A non-stationary multiscale data fusion framework for soil moisture estimation

Binayak P. Mohanty and Dhruva Kathuria

Biological & Agricultural Engineering
Texas A&M University, USA
Past decade has witnessed **global burgeoning of soil moisture datasets** from in situ networks and remote sensing platforms.

- On a daily scale, individual platforms have **limitations such as incomplete spatial coverage and errors in soil moisture retrievals.**

**Fusion of multiple soil moisture platforms** holds a significant potential in improving soil moisture predictions and understanding multiscale soil moisture dynamics.

**Data fusion** is the process of combining information from heterogeneous sources into a single composite picture of the relevant process, such that the composite picture is generally more accurate and complete than that derived from any single source alone.
Multiscale data fusion: a spatial hierarchical approach

Spatial Hierarchical Model

Data model

Process model

Parameter model

Multiscale data fusion

Parameter inference

Prediction across scales

Posterior distribution for parameters

Soil moisture observations at multiple scales

Elevation

Soil texture

Vegetation

Rainfall

Controls affecting soil moisture distribution

Prior distribution for parameters

Elevation

Soil texture

Vegetation

Rainfall

Controls affecting soil moisture distribution

Prior distribution for parameters

Data model

Process model

Parameter model

Multiscale data fusion

Parameter inference

Prediction across scales

Posterior distribution for parameters
Data fusion across scales

- For the non-stationary process model at point scale we define the corresponding soil moisture distribution at a spatial support A as:

\[ y(A) = \frac{1}{|A|} \int_A y(s) ds \]

- We use a numerical approximation by assuming a fine-scale grid \( \mathcal{G} \) to approximate the stochastic spatial integral \textbf{resulting in a change of support matrix} \( H \) and use it to perform data fusion of multiple soil moisture platforms.

\[ [z|y, P_z] \sim \mathcal{N}(\Delta + Hy(\mathcal{G}), \Sigma) \]

### Challenges in Multiplatform data fusion

- Inherent \textit{spatial variability of soil moisture} caused due to precipitation and land surface controls such as vegetation, soil texture and topography.

- \textbf{Systematic and random errors} in remote sensing retrievals.

- \textbf{Fusion of data platforms with different spatial supports}.

- \textbf{Massive size} of the datasets for continental scale fusion.

\[ z = (z_{A1}, ..., z_{An}) \] represents observed SM data from all the platforms.

\[ \Delta = (\Delta_{A1}, ..., \Delta_{An}) \] are the biases associated with each observation.

\( \Sigma \) is the \( n \times n \) error matrix.
Simultaneous validation at point, airborne and satellite scales

- The proposed scheme is applied to 5 days during SMEX02 and 10 days during SMAPVEX12.

**Figure 3**, Validation plots for hold-out data at point and airborne scales for A) SMEX02 and B) SMAPVEX12.

- For each day, we validate the fusion framework using hold-out data at point and airborne scales and back-predicting the satellite pixels.

**Table 1**, Root mean squared error (RMSE) values at point, airborne and satellite scales.