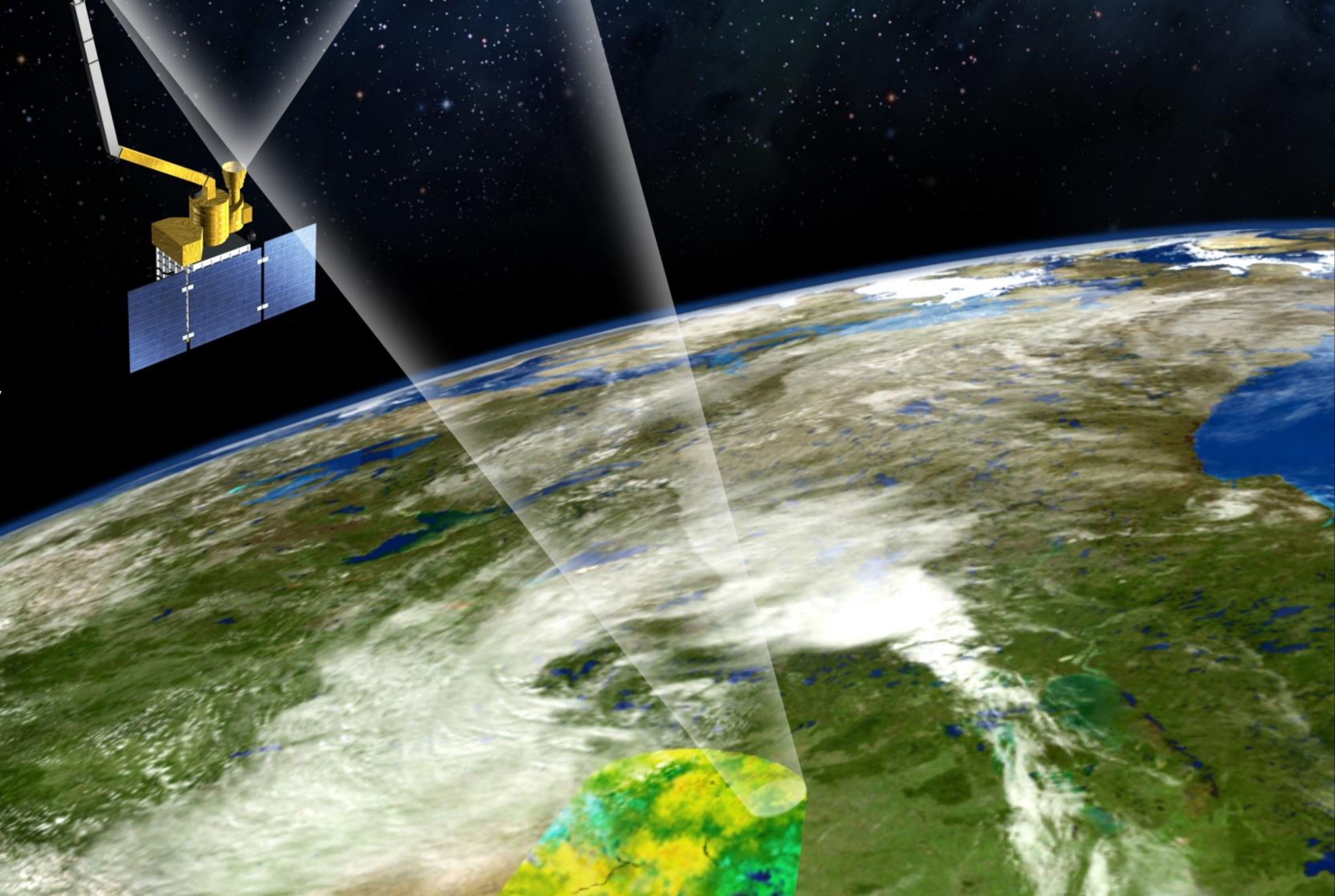
National Aeronautics and Space Administration

Soil Moisture Active Passive Mission

SMAP

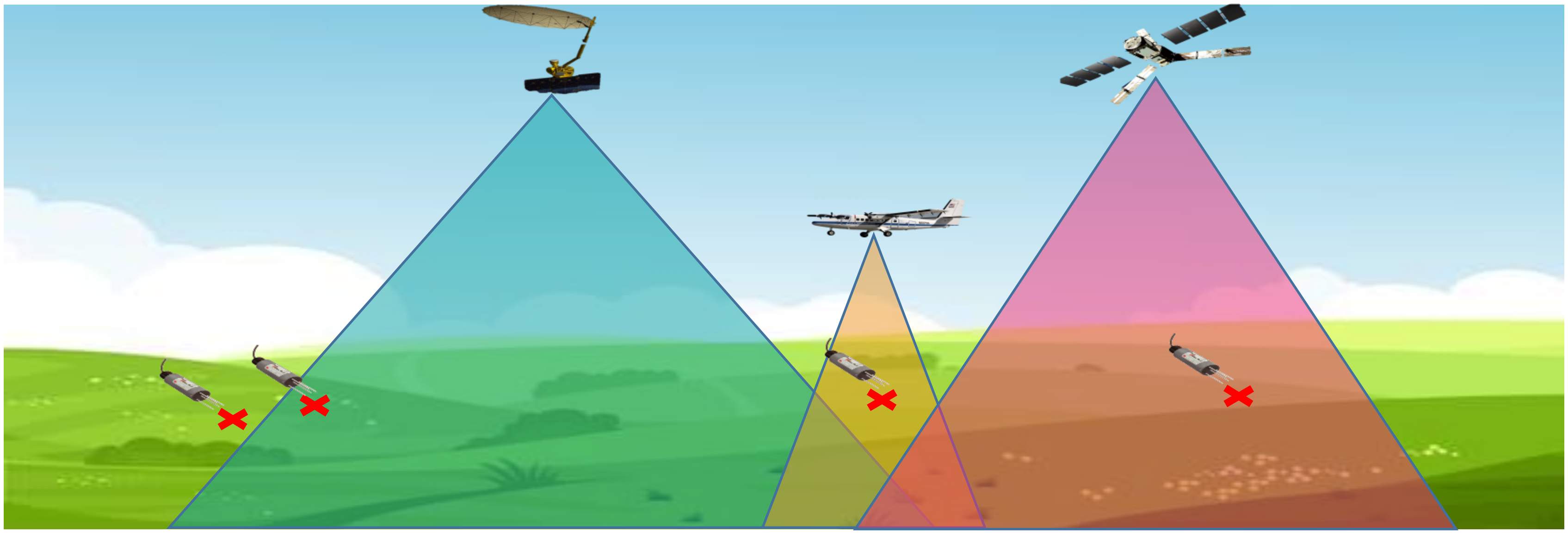
A Multiscale Spatio-Temporal Big Data Fusion Algorithm from Point to Satellite Footprint Scales



Dhruva Kathuria
Binayak P. Mohanty
Matthias Katzfuss
Texas A&M University

SMAP – ST Presentation
April 28, 2021

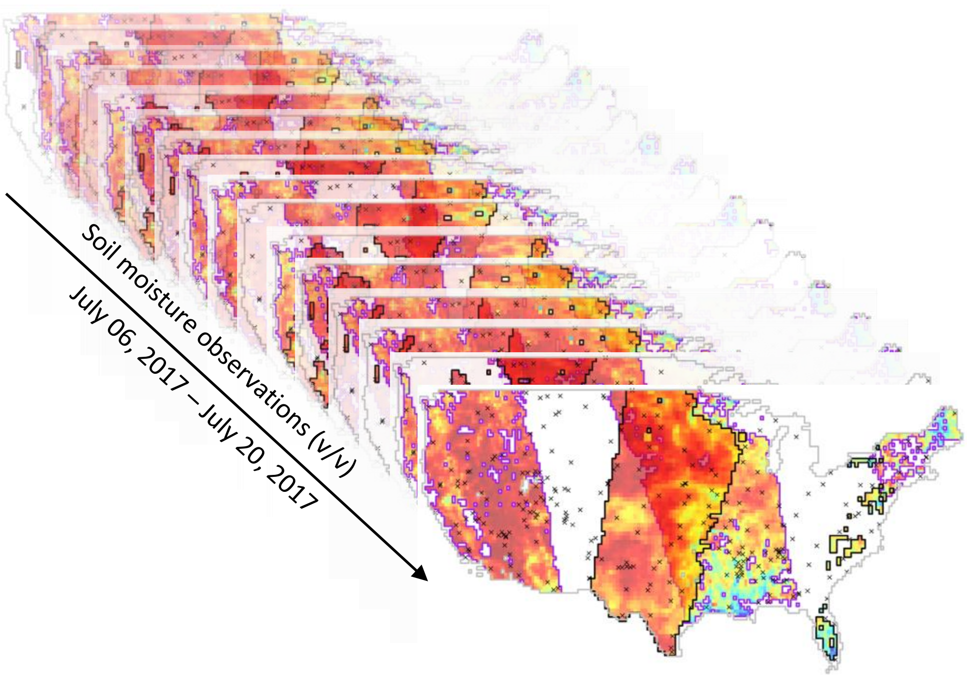
Introduction



- On a daily scale, individual platforms have limitations such as incomplete spatial coverage and errors in retrievals.
- Data fusion is the process of combining information from heterogeneous sources into a single composite picture of the relevant process.
- On an interpretation-prediction spectrum, physical models derived from the first laws of physics lie on one end while Machine Learning algorithms using black-box models fall on the other.

Geostatistical data fusion is an attractive alternative for SM inference in data-driven setting.

Study Area and Data

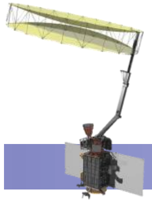


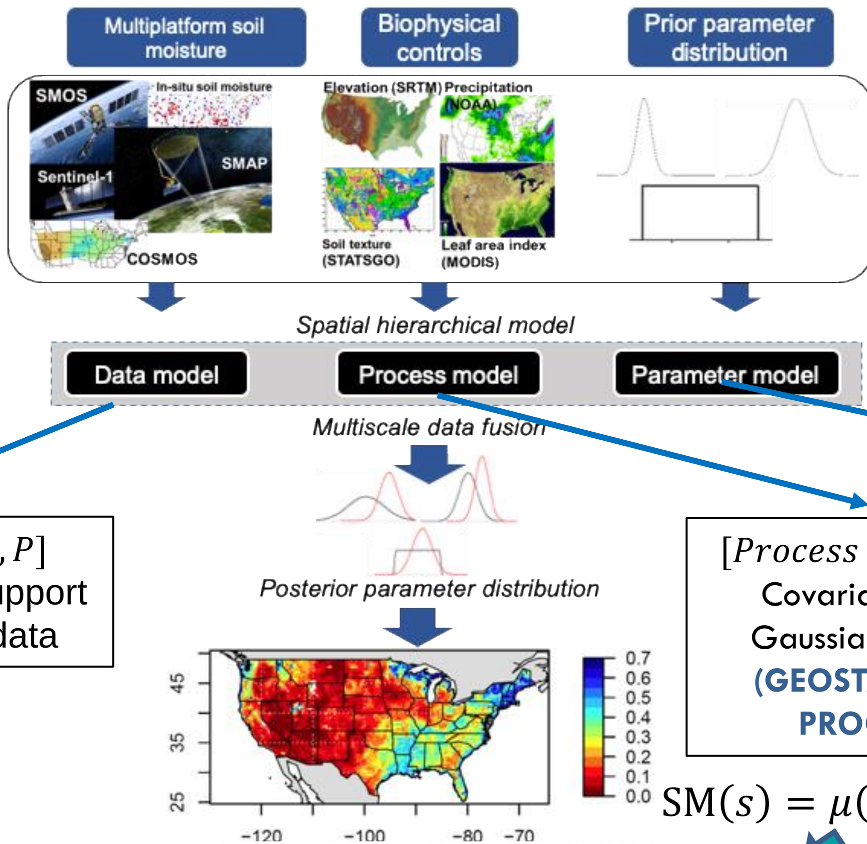
- We propose a data fusion scheme combining point and satellite soil moisture data for Contiguous US.

- **Soil moisture data**
  1) In-situ : USCRN and SCAN stations.
  2) Satellite: SMAP L3 (~ 36 km)
  3) Satellite: SMOS L3 (~ 25 km)

- **Covariate data**
  1) Rainfall (4 km): PRISM
  2) Soil Texture : SSURGO (1 km)
  3) Elevation: SSURGO (1 km)
  4) Leaf Area Index: MODIS (500 m)

Spatio-temporal Hierarchical Model





$[Observations|Process, P]$
Accounts for change of support and errors in observed data

$[Process\ (point)|P]$
Covariate-driven Gaussian Process
(GEOSTATISCAL PROCESS)

$[Parameters\ (P)]$
Parameters in the data and process models

$$\mathrm{SM}(s) = \mu(s) + e(s)$$

$$e(.) \sim GP(0, C)$$

deterministic mean function

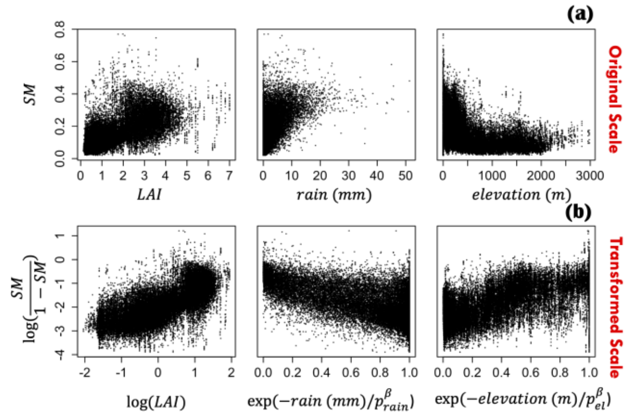
spatio-temporally dependent stochastic process

$$[Data, Process, Parameters] = [Data|Process, Parameters] \times [Process|Parameters] \times [Parameters]$$

Mean and covariance of SM

- The structure of the mean function is selected based on exploratory analysis of soil moisture data.

$$\mu(log(\frac{SM}{1-SM})) = \mu(SM') = \beta_0 + \beta_1 log(LAI) + \beta_2 exp(-\frac{rain}{p_{rain}^{\beta}}) + \beta_3 exp(-\frac{elevation}{p_{elevation}^{\beta}})$$



- The covariance function is modeled such that the covariance between any two locations is a function of the underlying covariate heterogeneity.

$$\begin{aligned} Cov(e(\mathbf{s_1}), e(\mathbf{s_2})) &= Cov(\sum_{j=1}^{M} w_j(X(\mathbf{s_1}))e_j(\mathbf{s_1}), \sum_{j=1}^{M} w_j(X(\mathbf{s_2}))e_j(\mathbf{s_2})) \\ &= \sum_{j=1}^{M} w_j(X(\mathbf{s_1}))w_j(X(\mathbf{s_2}))C_j|\mathbf{s_1} - \mathbf{s_2}| \\ &= C(\mathbf{s_1}, \mathbf{s_2}, X(\mathbf{s_1}), X(\mathbf{s_2})) \end{aligned}$$

$e(.)$ = Stochastic process governing spatial dependence

$X(s)$ = vector of controls (LAI, rain, clay, elevation) at point s (x, y, t).

C_j = j^{th} isotropic spacetime covariance function.

$w_j(s_1)$ = weighting function governing the effect of controls on C_j

M = number of isotropic covariance functions

Water Resources Research
RESEARCH ARTICLE
10.1029/2018WR023505

Key Points:
- Proposed a framework to assess spatial nonstationarity of soil moisture
- Optimal prediction and upscaling of soil moisture under nonstationarity
- Quantified the effects of soil texture and vegetation on the spatial variance/correlation of soil moisture

A Nonstationary Geostatistical Framework for Soil Moisture Prediction in the Presence of Surface Heterogeneity

Dhruva Kathuria[1], Binayak P. Mohanty[1], and Matthias Katzfuss[2]
[1]Biological and Agricultural Engineering, Texas A&M University, College Station, TX, USA, [2]Department of Statistics, Texas A&M University, College Station, TX, USA

Abstract Soil moisture is spatially variable due to complex interactions between geologic, topographic,

Spatio-temporal Data Fusion Model

Data model for pixel A_i: $z_j(A_i) = SM(A_i) + \delta(A_i) + \kappa(A_i)\, SM(A_i) + \epsilon(A_i)$

Process model at point scale: $SM(.) \sim GP(\mu, C)$

$\frac{1}{\,}\int$

Likelihood estimation consists of simulating and inverting the covariance matrix which scales quadratically with the number of assumed grid points and cubically with the number of observations.

$$-2log(f(z(\mathcal{A})|\theta) = log(det(\boxed{\Sigma_z})) + (z(\mathcal{A}) - \mu_z)^T \boxed{\Sigma_z^{-1}} (z(\mathcal{A}) - \mu_z) + nlog(2\pi),$$

$$\mu_{z,i} \approx (h^{\kappa}_{A_i})^T \mu_{A_i} + \delta(A_i),$$

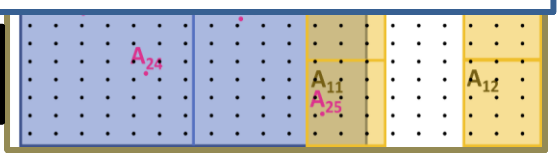
$$\Sigma_{z,ij} \approx (h^{\kappa}_{A_i})^T (C(\mathcal{G}_{A_i}, \mathcal{G}_{A_j})) h^{\kappa}_{A_j} + \tau^2_{A_{i,j}},$$

Computationally infeasible for big datasets and vast study domains

$g_l \in \mathcal{G} \cap A_i$

$$= h_{A_i} SM_{\mathcal{G}_i}$$

$$C(A_i, A_k) = h_{A_i} C(SM_{\mathcal{G}_i}, SM_{\mathcal{G}_k}) h_{A_k}$$



n_{A_i} = number of grid points inside $A_i = |\mathcal{G} \cap A_i|$

h_{A_i} = vector of length $n_{A_i} = ({}^{1}/_{n_{A_i}}, \ldots, {}^{1}/_{n_{A_i}})$

$SM_{\mathcal{G}_i}$ = vector of length $n_{A_i} = \{SM(g_l): g_l \in \mathcal{G} \cap A_i\}$


Water Resources Research

RESEARCH ARTICLE

10.1029/2018WR024581

Multiscale Data Fusion for Surface Soil Moisture Estimation: A Spatial Hierarchical Approach

Dhruva Kathuria[1], Binayak P. Mohanty[1], and Matthias Katzfuss[2]

Key Points:
- Proposed a multi-scale data fusion framework accounting for spatial variance/correlation of soil moisture
- The proposed framework optimally separates the inherent soil moisture dynamics and measurement errors in instruments
- The framework is applied to combine point, airborne and satellite data in a heterogeneous watershed

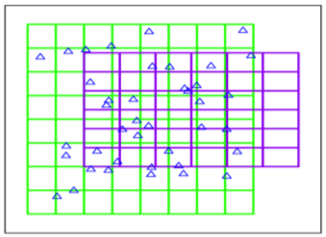
Abstract Surface soil moisture (SSM) has been identified as a key climate variable governing hydrologic and atmospheric processes across multiple spatial scales at local, regional, and global levels. The global burgeoning of SSM datasets in the past decade holds a significant potential in improving our

Multiscale approximation for Big Data

- Likelihood of fusion model = $-2log\boxed{(f(z(\mathcal{A})|\theta)} = log(det(\Sigma_z)) + (z(\mathcal{A}) - \mu_z)^T \Sigma_z^{-1}(z(\mathcal{A}) - \mu_z) + nlog(2\pi),$

Cost of computation = $\mathcal{O}(n^3) + \mathcal{O}(n_{\mathcal{G}}^2)$

Hypothetical Example



Data

1) Areal data R_1: 64 pixels (Green)
2) Areal data R_2: 36 pixels (Purple)
3) Point data P_1: 40 (Blue triangles)

Total data pixels $= \mathcal{A} = \{A_1, \dots, A_n\}; n = 140$

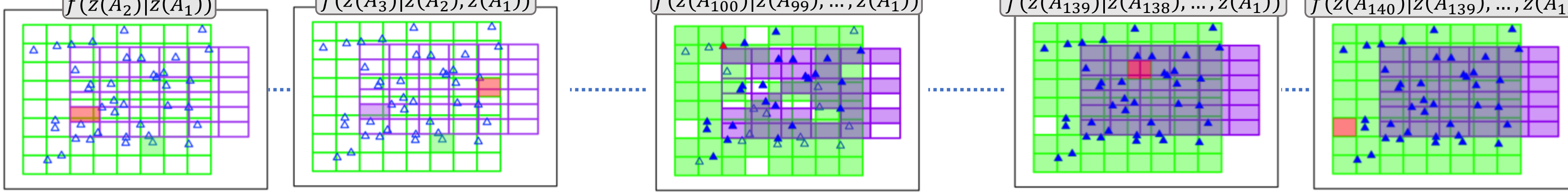
For 15-day Contiguous US analysis

$n = 100,386$

$n_{\mathcal{G}} = 1{,}500{,}000$

Exact Likelihood: $f(z(\mathcal{A})|\theta) = f(z(A_1)|\theta) \times \prod_{i=2}^{n} f(z(A_i)|\mathbf{z}(\mathbf{A}_{1:i-1}), \theta),$

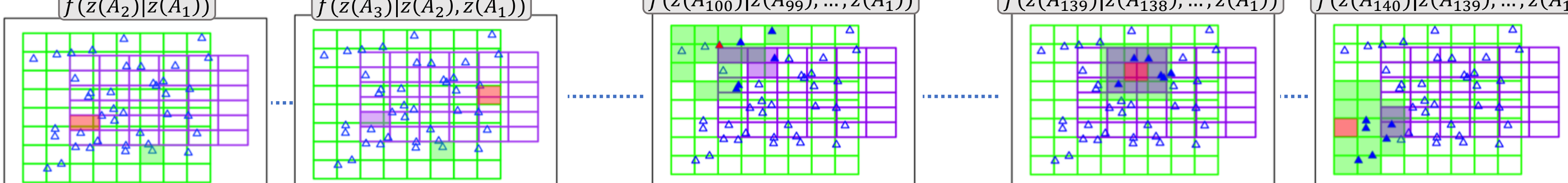
$f(z(A_2)|z(A_1))$ ... $f(z(A_3)|z(A_2), z(A_1))$ ... $f(z(A_{100})|z(A_{99}), \dots, z(A_1))$ ... $f(z(A_{139})|z(A_{138}), \dots, z(A_1))$ ... $f(z(A_{140})|z(A_{139}), \dots, z(A_1))$



Approximate Likelihood: $\hat{f}(z(\mathcal{A})|\theta) = f(z(A_1)|\theta) \times \prod_{i=2}^{n} f(z(A_i)|\mathbf{z}(\mathbf{A}_{m_i}), \theta),\ m_i = \begin{cases} i-1, i \le m \\ m, i > m \end{cases}$

Here, $m = 20$

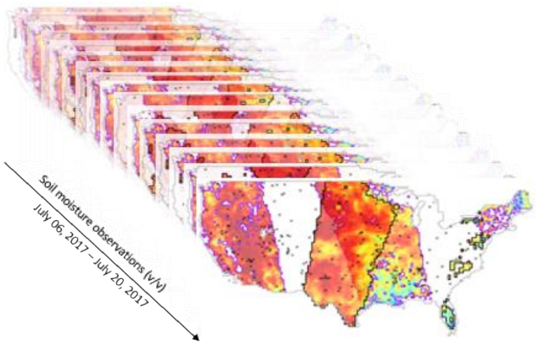
$\hat{f}(z(A_2)|z(A_1))$ ... $\hat{f}(z(A_3)|z(A_2), z(A_1))$ ... $\hat{f}(z(A_{100})|z(A_{99}), \dots, z(A_1))$ ... $\hat{f}(z(A_{139})|z(A_{138}), \dots, z(A_1))$ ... $\hat{f}(z(A_{140})|z(A_{139}), \dots, z(A_1))$

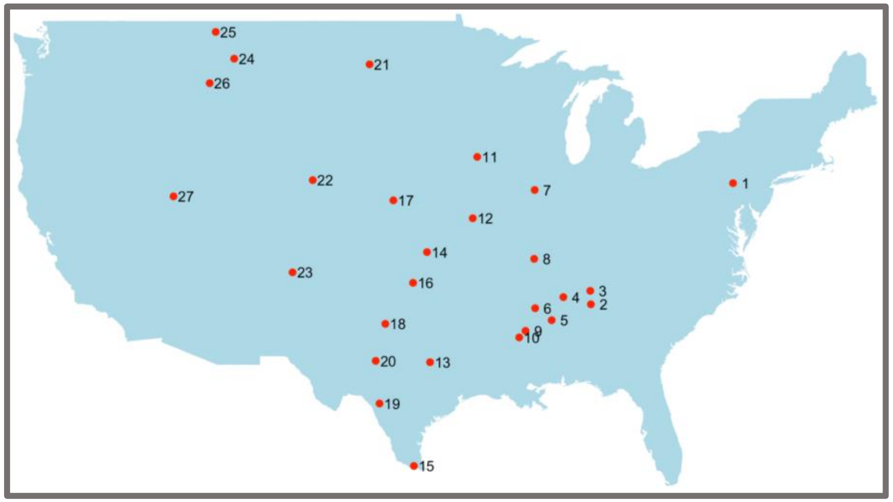


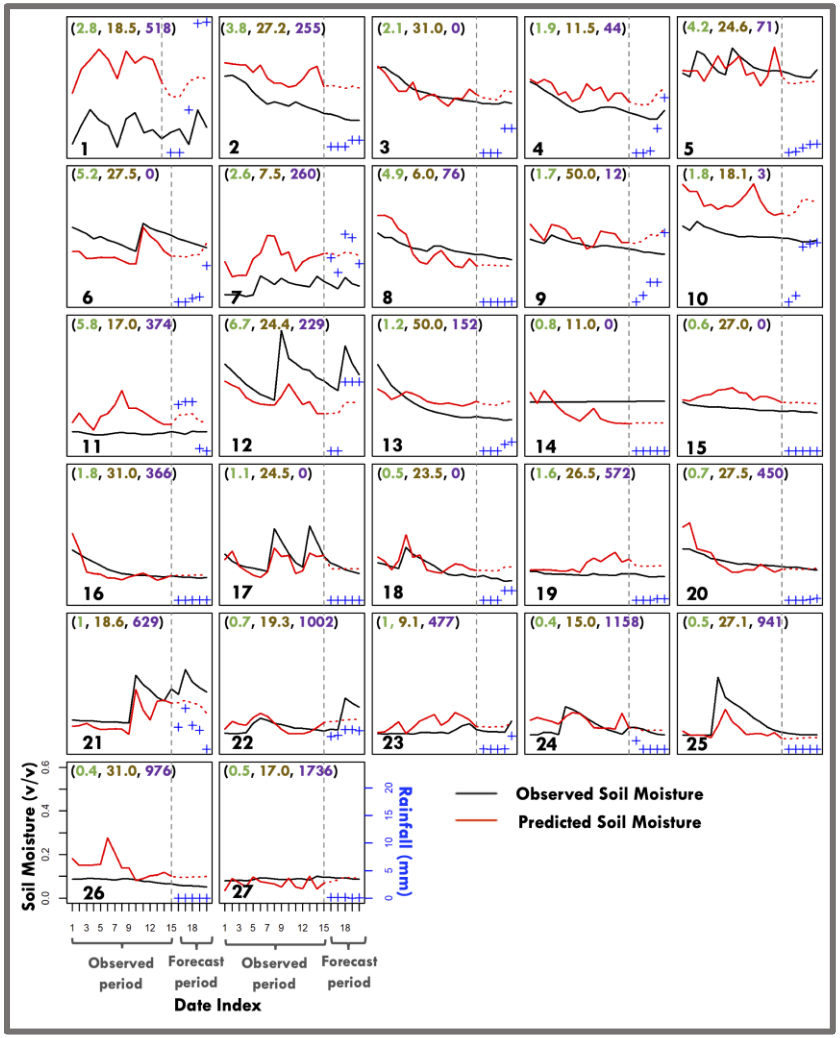
Kathuria, D., Mohanty, B. P. and Katzfuss, M. (2021). *A Multiscale Spatio-Temporal Big Data Fusion Algorithm from Point to Satellite Footprint Scales.* (under review RSE).

Multiscale predictions and forecasts

POINT SCALE

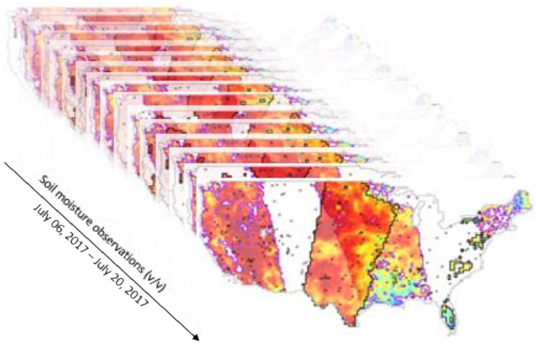




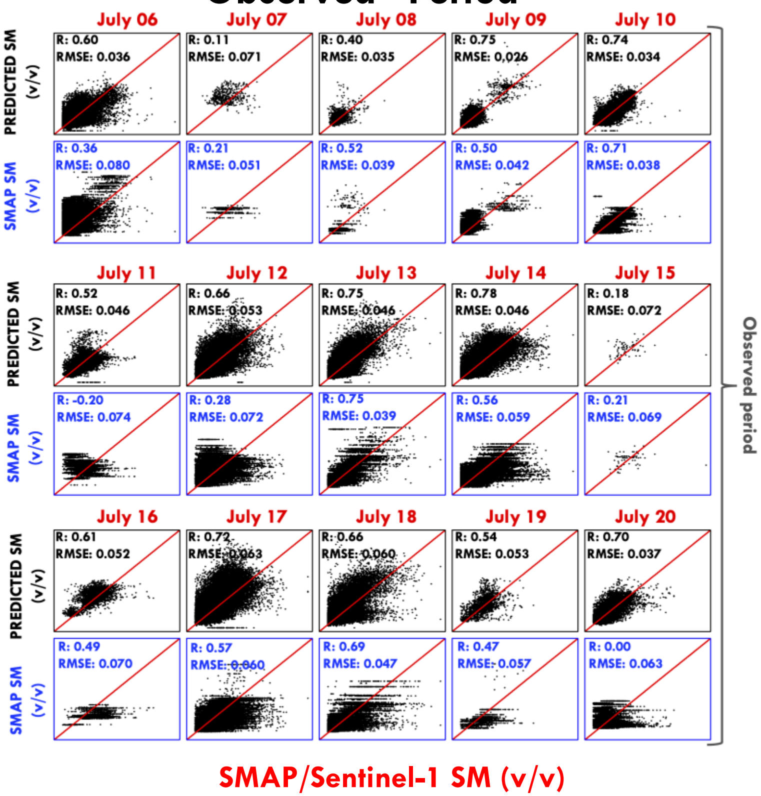


Multiscale predictions and forecasts

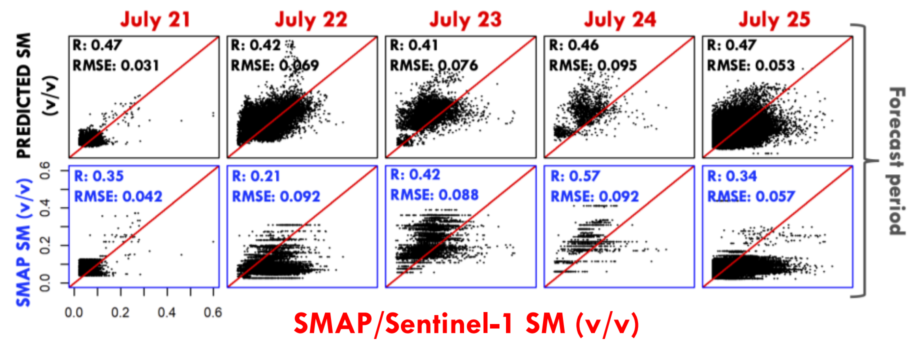
SMAP/SENTINEL-1 SM
(~ 3 km)



"Observed" Period



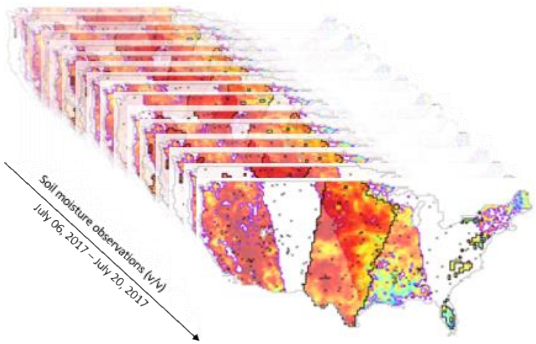
"Forecast" Period

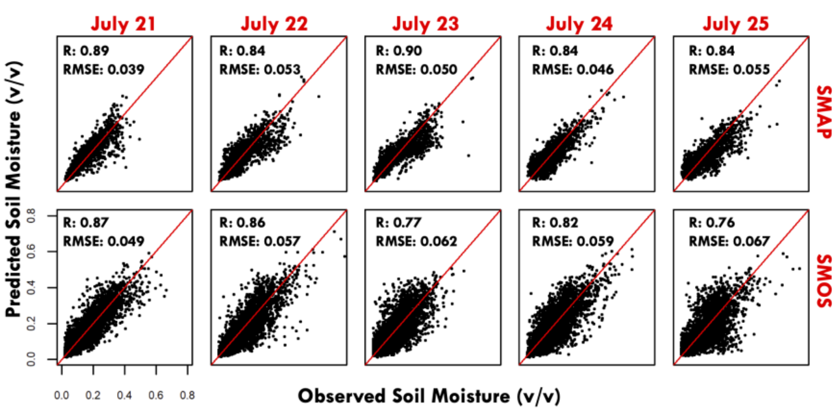


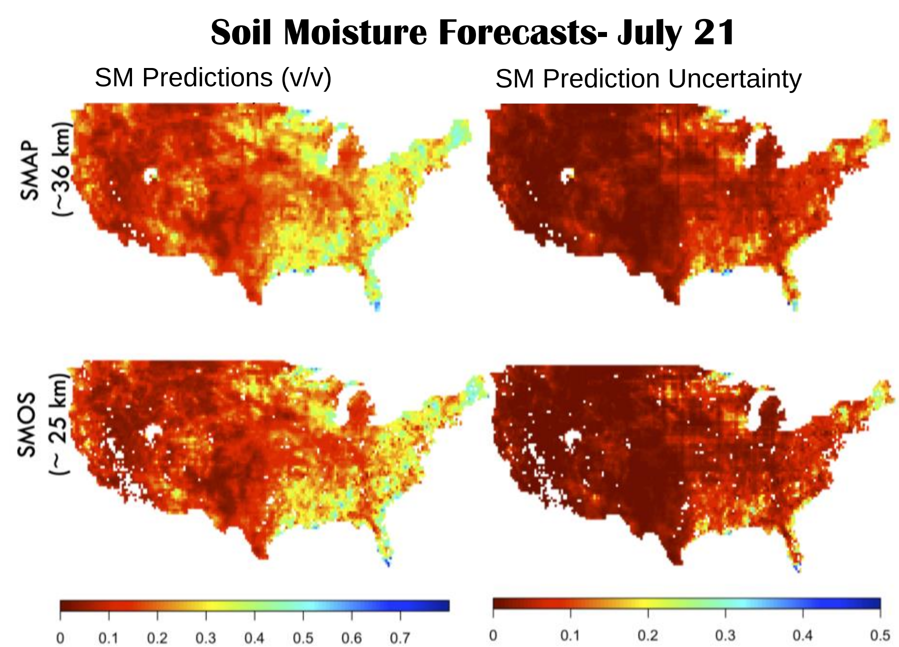
- For most of the days the soil moisture predictions agree well with the SMAP/Sentinel-1 product outperforming the base SMAP product even for the forecast period.

Multiscale predictions and forecasts

SMAP and SMOS



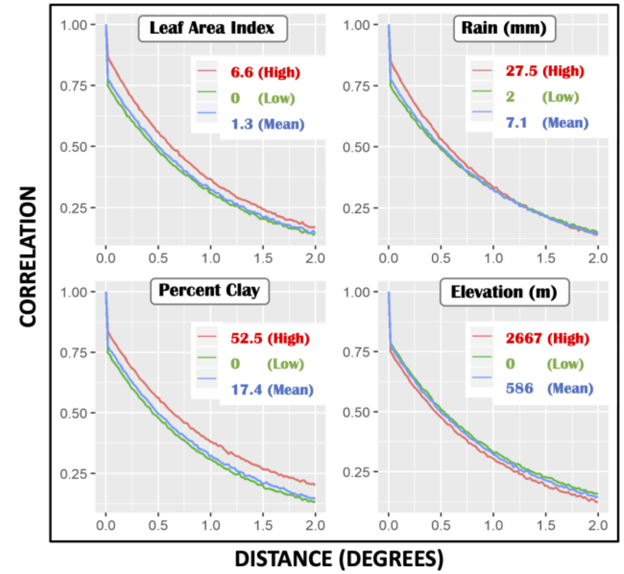


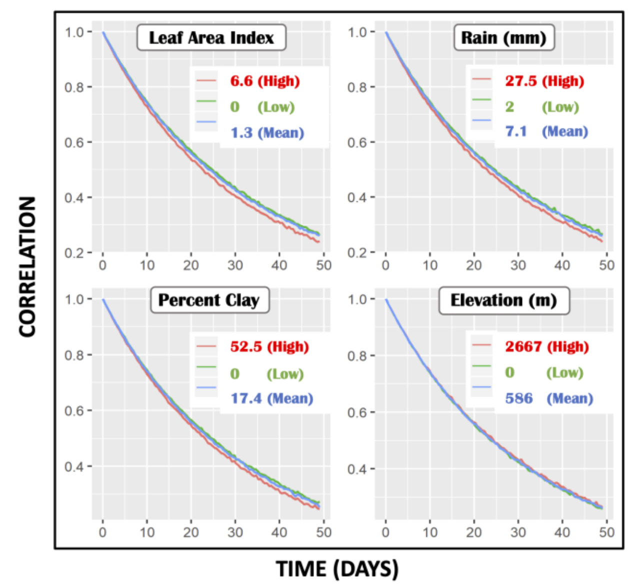


- Five-day forecasts of SM have satisfactory accuracy.
- The predictions are accompanied by prediction uncertainty.

Effect of covariates on soil moisture

- The mean soil moisture is affected by antecedent rainfall, vegetation and elevation.
- The spatial covariance of soil moisture is affected by vegetation, rainfall, percent clay and elevation.
- The temporal covariance was not affected by the chosen covariates in the analyzed 15-day data. Longer temporal data may be required.

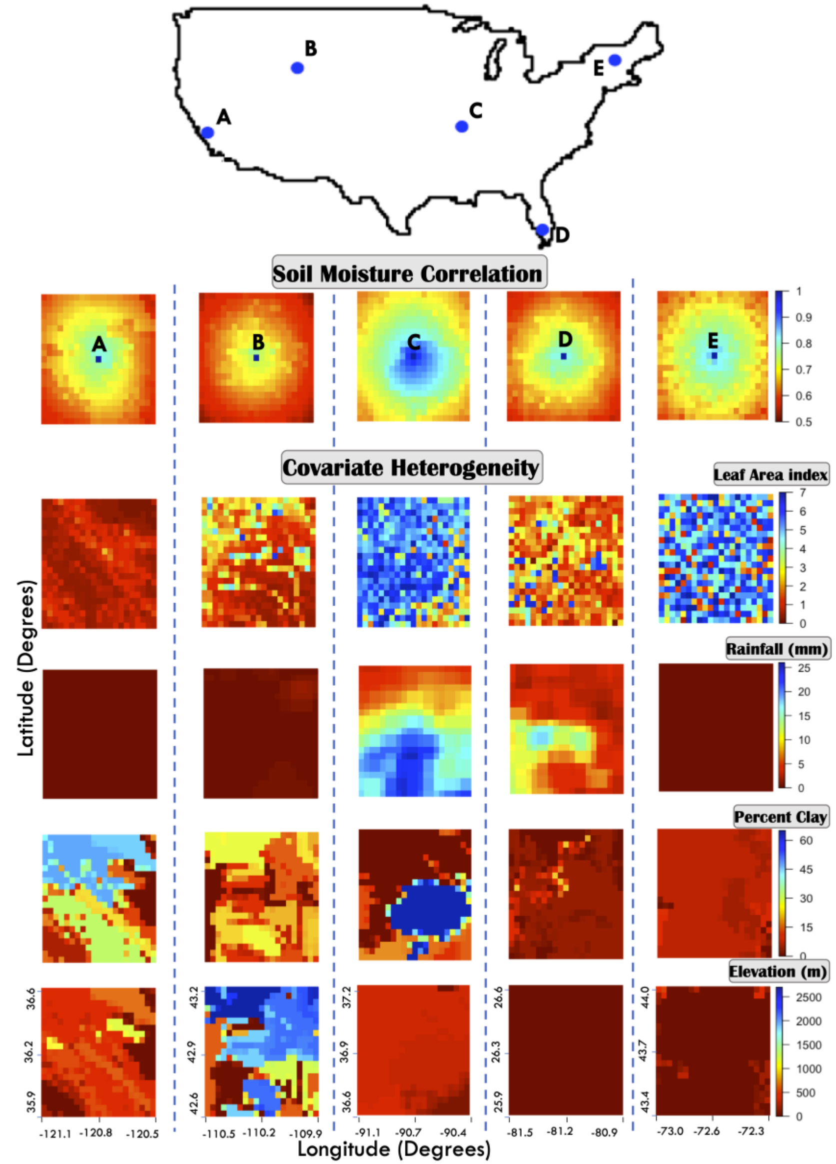




Kathuria, D., Mohanty, B. P. and Katzfuss, M. (2021). *A Multiscale Spatio-Temporal Big Data Fusion Algorithm from Point to Satellite Footprint Scales.* (under review RSE).

Effect of covariates on soil moisture

- The mean soil moisture is affected by antecedent rainfall, vegetation and elevation.
- The spatial covariance of soil moisture is affected by vegetation, rainfall, percent clay and elevation.
- The temporal covariance was not affected by the chosen covariates in the analyzed 15-day data. Longer temporal data may be required.
- The spatial covariance/correlation exhibits non-stationary behavior across CONUS driven by physical controls.



Kathuria, D., Mohanty, B. P. and Katzfuss, M. (2021). *A Multiscale Spatio-Temporal Big Data Fusion Algorithm from Point to Satellite Footprint Scales.* (under review RSE).

Conclusions

- We propose a novel geostatistical framework for fusing multiscale Big Data.
- We apply the fusion scheme to combine point and satellite soil moisture data for CONUS.
- We validate soil moisture predictions and forecasts across multiple scales.
- We quantify the effects of physical controls on soil moisture distribution.
- The proposed algorithm is general and can be used to fuse other environmental variables.