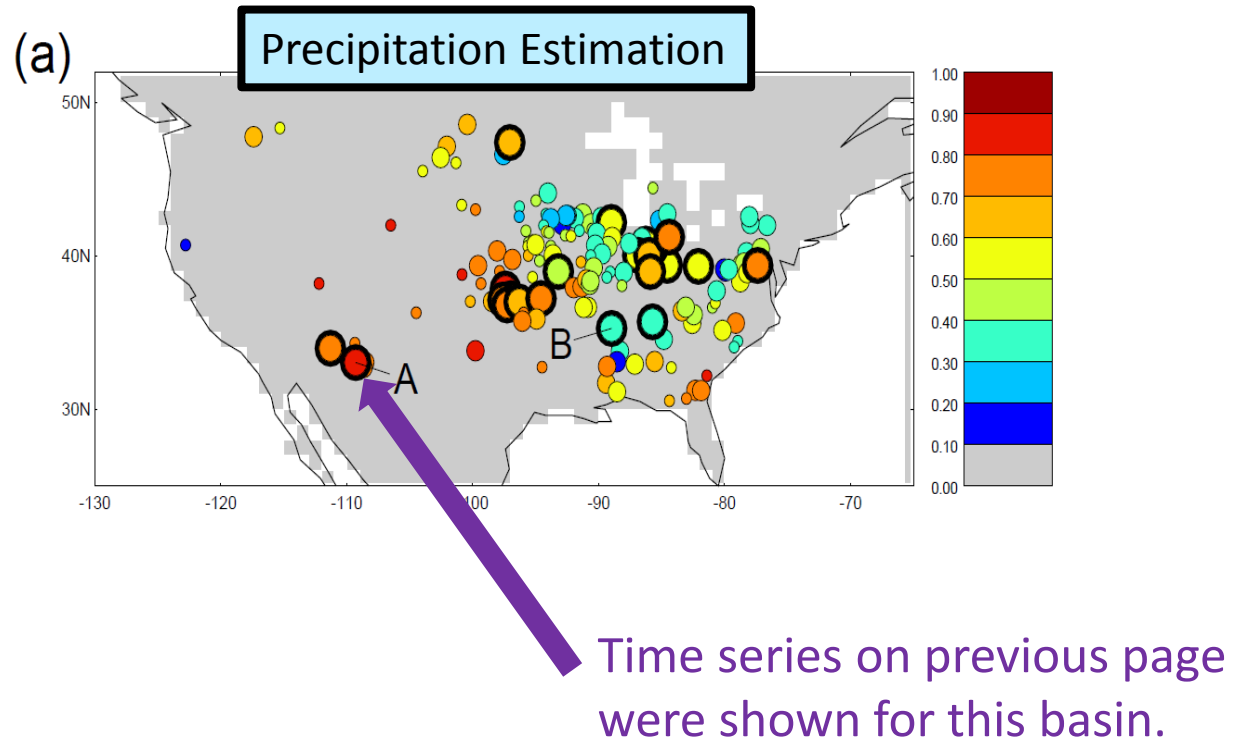
W_{ave} = average of the two consecutive retrievals

Some results! (One of the better estimations):

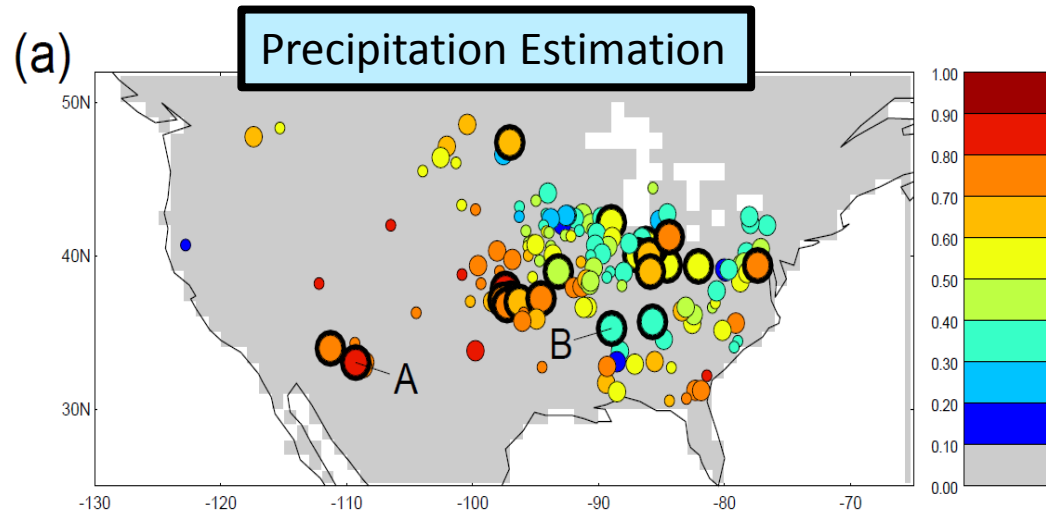
We can characterize the agreement in these time series with the square of the correlation coefficient, r^2 .



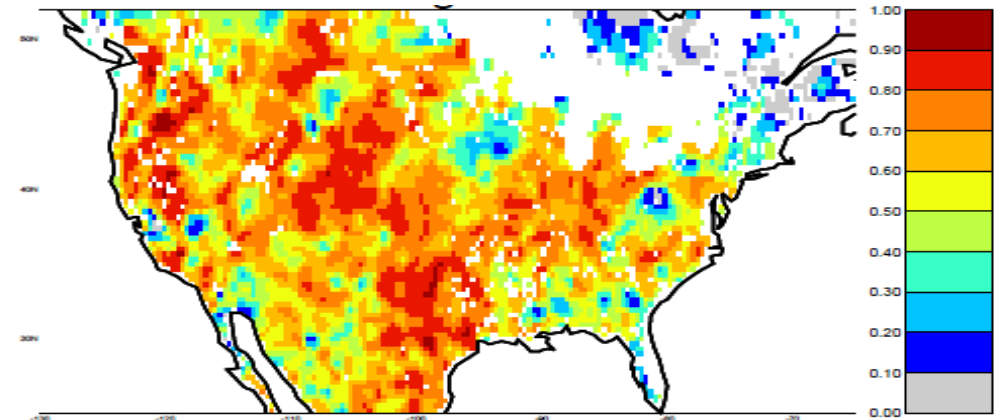
Basin level skill scores (time series of 10-day precipitation totals: r^2 vs observations)



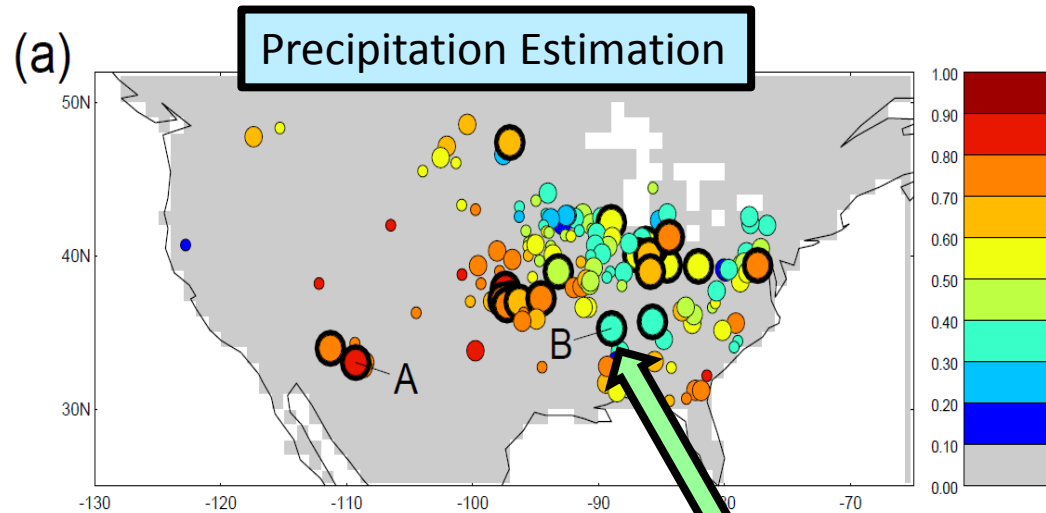
Basin level skill scores (time series of 10-day precipitation totals: r^2 vs observations)



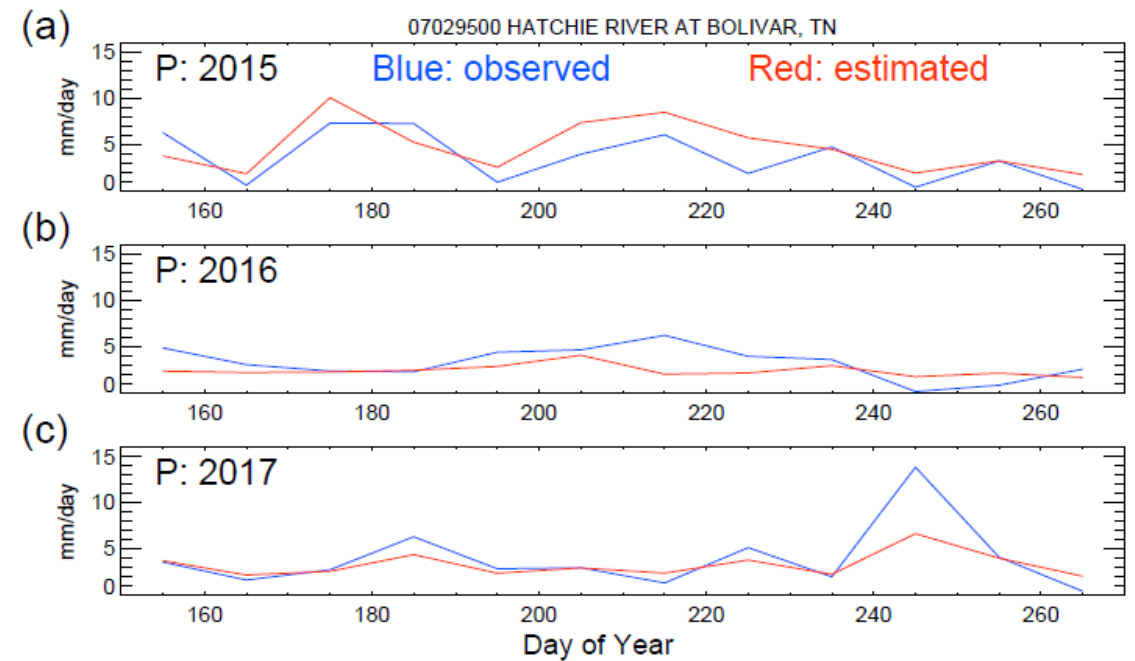
Aside: extending this analysis across the US, beyond “basins”, indicates high skill throughout the west.



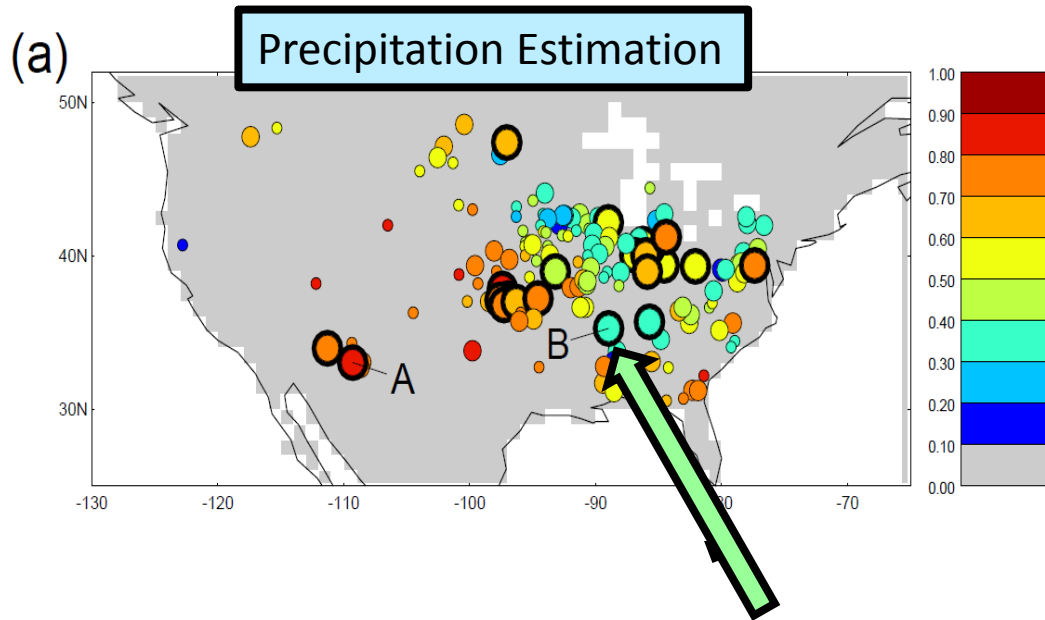
Basin level skill scores (time series of 10-day precipitation totals: r^2 vs observations)



The algorithm works relatively poorly in this basin. Still, there is some valid information in the estimates

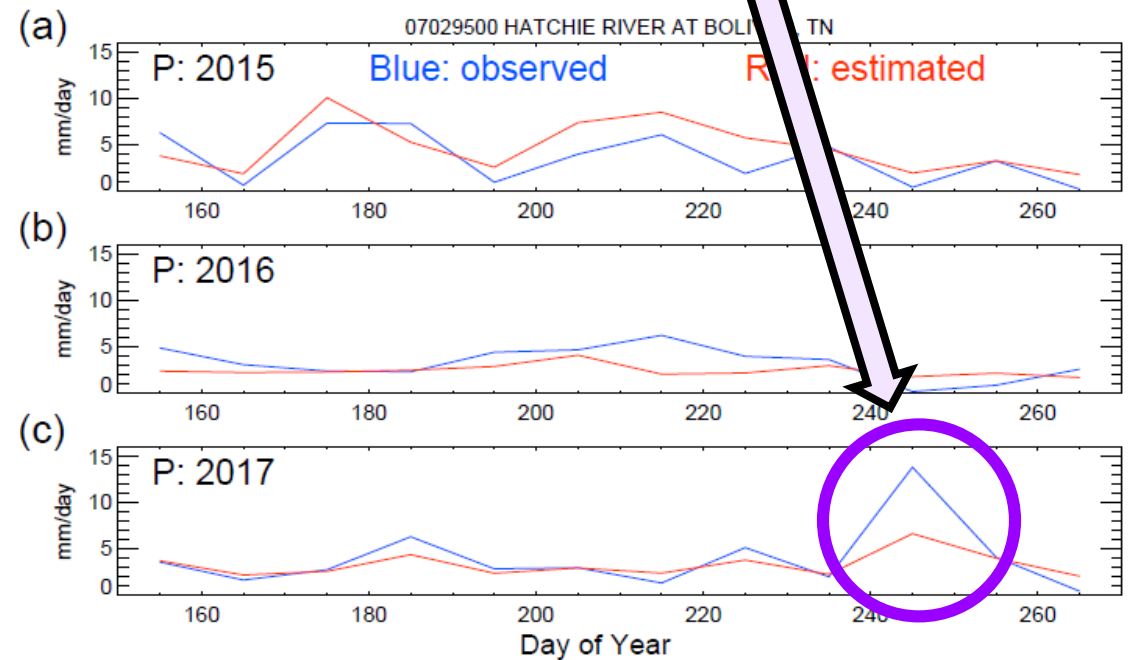


Basin level skill scores (time series of 10-day precipitation totals: r^2 vs observations)



The algorithm works relatively poorly in this basin. Still, there is some valid information in the estimates

Pardon the pun:
High rainfall saturates the soil moisture signal, making rainfall estimation difficult.



Step 2: Streamflow Estimation

Consider:

- The fraction of rainfall, P , that is converted to surface runoff, Q_{fast} , increases with surface soil moisture, W_{surf} :

$$Q_{\text{fast}} / P = f_1 (W_{\text{surf}})$$

- Drainage of moisture, Q_{slow} , to the water table (and eventually into streams) increases with increased soil moisture, W :

$$Q_{\text{slow}} = f_2 (W)$$

Step 2: Streamflow Estimation

Apply multiple regression:

$$Q = Q_{\text{fast}} + Q_{\text{slow}} = a P W + b W + c$$

The diagram shows the equation $Q = Q_{\text{fast}} + Q_{\text{slow}} = a P W + b W + c$. A red arrow points from the text 'basin-scale P' to the variable 'P' in the equation. Two green arrows point from the text 'basin-scale W' to the variable 'W' in the equation.

Note – in practice, more complex and accurate approaches would be used. This simple approach has the advantage, though, of demonstrating unequivocally that SMAP data hold relevant information.

Cross-validate!

Calibrate (i.e., find a, b, and c) using two years of observations:

obs Q from 2 SMAP years

obs P from 2 SMAP years

SMAP W from 2 SMAP years

$$Q = a P W + b W + c$$

Validate against data from third year:

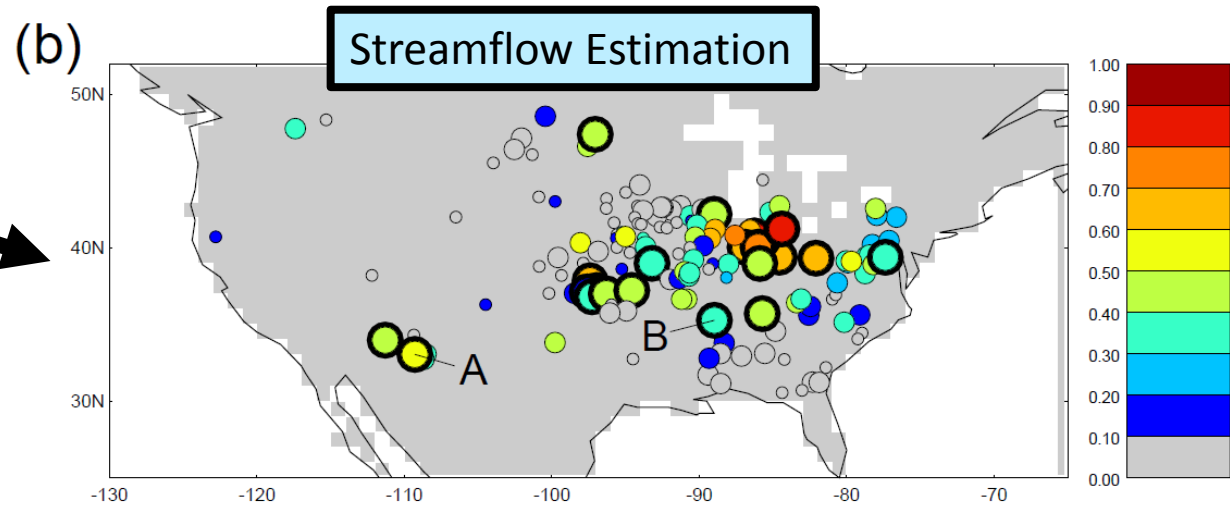
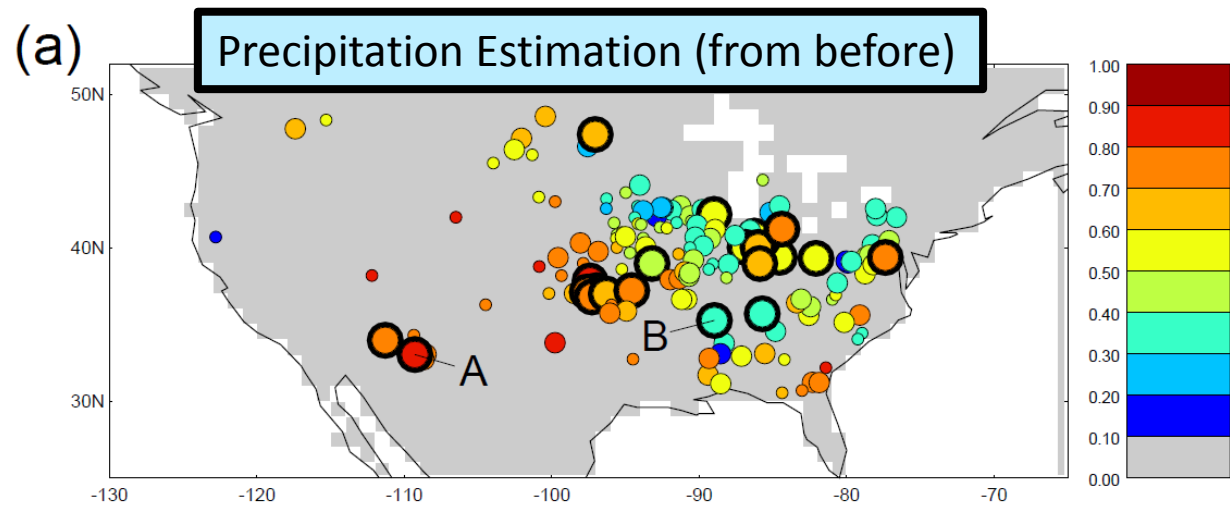
SMAP-based P estimate for 3rd year

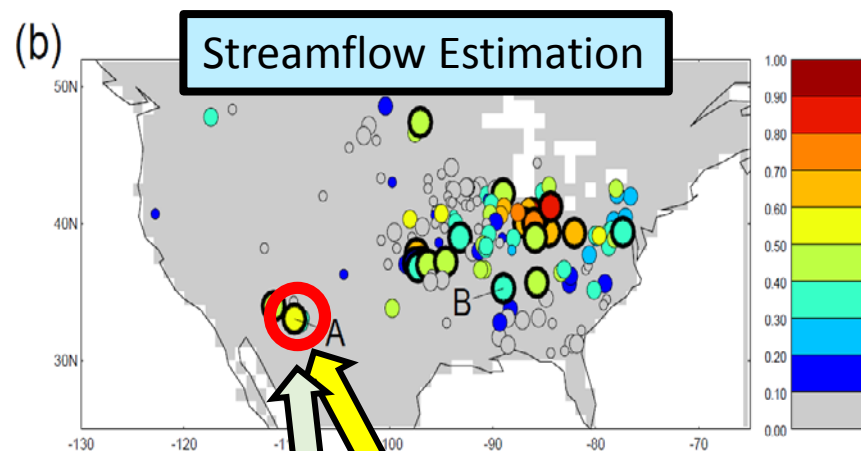
SMAP W in 3rd year

$$Q = a P W + b W + c$$

estimate of Q in 3rd year derived solely from SMAP data

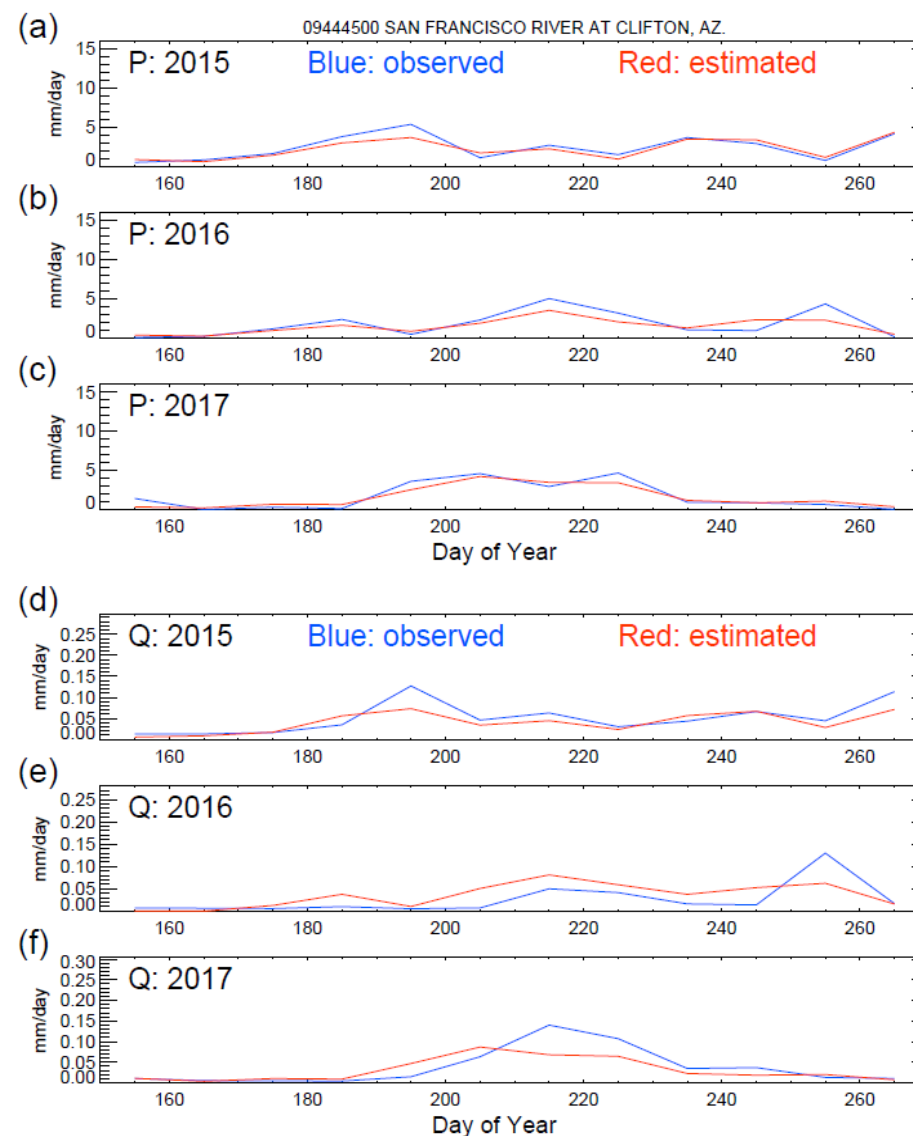
**Estimation skill
(time series of 10-day
streamflow totals:
 r^2 vs observations)**

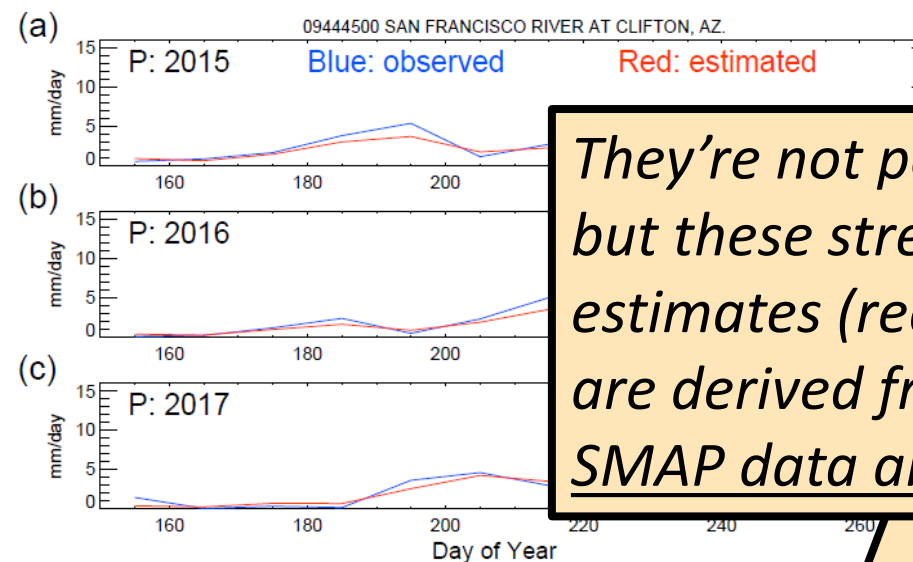
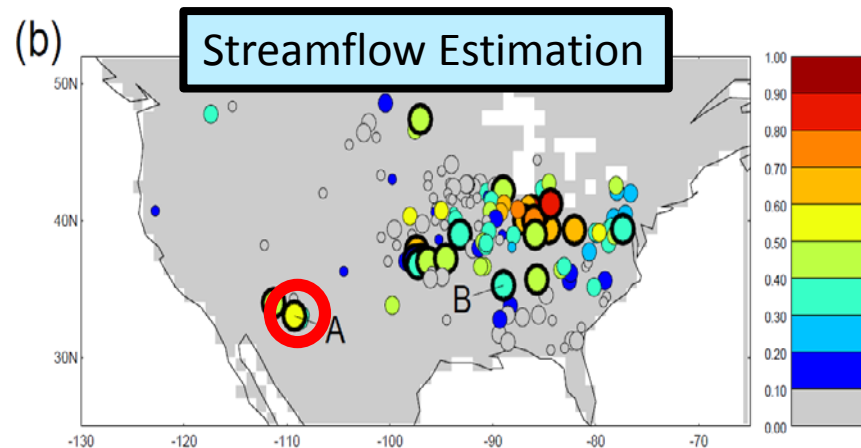




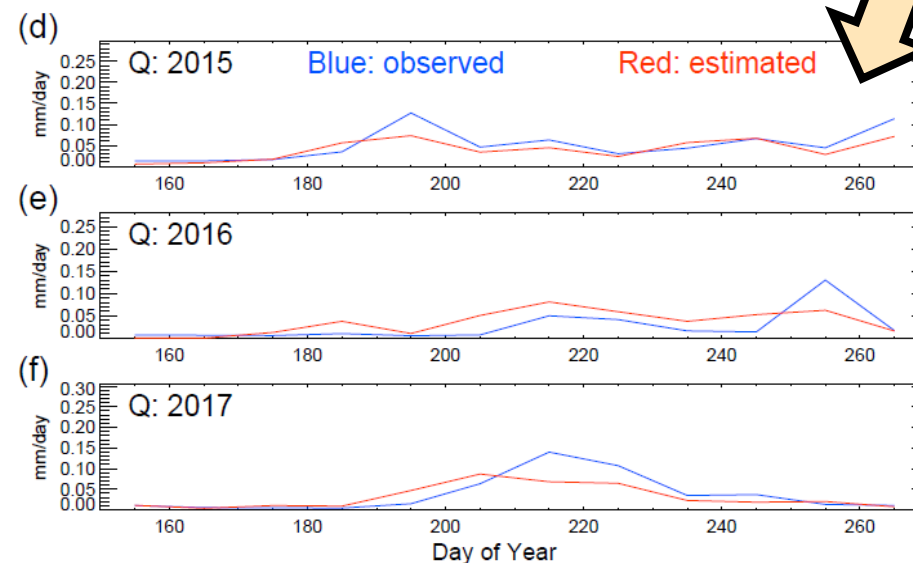
Precipitation estimates
(from before)

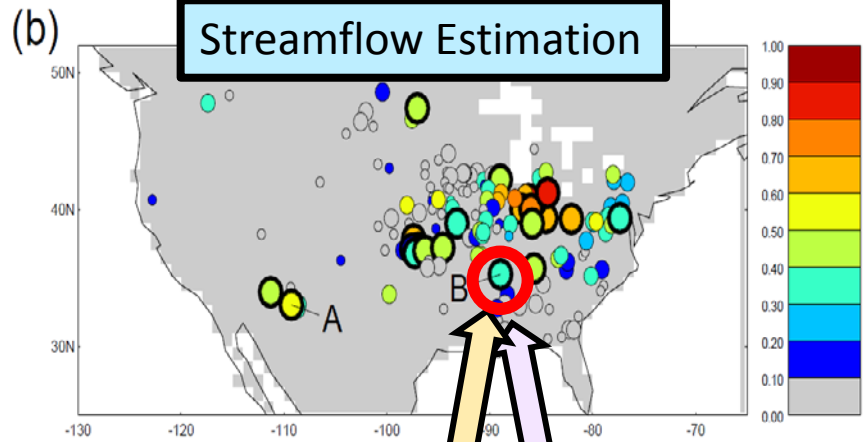
Streamflow estimates
capture some of the
observed behavior.





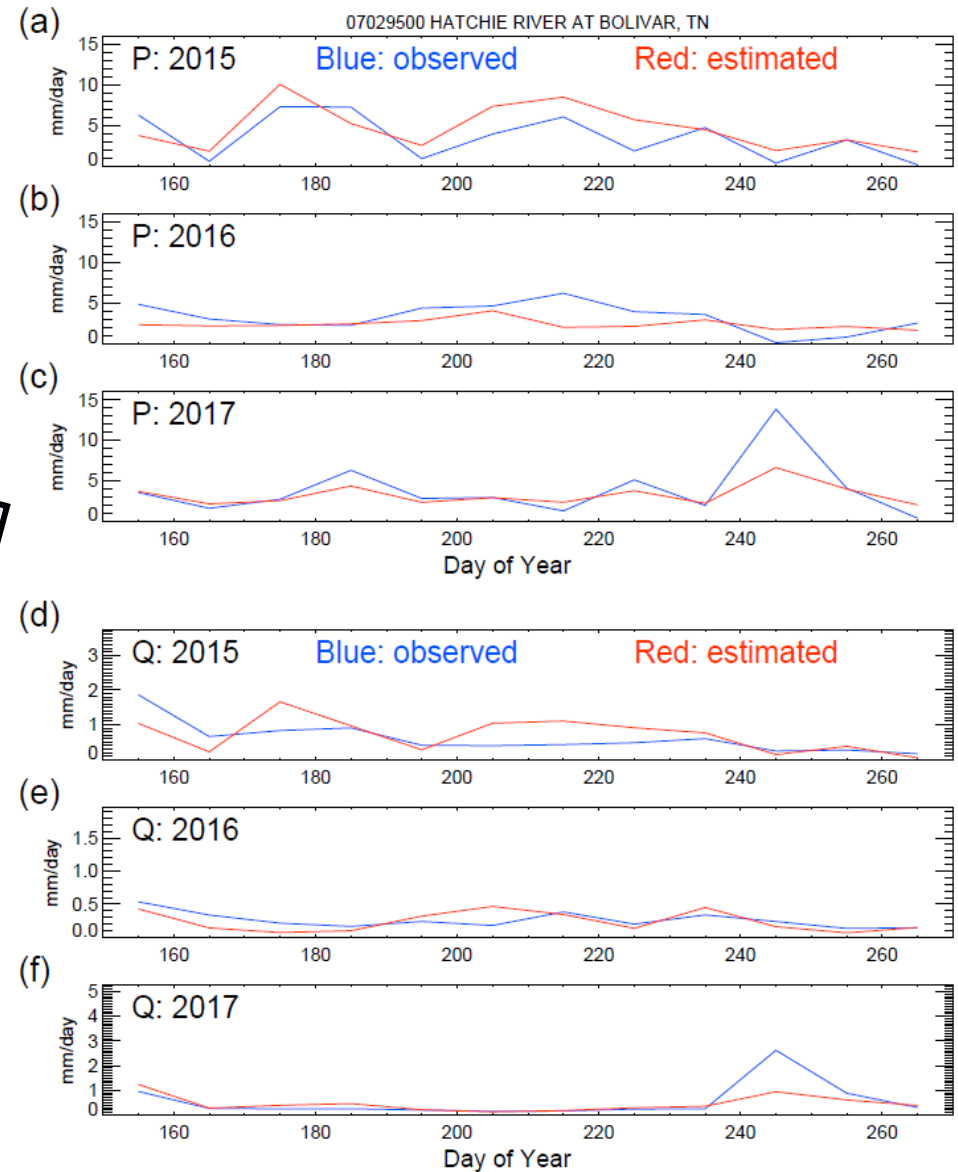
They're not perfect, but these streamflow estimates (red lines) are derived from SMAP data alone!

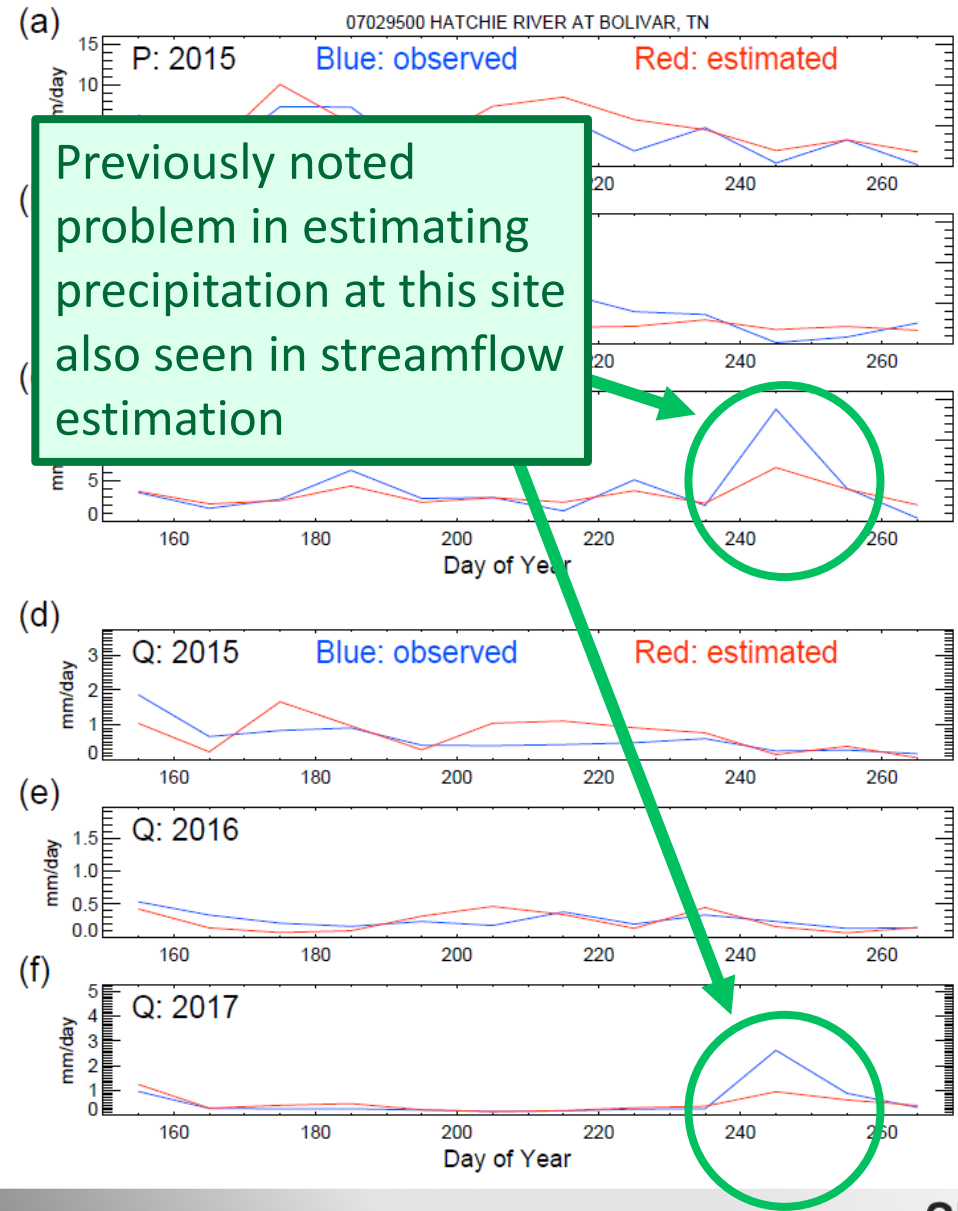
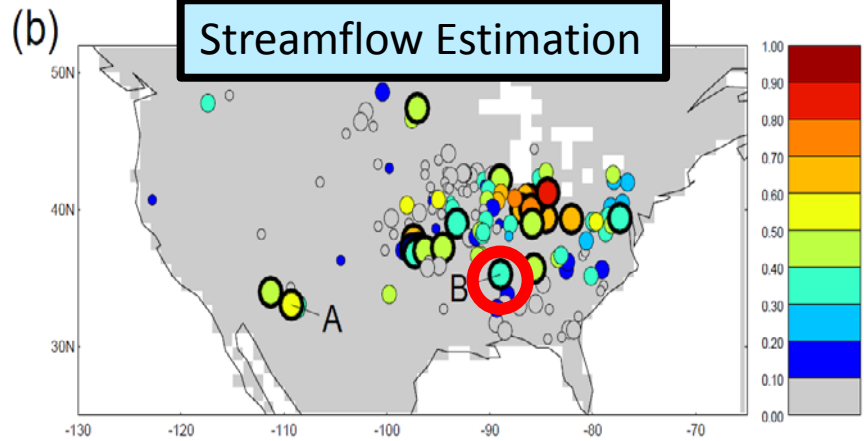




Precipitation estimates in
“poorer basin” (from before)

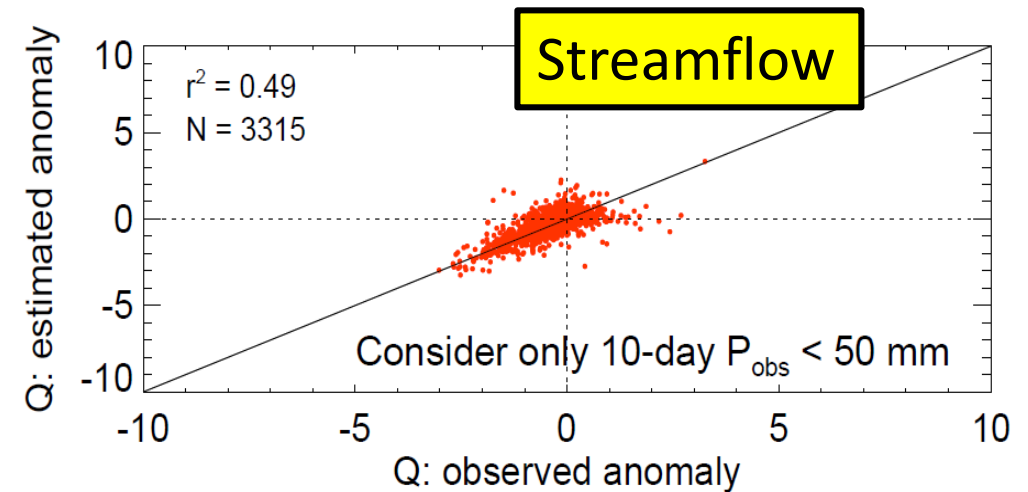
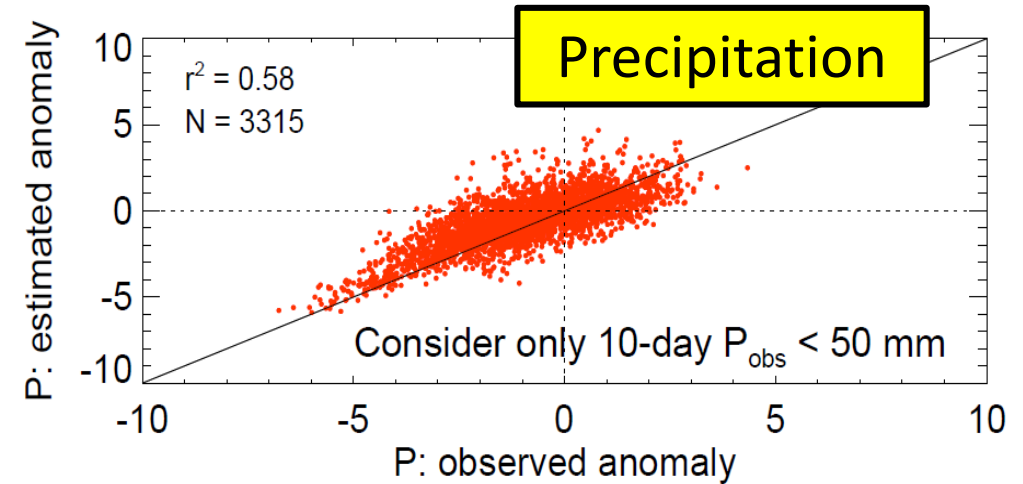
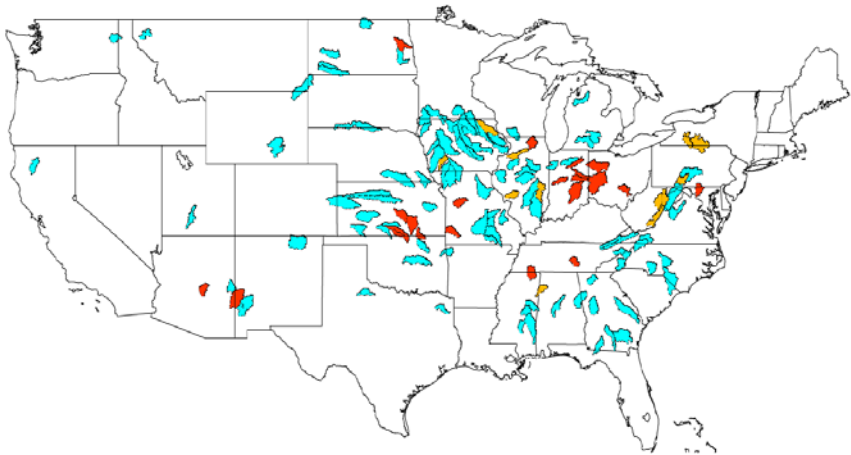
Streamflow estimates
capture less of the
observed behavior.





Inter-basin analysis

(throw results for all considered basins onto the same scatterplot)

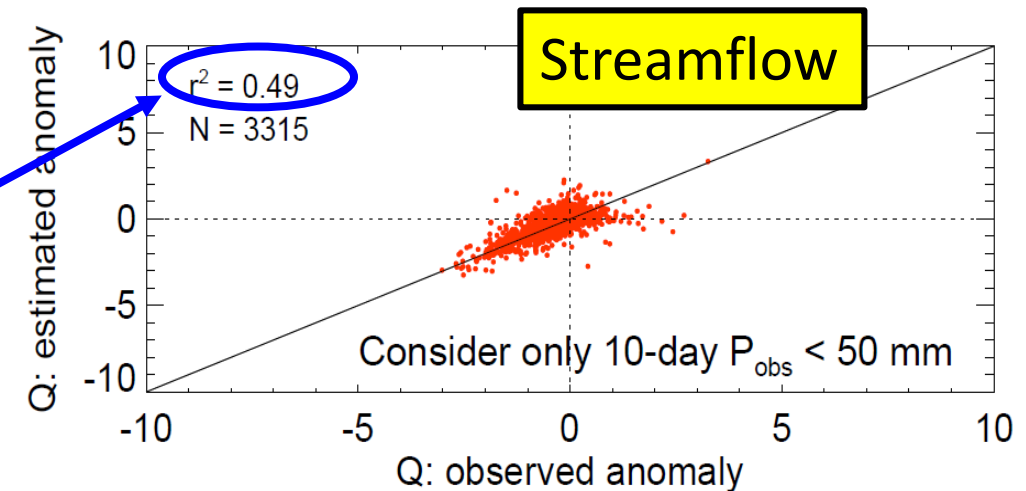
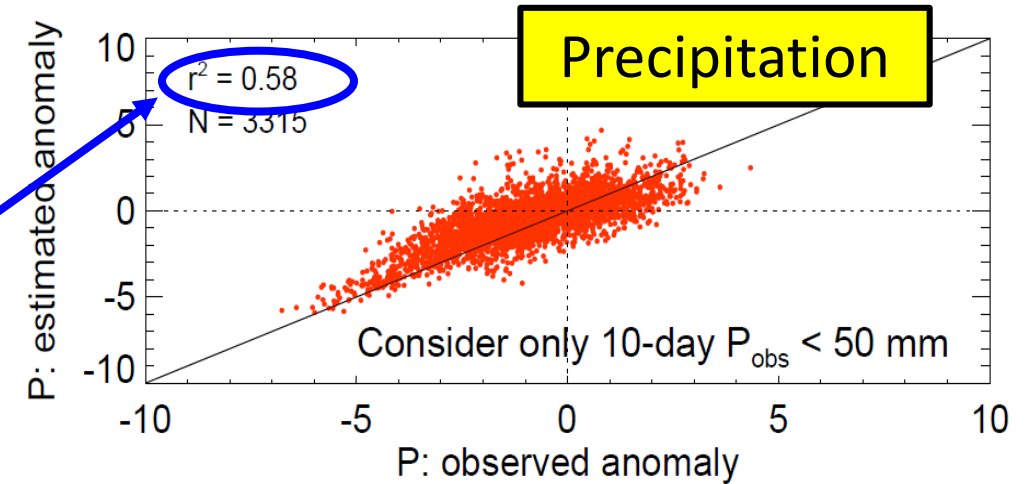


Inter-basin analysis

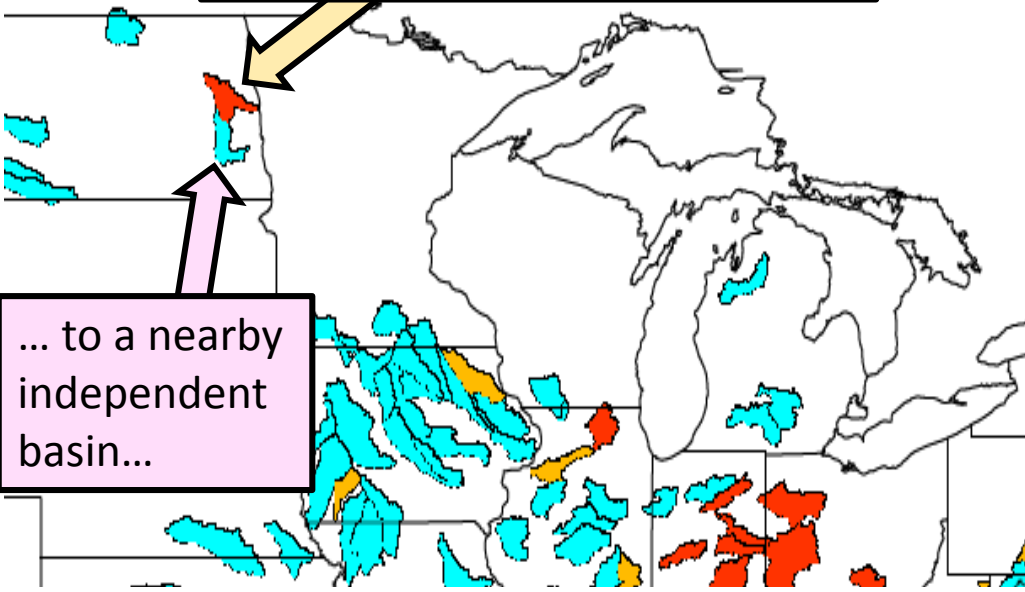
(throw results for all considered basins onto the same scatterplot)

For 10-day periods in which rainfall does not exceed 50 mm), SMAP data “explain” 58% of precipitation variance...

... and 49% of the streamflow variance



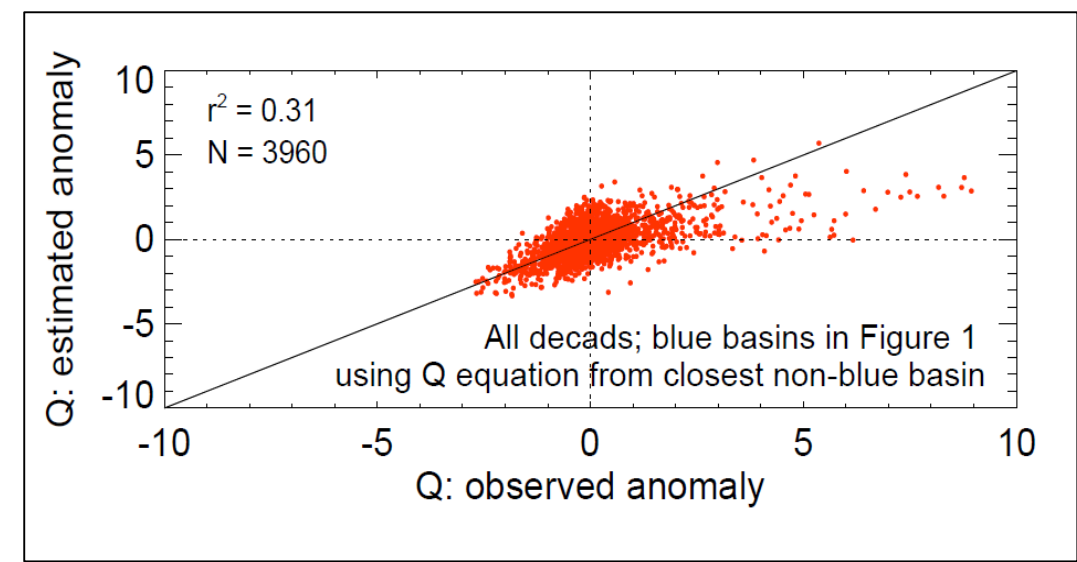
Applying streamflow equations
calibrated in one basin...



... to a nearby
independent
basin...

Transferability (streamflow)

... produces streamflow estimates with some skill
⇒ *there's hope for estimating streamflow in
basins that never had a streamflow gauge.*



Main Finding:

The SMAP estimates of rainfall and streamflow are not perfect, but they do contain relevant information.

⇒ At the very least, they should prove useful for constraining, or otherwise contributing to, rainfall and streamflow estimates obtained with more conventional approaches.

A final comment regarding this last study...

Obvious question: What is the potential for examining other basin water budget components?

$$\overset{\checkmark}{P} - \overset{?}{E} - \overset{\checkmark}{Q} - \overset{?}{\Delta\text{storage}} = 0$$

We know that evapotranspiration is a strong function of soil moisture
⇒ SMAP data could, in theory, be used to estimate it.

SMAP actually measures directly some of the storage change.

⇒ *The potential for estimating the other components as well is indeed there.*

To summarize this entire seminar:

1. “Loss functions”: the characterization of soil moisture dynamics.
⇒ derived loss functions can extend SMAP measurements to real time and into the future, which is valuable for applications.
2. Data assimilation and calibration: do they access the same information?
⇒ these approaches access complementary information; SMAP data indeed contain a wealth of hydrological info.
3. Estimating rainfall and streamflow with SMAP data.
⇒ SMAP data do contain information on large-scale water balance quantities!