Land Data Assimilation and the (Coordinated) National Soil Moisture Network

Wade Crow
USDA ARS Hydrology and Remote Sensing Laboratory

Alexander Gruber, Wouter Dorigo
TU-Wien, Department of Geodesy and Geo-information

USDA-ARS HRSL
Hydrology & Remote Sensing Lab
Beltsville, Maryland, USA

MOISST/NSMN Workshop, Lincoln, NE June 2018

TU WIEN GEO

Background

Three sources of large-scale soil moisture information:

1) Land surface modeling
2) Remote sensing (RS) products
3) Ground-based soil moisture observations

Soil moisture data assimilation: Updating dynamic and continuous model state predictions ($d\boldsymbol{S}/dt$) using sporadic (in time and space) soil moisture observations (θ).

S = Profile soil moisture and temperature states within a land surface model

θ = Soil moisture retrievals from RS and/or ground observations

Motivation:

1) Provides a spatially and temporally continuous soil moisture analysis.
2) Random errors in analysis $\leq$ those found in underlying model/observations.
3) Provides a mathematical basis for updating unobserved states.

Background

Three sources of large-scale soil moisture information:

1) Land surface modeling
2) Remote sensing (RS) products
3) Ground-based soil moisture observations

Operational (RS + modeling) systems:

- SMAP Level 4 surface and root-zone soil moisture analysis
  NASA GMAO/NASA SMAP
  [SMAP/CLSM, Global, 9-km, hourly, 2-3 day latency, percentile product]

- H14/SM-DAS-2 root-zone soil moisture
  ECMWF/EUMETSAT
  [ASCAT/HTESSEL, Global, 25-km, daily, <12 hour latency]

- NASA GSFC/USDA ARS/USDA FAS root-zone product
  [SMAP/SMOS/Palmer, Global, 25-km, daily, 2-3 day latency, anomaly product]

For all three products, ground observations are withheld for validation...

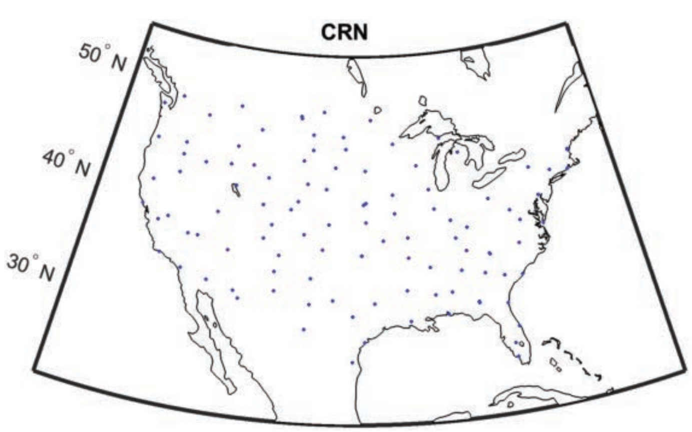
Background

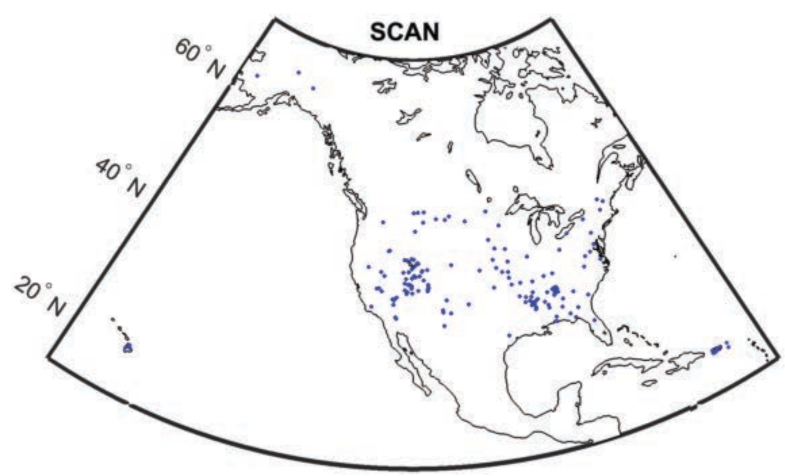
Three sources of large-scale soil moisture information:

1) Land surface modeling
2) Remote sensing (RS) products
3) Ground-based soil moisture observations

One possible conception of the USMN is a data assimilation system designed to simultaneously assimilate *both RS and Ground-based soil moisture.*

What new statistical information is required to add ground-based observations into state-of-the-art land data assimilation systems?

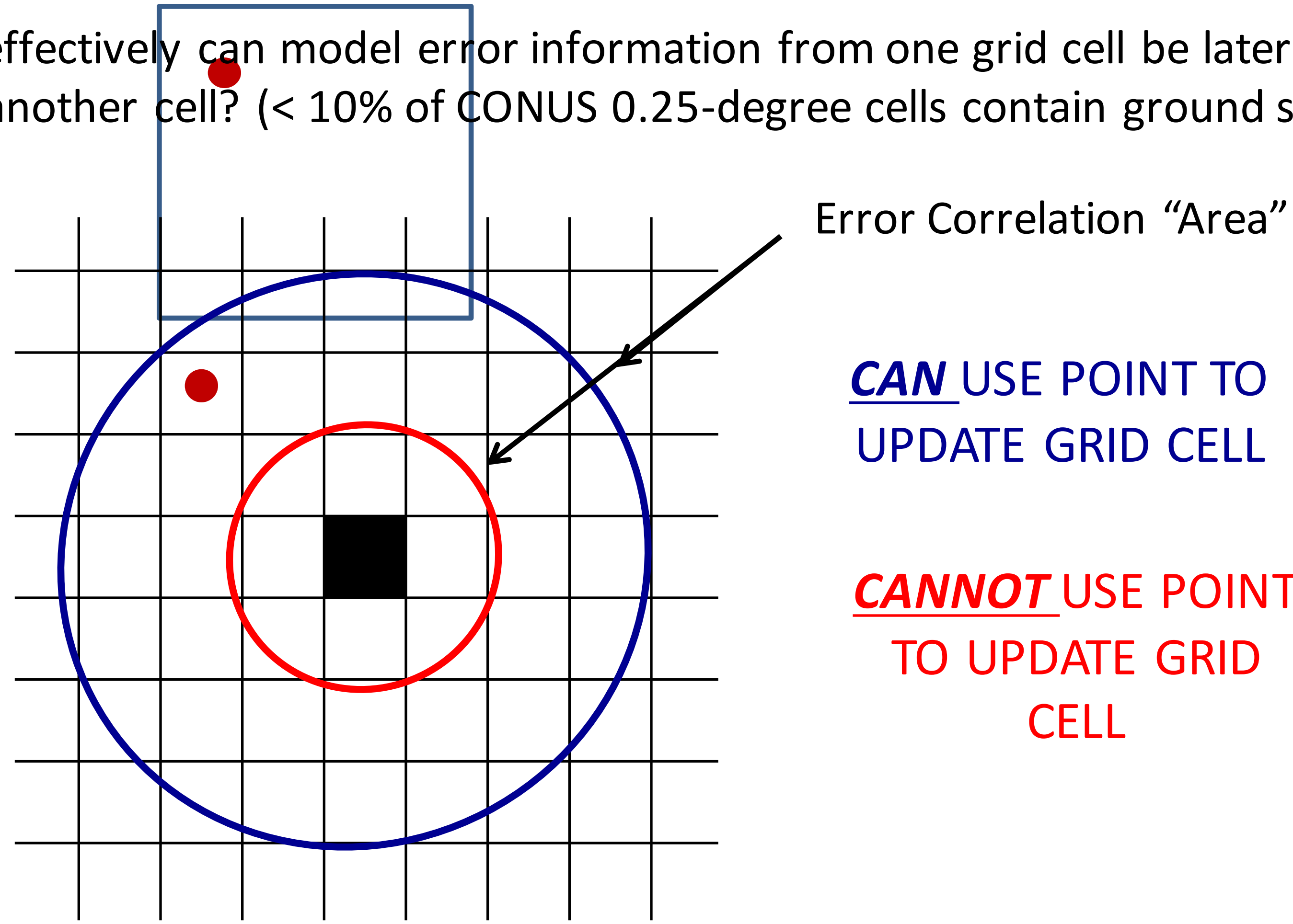




Two key issues for the assimilation of point-scale observations:

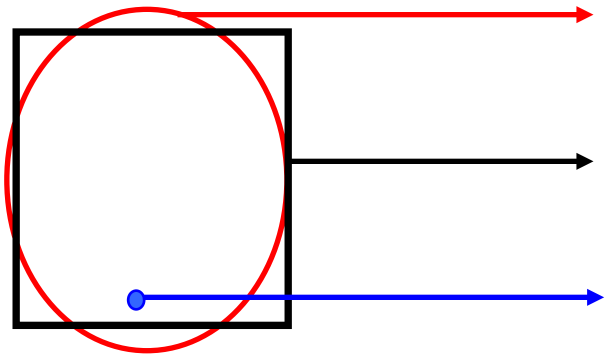
Issue #1: How well does a point-scale observation capture the grid-scale mean?

Issue #2: How effectively can model error information from one grid cell be laterally propagated to another cell? (< 10% of CONUS 0.25-degree cells contain ground sites)?



Application of Triple Collocation

1) Obtain three independent (and uncertain) estimates of *footprint-scale* soil moisture:



θ_{RS} Remote Sensing (RS)

θ_{LSM} Land Surface Model (LSM)

θ_{G} Sparse Ground Observation (G)

2) Assume **anomaly** products can be modeled as:

$$\theta_{RS} = \alpha_{RS}\theta_{True} + \varepsilon_{RS}$$
$$\theta_{LSM} = \alpha_{LSM}\theta_{True} + \varepsilon_{LSM}$$
$$\theta_{G} = \alpha_{G}\theta_{True} + \varepsilon_{G}$$

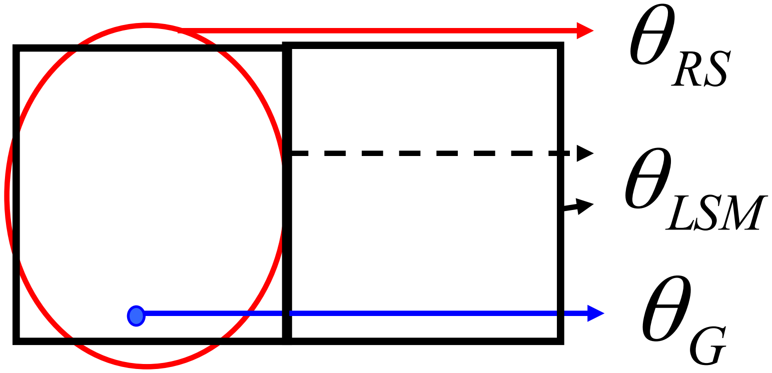
3) Triple collocation can provide:

a) Ratios: α_{LSM}/α_{RS} , α_{LSM}/α_{G} , and α_{G}/α_{RS}

b) Variances of: ε_{RS}, ε_{LSM}, and ε_{G}

Application of Extended Triple Collocation

1) Obtain three independent (and uncertain) estimates of ***footprint-scale*** soil moisture:



Remote Sensing (RS)

Land Surface Model (LSM_1 and LSM_2)

Sparse Ground Observation (G)

2) Assume **anomaly** products can be modeled as:

$$\theta_{RS} = \alpha_{RS}\theta_{True} + \varepsilon_{RS}$$
$$\theta_{LSM} = \alpha_{LSM}\theta_{True} + \varepsilon_{LSM}$$
$$\theta_{G} = \alpha_{G}\theta_{True} + \varepsilon_{G}$$

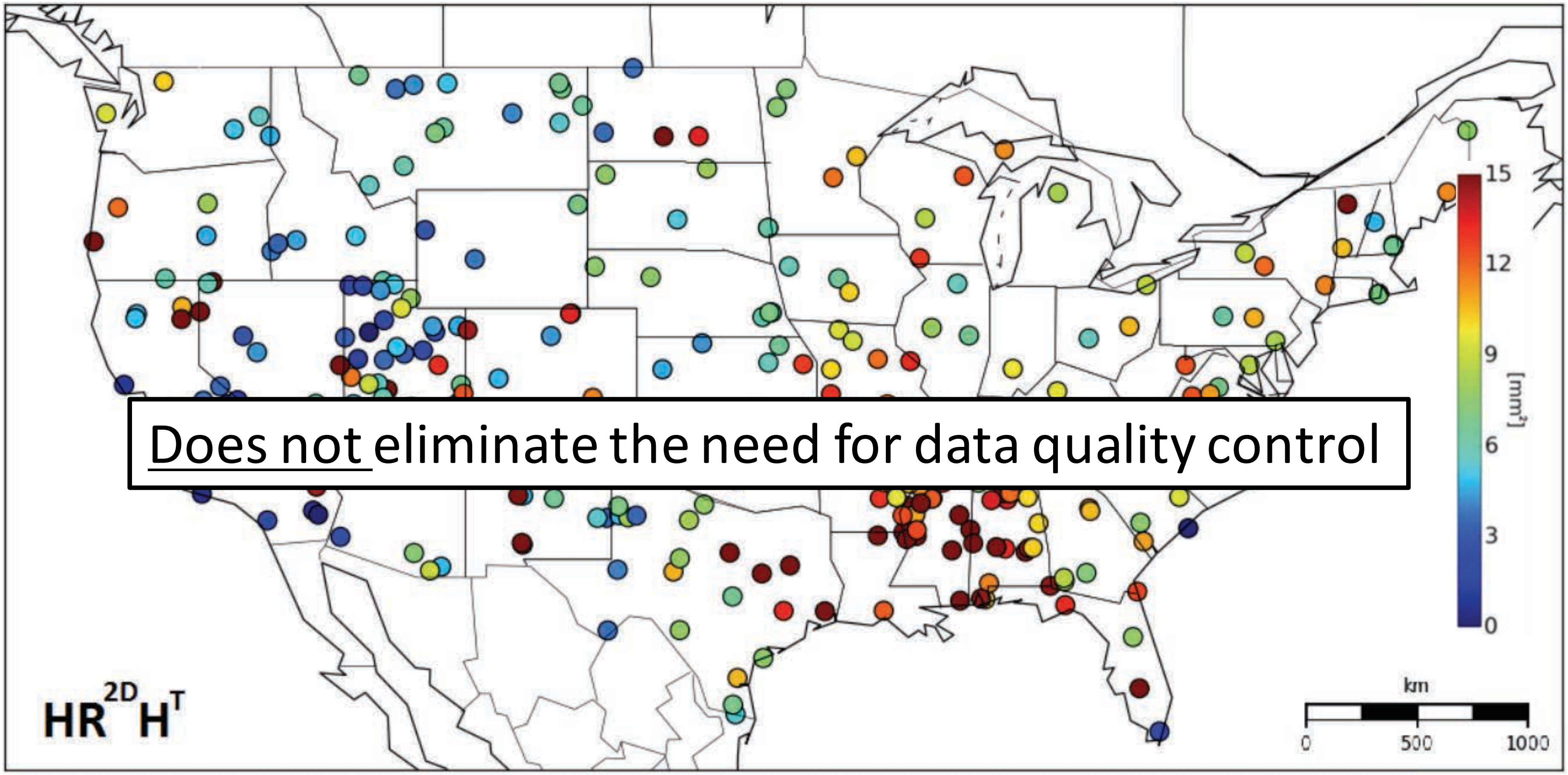
3) Extended triple collocation can provide:

a) Ratios: α_{LSM}/α_{RS} , α_{LSM}/α_{G} , and α_{G}/α_{RS}

b) Variances of: ε_{RS}, ε_{LSM}, and ε_{G} **plus Cov(ε_{LSM1}, ε_{LSM2})**

Issue #1: How well does a point-scale observation capture the grid-scale mean?

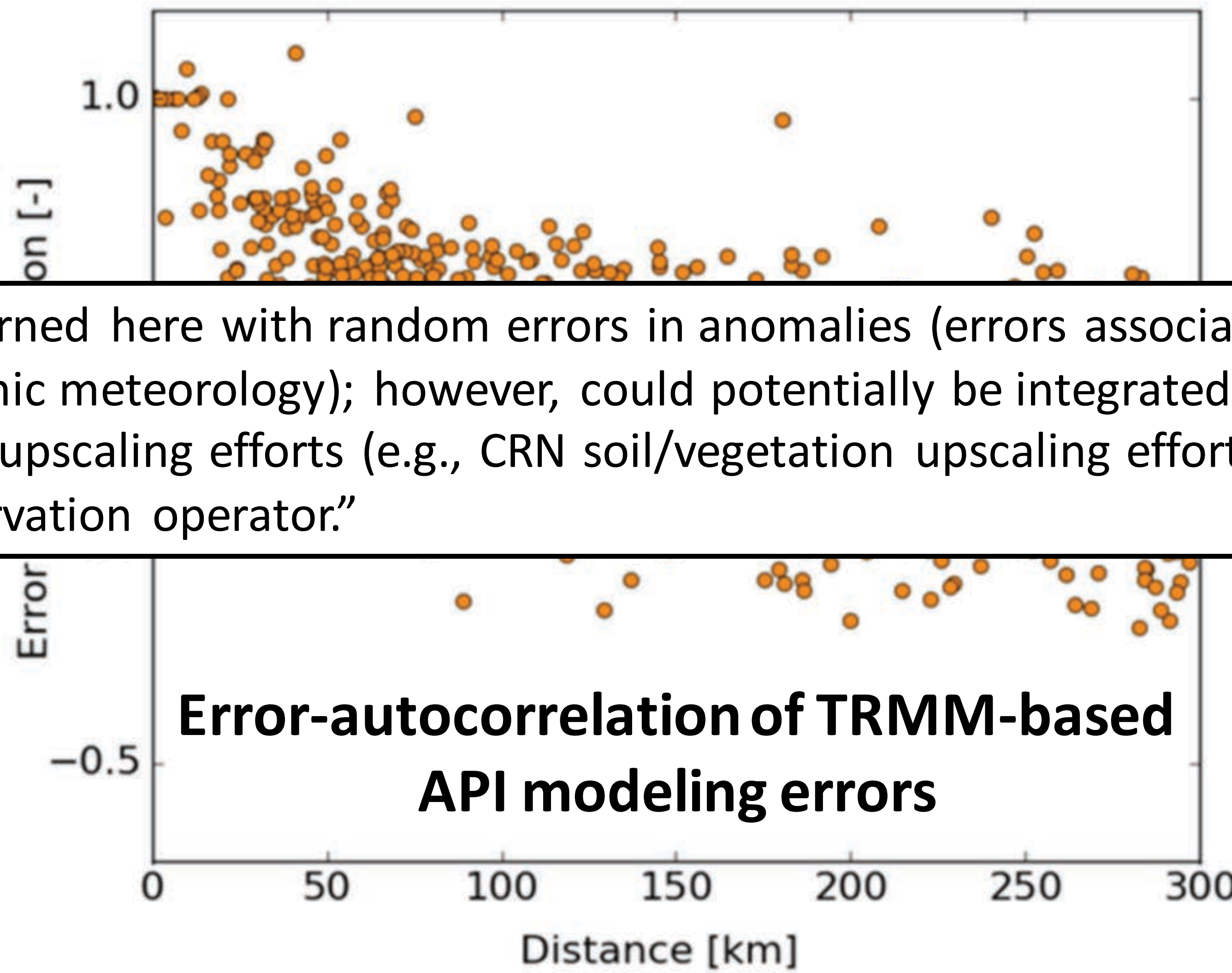
Error variance for point-to-grid (0.25°) upscaling of CRN and SCAN



Does not eliminate the need for data quality control

2008-2015 TC[ASCAT, SCAN+CRN, API]

Issue #2: Can model error information from be laterally propagated?

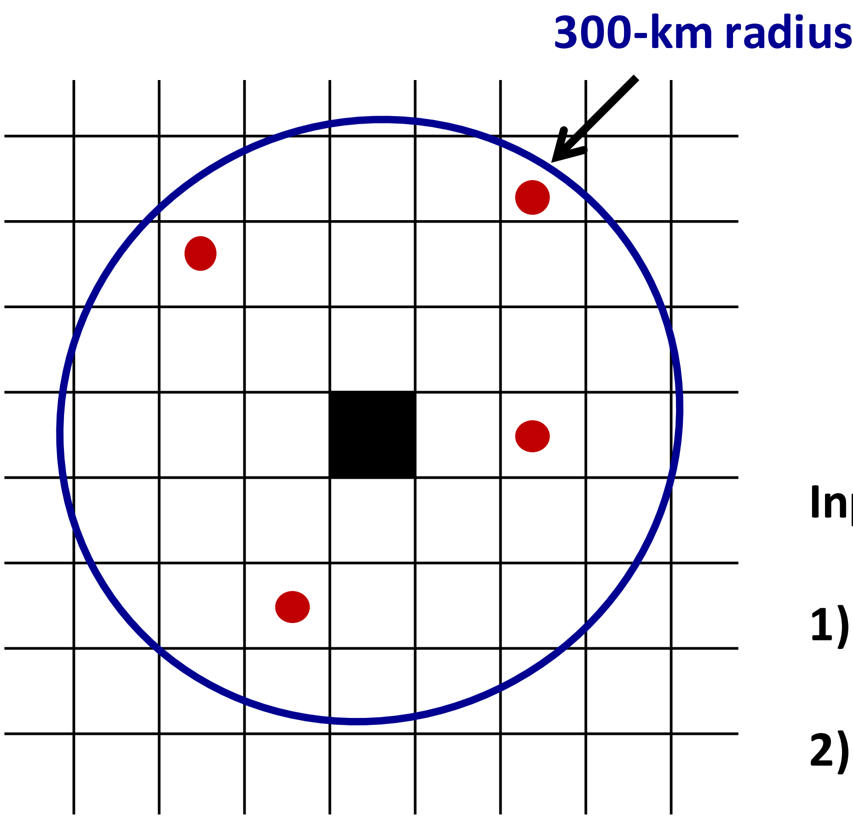


Concerned here with random errors in anomalies (errors associated with dynamic meteorology); however, could potentially be integrated with other upscaling efforts (e.g., CRN soil/vegetation upscaling effort) as an "observation operator."

Error-autocorrelation of TRMM-based API modeling errors

2008-2015 ETC[ASCAT, SCAN+CRN, API]

Data Assimilation Results



For each 0.25-degree grid:

Obs. Space = *N* observations within 300-km radius.

State Space = Grids with obs. + center grid (*N*+1)

Inputs that are needed for this system:

1) *R* = (*N* x *N***)** covariance matrix for observation errors.

2) *Q* = (*N*+1 x *N*+1**)** covariance matrix for LSM noise.

3) *H* = Transform between observations and model

Gruber, A., Crow, W.T., and Dorigo, W. Assimilation of spatially sparse in situ soil moisture networks into a continuous model domain. *Water Resources Research*. 54:1353-1367. 10.1002/2017WR021277. 2018.

What might a USMN DA system look like?

- Based on a state-of-the-art land surface model
e.g. Noah-MP (National Water Model) or CLSM (SMAP L4 product).

- 1/8-degree, hourly, profile soil moisture, 2-3 day latency.

- RS products: Assimilate 9-km SMAP L3 passive-only, enhanced product. Fall back on EUMETSAT ASCAT or JAXA AMSR2 products (less accurate but stronger continuity commitment).

- Ground observations: Assimilate all ground network observations with < 1 day data latency.

- Reserve all other ground-based soil moisture observations (citizen science inputs?) for retrospective validation and calibration purposes (contextualize information with climatology information).

- In addition, can run in retrospective/re-analysis mode. NLDAS-2 (North American met. Forcing) and ESA CCI (remote sensing) both go back to 1979.

Resource Requirements

Operational DA systems are not easy to construct and require on-going support. Based experience with existing systems:

~2 FTE for 3 years to develop

Development period would need to include major calibration activities. DA only resolve random errors, does not correct systematic errors in products.

~1 FT for every year of on-going operation

Need to maintain inputs into the system (e.g., deal with data input disruptions and data version changes).

Generate long-term, re-analysis (1979 onward?).

Computational aspects: Not overwhelming, likely ~20% CPU time of the current SMAP L4 system (back-of-envelop calculation). A 30-year re-analysis would likely take days to weeks.

Summary

A data assimilation system is one possible conception of what the NSMN might entail.

1) Spatial characteristics of modeling errors appears conducive to the assimilation of sparse, ground-based soil moisture observations.

2) Data assimilation represents the most mature and efficient method for integrating multiple observations types (and dynamic model predictions) in an unified and continuous product (numerous example in the atmospheric and ocean sciences).

3) Costs are substantial, must (of course) be weighed against over high-priority activities. Less costly approaches may be good enough.

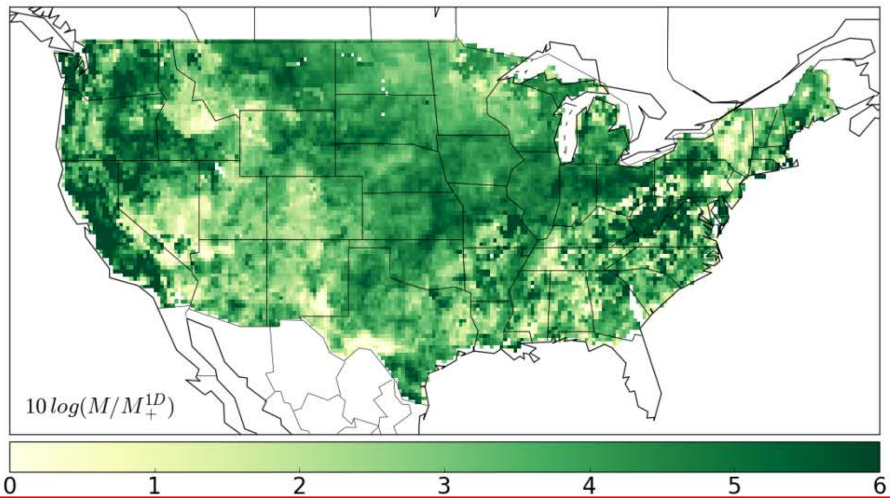
4) Have 1 FTE of visiting graduate student labor 12/18-12/19…happy to orientate that labor towards NSMS DA activities. Also happy to serve as on-going point-of-contact with the land DA community.

Thank you…

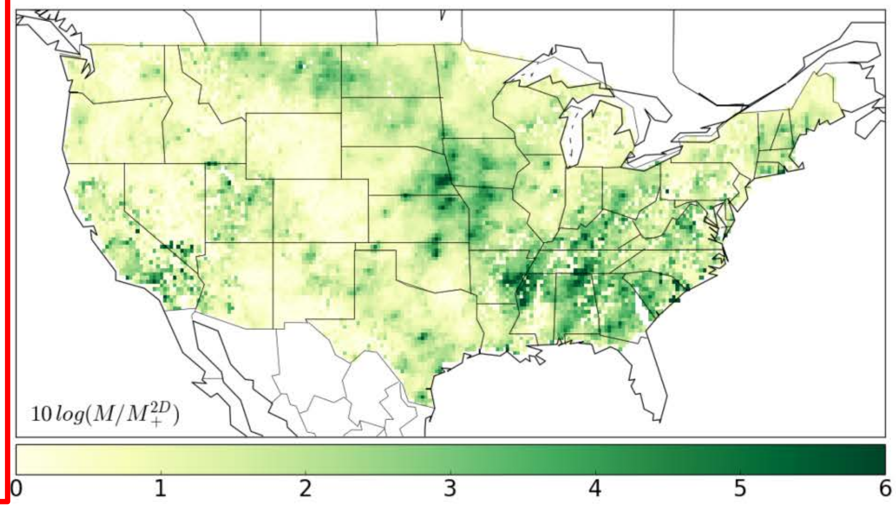
Gruber, A., Crow, W.T., and W. Dorigo. Assimilation of spatially sparse in situ soil moisture networks into a continuous model domain. *Water Resources Research*. 54:1353-1367. 10.1002/2017WR021277. 2018.

Surface Soil Moisture Data Assimilation Results

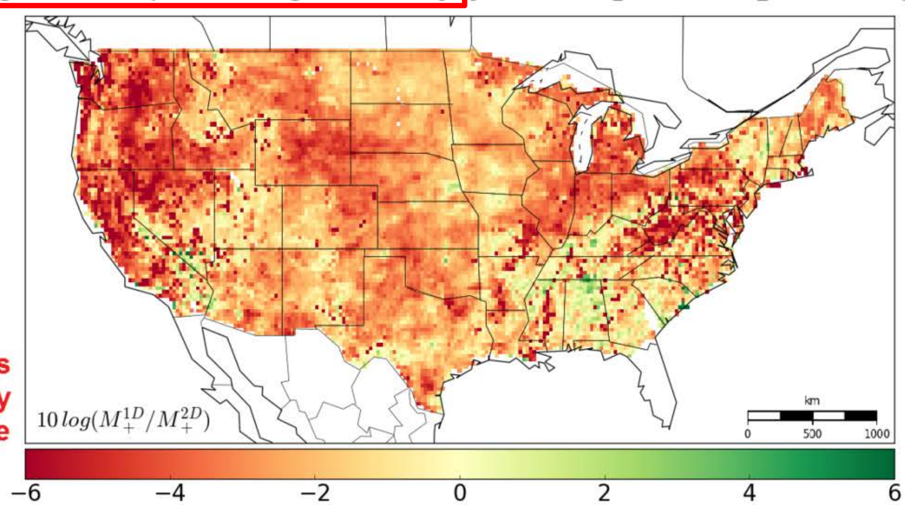
Reduction of error for 2D in situ DA



Reduction of error for 1D ASCAT DA



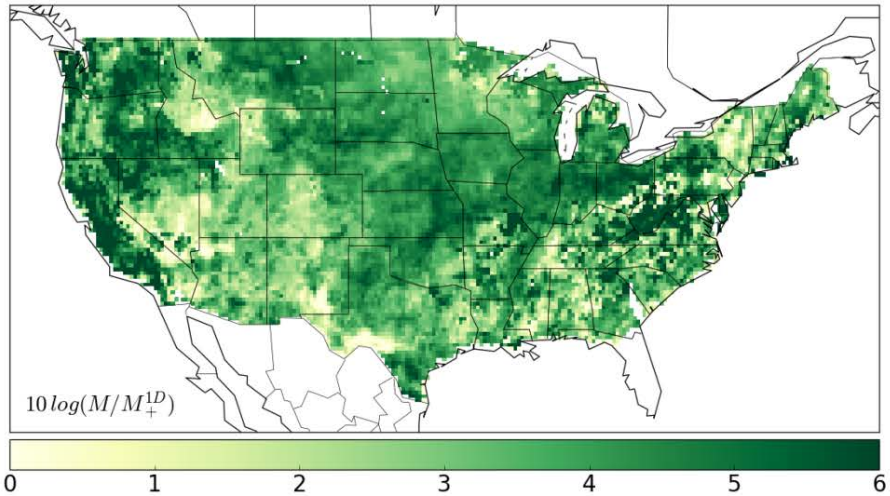
ASCAT is reducing uncertainty more



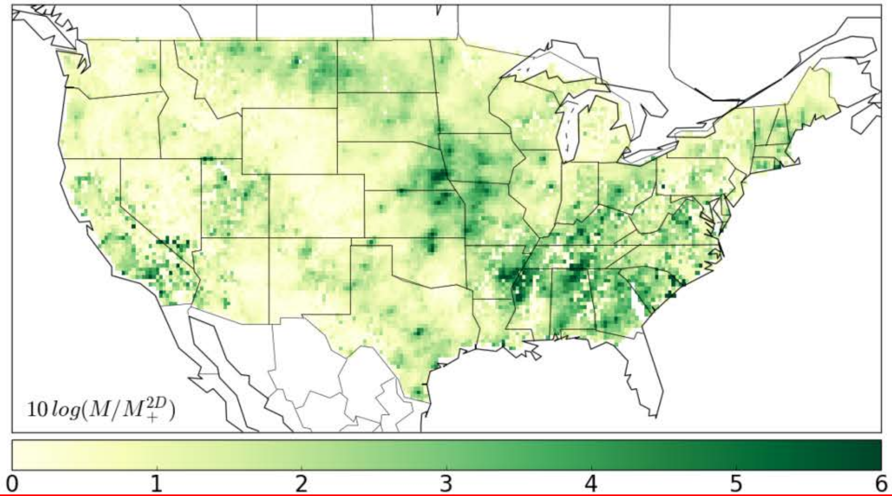
SCAN + CRN are reducing uncertainty more

Surface Soil Moisture Data Assimilation Results

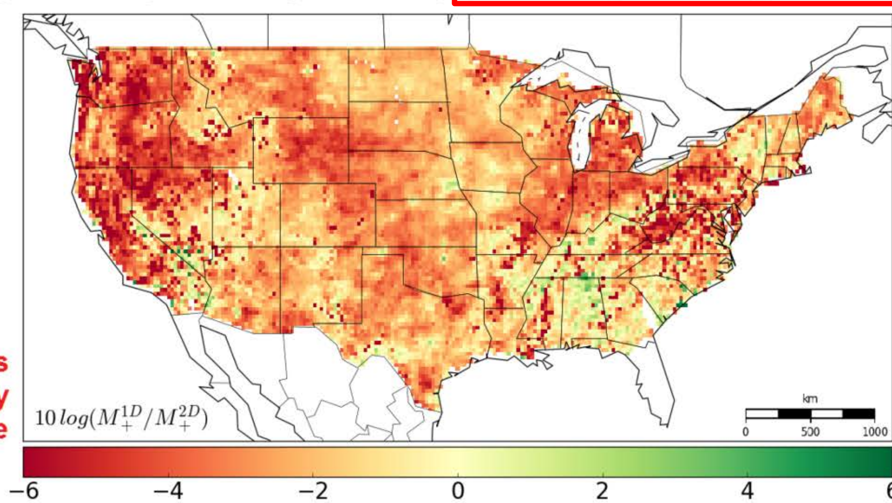
Reduction of error for 2D in situ DA



Reduction of error for 1D ASCAT DA



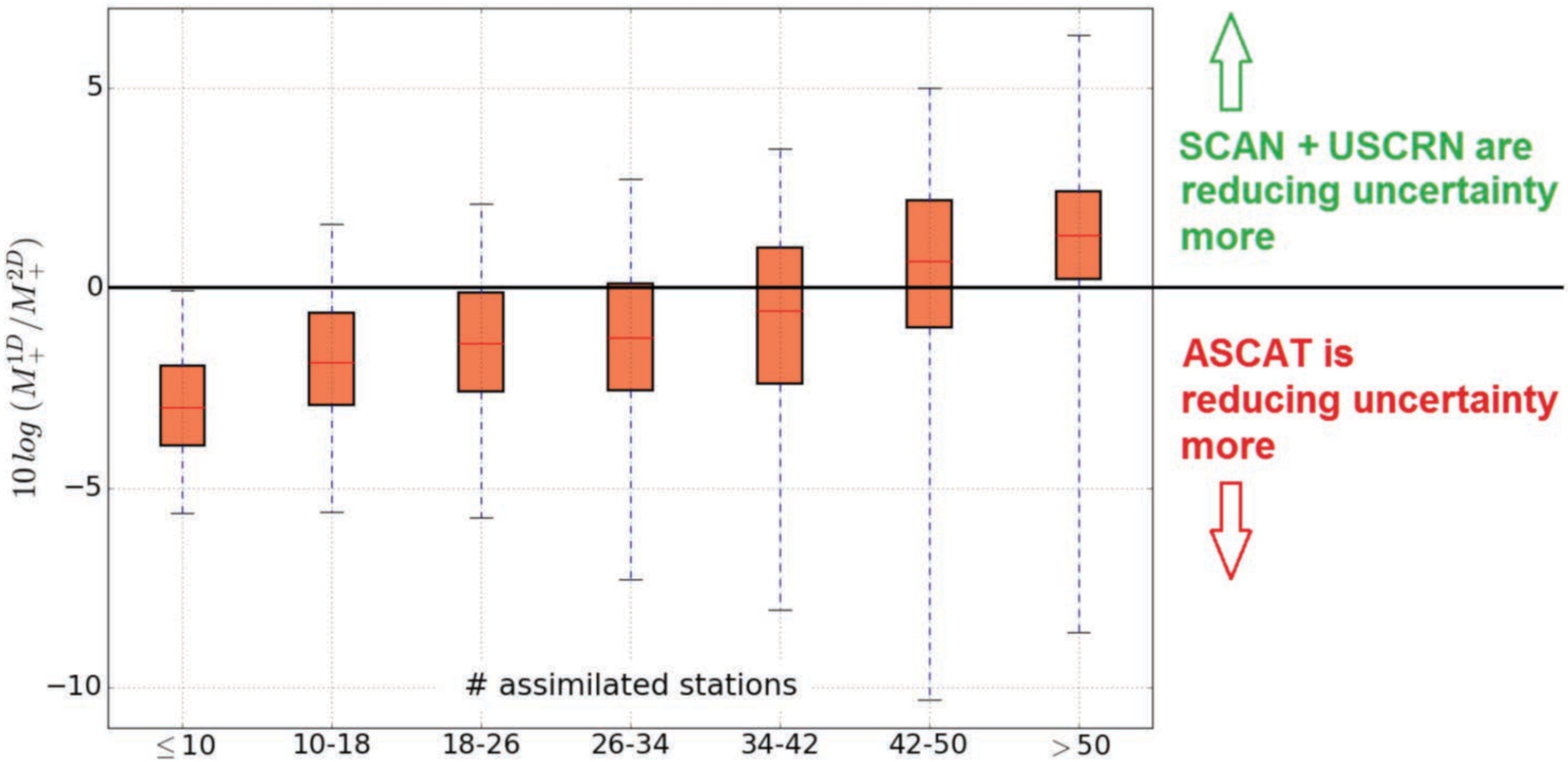
ASCAT is reducing uncertainty more



SCAN + CRN are reducing uncertainty more

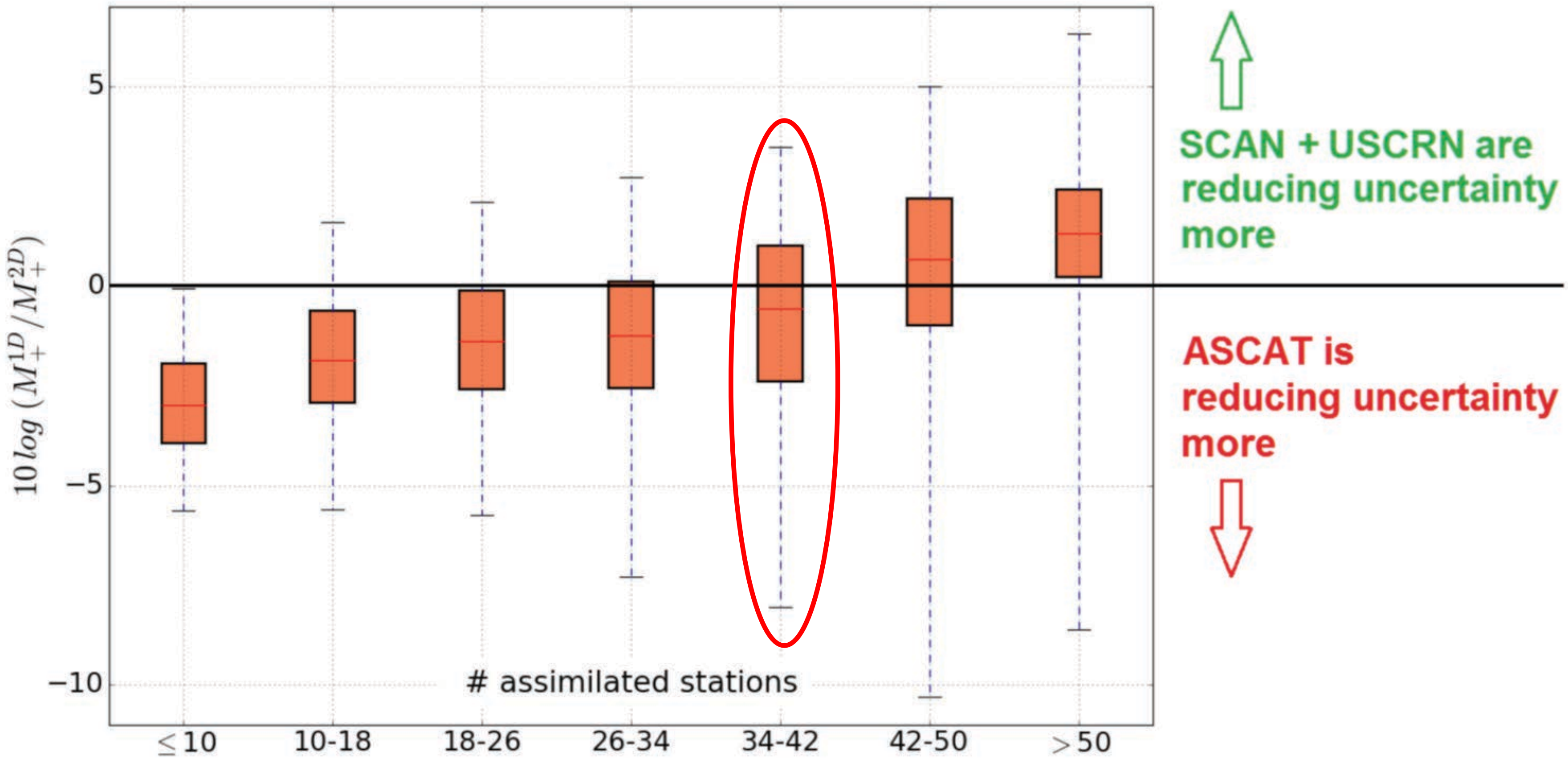
Relationship with Station Density

CONUS synthetic network analysis



Number of stations within 300-km assimilation zone

Relationship with Station Density



Number of stations within 300-km assimilation zone

~40 stations/300-km radius required to match ASCAT or ~ 1 station/84^2 km^2

OK Mesonet = ~1 station/35^2 km^2 SCAN +CRN = ~1 station/160^2 km^2

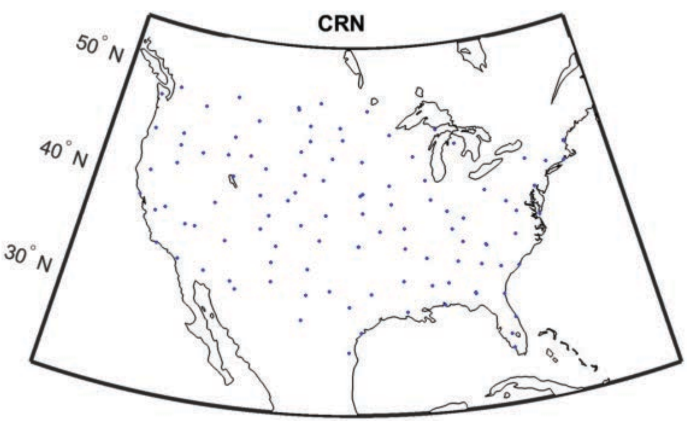
<u>Surface (0-5 cm) only…number may change for root-zone soil moisture</u>

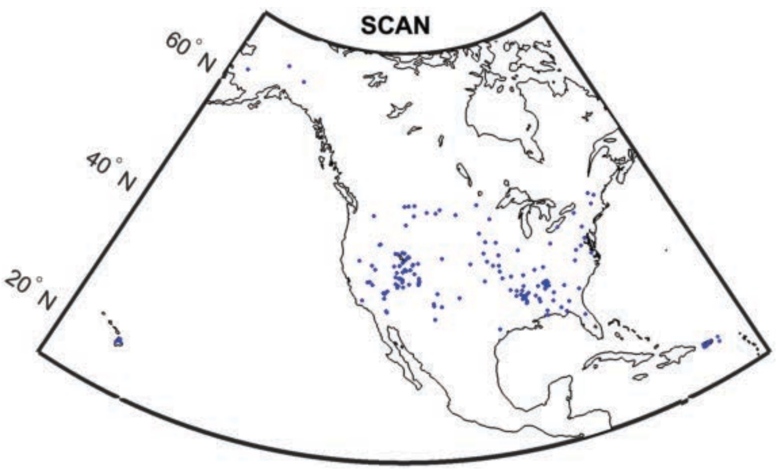
Background

• Recent interest in the development of an integrated North American soil moisture monitoring network.

• Optimal design for such a land data assimilation system capable of integrating:

1) Land surface modeling
2) Remote sensing (RS) products
3) Sparse ground-based soil moisture networks



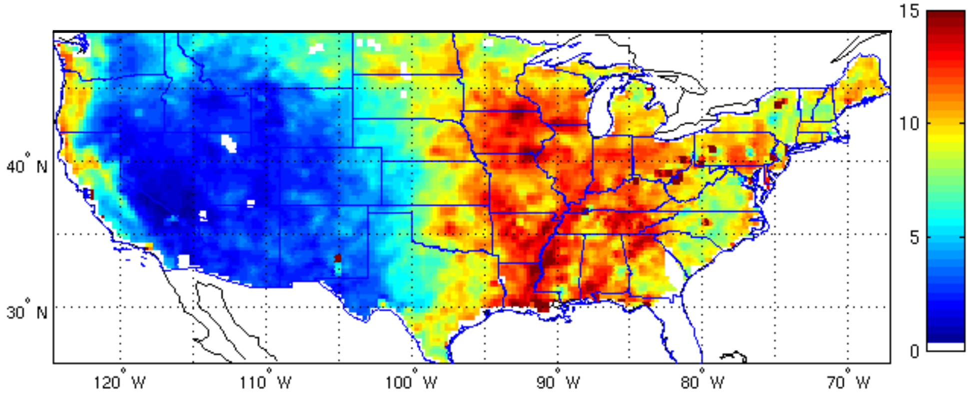


• Development of systems capable of integrating all three has been slow due (in part) to difficulty of statistically characterizing errors in each soil moisture product (especially spatial auto-correlation structure of errors).

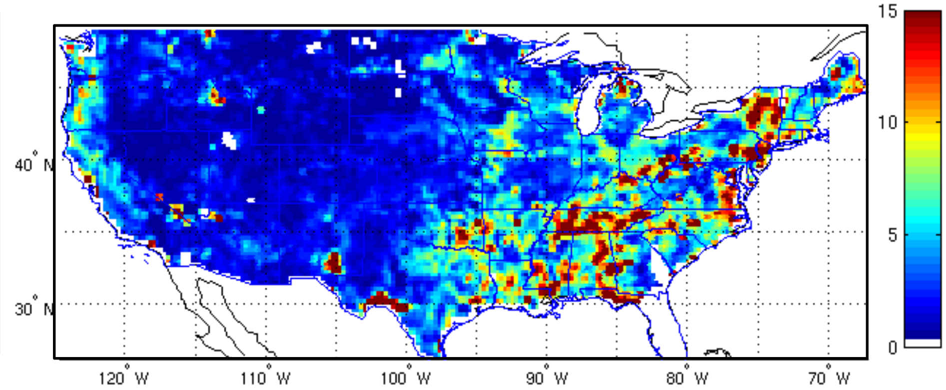
• As a result, we currently cannot provide useful advice regarding the relative value of RS retrievals versus ground-based observations for large-scale drought monitoring.

TC-based error products (KF inputs)

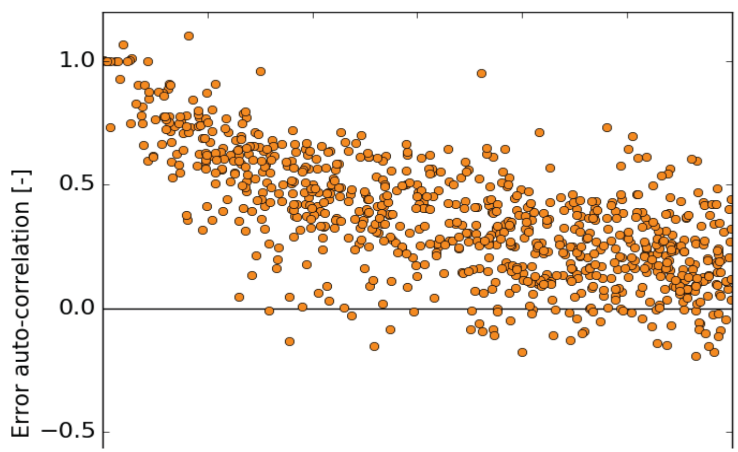
Sqrt(Q) (API/TRMM) [mm]



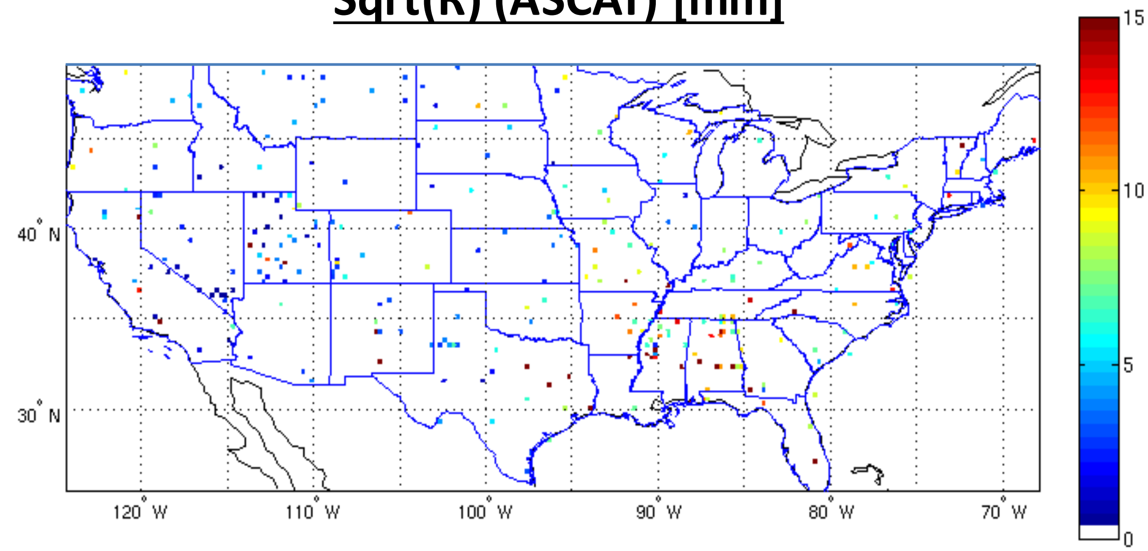
Sqrt(R) (ASCAT) [mm]

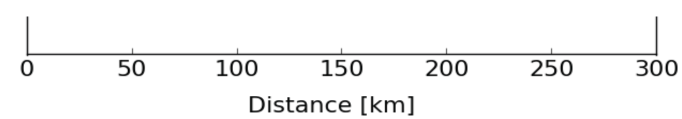


Spatial Error Autocorrelation (API/TRMM) [-]



Sqrt(R) (ASCAT) [mm]

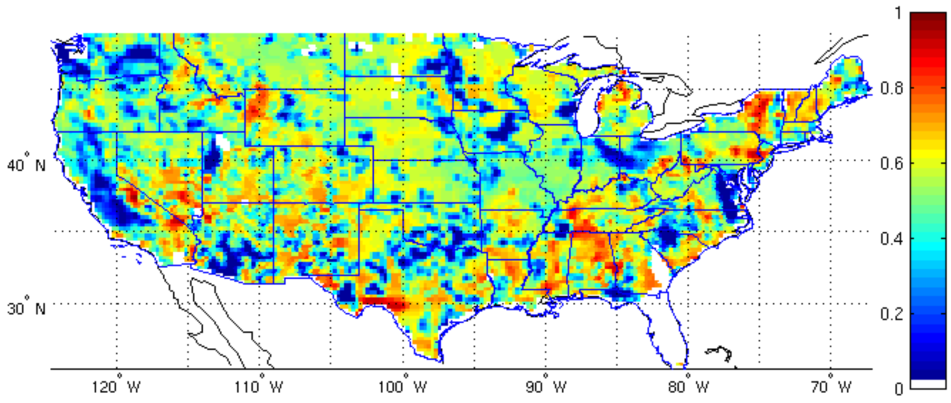




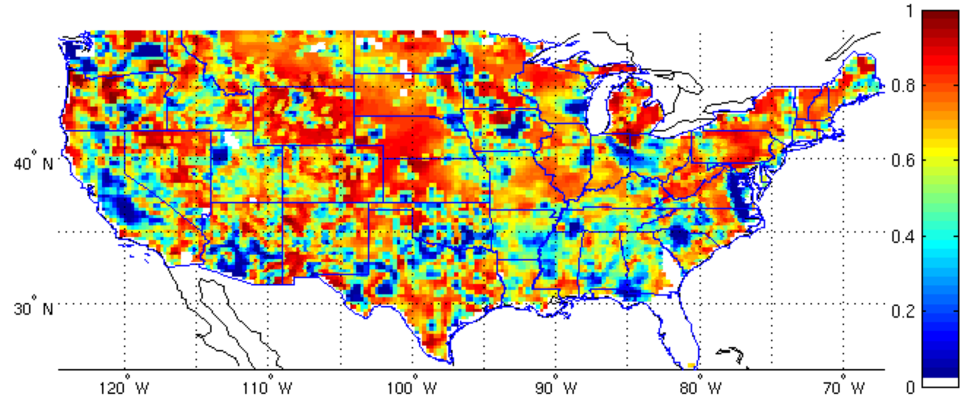
Standard error in daily KF analysis products

Normalized by open loop standard error (**blue is good**/**red is bad)**

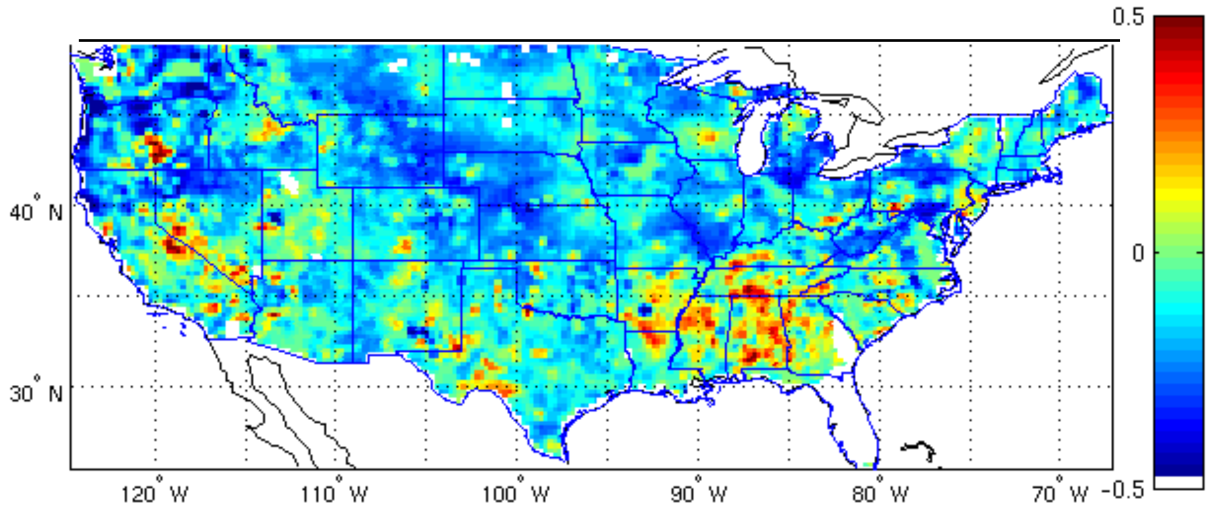
1D RS DA (ASCAT) [-]



2D in situ (SCAN + CRN) [-]



1D minus 2D DA [-]



SCAN+CRN is reducing uncertainty more

ASCAT is reducing uncertainty more

Data Assimilation System

Key details:

- Kalman Filter (KF) assimilation into TRMM-driven linear API model (1-layer, daily).
- Daily/0.25-degree analysis over CONUS.
- Apply 300-km localization to 2D *in situ* DA.
- All modeling and TC applied in anomaly space (relative to 31-day MA climatology).
- Assume zero error auto-correlation for in situ observations in 2D *in situ* DA.
- TC based on [LSM, ASCAT, SMOS] = [diagonal components of **Q** matrix].
- Baseline: 1D ASCAT-based assimilation (also parameterized using TC). Note: ASCAT is on a quasi-operational platform.

Target evaluation metric = KF-predicted analysis (i.e., post-update) surface error variance (based on TC-predicted ***Q***, ***R*** and ***H***).