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

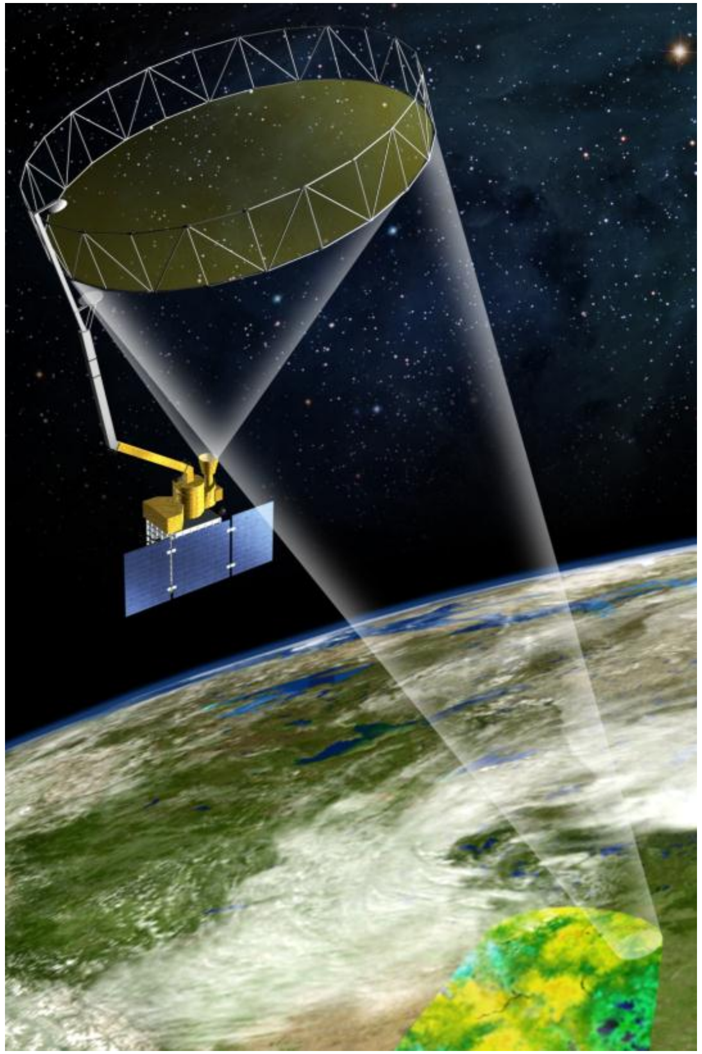
*with substantial help from Wade Crow, Rolf Reichle, and Sarith Mahanama

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/

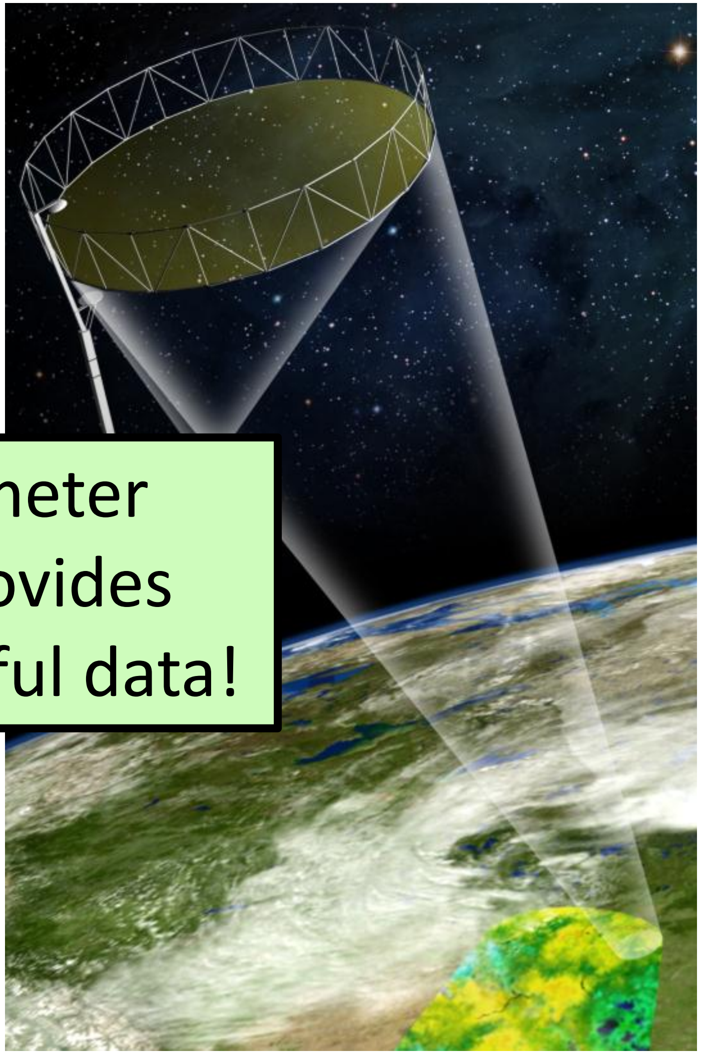
SMAP provides a capability for global mapping of soil moisture with unprecedented accuracy, resolution, and coverage.

Original specs:

Spatial resolution ~~3-km~~ ~~9-km~~ and 36-km
Temporal resoluti : ev ry 3 days (at leas)
How deep into so s eral cm (Level 1-3)
1 m (Level 4)
Accura umetric percent
Latenc s, depending on
Baseli n: 3 years
(just extended)
Orbit specs: Sun synchronous, 6 AM/PM,
8-day exact repeat with 2-3 day revisit

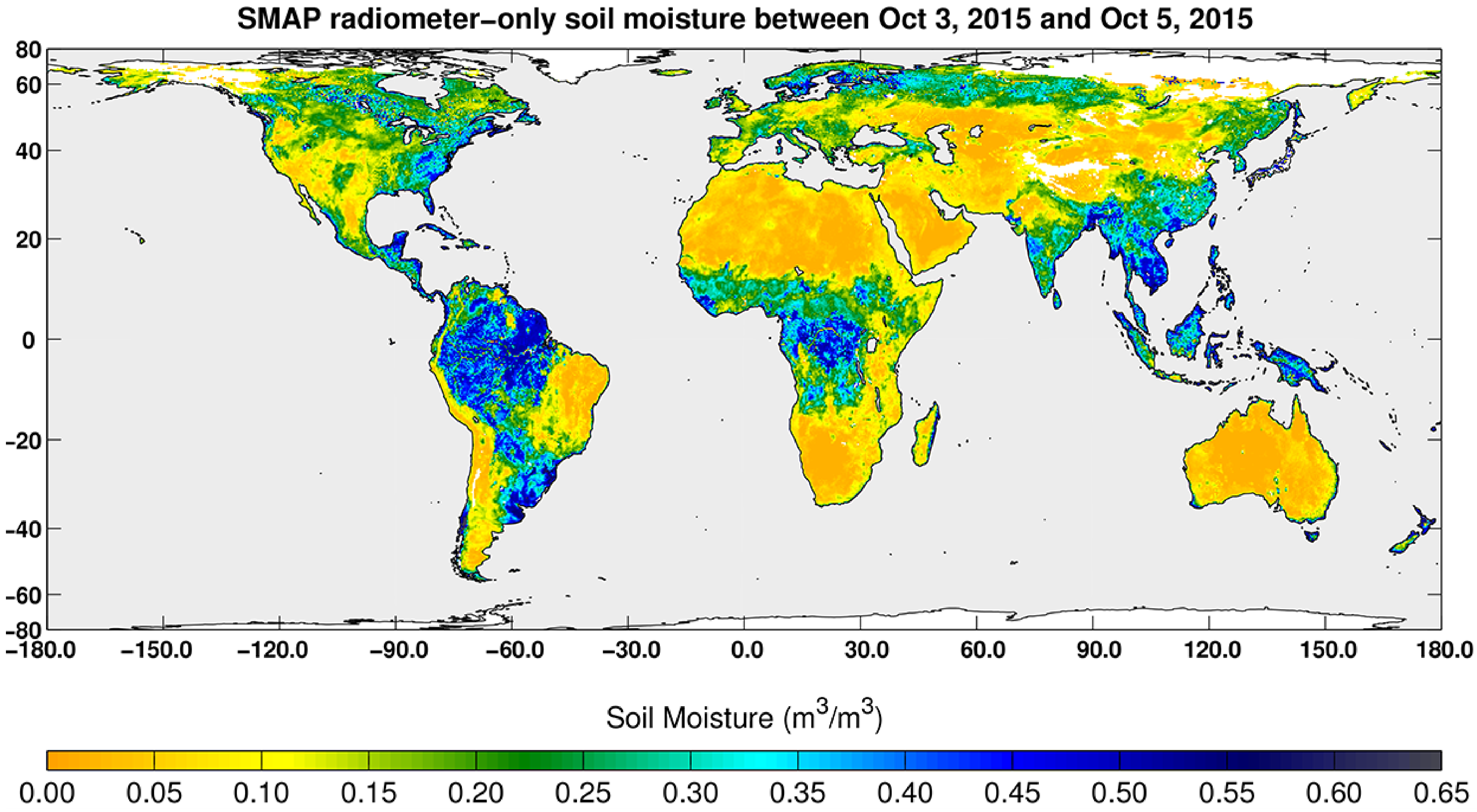
Radar instrument broke!

Radiometer still provides beautiful data!



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

Example:



SMAP radiometer–only soil moisture between Oct 3, 2015 and Oct 5, 2015
80
60
40
20
0
-20
-40
-60
-80
–180.0
–150.0
–120.0
–90.0
–60.0
–30.0
0.0
30.0
60.0
90.0
120.0
150.0
180.0
Soil Moisture (m3/m3)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65

So…. What can we do will 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.

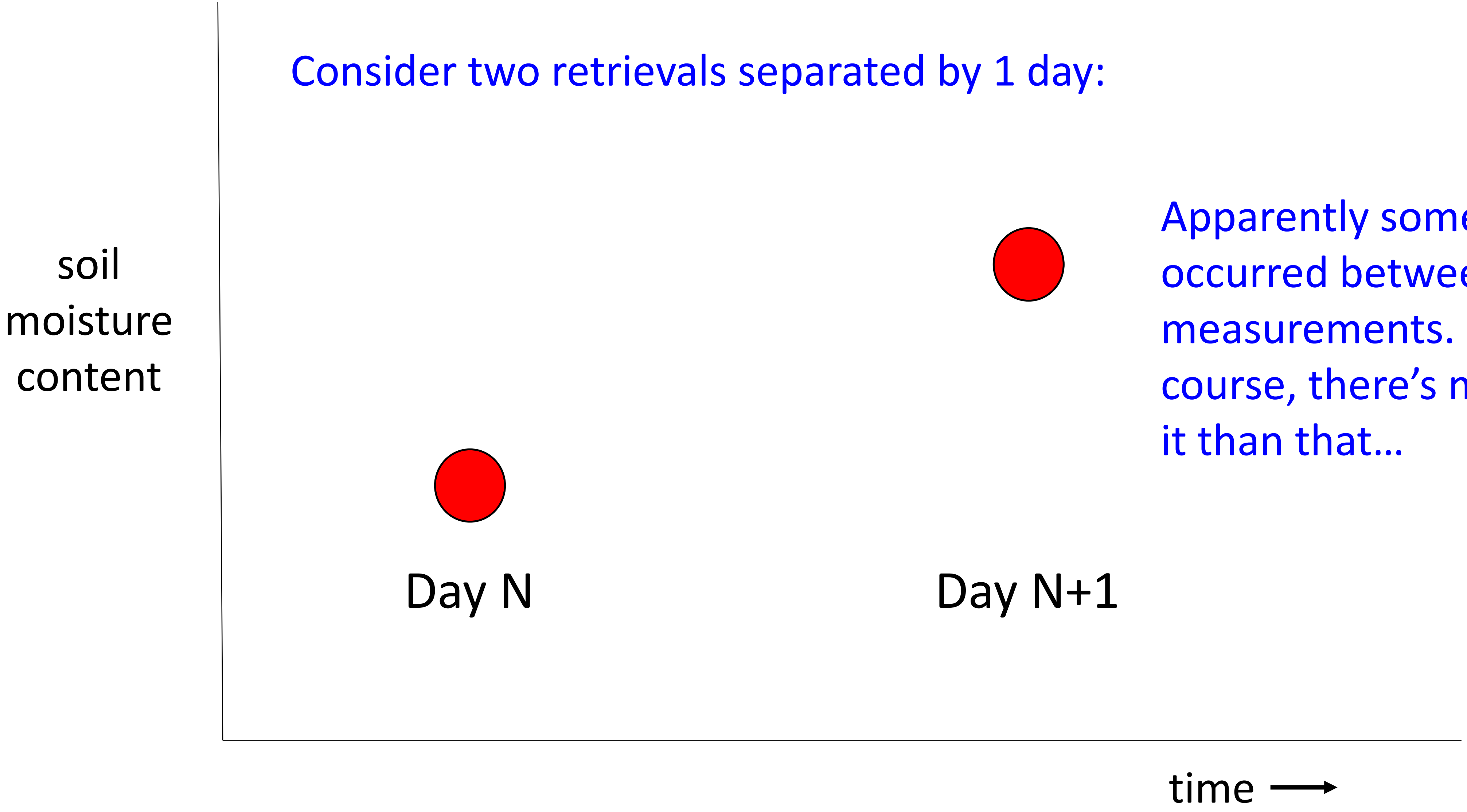
Published paper: Koster et al., J. Hydromet., 18, 837-843, 2017; DOI: 10.1175/JHM-D-16-0285.1

2. Data assim ... they access the same informatio

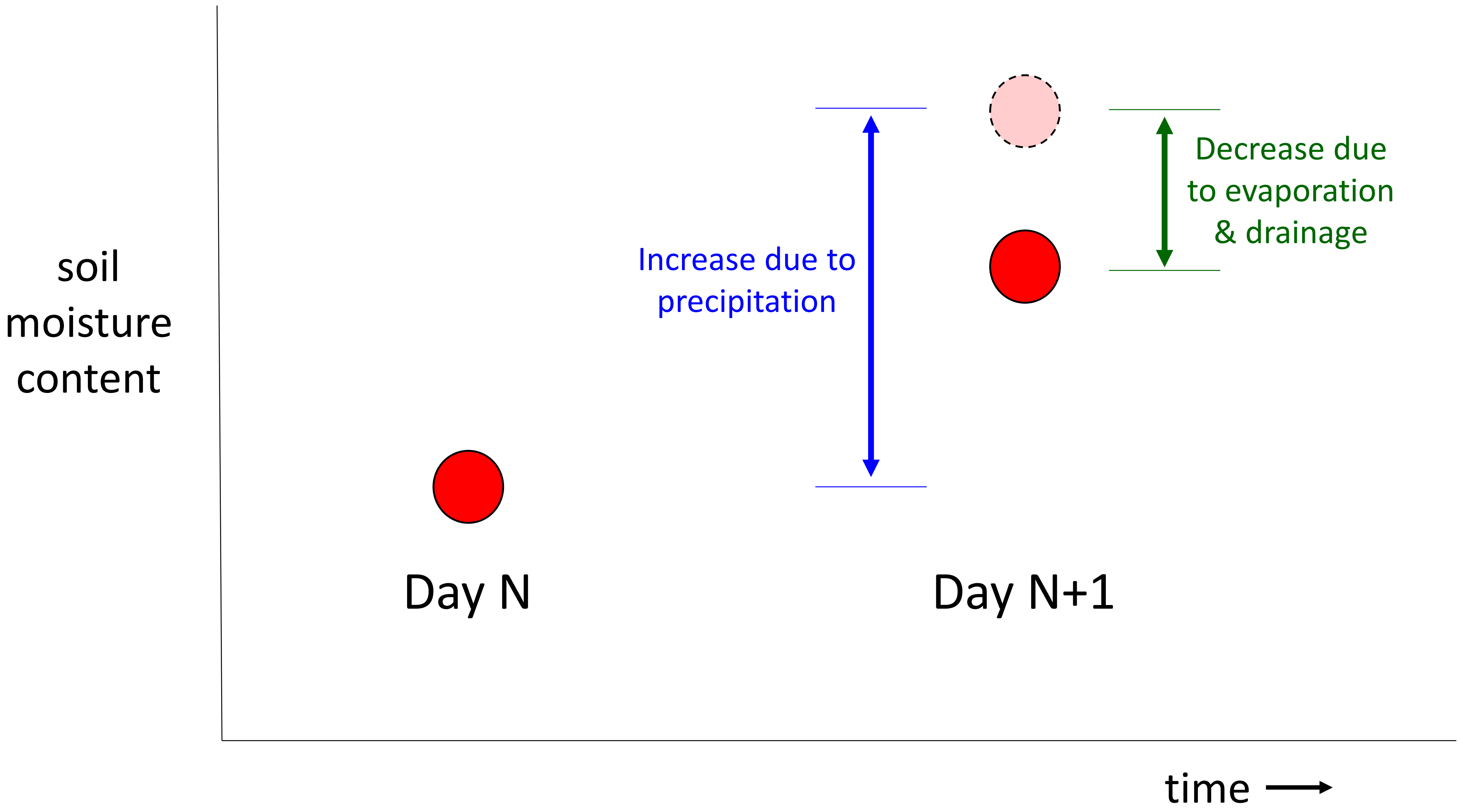
3. Estimating rainfall and streamflow with SMAP data.

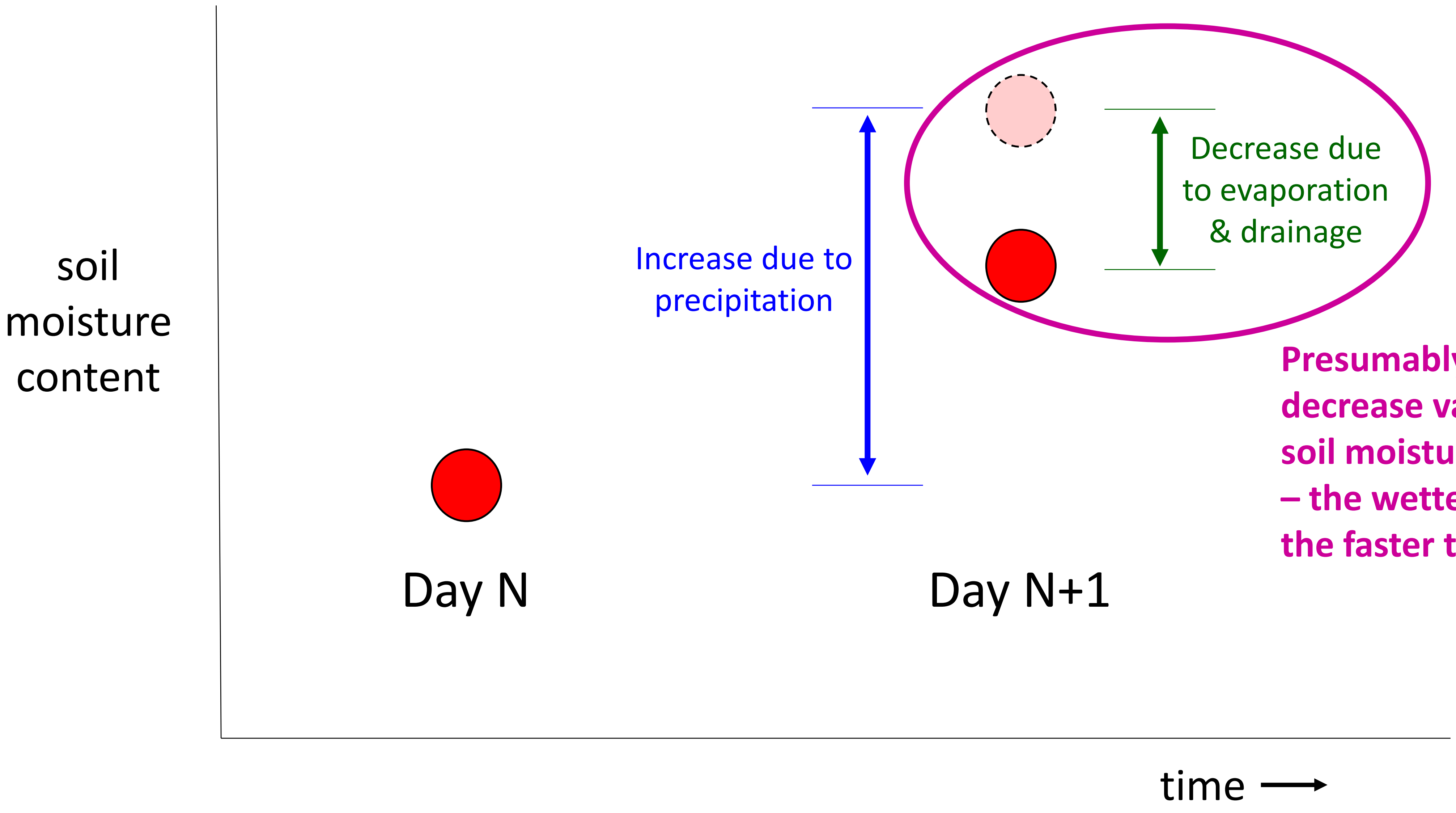
What is a loss function?

Consider two retrievals separated by 1 day:

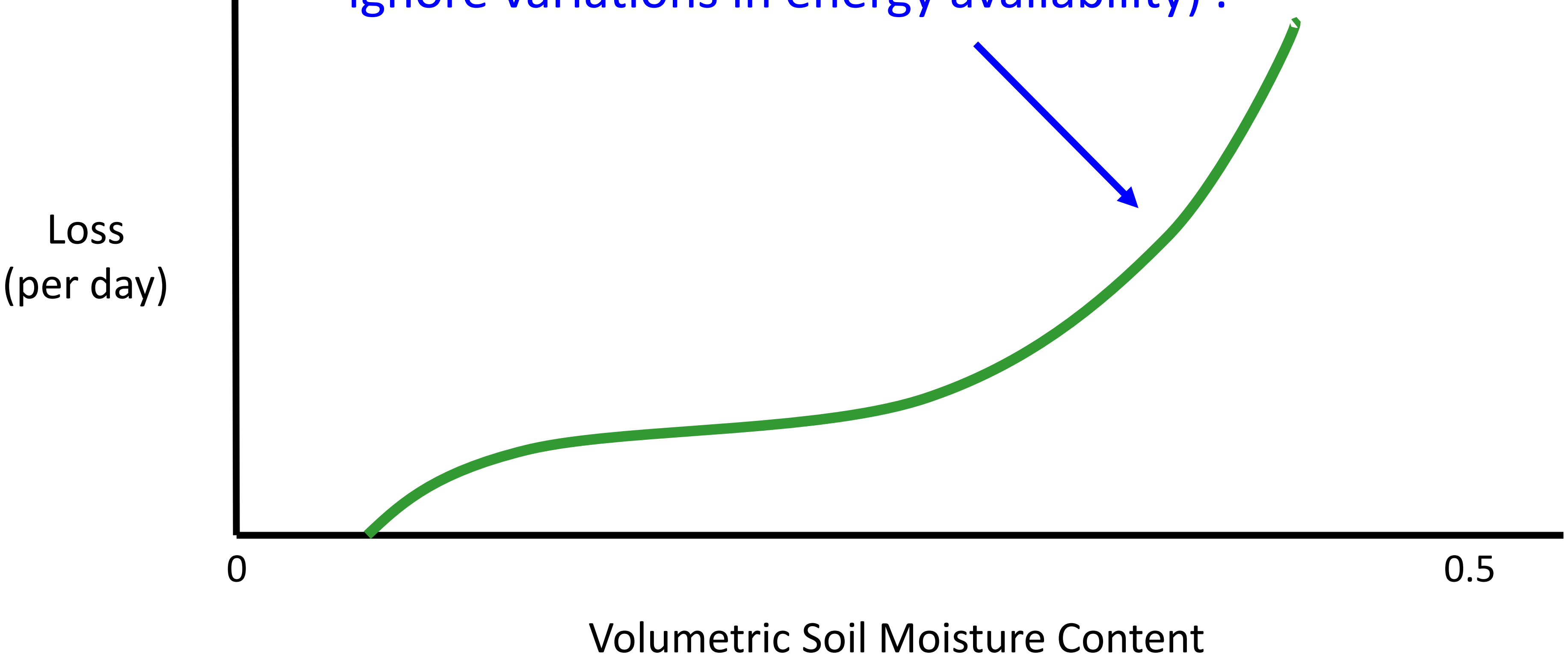


Apparently some rain occurred between the measurements. Of course, there’s more to it than that…

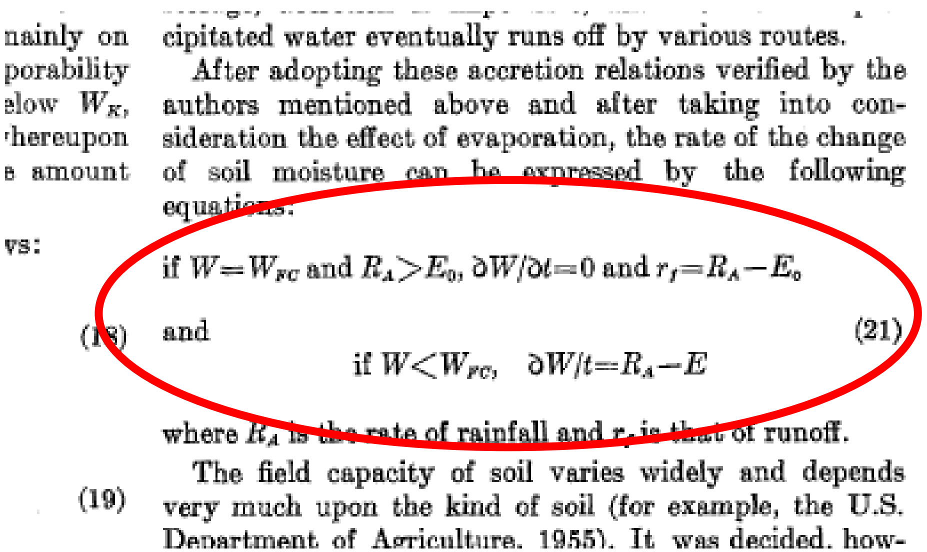

soil
moisture
content
Increase due to
precipitation
Decrease due
to evaporation
& drainage
Day N
Day N+1
time


soil moisture content
Increase due to precipitation
Decrease due to evaporation & drainage
Presumably this decrease varies with soil moisture content – the wetter the soil, the faster the loss.
Day N
Day N+1
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…



nainly on porability elow W_K, hereupon amount

vs:

(18)

(19)

cipitated 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:

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

and (21)

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

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

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-

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

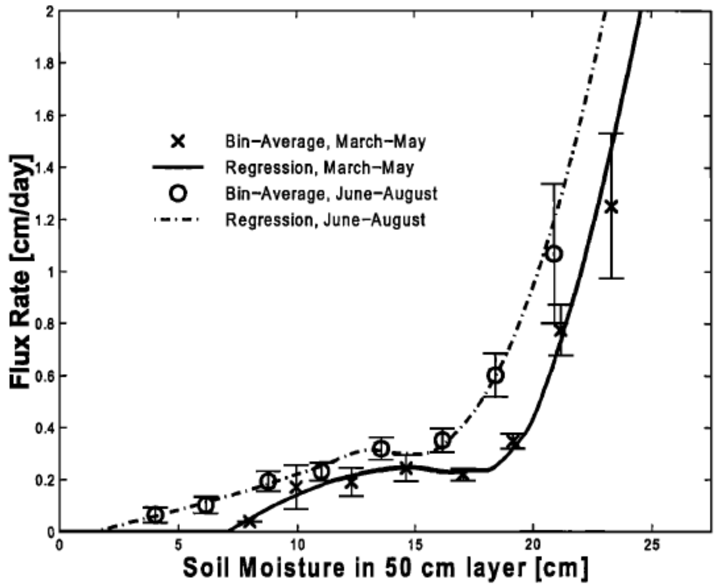


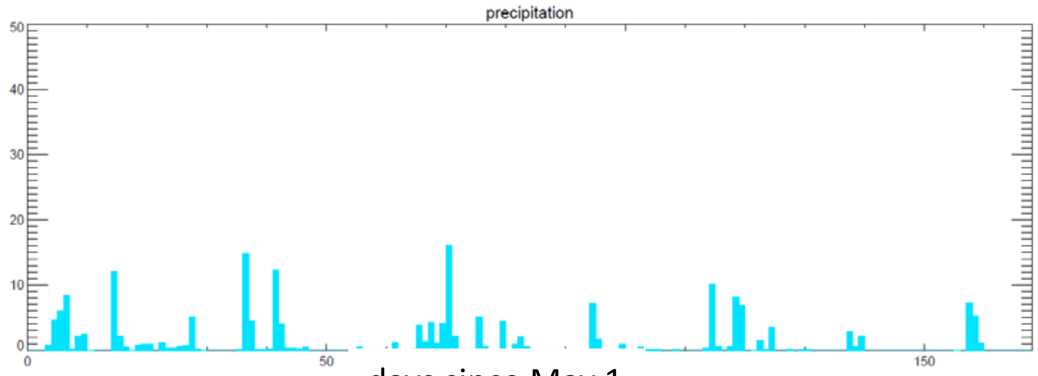
Figure 3. Estimation of moisture-dependent water loss from conditional mean precipitation for sites in Illinois. Note the plateau, presumably corresponding to potential evapotranspiration, and the sharp rise near saturation, presumably corresponding to percolation.

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

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

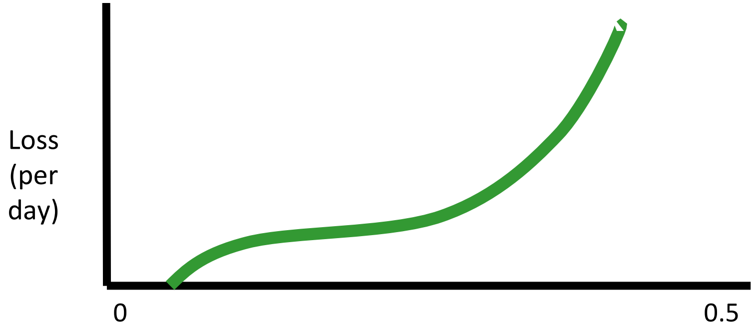
What is a simple way of determining a loss function from SMAP Level 2 soil moisture retrievals?

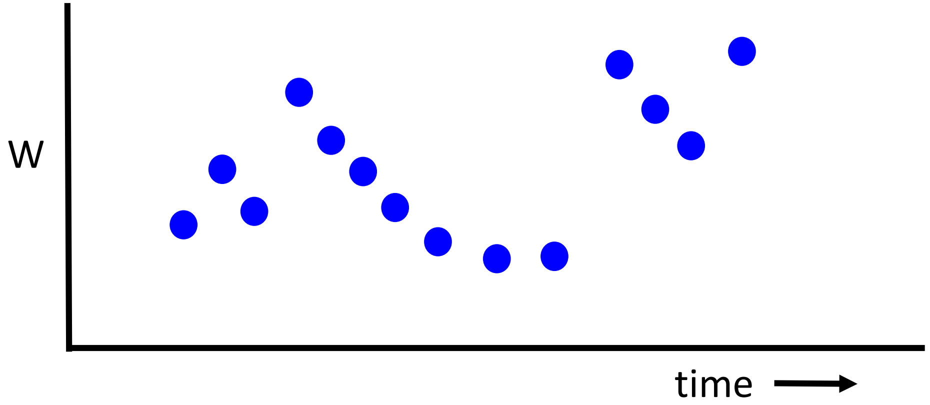
For a given time series of precipitation…



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

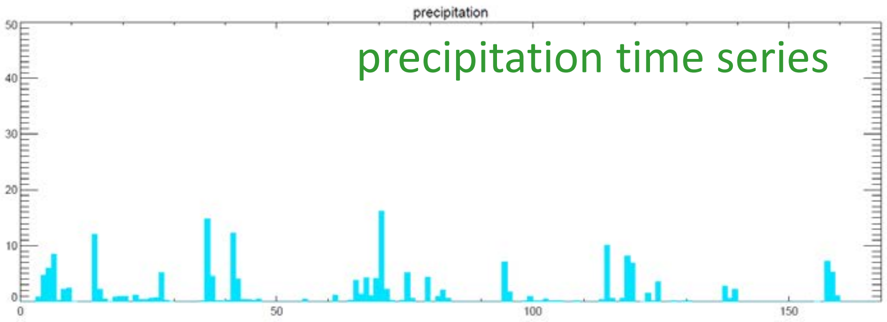
…and an assumed loss function…

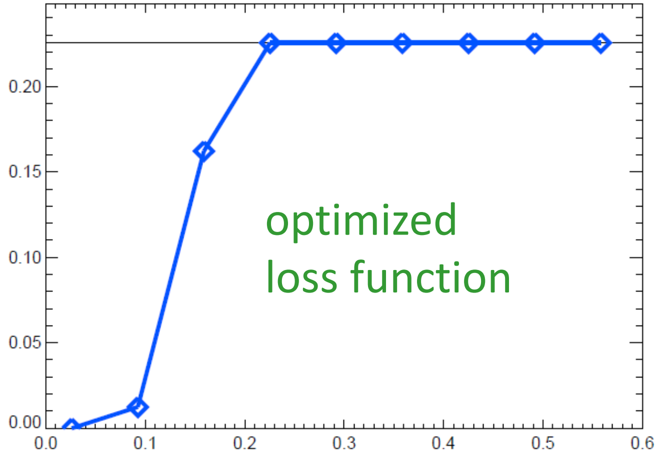


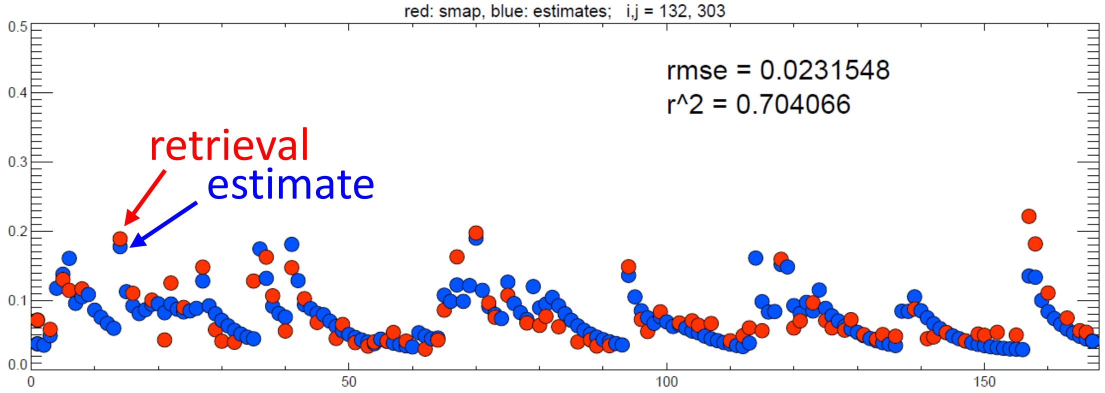


(generated with P data and assumed function alone – no retrievals used)

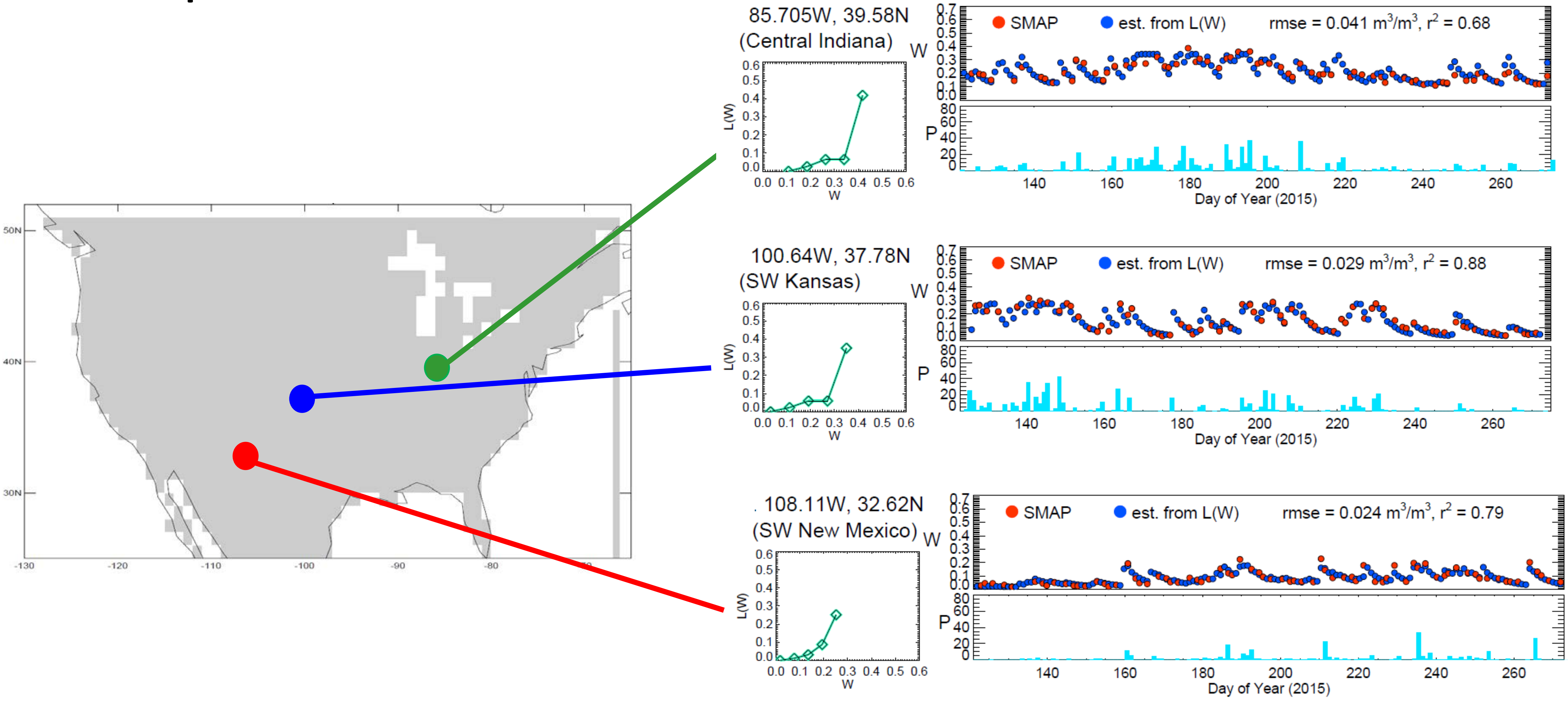
Procedure: through brute force, we find the loss function that produces the best reproduction (in terms of RMSE) of the SMAP Level 2 soil moisture time series.







Examples of Loss Functions

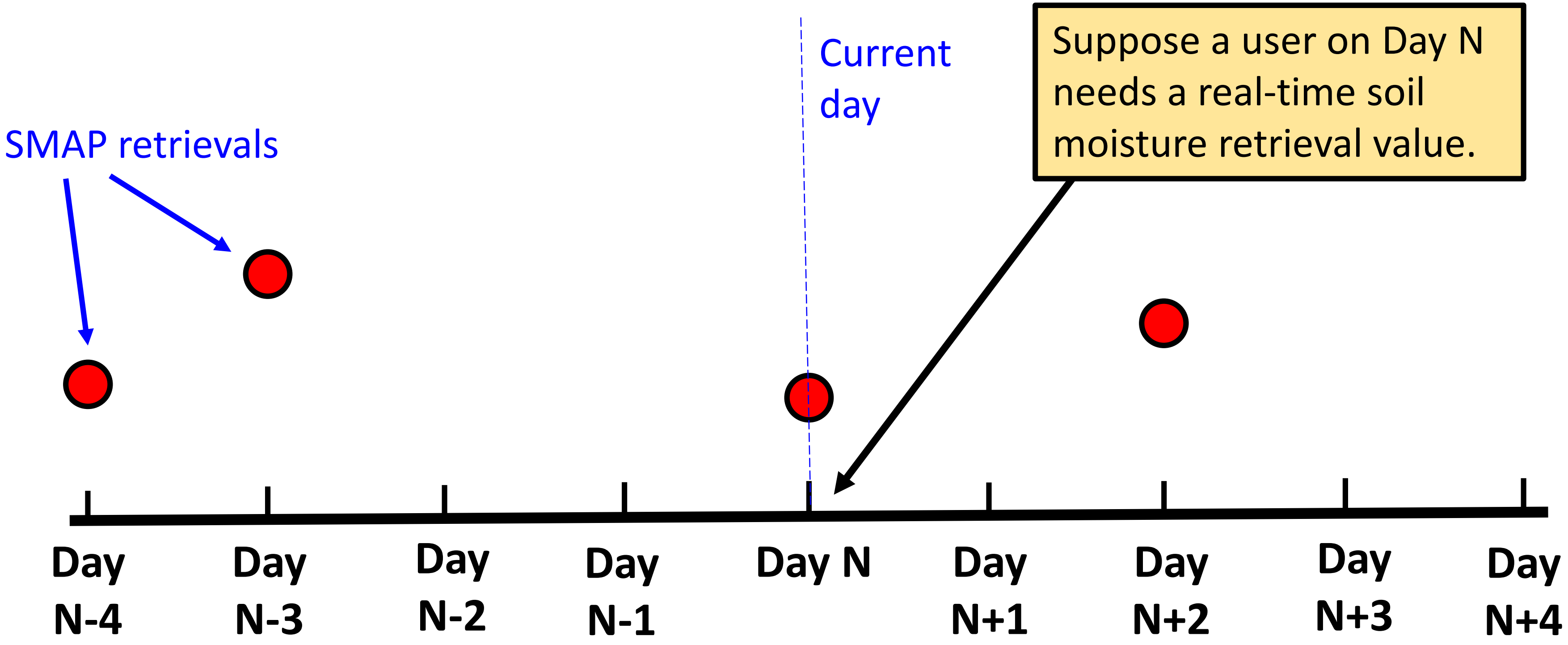


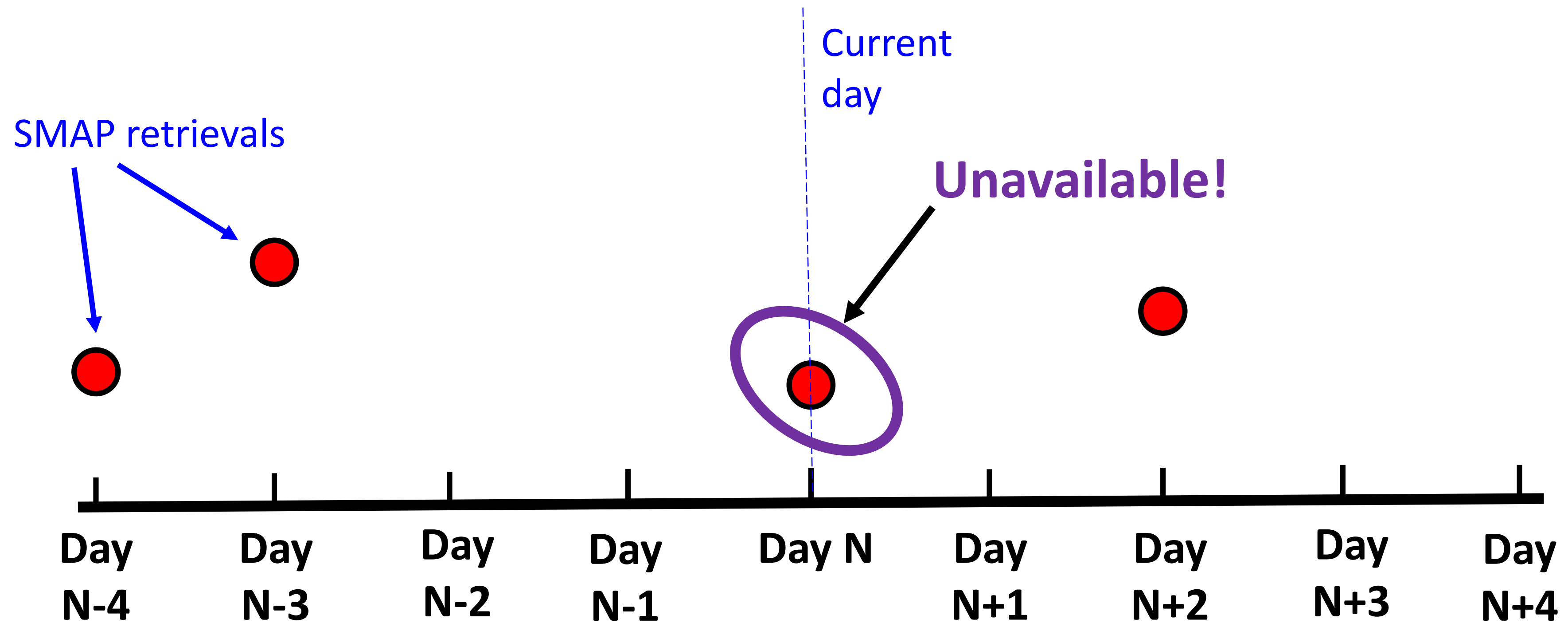
What can we do with these loss functions?

Answer: Many things. Discussed here:

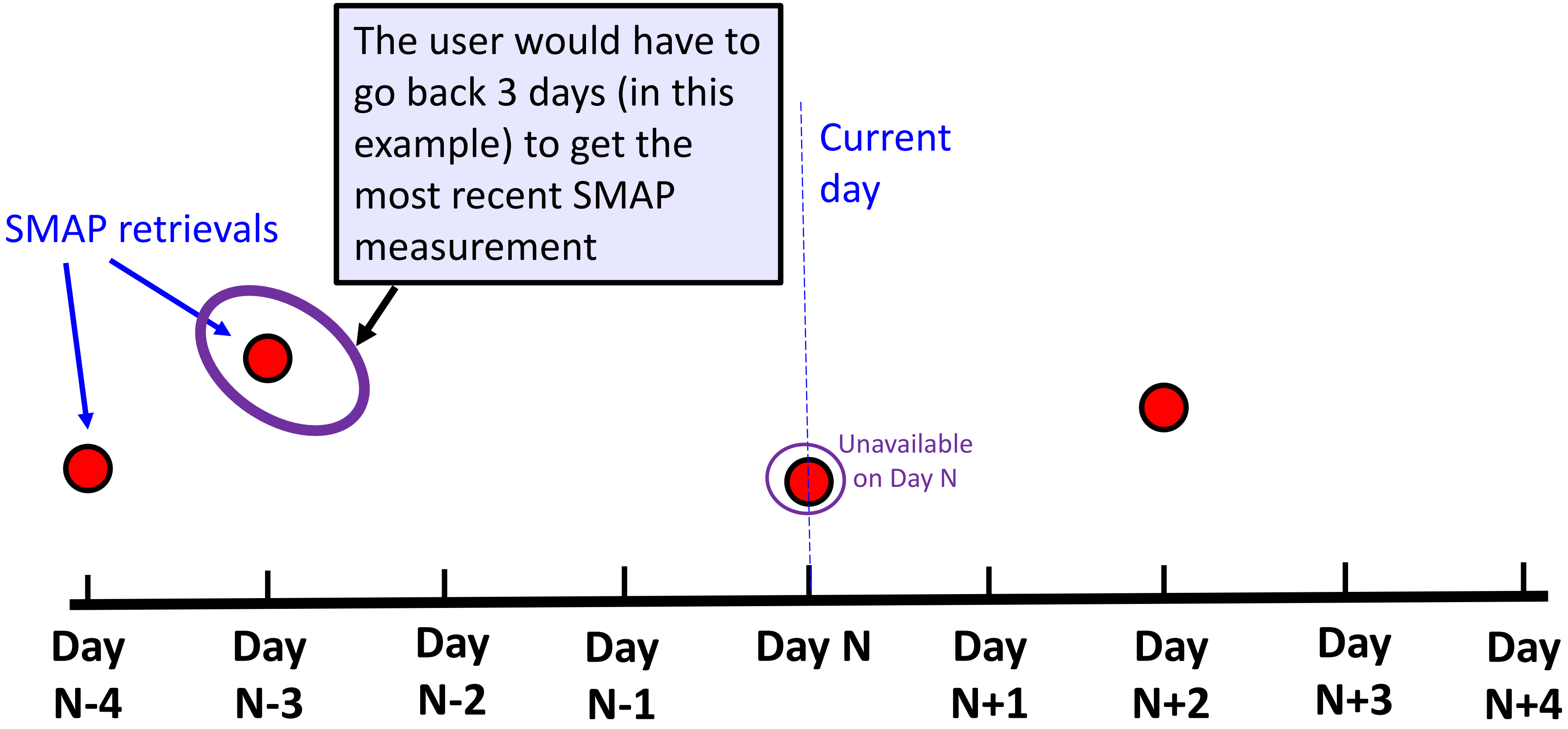
-- Decreasing the latency of SMAP retrieval information

-- Providing soil moisture forecasts

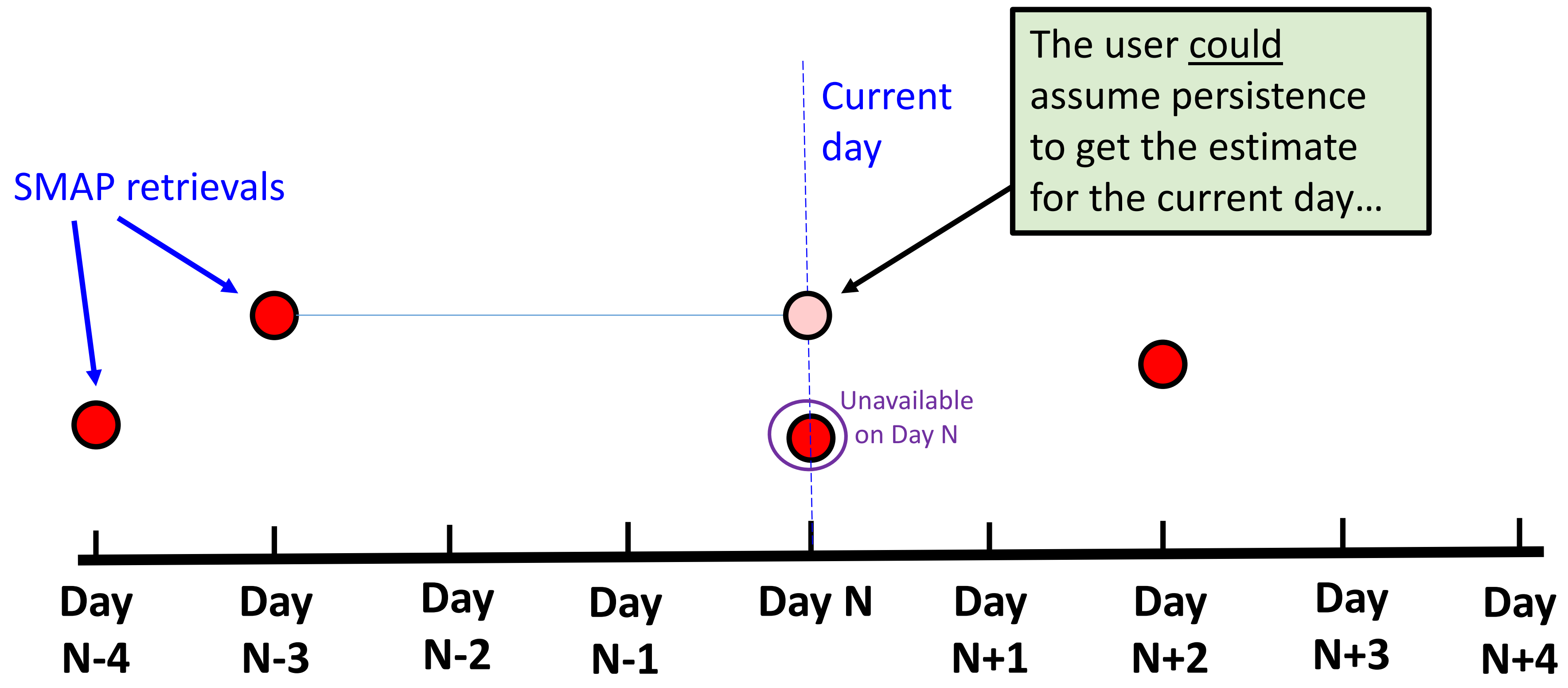

SMAP retrievals
Current day
Suppose a user on Day N needs a real-time soil moisture retrieval value.
Day N-4
Day N-3
Day N-2
Day N-1
Day N
Day N+1
Day N+2
Day N+3
Day N+4

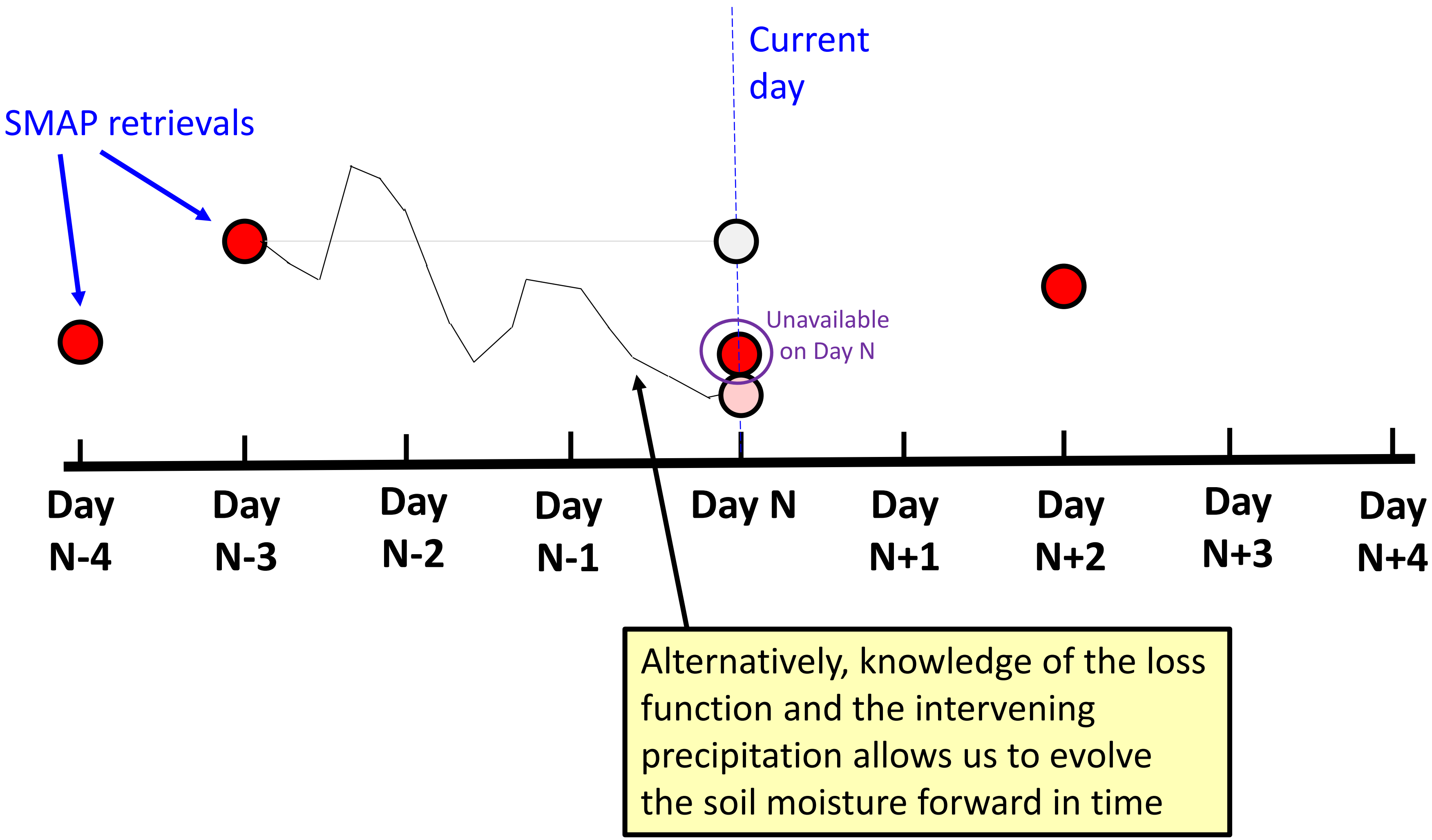


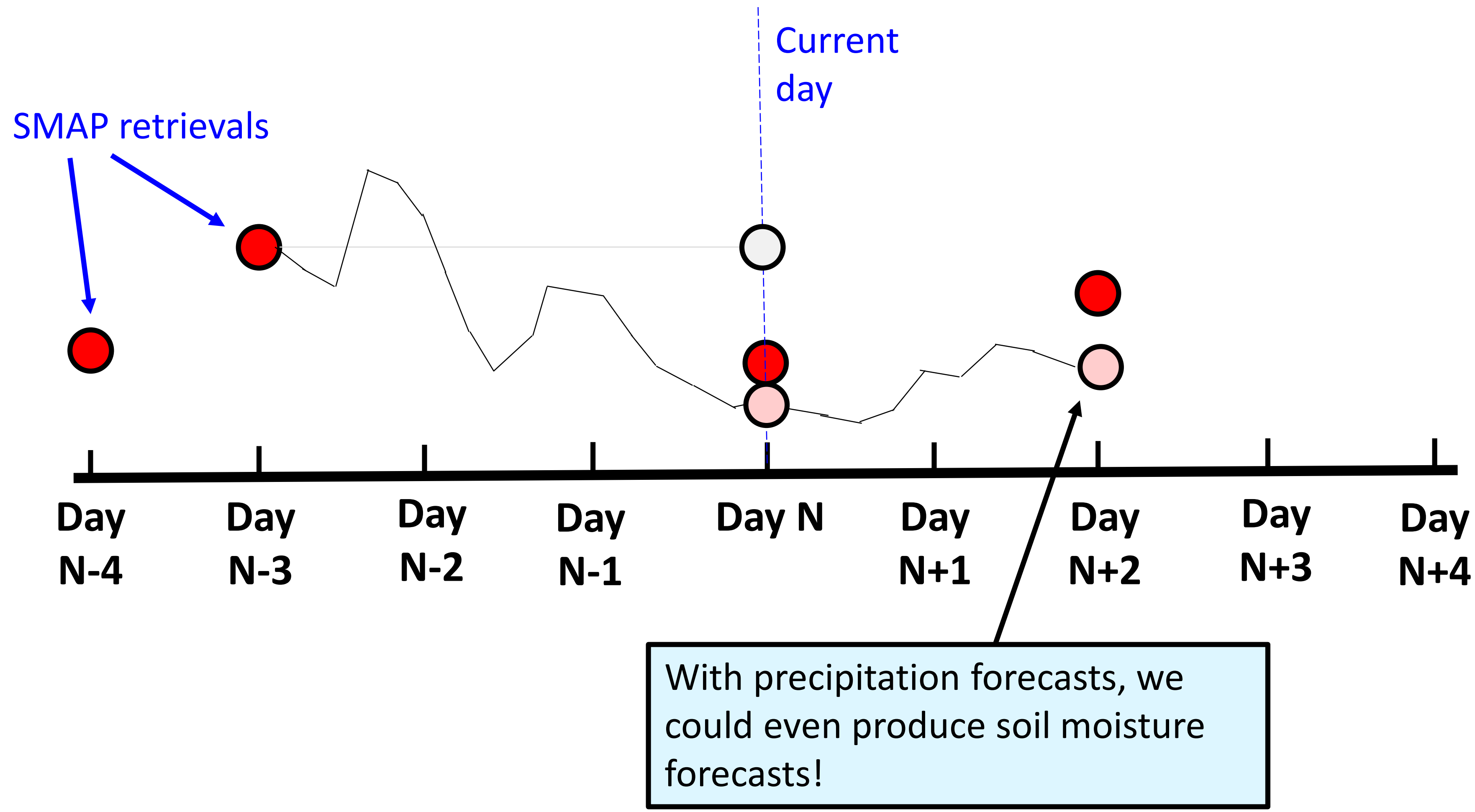
Given product latency, this retrieval value would not be available until Day N+1

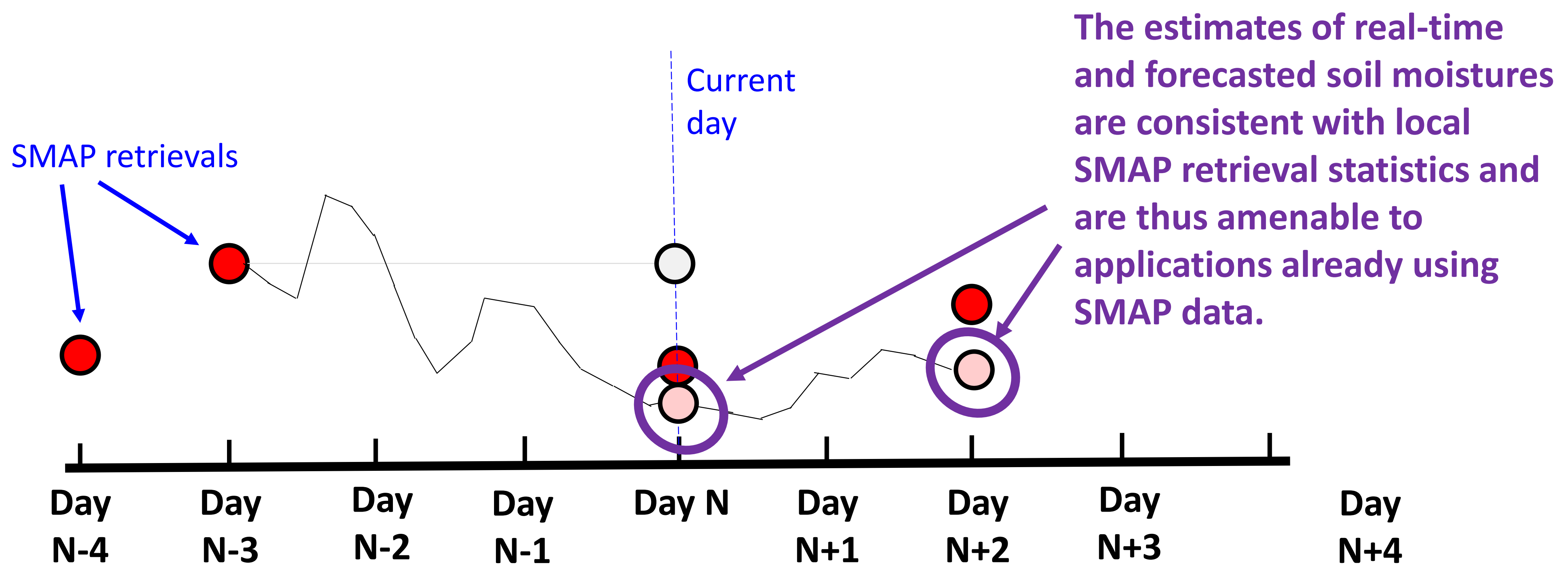


The user would have to go back 3 days (in this example) to get the most recent SMAP measurement
SMAP retrievals
Current day
Unavailable on Day N
Day N-4
Day N-3
Day N-2
Day N-1
Day N
Day N+1
Day N+2
Day N+3
Day N+4


SMAP retrievals
Current day
The user could assume persistence to get the estimate for the current day…
Unavailable on Day N
Day N-4
Day N-3
Day N-2
Day N-1
Day N
Day N+1
Day N+2
Day N+3
Day N+4


SMAP retrievals
Current
day
Unavailable
on Day N
Day
N-4
Day
N-3
Day
N-2
Day
N-1
Day N
Day
N+1
Day
N+2
Day
N+3
Day
N+4
Alternatively, knowledge of the loss function and the intervening precipitation allows us to evolve the soil moisture forward in time

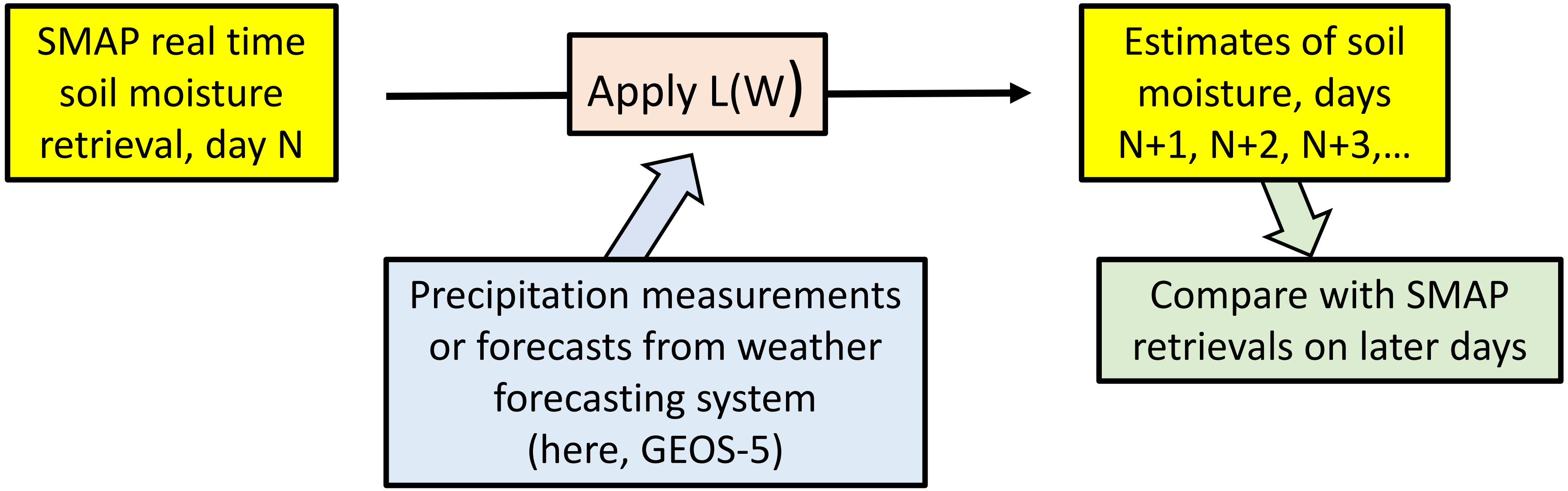

Current day
SMAP retrievals
Day N-4
Day N-3
Day N-2
Day N-1
Day N
Day N+1
Day N+2
Day N+3
Day N+4
With precipitation forecasts, we could even produce soil moisture forecasts!

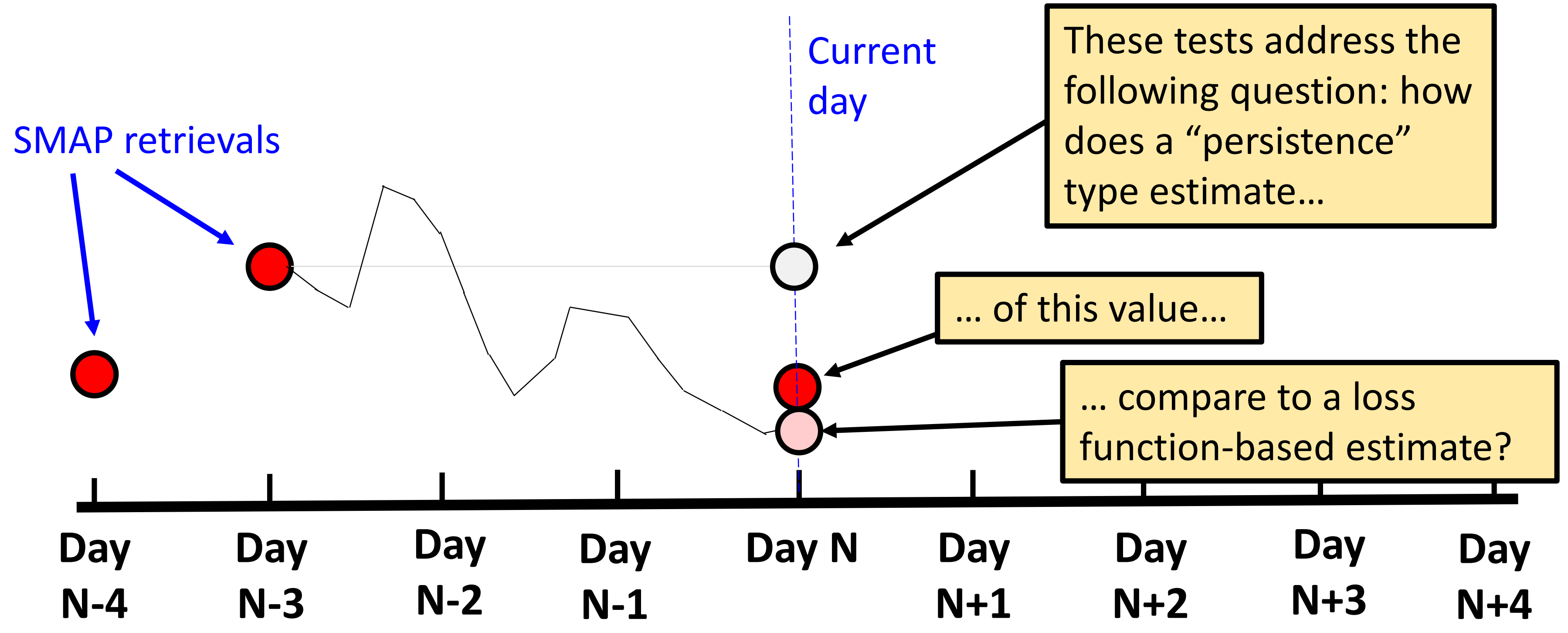

The estimates of real-time and forecasted soil moistures are consistent with local SMAP retrieval statistics and are thus amenable to applications already using SMAP data.
Current day
SMAP retrievals
Day N-4
Day N-3
Day N-2
Day N-1
Day N
Day N+1
Day N+2
Day N+3
Day N+4

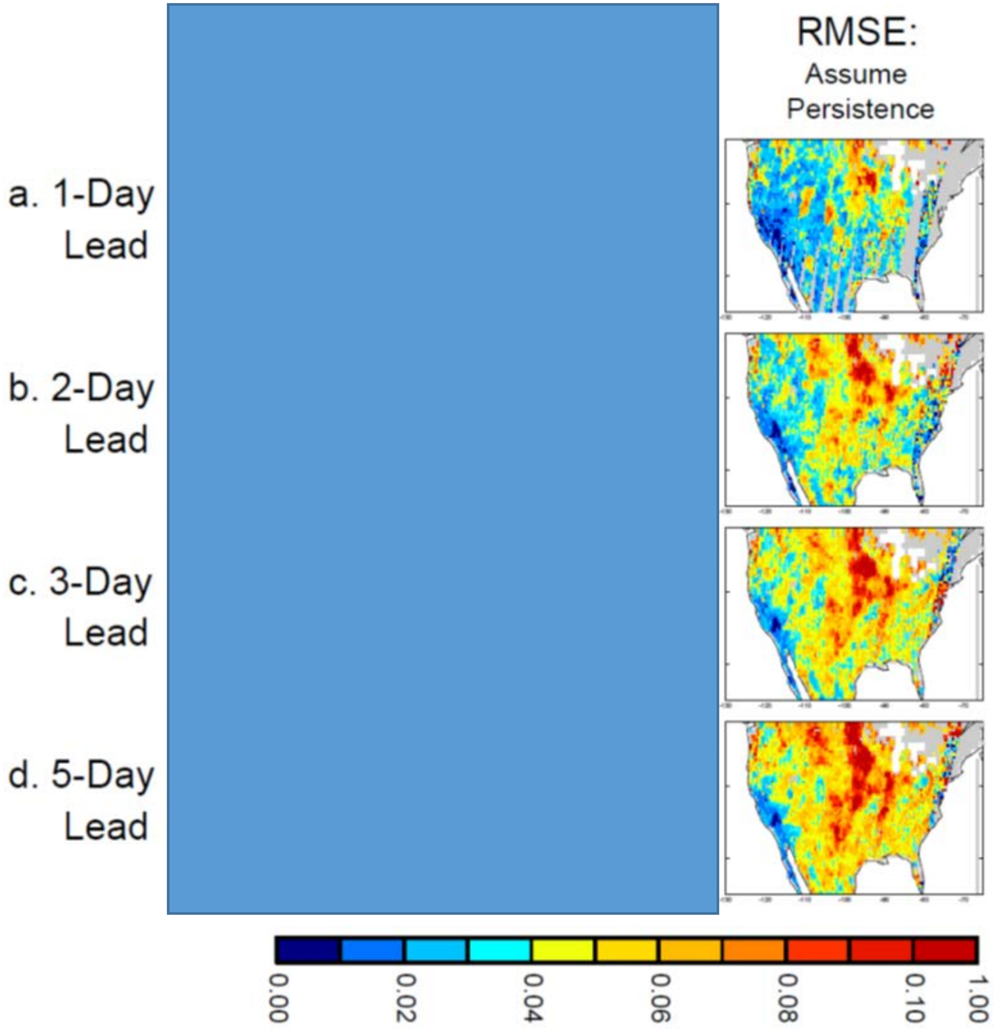
Overall strategy to test such soil moisture estimation:

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.):



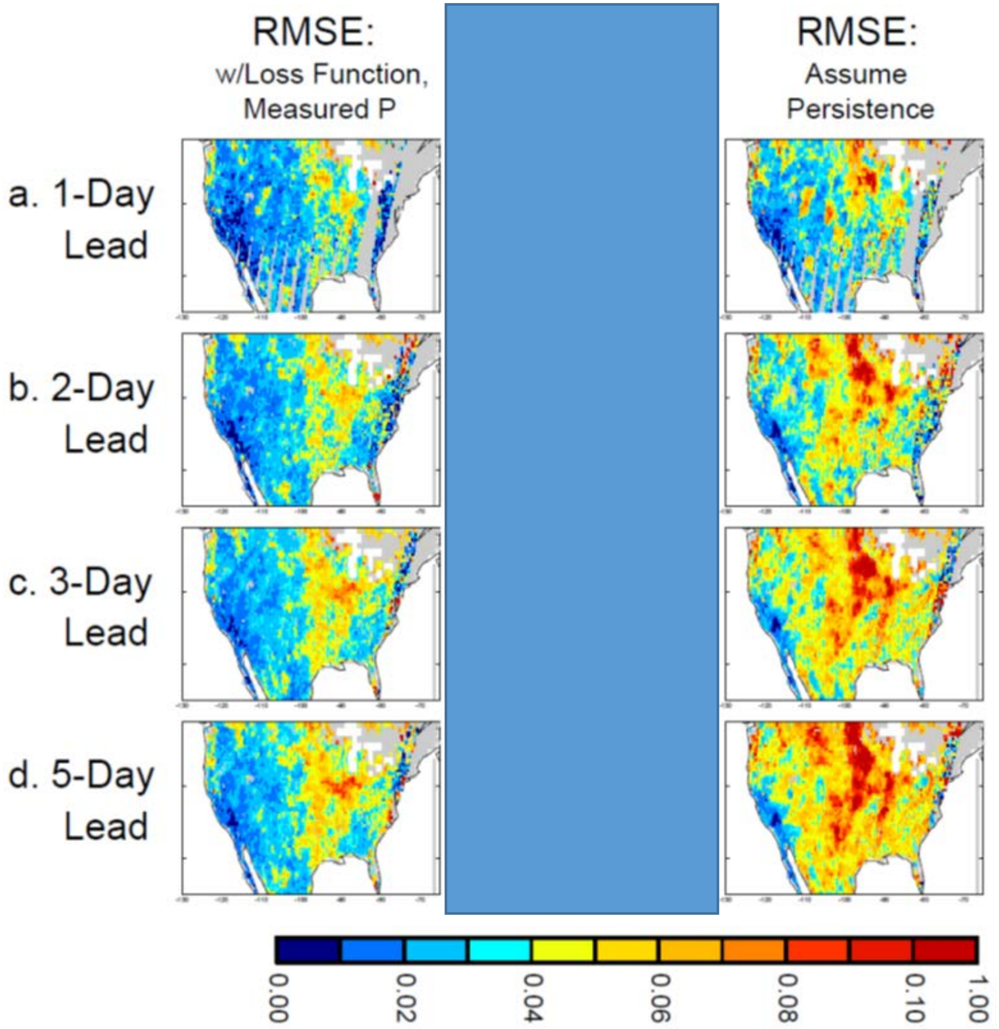

SMAP retrievals
Current day
These tests address the following question: how does a "persistence" type estimate…
… of this value…
… compare to a loss function-based estimate?
Day N-4
Day N-3
Day N-2
Day N-1
Day N
Day N+1
Day N+2
Day N+3
Day N+4



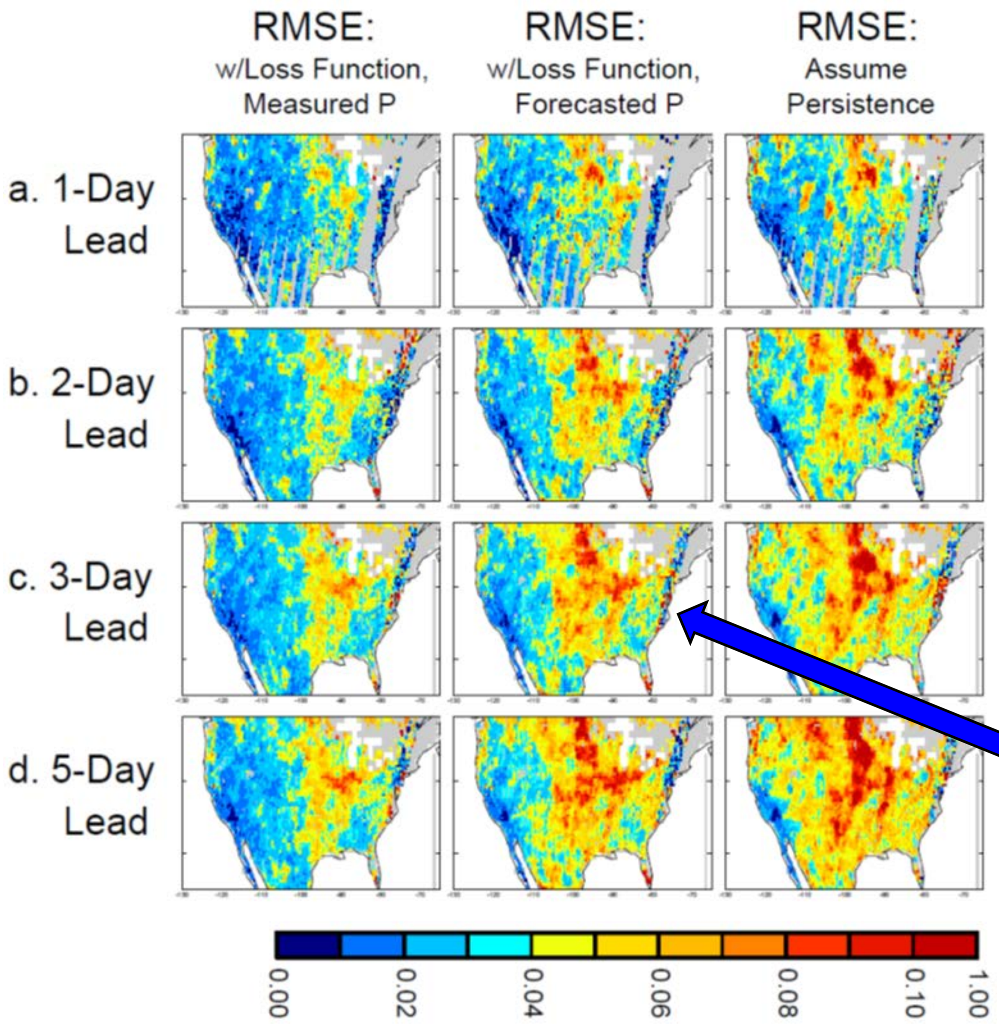
Persistence: If you assign the most recent SMAP retrieval value to the current day, these are the errors you obtain.

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

<u>**Loss function:**</u> 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:

- Loss functions (descriptions of how soil moisture decreases with evaporation and drainage) can be derived from joint analysis of SMAP data and precipitation data.

- 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. Estimati… 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!)

Input: meteorological forcing data (rainfall, snow, wind speed, air temperature…)

Input: energy and water states from previous time step

Land surface model:

Uses conservation equations, parameterizations to transfer water and energy between reservoirs, compute fluxes.

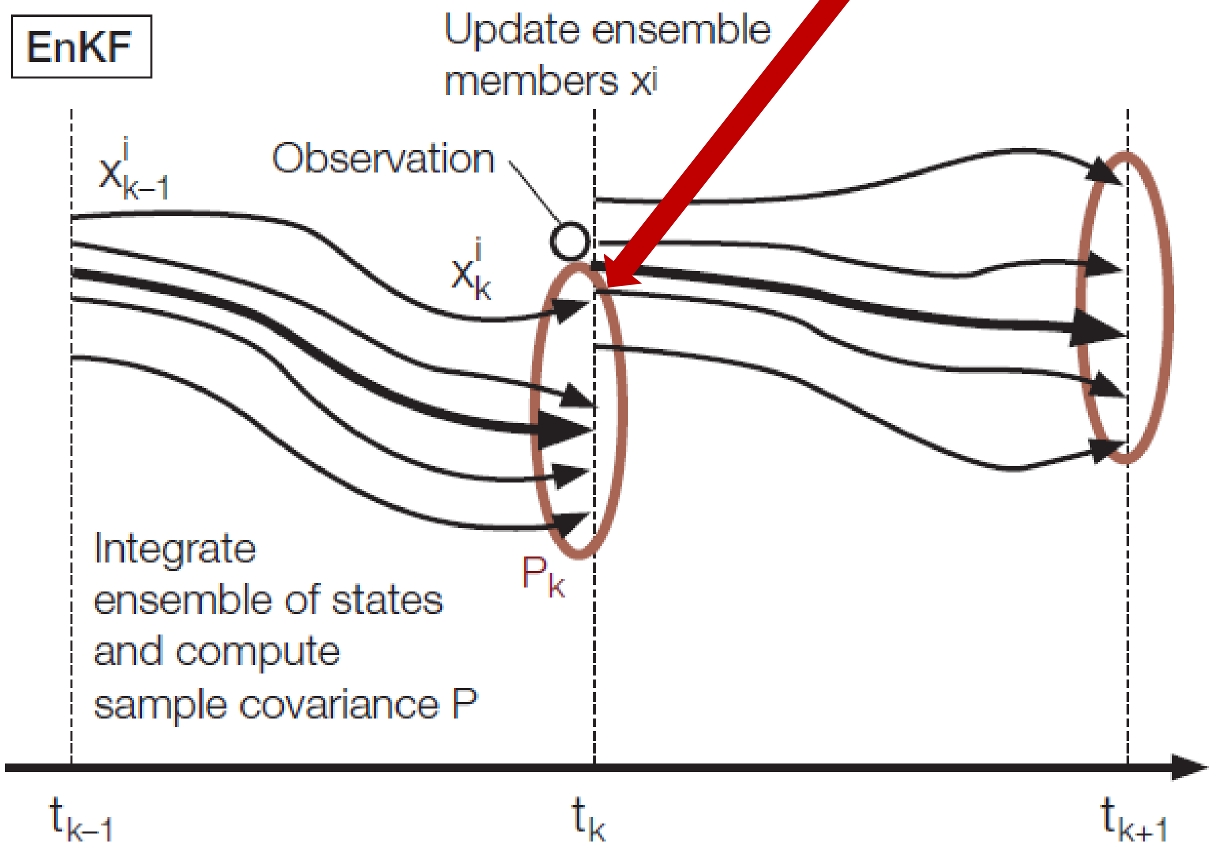
Output: Diagnostic surface fluxes (ET, runoff, sensible heat…)

Output: updated states for use in next time step

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

Approach 1:
Data assimilation

SMAP data (T_b) are used to update model soil moisture state variables at specific times



model simulation time ⟶

With data assimilation:

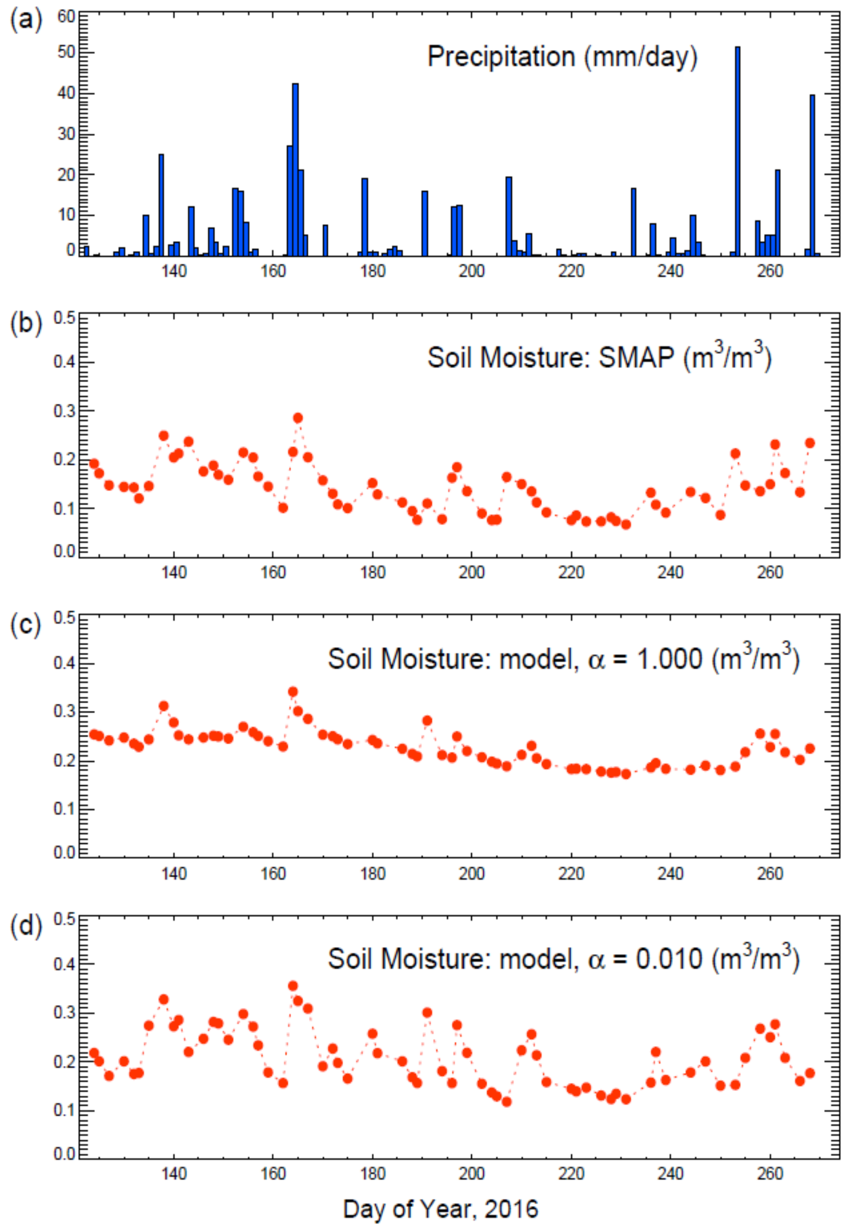
1) the land model is used "as is".
2) Only the prognostic states are modified.

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 <u>not</u> 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 <u>complementary</u> information?

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

BL (baseline): uncalibrated open-loop simulation

Control

BL_DA: uncalibrated simulation with data assimilation

OPT: open-loop simulation with calibrated parameters

OPT_DA: simulation with data assimilation and calibrated parameters

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

(use SMAP data to adjust the states)

Test improvements from model calibration

(use SMAP data to adjust the model)

BL (baseline): uncalibrated open-loop simulation

Control

BL_DA: uncalibrated simulation with data assimilation

OPT: open-loop simulation with calibrated parameters

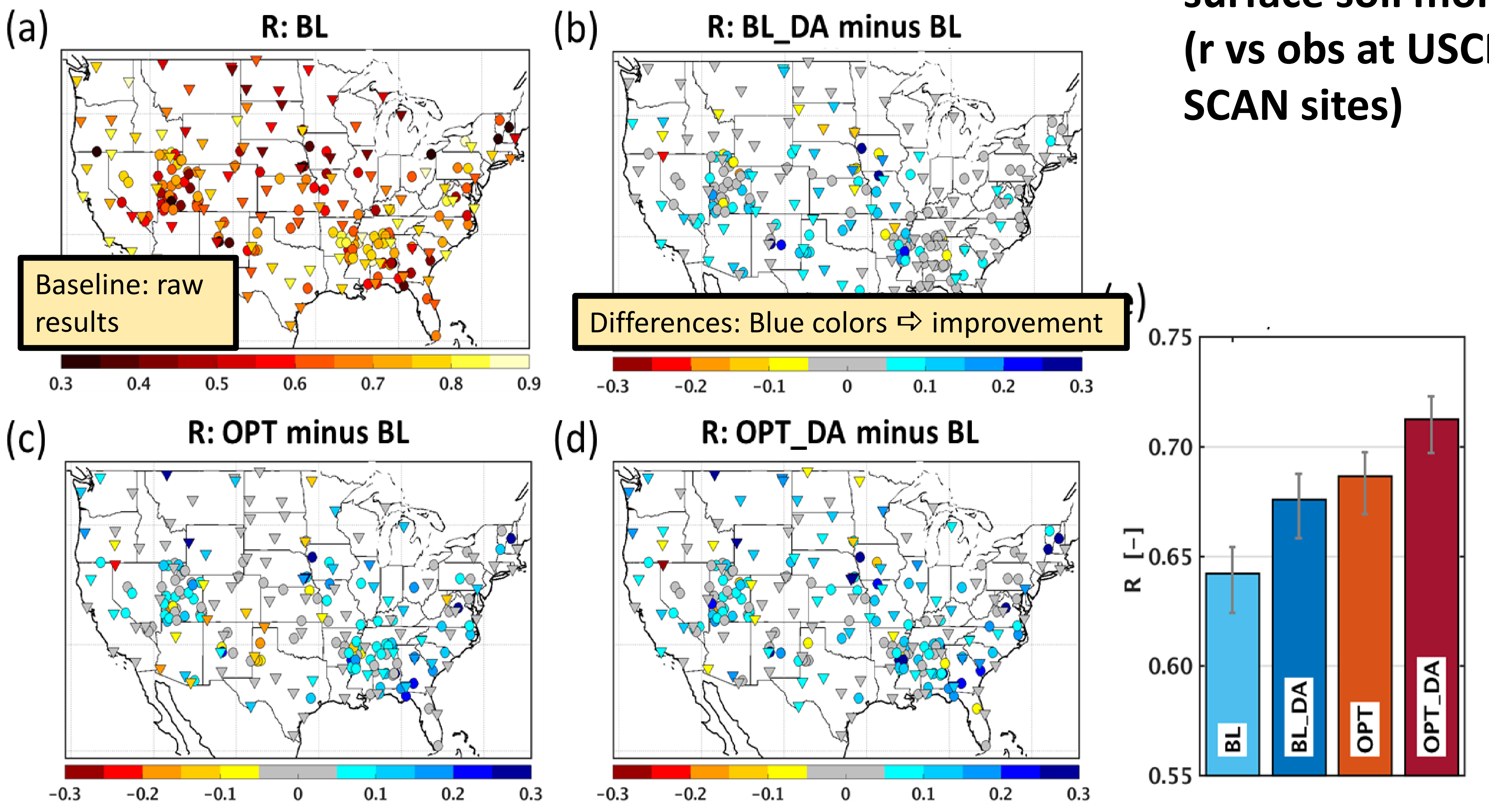
OPT_DA: simulation with data assimilation and calibrated parameters

Does this run do the best of all? ⇨ *Is the extraction of information from SMAP data through the two approaches complementary?*

Test improvements from data assimilation

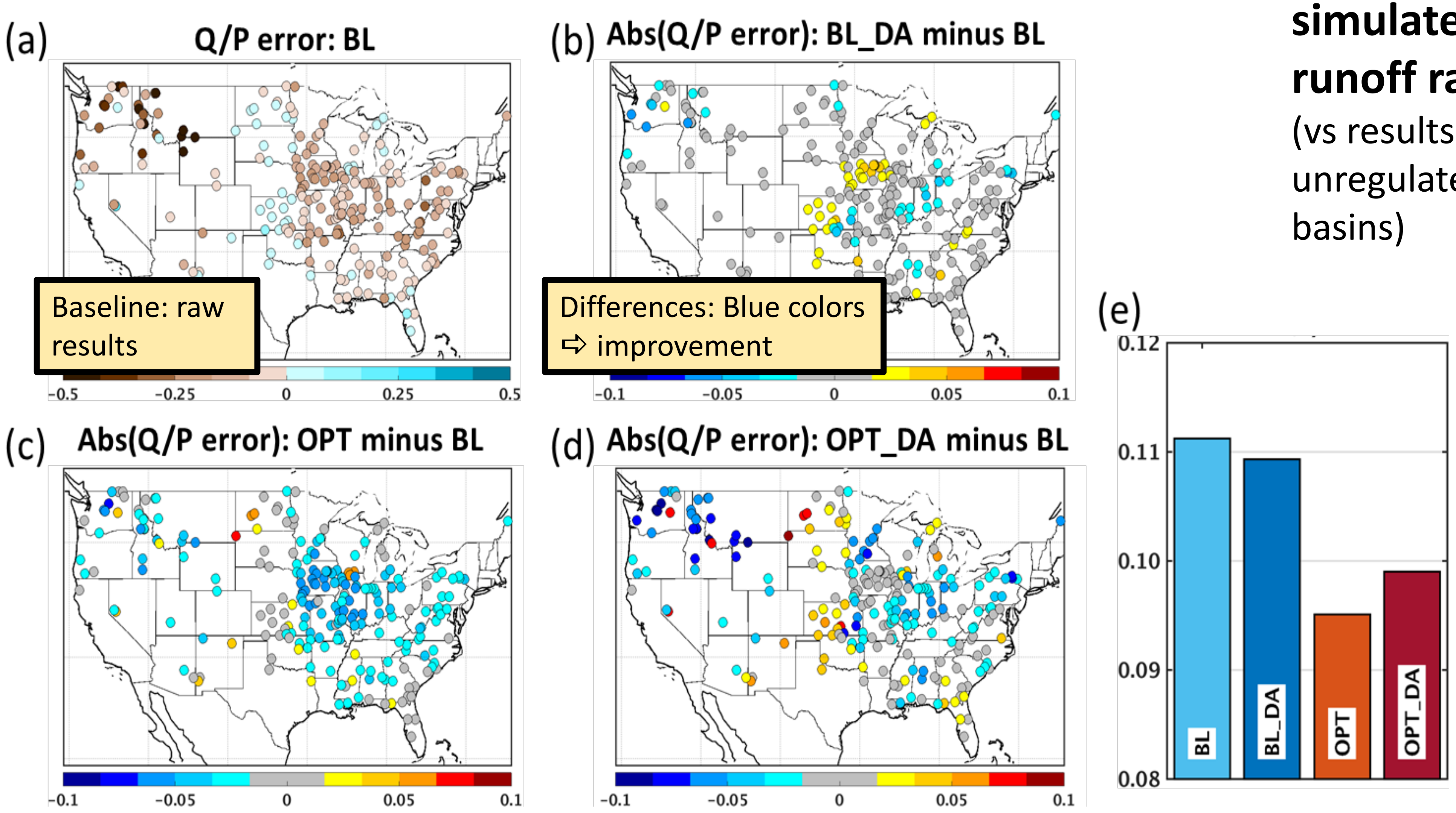
Test improvements from model calibration

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



Test improvements from data assimilation

Test improvements from model calibration



Error (bias) of simulated runoff ratio

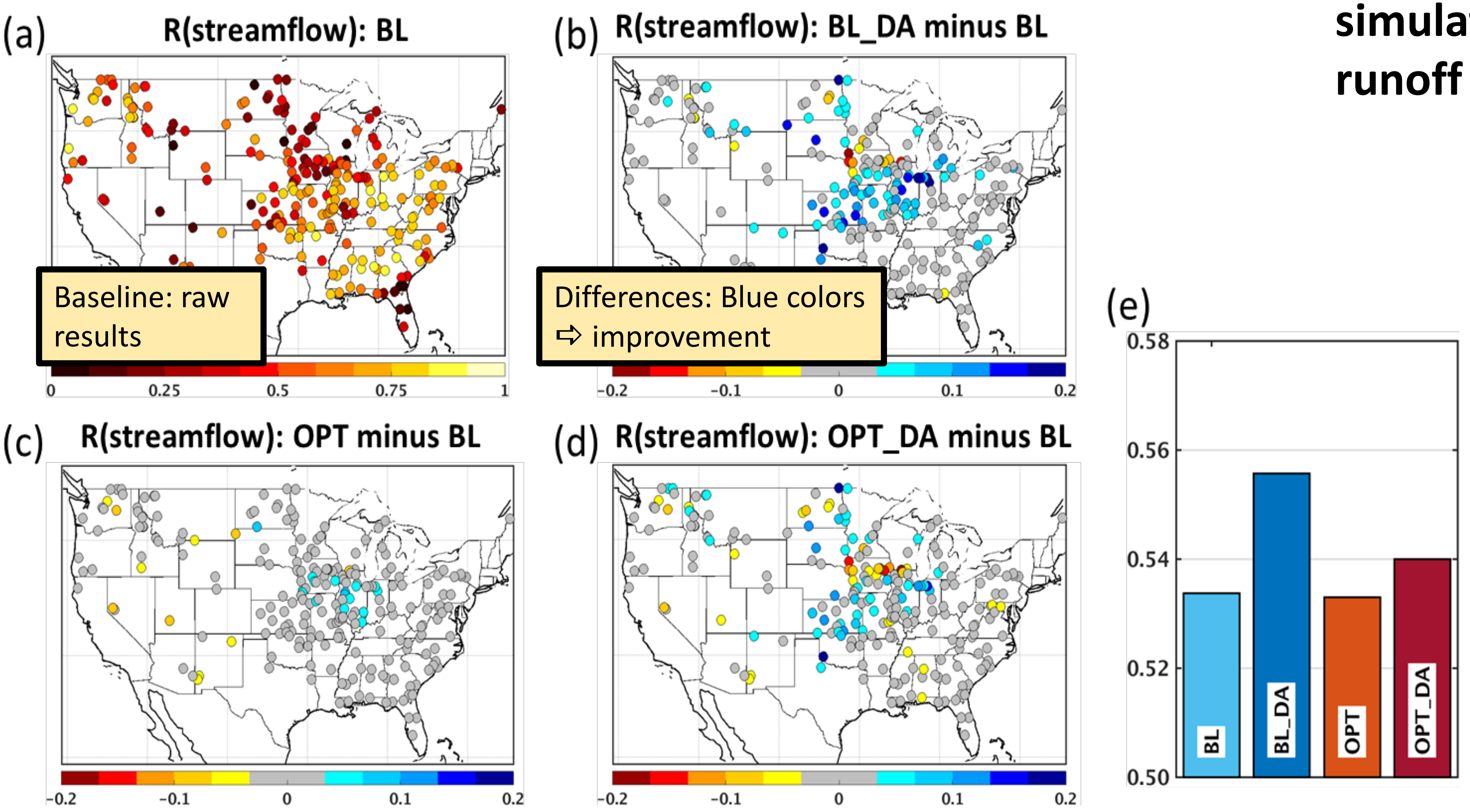
(vs results at unregulated USGS basins)

What's their favorite dessert?

Test improvements from data assimilation

Timing of simulated runoff (r vs obs)

Test improvements from model calibration



To summarize the impacts:

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

As it turns out, we can explain at least <u>some</u> of this complementarity…

Streamflow

	Data assimilation	Model calibration
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 ⇨ little impact on the climatology of the model.

With model calibration, the nature of the model itself is changed ⇨ 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.

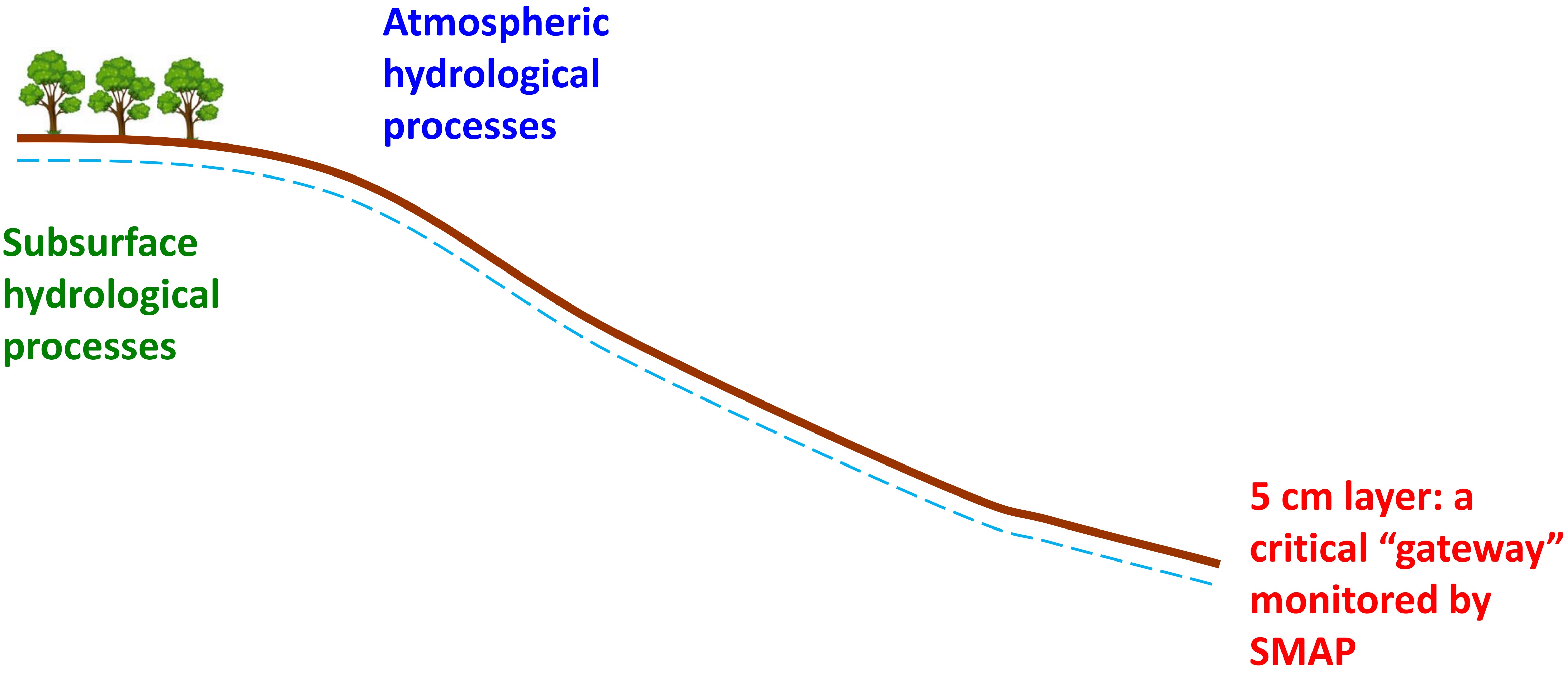
Three topics:

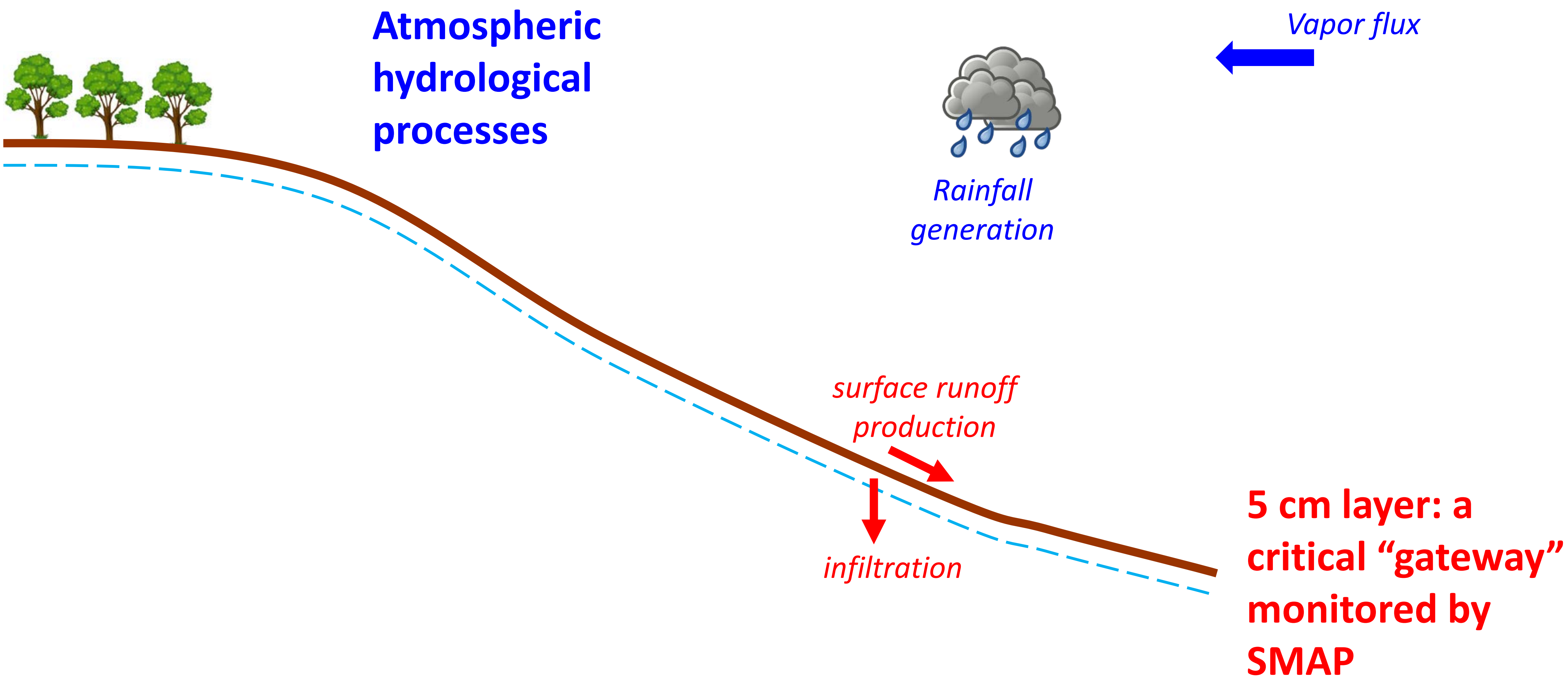
1. "Loss functions": the characterization of soil moisture dynamics.

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

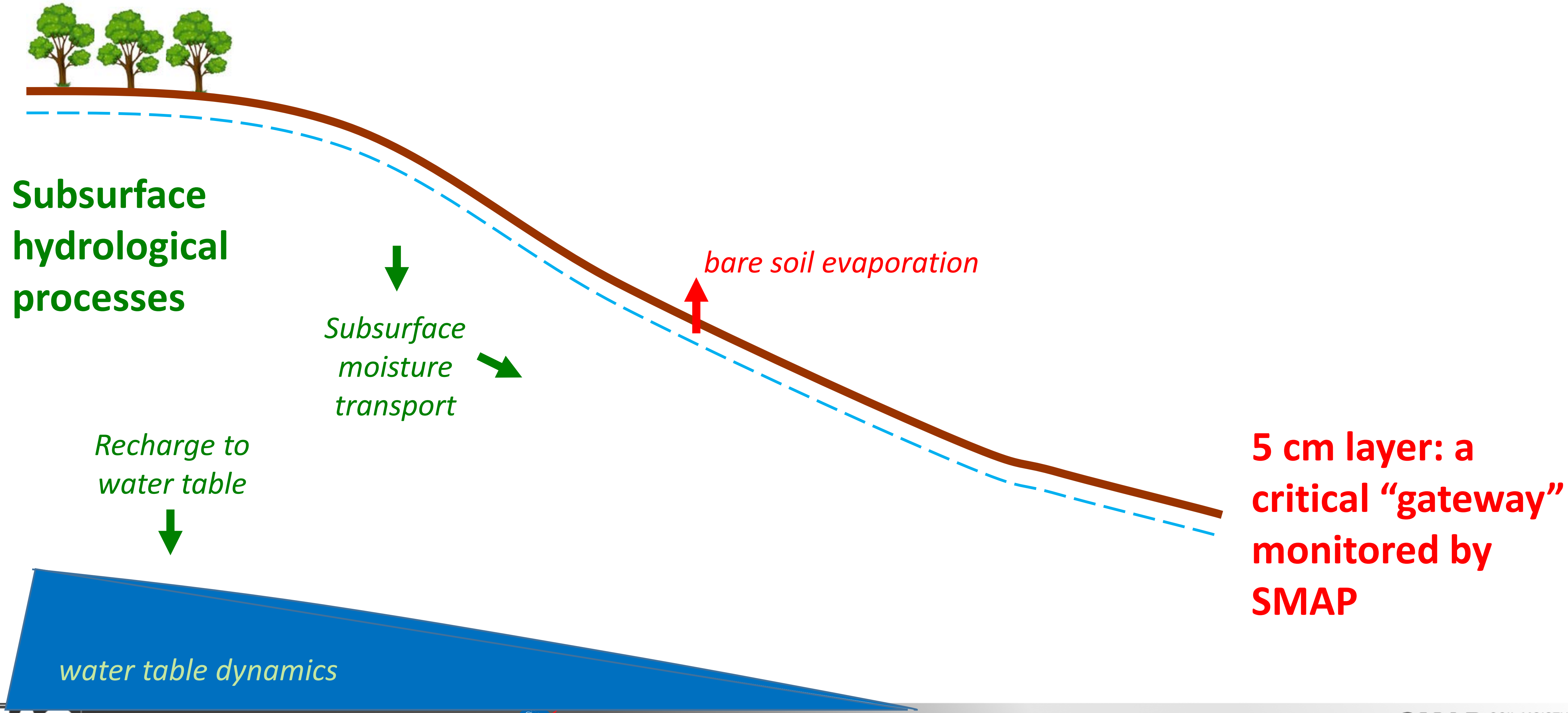
3. Estimating rainfall and streamflow with SMAP data.

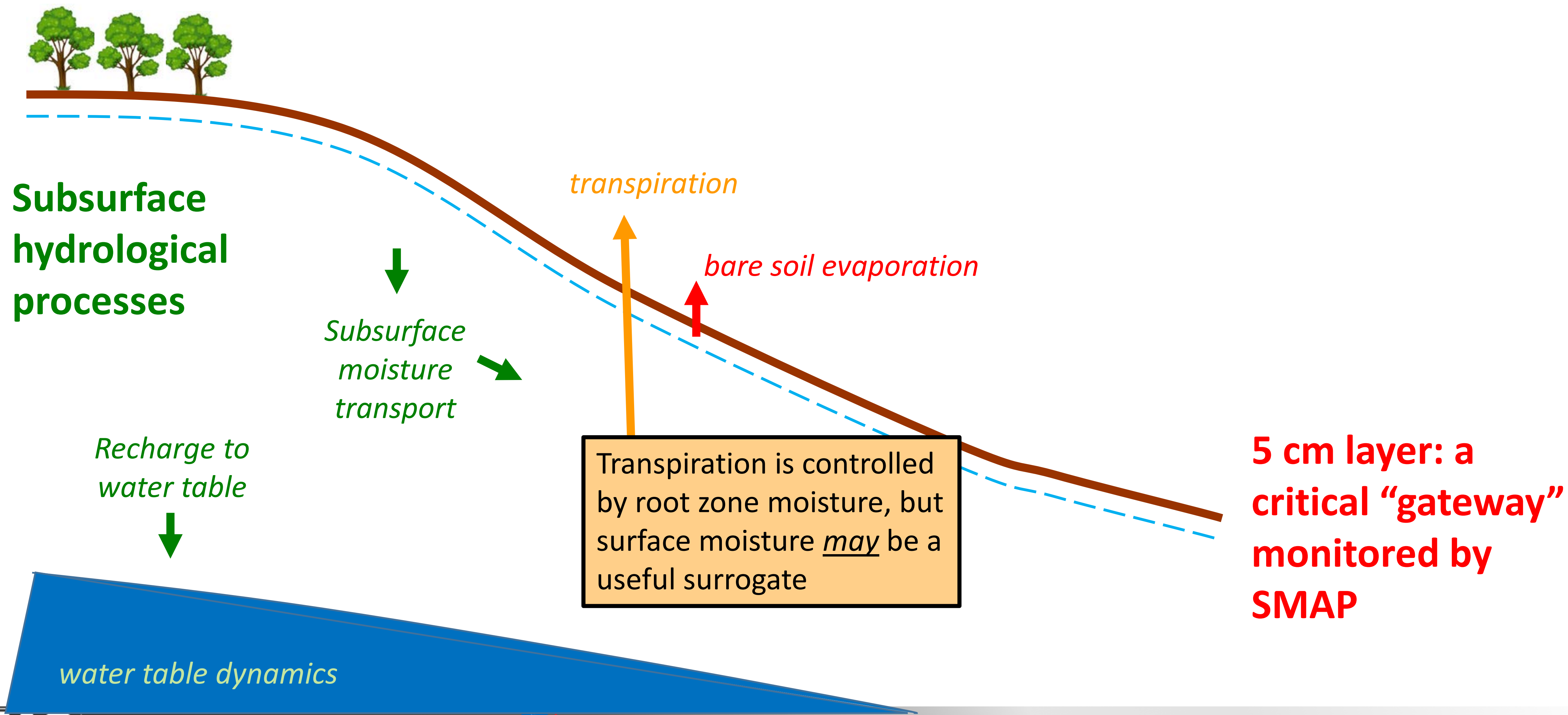
Paper under review in Water Resources Research


Atmospheric hydrological processes
Subsurface hydrological processes
5 cm layer: a critical “gateway” monitored by SMAP



Atmospheric hydrological processes
Vapor flux
Rainfall generation
surface runoff production
infiltration
5 cm layer: a critical "gateway" monitored by SMAP

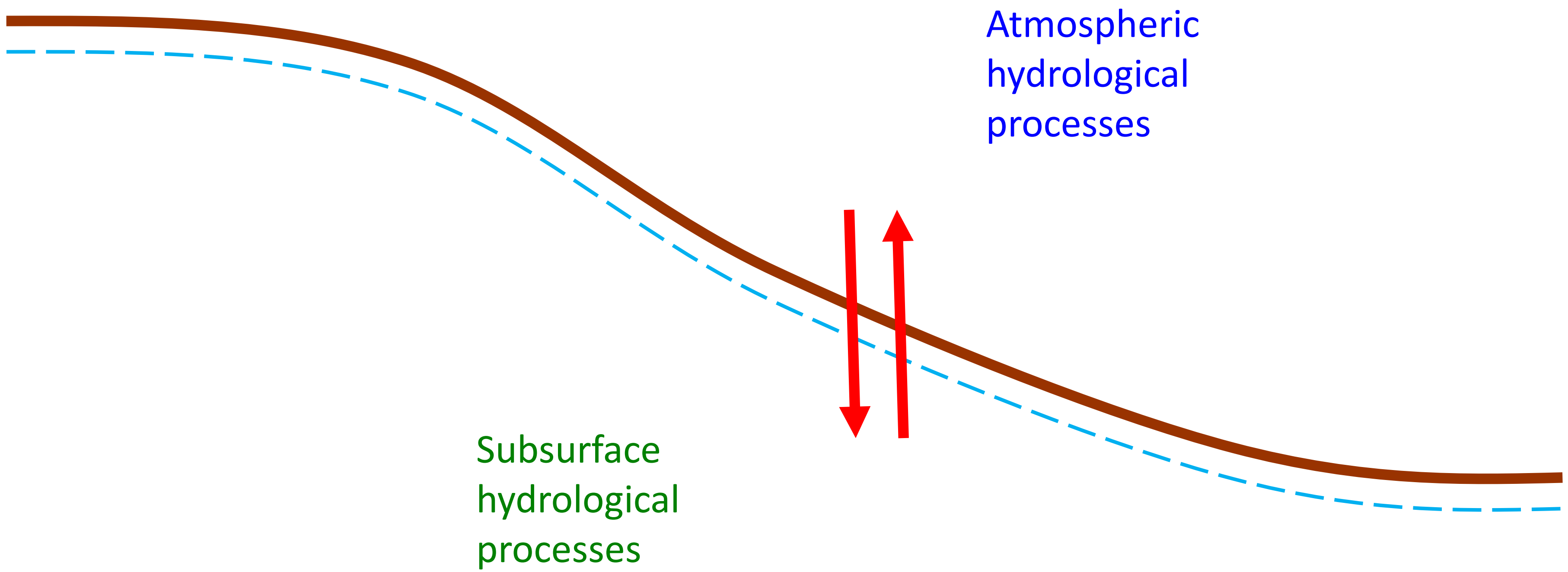

Subsurface hydrological processes
Subsurface moisture transport
bare soil evaporation
Recharge to water table
water table dynamics
5 cm layer: a critical “gateway” monitored by SMAP



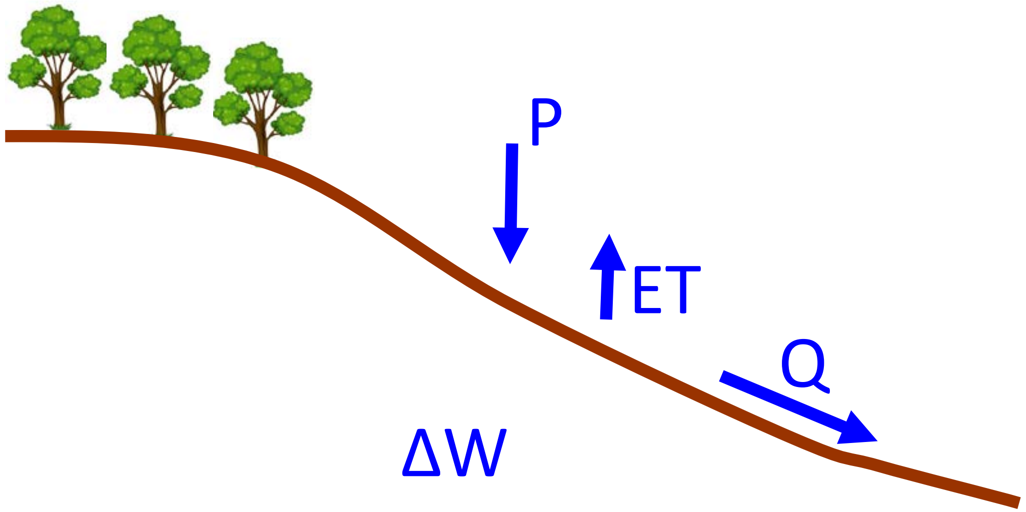
Subsurface hydrological processes
transpiration
bare soil evaporation
Subsurface moisture transport
Recharge to water table
Transpiration is controlled by root zone moisture, but surface moisture *may* be a useful surrogate
5 cm layer: a critical "gateway" monitored by SMAP
water table dynamics

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.



precipitation

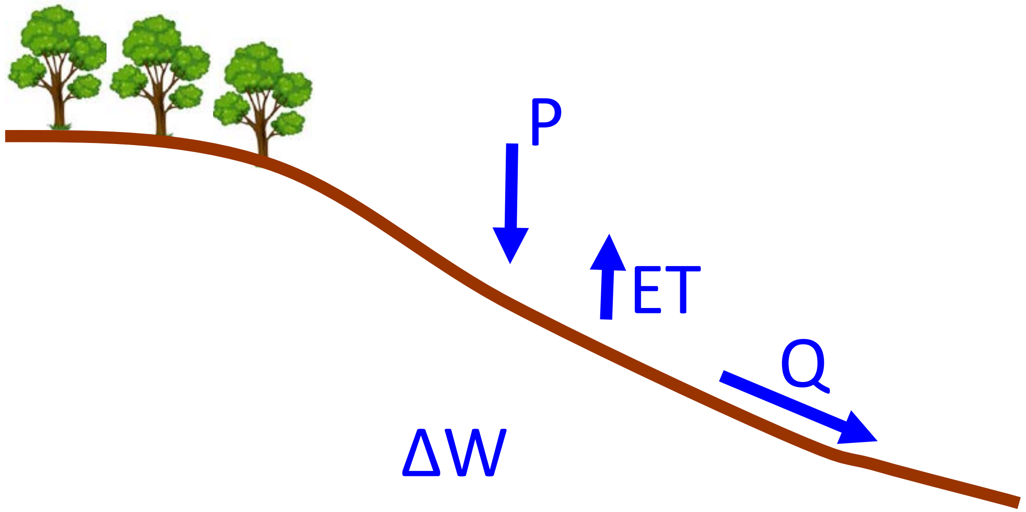
runoff (streamflow)

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

evapotranspiration

change in storage

<u>Sample Hydrological Application:</u> Estimating terms in a basin’s water balance.



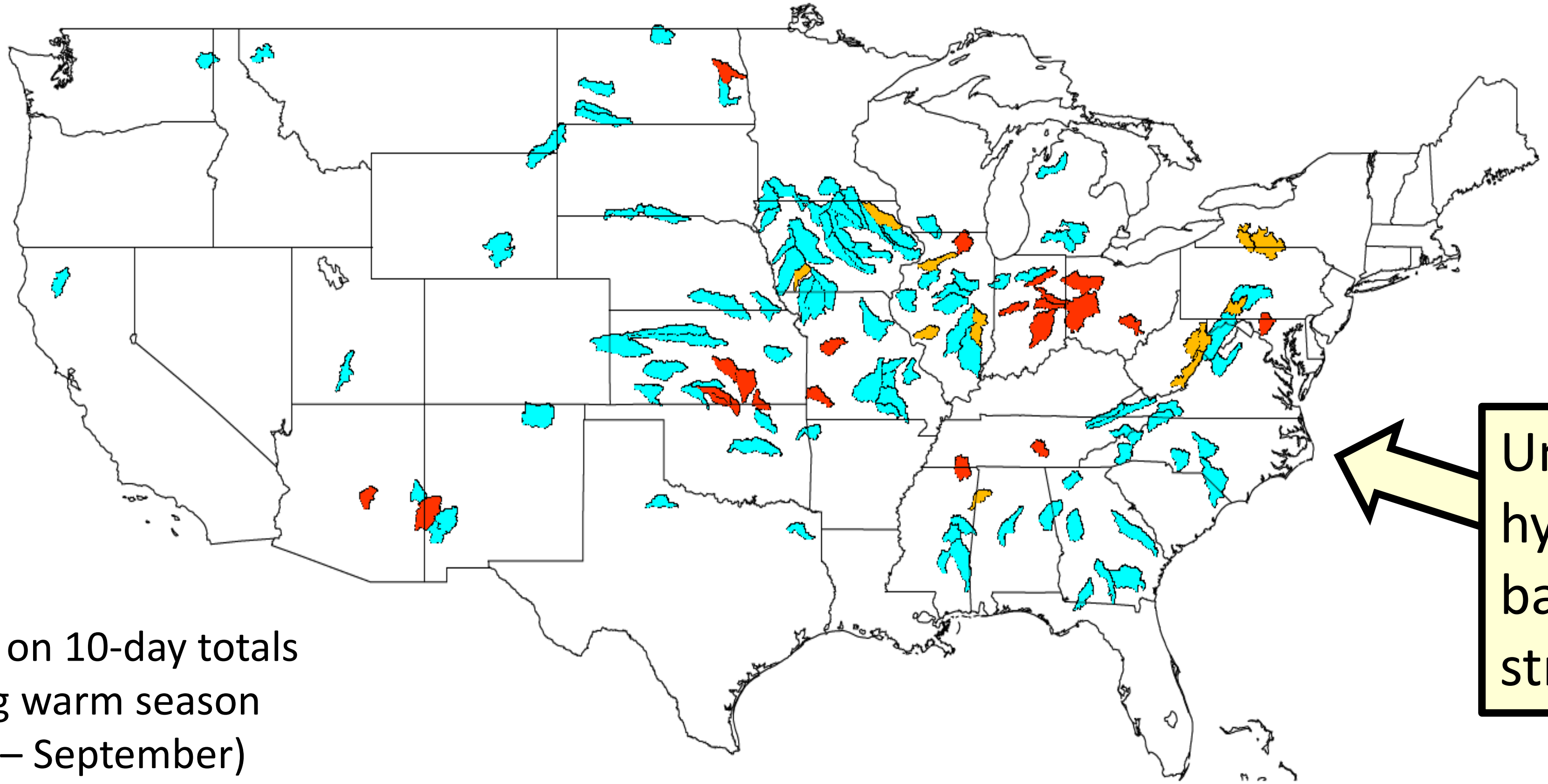
$$P = ET + Q + \Delta W$$

Today: a demonstration of how these terms can be estimated with SMAP soil moisture data

(not perfectly, but with <u>some</u> skill)

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

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

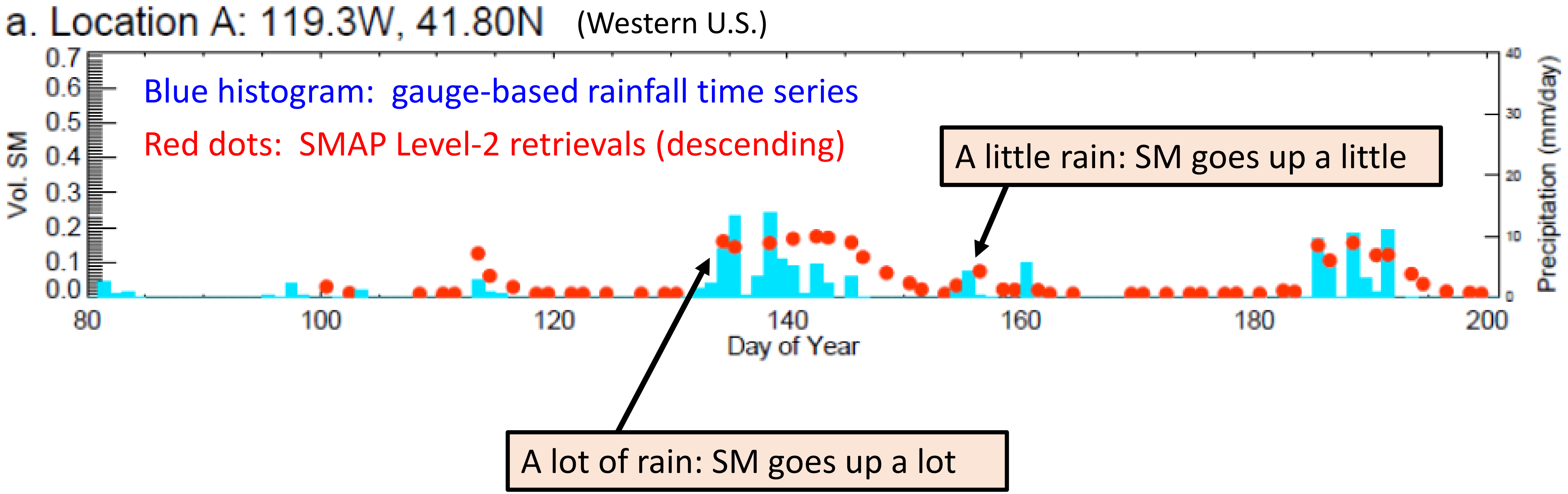


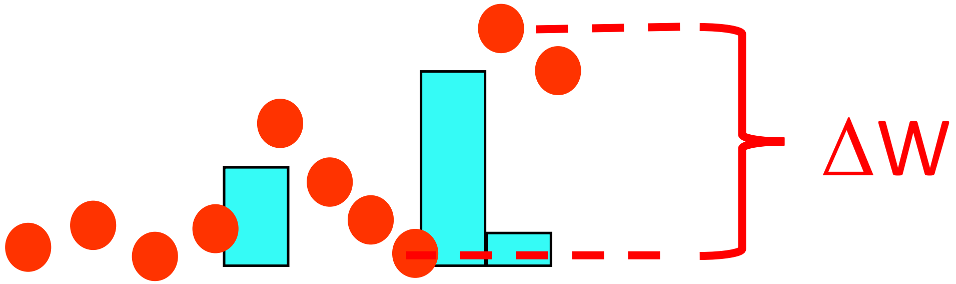
Unregulated hydrological basins with USGS stream gauges

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

Step 1: Precipitation Estimation

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

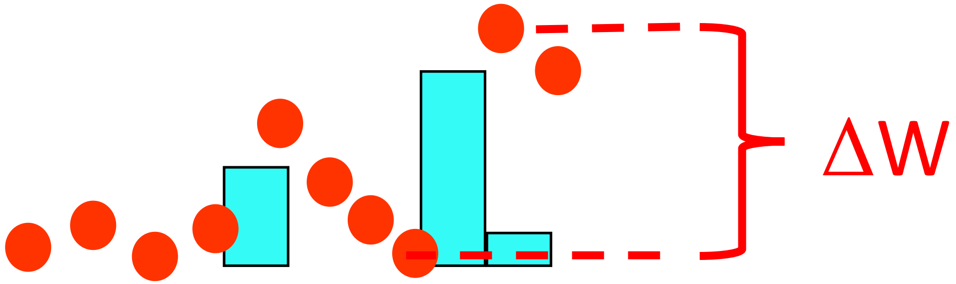




Precipitation estimation approach from Brocca et al. (2013):

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

W_{ave} = average of the two consecutive retrievals



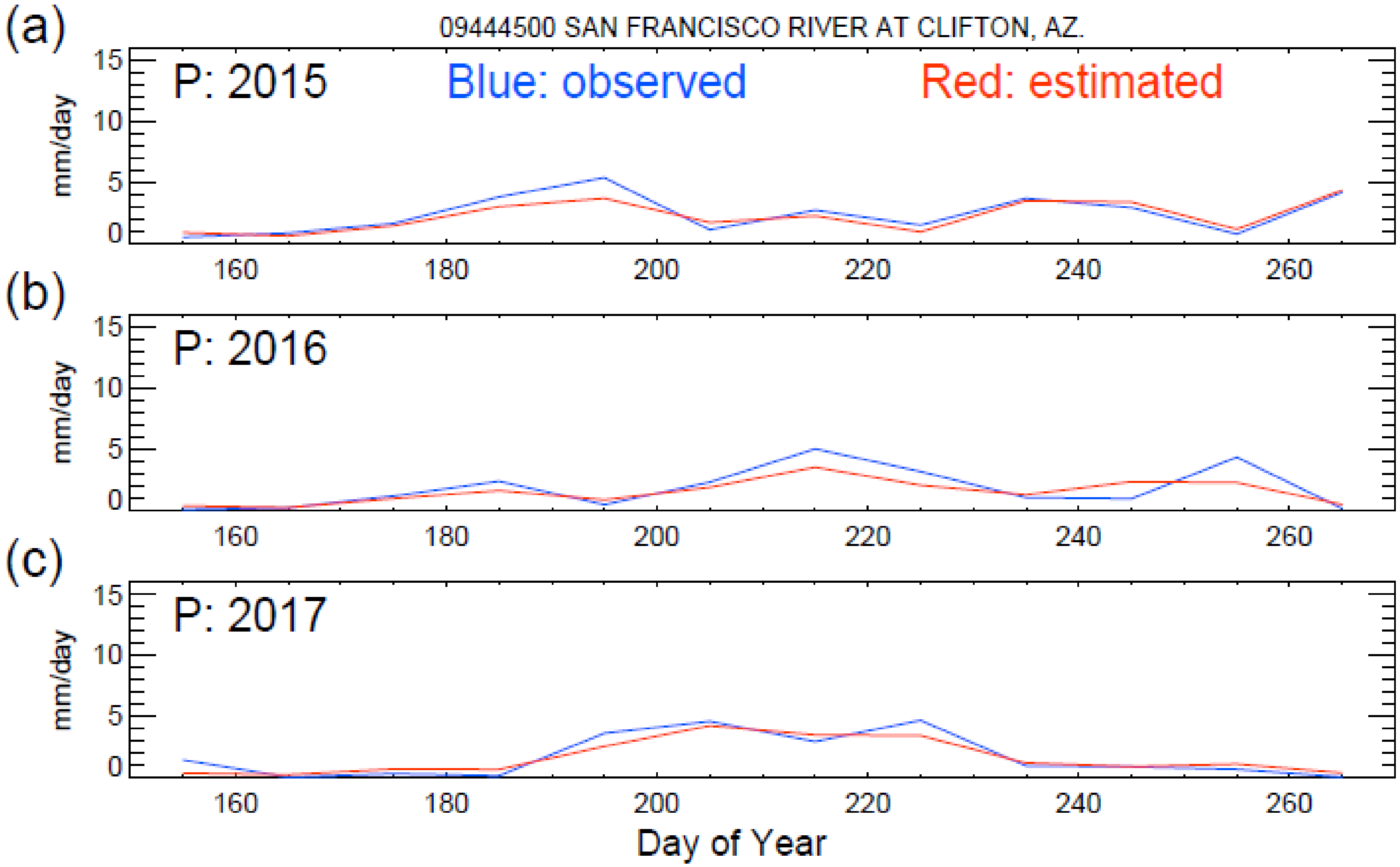
Updated approach:
This term replaced with calibrated loss function

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

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

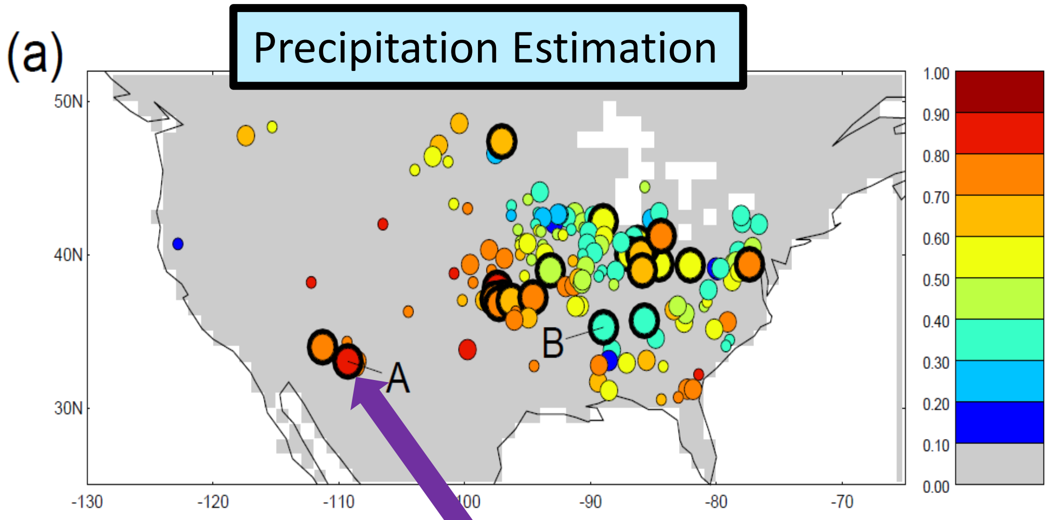
W_{ave} = average of the two consecutive retrievals

Some results!   (One of the better estimations):



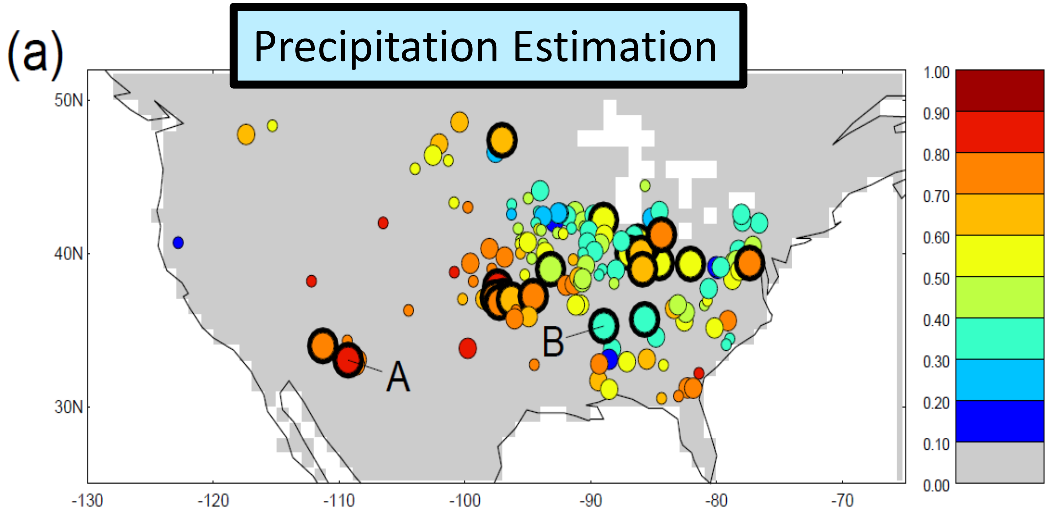
We can characterize the agreement in these time series with the square of the correlation coefficient, r^2.

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

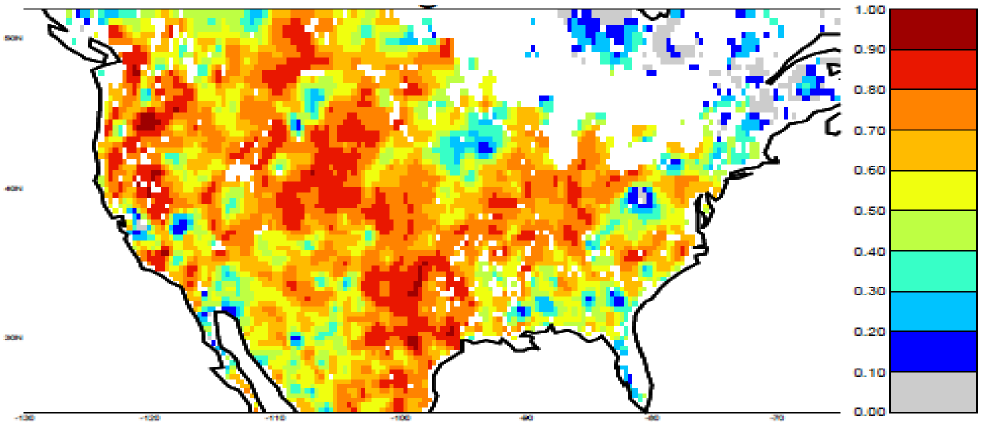


Time series on previous page
were shown for this basin.

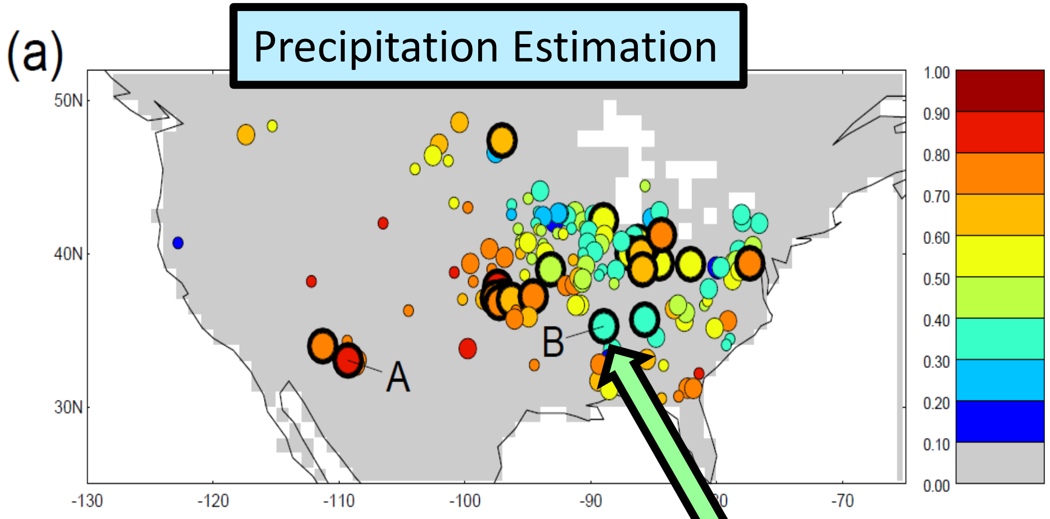
Basin level skill scores (time series of 10-day precipitation totals: r^2 vs observations)



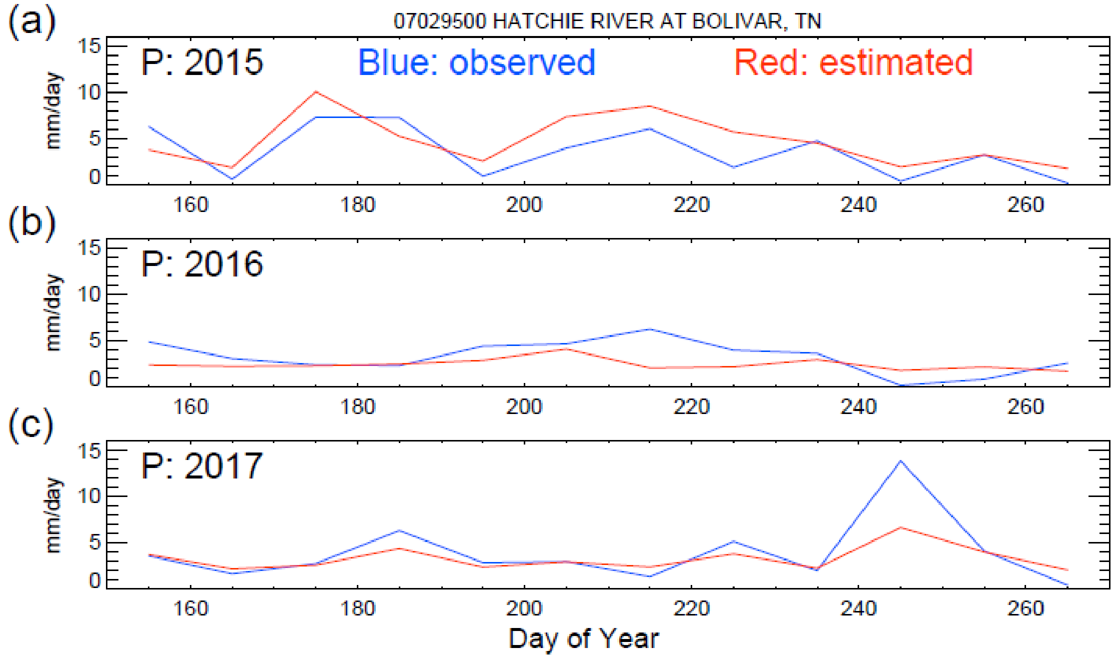
Aside: extending this analysis across the US, beyond "basins", indicates high skill throughout the west.



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

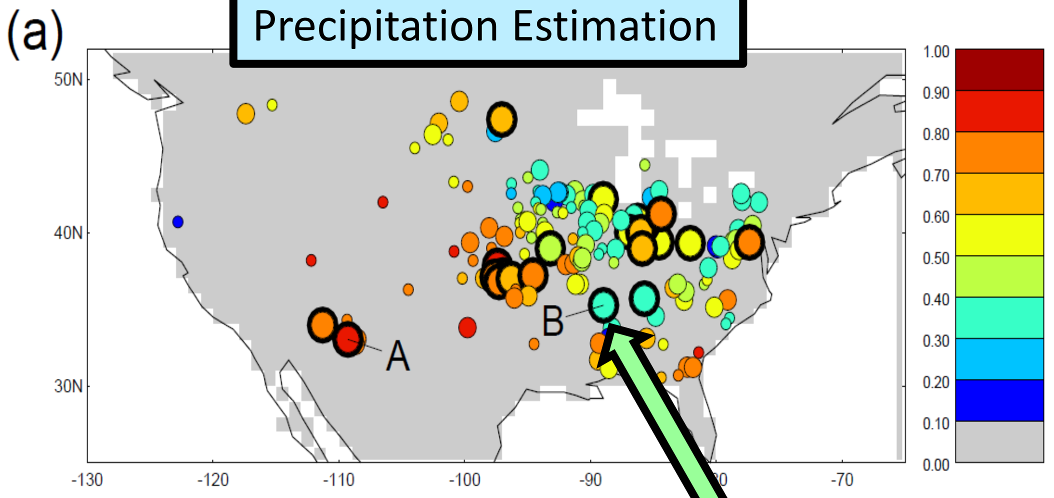


The algorithm works relatively poorly in this basin. Still, there is some valid information in the estimates



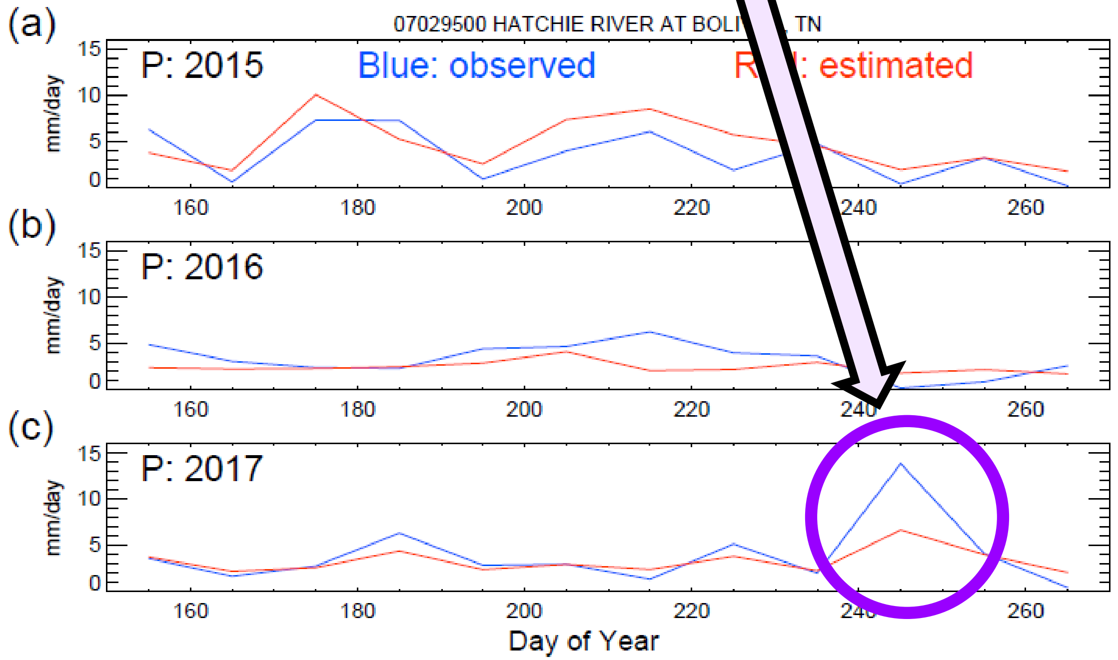
Basin level skill scores
(time series of 10-day precipitation totals: r^2 vs observations)

Precipitation Estimation



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} :

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

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

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

Step 2: Streamflow Estimation

Apply multiple regression:

basin-scale P

basin-scale W

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

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:

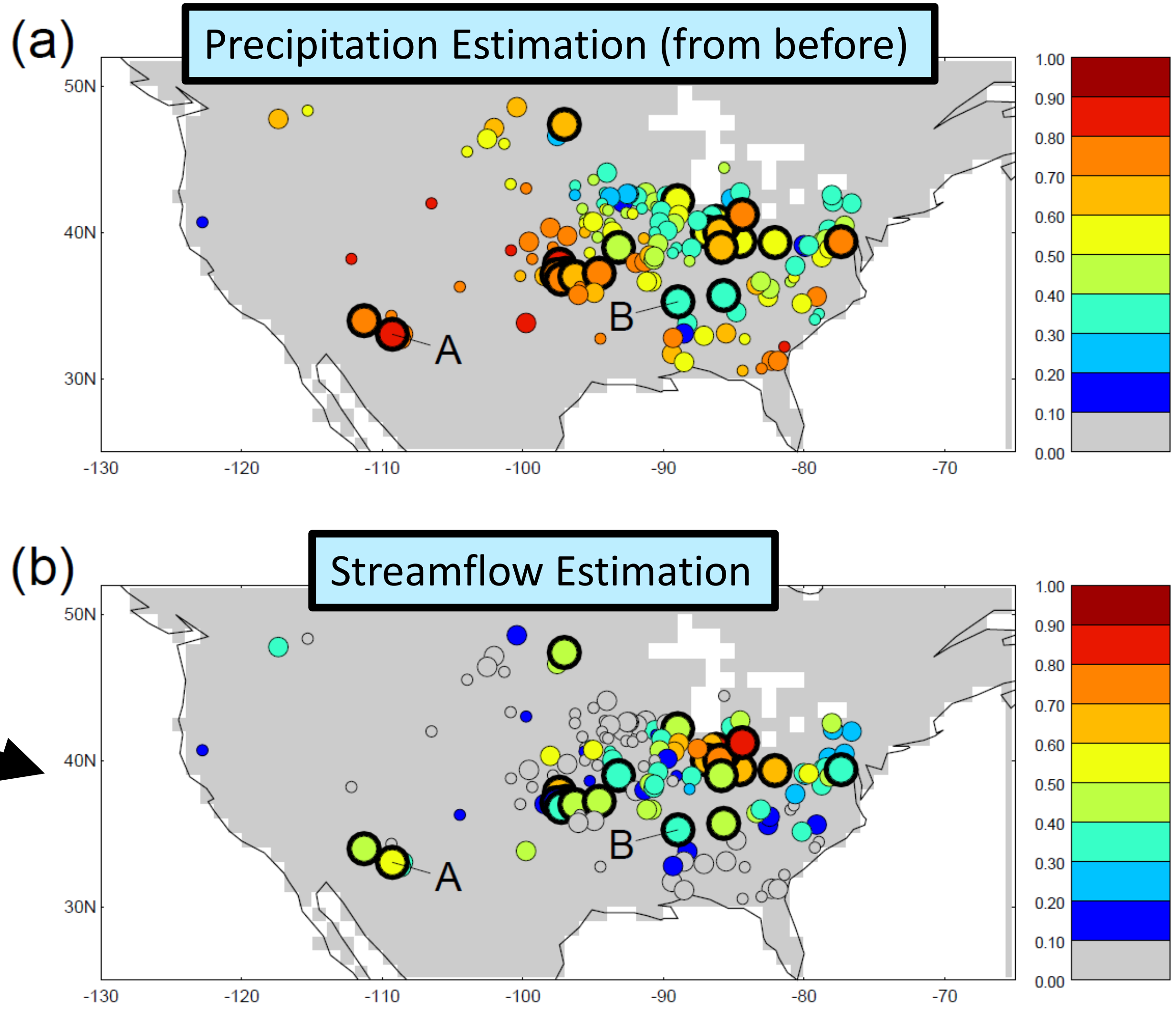
SMAP-based P estimate for 3rd year

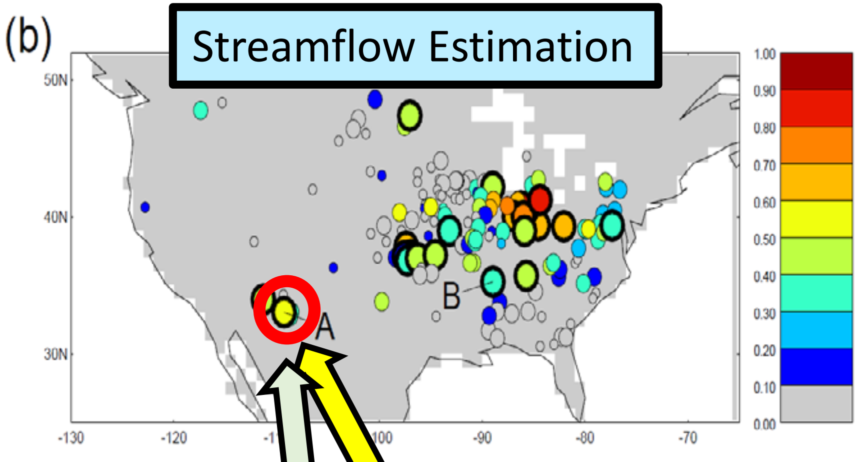
SMAP W in 3rd year

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

estimate of Q in 3rd year derived <u>solely</u> from SMAP data

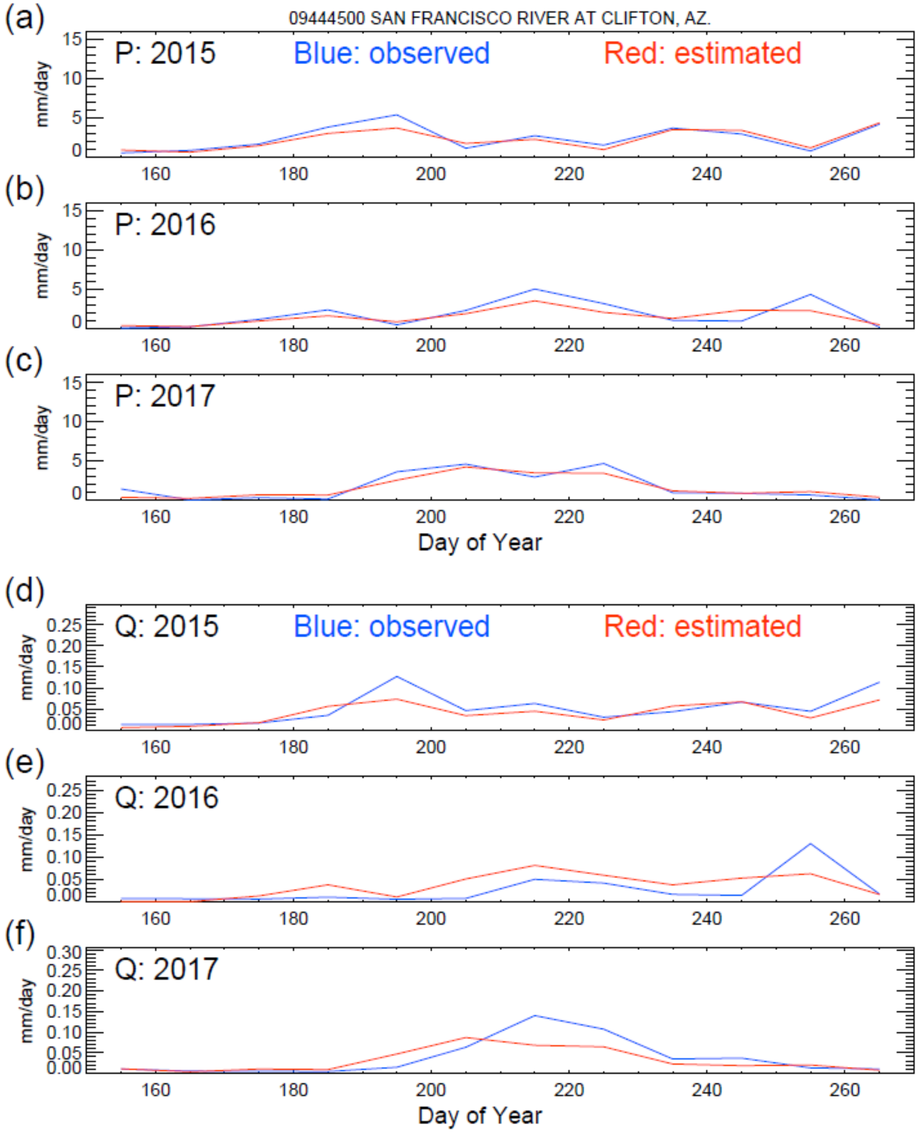
Estimation skill (time series of 10-day streamflow totals: r^2 vs observations)

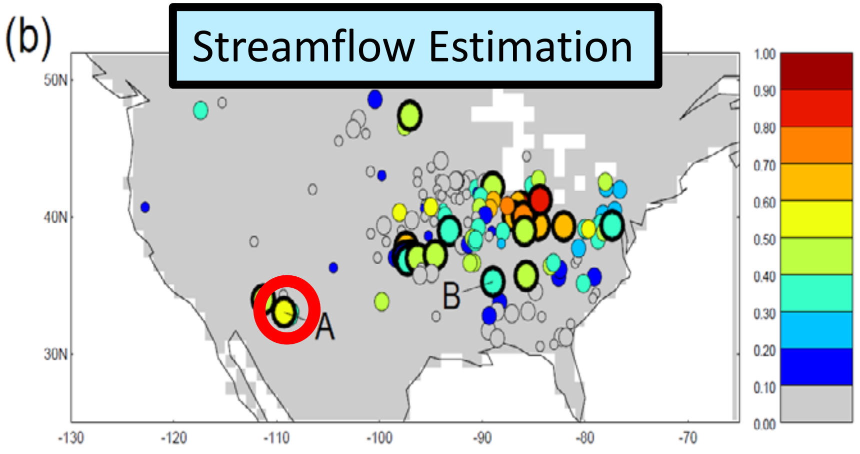


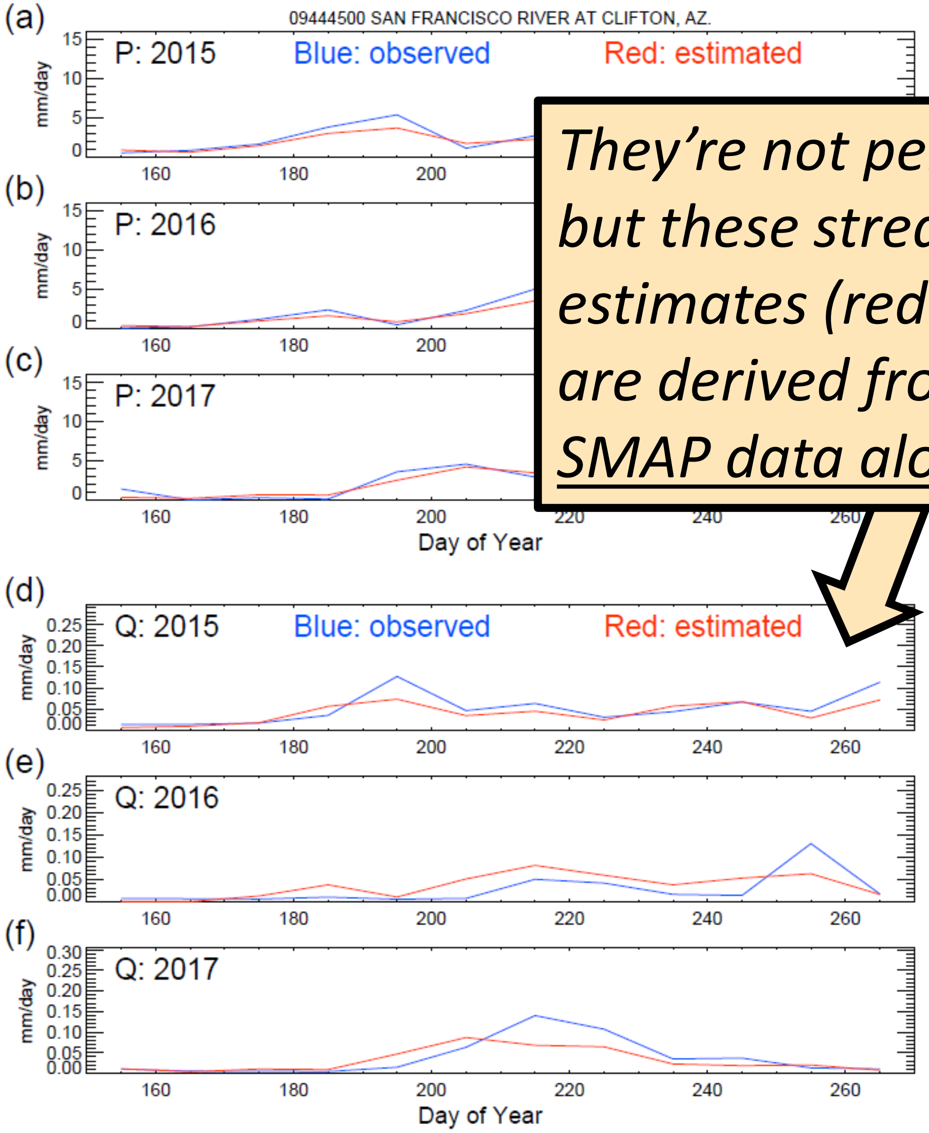


Precipitation estimates (from before)

Streamflow estimates capture some of the observed behavior.

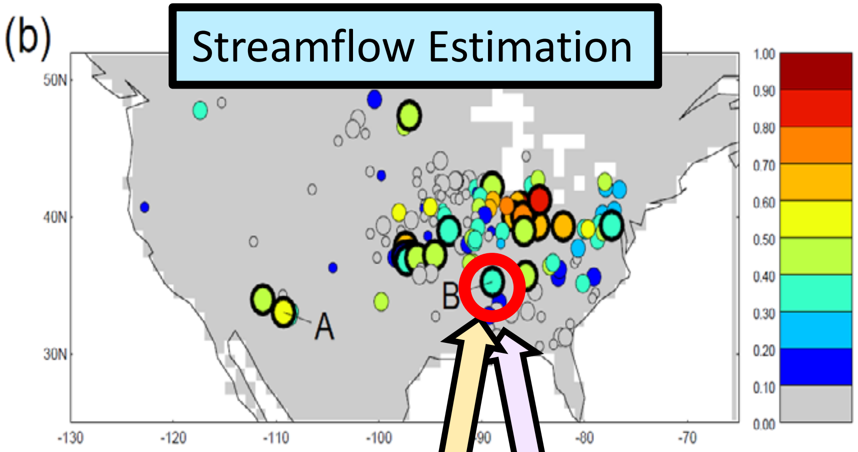


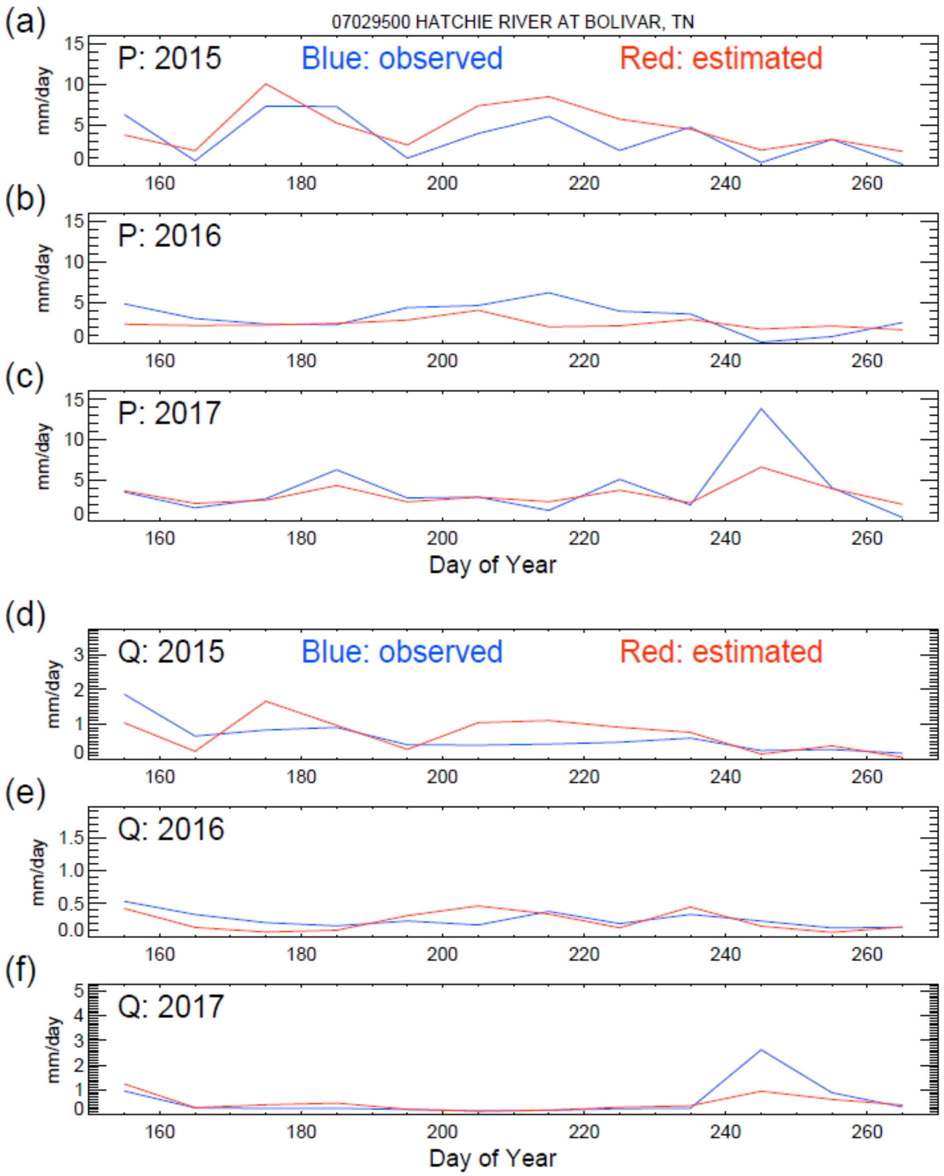




They're not perfect, but these streamflow estimates (red lines) are derived from <u>SMAP data alone</u>!

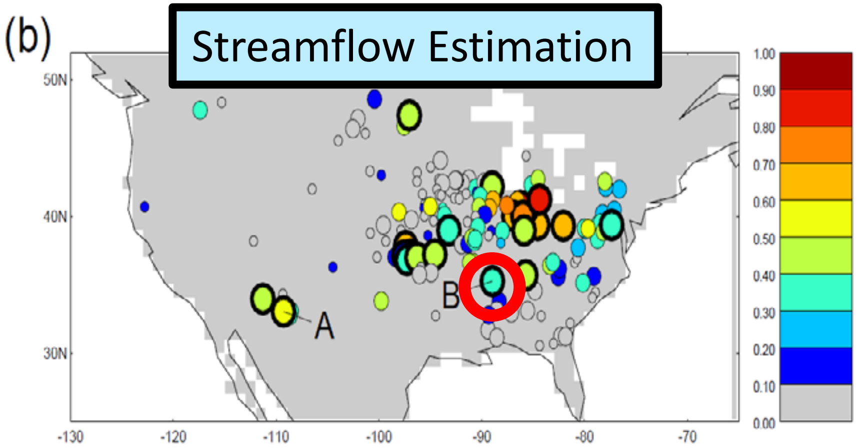
Streamflow Estimation

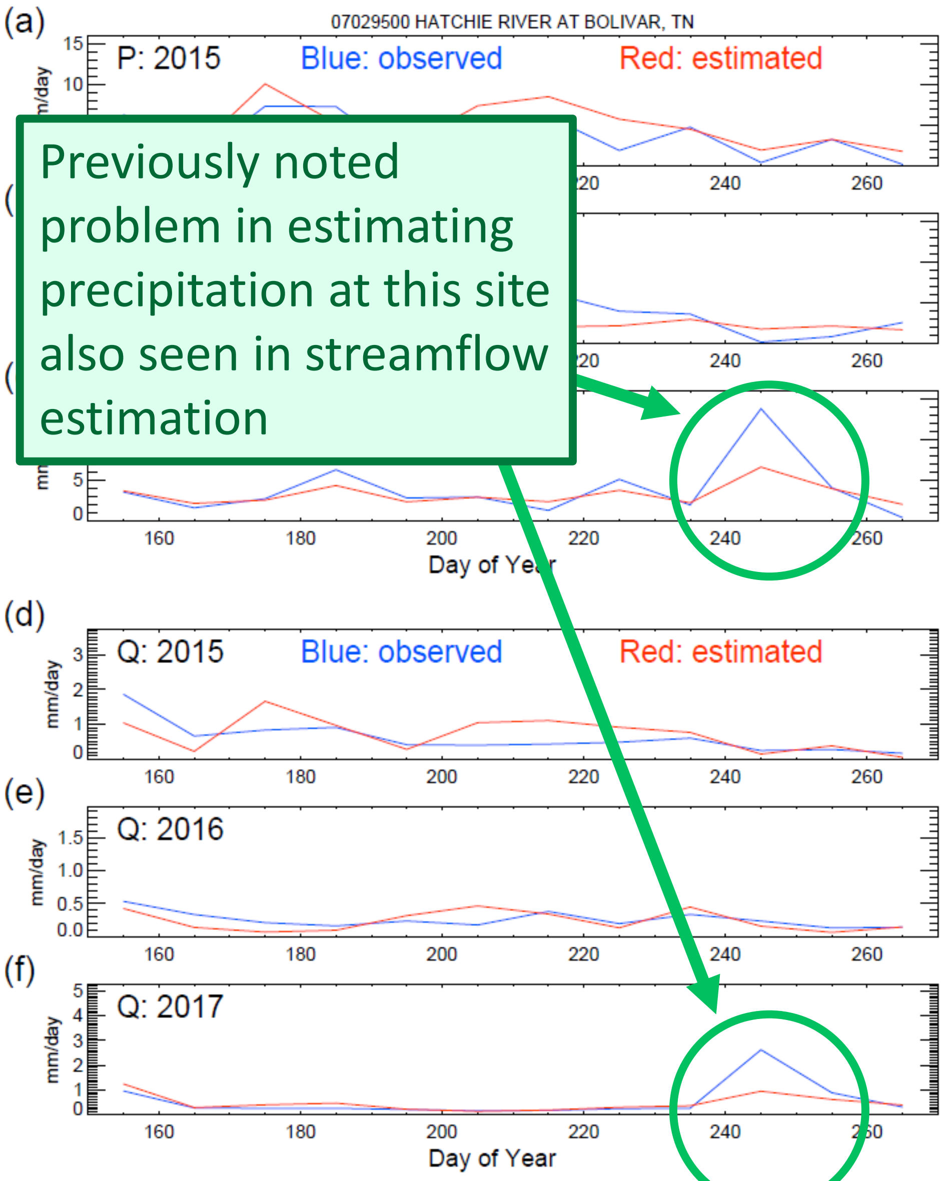




Precipitation estimates in “poorer basin” (from before)

Streamflow estimates capture less of the observed behavior.

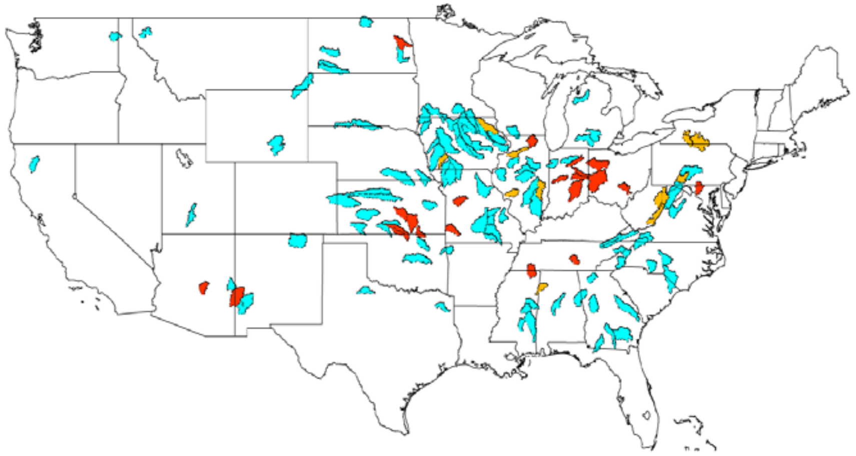


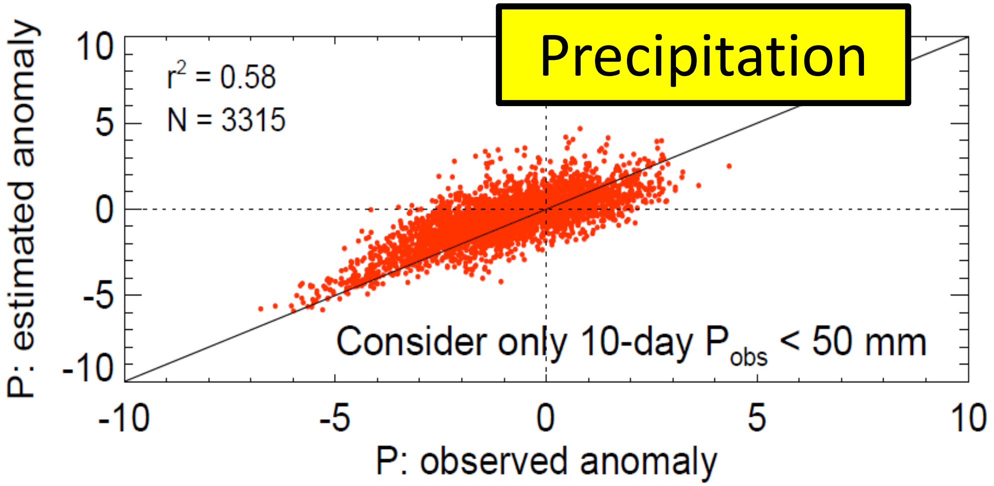


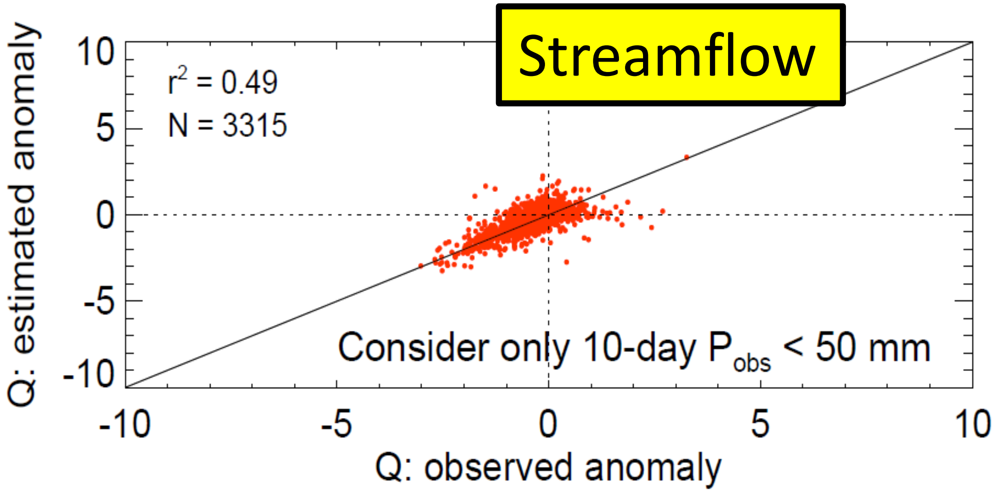
Previously noted problem in estimating precipitation at this site also seen in streamflow estimation

Inter-basin analysis

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





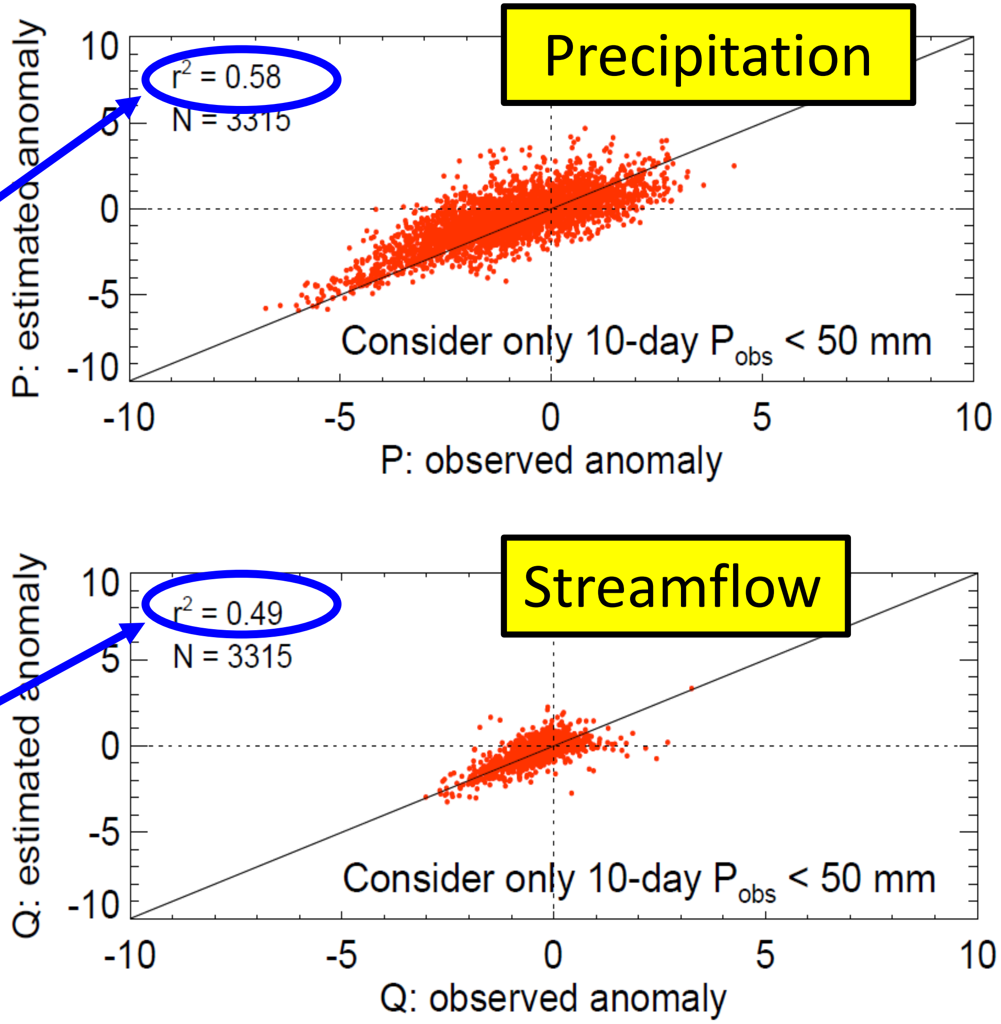


Inter-basin analysis

(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…

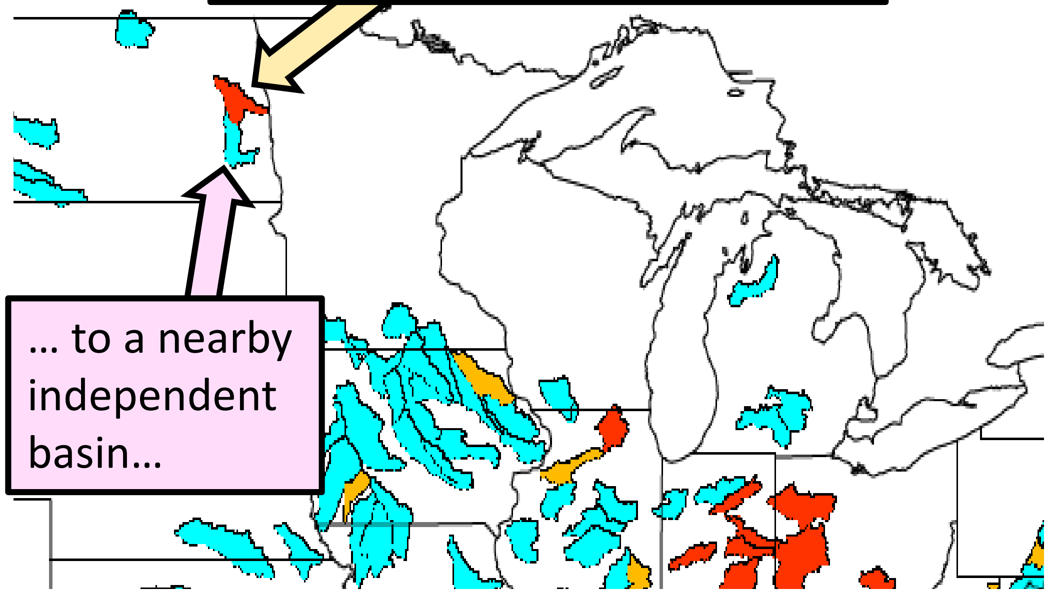
… and 49% of the streamflow variance



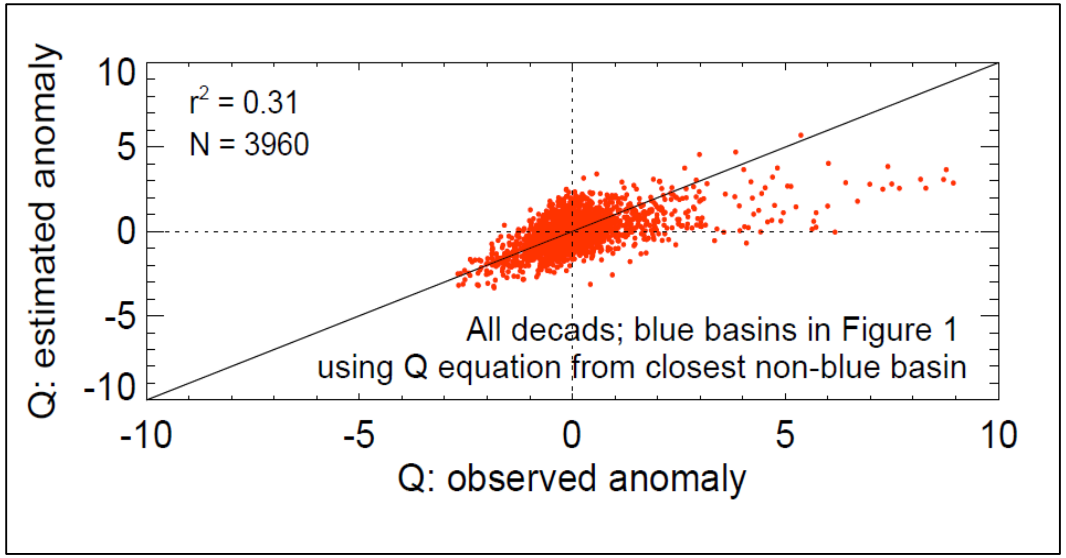
Transferability (streamflow)

Applying streamflow equations calibrated in one basin…

… to a nearby independent basin…



… 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 <u>complementary</u> information; SMAP data indeed contain a wealth of hydrological info.

3. Estimating rainfall and streamflow with SMAP data.

   ➪ SMAP data <u>do</u> contain information on large-scale water balance quantities!