Improving the SMAP Daily Soil Moisture Time Series with Land Surface Model Datasets Using Power Spectrum-Adjustment Techniques

Nazanin Tavakoli[1] and Paul Dirmeyer[1,2]

[1]George Mason University
[2]Center for Ocean-Land-Atmosphere Studies (COLA)

January 16, 2024

Title: Improving the SMAP Daily Soil Moisture Time Series with Land Surface Model Datasets Using Power Spectrum-Adjustment Techniques

Nazanin Tavakoli[1](ntavakol@gmu.edu), Paul Dirmeyer[1,2]

[1]George Mason University, [2]Center for Ocean-Land-Atmosphere Studies (COLA)

Abstract:

Land-atmosphere feedbacks act through process chains that link variables in the land-atmosphere system. For the global energy and water cycles, the first link in the chain is soil moisture. Flux tower sites provide in-situ observations, including land surface states, surface fluxes, and near-surface atmospheric states, to validate these links; however, they are unevenly distributed over the globe. Therefore, to obtain a global view of observationally based land-atmosphere coupling metrics, satellite data are useful.

Among satellite products, the Soil Moisture Active Passive (SMAP) satellite provides the closest match to in-situ observations. However, SMAP exhibits stochastic random noise that can deflate coupling estimates. Since soil moisture variability closely follows a first-order Markov process, it typically has a distinct red noise spectrum. Satellite data with random noise has a whiter spectrum at high frequencies that can be compared to the expected red spectrum. Also, missing data in SMAP are not entirely random; its 8-day repeating polar orbit creates a cadence of missing data for both ascending and descending overpasses, depending on the location. This creates additional artifacts in the power spectrum, calculated through lagged autocovariance in the time series, with harmonic spikes at 8, 4 (8/2), 2 2/3 (8/3), and 2 (8/4) days that broaden due to the satellite's orbital variations. To be optimally useful for quantifying land-atmosphere feedbacks, the effects of random noise and periodic missing data must be minimized.

A power spectrum adjustment technique has been designed to remove the orbital harmonic spikes from Level 3 (L3) SMAP data. This is achieved by fitting and removing a catenary function to the power spectrum between harmonic spikes. This adjusted spectrum is then scaled to match surface layer soil moisture observations at sites of the AmeriFlux network (in-situ data), which exhibit relatively low noise and have spectra that are very similar to those produced by offline land surface models (LSMs). Utilizing validated spectral data from gridded LSM-based datasets, a global L3 SMAP product with removed noise and harmonic effects is being produced. We will present results quantifying the extent to which this technique improves SMAP data and its temporal correlation with observations.

Improving the SMAP Daily Soil Moisture Time Series with Land Surface Model Datasets Using Power Spectrum-Adjustment Techniques

Nazanin Tavakoli[1](ntavakol@gmu.edu), Paul A. Dirmeyer[1,2]
[1]George Mason University, [2]Center for Ocean-Land-Atmosphere Studies (COLA)

Introduction

Soil moisture (SM) influences land-atmospheric coupling by partly controlling the evaporation of accumulated precipitation water, thereby playing a significant role in global water, energy, and carbon cycles. Although flux tower sites offer accurate, observationally-based data for SM analysis and evaluation of land-atmosphere interactions, these stations are not evenly distributed globally. Consequently, satellite systems that accurately estimate SM variability hold great potential for predicting and managing water resources.
The **Soil Moisture Active Passive (SMAP)** satellite, aligning closely with *in-situ* observations, is a valuable resource in this regard. However, SMAP L3 data is prone to unpredictable random distortions, deviating from the expected fundamental Markov process that manifests as a distinct 'red noise' pattern in soil moisture variability. This noise can degrade the correlation of SM with other variables. This study addresses this scientific question:

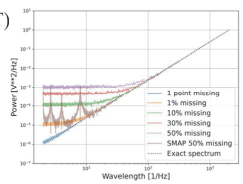
Is there a way that we can identify and remove the noise?

Effect of Random Missing Data

- Remove different percentages of points from idealized time series constructed of superposed waves

Fourier Decomposition won't work on Time series with gaps

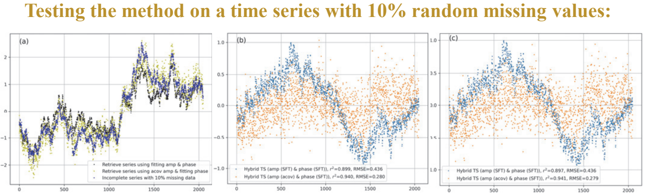
- Calculate Power Spectrum Density (PSD) based on Lagged Autocovariance (works with missing data)
- Get phases using "Slow" Fourier Transform (SFT) (Fitting Sine & Cosine to the Time Series)
- Retrieve the time series using the Backward Fourier Transform



Power spectrum density for different fractions of random missing data. SMAP 50% missing PSD has a pattern of 1-0-1-1-0-1-0-0

- The fraction of missing data has a systematic deviation at high frequencies.
- Random missing data superposes a whitening on the short end of the spectrum.
- SMAP has a distinct orbital pattern with an 8-day repeat cycle that depends on the location. This leads to power spectrum spikes in L3 data at wavelengths of 8, 4 (8/2), 2 2/3(8/3), and 2 (8/4), which are **orbital harmonics**.

Testing the method on a time series with 10% random missing values:



(a) Incomplete series with 10% missing data, invented series using phase and amplitude from SFT and lagged autocovariance, (b) deviation of invented series from complete, and (c) incomplete series.

- For **random time series (not having a SMAP pattern),** there is more possibility of the larger segment of missing data, so there will be more opportunity for the filling algorithm to be lost.
- Invented series that use the magnitude from lagged-autocovariance capture the original series more accurately.
- Correlation is picking up mostly large-scale patterns, where most energy lies, while small variations contribute very little to the correlation.
- Instead of correlation, checking day-to-day variation at different lags may be a more robust test to compare original and invented time series.

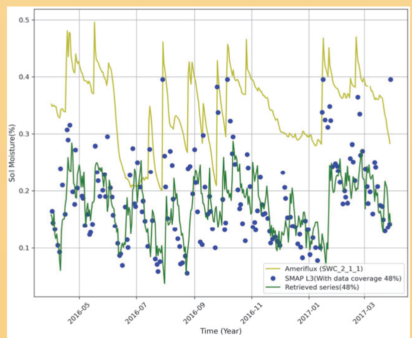
Spectrum-Adjustment Technique

At High Frequencies…

Impossible to Get Phases

Less Variability

Cannot Predict Dry-Downs



Daily soil moisture time series for SMAP L3, Ameriflux, and retrieved SMAP from 2016-04 to 2017-03 in ARM Southern Great Plains – Lemont.

View this poster online

Grants: 80NSSC20K1803
80NSSC21K1801

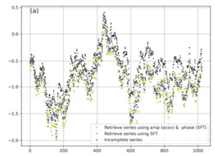
Future work

- Find a better solution to retrieve phases
- Apply this approach to more in-situ locations
- Produce a global gridded noise-reduced SMAP L3 product
- Apply this approach to CCI multi-platform SM product
- Calculate Coupling Metrics between retrieved SMAP and surface fluxes

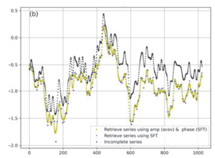
Checking day-to-day variation of series with 50% and 87.5% random-SMAP pattern missing data:

Lags	Fraction of common sign between complete and retrieved series using fitting amp & phase	Fraction of common sign between complete and retrieved series using acov amp & fitting phase
1	0.63	0.59
2	0.67	0.62
3	0.70	0.64
4	0.73	0.67
5	0.77	0.70
6	0.79	0.72
7	0.80	0.75
8	0.82	0.81

Lags	Fraction of common sign between complete and retrieved series using fitting amp & phase	Fraction of common sign between complete and retrieved series using acov amp & fitting phase
1	0.62	0.62
2	0.63	0.64
3	0.66	0.66
4	0.68	0.70
5	0.70	0.71
6	0.73	0.74
7	0.74	0.76
8	0.75	0.77

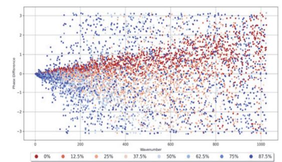


Right:
Incomplete series with 50% (a) and 87.5% (b) missing data, invented series using phase and amplitude from SFT and lagged autocovariance (acov).



Left:
Tables show how many times the sign of day-to-day changes matches between the complete and invented series.

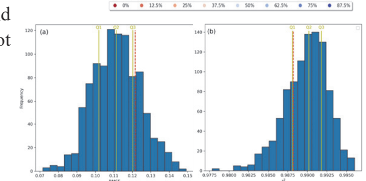
- Retrieved **random series with SMAP patterns** is significantly different from 50% since it will automatically filter out high-frequency phases.
- When phases are retrieved from SFT, in some frequencies, these phases and the real one have differences between $-\pi$ and π.



Phase Difference (Actual-SFT) for different percentages of missing data.

Non-parametric test to check whether this method can do gap-filling:

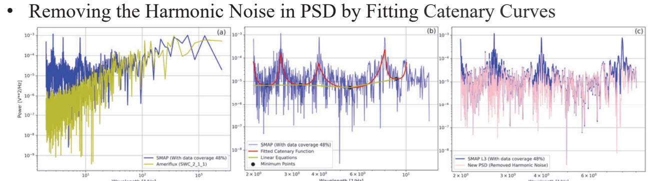
- Choose $\frac{k}{10}$ days randomly from 50% data used in the retrieval, find r^2 and RMSE for 1000 samples, and plot the distribution.



Distribution of (a) RMSE and (b) r^2 for invented series with 50% missing SMAP pattern values. RMSE and r^2 between invented and missing values from the complete series. Q1: first quantile, Q2: median, and Q3: upper quantile.

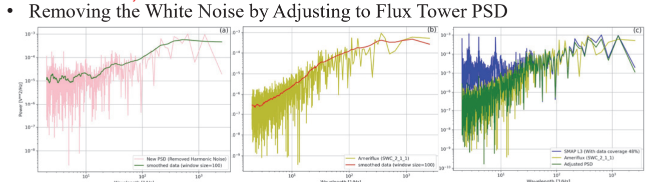
Real SMAP Data

- Removing the Harmonic Noise in PSD by Fitting Catenary Curves



(a) PSD for SMAP L3 and Ameriflux from 2015 to 2022 in ARM Southern Great Plains – Lamont, (b) fitted catenary to PSD and lines connected minimum points of each curve, (c) new PSD in which harmonic noise is removed by dividing lines to fitted catenary.

- Removing the White Noise by Adjusting to Flux Tower PSD



Removed harmonic noise PSD for SMAP L3 (a) and Ameriflux (b) with their smoothed spectra also plotted. (c) Adjusted PSD in which white noise is removed by dividing smoothed removed harmonic noise PSD to smoothed Ameriflux PSD and then removed harmonic noise PSD divided by this ratio.

38% of PSD is removing >>> Harmonic Noise
78% of PSD is removing >>> White Noise