Enhancing Soil Moisture Resolution: Downscaling of ESA CCI Data over West Africa using Hybrid Model

Odunayo David Adeniyi
Africa Research Fellow
European Space Agency
29/05/2024

Motivation

- Soil moisture (SM) is a critical factor in environmental processes.
- It affects plant growth, hydrology, and climate dynamics.
- Understanding soil moisture variations is essential for sustainable land management.
- However, traditional ground-based SM measurements lack spatiotemporal continuity.

Motivation

- The European Space Agency Climate Change Initiative (ESA CCI) provides critical long-term climate data records, including SM products, which are vital for climate research and environmental monitoring.

- The coarse spatial resolution of these products limits their regional application, especially in West Africa, where high-resolution data are needed to assess climate impacts and support adaptation efforts.

Motivation

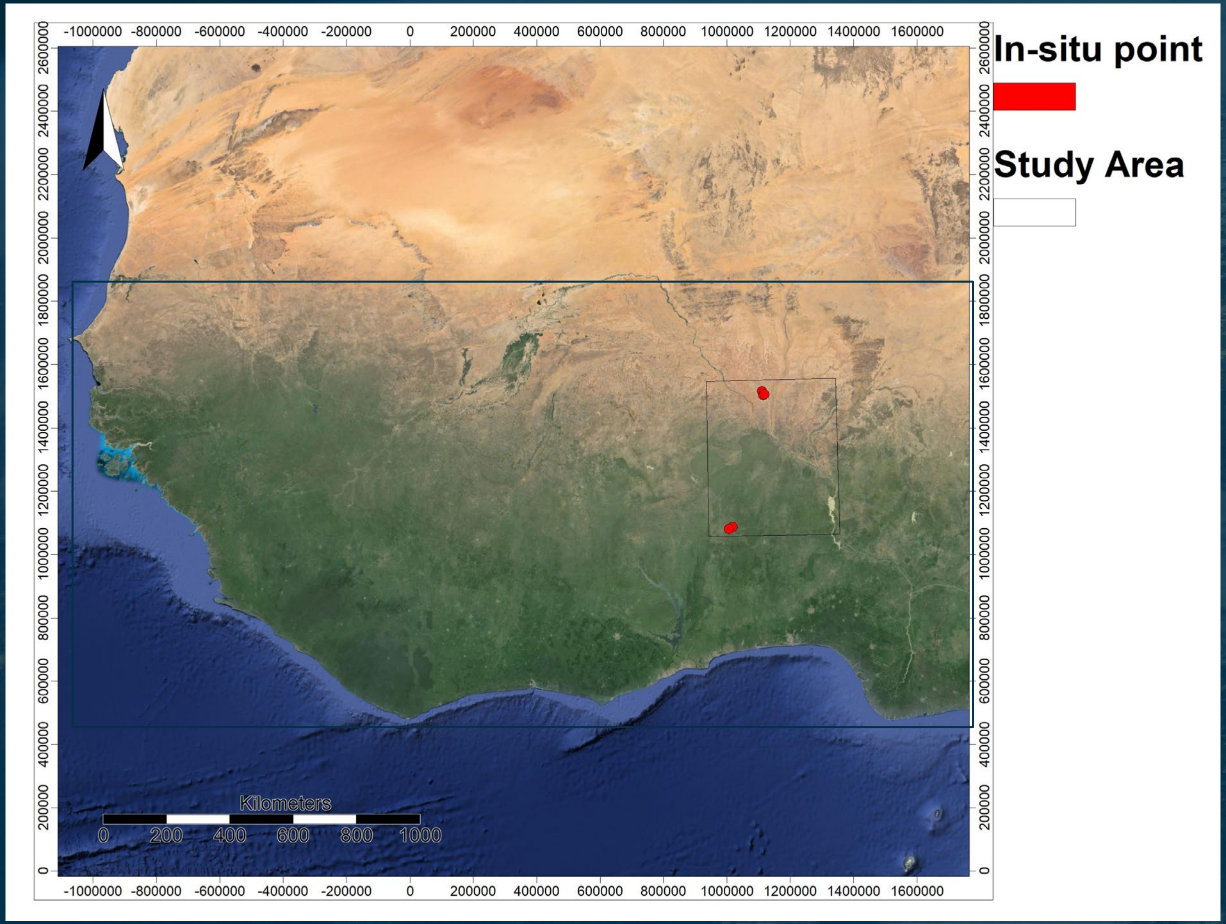
- Several downscaling techniques have been developed to enhance the spatial resolution of Earth Observation SM products.
- Limitations of traditional methods include:
  - High computational demand
  - Difficulty in capturing non-linear relationships
  - Assuming stationarity
- These limitations have led to the adoption of machine learning (ML) techniques for downscaling.

Goal

- Challenge:
  - ML model architecture and variable choice can contribute to model error, evidenced by residuals ($y_{actual\ SM} - y_{predicted\ SM}$). This residuals could have spatial autocorrelation with auxiliary variables and can be interpolated using geostatistical methods.

Therefore, the aim of this study is to combine ML Model and Geostatistical Method as a hybrid model to downscale ESA CCI SM products to 1 km over West Africa.

Study Area



	Sites	Land cover	Elevation (m)
Benin	Belefoungou	Trees	414
	Nalohou	Cropland	441
Niger	Banizoumbou	Cropland	208
	Tondikiboro	Cropland	235
	Wankama	Cropland	230

Data set

- Target product: ESA CCI SM (Combine Active and Passive MW – 25 km)
- MODIS DATA: MOD11A1/061 (LST) and MOD09GA/006/NDVI (NDVI)
- Target resolution: 1 km resolution
- Temporal resolution: Daily (April to August, 2016 – Raining Season)
- Derived indices: TVDI

  - $TVDI = \frac{LST - LST_{min}}{LST_{max} - LST_{min}}$
  - $LST_{max} = a + b * NDVI$
  - $LST_{min} = c + d * NDVI$

Where, NDVI represents the vegetation index corresponding to LST image. LST_{min} is the minimum LST corresponding to the same NDVI value, and LST_{max} is the maximum LST corresponding to the same NDVI value. Here, a, b, c, and d are the coefficients of the TVDI dry-wet edge equation (Sandholt et al., 2002).

- In situ data: ISMN database covering West Africa
- Date splited to 80% Training, 10% Validation, 10% Testing data sets

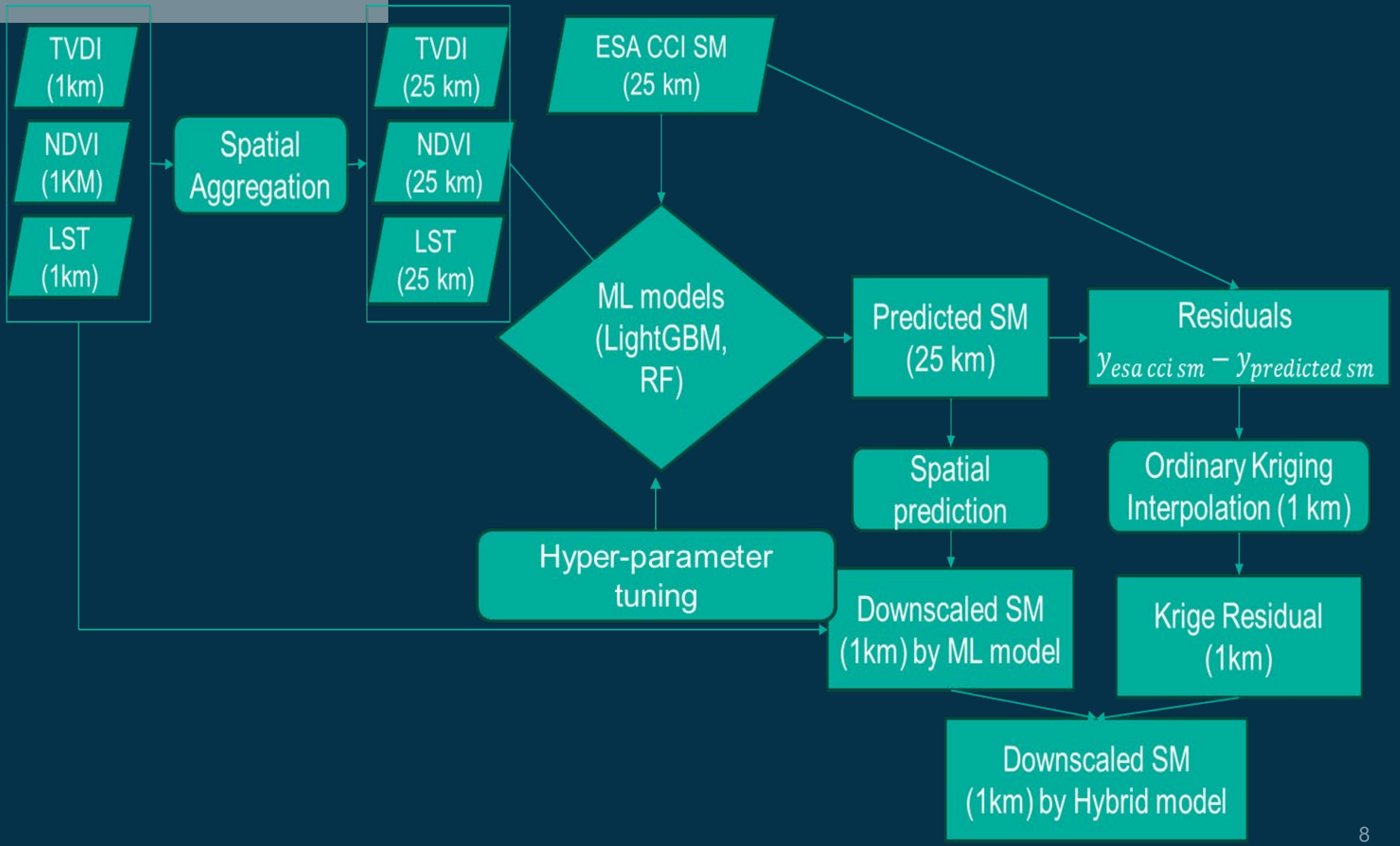
Downscaling Approach

- Concept of the Hybrid Model:

Machine Learning with Residual Kriging (MLRK)

$$\hat{Z}_{MLRK} = \hat{Z}_{ML}(x) + \hat{e}_{RK}(x) = f_x\left(V_g(x)\right) + \sum_{k=1}^{n} \lambda_k \,.\, e(x_k); g = 1,2,\dots,s\ k = 1,2,\dots,n$$

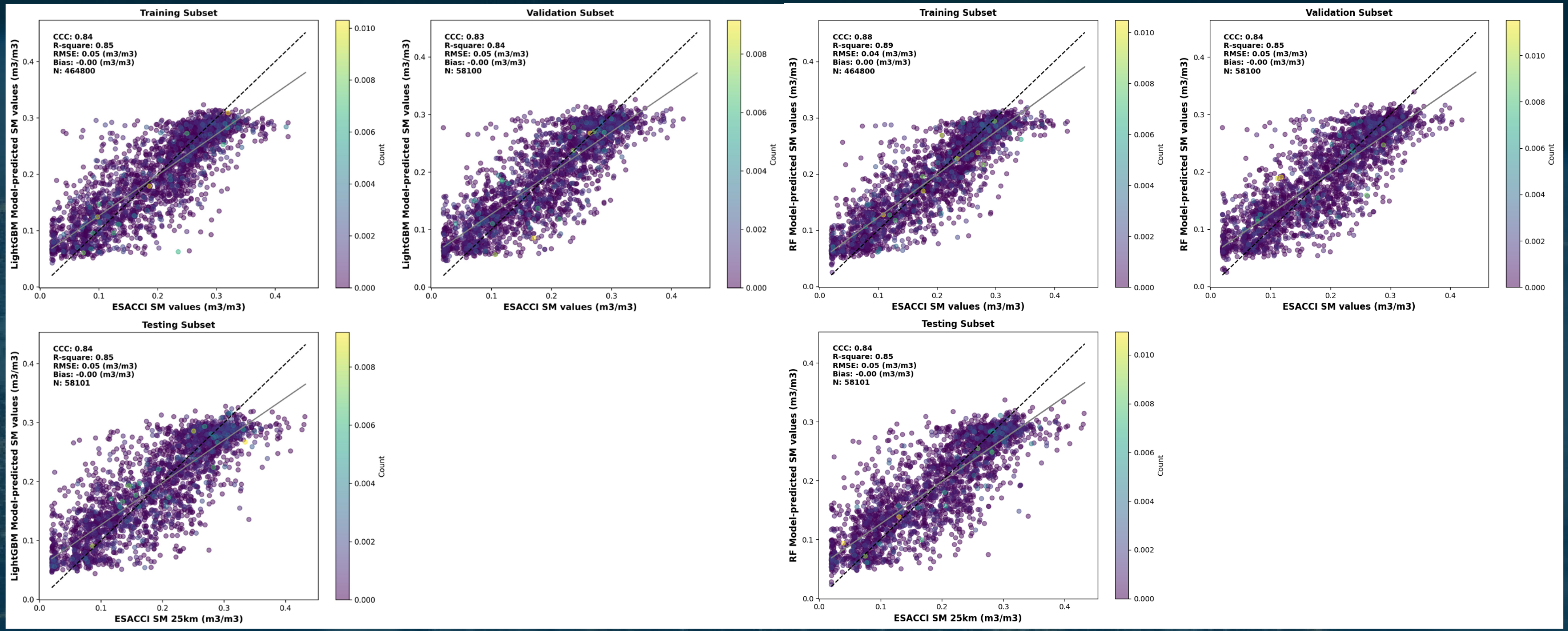
Where $Z\hat{\ }\_MLRK$ refers to the MLRK predicted values, $Z\hat{\ }\_ML\ (x)$ refers to the trend prediction, $e\hat{\ }\_RK\ (x)$ refers to the interpolated trend residual at point x, $f\_x\ (V\_g\ (x))$ refers to the functional relationship between soil and environmental variables V_g at the point x, λ_k refers to kriging weights which is determined by the spatial dependence structure of the trend residual, and $e(x\_k\ )$ refers to the trend residual at the sampling point x_k.



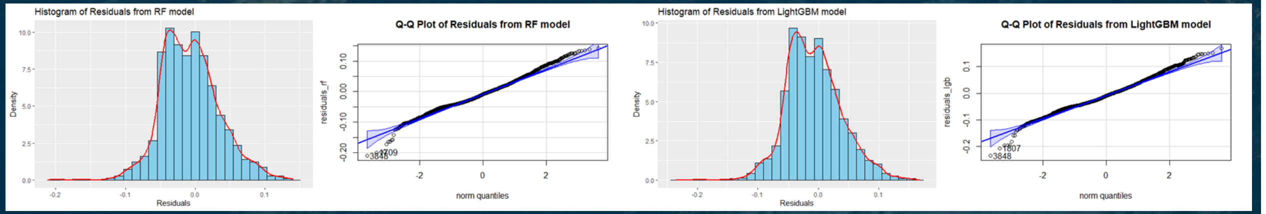
- Residual semi variance analysis

Results

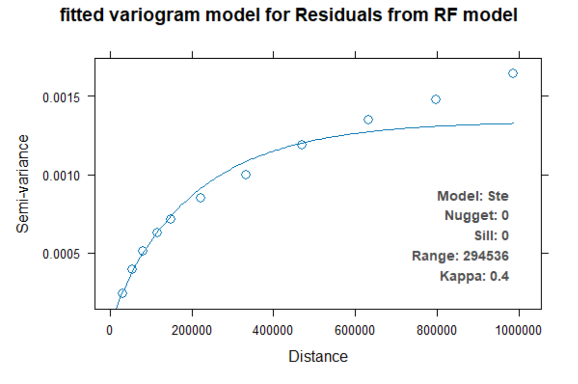
- Model Performance

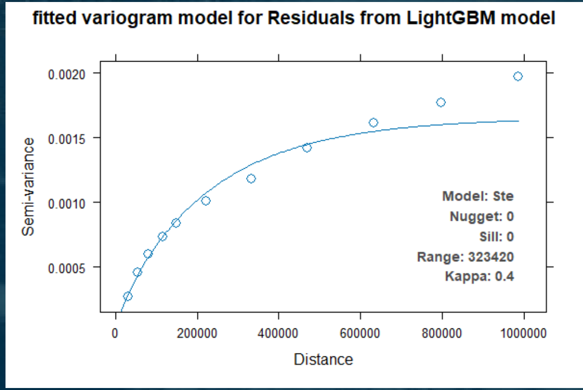


- Residuals $(y_{ESA\ CCI\ SM} - y_{predicted})$



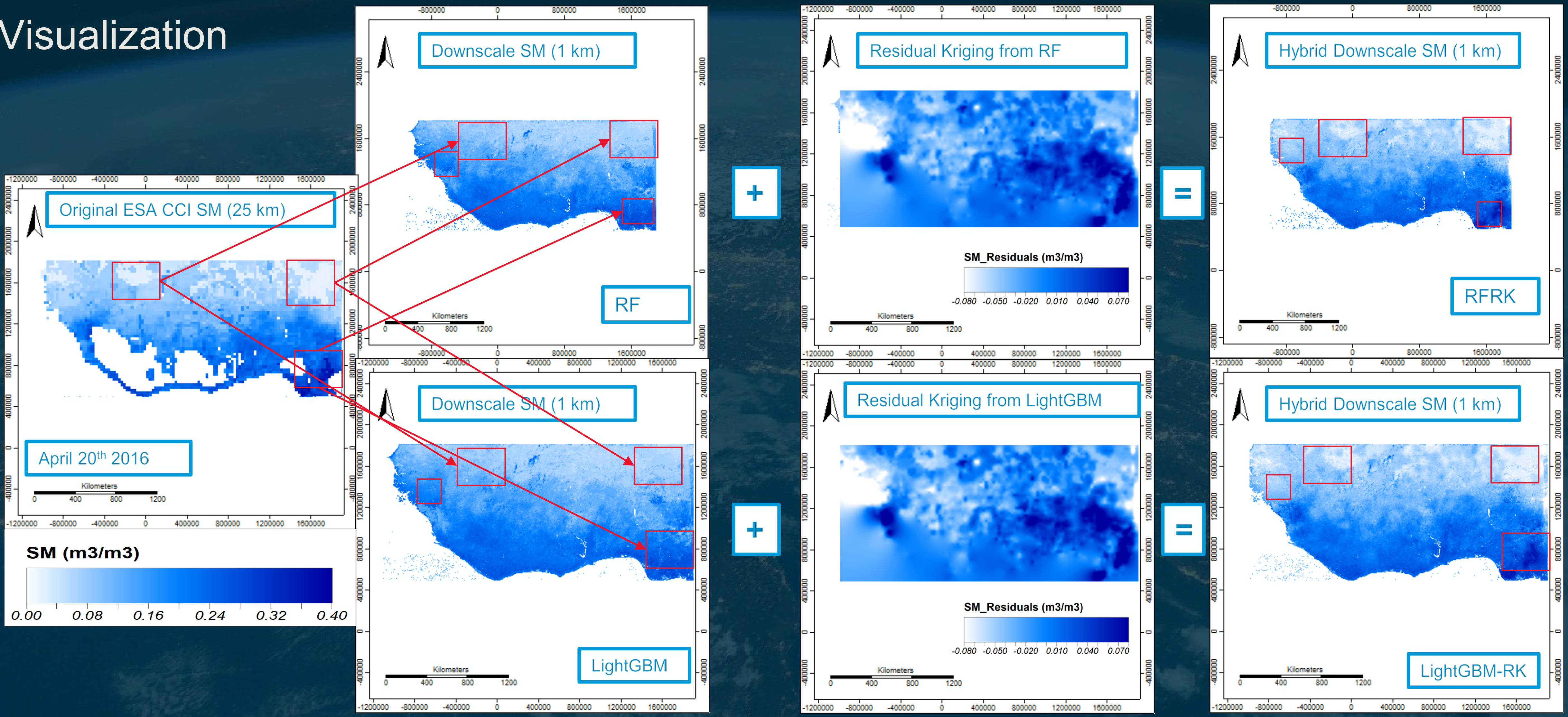
April 20, 2016





Results

- Visualization



Results

- Validation of Models with Original ESA CCI SM

	CCC	RMSE (m3/m3)	Bias (m3/m3)
LightGBM	0.86	0.04	-0.023
RF	0.86	0.04	-0.023
LightGBM-RK	0.98	0.017	-0.0005
RF-RK	0.97	0.017	-0.0011

- Validation of Models with Insitu data

	CCC	RMSE (m3/m3)	Bias (m3/m3)
Original ESA CCI SM	0.35	0.12	0.11
LightGBM	0.42	0.10	0.087
RF	0.43	0.10	0.087
LightGBM-RK	0.34	0.12	0.11
RF-RK	0.34	0.12	0.11

Take home points

- Combining ML models and geostatistical methods (Hybrid Model) leverages the spatial data of auxiliary datasets, improving prediction accuracy and reducing model error through residual interpolation with spatial autocorrelation.

- Integrating these methods into the downscaling framework corrects spatial errors and enhances the performance of original SM products.

- The ML models significantly improved the correlation of SM products (downscale SM) with in-situ measurements compared to the original ESA CCI SM product.

- The hybrid downscale SM follows with the original ESA CCI SM product with low correlation with the in-situ measurements.

Conclusion

- The hybrid model enhances prediction accuracy and reduces model error through residual interpolation.

- This approach corrects spatial errors in the downscaling framework and improves the performance of original soil moisture products. Although the correlation with ground measurements remains low due to the original data, the ML models' predicted soil moisture shows a higher correlation with in-situ measurements, proving the hybrid approach's effectiveness.