

# Rethinking Satellite Data Merging: From Averaging to SNR Optimization

Seokhyeon Kim<sup>ID</sup>, Ashish Sharma<sup>ID</sup>, Yi Y. Liu<sup>ID</sup>, and Sean I. Young<sup>ID</sup>

**Abstract**—Merging of multiple satellite datasets is a simple yet effective way to reduce prediction error. However, most merging methods for satellite data today are based on weighted averaging first proposed in 1969 for economic forecasting, which does not provide optimal outcomes when applied to satellite data. If our aim is to produce a merged data product that minimizes the prediction errors against a prediction target, there is no reason to insist that the merged product be an average of the parent datasets. A more disciplined approach based on mathematical optimization would be to minimize prediction errors. However, formulating merging as an optimization problem is insufficient by itself as the statistics needed for optimization, e.g. signal-to-noise ratio (SNR) of parent products, are often unavailable in practice and must be estimated jointly. In this paper, we address both of these problems for data merging. We first formulate optimal merging of satellite data as a SNR optimization (SNR-opt), and propose an estimation method to jointly estimate the required SNRs. This SNR-based approach has a natural interpretation as a multi-input single-output Wiener filter. Through extensive experimental validation on three global-scale satellite-derived soil moisture and land surface temperature products, we demonstrate that our SNR optimization significantly improves merging results over weighted averaging schemes.

**Index Terms**—data merging, weighted averaging, SNR, signal-to-noise-ratio, optimization, satellite remote sensing.

## I. INTRODUCTION

SATELLITE-DERIVED data are often merged (or “fused”) to improve predictions of geophysical variables of interest by averaging out errors. In many situations, different datasets can exhibit different types of errors under varying physical and climatological retrieval conditions [1, 2]. A weighted average of multiple datasets can be a simple and expedient way to form

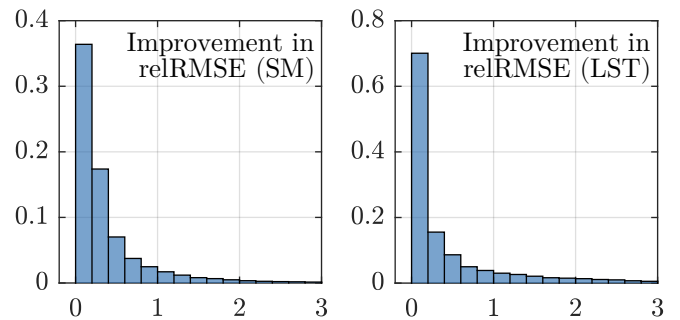


Fig. 1. Histograms of improvement in relRMSE due to our proposed method (SNR-opt) for global-scale soil moisture (left) and land surface temperature (right) predictions. Both use ERA5-Land as their prediction target. SNR-opt produces consistently lower errors than the weighted average merging scheme of Bates and Granger [3].

improved predictions. The key idea behind weighted averaging is to take independent information from data sources in hope of deriving an improved prediction via cancellation of random errors, the effectiveness of which depends on the independence of the separate data sources considered [3]. Following the work of Bates and Granger [3] who proposed optimal combination of forecasts using a mean square error (MSE) criterion, weighted averaging has been used in various fields of research including economics [4–6], ecology [7, 8], hydrometeorology [9–11] and others [12, 13]. While data merging using deep learning-based techniques have become more popular of late [14, 15], weighted averaging continues to be extremely useful for its simplicity and applicability when data is limited, and interpretability of results important in decision and policy making [6, 16].

Counterintuitively, however, weighted averaging sometimes increases prediction error [17, 18]. The averaging scheme of [3] is meant specifically for economic forecasts where statistics of the future prediction target are assumed inaccessible—merging of economic forecasts must rely solely on the errors of already past predictions. For satellite data merging, however, statistics of prediction targets are usually available [2, 19]—we have not only the variance of prediction errors but also that of the target itself to bound prediction errors. Weighted averaging imposes instead a unit-sum constraint on the merging weights, but this constraint is neither necessary nor beneficial in the general case (in fact, merge weights should sum to a value slightly less than one for optimal outcomes—as is shown later). Several authors further constrain the individual merging weights to lie between

Manuscript received February 11, 2021; revised February 11, 2021, and February 11, 2020. This work has been undertaken as part of a Discovery Project (DP200101326, *Assessing Water Supply Security in a Nonstationary Environment*) funded by the Australian Research Council. (Corresponding authors: Seokhyeon Kim; Sean I. Young.)

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

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

S. I. Young is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: sean0@stanford.edu).

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.2021.

zero and one [3, 20] but this too comes from a misunderstanding of merging weights as some quantitative metric—satellite data exhibit considerable inter-dataset error covariances, which lead frequently to valid negative merging weights [20, 21].

If one's objective is to produce a merged data product which minimizes the prediction error against a prediction target, there is no reason to insist that the merged product be an average of parent datasets. A more disciplined approach to merging parent datasets is to directly minimize the mean square error (MSE) of prediction for example (weighted averages, by contrast, do not minimize the MSE of prediction). However, having formulated merging as an optimization problem is not sufficient on its own since the statistics needed to solve the problem are unavailable and must be found jointly. In this work, we address both of the above problems by first formulating optimal data merging as a signal-to-noise (SNR) optimization problem, then proposing a technique to jointly find the required SNR statistics without the ground truth. Fig. 1 shows typical prediction improvements we obtain over the weighted average scheme of [3] when merging soil moisture and land surface temperature datasets. Intuitively speaking, these significant improvements come from imposing the condition that prediction error power should not be greater than the power of the quantity being predicted.

Our contributions are as follows. First, we propose a method (SNR-opt) to ascertain minimum MSE merging weights using SNR of the parent products and derive a relationship between these optimum weights and those based on averaging (Section III). Second, we propose a method to jointly estimate the SNR for the above-mentioned optimization for an arbitrary number of parent products. Our approach contrasts with schemes such as Triple Collocation (TC), which is usually applicable only to three or four parent datasets (Section III). Finally, we validate our method on satellite-derived soil moisture and land surface data (Section IV). We believe that our work will be beneficial especially to researchers who rely on optimally merged data for downstream tasks. Code for SNR-opt and figures in this paper are available at <http://www.github.com/steelp/snr-opt>.

## II. RELATED WORK

We briefly review classical data merging schemes as well as their application to satellite datasets.

### A. Merging Techniques

Taking simple averages of multiple datasets or forecasts has been the preferred way of merging data due to its simplicity and improvements seen in practice [4, 22, 23]. Bates and Granger [3] propose to linearly combine forecasts to minimize the MSE of the combined forecast, where unbiased and stationary parent forecasts are combined using weights constrained between zero and one with unit sum. However, this constrained approach was later found to be suboptimal compared to related methods such as regression [20]. Granger and Ramanathan [17] also show the sub-optimality of constraining the weights and propose a more flexible merging technique by relaxing weight constraints and adding a constant term, essentially turning combination into an ordinary least squares regression. Interestingly, all these works are predated by the seminal work of Wiener [24], who derived optimal prediction 20 years before [3]. In Section III, we show

our SNR optimization allows an interpretation of merging as a multi-input single-output Wiener filter, itself closely related to general ordinary least squares regression.

To address the non-stationarity of parent forecasts, Diebold and Pauly [25] incorporate time-varying weights using moving subsets of parent datasets and report significant improvements in the merged products. Coulson and Robins [26] first identify sources of dynamics and take them into account in the forecast combination. Min and Zellner [27] propose a Bayesian method for combination by adopting posterior odds defined as the ratio of the posterior probabilities of a fixed parameter model and a time-varying parametric model. Then with the posterior odds in one year, one of the models can be chosen and optimal weights calculated. Deutsch *et al.* [28] test dynamic combinations with immediate and gradual changes in the weights and demonstrate better performance compared to the static model. Diebold and Pauly [29] propose to overcome uncertainties in the estimation of weights using Bayesian shrinkage techniques, incorporating a prior on errors into weight estimation. Palm and Zellner [23] propose a Bayesian approach which utilizes informative priors on the errors when little information on the performance of the individual forecasts is available. Whereas Bayesian estimation techniques can be beneficial, our work considers optimization of weights in the prior-free scenario of [3].

Terui and van Dijk [30] evaluate combination of linear and nonlinear timeseries model-derived forecasts for both constant and time-varying methods and show that the time-varying one performs better. Several authors show that linear averaging can be better than more complex methods when the optimal weights are difficult to estimate precisely [6, 22, 31–33]. Similar to the averaging schemes, linearity of our proposed method allows us simpler interpretation and computation of optimal weights. We refer the reader to [4] for a review in forecast merging, and also to [5, 6, 8] for a more recent chronology.

### B. Finding Missing Statistics

Another critical factor for weighted averaging to be useful is correct characterization of the errors used in weight estimation [32]. Weighted average schemes require second-order statistics of the ground truth and prediction error. Triple collocation (TC) [19, 34, 35] finds error covariances and data-truth correlations from the covariances of three datasets without the ground truth provided that we have: data-ground truth linearity, stationarity of the ground truth and error, error-truth orthogonality and zero error crosscorrelation (ECC) [35]. Yilmaz and Crow [21] show that TC tends to underestimate error variances whenever the zero ECC assumption is violated. Gruber *et al.* [36] propose a quadruple collocation (QC) method based on a data quadruple to estimate some of the ECCs by relaxing the assumption of full error independence.

Several works [37–39] tackle the difficulties associated with obtaining three or more datasets with uncorrelated errors. Su *et al.* [37] propose the single instrumental variable (IVs) method that only requires two independent datasets where a temporally lagged time series of one dataset is used as a third one. Dong *et al.* [38] propose the double instrumental variables (IVd) method that uses two lagged time series of two datasets for the results to be more reliable. Dong *et al.* [39] subsequently develop the

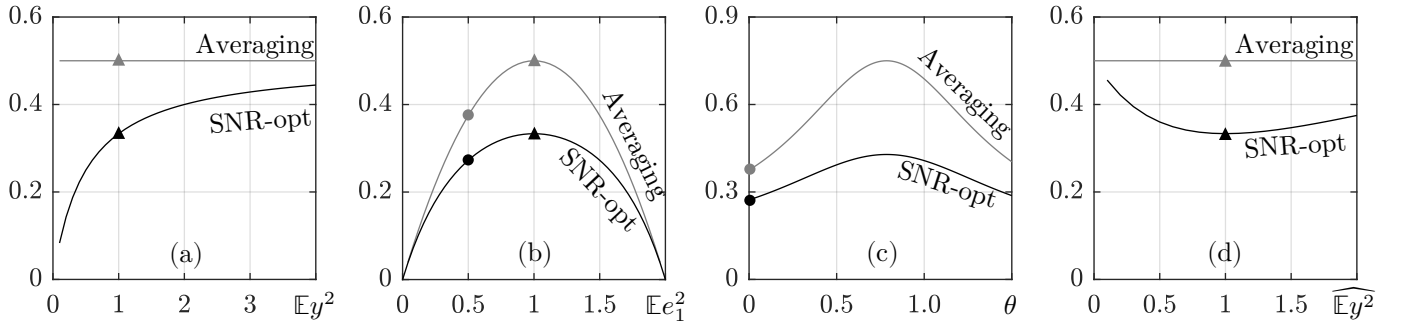


Fig. 2. Mean square error of weighted average and SNR-opt predictions. We vary the elements of the SNR matrix  $\mathbf{N} = \mathbb{E}(\mathbf{e}\mathbf{e}^T)/\mathbb{E}(y^2) \in \mathbb{S}_+^{2 \times 2}$  by (a) scaling the signal power  $\mathbb{E}(y^2)$  with fixed  $\mathbb{E}(\mathbf{e}\mathbf{e}^T) = \mathbf{I}$ , (b) scaling  $\mathbb{E}(\mathbf{e}\mathbf{e}^T) = \mathbf{I}$  element-wise with  $\mathbf{1}^T \mathbb{E}(\mathbf{e}\mathbf{e}^T) \mathbf{1} = 2$ , and (c) rotating  $\mathbb{E}(\mathbf{e}\mathbf{e}^T) = \text{diag}(0.5, 1.5)$  by angle  $\theta$ . In (d), we use  $\widehat{\mathbb{E}}y^2$  as our estimate of  $\mathbb{E}(y^2)$ . We fix  $\mathbb{E}(y^2) = 1$  in (b)–(d). Markers indicate locations with identical elements of  $\mathbf{N}$ .

extended double instrumental variable (EIVd) method that can estimate an unignorable ECC in a data triplet using their lagged representations. Despite such advances, efficient estimation of ECCs for an arbitrary number of datasets still remains an open problem. In Section III-C, we propose a technique to efficiently estimate error covariances for any number of products.

### C. Application to Satellite Datasets

Satellite datasets which have previously been considered for merging include soil moisture [40–45], sea surface temperature [18, 46], precipitation [9, 11, 47], and others [12, 13]. In most of these merging applications, the weighted average scheme [3] is used for its simplicity and compatibility with TC-based error estimation. Notably, Gruber *et al.* [41] propose a framework for merging satellite soil moisture data based on TC-derived error statistics. Kim *et al.* [45] use TC with EIVd to minimize errors in the merged soil moisture data. Xu *et al.* [47] merge multiple monthly precipitation datasets using error variances estimated from the collocation-based approach “three-cornered hat”, also compatible with the weighted average framework of [3].

Another reason for the popularity of weighted averages may be due to the supposed quantitative meaning of weights as the relative contribution factors of parent datasets. Khan *et al.* [18] present time-varying optimal weights between zero and one of five sea surface temperature datasets as their relative measures of importance. Kim *et al.* [48] present the spatial distributions of optimal weights on two satellite soil moisture datasets and also interpret them as their relative importance. However, it is difficult to interpret these weights quantitatively since they are simply regression coefficients [17]. Furthermore, these weights produce suboptimal merging results in many cases. Section III illustrates the suboptimality of the weighted averaging scheme using soil moisture datasets.

## III. MATHEMATICAL FRAMEWORK

### A. Weighted Averaging

To motivate optimization-based satellite data merging, let us first analyze the weighted average scheme [3, 49] together with its underlying assumptions. Our goal is to predict an unknown quantity (e.g. soil moisture across time at some location)  $y \in \mathbb{R}$  as a weighted average of  $N$  given parent datasets or products  $(x_1, \dots, x_N) = \mathbf{x} \in \mathbb{R}^N$ , related to the unknown quantity  $y$  as

$\mathbf{x} = y\mathbf{1} + \mathbf{e}$  with additive noise  $\mathbf{e}$ . We assume that  $\mathbf{e}$  is jointly Gaussian with zero mean while  $y$  is a weakly stationary process also with zero mean. Once a prediction of  $y$  is formed, we can add the correct offset (mean) obtained from climatological data [50, 51] back to our prediction.

The averaging weights for the prediction  $\mathbf{x} = y\mathbf{1} + \mathbf{e}$  can be expressed as the solution to the problem

$$\begin{aligned} & \text{minimize} \quad g(\mathbf{u}) = \mathbb{E}(\mathbf{e}^T \mathbf{u})^2 = \mathbb{E}(\mathbf{x}^T \mathbf{u} - y\mathbf{1}^T \mathbf{u})^2 \\ & \text{subject to} \quad h(\mathbf{u}) = \mathbf{1}^T \mathbf{u} = 1 \end{aligned} \quad (1)$$

[49]. The solution of the above optimization problem becomes

$$\mathbf{u}^\dagger = (\mathbf{1}^T \mathbb{E}(\mathbf{e}\mathbf{e}^T)^{-1} \mathbf{1})^{-1} \mathbb{E}(\mathbf{e}\mathbf{e}^T)^{-1} \mathbf{1}, \quad (2)$$

noting that  $\mathbf{1}^T \mathbb{E}(\mathbf{e}\mathbf{e}^T)^{-1} \mathbf{1}$  is simply the sum of elements of the matrix  $\mathbb{E}(\mathbf{e}\mathbf{e}^T)^{-1}$ . Therefore, the first term of (2) guarantees the elements of  $\mathbf{u}^\dagger$  have unit sum by construction. Looking now at the last expression of  $g(\mathbf{u})$ , the constraint  $\mathbf{1}^T \mathbf{u} = 1$  may seem to correctly produce an unbiased prediction of  $y$  and thereby a minimum mean square error, but this is true only in the trivial case where  $\mathbf{e} = \mathbf{0}$ . To see this, consider an extreme case where  $\mathbb{E}(y^2) = 0$ , that is,  $y = 0$ . When  $\mathbf{e} \neq \mathbf{0}$ , the only weights which minimize the last expression of  $g(\mathbf{u})$  is  $\mathbf{u} = \mathbf{0}$ , but this would not be allowed under the unit-sum constraint. This is a hint that both signal and noise powers,  $\mathbb{E}(y^2)$  and  $\mathbb{E}(\mathbf{e}\mathbf{e}^T)$ , play a role in the determination of minimum-mean-square-error weights.

Several works [3, 20] additionally trim negative weights and normalize the remaining non-negative ones but such a practice cannot be justified even from the point of view of (2). Several authors also ad hoc assume errors to be independent from each other [31, 52] in which case the merge weights simply become inverses of error variances.

### B. SNR Optimization

Here, we propose an SNR-based optimization method which minimizes the mean square error of the prediction. Notice first that the vector  $\mathbf{u} \in \mathbb{R}^N$  which minimizes the mean square error of prediction  $\hat{y} = \mathbf{x}^T \mathbf{u}$  is obtained by solving the problem

$$\text{minimize} \quad f(\mathbf{u}) = \mathbb{E}(\mathbf{x}^T \mathbf{u} - y)^2 \quad (3)$$

[17]. The solution of the above problem is

$$\mathbf{u}^* = \mathbb{E}(\mathbf{x}\mathbf{x}^T)^{-1} \mathbb{E}(y\mathbf{x}), \quad (4)$$

in which we assume  $\mathbb{E}(\mathbf{x}\mathbf{x}^T)$  is always positive definite. Noting

that  $\mathbb{E}(\mathbf{x}\mathbf{x}^T) = \mathbb{E}(\mathbf{e}\mathbf{e}^T) + \mathbb{E}(y^2)\mathbf{1}\mathbf{1}^T$  and  $\mathbb{E}(y\mathbf{x}) = \mathbb{E}(y^2)\mathbf{1}$ , we can rewrite the above solution as

$$\mathbf{u}^* = (\mathbb{E}(\mathbf{e}\mathbf{e}^T)/\mathbb{E}(y^2) + \mathbf{1}\mathbf{1}^T)^{-1}\mathbf{1}, \quad (5)$$

showing that the optimum weights  $\mathbf{u}^*$  depend on the matrix of noise-to-signal ratios  $\mathbf{N} = \mathbb{E}(\mathbf{e}\mathbf{e}^T)/\mathbb{E}(y^2)$ . We can interpret (5) as the coefficients of a multi-input single-output Wiener filter that assumes flat signal and noise power spectra.

From (5), one can see that the elements of  $\mathbf{u}^*$  need not sum to 1 in the general case (consider the case  $\mathbf{N} = \mathbf{I}$  for example) and that the sum only approaches 1 in the limit as the matrix of noise-to-signal ratios  $\mathbf{N}$  goes to  $\mathbf{0}$ . Similarly, examples can be constructed to show that not all elements of  $\mathbf{u}^*$  are positive in general. Since the weighted averaging weights  $\mathbf{u}^\dagger$  are different from  $\mathbf{u}^*$  in general, they must additionally be sub-optimal with respect to the MSE objective of (3). Fig. 2 plots the optimality gap between the two MSEs as  $\mathbf{N} = \mathbb{E}(\mathbf{e}\mathbf{e}^T)/\mathbb{E}(y^2) \in \mathbb{S}_+^{2 \times 2}$  is subjected to scaling of  $\mathbb{E}(y^2)$  (a), or transformations in  $\mathbb{E}(\mathbf{e}\mathbf{e}^T)$  by element-wise scaling (b) or rotations (c). SNR optimization still produces a lower MSE than the weighted average scheme even when  $\mathbf{N}$  is grossly under- or over-estimated (d).

In the general case, datasets  $(x_1, \dots, x_N) = \mathbf{x} \in \mathbb{R}^N$  relate to a random quantity  $y$  as  $\mathbf{x} = y\mathbf{a} + \mathbf{b} + \mathbf{e}$  with multiplicative and additive factors  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$  and a zero-mean additive noise  $\mathbf{e} \in \mathbb{R}^N$  that we continue to assume to be jointly Gaussian. We may safely assume  $\mathbf{b} = \mathbf{0}$  since  $\mathbf{x}$  can be de-biased in advance using  $\mathbf{b} = \mathbb{E}[\mathbf{x}]$ . In this case, the optimal weights are given by

$$\mathbf{u}^* = (\mathbf{N} + \mathbf{a}\mathbf{a}^T)^{-1}\mathbf{a}, \quad (6)$$

in which  $\mathbf{N} = \mathbb{E}[\mathbf{e}\mathbf{e}^T]/\mathbb{E}(y^2)$  as before. Therefore, normalizing the parent products  $\mathbf{x}$  to  $\mathbf{x}' = y\mathbf{1} + \mathbf{e}'$  in advance reduces the general solution (6) to the special solution (5) where the scale factors are 1. Formulating merging as problem (1) can be more useful if the statistic  $\mathbb{E}(y^2)$  is not available, but the availability of climatology data suggests this is not a concern in such tasks as soil moisture or land surface temperature estimation.

Despite their cosmetic differences, there is a simple scaling relationship between the two weights  $\mathbf{u}^\dagger$  and  $\mathbf{u}^*$ . Applying the Sherman–Morrison formula to  $\mathbf{u}^* = (\mathbf{N} + \mathbf{1}\mathbf{1}^T)^{-1}\mathbf{1}$  yields

$$\mathbf{u}^* = \mathbf{N}^{-1}\mathbf{1} - \frac{\mathbf{N}^{-1}\mathbf{1}\mathbf{1}^T\mathbf{N}^{-1}\mathbf{1}}{1 + \mathbf{1}^T\mathbf{N}^{-1}\mathbf{1}} = \frac{\mathbf{N}^{-1}\mathbf{1}}{1 + \mathbf{1}^T\mathbf{N}^{-1}\mathbf{1}} \quad (7)$$

and substituting  $\mathbf{N} = \mathbb{E}(\mathbf{e}\mathbf{e}^T)/\mathbb{E}(y^2)$  into the right-hand side of (7) reveals that the optimal weights

$$\mathbf{u}^* = (\mathbb{E}(y^2)^{-1} + \mathbf{1}^T\mathbb{E}(\mathbf{e}\mathbf{e}^T)^{-1}\mathbf{1})^{-1}\mathbb{E}(\mathbf{e}\mathbf{e}^T)^{-1}\mathbf{1}, \quad (8)$$

casting expression (5) into a form more similar to (2). We thus have a simple linear relationship  $\mathbf{u}^* = s\mathbf{u}^\dagger$  where

$$0 < s = \frac{\mathbf{1}^T\mathbb{E}(\mathbf{e}\mathbf{e}^T)^{-1}\mathbf{1}}{\mathbb{E}(y^2)^{-1} + \mathbf{1}^T\mathbb{E}(\mathbf{e}\mathbf{e}^T)^{-1}\mathbf{1}} < 1 \quad (9)$$

is a shrinkage factor that depends on the error covariances and the signal power. One corollary of the relationship  $\mathbf{u}^* = s\mathbf{u}^\dagger$  is that problem (3) can be written in the form (1) where the linear constraint becomes  $h(\mathbf{u}) = \mathbf{1}^T\mathbf{u} = s$ ; see Fig. 3 for geometric interpretations of the weights produced by weighted averaging and SNR optimization approaches.

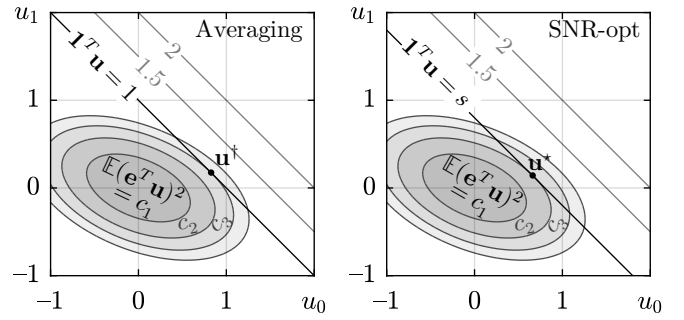


Fig. 3. Geometric interpretation of weights in  $\mathbb{R}^2$  for weighted averaging (left) and SNR optimization (right). Weighted averaging weights  $\mathbf{u}^\dagger$  are given by the point where the contour of  $\mathbb{E}(\mathbf{e}^T \mathbf{u})^2 = c$  tangentially touches the constraint line  $\mathbf{1}^T \mathbf{u} = 1$ . SNR optimal weights  $\mathbf{u}^*$  are given by the point where the contour of  $\mathbb{E}(\mathbf{e}^T \mathbf{u})^2 = c$  tangentially touches  $\mathbf{1}^T \mathbf{u} = s$  with  $s$  given by (9).

Since  $\mathbf{u}^\dagger$  is a scaled version of the optimal  $\mathbf{u}^*$ , both weights maximize the correlation between the merge prediction and the prediction target. However, if one's object is to find any merge prediction which is only maximally correlated with the target, a more appropriate characterization of their associated weights is through simpler, Rayleigh quotient maximization, which does not involve  $\mathbb{E}(\mathbf{e}\mathbf{e}^T)$ . In the case where  $\mathbf{x} = y\mathbf{1} + \mathbf{e}$ , the vector of weights  $\mathbf{u}^\dagger$  is maximally correlating if it solves the problem

$$\text{maximize } r(\mathbf{u}) = \frac{\mathbb{E}(y(\mathbf{x}^T \mathbf{u}))}{\sqrt{\mathbb{E}(y^2)}\sqrt{\mathbb{E}((\mathbf{x}^T \mathbf{u})^2)}}, \quad (10)$$

or, equivalently, the problem

$$\text{maximize } r^2(\mathbf{u}) \propto \frac{\mathbf{u}^T \mathbb{E}(y\mathbf{x})\mathbb{E}(y\mathbf{x})^T \mathbf{u}}{\mathbf{u}^T \mathbb{E}(\mathbf{x}\mathbf{x}^T) \mathbf{u}} = \frac{\mathbf{u}^T \mathbf{1}\mathbf{1}^T \mathbf{u}}{\mathbf{u}^T \mathbb{E}(\mathbf{x}\mathbf{x}^T) \mathbf{u}}, \quad (11)$$

in which the last expression in the objective of (11) is known as a generalized Rayleigh quotient [53].

Writing the two matrices as  $\mathbf{A} = \mathbf{1}\mathbf{1}^T$  and  $\mathbf{B} = \mathbb{E}(\mathbf{x}\mathbf{x}^T)$  for brevity, the maximally correlating weight vector  $\mathbf{u}^\dagger$  is obtained as any leading eigenvector of the pair of matrices  $(\mathbf{A}, \mathbf{B})$ , that is, by solving the eigenvalue problem  $\mathbf{A}\mathbf{u} = \lambda\mathbf{B}\mathbf{u}$  (only one of the eigenvectors has a non-zero eigenvalue since the rank of  $\mathbf{A}$  is one). From the one-dimensional space of solutions  $\mathbf{u}^\dagger$ , only the vector  $\mathbf{u}^*$  jointly maximizes correlation and minimizes the error variance such that the choice of merge weights  $\mathbf{u}^\dagger$  (2) in preference to  $\mathbf{u}^*$  (5) is difficult to justify.

We can further relate merging to a multi-input single-output (MISO) Wiener filter [54]. Here, cross power spectral densities of errors and the power spectral density of the signal— $\mathbf{S}_{ee}(\omega)$  and  $S_{yy}(\omega)$ —are used in place of  $\mathbb{E}(\mathbf{e}\mathbf{e}^T)$  and  $\mathbb{E}(y^2)$  such that merging weights are adapted to the frequency contents of error and signal. Writing  $\mathbf{N}(\omega) = \mathbf{S}_{ee}(\omega)/S_{yy}(\omega)$ , we have

$$\mathbf{u}^*(\omega) = (\mathbf{N}(\omega) + \mathbf{1}\mathbf{1}^T)^{-1}\mathbf{1}, \quad \omega \in [-\pi, +\pi], \quad (12)$$

in which  $\mathbf{u}^*(\omega)$  denotes weights at frequency  $\omega$ . In the discrete domain, one way to obtain the MISO Wiener-filtered output is by first taking an  $M$ -point Discrete Fourier Transform (DFT) of the parent products, merging them in the Fourier domain by linearly combining their  $m$ th Fourier coefficients with weights  $\mathbf{u}^*[m] = \mathbf{u}^*(2\pi m/M)$  for  $0 \leq m < M$ , then finally taking the inverse DFT of the  $M$ -vector of combined coefficients. We can

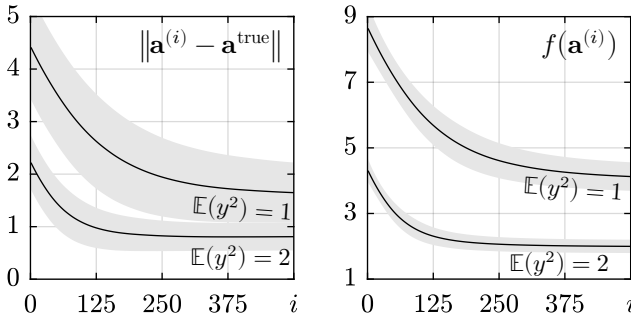


Fig. 4. Convergence of the optimization objective (left) and the relative mean absolute error of scale factors  $\mathbf{a}$  (right) across iterations. (The optimization fails to converge when  $\mathbf{a}$  is initialized randomly.) Bold lines and shaded areas show the mean and the standard deviation of the optimization objective (left), and the relative MAE (right) of the scale factors.

see that the merging weight vector (5) is a special case of (12) in which  $\mathbf{N}(\omega)$  is constant in  $\omega$  (flat error cross-power spectra and flat signal power spectra). While we assume flat noise and signal power spectra in this work, using fully estimated power spectra  $\mathbf{S}_{ee}(\omega)$  and  $S_{yy}(\omega)$  could further improve merging.

### C. SNR Estimation

To compute the general solution (6), we need to estimate the elements of symmetric matrix  $\mathbf{N}$  along with scale factors  $\mathbf{a}$ . In general, we can form an empirical covariance matrix of  $\mathbf{x}$  from measurements of the parent products (with their additive biases removed). Since we assume  $\mathbb{E}(y^2)$  is known from climatology data, we equate the empirical covariances to the unknowns as

$$\mathbf{C} = \mathbb{E}(\mathbf{x}\mathbf{x}^T) / \mathbb{E}(y^2) \approx \mathbf{N} + \mathbf{a}\mathbf{a}^T, \quad (13)$$

our objective being to simultaneously estimate the elements of matrix  $\mathbf{N}$  together with those of  $\mathbf{a}$ . In the case where prediction errors  $\mathbf{e}$  are uncorrelated as sometimes assumed [3], we could form and solve a non-linear system of  $\frac{(N+1)N}{2}$  equations in  $2N$  unknowns  $a_1, \dots, a_N, \mathbb{E}(e_1^2), \dots, \mathbb{E}(e_N^2)$ . Unfortunately, such a method is unable to handle cases where there are significant error covariances not captured by the unknowns. On the other hand, some schemes assume the scales  $\mathbf{a} = \mathbf{1}$  and find the full error covariance matrix [47]. However, unit scale factors are not observed in practice, also rendering such methods less useful.

Here, we do not assume particular scale factors or structure for the matrix  $\mathbf{N}$ . However, when  $N$  is large and parent datasets are gathered independently, it is reasonable to assume that the off-diagonal elements of  $\mathbf{N}$  are small. We propose to estimate the SNR matrix  $\mathbf{N}$  and the scale factors  $\mathbf{a}$  using the following  $\ell^1$ -norm minimization problem (which we coin SNR-est):

$$\begin{aligned} & \text{minimize} \quad f(\mathbf{a}) = \mathbf{1}^T ((\mathbf{1}\mathbf{1}^T - \mathbf{I}) \circ \text{abs}(\mathbf{C} - \mathbf{a}\mathbf{a}^T)) \mathbf{1} \\ & \text{subject to} \quad h(\mathbf{a}) = \text{diag}(\mathbf{a}\mathbf{a}^T) - \text{diag}(\mathbf{C}) \leq \mathbf{0} \end{aligned} \quad (14)$$

in hopes that the off-diagonals of  $\mathbf{N}^* = \mathbf{C} - \mathbf{a}^* \mathbf{a}^{*T}$  produced by the solution  $\mathbf{a}^*$  are small. Whereas (8) is not convex, related non-convex least-squares problems for phase retrieval [55, 56] are commonly solved using gradient descent with success, and this suggests that we too may be able to solve problem (14) with gradient descent methods provided that a good initialization for  $\mathbf{a}$  is used to avoid getting trapped in local minima.

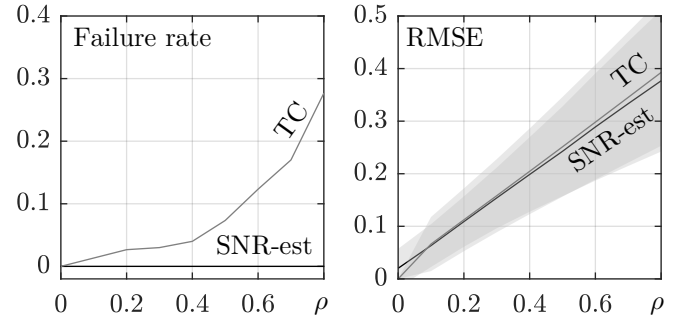


Fig. 5. Failure rate (left) and the RMSE of estimates of SNR matrix  $\mathbf{N}$  (right) from TC and our SNR estimation (SNR-est) schemes. We generate covariance matrices  $\mathbf{C} = \mathbf{N} + \mathbf{a}\mathbf{a}^T$  randomly from  $\mathbf{N} \in \mathbb{S}_+^{3 \times 3}$ ,  $\mathbf{a} \in \mathbb{R}^3$ , and a global error correlation coefficient  $\rho \in [0, 1]$ . RMSE averaged across 1000 estimations for each value of  $\rho$ . We use  $\beta = 0.1$  and 1000 iterations of gradient descent.

If the errors are uncorrelated and identically distributed, we have  $\mathbf{N} \approx \beta \mathbf{I}$  for some value of  $\beta$  and  $\mathbf{C} - \beta \mathbf{I} \approx \mathbf{a}\mathbf{a}^T$ . A good initializer for  $\mathbf{a}$  is thus a leading eigenvector of  $\mathbf{C} - \beta \mathbf{I}$ , scaled by the square root of the corresponding eigenvalue (we treat  $\beta$  as a tuneable parameter). To update the estimate of  $\mathbf{a}$ , one can use projected subgradient descent steps

$$\mathbf{a} \leftarrow \mathbf{a} - \eta \mathbf{1}^T ((\mathbf{1}\mathbf{1}^T - \mathbf{I}) \circ \text{sgn}(\mathbf{a}\mathbf{a}^T - \mathbf{C})) \mathbf{a} \quad (15d)$$

$$\mathbf{a} \leftarrow \mathbf{a} - \text{sgn}(\mathbf{a}) \circ \sqrt{\max(\text{diag}(\mathbf{a}\mathbf{a}^T - \mathbf{C}), \mathbf{0})} \quad (15p)$$

to first descend in the direction of the negative gradient using a stepsize  $\eta$  (15d) and also project the result so that the diagonals of  $\mathbf{a}\mathbf{a}^T$  do not exceed those of  $\mathbf{C}$  (15p). In Fig. 5, we graph the convergence properties of our gradient descent scheme. We see that descent converges with spectral initialization ( $\beta = 0.1$  and  $N = 6$  used, curves averaged across 1000 runs of varying error covariance matrices  $\mathbf{N}$  with a total power of one).

We also evaluate the performance of SNR-est in recovering  $\mathbf{N}$  and  $\mathbf{a}$  from random  $\mathbf{C}$ . We choose  $\mathbb{E}(y^2) = 1$  and  $N = 3$  to directly compare SNR-est against Triple Collocation. We first generate  $n_{ii} \in [0, 1]$ ,  $1 \leq i \leq 3$  with uniform distribution then assign  $n_{ij} = \rho \sqrt{n_{ii} n_{jj}}$  for chosen value of  $\rho$ . We generate the true  $a_i \in [0, 1]$ ,  $1 \leq i \leq 3$  similarly. We carry out 1000 runs of the experiment for each value of  $\rho$  to obtain average RMSE and a confidence interval. We use parameters  $\eta = 0.1$ ,  $\beta = 0.1$  and 1000 iterations of descent. Fig. 5 (left) shows that SNR-est and TC produce comparable errors in  $\mathbf{N}$ . RMSE of  $\mathbf{a}$  (not plotted) have a similar trend across  $\rho$  since  $\mathbf{N}$  and  $\mathbf{a}$  are linearly related through (13). In Fig. 5 (right), we show the failure rates of both methods across  $\rho$ , where an estimation is deemed to have failed if the estimates of  $n_{ii}$  come out negative or those of the factors  $a_i$  complex. Observe that SNR-est never produces an infeasible estimate due to our explicitly imposing a feasibility constraint in (14). On the other hand, TC begins to fail with an increase in the correlation value  $\rho$  due to its zero-ECC assumption.

## IV. EXPERIMENTAL EVALUATION

This work analyzes the accuracy of predictions derived from datasets combined using the weighted average and the proposed SNR optimization (SNR-opt) schemes. For evaluation, we run experiments where three satellite soil moisture (SM) products

TABLE I  
SUMMARY OF PARENT, REFERENCE, ANCILLARY AND VALIDATION DATASETS USED IN THIS WORK

	Product name	Resolutions (temporal/spatial)		Details
Parent datasets	<b>ASCAT</b> (Advanced SCATterometers) Level 2 SM index product H103 [61]	Daily/overpass (ascend/descend) at 9:30 PM/AM (local time)	25km×25km	Active C-band (5.3 GHz) radar backscatter
	<b>SMAP</b> (Soil Moisture Active Passive) Level 3 Radiometer Global Daily SM/LST V3 [62]	Daily/overpass (ascend/descend) at 6 PM/AM (local time)	36km×36km EASEv2-Grid	Passive L-band (1.41 GHz)
	<b>LPRM</b> (Land Parameter Retrieval Model) Level 3 Daily SM/LST V1 from AMSR2 on GCOM-W [63]	Daily/overpass (ascend/descend) at 1:30 PM/AM (local time)	0.25°×0.25°	Passive X-band (10.65 GHz)
Ref	<b>ERA5-Land</b> (European Centre for Medium-Range Weather Forecasts Re-Analysis 5 Land) [51]	Hourly	0.1°×0.1°	Volumetric soil water content in layer 1 (0–0.07m), Skin temperature
Ancillary data	<b>CZ</b> Updated Köppen–Geiger climate classification [64]	–	0.25°×0.25°	5 primary classes: tropical, arid, temperate, cold, and polar regions
	<b>MCD12C1</b> (Moderate Resolution Spectroradiometer Land Cover) V1 [65]	Yearly (2015)	0.05°×0.05°	6 primary classes: forest, shrublands, woodlands, grasslands, croplands, and unvegetated regions
Validation data	<b>ESA CCI SM</b> (European Space Agency Climate Change Initiative Soil Moisture) V05.2 [41, 66, 67]	Daily	0.25°×0.25°	Active–passive combined surface SM product
	<b>ISMN</b> (International Soil Moisture Network) [68, 69] and [70–82]	Hourly	425 stations across 18 networks:	SNOTEL <sup>128</sup> , SCAN <sup>111</sup> , PBO-H2O <sup>79</sup> , USCRN <sup>51</sup> , RSMN <sup>18</sup> , OZNET <sup>8</sup> , SMOSMANIA <sup>7</sup> , RISMA <sup>5</sup> , SOILSCAPE <sup>2</sup> , GROW <sup>3</sup> , REMEDHUS <sup>3</sup> , AMMA-CATCH <sup>2</sup> , HOBE <sup>2</sup> , IPE <sup>2</sup> , BIEBRZA-S-I <sup>1</sup> , COSMOS <sup>1</sup> , DAHRA <sup>1</sup> , and TERENO <sup>1</sup>

(ASCAT, SMAP, and LPRM) are merged across the five-year period from April 1, 2015 to March 31, 2020, with and without access to some reference (that is, assumed truth) reanalysis SM product (ERA5-Land). The details of the parent datasets and the references, including their unabbreviated names, are provided in Table I. Since the two merge predictions have the same bias and Pearson correlation against their prediction target, only the root mean squared error (RMSE) of predictions is compared in this work. We now describe the details of the data preprocessing and merging processes used in this work.

#### A. Data Preprocessing

The following preprocessing is applied to SM products. For ASCAT, pixels are used if they have less than 10% probability of snow, frozen ground and 50% of estimated retrieval error as per the recommendation in [57]. For SMAP, pixels are used if their open water fraction and frozen fraction are less than 10% and vegetation water content is less than 5kg/m<sup>2</sup>, as suggested in [58]. For LPRM data, pixels are used only if their vegetation optical depth in the C-band is less than 0.8 [59]. All ascending and descending data are averaged on a daily basis to maximize their spatial coverage. Since the ERA5-Land reference product is sampled hourly, we select for each pixel, SM values closest to daily average scan times of the three SM datasets. Products that have variable spatial resolutions are first resampled onto the global cylindrical 36 km Equal-Area Scalable Earth, version 2 (EASEv2) grid [60] using nearest neighbor resampling.

The combination results are additionally evaluated against ground-based SM obtained via the International Soil Moisture Network (ISMN) [68, 69]. To minimize systematic differences between the satellite- and the ground-based data, strict filtering guided by ancillary data is applied to the ground-based data as per [83]. This process involves applying quality flags; limiting depths (< 10cm); picking values temporally closest to the daily average scan time. For selecting the most representative station

in a grid cell, we use the average of the correlation of the *in situ* data against the three SM datasets and the reference. Following [83], we discard stations having significant ( $p = 0.05$ ) negative correlations with three or more of the four datasets since these stations are so not representative of their associated cells. After the filtering process, only stations that have at least fifty paired observations with each SM product are finally selected, giving us a total of 425 stations across 18 networks—refer to the last row of Table I.

Whereas our parent data represent soil moisture within a few centimeters of the top-soil layer, the *in situ* sensors of ISMN are installed at deeper depths, usually in the range of 5–10cm. The exponential filter of Wagner *et al.* [84] has been used in earlier studies [42, 85] to first convert the profiles of satellite-derived surface soil moisture to the depths of *in situ* sensors, mitigating the influence of depth discrepancies on evaluation results. We also adopt this filter-based approach in our work. In recursive form, the exponential filter of [84] can be expressed as

$$W(t_n) = W(t_{n-1}) + K_n(M(t_n) - W(t_{n-1})) \quad (16)$$

in which  $W(t_n)$  and  $M(t_n)$  denote the soil water index and the soil moisture at time  $t_n$ , respectively, and  $W(t_1) = M(t_1)$  for the initial conditions [86]. The gain  $K_n$  at  $t_n$  is given by

$$K_n = \frac{K_{n-1}}{K_{n-1} + \exp(-(t_n - t_{n-1})/T)} \in [0, 1], \quad (17)$$

in which  $K_1 = 1$ , and parameter  $T \in [1, 100]$ , the time scale of soil moisture variations. In this work, we find the optimal  $T$  by maximizing the Nash–Sutcliffe efficiency [42, 84, 86] between the soil water index from the reference soil moisture data and *in situ* data minus its mean at each station. Once the optimal  $T$  is computed, we apply gains (17) to both merged products and compare them with the corresponding *in situ* data.

For reference, we also compare the merged products with the soil moisture product ESA CCI SM [41, 66, 67]—see Table I



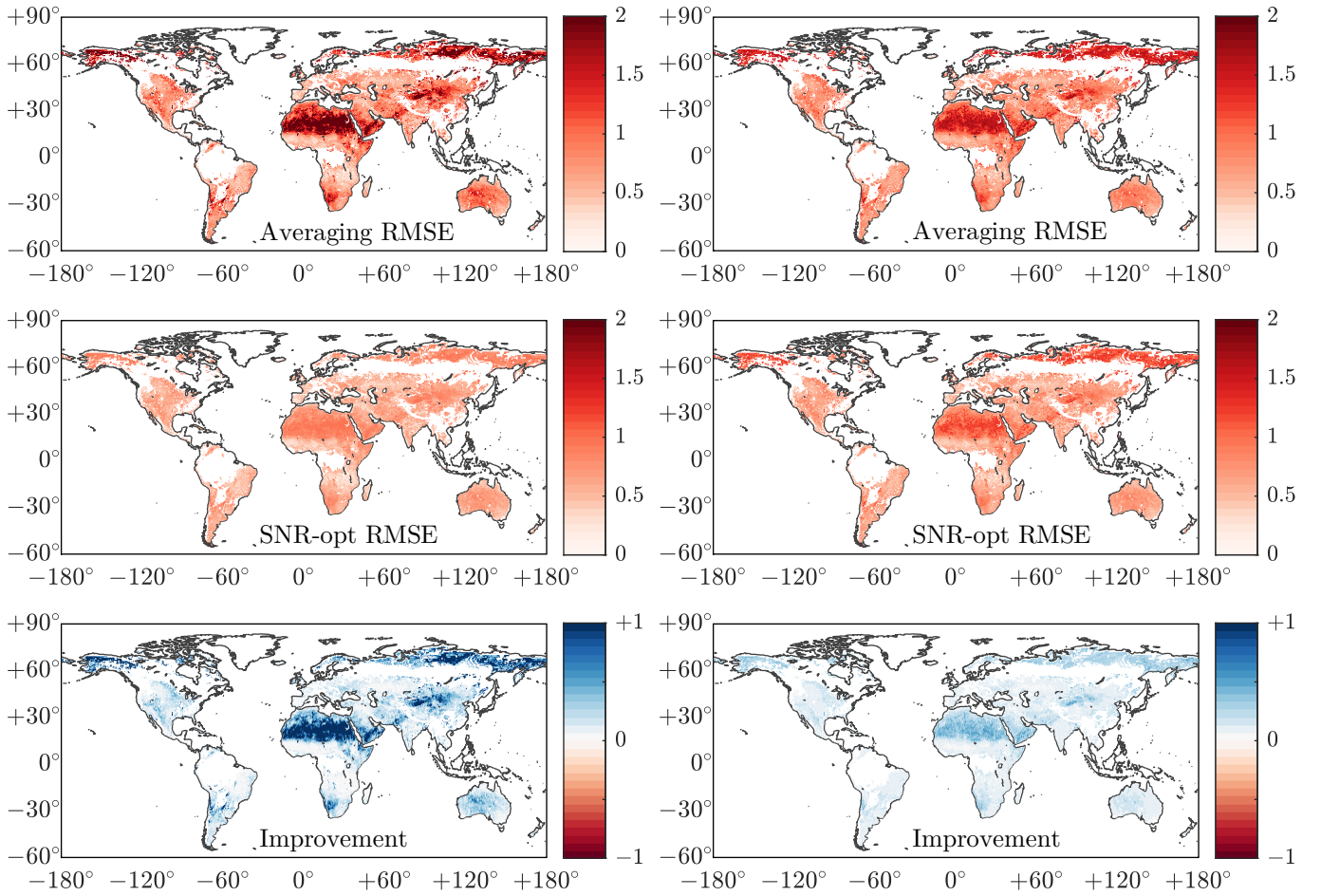


Fig. 6. Relative RMSE of soil moisture prediction against the reference (ERA5-Land) over the five-year study period April 2015–March 2020, based on true (left) and estimated (right) scale factors and SNRs. The top and the middle plots show the RMSE of the averaging and our SNR-opt predictions, respectively, and the bottom plots show the improvement in RMSE due to SNR-opt (higher is better). SNR-opt produces consistently lower RMSE than weighted averaging.

for details. (ESA) CCI SM contains daily surface soil moisture data at  $0.25^\circ$  spatial resolution spanning a 41-year period from November 1978 to December 2019. CCI SM is algorithmically fused through harmonization (data scaling) and combination of multiple soil moisture retrievals derived from four active and eight passive microwave sensors. The datasets are fused using resampling to  $0.25^\circ$  at 0:00 Coordinated Universal Time (UTC) daily; cumulative distribution function matching (scaling) to a long-term and consistent reference (satellite or reanalysis data) and weighted averaging based data merging using TC-derived error variances. The active and the passive sensor-derived data are merged independently and the two merges are then merged again to active–passive combined data with a TC-based merge scheme. This weighted averaging is implemented only when the correlations among active, passive, and modeled soil moisture time series data are statistically significant, suggesting both the active and the passive data are reliable at a given pixel [41]. If not, the merge algorithm returns either the passive or the active data, their unweighted average, or even discard the data at that location altogether. We similarly combine active (ASCAT) and passive (SMAP and LPRM) data, but from 2015–2020.

To analyze the performance of satellite SM products under different retrieval conditions, we also condition the evaluation results on climate zone (CZ) and land cover (LC). For this, we

used the five primary CZ classes from updated Köppen–Geiger climate classification [64]: tropical, arid, temperate, cold, and polar regions. For LC, the six classes from MODIS MCD12C1 [65] are: forest, shrublands, woodlands, grasslands, croplands and unvegetated regions.

### B. Merging Pipeline

We now merge the preprocessed datasets using the following pipeline. First, systematic biases in datasets are corrected prior to merging. We normalize all parent products to zero mean and unit standard deviation. For comparisons, the reference is also normalized in the same way (therefore, the signal variance of  $\mathbb{E}(y^2) = 1$  is always assumed). Second, we estimate for every pixel the SNR matrix  $\mathbf{N} = \mathbb{E}(\mathbf{e}\mathbf{e}^T)/\mathbb{E}(y^2)$  and the scale factors  $\mathbf{a}$ . We experiment with both the true and the estimated versions of  $\mathbf{a}$  and  $\mathbf{N}$ , with the estimated ones obtained using the method described in Section III-C. We find the initial solution of (14) good enough as  $\mathbf{a}$ . For initialization, we use  $\beta = 0.6$  based on a grid search for the optimum RMSE of both merge predictions using global spatially aggregated five-year period soil moisture data. This approximate value of  $\beta$  is intended to show that the proposed SNR-opt outperforms the weighted average even with rough estimates of  $\mathbf{a}$ . More careful estimates of  $\mathbf{a}$  can produce better outcomes—at worst, SNR-opt degenerates to a weighted

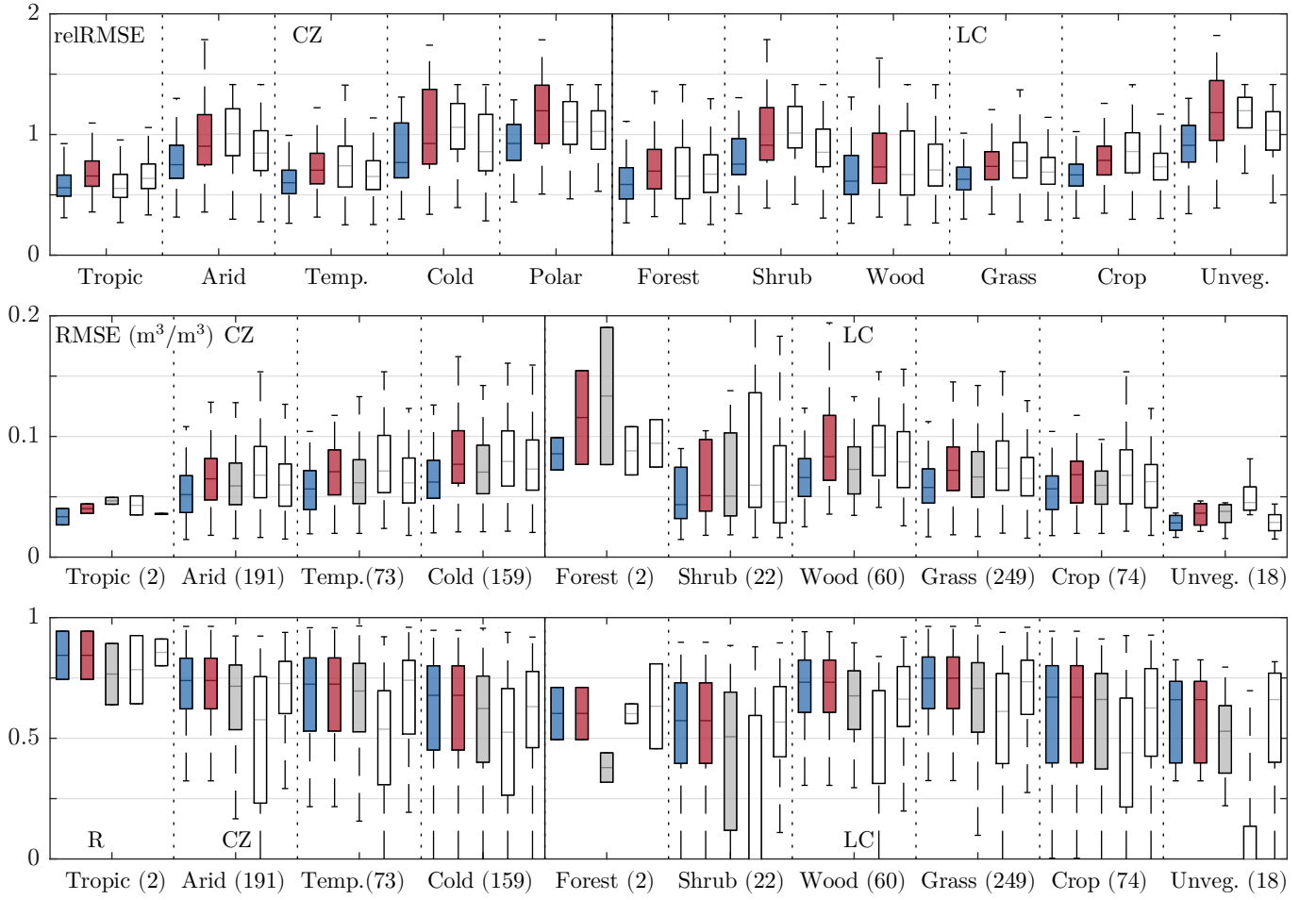


Fig. 7. RMSE of soil moisture predictions produced by SNR-opt (blue) and weighted averaging (red), stratified by CZ, and LC classes. Estimated scale factors and SNR are used. We include results for CCI SM (gray), and two parents ASCAT and SMAP (white) for reference. The first row shows the relative RMSE of SNR-opt and weighted averaging predictions on ERA5-Land. The second and the third rows show the RMSE and R of the predictions against the ISMN ground measurements, respectively (lower RMSE and higher R are better). Numbers in the parentheses indicate the number of ground stations. Outliers omitted for clarity.

averaging. Observe that weighted averaging (2) assumes  $\mathbf{a} = \mathbf{1}$  but SNR-opt (5) allows arbitrary  $\mathbf{a}$  for weight computation. To facilitate fairer comparisons between the two methods, we first estimate  $\mathbf{a}$  via (14) then rescale parent products  $\mathbf{x} = \mathbf{y}\mathbf{a} + \mathbf{e}$  to  $\mathbf{x}' = \mathbf{y}\mathbf{1} + \mathbf{e}'$  with  $\mathbf{N}' = \mathbf{C}' - \mathbf{1}\mathbf{1}^T$ ,  $\mathbf{C}' = \mathbb{E}(\mathbf{x}'\mathbf{x}'^T)/\mathbb{E}(y^2)$  so that  $\mathbf{N}' = \mathbb{E}(\mathbf{e}'\mathbf{e}'^T)/\mathbb{E}(y^2)$  can be used for the computation of both weighted averaging and SNR-opt weights. Such rescaling to  $\mathbf{x}'$  is not necessary if solely implementing SNR-opt—once  $\mathbf{a}$  has been estimated,  $\mathbf{u}^*$  can be obtained for  $\mathbf{x}$  directly from (6) using  $\mathbf{N} = \mathbf{C} - \mathbf{a}\mathbf{a}^T$ ,  $\mathbf{C} = \mathbb{E}(\mathbf{x}\mathbf{x}^T)/\mathbb{E}(y^2)$ . If needed, we also flip the signs of the final merged products to guarantee that they are positively correlated with the reference.

### C. Additional Validation

As further validation of the proposed method, we apply our merging scheme additionally to land surface temperature (LST) datasets. These LST datasets are ancillary data accompanied by SM ones. Since LST is not available from ASCAT, we merge LST datasets of SMAP and LPRM, using the skin temperature data of ERA5-Land as the reference.

The two satellite LST datasets SMAP and LPRM are derived from the microwave brightness temperature observed through their descending overpasses, respectively at 6 am, and 1:30 am

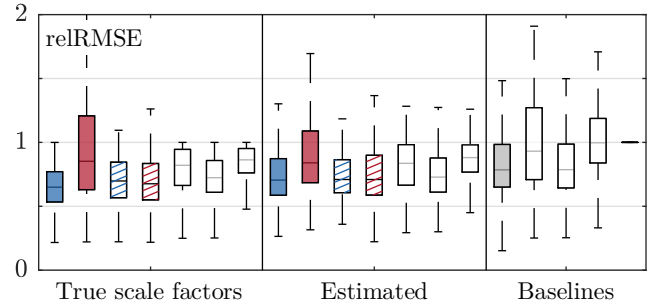


Fig. 8. RMSE of soil moisture predictions using weighted averaging (red) and SNR-opt (blue) on scaled (solid) and unscaled (striped) parents. ERA5-Land is the prediction target. We show predictions based on true (left) and estimated (middle) scale factors and SNR, and using individual baselines (right). We include ASCAT, SMAP and LPRM (white, in that order), ESA CCI (gray) and mean prediction (collapsed box).

local time. For the ERA5-Land skin temperature data, we use the value temporally closest to the daily average scan times of SMAP and LPRM. Unlike the SM case, however, we apply no filters on the LST data and evaluate the results directly against ERA5-Land, conditioned on the CZ, and the LC classes. Other aspects of the merging process remain the same as for the SM case, including the use of the initialization parameter  $\beta = 0.6$ .



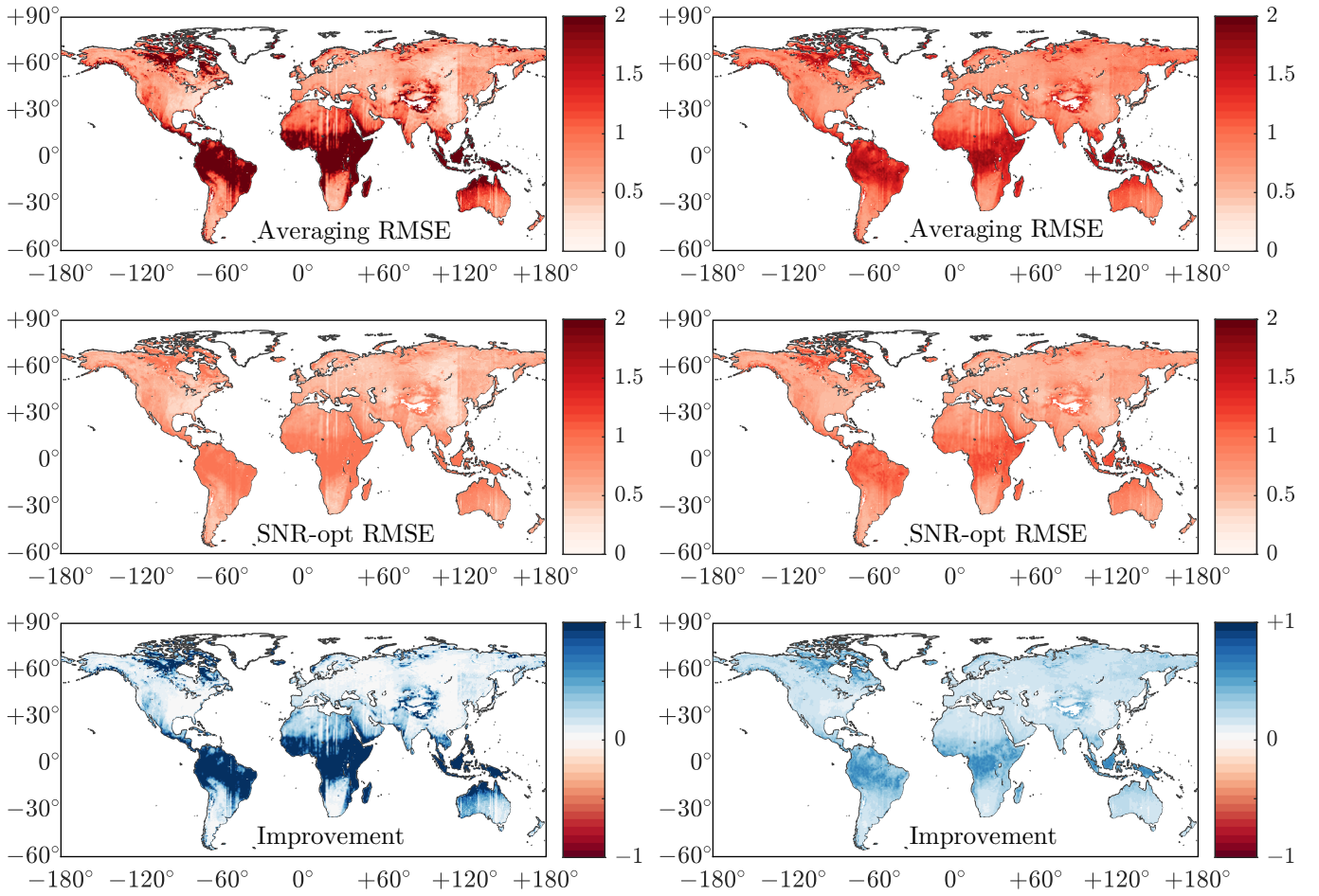


Fig. 9. Relative RMSE of land surface temperature prediction against the reference (ERA5-Land) over the five-year study period April 2015–March 2020, based on true (left) and estimated (right) scale factors and SNRs. The top and the middle plots show the RMSE of the weighted averaging and our SNR-opt predictions, respectively, and the bottom plots the improvement in RMSE due to SNR-opt (higher is better). SNR-opt always has lower RMSE than weighted averaging.

## V. RESULTS AND DISCUSSION

### A. Soil Moisture Prediction

Fig. 6 shows global spatial distributions of relative RMSE of predictions produced by weighted averaging and SNR-opt. We use ERA5-Land as the prediction target. We evaluate using the true (left) and the estimated (right) scale factors and SNR. Both predictions generally show similar spatial error patterns except for high-latitude regions in the northern hemisphere, the Sahel regions of Africa, and the western edges of South America and South Africa, where difficulties have frequently been reported in retrieving SM data from satellite observations [87, 88]. Over these problematic regions, SNR-opt generally produces better results than weighted averaging. This is seen in the two bottom subplots of Fig. 6, where spatial distributions of the difference in RMSE (RMSE of weighted average prediction minus RMSE of SNR-opt prediction) are notably positive (bluish). While the case with true scale factors and SNR matrix (left) shows larger RMSE differences, superiority of SNR-opt may still be seen in the case where their estimated quantities are used (right).

For further analyses, differences in the relative RMSE of the two predictions (based on estimated scale factors and SNR) and two (ASCAT and SMAP) of the parents are given as box plots in Fig. 7 (first row). We stratify all RMSEs on the CZ, and the

LC classes mentioned previously. LPRM is excluded from the plots for clarity but has larger errors than the other two. In each class, the difference between the SNR-opt (blue) and weighted averaging (red) predictions can be visually distinguished. Such differences are pronounced in regions with retrieval difficulties (arid, cold and polar climate zones; shrub and unvegetated land covers) and even more pronounced if true scale factors and SNR are used to form the merged products (not shown). While the SNR-opt predictions show consistently lower RMSE compared to parents (white) over all classes, this is not true for weighted averaging. This propensity is more pronounced for SMAP in the above-mentioned areas where SM retrieval is problematic.

As seen in the box plots of the second (RMSE), and the third (Pearson’s R) rows of Fig. 7, we further compare the errors of merged products, the active–passive combined product of CCI SM and (ASCAT and SMAP) against the ground measurements from the 432 ISMN stations as the truth. Again, CCI SM results have been included only a reference point since their approach and materials are different from ours. Here, our intention is to show that our proposed approach produces results comparable to other well-established products. For RMSE (second row of Fig. 7), SNR-opt produces results that are consistently better than the others. Note the weighted averaging results are as good as or worse than CCI SM (itself based on a weighted averaging

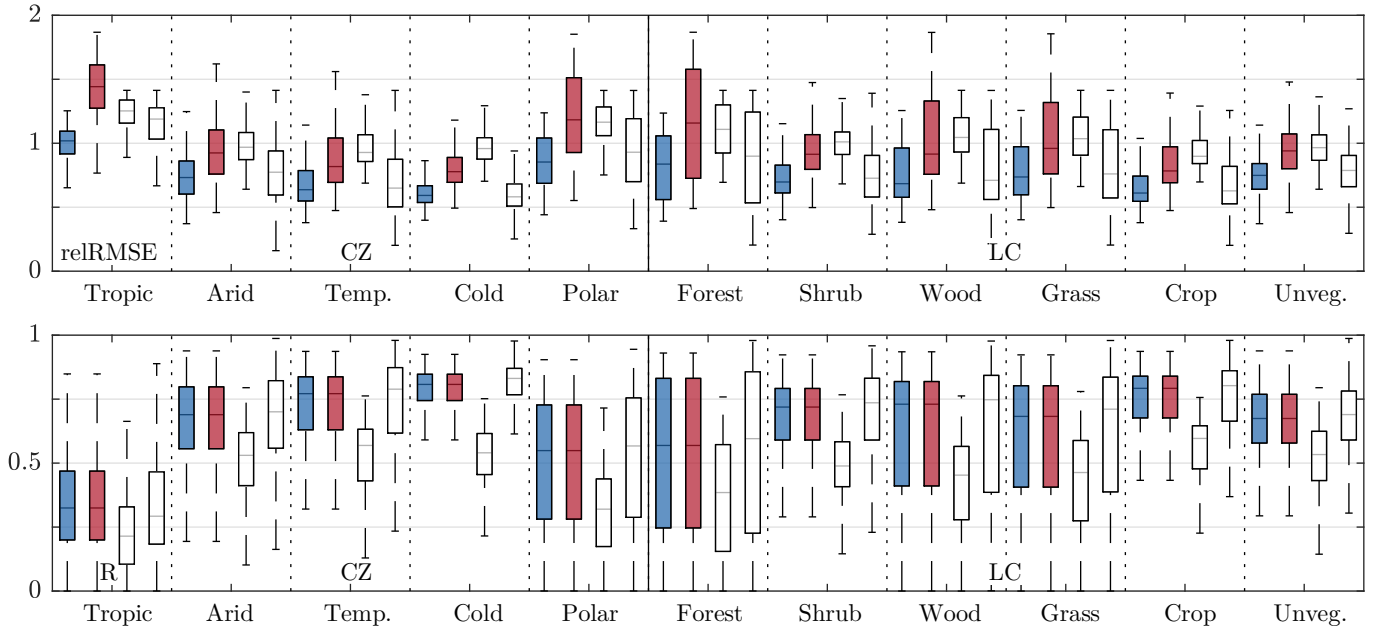


Fig. 10. Relative RMSE of merge predictions formed using SNR-opt (blue) and weighted averaging (red) with land surface temperature (ERA5-Land) as the prediction target (estimated scale factors and SNR used). We condition the RMSE on the CZ (left plots) and the LC (right plots) classes. We include the errors of SMAP and LPRM (first and second white of each class) for reference. For simpler presentation, we omit outliers from all box plots.

scheme and TC) and underperforms SMAP especially under the problematic retrieval conditions. The third row of Fig. 7 verifies that SNR-opt and weighted averaging yield the same R value—see Section III-B for explanation. We observe that our R values are slightly higher than that of CCI SM. SMAP already has high R values overall, but SNR-opt compensates for the degradation of SMAP under the problematic conditions such as cold climate and woodland.

Fig. 8 provides a comprehensive set of plots for the RMSE of soil moisture predictions and baselines with ERA5-Land as the prediction target. We show predictions using the true (left) and the estimated (middle) scale factors and SNR. Within each group, we plot RMSE results for SNR-opt- (blue) and weighted average- (red) merging of scaled (solid) and original unscaled (striped) parents, together with the results for applying only the scaling to the individual parents ASCAT, SMAP, and LPRM (white, in that order). In addition, we also present baselines for comparison (right) including CCI SM, the original parents, and mean prediction. The predictions from both groups generally outperform the baselines, except for the weighted average of the scaled parents (red) in both groups, even compared to the single parent predictions (white) and some of the original parents in the baseline group. The predictions from the group using the true scale factors and SNR surpass those from the other group using the estimated values, except for the weighted average of the scaled parents (red) showing the worst performance. When true scale factors and SNR are used, SNR-opt prediction RMSE never exceed unity (variances of prediction targets). Weighted averaging prediction RMSE sometimes exceeds unity and can actually be worse than simple mean prediction.

Note that weighted averaging benefits from using estimated scale factors and SNR. Predictions formed by merging unscaled parents using SNR-opt (striped-blue) and weighted averaging (striped-red) are similar in this case and comparable with those

of SNR-opt over both groups. However, this kind of gain is not theoretically justified and somewhat coincidental. Given scale factors  $\mathbf{a}$ , the weighted average merging of unscaled parents  $\mathbf{x}$  using weights  $\mathbf{u}^\dagger$  is identical to merging scaled parents  $\mathbf{x}'$  using  $\mathbf{u}^{\dagger'} = \mathbf{a} \cdot \mathbf{u}^\dagger$ . If by chance  $\mathbf{a} \approx \mathbf{s}\mathbf{1}$  (9),  $\mathbf{u}^{\dagger'}$  becomes closer to  $\mathbf{u}^*$  so using  $\mathbf{u}^{\dagger'}$  improves the RMSE performance of weighted averaging. Similarly,  $\mathbf{u}^*$  reduces to  $\mathbf{u}^{*'} = \mathbf{a} \cdot \mathbf{u}^*$  but here the performance loss in SNR-opt due to using  $\mathbf{u}^{*'}$  is not as dramatic as the improvement for weighted averaging. Since one does not generally have  $\mathbf{a} \approx \mathbf{s}\mathbf{1}$ , this type of ad hoc RMSE performance gain for weighted averaging is not always guaranteed.

### B. Land Surface Temperature Prediction

In Fig. 9, we show global distributions of RMSE of the two merged LST datasets through weighted averaging and SNR-opt using ERA5-Land as a reference. Similar to the SM case, the two sets of results using the true (left) and the estimated (right) scale factors and SNR are shown. We observe swath tracks due to the LST difference at the different overpass times of SMAP (6 am) and LPRM (1:30 am). These are not observed for SM.

While SNR-opt does generally perform better than weighted averaging, RMSE differences are even more drastic when true scale factors and SNRs are used. Fig. 9 (left) shows that largest differences can be observed in densely vegetated areas such as the Amazon, Central East Africa, Southeast Asia, and Northern Australia. In these areas, microwave signals from land surface tend to be attenuated severely by dense vegetation [89, 90]. In Fig. 10, we further investigate differences in RMSE (top) and R (bottom) of the two predictions (using estimated scale factors and SNR) conditioned on the CZ and LC classes. Again, RMSE differences between the two predictions are most notable in the tropical areas. While SNR-opt outperforms the others, weighted averaging is even worse than parent results. In terms of R, both merged products are always slightly better than the two parents

if true scale factors and SNR are used (not shown). These small improvements diminish, however, if estimated scale factors and SNR are used. These are the results shown in Fig. 10.

### C. Limitations and Future Work

Given the promising results presented here, we have already conceived several ways to extend this work. First, we intend to evaluate SNR-opt further by merging a wider variety of signals such as satellite-derived flood signal [91, 92], radar reflectivity [93] and vegetation optical depth [94–96]. Application of SNR-opt to such types of signals may further elucidate the strengths and weaknesses of our approach. Second, recent advances with triple collocation and related approaches [36, 39] may provide alternative more accurate ways to compute the parameters for estimating weights. Third, it may be beneficial to see if SNR-opt can be used for developing long-term and consistent global datasets to replace the individual products. Finally, SNR-opt is a simplified case of MISO Wiener filtering. Full MISO Wiener filtering (see Section III) may improve merging results further.

## VI. CONCLUSION

Weighted averaging has traditionally been used for merging satellite data. However, sub-optimality of weighted averaging has not fully been discussed by earlier studies. In this work, we first demonstrated the sub-optimality of weighted averages and proposed an optimization-based merging method which derives its merging weights from the signal-to-noise ratio of the parent products. For this reason, we referred to our merging method as SNR-opt. We validated SNR-opt using soil moisture and land surface temperature, and additionally compared SNR-opt with traditional weighted averaging. We compared the performance of predictions produced by SNR-opt and weighted averaging against reference products and ground measurements. SNR-opt consistently outperformed weighted averaging especially over regions where difficulties are faced in retrieving soil moisture and land surface temperature from satellite observations. In the case of soil moisture datasets, the merged product produced by SNR-opt showed performance comparable to well-established soil moisture product ESA CCI SM. Our findings suggest that SNR-opt may be widely applicable also to the merging of other satellite datasets not specifically discussed in this paper.

### AUTHOR CONTRIBUTIONS

S. Kim implemented the SNR-opt procedure and performed merging experiments. A. Sharma and Y. Y. Liu gave advice on experiments. S. I. Young conceived SNR-opt and supervised all aspects of the project. All authors contributed to designing the experiments and writing the paper.

## REFERENCES

- [1] R. Zhang, S. Kim, A. Sharma, and V. Lakshmi, "Identifying relative strengths of SMAP, SMOS-IC, and ASCAT to capture temporal variability," *Remote Sens Environ*, vol. 252, p. 112126, 2021.
- [2] R. M. Parinussa, A. G. C. A. Meesters, Y. Y. Liu, W. Dorigo, W. Wagner, and R. A. M. de Jeu, "Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model," *IEEE Geosci Remote Sens Lett*, vol. 8, no. 4, pp. 779–783, 2011.
- [3] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *J Oper Res Soc*, vol. 20, no. 4, pp. 451–468, 1969.
- [4] R. T. Clemen, "Combining forecasts: A review and annotated-bibliography," *Int J Forecast*, vol. 5, no. 4, pp. 559–583, 1989.
- [5] D. K. Barrow and N. Kourentzes, "Distributions of forecasting errors of forecast combinations: Implications for inventory management," *Int J Prod Econ*, vol. 177, pp. 24–33, 2016.
- [6] G. Elliott and A. Timmermann, *Economic Forecasting*. Princeton, New Jersey 08540, USA: Princeton University Press, 2016.
- [7] M. B. Araujo and M. New, "Ensemble forecasting of species distributions," *Trends Ecol Evol*, vol. 22, no. 1, pp. 42–47, 2007.
- [8] C. F. Dormann *et al.*, "Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference," *Ecol Monogr*, vol. 88, no. 4, pp. 485–504, 2018.
- [9] H. E. Beck *et al.*, "MSWEP: 3-hourly 0.25 degrees global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data," *Hydrol Earth Syst Sc*, vol. 21, no. 1, pp. 589–615, 2017.
- [10] R. D. Bhowmik, A. Sharma, and A. Sankarasubramanian, "Reducing model structural uncertainty in climate model projections-A rank-based model combination approach," *J Clim*, vol. 30, no. 24, pp. 10139–10154, 2017.
- [11] C. Massari *et al.*, "A daily 25 km short-latency rainfall product for data-scarce regions based on the integration of the Global Precipitation Measurement mission rainfall and multiple-satellite soil moisture products," *Hydrol Earth Syst Sc*, vol. 24, no. 5, pp. 2687–2710, 2020.
- [12] S. Chowdhury and A. Sharma, "Multisite seasonal forecast of arid river flows using a dynamic model combination approach," *Water Resour Res*, vol. 45, no. 10, p. W10428, 2009.
- [13] 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, 2018.
- [14] T. Meng, X. Y. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Inf Fusion*, vol. 57, pp. 115–129, 2020.
- [15] J. Gao, P. Li, Z. Chen, and J. Zhang, "A survey on deep learning for multimodal data fusion," *Neural Comput*, vol. 32, no. 5, pp. 829–864, 2020.
- [16] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 Competition: Results, findings, conclusion and way forward," *Int J Forecast*, vol. 34, no. 4, pp. 802–808, 2018.
- [17] C. W. J. Granger and R. Ramanathan, "Improved methods of combining forecasts," *J Forecast*, vol. 3, no. 2, pp. 197–204, 1984.
- [18] M. Z. K. Khan, R. Mehrotra, A. Sharma, and A. Sankarasubramanian, "Global sea surface temperature forecasts using an improved multimodel approach," *J Clim*, vol. 27, no. 10, pp. 3505–3515, 2014.
- [19] 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, 2014.
- [20] P. Radchenko, A. L. Vasnev, and W. Wang, "Too similar to combine? On negative weights in forecast combination," *Business Analytics*, 2020.
- [21] M. T. Yilmaz and W. T. Crow, "Evaluation of assumptions in soil moisture triple collocation analysis," *J Hydrometeorol*, vol. 15, no. 3, pp. 1293–1302, 2014.
- [22] V. R. R. Jose and R. L. Winkler, "Simple robust averages of forecasts: Some empirical results," *Int J Forecast*, vol. 24, no. 1, pp. 163–169, 2008.
- [23] F. C. Palm and A. Zellner, "To combine or not to combine? issues of combining forecasts," *J Forecast*, vol. 11, no. 8, pp. 687–701, 1992.
- [24] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. Cambridge, Massachusetts: MIT Press, 1949.
- [25] F. X. Diebold and P. Pauly, "Structural change and the combination of forecasts," *J Forecast*, vol. 6, no. 1, pp. 21–40, 1987.
- [26] N. E. Coulson and R. P. Robins, "Forecast combination in a dynamic setting," *J Forecast*, vol. 12, no. 1, pp. 63–67, 1993.
- [27] C. K. Min and A. Zellner, "Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates," *J Econom*, vol. 56, no. 1–2, pp. 89–118, 1993.
- [28] M. Deutsch, C. W. J. Granger, and T. Terasvirta, "The combination of forecasts using changing weights," *Int J Forecast*, vol. 10, no. 1, pp. 47–

- 57, 1994.
- [29] F. X. Diebold and P. Pauly, "The use of prior information in forecast combination," *Int J Forecast*, vol. 6, no. 4, pp. 503–508, 1990.
- [30] N. Terui and H. K. van Dijk, "Combined forecasts from linear and nonlinear time series models," *Int J Forecasting*, vol. 18, no. 3, pp. 421–438, 2002.
- [31] R. T. Clemen and R. L. Winkler, "Combining economic forecasts," *J Bus Econ Stat*, vol. 4, no. 1, pp. 39–46, 1986.
- [32] J. Smith and K. F. Wallis, "A simple explanation of the forecast combination puzzle," *Oxf Bull Econ Stat*, vol. 71, no. 3, pp. 331–355, 2009.
- [33] V. Genre, G. Kenny, A. Meyler, and A. Timmermann, "Combining expert forecasts: Can anything beat the simple average?," *Int J Forecast*, vol. 29, no. 1, pp. 108–121, 2013.
- [34] 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, 1998.
- [35] 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 Obs Geoinf*, vol. 45, pp. 200–211, 2016.
- [36] A. Gruber, C. H. Su, W. T. Crow, S. Zwieback, W. A. Dorigo, and A. 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, 2016.
- [37] 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, 2014.
- [38] J. Dong, W. T. Crow, Z. Duan, L. N. Wei, and Y. Lu, "A double instrumental variable method for geophysical product error estimation," *Remote Sens Environ*, vol. 225, pp. 217–228, 2019.
- [39] J. Dong, L. N. 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, p. 124413, 2020.
- [40] Y. Y. Liu *et al.*, "Trend-preserving blending of passive and active microwave soil moisture retrievals," *Remote Sens Environ*, vol. 123, no. 0, pp. 280–297, 2012.
- [41] 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, 2017.
- [42] 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, 2018.
- [43] H. E. Beck *et al.*, "Evaluation of 18 satellite-and model-based soil moisture products using in situ measurements from 826 sensors," *Hydrol Earth Syst Sc*, vol. 25, no. 1, pp. 17–40, 2021.
- [44] D. F. T. Hagan *et al.*, "Maximizing temporal correlations in long-term global satellite soil moisture data-merging," *Remote Sens*, vol. 12, no. 13, p. 2164, 2020.
- [45] S. Kim, H. Pham, Y. Y. Liu, L. Marshall, and A. Sharma, "Improving the combination of satellite soil moisture datasets by considering error cross-correlation: A comparison between triple collocation (TC) and extended double instrumental variable (EIVD) alternatives," *IEEE Trans Geosci Remote Sens*, vol. Early Access, pp. 1–11, 2020.
- [46] S. Chowdhury and A. Sharma, "Global sea surface temperature forecasts using a pairwise dynamic combination approach," *J Clim*, vol. 24, no. 7, pp. 1869–1877, 2011.
- [47] L. Xu, N. Chen, H. Moradkhani, X. Zhang, and C. Hu, "Improving global monthly and daily precipitation estimation by fusing gauge observations, remote sensing, and reanalysis data sets," *Water Resour Res*, vol. 56, no. 3, p. e2019WR026444, 2020.
- [48] S. Kim, R. M. Parinussa, Y. Y. Liu, F. M. Johnson, and A. Sharma, "Merging Alternate Remotely-Sensed Soil Moisture Retrievals Using a Non-Static Model Combination Approach," *Remote Sens*, vol. 8, no. 6, p. 518, 2016.
- [49] A. Timmermann, "Forecast Combinations," in *Handbook of economic forecasting*, vol. 1, 2006, pp. 135–196.
- [50] H. E. Beck *et al.*, "Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments," *J Clim*, vol. 33, no. 4, pp. 1299–1315, 2020.
- [51] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Q J R Meteorol Soc*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [52] C. Jiménez, B. Martens, D. M. Miralles, J. B. Fisher, H. E. Beck, and D. Fernández-Prieto, "Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements," *Hydrol Earth Syst Sc*, vol. 22, no. 8, pp. 4513–4533, 2018.
- [53] B. N. Datta, *Numerical Linear Algebra and Applications*. Siam, 2010.
- [54] Y. Huang, J. Benesty, and J. Chen, "Wiener filter and basic adaptive algorithms," pp. 31–57, 2006.
- [55] E. J. Candès, X. D. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: theory and algorithms," *IEEE Trans Inf Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.
- [56] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," *IEEE Signal Process Mag*, vol. 32, no. 3, pp. 87–109, 2015.
- [57] 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, 2018.
- [58] N. N. Das, D. Entekhabi, E. G. Njoku, J. J. Shi, J. T. Johnson, and A. Colliander, "Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data," *IEEE Trans Geosci Remote Sens*, vol. 52, no. 4, pp. 2018–2028, 2013.
- [59] R. A. M. de Jeu, W. Wagner, T. R. H. Holmes, A. J. Dolman, N. C. van de Giesen, and J. Friesen, "Global soil moisture patterns observed by space borne microwave radiometers and scatterometers," *Surv Geophys*, vol. 29, no. 4–5, pp. 399–420, 2008.
- [60] M. J. Brodzik, B. Billingsley, T. Haran, B. Raup, and M. H. Savoie, "EASE-Grid 2.0: Incremental but significant improvements for earth-gridded data sets," *Isprs Int J Geo-Inf*, vol. 1, no. 1, pp. 32–45, 2012.
- [61] 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, 2009.
- [62] 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)*, NASA National Snow and Ice Data Center Distributed Active Archive Center.
- [63] R. M. Parinussa, T. R. H. Holmes, N. Wanders, W. A. Dorigo, and R. A. M. de Jeu, "A preliminary study toward consistent soil moisture from AMSR2," *J Hydrometeorol*, vol. 16, no. 2, pp. 932–947, 2015.
- [64] M. C. Peel, B. L. Finlayson, and T. A. McMahon, "Updated world map of the Köppen-Geiger climate classification," *Hydrol Earth Syst Sc*, vol. 11, no. 5, pp. 1633–1644, 2007.
- [65] 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, 2010.
- [66] W. Dorigo *et al.*, "ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions," *Remote Sens Environ*, vol. 203, pp. 185–215, 2017.
- [67] A. Gruber, T. Scanlon, R. van der Schalie, W. Wagner, and W. Dorigo, "Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology," *Earth Syst Sci Data*, vol. 11, no. 2, pp. 717–739, 2019.
- [68] W. A. Dorigo *et al.*, "The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements," *Hydrol Earth Syst Sc*, vol. 15, no. 5, pp. 1675–1698, 2011.
- [69] 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, pp. 1–21, 2013.
- [70] T. Lebel *et al.*, "AMMA-CATCH studies in the Sahelian region of West-Africa: An overview," *J Hydrol*, vol. 375, no. 1–2, pp. 3–13, 2009.
- [71] M. Zreda, D. Desilets, T. P. A. Ferre, and R. L. Scott, "Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons," *Geophys Res Lett*, vol. 35, no. 21, p. L21402, 2008.
- [72] K. Yang *et al.*, "A multiscale soil moisture and freeze-thaw monitoring network on the third pole," *B Am Meteorol Soc*, vol. 94, no. 12, pp. 1907–1916, 2013.

- [73] T. Tagesson *et al.*, “Ecosystem properties of semiarid savanna grassland in West Africa and its relationship with environmental variability,” *Glob Chang Biol*, vol. 21, no. 1, pp. 250–264, 2015.
- [74] 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 Sc*, vol. 16, no. 5, pp. 1445–1463, 2012.
- [75] A. B. Smith *et al.*, “The Murrumbidgee soil moisture monitoring network data set,” *Water Resour Res*, vol. 48, no. 7, p. W07701, 2012.
- [76] 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.
- [77] 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, pp. 1–12, 2015.
- [78] 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. IGARSS’07*, 2007, pp. 1196–1199.
- [79] G. Leavesley *et al.*, “A modeling framework for improved agricultural water supply forecasting,” in *AGU Fall Meeting Abstracts*, 2008, vol. 1, p. 0497.
- [80] 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 Sci Technol Forum’10*, 2010.
- [81] S. Zacharias *et al.*, “A network of terrestrial environmental observatories in Germany,” *Vadose Zone J*, vol. 10, no. 3, pp. 955–973, 2011.
- [82] J. E. Bell *et al.*, “US Climate Reference Network soil moisture and temperature observations,” *J Hydrometeorol*, vol. 14, no. 3, pp. 977–988, 2013.
- [83] 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, 2014.
- [84] W. Wagner, G. Lemoine, and H. Rott, “A method for estimating soil moisture from ERS scatterometer and soil data,” *Remote Sens Environ*, vol. 70, no. 2, pp. 191–207, 1999.
- [85] L. Brocca, F. Melone, T. Moramarco, W. Wagner, and S. Hasenauer, “ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy,” *Remote Sens Environ*, vol. 114, no. 11, pp. 2745–2755, 2010.
- [86] C. Albergel *et al.*, “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 Sc*, vol. 12, no. 6, pp. 1323–1337, 2008.
- [87] 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, 2019.
- [88] A. Al-Yaari *et al.*, “Evaluating soil moisture retrievals from ESA’s SMOS and NASA’s SMAP brightness temperature datasets,” *Remote Sens Environ*, vol. 193, pp. 257–273, 2017.
- [89] F. T. Ulaby and E. A. Wilson, “Microwave attenuation properties of vegetation canopies,” *IEEE Trans Geosci Remote Sens*, vol. 23, no. 5, pp. 746–753, 1985.
- [90] T. Jackson and T. Schmugge, “Vegetation effects on the microwave emission of soils,” vol. 36, no. 3, pp. 203–212, 1991.
- [91] T. De Groeve, Z. Kugler, and G. R. Brakenridge, “Near real time flood alerting for the global disaster alert and coordination system,” in *Proc. ISCRAM’07*, 2007, pp. 33–39.
- [92] 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, 2019.
- [93] E. A. Brandes, J. Vivekanandan, and J. W. Wilson, “A comparison of radar reflectivity estimates of rainfall from collocated radars,” *J Atmos Ocean Technol*, vol. 16, no. 9, pp. 1264–1272, 1999.
- [94] Y. Y. Liu, R. A. M. de Jeu, M. F. McCabe, J. P. Evans, and A. I. J. M. van Dijk, “Global long-term passive microwave satellite-based retrievals of vegetation optical depth,” *Geophys Res Lett*, vol. 38, no. 18, p. L18402, 2011.
- [95] 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, 2017.
- [96] L. Moesinger *et al.*, “The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA),” *Earth Syst Sci Data*, vol. 12, no. 1, pp. 177–196, 2020.

**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.

**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.

**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.

**Sean I. Young** received B.Com and B.E. degrees in software engineering from the University of Auckland in 2008, and the M.Eng.Sc and Ph.D. degrees from the University of New South Wales, Sydney, in 2011 and 2018, respectively. He is currently a postdoctoral researcher at Stanford University, Stanford, CA. In 2016, he was a visiting researcher at InterDigital Communications, San Diego, CA. His research interests are large-scale optimization and inverse problems in image processing. He received the APRS/IAPR best paper award at Australian Pattern Recognition Society’s DICTA 2018 conference, together with David Taubman.