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

- 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.
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

Parameters in the data and process models

\[ \text{Multiplatform soil moisture} \]
\[ \text{Biophysical controls} \]
\[ \text{Prior parameter distribution} \]

Spatial hierarchical model

Data model  Process model  Parameter model

Multiplatform data fusion

\[ \text{Observations} | \text{Process, } P \]
Accounts for change of support and errors in observed data

Posterior parameter distribution

\[ \text{Process (point)} | P \]
Covariate-driven Gaussian Process (GEOSTATISTICAL PROCESS)

\[ \text{SM}(s) = \mu(s) + e(s) \]
deterministic mean function

\[ e(.) \sim GP(0, C) \]
spatio-temporally dependent stochastic process

\[ \text{Data, Process, Parameters} = [\text{Data}|\text{Process, Parameters}] \times [\text{Process}|\text{Parameters}] \times [\text{Parameters}] \]
Mean and covariance of SM

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

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

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

\[
\text{Cov}(e(s_1), e(s_2)) = \text{Cov}\left(\sum_{j=1}^{M} w_j(X(s_1))e_j(s_1), \sum_{j=1}^{M} w_j(X(s_2))e_j(s_2)\right)
= \sum_{j=1}^{M} w_j(X(s_1))w_j(X(s_2))C_j|s_1 - s_2|
= C(s_1, s_2, X(s_1), X(s_2))
\]

\(e(\cdot) = \) 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
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: $SM(.) \sim GP(\mu, C)$

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.

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

$$\mu_{z,i} \approx (h^\kappa_{Ai})^T \mu_A + \delta(A_i),$$

$$\Sigma_{z,ij} \approx (h^\kappa_{Ai})^T (C(G_A, G_A)) h^\kappa_{Aj} + \tau^2_{Ai,j},$$

$$= h_{Ai} SM_{Gi}$$

$$C(A_i, A_k) = h_{Ai} C(SM_{Gi}, SM_{Gk}) h_{Ak}$$

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

$h_{Ai} = \text{vector of length } n_{Ai} = (1/n_{Ai}, \ldots, 1/n_{Ai})$

$SM_{Gi} = \text{vector of length } n_{Ai} = \{SM(g_i): g_i \in G \cap A_i\}$
Likelihood of fusion model = 

\[-2 \log(f(z(A)|\theta)) = \log(\det(\Sigma_z)) + (z(A) - \mu_z)^T \Sigma^{-1}_z (z(A) - \mu_z) + n \log(2\pi),\]

Cost of computation = \(O(n^3) + O(n_G^2)\)

### 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 = \(A = \{A_1, \ldots, A_n\}; n = 140\)

#### Exact Likelihood:

\[f(z(A)|\theta) = f(z(A_1)|\theta) \times \prod_{i=2}^n f(z(A_i)|z(A_{1:i-1}), \theta),\]

#### Approximate Likelihood:

\[\hat{f}(z(A)|\theta) = f(z(A_1)|\theta) \times \prod_{i=2}^n f(z(A_i)|z(A_{m_i}), \theta), \quad m_i = \{i-1, i \leq m \mid m_i, i > m\}\]

For 15-day Contiguous US analysis

\[n = 100,386\]

\[n_G = 1,500,000\]

Multiscale predictions and forecasts

POINT SCALE

Map of the United States with points marked and graphs showing multiscale predictions and forecasts.
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.
Five-day forecasts of SM have satisfactory accuracy.

The predictions are accompanied by prediction uncertainty.
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.
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.

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.