Supplement of Nat. Hazards Earth Syst. Sci., 20, 2591–2607, 2020
https://doi.org/10.5194/nhess-20-2591-2020-supplement

Natural Hazards and Earth System Sciences
Open Access

Supplement of

Challenges in flood modeling over data-scarce regions: how to exploit globally available soil moisture products to estimate antecedent soil wetness conditions in Morocco

El Mahdi El Khalki et al.

Correspondence to: Yves Tramblay (yves.tramblay@ird.fr)

Extended collocation analysis:

An alternative technique to validate soil moisture products when ground truth is missing is the use of Triple Collocation (TC) analysis (Gruber et al. 2017). TC analysis requires the availability of three datasets with mutually independent errors and linear additive error model between the measurement systems and the unknown truth:

$$X = \alpha + \beta S + \varepsilon, \tag{1}$$

where X is the soil moisture estimate, S is the true soil moisture, α and β are additive and multiplicative biases, respectively. Eventually, ε is the zero-mean random error.

To build such a triplet, satellite and ground-based datasets can be combined with modeled soil moisture fields from reanalysis (e.g., ERA5). The reanalysis datasets ingest a number of satellite, atmospheric and ground observations which can potentially undermine their independence with respect to other members of the triplets. This creates doubts about the satisfaction of the null cross-correlation assumptions required to apply TC (Stoffelen, 1998). In a preliminary analysis (not shown), we used TC to characterize the error variance of the different soil moisture datasets by using different triplet combinations of the products. However, we observed substantial differences among the selected triplets likely due to error co-dependence. Based on that, we assumed the existence of non-null error cross correlation for the selected triplets (e.g. ERA5, SMOS and ASCAT).

When more than three products are available (i.e., N), the error can be estimated using an Extended Collocation (EC) approach (Gruber et al. 2017). The same assumptions for TC also apply for EC, but the number (N >3) datasets constitutes an over-constrained system, allowing the designation of N-3 non-zero error covariance terms which can be estimated with a least-squares solution (Pierdicca et al. 2015). Therefore, the zero TC assumption can be relaxed to allow non-zero correlation among N-3 data product pairs. For N = 4, the X, Y, Z, W measurement systems and assuming that non-zero EC exists only between X and Y, the least-squares solution for the QC problem is given by:

$$M = \begin{bmatrix} \sigma_X^2 \\ \sigma_Y^2 \\ \sigma_Z^2 \\ \sigma_W^2 \\ \sigma_{XY} \\ \sigma_{XZ}\sigma_{XW}/\sigma_{ZW} \\ \sigma_{YZ}\sigma_{YW}/\sigma_{ZW} \\ \sigma_{XZ}\sigma_{ZW}/\sigma_{XW} \\ \sigma_{YZ}\sigma_{ZW}/\sigma_{YW} \\ \sigma_{XW}\sigma_{ZW}/\sigma_{XZ} \\ \sigma_{YW}\sigma_{ZW}/\sigma_{YZ} \\ \sigma_{XZ}\sigma_{YW}/\sigma_{ZW} \\ \sigma_{XW}\sigma_{YZ}/\sigma_{ZW} \end{bmatrix} A = \begin{bmatrix} 1000010000 \\ 0100001000 \\ 0010000100 \\ 0001000010 \\ 0000100001 \\ 1000000000 \\ 0100000000 \\ 0010000000 \\ 0010000000 \\ 0001000000 \\ 0001000000 \\ 0000100000 \\ 0000100000 \end{bmatrix} S = \begin{bmatrix} \beta_X^2\sigma_T^2 \\ \beta_Y^2\sigma_T^2 \\ \beta_Z^2\sigma_T^2 \\ \beta_W^2\sigma_T^2 \\ \beta_X\beta_Y\sigma_T^2 \\ \sigma_{\varepsilon_X}^2 \\ \sigma_{\varepsilon_Y}^2 \\ \sigma_{\varepsilon_Z}^2 \\ \sigma_{\varepsilon_W}^2 \\ \sigma_{\varepsilon_X\varepsilon_Y} \end{bmatrix}, \tag{2}$$

where σ_T^2 is the true soil moisture variance, σ_ε^2 is the variance of the random error, and σ_(ε_X ε_Y) is the error covariance between X and Y.

And the least squares solution for the parameters in S is given as:

$$\hat{S} = (A^T A)^{-1} A^T M, \tag{3}$$

Which provide the error variance of each dataset as long as the error covariance terms. More details on the method and its mathematical derivation can be found in Gruber et al. (2017).The error variance provided by EC can also be expressed in normalised form as Signal-to-Noise Ratio (SNR). This overcomes the dependency on the chosen scaling reference and allows to compare the error variances between the data sets. SNR is usually given in decibel, which can be easily interpreted: a value of zero means that the signal variance is equal to the noise variance, and every 3dB increase(decrease) implies a doubling (halving) of the signal variance compared to the noise variance. The SNR (expressed in dB) can be computed using the following formulation:

$$SNR[db] = 10 \log \frac{\beta_i^2 \sigma_\theta^2}{MSE_i}, \tag{4}$$

with i, j in [X, Y, Z] and i ≠ j.

In some special cases, the MSEi can become negative and the SNR cannot be expressed in dB (logarithm of a negative number is undefined). The reason is that the relation of the covariances between the data sets become larger than the actual signal variance (e.g. #XY #XZ/#Y Z > #2X), which can be related numerical problems, wrong estimation of the covariances or a violation of the underlying assumptions of the error model in general.In our study we used two different configurations of the EC techniques. In particular, for the Issyl basin no in situ observations are available so we used quadruple collocation analysis with quadruplets constructed with ASCAT, SMOS, ERA5 and SMA and ASCAT, SMOS-IC, ERA5 and SMA. The choice of these quadruplets was based on the assumption of non-zero correlation between SMOS products and ERA5 so in the process we also estimated σ_(SMOS-ERA) (not shown). Similarly, for Rheraya we applied the methods by using five different datasets and assuming SMOS and ERA products and SMA and in situ observations characterized by non-null error cross-correlations. For both basins we used either SMOS or SMOS-IC in the configurations.

Results:

Table 1 shows the results obtained for the two basins and two configurations. For Issyl, it can be seen that SMOS-IC is the best performing product with SNR much larger 3DB, followed by ASCAT and SMA. Conversely ERA5 and SMOS are suboptimal having noise variance similar to the signal variance. For Rheraya SMOS-IC is the only product providing SNR>3DB followed by SMOS and ERA5 which are however are still suboptimal. Poor results are found for both SMA, in situ and ASCAT in this catchment. Overall, the results of this complementary analysis confirm the findings of previous sections.

Table 1: Signal to noise ratio for Rheraya and Issyl basins. The SNT = 0 : Error variance, SNR > 3 Signal variance double the noice variance (very good) and SNR < 3 Signal variance half noice variance (not good).

	ASCAT	SMOS	SMOS-IC	ERA5	SMA
Rheraya	-5.55		7.54		-1.99
	-6.16	4.31		1.16	-1.10
Issyl	4.23	1.90		2.33	5.03
	4.28		8.12	2.33	4.99

References:

Gruber, A., Dorigo, W. A., Crow, W. and Wagner, W.: Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals, IEEE Trans. Geosci. Remote Sens., 55(12), 6780–6792, doi:10.1109/TGRS.2017.2734070, 2017.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, J. Geophys. Res. Ocean., 103(3334), 7755–7766, doi:10.1029/97jc03180, 1998.