



# Impact of Rescaling Approaches in Simple Fusion of Soil Moisture Products

**DOI:**

[10.1029/2019WR025111](https://doi.org/10.1029/2019WR025111)

**Document Version**

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**

Afshar, M., Yilmaz, M. T., & Crow, W. T. (2019). Impact of Rescaling Approaches in Simple Fusion of Soil Moisture Products. *Water Resources Research*. <https://doi.org/10.1029/2019WR025111>

**Published in:**

Water Resources Research

**Citing this paper**

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



## Impact of Rescaling Approaches in Simple Fusion of Soil Moisture Products

M. H. Afshar<sup>1,2</sup>, M. T. Yilmaz<sup>2</sup>, and W. T. Crow<sup>3</sup>

<sup>1</sup>Department of Mechanical, Aerospace and Civil Engineering, University of Manchester, Manchester, UK

<sup>2</sup>Civil Engineering Department, Orta Dogu Teknik Universitesi, Ankara, Turkey

<sup>3</sup>USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA

Corresponding author: Mehdi H. Afshar ([mehdi.hafshar@gmail.com](mailto:mehdi.hafshar@gmail.com))

### Key Points:

- Precision of the fused products obtained using suitable linear rescaling methods is similar to the ones obtained using nonlinear methods.
- Selection of better reference dataset yield more precise fused product.
- Application of a smooth-deviance decomposition rescaling technique improves correlations.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1029/2019WR025111](https://doi.org/10.1029/2019WR025111)

## Abstract

In this study, the impact of various rescaling approaches in the framework of data fusion is explored. Four different soil moisture products (Advanced Scatterometer, ASCAT; Advanced Microwave Scanning Radiometer for EOS, AMSR-E; Antecedent Precipitation Index, API; Global Land Data Assimilation System, GLDAS-NOAH) are fused. The systematic differences between products are removed before the fusion utilizing various rescaling approaches focusing on different: methods (regression-, variance/CDF matching-, MARS-, and SVM-based), stationarity assumptions (constant or time-varying rescaling coefficients), and time frequency techniques (periodic or non-periodic high and low frequency components). Given that statistical descriptions (e.g., standard deviation, correlation coefficient) of reference datasets are utilized in rescaling approaches, the precision of the selected reference dataset also impacts the final fused product precision. Experiments are validated over 542 soil moisture monitoring sites selected from the International Soil Moisture Network datasets between 2007 and 2011. Overall, results highlight the importance of reference dataset selection - particularly that a more precise reference product yields a higher precision fused soil moisture product. This conclusion is neither sensitive to the number of fused products nor the rescaling procedure. Among rescaling approaches, the precision of fused products is most affected by the choice of rescaling stationary assumption and time frequency decomposition technique. Variations in rescaling methods have only a small impact on the precision of pair fused products. In contrast, utilizing a time-varying stationary assumption and non-periodic decomposition technique produces correlation improvements of 0.07 [-] and 0.02 [-], respectively, versus the other widely implemented rescaling approaches.

## 1 Introduction

Soil moisture is a critical environmental parameter that significantly regulates the terrestrial carbon, energy, and water cycles [Koster *et al.*, 2004]. Therefore, describing the spatial distribution and temporal changes of soil moisture contributes to the development of accurate climate, ecological and hydrological models at local, regional and global scales [Dorigo *et al.*, 2012].

Soil moisture can be estimated through multiple different methods (i.e., in-situ measurements, numerical modeling, and remote sensing). Direct monitoring methods, such as gravimetric sampling [Cosh *et al.*, 2016] and dielectric probes, provide soil moisture information at very fine spatial resolutions ( $< 10 \text{ cm}^2$ ) and high temporal sampling (e.g., every 30 min or 1 h). Even though such observations are critical for the satellite mission validation efforts, their use is often impractical for studies focusing on large spatial areas [Bulut *et al.*, 2019]. Instead hydrological model- or satellite remote sensing-based soil moisture datasets are commonly used for coarse-scale applications related such as drought monitoring [Afshar *et al.*, 2016], crop yield monitoring [Anderson *et al.*, 2015; Anderson *et al.*, 2016; Mladenova *et al.*, 2017], and

improvement of hydrological models via data assimilation [Houser *et al.*, 1998; Crow *et al.*, 2003; Yilmaz *et al.*, 2011; Lievens *et al.*, 2015].

Owing to the large number of studies using these datasets in coarse-scale applications, many soil moisture products have been produced using hydrological model- and remote sensing-based methods. Complex hydrological models (e.g., Mosaic, NOAH, the Community Land Model, and the Variable Infiltration Capacity) and much simpler hydrological models (e.g., the Antecedent Precipitation Index, API) offer spatiotemporally continuous soil moisture datasets. However, these models rely heavily on parameters that vary significantly in space and time and are often impractical to estimate. Remote sensing approaches also provide spatially continuous soil moisture datasets (e.g., the Advanced Scatterometer [ASCAT; Wagner *et al.*, 1999], Soil Moisture and Ocean Salinity [SMOS; Kerr *et al.*, 2001], and Soil Moisture Active Passive [SMAP; Entekhabi *et al.*, 2010] missions). However, these datasets reflect conditions within only the top layer of the soil (~ 3-5 cm), typically have temporal and spatial resolutions that are coarser than models, and rely on retrieval algorithms that depend on multiple land surface parameters (related to, e.g., land cover, topography, and radiative activities).

Given the common availability of multiple soil moisture datasets for the same time location, and the reality that all products have unique strengths and shortcomings, it is often desirable to fuse multiple products into an integrated estimate. Efforts to fuse satellite- and model-based soil moisture estimates are implemented using either simple merging methodologies [Liu *et al.*, 2011; Liu *et al.*, 2012; Wagner *et al.*, 2012; Gruber *et al.*, 2017; Dorigo *et al.*, 2017; Yilmaz *et al.*, 2012; van der Schalie *et al.*, 2017; van der Schalie *et al.*, 2018] or more complex data assimilation methodologies [Houser *et al.*, 1998; Reichle & Koster, 2005; Gevaert *et al.*, 2018].

However, before such merging methodologies can be implemented, systematic differences between soil moisture estimates obtained from different platforms and/or sensors must be rectified [Dirmeyer *et al.*, 2004; Reichle & Koster, 2004; Reichle & Koster, 2005; Yilmaz & Crow, 2013; Yin *et al.*, 2014; Su & Ryu, 2015; Afshar & Yilmaz, 2017]. To accomplish this, a number of different rescaling methods are commonly applied [Afshar & Yilmaz, 2017]. In all approaches, one dataset is selected as a reference, and the statistical moments of all other products are tuned to this reference dataset. The specific goals of this rescaling vary but can include: minimizing the variability of differences between the rescaled product and the reference product, maximizing the correlation between them or matching the total variability of an unscaled product to an arbitrary reference dataset [Hain *et al.*, 2011; Miralles *et al.*, 2011; Parinussa *et al.*, 2011; Scipal *et al.*, 2008; Stoffelen, 1998; Zwieback *et al.*, 2012; Liu *et al.*, 2011]. Based on above mentioned goals, rescaling methods can be either linear [e.g., first order linear regression (REG), variance matching (VAR), etc.] or nonlinear [e.g., CDF matching (CDFM) and nonlinear techniques, multi adaptive regression splines (MAR), support vector machines (SVM), artificial neural networks (ANN)] in form.

Once a particular rescaling method (e.g., VAR, REG, CDFM, MAR, SVM, ANN) is selected for a specific application, this method can be implemented using different approaches that consider parameter variability at a range of different time scales [Yilmaz *et al.*, 2016]. For example, a rescaling method can be tuned for the entire time series or, alternatively, for each month separately (that is, rescaling coefficients or product dependencies are assumed either to be constant or time-varying). If the reference product, that all other products are rescaled to, is more precise than the unscaled product, then the implementation of rescaling methods using time-varying coefficients (e.g., using monthly rescaling coefficients rather than using a single coefficient for the entire time series) may yield a more precise rescaled product than implementation of them using constant coefficients. The inverse is also true that in the case of a relatively imprecise reference product. That is, a weakly rescaled product (using, for example, only a single set of long-term rescaling parameters) will yield a more precise rescaled product [Yilmaz *et al.*, 2016].

Given that rescaling approaches typically use reference dataset statistics to minimize systematic differences between datasets, the precision (i.e., high temporal correlation with true soil moisture values [Koster *et al.*, 2009; Entekhabi *et al.*, 2010]) of the reference dataset selection can impact the rescaled product precision via the statistics estimated in the rescaling process (details of this impact are presented below in section 2.4). Even though earlier studies have investigated the performance of different rescaling methodologies [Yilmaz & Crow, 2013; Yilmaz *et al.*, 2016], no study has specifically investigated the impact of the reference dataset selection on rescaled product precision. Additionally, no study has investigated the impact of more complex, nonlinear rescaling techniques like SVM and MARS [Afshar & Yilmaz, 2017] in a data-fusion methodology.

Full data assimilation methodologies tend to be based on time-variant model error covariances corresponding to the moment when observations are ingested sequentially. In contrast, simple fusion methodologies often use a constant error variance (i.e., constant weight) assumption and can be implemented during the post-processing of datasets rather than on-line at each assimilation time step. Yilmaz & Crow [2013] investigated the optimality of the rescaling techniques using a simple API model and found that triple collocation-based rescaling consistently yielded better results than a regression-based rescaling methodology. The optimality difference between the two methodologies may be enhanced when datasets are rescaled to each other's space via nonlinear rescaling approaches. Hence, we expect differences to emerge between the optimality of the rescaling methodologies for complex data assimilation (i.e., TCA; [Yilmaz & Crow, 2013]) versus simpler merging methodologies.

The optimality of rescaling methods [Afshar & Yilmaz, 2017] depends in part on the goal of the rescaling methodology [Yilmaz & Crow, 2013]; however, the performance of such methods is also affected by the degree to which underlying assumptions of the rescaling methods are met [Yilmaz & Crow, 2013] and by their rescaling method/approach selection [Yilmaz *et al.*,

2016]. Hence, further investigation of the impact of various rescaling approaches over the rescaled/fused product precision is necessary for applications utilizing these rescaling approaches. Earlier studies [Yilmaz & Crow, 2013; Yilmaz *et al.*, 2016] decomposed time series into high- and low-frequency components with the assumption that low-frequency components are annually periodic (i.e., do not have inter-annual variability), while a relaxed assumption (i.e., where high- and low-frequency components could be non-periodic) has not been investigated before in a data merging methodology. In addition, while the impact of various rescaling methodologies (e.g., VAR, REG, CDFM) on simple merging framework has been previously investigated in a few studies [Anderson *et al.*, 2012; Yilmaz *et al.*, 2012; Dorigo *et al.*, 2017], the impact of rescaling stationarity assumptions (i.e., time varying versus constant coefficients) and time-frequency techniques (i.e. low versus high frequency decomposition) selection has not been investigated for simple merging methodologies.

Accordingly, the goal of this study is to investigate the impact of: 1) reference dataset selection, 2) rescaling methodology/stationarity/technique selection, 3) relaxed low-frequency decomposition assumptions, and 4) the SVM and MARS rescaling methods selection on fused-product skill in a simple soil moisture merging methodology. The methodologies and the datasets used in this study are given in the second section. Results are presented and discussed in the third section, and concluding remarks are provided in the last section.

## 2 Rescaling

Soil moisture time series (e.g.,  $X$  and  $Y$ ) are commonly represented via a linear, additive model as a combination of their mean, true anomaly, and random errors:

$$X = \mu_X + \alpha_X T + \epsilon_X \quad (1)$$

$$Y = \mu_Y + \alpha_Y T + \epsilon_Y \quad (2)$$

where  $\mu_X$  and  $\mu_Y$  are the mean values of  $X$  and  $Y$ ;  $T$  is the true soil moisture product;  $\alpha_X$  and  $\alpha_Y$  are the scaling factors of  $X$  and  $Y$  to the space of  $T$ ;  $\alpha_X T$  and  $\alpha_Y T$  are the real signal components of  $X$  and  $Y$ , and  $\epsilon_X$  and  $\epsilon_Y$  are the zero-mean random errors (i.e., the noise components) of  $X$  and  $Y$ .

In data fusion, the goal is to obtain a fused child product ( $F$ ) with a lower noise component (i.e., error variance) than the parent products  $X$  and  $Y$ . To reach this goal, the differences between the signal components of pair products need to be minimized before application of a particular data fusion methodologies. This is often achieved by multiplying the pair products (e.g.,  $Y$ ) by a linear coefficient that is found by utilizing statistical descriptions (e.g., mean, and standard deviation) of a reference product (e.g.,  $X$ ):

$$Y^* = \mu_X + (Y - \mu_Y)c_Y \quad (3)$$

where  $Y^*$  is the rescaled form of  $Y$ , and  $c_Y$  is the linear scaling coefficient. Here, it is stressed that nonlinear rescaling methodologies exist as well; however, the linear form is shown here for

brevity. Once the products are transformed into the same space (using the different rescaling methodologies elaborated below), data fusion can be expressed in its form as:

$$F = w_X X + w_{Y^*} Y^* \quad (4)$$

where  $w_X$  and  $w_{Y^*}$  are weights for  $X$  and  $Y^*$ . Here, different type of weights (e.g., time dependent or constant) can be used for  $w_X$  and  $w_{Y^*}$  where the process of optimization of weights may involve different evaluation criteria. In addition, two different approaches are used for weighting of the soil moisture products. The first approach is based on simple merging which assigns equal weighting to each rescaled product. Our second approach uses triple collocation analysis (TCA) to assign (different) weights to the rescaled products as following:

$$w_X = \frac{\sigma_{\varepsilon_Y}^2}{\sigma_{\varepsilon_X}^2 + \sigma_{\varepsilon_Y}^2} \quad (5)$$

$$w_Y = \frac{\sigma_{\varepsilon_X}^2}{\sigma_{\varepsilon_X}^2 + \sigma_{\varepsilon_Y}^2} \quad (6)$$

where  $w_X$  and  $w_{Y^*}$  are the weights for  $X$  and  $Y$ , and  $\sigma_{\varepsilon_X}^2$  and  $\sigma_{\varepsilon_Y}^2$  are TCA-derived error variances for the  $X$  and  $Y$  products. In the TCA-based weighting approach, a third collocated independent product is required to derive error variances. In soil moisture studies this is often acquired from a land surface model. Here, surface soil moisture acquired from the NOAH land surface model is used for the third product. It is also used as the reference to derive weights for product pair to be fused. Based on TCA, error variances for the  $X$  and  $Y$  products can be found in the form of:

$$\sigma_{\varepsilon_X}^2 = \sigma_X^2 - \frac{\sigma_{X,Y}\sigma_{X,R}}{\sigma_{Y,R}} \quad (7)$$

$$\sigma_{\varepsilon_Y}^2 = \sigma_Y^2 - \frac{\sigma_{Y,X}\sigma_{Y,R}}{\sigma_{X,R}} \quad (8)$$

where  $\sigma_X^2$  and  $\sigma_Y^2$  are the variances of  $X$  and  $Y$  products;  $\sigma_{X,Y}$  and  $\sigma_{Y,X}$  are the covariance between  $X$  and  $Y$  products, and  $\sigma_{X,R}$ , and  $\sigma_{Y,R}$  are the covariance between pair products of  $X$  and  $Y$  that are aimed to be fused over space of reference product which is denoted with subscript ( $R$ ). For more information about this approach, readers are referred to the study of *Gruber et al.* [2017].

## 2.1 Rescaling methods – linear and nonlinear methods

There are several different linear and nonlinear rescaling methods commonly applied in studies focusing on removing systematic differences between soil moisture products. Among them, the CDFM [*Reichle & Koster, 2004*] is arguably the most common. In addition, linear regression- [*Crow & Zhan, 2007*], variance matching- [*Draper et al., 2009*], TCA- [*Yilmaz & Crow, 2013*], copula- [*Leroux, et al., 2014*], wavelet- [*Su & Ryu, 2015*], quadratic polynomial- [*Zwieback et al., 2016*], and GP-, MAR-, SVM-, ANN- [*Afshar & Yilmaz, 2017*] based methods have been also proposed. Inter-comparison of these rescaling methods have demonstrated that

MARS and SVM result in more precise rescaled products relative to other methods [Afshar & Yilmaz, 2017]. Accordingly, in addition to the linear rescaling methods (REG and VAR), three nonlinear methods (CDFM, MARS, and SVM) are also applied here. Since errors in the statistics (e.g., mean, standard deviation, correlation, covariance, etc.) underlying these rescaling methods depend, in part, on the precision of  $X$  and  $Y$  (for details, see below in section 2.4), the selection of a better-quality reference dataset may yield more accurate statistics, and thus, more precise fused products.

### 2.1.1 Regression- and variance-based rescaling methods

Methods based on the assumption that there is a linear relation between unscaled and reference soil moisture products, can be implemented with  $c_Y$  (eq. 3) derived via either a linear regression (REG) or variance matching (VAR) approach [Yilmaz & Crow, 2013]:

$$c_Y^R = |\rho_{XY}| \sigma_X / \sigma_Y \quad (9)$$

$$c_Y^V = \sigma_X / \sigma_Y \quad (10)$$

where  $X$  is selected as the reference;  $Y$  is rescaled to the space of  $X$ ;  $c_Y^R$  and  $c_Y^V$  are the rescaling factors for the REG and VAR methods, respectively;  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of  $X$  and  $Y$ , respectively, and  $\rho_{XY}$  is the correlation coefficient between  $X$  and  $Y$ .

For the case where the relation between  $X$  and  $Y$  are represented using linear methods, it is trivial to analytically show such relations. It is more challenging when nonlinear methods are used. Nevertheless, we attempt to express such relationships using nonlinear methods described below.

### 2.1.2 CDF matching

CDFM [Reichle & Koster, 2004; Afshar & Yilmaz, 2017] is one of the earliest and most commonly used approach in soil moisture rescaling studies. The aim of this method is to eliminate differences between all the statistical moments of two soil moisture datasets. To this end, CDFM simply matches the CDF of unscaled product (e.g.,  $Y$ ) to the CDF of the reference product ( $X$ ). There are different ways available to use CDFM in order to remove the systematic differences between unscaled and reference soil moisture datasets [Zwieback *et al.*, 2016]. In this study, the CDFM method is applied through the calculation of the CDF for unscaled and reference soil moisture products based on ranks of their observations and conveyed into the space of reference products through the inverse relation between real observation and the CDF of the reference product. For more details about CDFM and its application readers are referred to Afshar & Yilmaz, [2017].



### 2.1.3 Multivariate adaptive regression splines

MARS [Friedman, 1991] is an advanced form of stepwise regression that makes no assumptions regarding to the functional relationship between dependent and independent variables (here unscaled and reference products). Instead MARS uses a series of basis functions to model nonlinearities in the relationship between the independent and dependent variables. The principle of MARS is to split the unscaled product's space into distinct intervals and fit an individual spline (basis function) to each interval separately. The final model is built by connecting these basis functions at knot points looked at the endpoint of individual intervals. The general MARS model of reference ( $X$ ) product with  $M$  basis functions can be written as:

$$Y^* = a_0 + \sum_{m=1}^M a_m B_m(Y) \quad (11)$$

where  $a_0$  is a constant coefficient;  $a_m$  is the coefficient of the  $m^{th}$  basis function;  $B_m(Y)$  is the  $m^{th}$  basis function in the form of  $\max(0, Y - t)$  or  $\max(0, t - Y)$  with a knot occurring at value  $t$ , and  $M$  is the number of basis functions built in the model. The MARS algorithm consists of separate forward and backward stepwise procedures. In the forward phase, the model adds basis functions and tries to find potential knots to reduce errors between rescaled and reference products in terms of mean square error (MSE), resulting in a complicated and (likely) over-fitted model. In the backward phase, the MARS model prunes the least effective terms among the previously added basis functions based on a generalized cross-validation (GCV) measure of MSE. The GCV procedure determines which basis function to keep in the model and which to eliminate by introducing a penalty to the system based on the number of terms (including intercepts) over the maximum number of terms allowed to be remained in pruned model (this threshold is semi-automatically calculated based on the number of variables). Here, the fitting phase of MARS to the unscaled and reference products is conducted in the R programming environment by using the Earth package [Miborrow, 2016]. For more information about MARS, see Hastie et al. [2009] and Sharda et al. [2008].

### 2.1.4 Support vector machines (SVM)

The SVM [Vapnik & Chervonenkis, 1974; Vapnik, 1998] is a technique based on statistical learning theory that uses the principle of structural risk minimization [Hernández et al., 2009]. In the regression model of SVM, a function associated with the dependent variable (here  $X$ ), which itself is a function of independent variable (here  $Y$ ), is estimated [Olson & Delen, 2008]. As with other rescaling methods, it is assumed that the relationship between  $X$  and  $Y$  (which are assumed to be independent and identically distributed products) can be characterized as an algebraic function in the form:

$$f(y) = W^T \varphi(y) + b \quad (12)$$

$$Y^* = f(y) + noise \quad (13)$$

where  $Y^*$  is the rescaled product;  $\phi(y)$  is a nonlinear mapping of the unscaled soil moisture products, and  $W$  and  $b$  are the weights and the bias values of regression function, respectively. These values are determined by minimizing the objective function of:

$$\min_{w,e,b} j(w, e) = \frac{1}{2} W^T w + \frac{\gamma}{2} \sum_{i=1}^N e_i^2 \quad (14)$$

subject to

$$e_i = X_i - Y_i^*, \quad i = 1, 2, \dots, N \quad (15)$$

where  $\gamma$  is the real positive number that is used for penalizing an occurred error during calibration;  $e_i$  is the amount of error at time step  $i$ ;  $X_i$  is the reference product at time step  $i$ , and  $N$  is the number of observations. The SVM solves this minimization problem via the Lagrange multipliers method and transforms it into the form:

$$f(y) = \sum_{i=1}^N a_i K(y, y_i) + b \quad (16)$$

where  $a_i$  is average of the Lagrange multipliers, and  $K(y, y_i)$  is the kernel function that can be written as an inner product in a feature space by following Mercer's theorem. There are a number of different types of kernel functional forms (e.g., linear, polynomial, radial basis, and sigmoid) that the SVM method can utilize. Here, the radial basis kernel function type was chosen based on results in *Pasolli et al.* [2011] and *Afshar & Yilmaz* [2017]. The above-mentioned optimized problems are solved with the E1071 package [*Meyer et al.*, 2015] in the R programming environment. Parameters of the kernel functions are found based on cross-validation (obtained optimal parameter values are not shown). For more information about the SVM and its technical details, readers are referred to the studies of *Vapnik* [1998] and *Smola and Scholkopf* [2004].

## 2.2 Rescaling stationarity – use of time-varying coefficients

The use of time-varying rescaling coefficients can improve the precision of a rescaled time series [*Yilmaz et al.*, 2016] owing to nonstationary relationships that may exist between the products. Accordingly, the rescaling of soil moisture time series can be implemented using either constant or time-varying rescaling coefficients (recognized as relatively less and more aggressive approaches, respectively [*Yilmaz et al.*, 2016]). Most studies use a less-aggressive rescaling approach, while more aggressive rescaling methodologies apply rescaling coefficients that vary periodically in time (e.g., fitting 12 different rescaling coefficients separately for each month of the year). Accordingly, two different rescaling stationarity selections are applied here: i) a single rescaling coefficient applied constantly in time and ii) rescaling coefficients that vary monthly. For the time-varying case, all soil moisture values obtained for a given month are rescaled against corresponding soil moisture values of the reference dataset for the same month-of-year (during all years of the data period). The process is then repeated separately for all 12 months to form a continuous monthly time series. Note that the application of different monthly rescaling

coefficients generally creates spurious temporal breaks in the resulting rescaled soil moisture time series.

Here the errors of the statistics (e.g., mean, standard deviation, correlation, and covariance) used in the calculation of time-varying coefficients depend on the precision of  $X$  and  $Y$ . This is particularly true when the datasets are subdivided to calculate various numbers of time-varying rescaling coefficients (due to the associated reduction in sample size). Hence, as with the case of rescaling method selection, the selection of a better-quality reference dataset will result in the more accurate sampling of time-varying rescaling coefficients.

### 2.3 Rescaling techniques - seasonality-anomaly vs. smooth-deviance decomposition

As discussed above, rescaling methodologies are commonly implemented for the entire soil moisture time series. However, under the assumption that the rescaling coefficients of soil moisture time series can vary for the different time frequency components, rescaling methods can also be implemented separately for each component [Su & Ryu, 2015]. In such an approach, the time series is decomposed into its frequency components and rescaling methods are applied separately to each component using different rescaling coefficients ( $c_{Y_j}$ ). Next, for the case of different rescaling coefficient scenario and linearly rescaling, these components are summed to derive a final rescaled soil moisture product as:

$$Y = \sum_{j=1}^k Y_j \quad (17)$$

$$Y^* = \sum_{j=1}^k Y_j^* \quad (18)$$

$$Y_j^* = \mu_{X_j} + (Y_j - \mu_{Y_j})c_{Y_j} \quad (19)$$

where  $Y^*$  and  $Y$  are decomposed in to their different frequency components  $Y_j^*$  and  $Y_j$ , respectively (e.g.,  $Y_j$  is the  $j^{th}$  component of  $Y$ ), and  $k$  represents the total number of frequency components. Although some studies have utilized a large number of frequency components (e.g., 8 components in the study of [Su & Ryu, 2015]), most studies have adapted a simply, two-frequency component scenario [Yilmaz & Crow, 2013; De Lannoy & Reichle, 2016]. Accordingly, this study also adapts a two-frequency component solution.

Such high- and low-frequency decomposition can be done in several different ways. In many studies, the low-frequency component is assumed to be seasonally periodic [Yilmaz & Crow, 2013] and to lack inter-annual variability (i.e., calculated as the expected soil moisture value within a moving-window centered on a particular day-of-year). Alternatively, the low-frequency component can be also calculated as non-periodic and containing inter-annual variability. Once the low frequency component is acquired, the high-frequency component can be obtained from subtracting the low-frequency component from the (complete) native time series:

$$Y_{i,j}^{Low-Periodic} = \frac{\sum_{i=1}^n \sum_{j-14}^{j+14} Y_j}{29n} \quad (20)$$

$$Y_{i,j}^{Low-NonPeriodic} = \sum_{j-14}^{j+14} Y_{i,j} W_j \quad (21)$$

$$Y_{i,j}^{High} = Y_{i,j} - Y_{i,j}^{Low} \quad (22)$$

$$VR^{High} = \frac{var(Y^{High})}{var(Y)} \quad (23)$$

$$VR^{Low} = \frac{var(Y^{Low})}{var(Y)} \quad (24)$$

where  $i$  indexes a particular year (within total  $n$  years) and  $j$  the day of the year (DOY). The low frequency component of the products are represented by the  $Y_{i,j}^{Low}$  term for product  $Y$ . It can be calculated using either a ( $Y_{i,j}^{Low-Periodic}$ ; hereinafter called “seasonality”) or non-periodic ( $Y_{i,j}^{Low-NonPeriodic}$ ; hereinafter called “smooth”) approach. The term  $Y_{i,j}^{High}$  refers to the high-frequency time series component (hereinafter called “anomaly” and “deviance” for periodic and non-periodic cases, respectively). Moreover,  $W_j$  is the weight to be used for window-averaging on DOY  $j$ . The variability of low and high frequency components are represented by the terms of  $VR^{Low}$ , and  $VR^{High}$  (the ratio of the variance of low/high components to the variance of entire series). Note that, for this study, the  $Y_{i,j}^{Low}$  and  $Y_{i,j}^{High}$  components of  $Y$  vary on a daily timescale; however, another timescale could be easily implemented for a different application.

The seasonality of the dataset ( $Y$ ) for any DOY is found as the average of a 29-day moving average window centered on a specific DOY and sampling across all available years of the dataset. However, the smooth component has been calculated by passing a smoothing filter over the time series using a weighted moving average window centered on a time step.

Daily weights for the calculation of the average value for any particular window differ in equations (20-21). While the seasonality/anomaly decomposition assumes equal weighting for the available days in any given window for any given DOY, the smooth/deviance decomposition assumes varying weights for any given DOY. Given soil moisture has high auto-correlation and is commonly over-sampled in time, an inverse relation is assumed here for daily weights based on their temporal distance from the center point of the window. Thus, days closer to the center of the 29-day window are assigned more weight than days towards the edges of the window:

$$\begin{cases} C_j = \left| \frac{1}{j} \right| & -14 < j < 0 \\ C_j = 1 & j = 0 \\ C_j = \frac{1}{j} & 0 < j < 14 \end{cases} \quad (25)$$

$$W_j = \frac{C_j}{\sum_{j-14}^{j+14} C_j} \quad (26)$$

where  $j$  refers to the day in the 29-day window;  $C_j$  is a coefficient which relates the day  $j$  to its weight ( $W_j$ ), and the time domain is between [-14 to +14] days.

Here, three different rescaling techniques are used for temporal decomposition selection (i.e., no decomposition, seasonality-anomaly decomposition, or smooth-deviance decomposition). Overall, the quality of the reference dataset impacts the precision of the decomposed high/low frequency components, which are further used in the calculation of other statistics related to our rescaling method or rescaling stationarity selection. This implies that the use of a better reference dataset selection will ultimately reduce the magnitude of errors and, subsequently, improve the precision of the decomposed and the rescaled datasets.

## 2.4 Reference Dataset Selection

Another aspect of rescaling procedure is related to the reference dataset selection (i.e.,  $X$ ). Rescaling methodologies utilize certain statistics (e.g., mean, standard deviation, correlation, and covariance) that are sampled from the reference dataset. Soil moisture dataset errors may reduce the precision of the rescaling parameters and the rescaled product, while unjustified assumptions (e.g., orthogonality and zero error cross correlations; [Yilmaz & Crow, 2014]) further diminish the precision. Accordingly, a more precise reference product may provide a better rescaling target. As a result, more precise reference datasets will yield improved rescaled products in terms of precision when more aggressive rescaling techniques (i.e., the use of different rescaling coefficients for the decomposed components or for different months) are used [Yilmaz *et al.*, 2016]. Hence, the choice of a rescaling technique should be made consistent with the reference dataset choice. For example, if a high-quality reference dataset is available - then a more aggressive methodology can be applied. In contrast, lower-skill products are likely to be better-utilized using less-aggressive rescaling techniques [Yilmaz *et al.*, 2016]. Although most soil moisture dataset fusion studies use model products as their reference [Dorigo *et al.*, 2017; Liu *et al.*, 2011; Yilmaz *et al.*, 2012], other references (i.e., remotely sensed based soil moisture products) are considered here to evaluate the performance of rescaling methods for difference reference choices. Accordingly, each data set (introduced below) is selected, in turn, to serve as the reference product for a separate fusion experiment.

When different products are used as reference, the precision of these products might impact the final rescaled product. This can be shown using regression-based rescaling coefficient starting with equation (9),

$$c_Y^R = \frac{cov_{XY}}{\sigma_Y^2} \quad (27)$$

Assuming the mean values do not impact the rescaling factors, (27) can be rewritten as:

$$c_Y^R = \frac{(\alpha_X T + \epsilon_X)(\alpha_Y T + \epsilon_Y)}{\sigma_Y^2} \quad (28)$$

$$c_Y^R = \frac{\alpha_X \alpha_Y \sigma_T^2}{\alpha_Y^2 \sigma_T^2 + 2\alpha_Y \overline{T\epsilon_Y}} + \frac{\alpha_X \overline{T\epsilon_Y} + \alpha_Y \overline{T\epsilon_X}}{\alpha_Y^2 \sigma_T^2 + 2\alpha_Y \overline{T\epsilon_Y}} + \frac{\overline{\epsilon_X \epsilon_Y}}{\alpha_Y^2 \sigma_T^2 + 2\alpha_Y \overline{T\epsilon_Y}} \quad (29)$$

where  $\sigma_T^2$  is the variance of the true soil moisture signal. The first term in equation (29) is independent of random errors in the reference product  $X$  (i.e.,  $\epsilon_X$ ), while the second and third terms are not. Assuming orthogonal ( $\overline{T\epsilon_X}$ ) and cross-correlated ( $\overline{\epsilon_X \epsilon_Y}$ ) covariances do not mutually cancel (Figure 1 of [Yilmaz and Crow, 2014]) (i.e., the strong seasonality in the soil moisture datasets results in non-orthogonal and cross-correlated error terms), the second and third terms in equation (29) will bias the rescaling coefficient  $c_Y^R$ . Accordingly, it is expected that products with lower errors form a better reference product than those with higher errors (either random or systematic) in terms of rescaled product precision even though the same product ( $Y$ ) is being rescaled. Similarly, it might be theoretically expected also for other rescaling methodologies that a reference product with lower errors might result in smaller uncertainty in the rescaling steps, though it may not be trivial to analytically prove it particularly for more complex methods (e.g., SVM).

### 3 Data sets

Various remote sensing and hydrological modelling soil moisture datasets (ASCAT, AMSR-E LPRM, NOAH, and API) available at a 0.25-degree spatial resolution are rescaled and fused. The validation of rescaled soil moisture products is based on station-based estimates of in-situ measurements obtained over 542 globally distributed sites between January 2007 and October 2011. Data set details are given below.

#### 3.1 AMSR-E LPRM

The Advanced Microwave Scanning Radiometer for EOS (AMSR-E) on-board the NASA Aqua satellite is a passive microwave radiometer that provided near-daily observations at six different frequencies (between 6.9 and 89.0 GHz) in both horizontal and vertical polarizations between 2002 and 2011. The AMSR-E measured brightness temperature with daily ascending and descending overpasses over a swath width of 1445 km. These measurements have been converted into volumetric soil moisture estimates through various algorithms (e.g., REG [Al-Yaari *et al.*, 2016], NN [Rodríguez-Fernández *et al.*, 2016]) resulting in multiple different soil moisture datasets [Mladenova *et al.*, 2014].

Among the different soil moisture datasets derived from AMSR-E observations, the Land Parameter Retrieval Method version 05 (LPRM; Owe *et al.*, [2001, 2008])-based AMSR-E dataset is used in this study. LPRM is a radiative transfer model that retrieves soil moisture and vegetation water content simultaneously utilizing either soil or canopy temperature as well as passive microwave-based X-band and C-band observations from AMSR-E for the retrieval of the surface soil moisture content. The LPRM-based soil moisture datasets used here are acquired from Vrije Universiteit Amsterdam [personal communication with Robert Parinussa, 2013] and

are available online in a gridded format and spatial resolution of  $0.25^\circ$  between June 2002 and October 2011 (<https://disc.gsfc.nasa.gov>). For more details on the LPRM retrieval method, see *Owe et al.* [2001 and 2008].

### 3.2 ASCAT

The soil moisture product derived from the Advanced Scatterometer sensor onboard of the Metop satellite [*Albergel et al.*, 2009; *Bartalis et al.*, 2007; *Wagner et al.*, 2007] retrieves estimates from C-band backscatter observations. This real-aperture radar sensor scans the globe using six antennas at different azimuth and viewing angles (three antennas on each side of the platform with  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$  viewing angles) and retrieves backscatter observations within an approximately 550-km swath width during 14 orbit revolutions per day - resulting in a  $\sim 1.5$ -day revisit time globally and twice per day over western Europe. The WARP 5.5 ASCAT soil moisture dataset used in this study is acquired from the Technical University of Vienna [*Wagner et al.*, 1999] with a spatial resolution of 25km. This product is accessible online for the time-period January 2007 to May 2012 at <http://hsaf.meteoam.it>. For more information about the ASCAT soil moisture retrieval algorithm, see *Wagner et al.* [1999 and 2007].

### 3.3 NOAH GLDAS

The NOAH land surface model [*Chen et al.*, 1996; *Koren et al.*, 1999; *Ek et al.*, 2003] is a 1-D column model that can be executed in both offline and coupled modes. NOAH uses atmospheric data as well as soil and vegetation parameters to solve for the energy and the water balance equations for different layers of soil profile. NOAH soil moisture datasets used in this study are offline simulations provided by Global Land Data Assimilation System (GLDAS Version 2; [*Rodell et al.*, 2004]) using NOAH v2.7 at spatial resolution of  $0.25^\circ$ . The GLDAS NOAH soil moisture datasets used here represent the top 10 cm of the soil column and are provided at three-hourly time steps. These soil moisture values were later averaged into daily values (0 to 24 UTC). The GLDAS NOAH soil moisture datasets used in this study are publicly available from January 2000 until present (<http://disc.sci.gsfc.nasa.gov>). For more details about NOAH and GLDAS simulations please, see *Rodell et al.* [2004].

### 3.4 API

The antecedent precipitation index (API; [*Choudhury & Blanchard*, 1983; *Blanchard et al.*, 1981; *McFarland & Blanchard*, 1977]) is a proxy that estimates surface soil moisture based on either rainfall or rainfall and runoff [*Blanchard et al.*, 1981]. The simple and practical form of the API has led to its popularity in exploratory data assimilation and fusion studies [*Crow & Ryu*, 2009]. It assumed that soil moisture depletion can be simplified as an exponential function of the input moisture to the soil profile [*Lindsey et al.*, 1949; *Chow*, 1964]. This exponential relation can also be presented in a linear form of:

$$API_i = \gamma_i API_{i-1} + P_i \quad (30)$$

where  $API$  is the antecedent precipitation which will be considered as soil moisture;  $P$  is the daily precipitation or infiltration amount;  $\gamma$  is a dimensionless depletion rate, and  $i$  is the day of estimate. Here,  $\gamma$  is taken as a constant value of 0.85 following [Yilmaz & Crow, 2013]. As precipitation input, daily Tropical Rainfall Measuring Mission (TRMM) 3B42 version 7 product has been used [Huffman *et al.*, 2007]. This product is available at a spatial resolution of  $0.25^\circ$  and available online from January 1998 until present (<https://disc.gsfc.nasa.gov>). For more details about the API model and its alternative shapes, see Blanchard *et al.* [1981].

### 3.5 In situ measurements

For validation, in-situ soil moisture measurements were collected from 542 different soil moisture monitoring sites. These datasets were collected and archived within the International Soil Moisture Network (ISMN; a data host that collects the in-situ soil moisture measurements from networks and field validation sites all around the world and make them available online for users [Dorigo *et al.*, 2011 & 2013]). The datasets vary widely in terms of measuring depth, sensor types, and measurement period. Therefore, to make consistent time series of soil moisture measurements ready for the analysis, 2007-2011 observations flagged as “Good” in the ISMN system were collected at the highest (i.e., closest-to-surface) available depth of sensor measurement, ranging from zero to 0.3 meters, and converted into daily estimates. Moreover, to obtain robust statistical results in the evaluation part of the study, sites with less than 540 data points (~1.5 years of daily data) were masked out from 1033 available site in the ISMN system for the mentioned period. For further details about ISMN datasets used in this study, readers are referred to Table 1. Note that networks that have less than four sites later are combined to the “OTH” group (i.e., other) in the section 6 of this study.

## 4 Fusion Scenarios

After a given rescaling approach is applied, the resulting rescaled soil moisture products are fused within three scenarios presented in Figure 1. The impacts of various: rescaling methods, rescaling stationarity selections, rescaling techniques, and reference dataset selections are investigated separately in via multiple numerical experiments that fuse satellite- and a model-based estimates. Moreover, the impact of different weighting schemes are compared within fusion schemes using a simple equal weighting (i.e., naive merging [Yilmaz *et al.*, 2012]) and complex merging approaches (e.g., TCA-based [Yilmaz *et al.*, 2012; Gruber *et al.*, 2017]). In particular, four separate soil moisture products (ASCAT, AMSR-E, API, NOAH) are used for fusing two or more two products while the comparison of weighting schemes are based on the pairwise merging of ASCAT, AMSR-E, API soil moisture products over the NOAH space. Note that the uncertainty that fusion process reduces arises from a broad range of sources including: poor spatial resolution, retrieval algorithm error, or sensor error-based uncertainty. Accordingly, in most cases it is not possible to differentiate these components from each other.



## 5 Evaluation

Soil moisture products can have markedly different levels of temporal variability, implying that error variance or root mean square error measure alone may not be appropriate statistics to objectively measure the skill of the fused product. Alternatively, a dimensionless statistic, correlation coefficient, is used for the assessment of the skill of native or fused soil moisture products:

$$\rho_i = \frac{\text{cov}(Sta_i, SM_i)}{\sigma_{Sta_i} \sigma_{SM_i}} \quad (31)$$

where  $Sta$  and  $SM$  are the in-situ observations and the evaluated soil moisture products, respectively, and  $\rho_i$  is the Pearson correlation over the  $i^{th}$  site (542 in total). Higher precision for any product is indicated by higher  $\rho_i$ . Here, evaluation is performed for native soil moisture products (ASCAT, AMSR-E, API, NOAA) as well as for generated fused products. Notice that the calculation of correlation coefficients between fused products and in-situ measurements are based on all available input observations. Separate analysis have been performed to examine the use of mutually available input observations (i.e., collocated observations among products involved in the fusion process) over fused product precision compared against the use of all available input observations. These analysis (not shown) demonstrate that, on average, the results only change marginally (less than 0.01 correlation), hence in this study all available datasets are used for rescaling and fusing.

## 6 Results and Discussion

Based on the rescaling and the fusion setups described above, in scenarios one and two (see Figure 1), a 6-dimensional matrix (containing dimensions corresponding choices of: fused products, rescaling methods, rescaling techniques, rescaling stationarity selections, reference dataset selections, and validation locations) of correlation results versus ground-based soil moisture observations is constructed. While in a third scenario, the original 6-dimensional matrix is reduced to a 3-dimensional matrix (containing dimensions corresponding choices of: fused products, weighting scheme, and validation locations) by considering only the simplest rescaling approach (i.e., using of linear regression without considering any variation in stationarity assumption or decomposition technique). Below, results generated from the of above mentioned scenarios are presented in three subsections related to: the precision of the input datasets, the impact of rescaling options on the precision of fused products, and the precision gain of different products during the fusion process.

### 6.1 Precision of Input Datasets

Before products are rescaled and fused within their selected reference dataset space, their precision is summarized in Table 2 and Figure 2. Overall, relative to ground-based observations, NOAA and AMSR-E soil moisture time series have higher (correlation-based) precision than ASCAT and API owing to the precision of their low-frequency (i.e., seasonality and smooth)

components, while the high-frequency skill of ASCAT and API (anomaly and deviance) are comparable to that of AMSR-E. NOAH exhibits the highest correlation against in-situ measurements with regards to both high and low frequency components amongst all the products (Figure 2). Low- and high-frequency variability compared to the entire time series variability (eqs. 20-24) results for all products are different for various temporal decomposition techniques (see the top part of Table 2). In particular, the smooth component generally contributes a higher percentage of the total variability than the seasonality component for all products (Table 2a). In fact, the seasonality variability is lower than the anomaly variability, while the smooth component contributes more variability than the deviance component.

## 6.2 Impact of Rescaling Options

While our data fusion analysis generates a 6-dimensional correlation matrix, Figures 3 and 4 select two dimensions for display and averages correlation values across the remaining four dimensions of the matrix. One of the displayed dimensions (i.e., the x-axis) always reflects variations in the reference dataset choice. The second displayed dimension (expressed via color variations summarized in the legend of each panel) varies from panel to panel. For example, the impact of rescaling method selection in a data fusion framework is presented in panel A of Figures 3 and 4. Overall, under identical experiment designs, the selection of NOAH as the reference dataset results in a more precise fused product than the selection of other datasets as reference (Figure 3 and 4). This tendency is not impacted by the number of fused products (two or more than two cases). The reference dataset selection also impacts the rescaling method performance. In particular, the nonlinear MAR and SVM methods yield a more precise fused product when NOAH (i.e., a more precise product) is used as the rescaling reference. On the other hand, the linear REG and VAR methods yield a more precise fused product when less correlated products (i.e., against in-situ measurements) are used as the rescaling reference (see correlation the difference between methods over different references in panel A of Figures 3 and 4).

The impact of rescaling stationarity selection (i.e., the use of a constant or time-varying scaling coefficients) over the precision of fused products is shown in panel B of Figures 3 and 4 for fusion of two and more than two products, respectively. Confirming the results of *Yilmaz et al.* [2016] the use of more aggressive (time-varying) rescaling stationarity selection results in degraded fused product in terms of precision when an imprecise reference dataset is used (see the correlation difference between time-varying stationarity assumptions over API and NOAH references presented with red color in panel B of Figures 3 and 4). In general, higher correlations are associated with the application of constant rescaling parameters, while the use of a more aggressive rescaling stationarity selection does not improve the correlation of the final fused product as much - particularly when the overall precision of the reference product is low (e.g., ASCAT and API). On the other hand, when products are fused using a good reference product (e.g., NOAH), the use of time-varying stationarity assumption improves the correlation of final

fused products. While this increment can be seen in both scenarios regardless of number of products involved in fusion process.

For all products, time series deviance components have lower autocorrelation and less cross-correlation (see the middle and bottom portions of Table 2, respectively) than comparable anomaly components, while the smooth components have higher cross-correlation and variability than the seasonality component (see the variability ratio difference between fifth and third column in top part of Table 2). Accordingly, smooth-deviance decomposition is expected to be more beneficial as a pre-processing step than seasonality-anomaly decomposition. Accordingly, when the autocorrelation and variability of decomposition techniques is compared in Table 2, the deviance high frequency product has lower variability and resembles a random noise process more than the seasonality-anomaly-based product (i.e., the deviance component is closer to a random noise process than the anomaly component for the same product). Expectedly, once these components are rescaled and fused, the smooth-deviance decomposition-based fused product results in a more precise fused product versus independent ground observations (see correlation difference between smooth-deviance and seasonality-anomaly techniques in panel C of Figures 3 and 4) than the comparable fused product derived via seasonality-anomaly decomposition.

Investigating the impact of product selection on the precision of the fused product (panel D of Figures 3 and 4) reveals that the rescaling procedure and reference dataset selection are critical for the fusion process - regardless of the number of products that are fused. The increment trend in correlation of fused products realized upon improving the precision of reference product is detectable in panel D of both Figures 3 and 4. On the other hand, adding additional products into the fusion process generally improves its precision. For example, the comparison of products based on the fusion of AMSR-E, ASCAT, and API (indicated with blue coloring in panel D of Figure 4) with products based on AMSR-E and ASCAT fusion (represented with blue coloring in panel D of Figure 3) shows that adding the API product into the fusion process provides a positive effect to the precision of fused product (see correlation differences shown in panel D of Figures 3 and 4). As expected, this improvement is more pronounced over sites where the products are less correlated with ground truth (that is, the lowest 25% quantile shown in boxplots have higher improvements than higher quantiles).

### 6.3 The Precision Gain from the Fusion Process

On average, when sampled across all possible reference choices, the simple fusion of products on average improves the correlation of the individual products by 0.06 when two products are fused (AMSR-E 0.108, ASCAT 0.114, API 0.121, and NOAA -0.097) and 0.10 when more than two products are fused (AMSR-E 0.152, ASCAT 0.162, API 0.168, and NOAA -0.082). However, when NOAA is selected as the reference dataset, these correlation improvements increase to 0.10 when two products are fused (AMSR-E 0.15, ASCAT 0.16, API

0.16, and NOAH -0.08), and 0.22 when more than two products are fused (AMSR-E 0.27, ASCAT 0.27, API 0.28, and NOAH 0.04). Overall, AMSR-E, ASCAT, and API benefit from the fusion the most (i.e., correlation difference given in panel D of Figures 3 and 4, and Figure 2), particularly when more than two products are fused, and NOAH is selected as the reference.

Since the NOAH dataset is globally available, and the fused products have higher correlation with ground truth when NOAH is selected as the reference dataset, further analyses are performed for the case of using NOAH as the reference. Table 3 and Table 4 summarize the correlation statistics of fused products against in-situ measurements for the case of fusing two or more two products. On average, MAR and SVM perform better than other methods (CDFM, REG, and VAR). Although the correlation difference between nonlinear and linear methods is relatively marginal, when the reference dataset selection is changed from NOAH to in-situ, larger benefits are expected in the case of nonlinear methods since utilizing in-situ data as reference add less sampling uncertainty [Afshar & Yilmaz, 2017]. Overall, the implementation of time varying rescaling coefficients yields higher improvement compared to constant coefficients (i.e., on average the difference between the correlations under the columns of T and C in Tables 3 and 4 is 0.043 and 0.076, respectively). Similarly, the selection of decomposition techniques yields more precise fused products compared to the fusing of products without first decomposing them (i.e., on average the differences between the correlations in front of the rows related to the seasonality-anomaly- and no decomposition- techniques are 0.013 and 0.006 in Tables 3 and 4, while the differences between rows associated with smooth-deviance- and no decomposition- techniques are 0.02 and 0.015 in Tables 3 and 4, respectively).

Given that the performance of a smooth-deviance decomposition technique is better than a seasonality-anomaly technique, and the fusing of AMSR-E and ASCAT soil moisture products provides an idea about the utility of rescaling approaches for the fusion of active and passive soil moisture products, Figure 5 investigates the performance of AMSR-E – ASCAT fusion over different networks separately. This investigation is based on comparison of the cross-correlations between the unscaled products and the fused products against in-situ datasets over different networks when different rescaling method and technique selections are used. Overall, the correlation improvement achieved by the smooth-deviance decomposition technique is observed over all the networks. The AMSR-E – ASCAT fusion product is more precise than its parent products individually. The single exception to this tendency is found in the UDC\_SMOS network where both unscaled and reference soil moisture products are of low quality. The reduced precision of fused product with respect to the pair products is more pronounced when products are being rescaled aggressively. This illustrates – once again – the failure of aggressive approaches (i.e., nonlinear methods or time-varying stationarity assumptions) in cases where the reference product is inferior to the unscaled products that are being fused.

Overall, a time-varying rescaling coefficient assumption (e.g., monthly coefficients) clearly results in an improved fused product when the reference product is more correlated with the in

situ measurements [Yilmaz *et al.*, 2016]. While aggressive implementation of the rescaling methods (including the time-varying assumptions) over a reference which is relatively less correlated with the truth, results in reduced fused product precision due to the over-fitting of the unscaled product to the reference product. For such cases, simpler linear rescaling methods can remove systematic differences in the unscaled dataset without compromising the precision of the fused product.

The comparison of different weighting approaches in data fusion framework is shown in Figure 6, where the AMSR-E, ASCAT, and API soil moisture products are fused in pairwise schemes based on two different weighting approaches (i.e., naive- and TCA-based). The overall average of three product pairs (see the last pair boxplots of Figure 6), indicate that the fusion framework of this study does not measurably benefit from TCA-based weighting of fused products. Although, these results are consistent with the results of Yilmaz *et al.*, [2012] who suggest that there is rarely enough contrast between the precision of fused products for TCA-based merging to be clearly beneficial, other studies have found that TCA-based weighting approach can improve the precision of fused products [Gruber *et al.*, 2017 & 2019]. The TCA methodology for accurately estimating of errors, and hence weights, requires specific error assumptions to be met. Accordingly, the optimality of TCA implementations and the gained utility from this methodology may reasonably vary for different implementations.

## 7 Conclusions

Most soil moisture applications require soil moisture product with a good precision. For such studies, an improved soil moisture dataset can be obtained by appropriately removing systematic differences between the multiple products and merging them with a proper methodology (i.e., simple fusion or data assimilation). However, the approach taken during the removal of systematic differences (i.e., rescaling) can also affect the precision of the final fused product. The present study investigates the impact of rescaling method choice and the implementation approaches with the aim of optimizing a simple data fusion framework for soil moisture. The fused products are compared against in situ measurements over 542 ISMN sites to evaluate the impact of various rescaling approaches (i.e., different rescaling methodologies, techniques, and stationarity selections) during the period of 2007-2011 on the precision of the resulting fused soil moisture product.

Here, in this study it is concluded that the precision of the fused product is most sensitive to variations in the dataset selected to serve as a reference (assuming other rescaling related choices are the same). After the reference dataset selection, the product selection (i.e., the selection of data products to be fused) matters the most. For example, AMSR-E and ASCAT precisions individually are not superior to NOAA - yet the fusion of them over a proper reference product and using a proper rescaling approach yields a superior fused product than the NOAA product. Additionally, among rescaling approaches, the implementation of nonlinear methods

(e.g., CDFM, SVM) marginally improve the precision of fused products (~0.01 correlation improvement) as compared to linear methods (e.g., REG); however, the implementation of linear methods with a proper decomposition technique (e.g., smooth-deviance decomposition) can yield higher correlations than use of nonlinear methods that are not enhanced with decomposition techniques and time-varying stationarity assumption. Overall, rescaling the high- and the low-frequency components of soil moisture datasets using a smooth/deviance technique results in a more precise fused product than use of seasonality/anomaly technique. Nevertheless, the relative performance of different decomposition techniques, and also stationarity assumptions, should be further investigated in detail via follow-on studies utilizing more experiments with other different methods before making a general conclusion about the optimality of the high/low frequency decomposition technique and also different stationarity assumptions across the entire spectrum of soil moisture applications.

### Acknowledgments

We thank Robert Parinussa of Vrije Universiteit Amsterdam for the AMSR-E LPRM datasets, Wouter Dorigo of Technische Universität Wien for the ASCAT datasets, NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) for the NOAA datasets, and ISMN for providing the in-situ based soil moisture estimates. This research is funded by Turkish Scientific and Technical Research Council (TUBITAK) under grant number 114Y676 and by EU Marie Curie Seventh Framework Programme FP7-PEOPLE-2013-CIG project number 630110.

### References

- Afshar, M., & Yilmaz, M. (2016). Conditional Copula-Based Spatial–Temporal Drought Characteristics Analysis—A Case Study over Turkey. *Water*, 8(10), 426.
- Afshar, M., & Yilmaz, M. (2017). The added utility of nonlinear methods compared to linear methods in rescaling soil moisture products. *Remote Sensing of Environment*, 196, 224–237.
- Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Pignatelli, B., & Martin, E. (2008). From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in situ observations and model simulations, *Hydrol. Earth Syst. Sci.*, 12, 1323–1337.
- Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., & Hasenauer, S. (2009). An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France. *Hydrology and Earth System Sciences*, 13(2), 115–124.

- Al-Yaari, A., Wigneron, J., Kerr, Y., de Jeu, R., Rodriguez-Fernandez, N., van der Schalie, R., Al Bitar, A., Mialon, A., Richaume, P., Dolman, A., & Ducharne, A. (2016). Testing regression equations to derive long-term global soil moisture datasets from passive microwave observations. *Remote Sensing of Environment*, 180, 453-464.
- Anderson, M., Zolin, C., Hain, C., Semmens, K., Yilmaz, M., & Gao, F. (2015). Comparison of satellite-derived LAI and precipitation anomalies over Brazil with a thermal infrared-based Evaporative Stress Index for 2003–2013. *Journal of Hydrology*, 526, 287-302.
- Anderson, M., Zolin, C., Sentelhas, P., Hain, C., Semmens, K., Yilmaz, M., Gao, F., Otkin, J.A., & Tetrault, R. (2016). The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. *Remote Sensing of Environment*, 174, 82-99.
- Anderson, W., Zaitchik, B., Hain, C., Anderson, M., Yilmaz, M., Mecikalski, J., & Schultz, L. (2012). Towards an integrated soil moisture drought monitor for East Africa. *Hydrology and Earth System Sciences*, 16, 2893-2913.
- Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., & Anderson, C. (2007). Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophysical Research Letters*, 34(20), L20401.
- Bell, J. E., Palecki, M. A., Baker, C. B., Collins, W. G., Lawrimore, J. H., Leeper, R. D., Hall, M. E., Kochendorfer, J., Meyers, T. P., Wilson, T., & Diamond, H. J. (2013). U.S. Climate Reference Network soil moisture and temperature observations, *J. Hydrometeorol.*, 14, 977-988.
- Blanchard, B., McFarland, M., Schmugge, T., & Rhoades, E. (1981). Estimation of Soil-Moisture with API Algorithms and Microwave Emission. *Water Resources Bulletin*, 17(5), 767-774.
- Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martinez-Fernández, J., Llorens, P., Latron, J., Martin, C., & Bittelli, M. (2011). Soil moisture estimation through ASCAT and AMSR-E sensors: an inter-comparison and validation study across Europe, *Remote Sensing of Environment*, 115, 3390–3408.
- Bulut, B., Yilmaz, M.T., Afshar, M.H., Şorman, A.Ü., Yücel, İ., Cosh, M.H., & Şimşek, O. (2019). Evaluation of Remotely-Sensed and Model-Based Soil Moisture Products According to Different Soil Type, Vegetation Cover and Climate Regime Using Station-Based Observations over Turkey, *Remote Sensing*, 11(16), p.1875.

- Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., & Piguet, B. (2007). In situ soil moisture observations for the CAL/VAL of SMOS: the SMOSMANIA network, International Geoscience and Remote Sensing Symposium, IGARSS, Barcelona, Spain, 23-28 July 2007, 1196-1199.
- Cappelaere, B., Descroix, L., Lebel, T., Boulain, N., Ramier, D., Laurent, J.-P., Favreau, S., , E., Boubkraoui, S., Boucher, M., Bouzou Moussa, I., Chaffard, V., Hiernaux, P., Issoufou, HBA., Le Breton, E., Mamadou, I., Nazoumou, I., Oi, M., Otle, C., & Quantin, G. (2009). The AMMA Catch observing system in the cultivated Sahel of South West Niger-Strategy, Implementation and Site description, *Journal of Hydrology*, 375(1-2), 34-51.
- Chow, V. (1964). Handbook of Applied Hydrology. In V. Chow, Handbook of Applied Hydrology (pp. (21): 1-5). New York: McGraw-HiU Book Company, Inc.
- Cosh, M., Jackson, T., Starks, P., & Heathman, G. (2006). Temporal stability of surface soil moisture in the Little Washita River watershed and its applications in satellite soil moisture product validation. *Journal of Hydrology*, 323(1-4), 168-177.
- Cosh, M., Ochsner, T., McKee, L., Dong, J., Basara, J., Evett, S., Hatch, C.E., Small, E.E., Steele-Dunne, S.C., Zreda, M., & Sayde, C. (2016). The Soil Moisture Active Passive Marena, Oklahoma, In Situ Sensor Testbed (SMAP-MOISST): Testbed Design and Evaluation of In Situ Sensors. *Vadose Zone Journal*, 15(4), doi:10.2136/vzj2015.09.0122.
- Crow, W., & Ryu, D. (2009). A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences*, 13(1), 1-16.
- Crow, W., & Zhan, X. (2007). Continental-Scale Evaluation of Remotely Sensed Soil Moisture Products. *IEEE Geoscience and Remote Sensing Letters*, 4(3), 451-455.
- De Rosnay, P., Gruhier, C., Timouk, F., Baup, F., Mougin, E., Hiernaux, P., Kergoat, L., & LeDantec, V. (2009). Multi-scale soil moisture measurements at the Gourma meso-scale site in Mali, *Journal of Hydrology*, 375, 241-252, 2009.
- De Lannoy, G., & Reichle, R. (2016). Global Assimilation of Multiangle and Multipolarization SMOS Brightness Temperature Observations into the GEOS-5 Catchment Land Surface Model for Soil Moisture Estimation. *Journal of Hydrometeorology*, 17, 669–691.
- Dirmeyer, P., Guo, Z., Gao, X., Dirmeyer, P., Guo, Z., & Gao, X. (2004). Comparison, Validation, and Transferability of Eight Multiyear Global Soil Wetness Products. *Journal of Hydrometeorology*, 5(6), 1011-1033.



- Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., & Jackson, T. (2011). The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, *Hydrol. Earth Syst. Sci.*, 15, 1675-1698.
- Dorigo, W.A., Xaver, A. Vreugdenhil, M. Gruber, A., Hegyiová, A. Sanchis-Dufau, A.D., Zamojski, D., Cordes, C., Wagner, W., & Drusch, M. (2013). Global Automated Quality Control of In situ Soil Moisture data from the International Soil Moisture Network. *Vadose Zone Journal*, 12, 3, doi:10.2136/vzj2012.0097.
- Dorigo, W., de Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., & Fernández-Prieto, D. (2012). Evaluating global trends (1988-2010) in harmonized multi-satellite surface soil moisture. *Geophysical Research Letters*, 39(18), L18405.
- Dorigo, W., Gruber, A., De Jue, R.A.M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R.M., & Kidd, R. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations, *Remote Sensing of Environment*, 162, 380-395.
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P.D., Hirschi, M., Ikonen, J., De Jeu, R., Kidd, R., Lahoz, W., Liu, Y.Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., Schalie, R.V.D., Seneviratne, S.I., Smolander, T., & Lecomte, P. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, 203, 185-215.
- Draper, C., Walker, J., Steinle, P., de Jeu, R., & Holmes, T. (2009). An evaluation of AMSR-E derived soil moisture over Australia. *Remote Sensing of Environment*, 113(4), 703-710.
- Friedman, J. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1-67.
- Gruber, A., Dorigo, W., Crow, W., & Wagner, W. (2017). Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 6780-6792.
- Gruber, A., Scanlon, T., an der Schalie, R., Wagner, W., & Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology, *Earth System Science Data* 11(2), 717-739.
- Hain, C., Crow, W., Mecikalski, J., Anderson, M., & Holmes, T. (2011). An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote

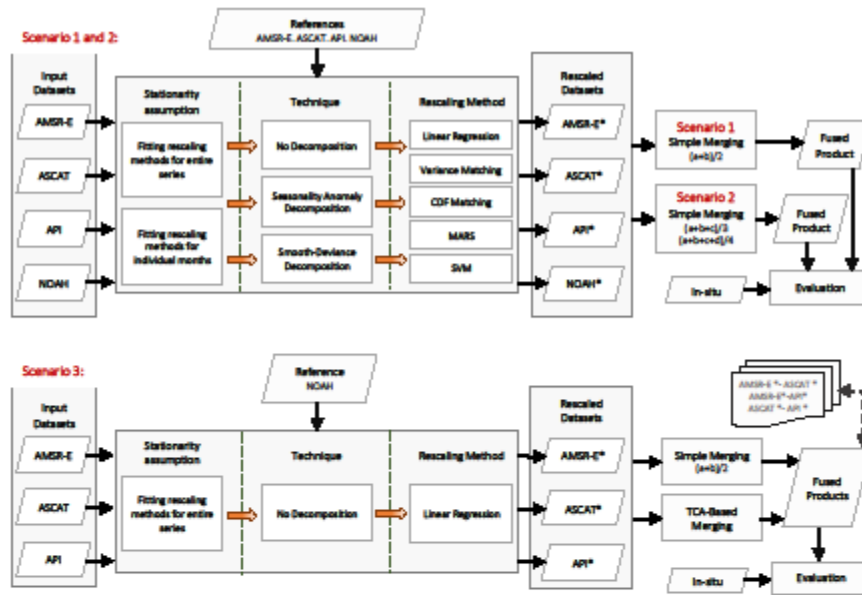
- sensing and land surface modeling. *Journal of Geophysical Research Atmospheres*, 116(15), 1-18.
- Hernández, N., Kiralj, R., Ferreira, M., & Talavera, I. (2009). Critical comparative analysis, validation and interpretation of SVM and PLS regression models in a QSAR study on HIV-1 protease inhibitors. *Chemometrics and Intelligent Laboratory Systems*, 98(1), 65-77.
- Hollinger, S.E., & Scott A. I. (1994). A soil moisture climatology of Illinois. *J. Climate*, 7, 822-833.
- Jackson, T.J., Cosh, M.H., Bindlish, R., Starks, P.J., Bosch, D.D., Seyfried, M.S., Goodrich, D.C., & Moran, M.S. (2010). Validation of Advanced Microwave Scanning Radiometer Soil Moisture Products. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 4256-4272.
- Jonas A. (2013). A 10-Year Dataset of Basic Meteorology and Soil Properties in Central Sudan. *Dataset Papers in Geosciences*, 2013, 297973.
- Koster, R., Dirmeyer, P., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C.T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y.C., Taylor, C.M., Verseghy, D., Vasic, R., Xue, Y., & Yamada, T. (2004). Regions of strong coupling between soil moisture and precipitation. *Science (New York, N.Y.)*, 305(5687), 1138-40.
- Larson, K. M., Small, E. E., Gutmann, E. D., Bilich, A. L., Braun, J. J., & Zavorotny V. U. (2008). Use of GPS receivers as a soil moisture network for water cycle studies. *Geophysical Research Letters*, 35, L24405.
- Lebel T., Cappelaere B., Galle S., Hanan N., Kergoat L., Levis S., Vieux B., Descroix L., Gosset, M., Mougin E., Peugeot C., & Seguis L. (2009). AMMA-CATCH studies in the Sahelian region of West-Africa: An overview. *Journal of Hydrology*, 375, 3-13.
- Leroux, D., Kerr, Y., Al Bitar, A., Bindlish, R., Jackson, T., Berthelot, B., & Portet, G. (2014). Comparison between SMOS, VUA, ASCAT, and ECMWF Soil Moisture Products Over Four Watersheds in U.S. *IEEE Transactions on Geoscience and Remote Sensing*, 52(3), 1562-1571.
- Leroux, D., Kerr, Y., Wood, E., Sahoo, A., Bindlish, R., & Jackson, T. (2014). An Approach to Constructing a Homogeneous Time Series of Soil Moisture Using SMOS. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1), 393-405.

- Leavesley, G. H., David, O., Garen, D. C., Lea, J., Marron, J. K., Pagano, T. C., Perkins, T. R., & Strobel, M. L. (2008). A Modeling Framework for Improved Agricultural Water Supply Forecasting, AGU Fall Meeting Abstracts.
- Lindsey, R., Kohler, Jr., M., & Paulhus, J. (1949). *Applied Hydrology*. New York: McGraw-Hill Book Company, Inc.
- Liu, Y., Dorigo, W., Parinussa, R., de Jeu, R., Wagner, W., McCabe, M., Evans, J.P., & van Dijk, A. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment*, 123, 280-297.
- Liu, Y., Parinussa, R., Dorigo, W., De Jeu, R., Wagner, W., van Dijk, A., McCabe, M., & Evans, J. (2011). Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences*, 15(2), 425-436.
- Loew, A., Dall'Amico, J. T., Schlenz, F., & Mauser, W. (2010). The Upper Danube soil moisture validation site: Measurements and activities. In *EGU General Assembly Conference Abstracts* (Vol. 12, p. 9706).
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C., & Lin, C. (2015). *Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, R package version 1.6-7. TU Wien.
- Miborrow, S. (2016). *earth: Multivariate Adaptive Regression Splines*. R package version 4.2.0. <http://CRAN.R-project.org/package=earth>.
- Miralles, D., De Jeu, R., Gash, J., Holmes, T., & Dolman, A. (2011). Magnitude and variability of land evaporation and its components at the global scale. *Hydrology and Earth System Sciences*, 15(3), 967-981.
- Mladenova, I., Bolten, J., Crow, W., Anderson, M., Hain, C., Johnson, D., & Mueller, R. (2017). Inter-comparison of soil moisture, evaporative stress and vegetation indices for estimating corn and soybean yields over the U.S. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(4), 1328-1343.
- Mladenova, I., Jackson, T., Njoku, E., Bindlish, R., Chan, S., Cosh, M., Holmes, T.R.H., De Jeu, R.A.M., & Santi, E. (2014). Remote monitoring of soil moisture using passive microwave-based techniques — Theoretical basis and overview of selected algorithms for AMSR-E. *Remote Sensing of Environment*, 144, 197-213.

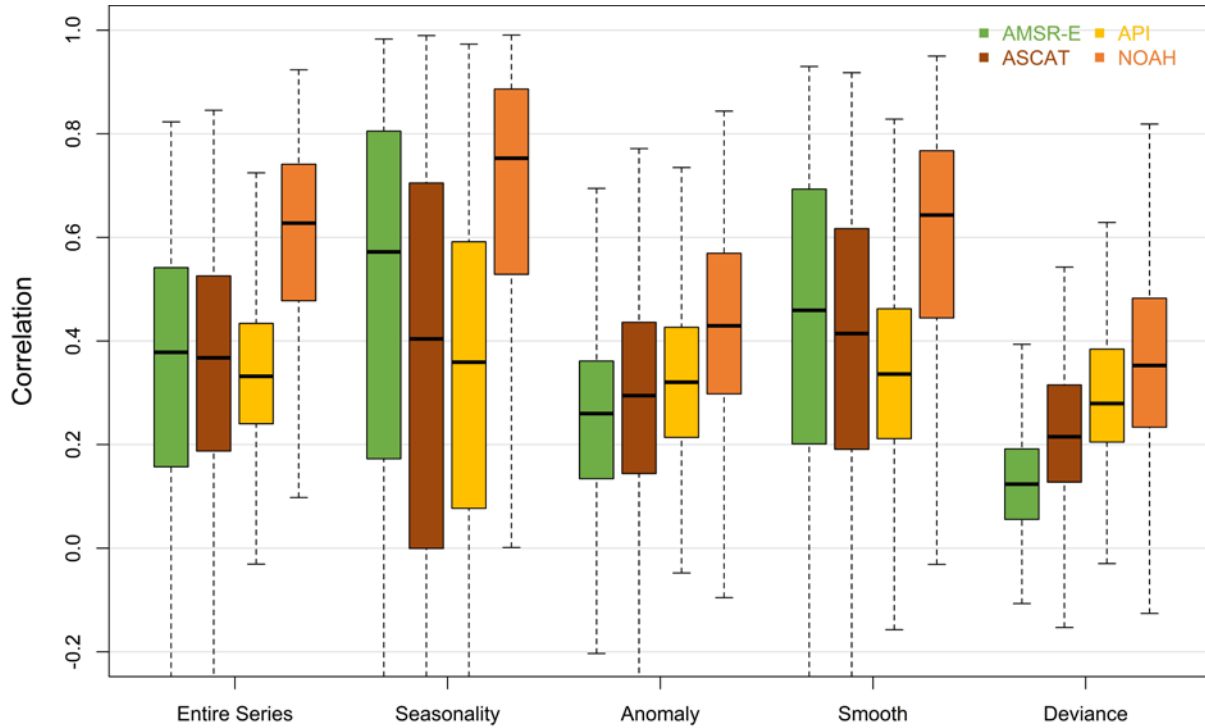
- Morbidelli, R., Saltalippi, C., Flammini, A., Rossi, E. & Corradini, C. (2014). Soil water content vertical profiles under natural conditions: Matching of experiments and simulations by a conceptual model, *Hydrological Processes*, 28(17), 4732-4742.
- Mougin, E., Hiernaux, P., Kergoat, L., Grippa, M., de Rosnay, P., Timouk, F., Le Dantec, V., Demarez, V., Lavenu, F., Arjounin, M., Lebel, T. et al. (2009). The AMMA-CATCH Gourma observatory site in Mali: Relating climatic variations to changes in vegetation, surface in press, hydrology, fluxes and natural resources. *Journal of Hydrology*, 375(1-2): 14-33.
- Olson, D., & Delen, D. (2008). *Advanced data mining techniques*. Springer.
- Parinussa, R., Holmes, T., Yilmaz, M., & Crow, W. (2011). The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Hydrology and Earth System Sciences*, 15(10), 3135-3151.
- Pellarin, T., Laurent, J.P., Cappelaere, B., Decharme, B., Descroix, L., & Ramier, D. (2009). Hydrological Modelling and Associated Microwave Emission of a Semi-Arid Region in South-Western Niger, *Journal of Hydrology*, 375(1-2), 262-272.
- Reichle, R., & Koster, R. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, 31(19), L19501.
- Reichle, R., & Koster, R. (2005). Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophysical Research Letters*, 32(2), L02404.
- Rodríguez-Fernández, N., Kerr, Y., van der Schalie, R., Al-Yaari, A., Wigneron, J.-P., de Jeu, R., Richaume, P., Dutra, E., Mialon, A., & Drusch, M. (2016). Long Term Global Surface Soil Moisture Fields Using an SMOS-Trained Neural Network Applied to AMSR-E Data. *Remote Sensing*, 8(11), 959.
- Sanchez, N., Martinez-Fernandez, J., Scaini, A., & Perez-Gutierrez, C. (2012). Validation of the SMOS L2 soil moisture data in the REMEDHUS network (Spain), *IEEE Transactions on Geoscience and Remote Sensing*, 50, 1602–1611.
- Schlenz, F., Dall’Amico, J., Loew, A., & Mauser, W. (2012). Uncertainty Assessment of the SMOS Validation in the Upper Danube Catchment, *IEEE Transactions on Geoscience and Remote Sensing*, 50(5), 1517–1529.

- Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., & Wagner, W. (2008). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters*, 35(24), L24403.
- Smith, A. B., Walker, J. P., Western, A. W., Young, R. I., Ellett, K. M., Pipunic, R. C., Grayson, R. B., Siriwardena, L., Chiew, F. H. S., & Richter, H. (2012). The Murrumbidgee soil moisture monitoring network data set, *Water Resour. Res.*, 48, W07701.
- Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans*, 103(C4), 7755-7766.
- Su, Z., Wen, J., Dente, L., van der Velde, R., Wang, L., Ma, Y., Yang, K., & Hu, Z. (2011). The Tibetan Plateau observatory of plateau scale soil moisture and soil temperature (Tibet-Obs) for quantifying uncertainties in coarse resolution satellite and model products, *Hydrol. Earth Syst. Sci.*, 15, 2303–2316.
- Su, C.-H., & Ryu, D. (2015). Multi-scale analysis of bias correction of soil moisture. *Hydrology and Earth System Sciences*, 19(1), 17-31.
- van der Schalie, R., de Jeu, R., Kerr, Y., Wigneron, J., Rodríguez-Fernández, N., Al-Yaari, A., Parinussa, R.M., Mecklenburg, S., & Drusch, M. (2017). The merging of radiative transfer based surface soil moisture data from SMOS and AMSR-E. *Remote Sensing of Environment*, 189, 180-193.
- van der Schalie, R., Jeu, R. d., Rodríguez-Fernández, N., Al-Yaari, A., Kerr, Y., Wigneron, J.-P., & Drusch, M. (2018). The Effect of Three Different Data Fusion Approaches on the Quality of Soil Moisture Retrievals from Multiple Passive Microwave Sensors. *Remote Sensing*, 10, 107. doi:<https://doi.org/10.3390/rs10010107>
- Vapnik, V. (1998). *Statistical learning theory*. Wiley.
- Vapnik, V., & Chervonenkis, A. (1974). *Theory of Pattern Recognition [in Russian]*. Nauka, USSR.
- Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., & Kerr, Y. (2007). Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Hydrology Research*, 38(1), 1-20.
- Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E., & Ertl, M. (2012). Fusion of Active and Passive Microwave Observations To Create an Essential Climate Variable Data Record on Soil Moisture. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-7(September), 315-321.

- Young, R., Walker, J., Yeoh, N., Smith, A., Ellett, K., Merlin, O., & Western, A. (2008). Soil moisture and meteorological observations from the murrumbidgee catchment, Department of Civil and Environmental Engineering, The University of Melbourne.
- Yilmaz, M., & Crow, W. (2013). The Optimality of Potential Rescaling Approaches in Land Data Assimilation. *Journal of Hydrometeorology*, 14(2), 650-660.
- Yilmaz, M., Crow, W., Anderson, M., & Hain, C. (2012). An objective methodology for merging satellite- and model-based soil moisture products. *Water Resources Research*, 48(11), n/a-n/a.
- Yilmaz, M., Crow, W., Ryu, D., Yilmaz, M., Crow, W., & Ryu, D. (2016). Impact of Model Relative Accuracy in Framework of Rescaling Observations in Hydrological Data Assimilation Studies. *Journal of Hydrometeorology*, 17(8), 2245-2257.
- Yin, J., Zhan, X., Zheng, Y., Liu, J., Hain, C., & Fang, L. (2014). Impact of quality control of satellite soil moisture data on their assimilation into land surface model. *Geophysical Research Letters*, 41(20), 7159-7166.
- Zreda M., Desilets D., Ferré Ty P.A., & Scott R.L. (2008). Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, *Geophysical Research Letters*, 35(21).
- Zreda, M., Shuttleworth, W.J., Zeng, X., Zweck, C., Desilets, D., Franz, T., & Rosolem, R. (2012). COSMOS: the COsmic-ray Soil Moisture Observing System, *Hydrology and Earth System Sciences*, 16, 4079-4099.
- Zwieback, S., Scipal, K., Dorigo, W., & Wagner, W. (2012). Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes in Geophysics*, 19(1), 69-80.
- Zwieback, S., Su, C.-H., Gruber, A., Dorigo, & W., Wagner. (2016). The Impact of Quadratic Nonlinear Relations between Soil Moisture Products on Uncertainty Estimates from Triple Collocation Analysis and Two Quadratic Extensions. *Journal of Hydrometeorology*, 17(6), 1725-1743.

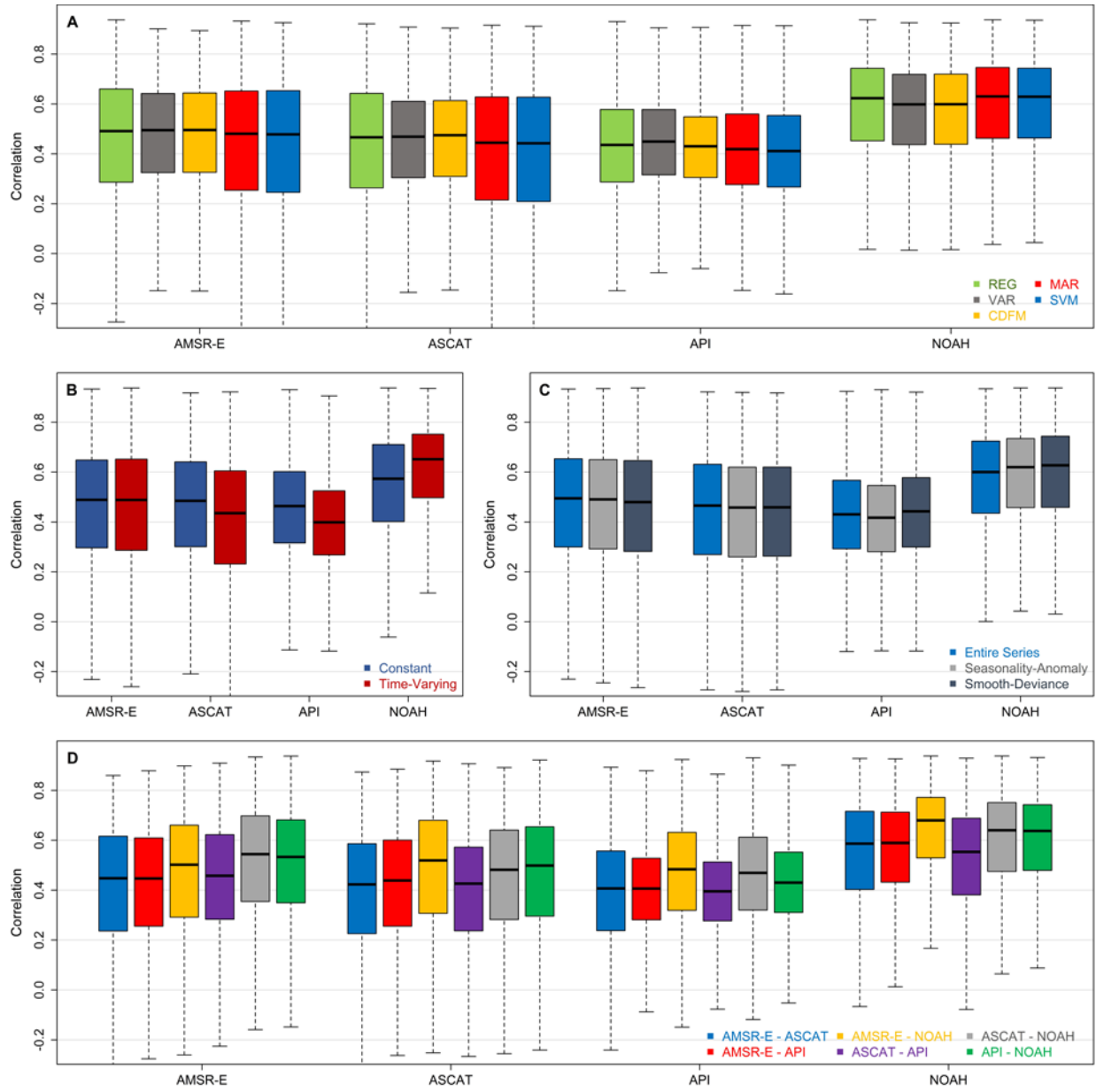


**Figure 1.** Schematic of various rescaling and data fusion procedures. Scenario 1 and 2 are focusing on impact of rescaling approaches on fusing of products based on simple merging scheme, while scenario 3 is focusing on comparison of merging schemes (simple- and TCA-based- merging) utilizing the simplest rescaling approach. The distinction between Scenario 1 and 2 is summarized in the number of products that they use in their fusion procedures. While Scenario 1 is developed based on fusing of pair products, scenario 2 is developed based on fusing of more than two (three or four) products.

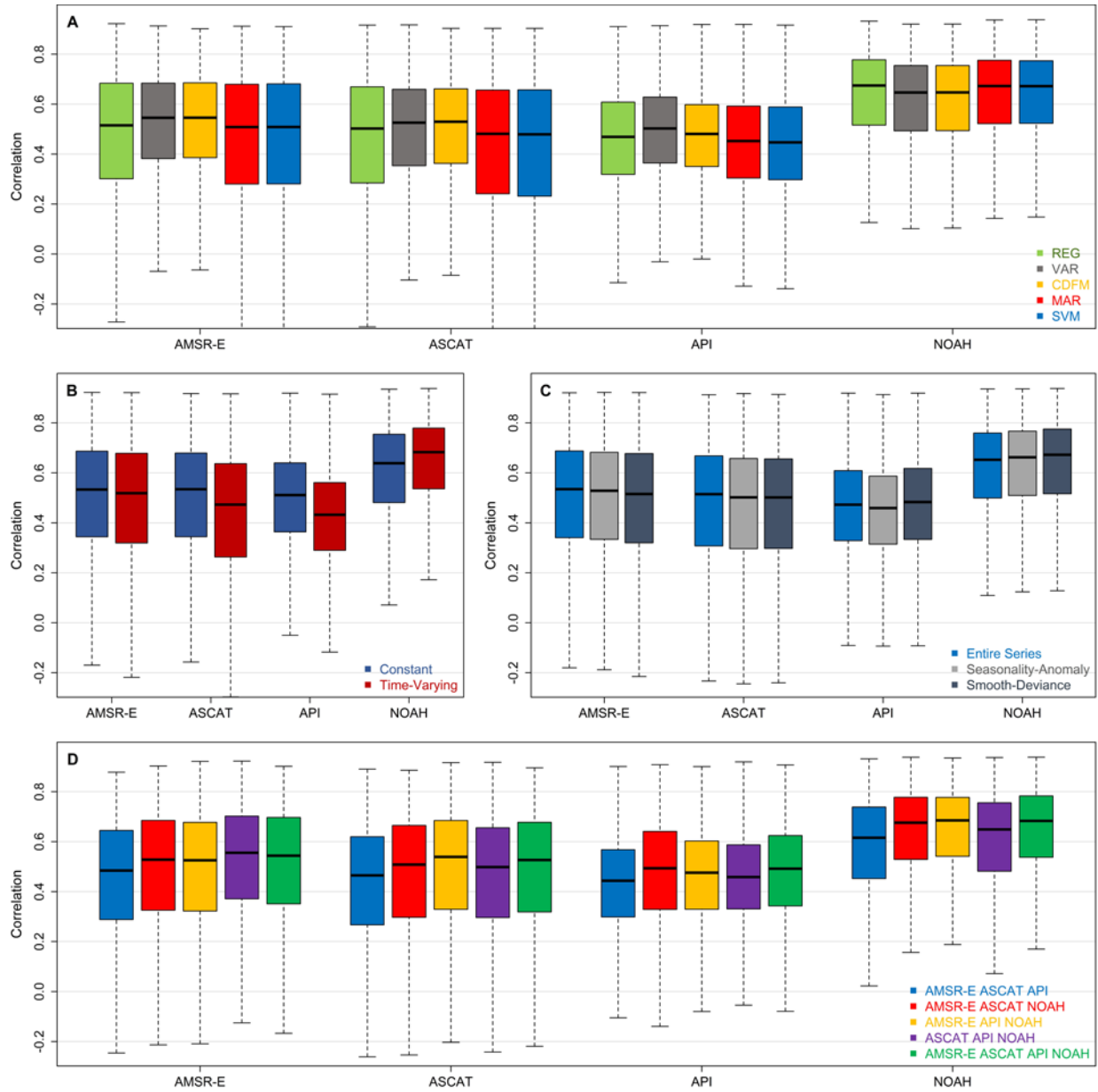


**Figure 2.** The correlation between in situ soil moisture observations and various remotely sensed and modelled soil moisture time series components defined in section 2.3 by eqs. 20-22.

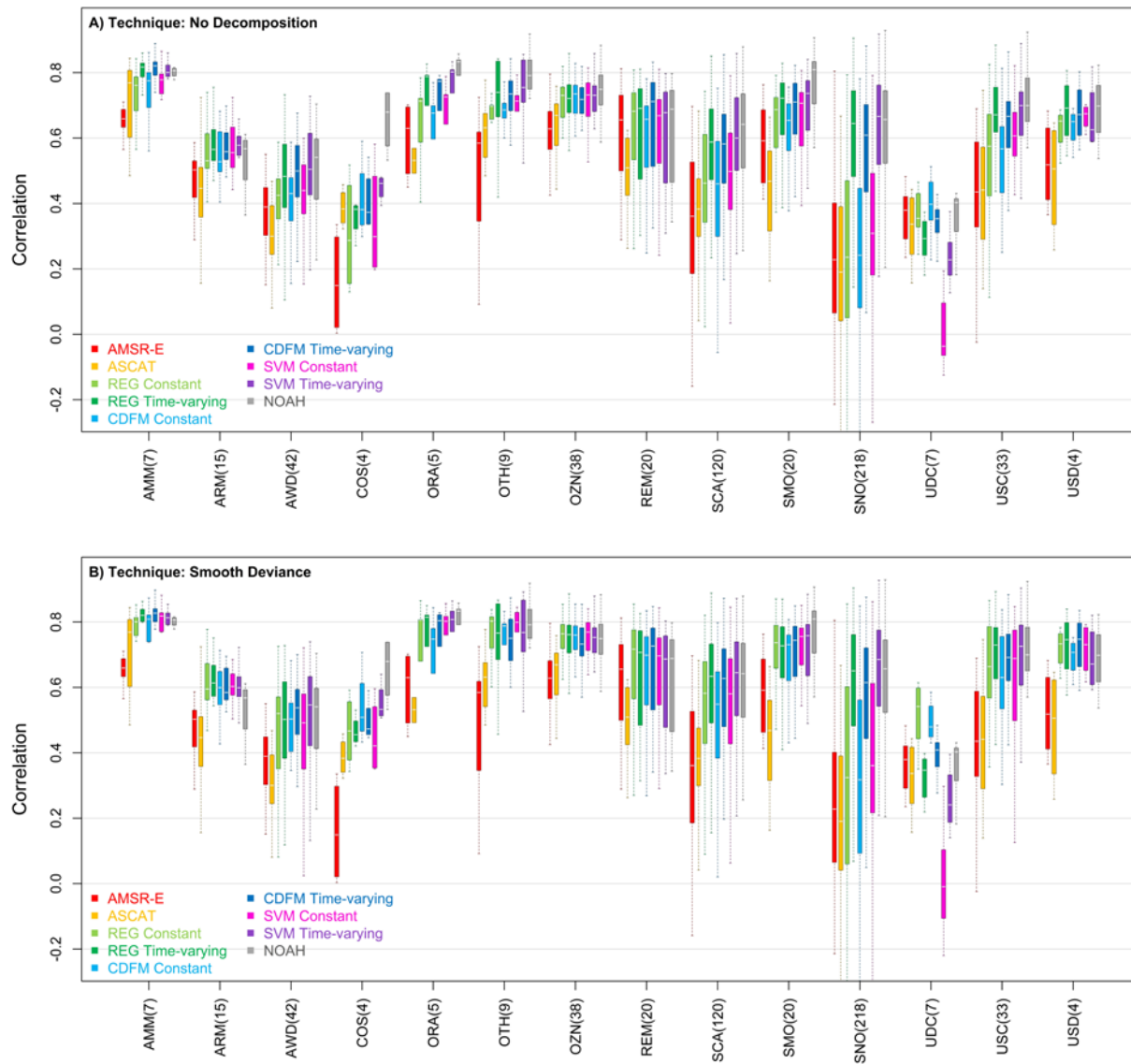




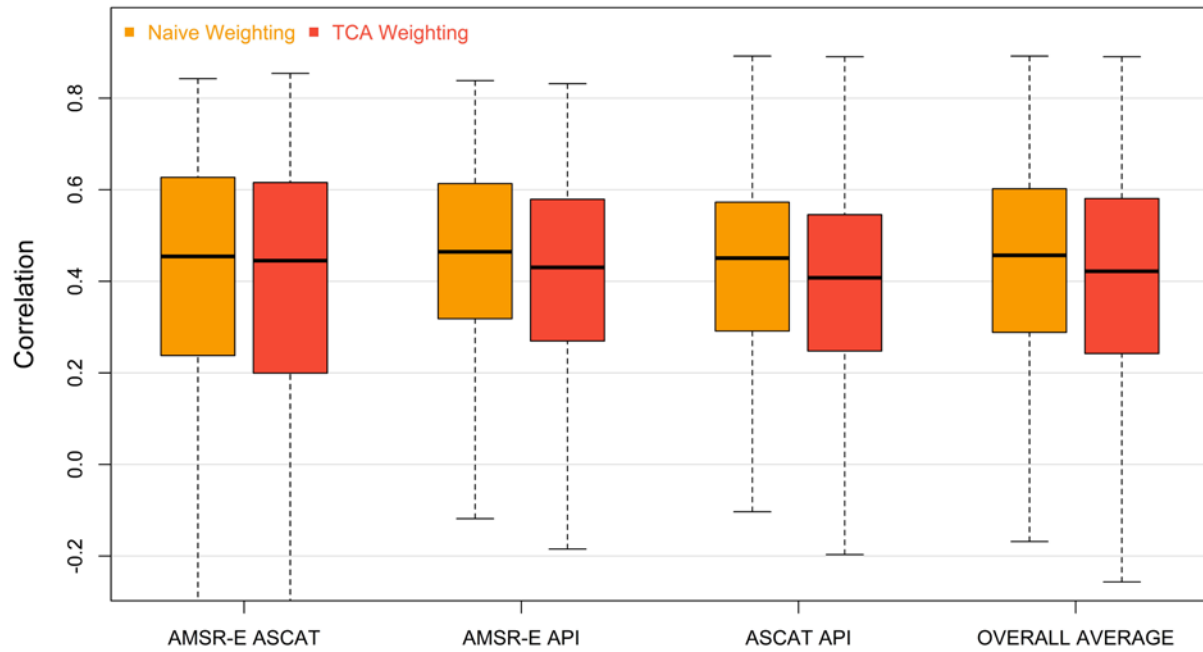
**Figure 3.** For various rescaling reference selections ( $x$ -axis), correlation of pairwise fusion results against in situ values as a function of a) rescaling method, b) stationarity assumption, c) technique, and d) pairwise parent product selection.



**Figure 4.** For various rescaling reference selections (x-axis), correlation of three-way fusion results against in situ values as a function of a) rescaling method, b) stationarity assumption, c) technique, and d) three- and four-way data fusion products.



**Figure 5.** Correlation results of both native and fused soil moisture products for various networks (those networks which have less than four sites are combined to the “OTH” group; full names of the networks plotted along the x-axis are available in Table 1). Fused products are based on the pairwise merger of AMSR-E and ASCAT datasets using a NOAH rescaling reference. Rescaling is based on both the a) no decomposition and b) smooth-deviance techniques.



**Figure 6.** Impact of different weighting approaches (defined in section 2 by eqs. 4-9) on the accuracy of pairwise fusion products.

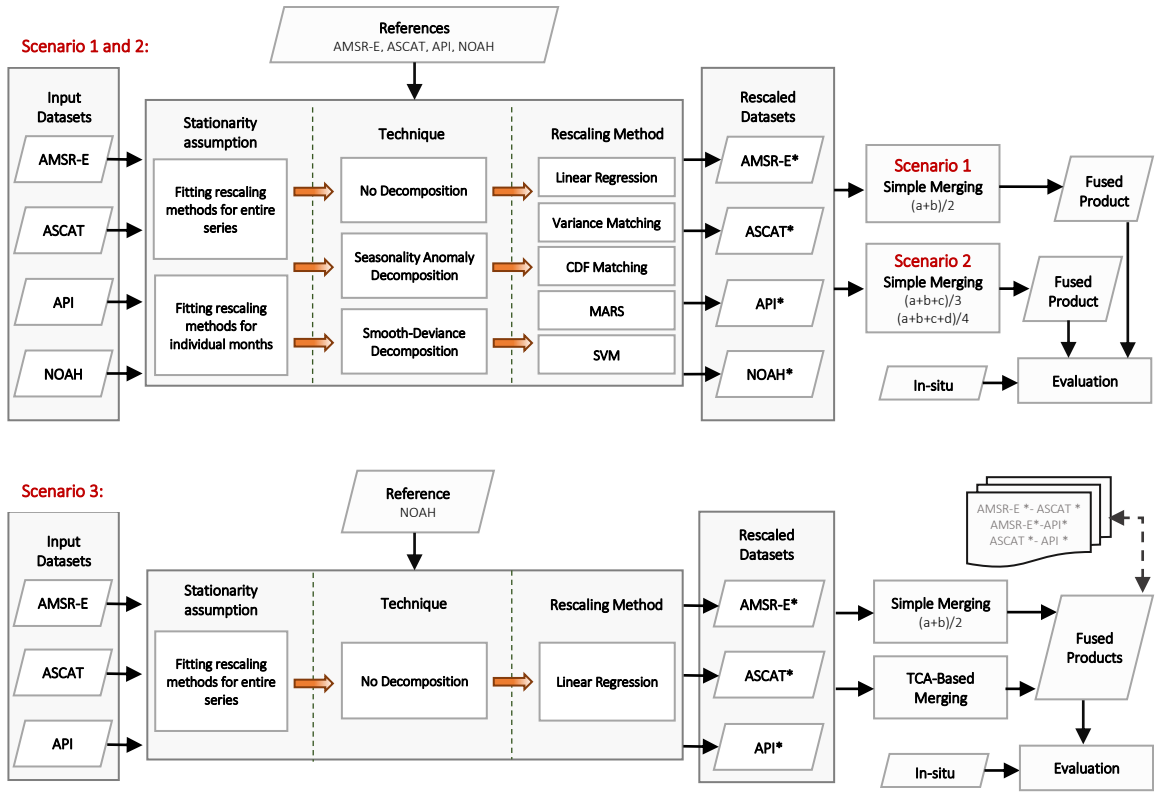
**Table 1.** Description of in situ soil moisture networks used in this study. (Notice that the sites in the USDA-ARS network are comprised of an average of many different measurement locations. While in other networks the used sites are just one single instrument at one single location.)

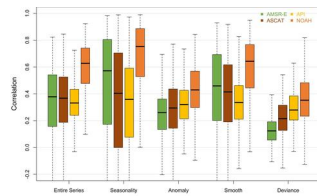
**Table 2.** Summary of statistical results comparing the variability and mutual dependency of different soil moisture products (statistics are averaged across 542 sites). a) The average ratio of both low- and high-frequency (see eqs. 23-24) soil moisture daily time series variance to daily total time series variance, b) The average auto-correlation of the soil moisture time series and their high-frequency components across 542 sites, and c) Average pairwise cross-correlation between different decomposed parts of soil moisture time series.

**Table 3.** Impact of rescaling methods, techniques, and stationarity assumptions on correlation-based accuracy (obtained from averaged comparisons with 542 in situ sites) of pairwise fusion data products (NOAH reference). Constant and time-varying stationarity assumptions are indicated with “C” and “T”, respectively (the best results in each row are written in bold).

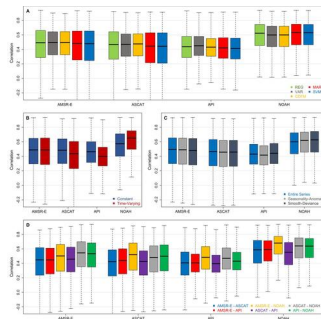
**Table 4.** Impact of rescaling methods, techniques, and stationarity assumptions on correlation-based accuracy (obtained from averaged comparisons with 542 in situ sites) of three- and four-way data fusion products (NOAH reference). Constant and time-varying stationarity assumptions are indicated with “C” and “T”, respectively (the best results in each row are written in bold).

Author Manuscript



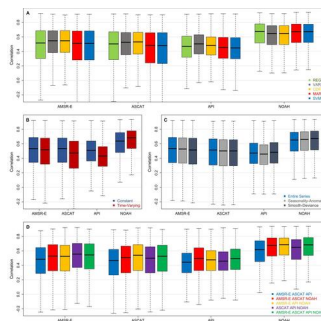


2019WR025111-f02-z-.jpg

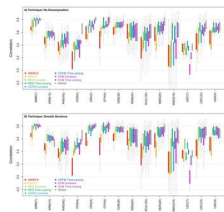


2019WR025111-f03-z-jpg

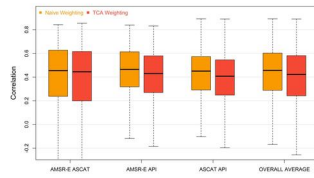




2019WR025111-f04-z-.jpg



2019WR025111-f05-z-.jpg



2019WR025111-f06-z-.jpg