Improving the Combination of Satellite Soil Moisture Data Sets by Considering Error Cross Correlation: A Comparison Between Triple Collocation (TC) and Extended Double Instrumental Variable (EIVD) Alternatives

Seokhyeon Kim, Hung T. Pham, Yi Y. Liu, Lucy Marshall, and Ashish Sharma

***Abstract*— Satellite-derived geophysical variables provide valuable information about the earth's functioning, but there are errors that limit their direct uses. Linearly combining two or more data sources is a simple and efficient method to reduce the uncertainty between the truth and observations. However, calculating the optimal weight for such a linear combination generally needs a reference "truth" that is rarely available in practical applications. To address this limitation, a triple collocation (TC) technique is often used to estimate data error by using a data triplet without the truth. The TC-based error variances are then applied to calculate the optimal weight calculation with an assumption of independence between errors. However, ignoring the error cross-correlations (ECCs) can lead to inaccurate optimal weight for a linear combination and hence may degrade the performance of calculations using merged products. Recently, an extended double instrumental variable (EIVD) method has been proposed to estimate ECC using only three products like TC. In this study, we examined the performance of an EIVD-based linear combination approach through two applications using synthetic data and actual satellite-derived soil moisture (SM) products. For direct comparison, the TC-based linear combination was also implemented using the same data sets. The verification showed that the EIVD-based products are generally better than the TC-based products and result in statistically significant differences in correlation coefficients. While the results here are based on SM, the proposed method is potentially applicable for reconstruction of other data sets that seldom have suitable references in large-scale regional settings.**

***Index Terms*— Cross correlation errors, extended double instrumental variable (EIVD), linear combination, mean squared error (mse), soil moisture (SM), triple collocation (TC).**

I. Introduction

DATA from various sources are often merged to offset random errors using optimal combination alternatives [10]. Linearly combining two or more data sets is a simple and efficient method to archive improvements via the merged product. The principle of linear combination is merging of independent information in each data source, improving over each individual product through cancellation of random errors, with effectiveness depending on the degree of independence across the data sources [12]. Since Bates and Granger [12] proposed minimizing the mean squared error (mse) through optimally combining two forecasts in the context of airline passenger operations, methodologies based on this have been widely applied across various disciplines [14]–[18].

In linear combination for minimizing mse, a key step is the calculation of the optimal weight for each original data set, which is a function of error variances and error cross correlation (ECC) of the original data being combined (hereafter referred to as parent data or parent product). However, the weight estimation process needs a reference "truth" that is seldom available. This combination approach may use an assumed truth (e.g., reanalysis data) to assess how the merged data vary in space and time, but practical uses of such an approach is limited because the expectation of improvement is subject to the selected reference data.

Triple collocation (TC) can estimate error variances and data-truth correlations without a reference by comparing across a triplet of independent data sources [20]–[22]. Given this advantage of TC, it has been widely used to estimate errors of geophysical variables following which the data sets are combined in various applications [23]–[26].

However, the TC-based method relies on the assumption of zero ECC which does not always hold. As shown by Yilmaz and Crow [29], the overall effect of ECC on TC is important and can underestimate the errors in soil moisture (SM) products. As a measure to resolve this issue,

Manuscript received July 16, 2020; revised September 2, 2020 and October 15, 2020; accepted October 16, 2020. This work was supported by the Australian Research Council (ARC) through the Discovery Project under Grant DP200101326. *(Corresponding author: Seokhyeon Kim.)*

Seokhyeon Kim, Lucy Marshall, and Ashish Sharma are with the School of Civil and Environmental Engineering, University of New South Wales (UNSW), Sydney, NSW 2052, Australia (e-mail: seokhyeon.kim@unsw.edu.au; lucy.marshall@unsw.edu.au; a.sharma@unsw.edu.au).

Hung T. Pham is with the Faculty of Water Resources Engineering, Danang University of Science and Technology, Da Nang 550000, Vietnam (e-mail: pthung@dut.udn.vn).

Yi Y. Liu is with the School of Geography and Remote Sensing, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: yi.liu@nuist.edu.cn).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.3032418

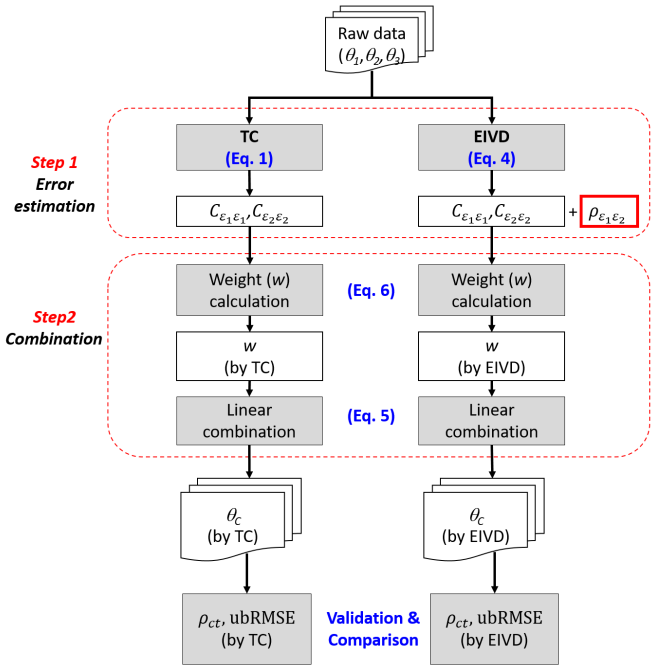


Fig. 1. Flowchart presenting error estimation, combination, and validation of geophysical variables, as applied in this study.

Gruber *et al.* [24] proposed the quadruple collocation (QC) that utilizes a quadruple of data to estimate a part of ECCs by partially adopting the error independence. However, the difficulty in obtaining four data sets for the QC implementation still remains. To ease the difficulty, Dong *et al.* [31] recently proposed the extended double instrumental variable (EIVD) method by which an ECC can be estimated using three data sets.

Given the availability of the abovementioned alternatives, the aim of this study is, therefore, to evaluate and compare TC- and EIVD-based linear combination approaches where the optimal weight minimizes the mse between the combined product and the unknown truth. For this, we first present the theoretical background of the TC, EIVD, and linear combination alternatives for minimizing mse in Section II and then verify the various methods through experiments using synthetic data in Section III-A. In Section III-B, satellite- and model-derived SM data are used as the basis for combination, and ground-based SM measurements are used for evaluating the combined products. Finally, in Section IV, a summary of the results and conclusions are presented.

II. Data and Method

Fig. 1 presents the process of implementing and comparing TC- and EIVD-based linear combination approaches. There are two main steps: 1) error estimation of geophysical variables against the (unknown) truth and 2) calculating the optimal weight for linear combination which minimizes the mse between the combined product and the (unknown) truth. Then, the combination results are validated against independent data sets which are regarded here as the truth.

Three data sets denoted $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_3$ are used for combining $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Here, TC is used to estimate the error variances ($C_{\varepsilon_1\varepsilon_1}$ and $C_{\varepsilon_2\varepsilon_2}$), and EIVD is used to estimate both error variances ($C_{\varepsilon_1\varepsilon_1}$ and $C_{\varepsilon_2\varepsilon_2}$) and ECC ($\rho_{\varepsilon_1\varepsilon_2}$). Then two products $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are linearly combined as $\boldsymbol{\theta}_c$ using the optimal weight (w) which is separately estimated using the TC- and EIVD-derived error metrics. More details of each step are presented in the following sections.

A. TC- and EIVD-Based Linear Combination for Minimizing mse

1) Step 1 (Error Estimation): TC can estimate the error variance of each product and the data-(unknown) truth correlation through intercomparisons between three independent products, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_3$. Each data is linearly related to the truth ($\boldsymbol{t}$) and presented by the sum of three terms, $\boldsymbol{\theta}_i = \alpha_i + \beta_i \boldsymbol{t} + \boldsymbol{\varepsilon}_i$ (linearity assumption), where $i = 1$, 2, and 3; constants α_i and β_i are the additive and multiplicative bias terms, respectively; and $\boldsymbol{\varepsilon}_i$ is random error with zero mean. With this triplet, the error variances ($C_{\varepsilon_i\varepsilon_i}$) and the data-truth Pearson correlation coefficients (ρ_{it}) can be calculated as

$$C_{\varepsilon_i\varepsilon_i} = \begin{bmatrix} \sqrt{C_{11} - \dfrac{C_{12}C_{13}}{C_{23}}} \\ \sqrt{C_{22} - \dfrac{C_{12}C_{23}}{C_{13}}} \\ \sqrt{C_{33} - \dfrac{C_{13}C_{23}}{C_{12}}} \end{bmatrix}, \quad \rho_{it} = \begin{bmatrix} \sqrt{\dfrac{C_{12}C_{13}}{C_{11}C_{23}}} \\ \sqrt{\dfrac{C_{12}C_{23}}{C_{22}C_{13}}} \\ \sqrt{\dfrac{C_{13}C_{23}}{C_{33}C_{12}}} \end{bmatrix} \tag{1}$$

where C_{ij} is the covariance ($i \neq j$) or variance ($i = j$) of the data (here, i and $j = 1$, 2, and 3). Detailed s and their derivations are available in [21].

Dong *et al.* [31] recently developed the EIVD, which is an improved version of the double instrumental variable method [32] and the single instrumental variable method [33]. From a triplet of data like TC, EIVD estimates the error variances, the data-truth Pearson correlation coefficients and an ECC between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, which can be solved through a linear system expressed by the following three matrices:

$$\boldsymbol{e} = \begin{bmatrix} \beta_3^2 C_{tt} \\ \beta_1^2 C_{tt} \\ \beta_2^2 C_{tt} \\ \beta_1\beta_2 C_{tt} \\ C_{\varepsilon_3\varepsilon_3} \\ C_{\varepsilon_1\varepsilon_1} \\ C_{\varepsilon_2\varepsilon_2} \\ C_{\varepsilon_1\varepsilon_2} \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{b} = \begin{bmatrix} C_{33} \\ C_{11} \\ C_{22} \\ C_{12} \\ C_{13}\sqrt{L_{33}/L_{11}} \\ C_{23}\sqrt{L_{33}/L_{22}} \\ C_{13}\sqrt{L_{11}/L_{33}} \\ C_{23}\sqrt{L_{22}/L_{33}} \\ C_{23}\sqrt{L_{11}/L_{33}} \\ C_{13}\sqrt{L_{22}/L_{33}} \end{bmatrix} \tag{2}$$

where $C_{\varepsilon_i\varepsilon_j}$ is the error covariance ($i \neq j$) or error variance ($i = j$) same to the case of TC (i and $j = 1$, 2, and 3); and L_{ii} is the autocovariances of variables ($i = 1$, 2, and 3). Given $\boldsymbol{Ae} = \boldsymbol{b}$, $\boldsymbol{e}$ can be solved by least squares as

$$\boldsymbol{e} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}. \tag{3}$$

In addition to the error variances in $\boldsymbol{e}$, the data-truth Pearson correlation and the ECC can be calculated using the remaining terms in $\boldsymbol{e}$ as

$$\rho_{it} = \sqrt{\frac{\beta_i^2 C_{tt}}{\beta_i^2 C_{tt} + C_{\varepsilon_i\varepsilon_i}}}, \quad \rho_{\varepsilon_i\varepsilon_j} = \frac{C_{\varepsilon_i\varepsilon_j}}{\sqrt{C_{\varepsilon_i\varepsilon_i} C_{\varepsilon_j\varepsilon_j}}} \tag{4}$$

where the subscripts (i and j) denote the data sets. ρ_{it} represents Pearson correlation between $\boldsymbol{\theta}_i$ and the unknown truth ($\boldsymbol{t}$); and $\rho_{\varepsilon_i\varepsilon_j}$ is the ECC of $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$, where $i = 1$ and $j = 2$ in this study.

TC is based on the following four assumptions [22]: 1) linearity between the truth and the errors; 2) stationarity for both the truth and the error; 3) zero ECC; and 4) error-truth orthogonality. Furthermore, as EIVD can be affected by unignorable statistical differences in error autocorrelations in the original data sets [31], this is considered as the fifth assumption in this study. The performances of TC and EIVD relies on how the above five assumptions are satisfied for the given set of variables, and as a result the outcomes of any linear combination based on this method can be accordingly affected. This is further discussed in Section III.

2) Step 2 (Linear Combination for Minimizing mse): Data sets $\boldsymbol{\theta}$ ($N \times P$) are simply combined by assigning weight w ($P \times 1$) to generate a combined product $\boldsymbol{\theta}_c$ ($N \times 1$) as

$$\boldsymbol{\theta}_c = \boldsymbol{\theta}\boldsymbol{w}, \quad \text{Subject to } \sum_{i=1}^{\mathrm{p}} w_i = 1 \tag{5}$$

where N is the sample size and P is the number of data sets. This process is demonstrated with two data sets ($P = 2$) below.

The optimal weight minimizing the mse for two unbiased data sets [12] is presented using error variances ($C_{\varepsilon_1\varepsilon_1}$ and $C_{\varepsilon_2\varepsilon_2}$) and the ECC ($\rho_{\varepsilon_1\varepsilon_2}$) as

$$w_1 = \frac{C_{\varepsilon_2\varepsilon_2} - \rho_{\varepsilon_1\varepsilon_2}\sqrt{C_{\varepsilon_1\varepsilon_1} C_{\varepsilon_2\varepsilon_2}}}{C_{\varepsilon_1\varepsilon_1} + C_{\varepsilon_1\varepsilon_1} - 2\rho_{\varepsilon_1\varepsilon_2}\sqrt{C_{\varepsilon_1\varepsilon_1} C_{\varepsilon_2\varepsilon_2}}}, \quad w_2 = 1 - w_1 \tag{6}$$

where w_1 and w_2 are the optimal weights for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. For a clearer demonstration of the $P = 2$ case, only w_1 is considered and simply denoted as w hereafter. We calculated two types of optimal weight by (6) where the error terms were estimated by TC [see (1)] and EIVD [see (4)], respectively, and then combined [see (5)]. Note that TC assumes $\rho_{\varepsilon_1\varepsilon_2}$ is zero but EIVD takes it into account in the weight calculation. Then we validated and compared the TC- and EIVD-based combination results through two analyses using synthetic data sets and real data sets (satellite- and model-derived SM products), of which details are described in the following section.

B. Evaluation of the Combined Data

The combination results are validated using the data-truth correlation (ρ_{ct}) and unbiased root mse metrics for the combined data (ubRMSE$_{ct}$); and compared against those for the parent data, ρ_{it} and ubRMSE$_{it}$ ($i = 1$, 2). In addition, the differences, $\Delta\rho$ and ΔubRMSE via (7) were also used for the evaluation. Positive $\Delta\rho$ (negative ΔubRMSE) values indicate improvements in the combined product, while zero difference means no improvement compared to the parent products

$$\begin{aligned} \Delta\rho &= \rho_{ct} - \max[\rho_{it}] \\ \Delta\text{ubRMSE} &= \text{ubRMSE}_{ct} - \min[\text{ubRMSE}_{it}]. \end{aligned} \tag{7}$$

1) Evaluation Using Synthetic Data: We evaluated how the TC/EIVD assumptions, described in Section II-A1, affect the linear combination results through the following experiments using synthetic data. For this, a true signal ($\boldsymbol{t}$) is first generated from a uniform distribution $U(0, 1)$ with a sample size of N. Then three synthetic products, $\boldsymbol{\theta}_i$ ($i = 1$, 2, and 3), were generated by adding zero-mean Gaussian errors ($\boldsymbol{\varepsilon}_i$) to $\boldsymbol{t}$ as

$$\boldsymbol{\theta}_i = \boldsymbol{t} + \boldsymbol{\varepsilon}_i. \tag{8}$$

Note that EIVD is based on using lag-1 autocovariance of the observations (i.e., temporal memory of the data). Therefore, the performance of EIVD can be undermined in the experiments using the true signal drawn from U(0,1) in which autocovariance is not explicitly reflected. Given this limit, the result interpretation should focus on patterns and tendencies with changes in control parameters.

We used this synthetic model to evaluate the two merging approaches. Our aim here is to demonstrate that: 1) the combined product outperforms any of the parent products and 2) even though the TC/EIVD assumptions are not perfectly valid, the outcome is still an improved product in comparison to any individual product being used. These arguments are illustrated using five synthetic experiments, Exp1–Exp5, where each assumption is violated one at a time by generating a suite of different additive errors ($\boldsymbol{\varepsilon}_i$).

For considering various noise ranges in the synthetic data, we used three values of signal-to-noise ratio (SNR) in decibels (i.e., SNR$_{\mathrm{dB}} = 0.1$, 1, and 2) in the five synthetic experiments as

$$\text{SNR} = 10^{\frac{\text{SNR}_{\mathrm{dB}}}{10}}, \quad P_n = \frac{P_t}{\text{SNR}} \tag{9}$$

where P_t and P_n are signal power and the noise power, respectively, and estimated as the expected values of square signal values and square noise values, respectively. In this study, we used P_n to generate noise with a specified variance according to the given SNR. The generated data were combined and validated through the two main steps (see Fig. 1), and then this process was repeated 500 times to capture the uncertainty in the combination. The results from the synthetic experiment are presented and discussed in Section III-A.

The first experiment (denoted "Exp1") tests how the sample size (N) violates the TC/EIVD assumptions and thus affects the data combination. Here, we specified six sample sizes

($N = 50$, 100, 200, 500, 1000, and 5000), and the errors defined as

$$\boldsymbol{\varepsilon_i} \sim \mathcal{N}(0, P_n) \tag{10}$$

where $\mathcal{N}(0, P_n)$ represents Gaussian random numbers of zero mean and variance P_n.

The second experiment (denoted "Exp2") aims to evaluate the effects of the stationarity assumption on the combination results. Here, the error mean is monotonically increased in time by including an additive term in (10) as

$$\boldsymbol{\varepsilon_i} = \boldsymbol{a} + \frac{H - 0.5N}{N} \cdot s \cdot E[\boldsymbol{t}], \quad H = 1:L \tag{11}$$

where $\boldsymbol{a}$ is random error drawn from $\mathcal{N}(0, P_n)$ of sample size of N; H values range from 1 to N with increment of 1; s is the increasing slope of the error mean with relation to the expected value of the true signal ($\boldsymbol{t}$) ($E[\boldsymbol{t}]$) and set as 0–10 increasing by 1. Note that, the sample size ($N = 500$) was fixed for Exp2 and the remaining experiments (Exp 3–Exp5).

The third experiment (denoted "Exp3") evaluates how the combination results are changed when the zero ECC assumption is violated. Increasing ECC breaches the TC assumption and also makes the data sets more dependent to each other (i.e., less information). The errors of three synthetic products were defined as

$$\boldsymbol{\varepsilon_i} \sim \mathcal{N}(0, P_n), \quad \boldsymbol{\varepsilon_j} = \rho_{\boldsymbol{\varepsilon_i}\boldsymbol{\varepsilon_j}} \boldsymbol{\varepsilon_i} + \sqrt{1 - \rho^2_{\boldsymbol{\varepsilon_i}\boldsymbol{\varepsilon_j}}}\boldsymbol{a} \tag{12}$$

where $i = 1$ and $j = 2$ and 3; $\rho_{\boldsymbol{\varepsilon_i}\boldsymbol{\varepsilon_j}}$ is cross correlation between $\boldsymbol{\varepsilon_i}$ and $\boldsymbol{\varepsilon_j}$ and set as 0–0.9 increasing by 0.1.

The fourth experiment (denoted "Exp4") investigates the effects of the error-truth orthogonality assumption on the combined product. The errors of the three synthetic data sets were generated as

$$\boldsymbol{\varepsilon_i} = \rho_{\boldsymbol{\varepsilon_i} t} \frac{\boldsymbol{t}}{\sqrt{\mathrm{SNR}}} + \sqrt{1 - \rho^2_{\boldsymbol{\varepsilon_i} t}}\boldsymbol{a} \tag{13}$$

where $i = 1$, 2, and 3; $\rho_{\boldsymbol{\varepsilon_i} t}$ is cross correlation between $\boldsymbol{\varepsilon_i}$ and the true signal ($\boldsymbol{t}$) and set as 0–0.9 increasing by 0.1.

Lastly, the fifth experiment (denoted "Exp5") tests the effects of the contrasted error autocorrelations on the combination results. The error of the first synthetic data set ($\boldsymbol{\varepsilon_1}$) was generated by a first-order autoregressive process with autocorrelation coefficient values ($L_{\boldsymbol{\varepsilon_1}\boldsymbol{\varepsilon_1}}$) varying from 0 to 0.9 increasing by 0.1, and $\boldsymbol{\varepsilon_2}$ and $\boldsymbol{\varepsilon_3}$ were simply generated by (10).

The details of the five experiments are summarized in Table I.

2) Evaluation Using SM Data Sets: We also applied the combination process depicted in Fig. 1 for multiple SM data sets. We used three satellite-derived and one reanalysis SM product as summarized in Table II. The four SM products cover a three-year period from April 2015 to March 2018, and are: 1) level 3 SM Ocean Salinity (SMOS) developed by Institut National de la Recherche Agronomique (INRA) and Centre d'Etudes Spatiales de la BIOsphère (CESBIO), SMOS-IC (Version 105); 2) level 2 Advanced SCATterometers (ASCAT) SM index product (Version 5); 3) level 3 SM Active Passive (SMAP) Radiometer Global Daily SM (Version 5); and 4) the volumetric SM content in the top layer (<0.05 m) from the Modern-Era Retrospective Analysis for Research and Applications Land version 2 reanalysis (MERRA2).

TABLE I
SUMMARY OF THE FIVE SYNTHETIC EXPERIMENTS

Exp.	Control Parameter	Related TC/EIVD Assumption	Related equation
Exp1	Sample size (N)	General	-
Exp2	Slope of the error mean (s)	Stationarity for both the truth and the error	(11)
Exp3	Error cross-correlation ($\rho_{\boldsymbol{\varepsilon_i}\boldsymbol{\varepsilon_j}}$)	Zero ECC	(12)
Exp4	Error-truth cross-correlation ($\rho_{\boldsymbol{\varepsilon_i t}}$)	Error-truth orthogonality	(13)
Exp5	Error autocorrelations ($L_{\boldsymbol{\varepsilon_1}\boldsymbol{\varepsilon_1}}$) only for θ_1	Non-contrast error autocorrelations	-

TABLE II
SUMMARY OF SM PRODUCTS USED IN THIS STUDY

Product name	Data type	Temporal resolution	Spatial resolution	Ref.
SMOS-IC	Passive L-band (1.400-1.427 GHz)	Daily/ Overpass (asc/des) at 6 AM/PM LST	25-km EASE-Grid	[1, 2]
ASCAT	Active C-band (5.3 GHz) radar backscatter	Daily/ Overpass (asc/des) at 9:30 PM/AM LST	25-km	[3, 4]
SMAP	Passive L-band (1.41 GHz)	Daily/ Overpass (asc/des) at 6 PM/AM LST	36-km EASE-Grid	[6]
MERRA2	Land surface model	Hourly (time-averaged)	0.5°×0.625°	[8]

* asc/des: ascending/descending; LST: local solar time. Data links: SMOS-IC: https://www.catds.fr/Products/Available-products-from-CEC-SM/SMOS-IC; ASCAT: https://www.eumetsat.int/website/home/Data/index.html; SMAP and MERRA2: https://earthdata.nasa.gov/

For the SM products, the following data-filtering/processing was consistently applied [25], [34]. For the three satellite-derived SM in both ascending and descending overpasses, we only used SMOS-IC data flagged as "recommended quality"; SMAP data of which open water fraction and frozen fraction are less than 10%, and vegetation water content is smaller than 5 kg/m^2; ASCAT data with less than 10% probability snow, frozen ground, and 50% of retrieval error. Since the error estimation and data combination in this study have been implemented when all data sets are available (i.e., only for paired data points), it has the same effect that all filtering criteria are commonly applied to all SM data. Then, both ascending and descending data for each product were averaged on a daily basis to maximize the spatial coverage of the data. Next, the SM products including MERRA2 at different spatial resolutions were resampled to the global cylindrical 36-km equal-area scalable Earth,

TABLE III
SUMMARY OF ISMN DATA USED IN THIS STUDY

Network	#Stations	Country	Ref.
BIEBRZA-S-1	1	Poland	http://www.igik.edu.pl/en
DAHRA	1	Senegal	[5]
FMI	1	Finland	http://fmiarc.fmi.fi/
HOBE	2	Denmark	[7]
PBO-H2O	36	USA	[9]
REMEDHUS	3	Spain	http://campus.usal.es/~hidrus/
RISMA	6	Canada	[11]
RSMN	18	Romania	http://assimo.meteoromania.ro
SCAN	85	USA	https://www.wcc.nrcs.usda.gov/scan/
SMOSMANIA	8	France	[13]
SNOTEL	42	USA	[19]
SOILSCAPE	3	USA	[27]
TERENO	1	Germany	[28]
USCRN	47	USA	[30]
Total	254		

version 2 (EASEv2) grid using bilinear interpolation. Finally, for the hourly MERRA2 data on each grid cell, we selected SM values closest to the averages of the scan time of the three satellite SM data on a daily basis.

We used ground-based SM obtained from the International Soil Moisture Network [35] (https://ismn.geo.tuwien.ac.at) for evaluating the combination results. For this, we applied a strict data filtering to the ground measurements for minimizing systematic differences to the coarse SM data [36]. The ground data filtering includes: applying the ISMN quality flags [37]; adopting SM series from the shallowest depth (<10 cm) with less topographic complexity and wetland fraction (<10%); selecting SM values temporally closest to the daily average scan time of the three satellite SM products; when there are multiple stations in a grid cell, picking up the most areal representative station which has the highest intercorrelations against the SM products over the grid cell [38]. In addition, we selected stations with at least 50 observations paired with each SM product. As a result, only 254 stations distributed over 15 networks remained as shown in Table III.

These four SM products SMOS-IC (hereafter simply referred to as SMOS), ASCAT, SMAP, and MERRA2 are used as $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_3$ in three cases of combination in this study. Those are C1: SMOS + ASCAT, C2: SMOS + SMAP, C3: ASCAT + SMAP. Note that MERRA2 is used as $\boldsymbol{\theta}_3$ in common for C1, C2, and C3.

Uncertainties in the retrieval process and/or limited equipment precision inevitably produce biases in the modeled dynamic ranges of satellite data, which cause systematic differences among the data sets. In this study, we linearly scaled the three satellite-derived SM products against MERRA2 (i.e., $\boldsymbol{\theta}_3$) to remove the systematic differences among them as

$$\boldsymbol{\theta}_i' = \left(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_i\right)\sigma_3/\sigma_i + \bar{\boldsymbol{\theta}}_3 \tag{14}$$

where $i = 1, 2$ in each combination case (C1, C2, and C3); $\boldsymbol{\theta}_i'$ is the scaled $\boldsymbol{\theta}_i$; σ_i, the standard deviation of $\boldsymbol{\theta}_i$; $\bar{\boldsymbol{\theta}}_i$, the mean of $\boldsymbol{\theta}_i$. Hereafter, $\boldsymbol{\theta}_i'$ is simply denoted as $\boldsymbol{\theta}_i$ for simplification.

The combination results using the 254 ISMN stations were also evaluated under various climate and land cover (LC) classes. For this, we use the three primary climate classes (arid, temperate, and cold) from the updated Koppen–Geiger climate classification [39] and five primary LC classes, forest, shrubland, woodland, grassland, and cropland, from a Moderate Resolution Imaging Spectroradiometer (MODIS)-derived yearly LC map from 2012 (MCD12C1, Version 051) [40].

III. RESULTS AND DISCUSSION

A. Synthetic Experiment Results

In this section, we first present EIVD-based combination results in Section III-A1 and then, in Section III-A2, compare the results with those of the TC-based approach by presenting differences in Pearson correlation and ubRMSE for the two approaches.

1) EIVD-Based Combination Results: As shown in Fig. 2, in terms of correlation, the three SNR cases (i.e., $\mathrm{SNR}_{\mathrm{dB}} =$ 0.1, 1, and 2) are generally similar to each other but the lower SNR is, the higher performances are for the ubRMSE. This suggests that higher SNR values lead to smaller error variations while smaller SNR values have higher variability in the error time series. This is an important factor to achieve synergistic results in the combined product.

In Exp1, the variability of $\Delta\rho$ and ΔubRMSE reduces with increasing sample sizes. This suggests that increased availability of data can result in less uncertainty in the improvement of the combined product. The findings suggest that a sample size of 500 or more should be used for the combination.

Exp2 evaluates changes in the combination results when the stationarity assumption is violated. Exp2 in Fig. 2(a) and (b) shows that the performance of the combined product in both $\Delta\rho$ and ΔubRMSE significantly degrades with increasing s. However, such conditions ($s > 2$) indicates that the error mean is greater than twice that of the truth and likely are not seen in practice.

Exp3 investigates the effects of the ECC ($\rho_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j}$) on the performance of the combined product [Exp3 in Fig. 2(a) and (b)]. In general, both $\Delta\rho$ and ΔubRMSE of all three SNR cases are linearly reduced with increasing $\rho_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j}$. The levels of ECC in SM data which have been reported in existing studies [24], [25], [29] suggest that the ECC should be carefully considered for a linear combination. This is one of the main points to be investigated in this study and is further discussed in Section III-B.

Exp4 tests how the error-truth orthogonality impacts the combination results [Exp4 in Fig. 2(a) and (b)]. The combination performance in ρ and ubRMSE is degraded when the

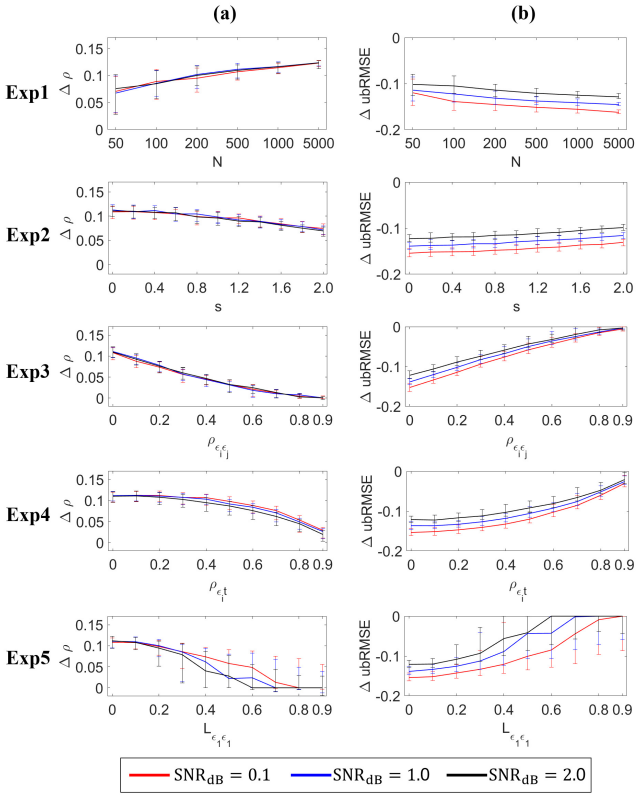


Fig. 2. Results of combining synthetic products in (a) Pearson correlation and (b) ubRMSE with Exp1: various sample sizes (N); Exp2: error means (s); Exp3: cross-correlations between the errors ($\rho_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j}$); Exp4: cross-correlations between the errors and the truth ($\rho_{\boldsymbol{\varepsilon}_i t}$); and Exp5: contrasted error autocorrelations ($L_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j}$). The y-axis represent (a) $\Delta\rho$ and (b) ΔRMSE defined by (7).

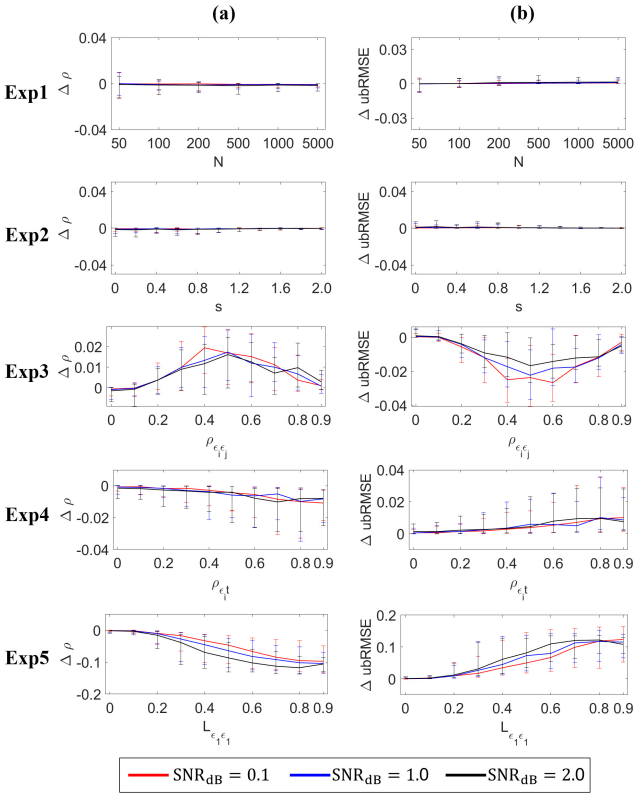


Fig. 3. Comparisons of the TC- and EIVD-combined products through Exp1–Exp5. Here, $\Delta\rho$ and ΔubRMSE on the y-axis represent differences in (a) Pearson correlation and (b) ubRMSE of the TC- and EIVD-combined products, i.e., EIVD minus TC.

cross correlation between error and the truth, i.e., $\rho_{\boldsymbol{\varepsilon}_i t}$, become significant. The degradation of $\Delta\rho$ and ΔubRMSE values of the two low SNR cases is sustained when $\rho_{\boldsymbol{\varepsilon}_i t}$ is smaller than 0.4, while the performance is disturbed when $\rho_{\boldsymbol{\varepsilon}_i t} > 0.4$. The more rapid degradation rates in Exp3 than Exp4 suggest that the ECC should be more importantly considered than $\rho_{\boldsymbol{\varepsilon}_i t}$ in the combination.

Finally, as shown in Exp5 in Fig. 2(a) and (b), high error autocorrelations in $\boldsymbol{\varepsilon}_1$ ($L_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1} > 0.2$) significantly degrades the combination performances. However, it should be noted that these degradations in the EIVD-based combined products are related to highly contrasting error autocorrelations among the original data sets rather than error autocorrelations itself [31]. Therefore, it is necessary to practically consider the contrast in error autocorrelations of products being merged, of which examples using SM products are presented and discussed in Section III-B1.

2) Differences Between EIVD- and TC-Derived Results: Fig. 3 presents differences in Pearson correlation and ubRMSE of the combined data by the two methods, i.e., EIVD minus TC.

While Exps 1 and 2 in Fig. 3 do not show any considerable differences between the EIVD and TC results, in Exp3 in Fig. 3 regarding the ECC ($\rho_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j}$), EIVD presents improvements in both Pearson correlation and ubRMSE. Best cases of these improvements appear around $\rho_{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j} = 0.5$; by which the EIVD-based data merging can somewhat complement the ignored ECC in the TC-based combination approach. For Exp 4 regarding cross-correlations between the errors and the truth ($\rho_{\boldsymbol{\varepsilon}_i t}$), the three SNR cases are stable when $\rho_{\boldsymbol{\varepsilon}_i t}$ is moderate (<0.3) but are getting instable with higher $\rho_{\boldsymbol{\varepsilon}_i t}$ values. A possible reason for these unstable results is that the error variances with high $\rho_{\boldsymbol{\varepsilon}_i t}$ values are relatively small and therefore error autocovariances among the parent data sets sensitively contrasted to each other through the randomly generated errors, which can violate the EIVD assumption, i.e., noncontrasted error autocovariances. More importantly, the results of Exp5 suggest limits of the EIVD-based data combination approach. As noted by Exp5 in Fig. 2, compared to the TC-based results, the EIVD-based results are highly degraded by high error autocorrelations in $\boldsymbol{\varepsilon}_1$ ($L_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1} > 0.3$) that can violate the assumption of noncontrasted error autocovariances in the EIVD approach.

In summary, the EIVD-based combination approach somewhat complements the TC-based method except for the special cases under high SNR, low s, and, most considerably, contrasted error autocorrelation values. Given that the EIVD-based approach has the advantages and disadvantages stated through the synthetic experiments, it is important to investigate how the approach performs with real data. Therefore, in the next section, we evaluate the method

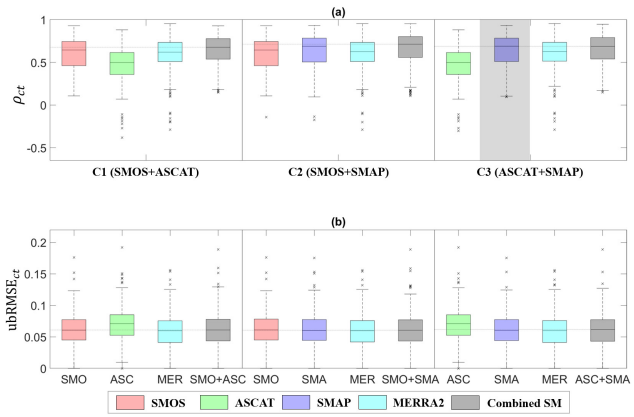


Fig. 4. Results of combining SM products with three combinations, C1–C3, in terms of (a) Pearson correlation and (b) ubRMSE using ground measurements as the truth. Here, the shaded box plot in (a) indicates that statistically insignificant differences in average values of correlation coefficients for combined and parent products according to one-sided paired t-tests at a significance level of 0.05.

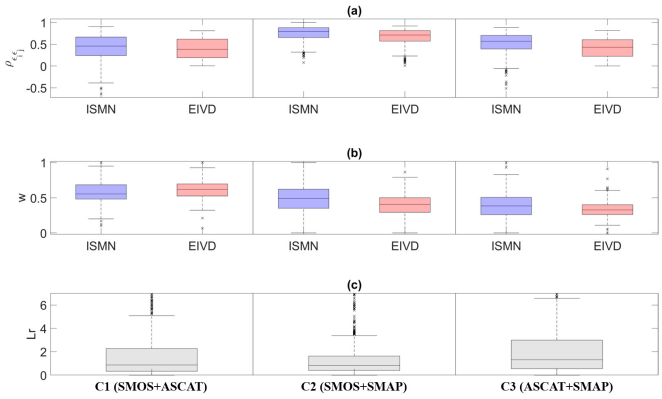


Fig. 5. Comparisons of ISMN data- and EIVD-derived (a) ECC ($\rho_{\varepsilon i \varepsilon j}$), (b) optimal weight ($w$) for C1, C2, and C3, and (c) ratios of error autocovariances of SM products.

with using the satellite-derived SM data as described in Section II-B2.

B. SM Combination Results

According to the results of the EIVD-based combination from the experiments using synthetic data, in this section we first present the EIVD-based SM combination results. Then we discuss the combination performances conditioned by various climate and LC classes and compare with TC-based results. Note that only statistically significant ($\alpha = 0.05$) correlation coefficients were included for the result presentation in this section.

1) EIVD-Based SM Combination Results: Fig. 4 presents the validation results of the three cases of combining SM products using ground measurements from the 254 ISMN stations as the truth.

Fig. 4(a) shows box plots of the Pearson correlation coefficients between the combined products and the ground measurements regarded as the truth. There is substantial improvement in the combined products compared to the ASCAT product and considerable enhancement compared to MERRA2 data which was not merged but used as the third product to implement TC and EIVD. The combined product has slightly better performance than SMOS and SMAP products. For this, one-sided paired t-tests at a significance level ($\alpha = 0.05$) were used to check whether differences in average values of correlation coefficients for combined and parent products over the ground stations is greater than zero (i.e., $\mu_d > 0$). Consequently, C3 shows statistically insignificant results against one of the two parent products, i.e., SMAP, presented as shaded box plots in Fig. 4(a). Numerical results of the paired t-tests are presented in Table S1 in Supporting Information. Unlike these differences in the Pearson correlation, the differences in ubRMSE are not considerably different from each other as shown in Fig. 4(b). This is because the original SM products were rescaled against MERRA2 by (14) to remove the systematic differences among them. Therefore, contents to be described hereafter are focused on the Pearson correlation.

Based on the synthetic experimental results, three possible reasons for the combination results are considered including short data lengths, nonnegligible ECC and highly contrasted error autocorrelations of the SM products. For the first reason, the number of paired SM data used for the combination at the global scale is just 187 ± 87 over the three-year period with shorter data lengths at mid-latitude regions. To investigate the second reason, we calculated ECC ($\rho_{\varepsilon_1\varepsilon_2}$) of the two parent products ($\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$) by using the ISMN data for the three combination cases: C1, C2, and C3. Errors of the satellite SM data were regarded as residuals from a linearly fit model for the SM data ($\boldsymbol{\theta}_i$) which is a function of the ground measurements ($\boldsymbol{t}$), i.e., $\boldsymbol{\theta}_i = f(\boldsymbol{t}) = \alpha_i + \beta_i \boldsymbol{t} + \boldsymbol{\varepsilon}_i$. Here, $i = 1$ and 2; constants α_i and β_i are additive and multiplicative biases to be estimated; $\boldsymbol{\varepsilon}_i$ is error series. For each combination case, the ISMN-derived $\rho_{\varepsilon_1\varepsilon_2}$ values are compared with the EIVD-derived $\rho_{\varepsilon_1\varepsilon_2}$ values in Fig. 5(a); the optimal weights calculated by the ISMN data and EIVD are compared in Fig. 5(b); the ratios of error autocovariances (Lr) in Fig. 5(c). Here, with three products ($\boldsymbol{\theta}_i = \alpha_i + \beta_i \boldsymbol{t} + \boldsymbol{\varepsilon}_i$, $i = 1$, 2, and 3) and error autocovariances ($L_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1}$, $L_{\boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2}$, and $L_{\boldsymbol{\varepsilon}_3\boldsymbol{\varepsilon}_3}$) considered in EIVD, three Lr values are defined as (15). For the EIVD approach to be effective, the three Lr values should be closer to one [31]

$$\mathrm{Lr} = \left[\frac{L_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1}\alpha_2^2}{L_{\boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2}\alpha_1^2} \; \frac{L_{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1}\alpha_3^2}{L_{\boldsymbol{\varepsilon}_3\boldsymbol{\varepsilon}_3}\alpha_1^2} \; \frac{L_{\boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2}\alpha_3^2}{L_{\boldsymbol{\varepsilon}_3\boldsymbol{\varepsilon}_4}\alpha_2^2} \right]^T. \tag{15}$$

As shown in Fig. 5(a) and Table S2 in Supporting Information, the highest interquartile range of ECC calculated by the ISMN data appeared at C2 which is from the triplet of SMOS, SMAP, and MERRA2, while C1 and C3 present relatively low ranges which are from the triplets of SMOS, ASCAT, and MERRA2; and ASCAT, SMAP, and MERRA2, respectively. These smaller ECC ranges in C1 and C3 are in line with lower ECCs among three SM products derived from passive and active microwave, and a land surface model, as found in [24] and [25]. The EIVD-derived ECCs are

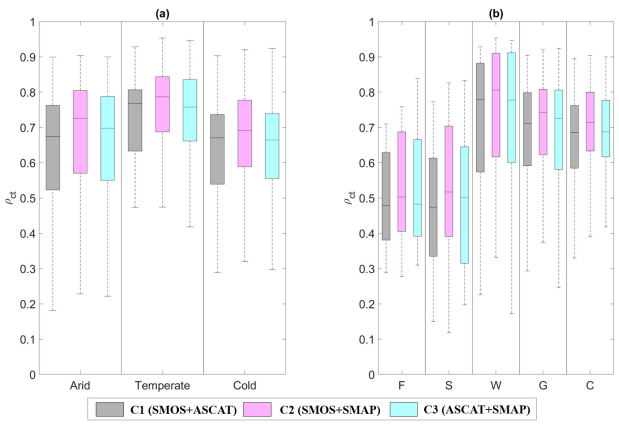


Fig. 6. Pearson correlation of EIVD-based combination over (a) climate (arid, temperate, and cold) and (b) LC (F = Forest, S = Shrubland, W = Woodland, G = Grassland, and C = Cropland) classes.

generally underestimated compared to the ISMN-based ECCs. This is because the ISMN-based ECC uses the ISMN as the "truth" and accordingly the ISMN-based ECC partly includes the error variance of itself. Note that TC regards these ECCs as zeros and such noticeable discrepancies in ECCs violating the TC assumption degrades the combination performances as presented in Exp3 of Fig. 3.

Fig. 5(b) presents the combination weight for each product in each case, calculated by (6). The weight can be interpreted as the relative strength between two products in terms of minimizing mse against the ground measurements in which SMOS and MERRA present similar results to each other, followed by ASCAT. As shown, the EIVD-derived weights are generally underestimated compared to the ISMN-based results, excepting C1.

Lastly, Fig. 5(c) shows box plots presenting ratios of error autocovariances [Lr, (15)] of two merged SM products and one reference product (i.e., MERRA2) for C1, C2, and C3. Their interquartile ranges are varied from 0.33 to 2.99 and it is observed extreme outliers (>6) which can degrade the combination performances based on the EIVD approach.

2) Performances of EIVD-Based Combined SM Products Over Climate and LC Classes: In this section, we investigate the performances of the EIVD-based results represented by the Pearson correlation (ρ_{ct}) over various climate and LC classes.

Fig. 6(a) presents ρ_{ct} over three climate classes, arid, temperate, and cold over which the ISMN stations are distributed. While C1 (SMOS + ASCAT) and C3 (ASCAT + SMAP) have similar interquartile ranges with each other, those of C2 (SMOS + SMAP) are higher than C1 and C3. The stations in temperate areas generally show the highest ρ_{ct}, followed by arid and cold regions. As shown in Fig. 6(b), the ρ_{ct} distributions are more contrasted to each other by the five LC classes: forest, shrubland, woodland, grassland, and cropland. Like the cases by the climate classes, C2 generally presents higher ρ_{ct} over all LC classes than C1 and C3. Here, the mean values of C1 and C3 are statistically indifferent except for the case of Grassland at a significant level ($\alpha = 0.05$).

TABLE IV
p-VALUES OF ONE-SIDED PAIRED t-TESTS TO TEST HYPOTHESES

Category		C1 (SMOS+ ASCAT)	C2 (SMOS+ SMAP)	C3 (ASCAT+ SMAP)
Overall		0.000	0.000	0.000
Climate	Arid	0.010	0.005	0.013
	Temperate	0.009	**0.134**	0.002
	Cold	0.003	0.006	0.029
Land cover	Forest	**0.207**	0.026	**0.263**
	Shrubland	**0.071**	**0.303**	**0.089**
	Woodland	**0.387**	**0.150**	**0.195**
	Grassland	0.003	0.003	0.002
	Cropland	0.000	0.032	0.000

* null hypothesis, $\mathcal{H}_0$: $\bar{\rho}_{\mathrm{ct,EIVD}} = \bar{\rho}_{\mathrm{ct,TC}}$; and alternative hypothesis $\mathcal{H}_1$: $\bar{\rho}_{\mathrm{ct,EIVD}} > \bar{\rho}_{\mathrm{ct,TC}}$. Here, the bold underlined values are larger than a significance level (α=0.05), by which the differences are statistically insignificant at the significance level.

Regarding the performance by LC classes, the highest ρ_{ct} values are observed from the stations in woodland, followed by grassland and cropland; and shrubland and forest. These combination results indicate that the complementarity by the different algorithms of SMOS and SMAP using L-band in common is more effective than that from the different SM retrieval types, i.e., active: ASCAT and passive: SMOS and SMAP.

3) Comparison With TC-Based Results: In this section, we compare the EIVD-derived results with those of the TC-based approach. For this, the student's one-sided paired t-tests were implemented to test if mean values of Pearson correlation of combined SM products based on EIVD ($\bar{\rho}_{\mathrm{ct,EIVD}}$) is larger than those of the TC-based approach ($\bar{\rho}_{\mathrm{ct,TC}}$). Here, null hypothesis, $\mathcal{H}_0 : \bar{\rho}_{\mathrm{ct,EIVD}} = \bar{\rho}_{\mathrm{ct,TC}}$; and alternative hypothesis $\mathcal{H}_1 : \bar{\rho}_{\mathrm{ct,EIVD}} > \bar{\rho}_{\mathrm{ct,TC}}$. Including the overall performances, it was also tested for the three climate zones and five LCs and the results are summarized as p-values in Table IV and related box plots are presented in Figs. S1–S3 in Supporting Information.

Generally, the EIVD-based combination approach generally produces better results than the TC-based approach as presented by the small p-values for overall C1, C2, and C3. In case of performances by climate zones, most of the cases, except for C2 in Temperate, show statistically significant results with small p-values. For the LCs, while the results are generally insignificant for forest, shrubland and woodland, the EIVD-based combined products over grassland and cropland generally present significant results. As a possible reason

for this, as presented in Fig. S4(b) in Supporting Information, compared to other LC classed, ranges of Lr values [see (15)] for forest and shrubland highly deviate from 1 which is regarded as the ideal condition for the EIVD-based approach, while the values by the climate classes are not contrasted to each other [see Fig. S4(a)]. An interesting point is that the p-values of the three combined products (C1, C2, and C3) are complementary to each other over the LC classes. For example, over shrubland, the p-value of C2 is higher than others, but the pattern is opposite for woodland. This complementarity suggests that merging the three products rather than two of them could enhance the performances.

C. Caveats and Outlook

The favorable results presented here suggest that the EIVD-based data merging is a viable way to improve satellite data without reference. Nevertheless, the following future works are worthwhile to further enhance its performance and applicability.

First, further to the case study using the three-year SM products in this study, the method should be tested with more various data sets with longer sample sizes and wider spatial domains. Such applications can further explain the strengths and weaknesses of the proposed method. In particular, considering more diverse parent data sets make us utilize unique features of each product, resulting from differences in retrieval algorithms, observation types (passive and active), electromagnetic bands, and so on.

In this study, the linear rescaling method by (14) was used to adjust the mean and standard deviation of the satellite SM data sets for the reference. It would be worth exploring in the future works whether other scaling methods could lead to further improvements to the combined data [32], [41], [42].

Lastly, it would be useful to investigate if the EIVD-based combination approach is applicable for developing long-term global data sets which are superior to individual products.

IV. Conclusion

In this study, we evaluated and compared the TC-and EIVD-based linear combination approaches which minimize mse against the hidden truth. For this, we implemented two sets of experiments using: 1) synthetic data sets and 2) variously sourced four SM data sets.

Three main conclusions drawn from the synthetic experiments are: 1) a reasonable sample size of data is necessary to properly characterize data weights; 2) the ECC should be considered in the combination process, for which the EIVD-based approach partially complements the ignored ECC in the TC-based method; and 3) the EIVD-based approach can be limited under contrasting error autocorrelation values.

The EIVD-based combination results using SM products showed that eight of nine cases presented significantly different average values of Pearson correlation coefficients for combined and parent products over the ground stations. In the results, compared to the ECCs and optimal weights derived from the ISMN stations for the two merged SM products, the EIVD generally underestimated both values because the ISMN-based ECC uses the ISMN as the "truth" by which the ISMN-based ECC partially contains its own error variance. The three combined SM products outperformed over temperate (climate) and woodland (LC), and the combination of SMOS and SMAP showed better performances than the cases of SMOS + ASCAT and ASCAT + SMAP. Compared to the TC-based results, the EIVD-based approach generally presents better results.

The findings suggest the EIVD-based combination can be potentially used for various geophysical variables especially in large spatial scales where the truth reference is seldom available, e.g., satellite-based global flood monitoring [43], [44].

References

[1] R. Fernandez-Moran *et al.*, "SMOS-IC: An alternative SMOS soil moisture and vegetation optical depth product," *Remote Sens.*, vol. 9, no. 5, p. 457, May 2017.

[2] R. Fernandez-Moran *et al.*, "A new calibration of the effective scattering albedo and soil roughness parameters in the SMOS SM retrieval algorithm," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 62, pp. 27–38, Oct. 2017.

[3] J. M. Wagner, U. Shamir, and D. H. Marks, "Water distribution reliability: Analytical methods," *J. Water Resour. Planning Manage.*, vol. 114, no. 3, pp. 253–275, May 1988, doi: 10.1061/(Asce)0733-9496(1988)114:3(253).

[4] V. Naeimi, K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner, "An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 1999–2013, Jul. 2009, doi: 10.1109/TGRS.2008.2011617.

[5] T. Tagesson *et al.*, "Ecosystem properties of semiarid Savanna grassland in West Africa and its relationship with environmental variability," *Global Change Biol.*, vol. 21, no. 1, pp. 250–264, Jan. 2015.

[6] P. E. O'Neill, S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish, *SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 5 (SPL3SMP)*. Washington, DC, USA: NASA National Snow and Ice Data Center Distributed Active Archive Center, 2020, doi: 10.5067/ZX7YX2Y2LHEB.

[7] S. Bircher, N. Skou, K. H. Jensen, J. P. Walker, and L. Rasmussen, "A soil moisture and temperature network for SMOS validation in Western Denmark," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 5, pp. 1445–1463, May 2012.

[8] Global Modeling and Assimilation Office (GMAO). *MERRA-2 tavg1_2d_lnd_Nx: 2d, 1-Hourly, Time-Averaged, Single-Level, Assimilation, Land Surface Diagnostics V5.12.4*. Greenbelt, MD, USA: Goddard Earth Sciences Data and Information Services Center (GES DISC). Accessed: Oct. 9, 2018, doi: 10.5067/RKPHT8KC1Y1T.

[9] K. M. Larson, E. E. Small, E. D. Gutmann, A. L. Bilich, J. J. Braun, and V. U. Zavorotny, "Use of GPS receivers as a soil moisture network for water cycle studies," *Geophys. Res. Lett.*, vol. 35, no. 24, 2008, Art. no. L24405.

[10] S. Kim, Y. Y. Liu, F. M. Johnson, R. M. Parinussa, and A. Sharma, "A global comparison of alternate AMSR2 soil moisture products: Why do they differ?" *Remote Sens. Environ.*, vol. 161, pp. 43–62, May 2015, doi: 10.1016/j.rse.2015.02.002.

[11] E. R. Ojo, P. R. Bullock, J. L'Heureux, J. Powers, H. McNairn, and A. Pacheco, "Calibration and evaluation of a frequency domain reflectometry sensor for real-time soil moisture monitoring," *Vadose Zone J.*, vol. 14, no. 3, Mar. 2015, Art. no. vzj2014.08.0114.

[12] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *J. Oper. Res. Soc.*, vol. 20, no. 4, pp. 451–468, Dec. 1969, doi: 10.1057/jors.1969.103.

[13] J.-C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, and B. Piguet, "*In situ* soil moisture observations for the CAL/VAL of SMOS: The SMOSMANIA network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2007, pp. 1196–1199.

[14] D. Choudhury, R. Mehrotra, A. Sharma, A. S. Gupta, and B. Sivakumar, "Effectiveness of CMIP5 decadal experiments for interannual rainfall prediction over Australia," *Water Resour. Res.*, vol. 55, no. 8, pp. 7400–7418, Aug. 2019.

[15] M. Araujo and M. New, "Ensemble forecasting of species distributions," *Trends Ecol. Evol.*, vol. 22, no. 1, pp. 42–47, Jan. 2007.

[16] A. Timmermann, "Forecast combinations," *Handbook Econ. Forecasting*, vol. 1, pp. 135–196, Jan. 2006.

[17] H. T. Pham, L. Marshall, F. Johnson, and A. Sharma, "A method for combining SRTM DEM and ASTER GDEM2 to improve topography estimation in regions without reference data," *Remote Sens. Environ.*, vol. 210, pp. 229–241, Jun. 2018, doi: 10.1016/j.rse.2018.03.026.

[18] C. Wasko, A. Sharma, and P. Rasmussen, "Improved spatial prediction: A combinatorial approach," *Water Resour. Res.*, vol. 49, no. 7, pp. 3927–3935, Jul. 2013, doi: 10.1002/wrcr.20290.

[19] G. Leavesley *et al.*, "A modeling framework for improved agricultural water supply forecasting," in *Proc. AGU Fall Meeting Abstr.*, vol. 1, 2008, p. 0497.

[20] A. Stoffelen, "Toward the true near-surface wind speed: Error modeling and calibration using triple collocation," *J. Geophys. Res., Oceans*, vol. 103, no. C4, pp. 7755–7766, Apr. 1998.

[21] K. A. McColl, J. Vogelzang, A. G. Konings, D. Entekhabi, M. Piles, and A. Stoffelen, "Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target," *Geophys. Res. Lett.*, vol. 41, no. 17, pp. 6229–6236, Sep. 2014, doi: 10.1002/2014gl061322.

[22] A. Gruber, C.-H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, "Recent advances in (soil moisture) triple collocation analysis," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 45, pp. 200–211, Mar. 2016, doi: 10.1016/j.jag.2015.09.002.

[23] K. Scipal, W. Dorigo, and R. deJeu, "Triple collocation— A new tool to determine the error structure of global soil moisture products," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 4426–4429, doi: 10.1109/IGARSS.2010.5652128.

[24] A. Gruber, C.-H. Su, W. T. Crow, S. Zwieback, W. A. Dorigo, and W. Wagner, "Estimating error cross-correlations in soil moisture data sets using extended collocation analysis," *J. Geophys. Res., Atmos.*, vol. 121, no. 3, pp. 1208–1219, Feb. 2016, doi: 10.1002/2015jd024027.

[25] F. Chen *et al.*, "Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation," *Remote Sens. Environ.*, vol. 214, pp. 1–13, Sep. 2018, doi: 10.1016/j.rse.2018.05.008.

[26] A. Gruber, W. A. Dorigo, W. Crow, and W. Wagner, "Triple collocation-based merging of satellite soil moisture retrievals," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6780–6792, Dec. 2017, doi: 10.1109/Tgrs.2017.2734070.

[27] M. Moghaddam *et al.*, "Soil moisture sensing controller and optimal estimator (SoilSCAPE): First deployment of the wireless sensor network and latest progress on soil moisture satellite retrieval validation strategies," in *Earth Science Technology Forum*, 2010, pp. 1–9.

[28] S. Zacharias *et al.*, "A network of terrestrial environmental observatories in Germany," *Vadose Zone J.*, vol. 10, no. 3, pp. 955–973, Aug. 2011.

[29] M. T. Yilmaz and W. T. Crow, "Evaluation of assumptions in soil moisture triple collocation analysis," *J. Hydrometeorology*, vol. 15, no. 3, pp. 1293–1302, Jun. 2014, doi: 10.1175/JHM-D-13-0158.1.

[30] J. E. Bell *et al.*, "US climate reference network soil moisture and temperature observations," *J. Hydrometeorol.*, vol. 14, no. 3, pp. 977–988, 2013.

[31] J. Dong, L. Wei, X. Chen, Z. Duan, and Y. Lu, "An instrument variable based algorithm for estimating cross-correlated hydrological remote sensing errors," *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124413.

[32] J. Dong, W. T. Crow, Z. Duan, L. Wei, and Y. Lu, "A double instrumental variable method for geophysical product error estimation," *Remote Sens. Environ.*, vol. 225, pp. 217–228, May 2019.

[33] C.-H. Su, D. Ryu, W. T. Crow, and A. W. Western, "Beyond triple collocation: Applications to soil moisture monitoring," *J. Geophys. Res., Atmos.*, vol. 119, no. 11, pp. 6419–6439, Jun. 2014, doi: 10.1002/2013JD021043.

[34] H. Kim *et al.*, "Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products," *Remote Sens. Environ.*, vol. 204, pp. 260–275, Jan. 2018, doi: 10.1016/j.rse.2017.10.026.

[35] W. A. Dorigo *et al.*, "The international soil moisture network: A data hosting facility for global *in situ* soil moisture measurements," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 5, pp. 1675–1698, May 2011, doi: 10.5194/hess-15-1675-2011.

[36] R. Zhang, S. Kim, and A. Sharma, "A comprehensive validation of the SMAP enhanced level-3 soil moisture product using ground measurements over varied climates and landscapes," *Remote Sens. Environ.*, vol. 223, pp. 82–94, Mar. 2019, doi: 10.1016/j.rse.2019.01.015.

[37] W. A. Dorigo *et al.*, "Global automated quality control of *in situ* soil moisture data from the international soil moisture network," *Vadose Zone J.*, vol. 12, no. 3, Aug. 2013, Art. no. vzj2012.0097, doi: 10.2136/vzj2012.0097.

[38] W. A. Dorigo *et al.*, "Evaluation of the ESA CCI soil moisture product using ground-based observations," *Remote Sens. Environ.*, vol. 162, pp. 380–395, Jun. 2015, doi: 10.1016/j.rse.2014.07.023.

[39] M. C. Peel, B. L. Finlayson, and T. A. Mcmahon, "Updated world map of the Köppen-Geiger climate classification," *Hydrol. Earth Syst. Sci. Discuss.*, vol. 4, no. 2, pp. 439–473, Mar. 2007.

[40] M. A. Friedl *et al.*, "MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 168–182, Jan. 2010, doi: 10.1016/j.rse.2009.08.016.

[41] Y. Y. Liu *et al.*, "Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 2, pp. 425–436, Feb. 2011, doi: 10.5194/hess-15-425-2011.

[42] J. Dong and W. T. Crow, "The added value of assimilating remotely sensed soil moisture for estimating summertime soil moisture-air temperature coupling strength," *Water Resour. Res.*, vol. 54, no. 9, pp. 6072–6084, Sep. 2018.

[43] T. De Groeve, Z. Kugler, and G. R. Brakenridge, "Near real time flood alerting for the global disaster alert and coordination system," in *Proc. 4th Int. ISCRAM Conf.*, B. Van de Walle, P. Burghardt, and C. Nieuwenhuis, eds. Delft, The Netherlands, 2007, pp. 33–39.

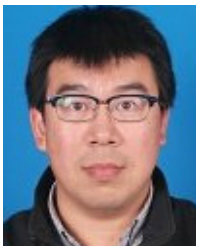
[44] S. Kim and A. Sharma, "The role of floodplain topography in deriving basin discharge using passive microwave remote sensing," *Water Resour. Res.*, vol. 55, no. 2, pp. 1707–1716, Feb. 2019, doi: 10.1029/2018wr023627.

Seokhyeon Kim received the B.E. and M.E. degrees from Korea University, Seoul, South Korea, in 2001 and 2008, respectively, and the Ph.D. degree from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2017.

He is a Research Associate with Water Research Centre, UNSW. His research interests include the use and refinement of satellite remote sensing-based Earth observation system methodologies, hydrological modeling and prediction, water resources system optimization, climate change projection of hydrological extremes, and hydrological spatiotemporal data analysis.

Hung T. Pham received the B.E. degree from the Danang University of Science and Technology, Da Nang, Vietnam, in 2007, the M.E. degree from The University of Adelaide, Adelaide, SA, Australia, in 2014, and the Ph.D. degree from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2019.

He is a Lecturer with the Danang University of Science and Technology. His research interests include the flood forecasting in ungauged areas by using remotely sensed data, drought prediction using machine learning techniques and satellite observations, quantifying uncertainty, and assessing the effects of uncertainty on hydrological extremes.



Yi Y. Liu received the B.E. degree from Jilin University, Changchun, China, in 2003, the M.Sc. degree from Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, in 2007, and the Ph.D. degree in remote sensing in hydrology from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2012.

In the past years, based on the observations from a series of passive microwave satellites, he has been working on generating long-term data sets of soil moisture and vegetation water content, investigating their long-term and interannual dynamics, and identifying the natural and human factors for these changes at regional to global scales. He is working with the School of Geography and Remote Sensing, Nanjing University of Information Science and Technology (NUIST), Nanjing, China. His research interests include using satellite-based observations to investigate the hydrological cycle, e.g., precipitation, radiation, atmospheric conditions, surface temperature, soil moisture, and vegetation dynamics, and to better understand the interactions between different ecological and hydrological components.

Dr. Liu was a recipient of the Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA) at the University of New South Wales.

Lucy Marshall received the B.E., M.E., and Ph.D. degrees from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2001, 2002 and 2006, respectively.

From 2006 to 2013, she was an Associate Professor of Watershed Analysis with Montana State University, Bozeman, MT, USA, where she worked on the interface of engineering and environmental science in quantifying uncertainty in hydrologic and environmental systems. She returned to UNSW as an Australian Research Council (ARC) Future Fellow in 2013, where she is an Associate Dean of Engineering (Equity and Diversity). She is also an Associate Professor with the School of Civil and Environmental Engineering, UNSW. Her research interests include conceptualizations of hydrologic processes and development of model diagnostics and uncertainty analysis methods.

Ashish Sharma received the bachelor's degree from the University of Roorkee, Roorkee, India, in 1989, the master's degree from IIT Delhi, New Delhi, India, in 1991, and the Ph.D. degree from Utah State University, Logan, UT, USA, in 1996.

He is a Professor with the School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW, Australia. His research interests include finding ways of meaningfully assessing impact of climate change on hydrology and water resources, assessing (and reducing) modeling uncertainty, estimating design floods in a more meaningful way, issuing seasonal forecasts for water resources management, developing better models that simulate both hydrology and ecology in an increasingly warming world, and a bunch of other things mostly aligned around his strengths in using statistical analysis, and methods for a range of hydrologic problems.

Dr. Sharma was a recipient of the Australian Research Council (ARC) Future Fellow.