Rethinking Satellite Data Merging: From Averaging to SNR Optimization

Seokhyeon Kim, Ashish Sharma, Yi Y. Liu, and Sean I. Young

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 article, 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-opt significantly improves merging over the existing weighted averaging scheme.

Index Terms—Data merging, optimization, satellite remote sensing, signal-to-noise-ratio (SNR), weighted averaging.

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 improved predictions. The key idea behind weighted averaging is to take independent information from data sources in hope of deriving an improved prediction via cancelation 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 are 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 zero and one [3], [20] but this too comes from a misunderstanding of merging weights as some quantitative metric—satellite data exhibit considerable variability.
interdataset 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 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 (SM) and land surface temperature (LST) 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 optimization (SNR-opt) to ascertain minimum MSE merging weights using SNRs of the parent products and derive a relationship between these optimum weights and the existing weighted average ones (Section III). Second, we propose 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 SM and land surface temperature (LST) 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.

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 suboptimality 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-square (OLSs) 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-opt allows an interpretation of merging as a multi-input single-output (MISO) Wiener filter, itself closely related to general OLS regression.

To address the nonstationarity 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], and [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. 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 cross correlation (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 variable (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 extended IVd (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 SM [40]–[44], 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 SM data based on TC-derived error statistics. Kim et al. [45] use TC with EIVd to minimize errors in the merged SM 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 SM 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 SM datasets.

III. MATHEMATICAL FRAMEWORK

A. Weighted Averaging

To motivate optimization-based satellite data merging, let us first analyze the popular weighted average scheme [3], [49] and its underlying assumptions. Our goal is to predict an unknown quantity (e.g., SM across time at some location) \( y \in \mathbb{R} \) as a weighted average of \( N \) given parent datasets or products \( (x_1, \ldots, x_N) = x \in \mathbb{R}^N \), related to the unknown quantity \( y \) as \( x = y1 + e \) with additive noise \( e \). We assume that \( e \) 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 historical data [50], [51] back to our prediction.

The averaging weights for the prediction \( x = y1 + e \) can be expressed as the solution to the problem

\[
\begin{align*}
\min g(u) &= \mathbb{E}(e^T u)^2 = \mathbb{E}(x^T u - y1^T u)^2 \\
\text{s.t. } h(u) &= 1^T u = 1
\end{align*}
\]

[49]. The solution of this problem is given by [3], [41], [49]

\[
u^* = (1^T \mathbb{E}(ee^T)^{-1} 1)^{-1} \mathbb{E}(ee^T)^{-1} 1
\]

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

Several works [3], [20] additionally trim negative weights and normalize the remaining nonnegative ones but such a practice cannot be justified even from the point of view of (2). Several authors also \( \text{ad hoc} \) assume errors to be independent of 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 MSE of the prediction. Notice first that the vector \( u \in \mathbb{R}^N \) which minimizes the MSE of prediction \( \hat{y} = x^T u \) is obtained by solving the problem

\[
\min f(u) = \mathbb{E}(x^T u - y)^2
\]

[17]. The solution of the above problem is

\[
u^* = \mathbb{E}(xx^T)^{-1}\mathbb{E}(yx)
\]

in which we assume \( \mathbb{E}(xx^T) \) is always positive definite. Noting that \( \mathbb{E}(xx^T) = \mathbb{E}(ee^T) + y^21^T1 \) and \( \mathbb{E}(yx) = \mathbb{E}(y^2)1 \), we can rewrite the above solution as

\[
u^* = (\mathbb{E}(ee^T)/\mathbb{E}(y^2) + 11^T)^{-1}1
\]

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

From (5), one can see that the elements of \( u^* \) need not sum to 1 in the general case (consider the case \( N = I \), for example) and that the sum only approaches 1 in the limit as the matrix of SNRs \( N \to 0 \); see supplement for numerical examples. We can also demonstrate that not all elements of \( u^* \) need be positive in general. Since the weighted averaging weights \( u^* \) are different from \( u^* \) in general, they must additionally be suboptimal with respect to the MSE objective of (3). Fig. 2 plots the optimality gap between the two MSEs as \( N = \mathbb{E}(ee^T)/\mathbb{E}(y^2) \in S^{2N}_{++} \) is subjected to scaling of \( \mathbb{E}(y^2) \) (a), or transformations in
\[ \mathbb{E}(ee^T) \] by element-wise scaling (b) or rotations (c). SNR-opt still produces a lower MSE than the weighted average scheme even when \( N \) is grossly under- or over-estimated (d). In the general case, datasets \( (x_1, \ldots, x_N) = x \in \mathbb{R}^N \) relate to a random quantity \( y \) as

\[ x = ya + b + e \]

with multiplicative and additive factors \( a, b \in \mathbb{R}^N \) and a zero-mean additive noise \( e \in \mathbb{R}^N \) that we continue to assume to be jointly Gaussian. We may safely assume \( b = 0 \) since \( x \) can be debiased in advance using \( b = \mathbb{E}[x] \). In this case, the optimal weights are given by

\[ u^* = (N + aa^T)^{-1}a \]

(6)

in which \( N = \mathbb{E}(ee^T)/\mathbb{E}(y^2) \) as before. Therefore, normalizing the parent products \( x \) to \( x' = y1 + e \) 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 historical or reanalysis data suggests this is not a concern in such tasks as SM estimation. In fact, Fig. 2(d) shows that SNR-opt is not sensitive to errors in the estimation of \( \mathbb{E}(y^2) \).

Despite their cosmetic differences, there is a simple scaling relationship between the two weights \( u^* \) and \( u^t \). Applying the Sherman–Morrison formula to \( u^t = (N + 11^T)^{-1}1 \) yields

\[ u^t = N^{-1}1 - \frac{N^{-1}11^TN^{-1}1}{1 + 1^TN^{-1}1} \]

(7)

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

\[ u^* = \left( \mathbb{E}(y^2)^{-1} + 1^T\mathbb{E}(ee^T)^{-1}1 \right)^{-1} \mathbb{E}(ee^T)^{-1}1 \]

(8)

casting expression (5) into a form more similar to (2). We thus have a simple linear relationship \( u^* = su^t \) where

\[ 0 < s = \frac{1^T\mathbb{E}(ee^T)^{-1}1}{\mathbb{E}(y^2)^{-1} + 1^T\mathbb{E}(ee^T)^{-1}1} < 1 \]

(9)

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

Since \( u^t \) is a scaled version of the optimal \( u^* \), 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 \( \mathbb{E}(ee^T) \). In the case where \( x = y1 + e \), the vector of weights \( u^t \) is maximally correlating if it solves the problem

\[ \max r(u) = \frac{\mathbb{E}(y(x^Tu))}{\sqrt{\mathbb{E}(y^2)} \sqrt{\mathbb{E}((x^Tu)^2)}} \]

(10)

or, equivalently, the problem

\[ \max r^2(u) \propto \frac{u^T\mathbb{E}(yx)E(yx)^T}{u^T\mathbb{E}(xx^T)u} = \frac{u^T11^Tu}{u^T\mathbb{E}(xx^T)u} \]

(11)

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

Writing the two matrices as \( A = 11^T \) and \( B = \mathbb{E}(xx^T) \) for brevity, the maximally correlating weight vector \( u^* \) is obtained as any leading eigenvector of the pair of matrices \((A, B)\), that
is, by solving the eigenvalue problem \( Au = \beta Bu \) (only one of the eigenvectors has a nonzero eigenvalue since the rank of \( A \) is one). From the 1-D space of solutions \( u^1 \), only the vector \( u^* \) jointly maximizes correlation and minimizes the error variance such that the choice of merge weights \( u^1(2) \) in preference to \( u^* \) (5) is difficult to justify.

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

\[
u^*(\omega) = (N(\omega) + 11^T)^{-1} 1, \quad \omega \in [-\pi, +\pi)
\]

in which \( u^*(\omega) \) 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 \( u^*[m] = u^*(2\pi m/M) \) for \( 0 \leq m < M \), then finally taking the inverse DFT of the \( M \)-vector of combined coefficients. We can see that the merging weight vector (5) is a special case of (12) in which \( N(\omega) \) 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 \( S_{er}(\omega) \) 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 \( N \) along with scale factors \( a \). In general, we can form an empirical covariance matrix of \( x \) from measurements of the parent products (with their additive biases removed). Since we assume \( E(y^2) \) is known, we can equate the empirical covariances to the unknowns as

\[
C = \frac{E(xx^T)}{E(y^2)} \approx N + aa^T
\]

our objective being to simultaneously estimate the elements of matrix \( N \) together with those of \( a \). In the case where prediction errors \( \mathbf{e} \) are uncorrelated as sometimes assumed [3], we could form and solve a nonlinear system of \((N+1)N/2\) equations in \( 2N \) unknowns \( a_1, \ldots, a_N, E(e_1^2), \ldots, E(e_N^2) \). 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 \( a = 1 \) 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 \( N \). However, when \( N \) is large and parent datasets are gathered independently, it is reasonable to assume that the off-diagonal elements of \( N \) are small. We propose to estimate the SNR matrix \( N \) and the scale factors \( a \) using the following \( \ell^1 \)-norm minimization problem [which we coin SNR estimation (SNR-est)]:

\[
\min f(a) = \| \mathbf{I}((\mathbf{1}^T) \circ \text{abs}(C - aa^T)) \mathbf{1} \|_1
\]

s.t. \( h(a) = \text{diag}(aa^T) - \text{diag}(C) \leq 0 \) (15)

in hopes that the off-diagonals of \( N^* = C - aa^T \) produced by the solution \( a^* \) are small. Whereas (8) is not convex, related nonconvex 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 \( a \) is used to avoid getting trapped in local minima.

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

\[
a \leftarrow a - \eta (\mathbf{1}^T \circ \text{sgn}(aa^T - C)) a
\]

(15d)

\[
a \leftarrow a - \text{sgn}(a) \sqrt{\max(\text{diag}(aa^T - C), 0)}
\]

(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 \( aa^T \) do not exceed those of \( C \) (15p). In Fig. 4, 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 \( N \) with a total power of one).
We also evaluate the performance of SNR-est in recovering \( N \) and \( a \) from random \( C \). We choose \( \mathbb{E}(y^2) = 1 \) and \( N = 3 \) to directly compare SNR-est against TC. We first generate \( n_{ii} \in [0, 1], 1 \leq i \leq 3 \) with uniform distribution then assign \( n_{ij} = \rho(n_{ii}n_{jj})^{1/2} \) 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 root mean square error (RMSE) and a confidence interval. We use parameters \( \eta = 0.1, \beta = 0.1 \) and 1000 iterations of descent. In Fig. 5 (left), we show the failure rates of both methods across \( \rho \), where an estimation run is said to have failed if the estimates of \( n_{ii} \) come out negative or those of the factors \( a_i \), complex, as can happen with TC. Observe that SNR-est never produces infeasible estimates due to the explicit feasibility constraint in (14). By contrast, TC begins to fail with increasing \( \rho \) value owing to its zero-ECC assumption. Fig. 5 (right) shows that SNR-est and TC produce comparable errors in \( N \). RMSE of \( a \) (not plotted) have a similar trend across \( \rho \) since \( N \) and \( a \) are linearly related through (13).

IV. EXPERIMENTAL EVALUATION

This work analyzes the accuracy of predictions derived from datasets combined using the weighted average and the proposed SNR-opt schemes. For evaluation, we run experiments where three satellite SM products (ASCAT, SMAP, and LPDM) 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 (ERAS-Land), previously demonstrated to have good performance [43], [44]. 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, the RMSE of predictions is mainly 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 5 kg/m², as suggested in [58]. For LPDM 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 merging results are additionally evaluated against ground-based SM obtained via the International SM 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 (<10 cm); 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—see Table I.

Whereas our parent data represent SM 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 5–10 cm. 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 SM 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

\[
W(t_n) = W(t_{n-1}) + K_n (M(t_n) - W(t_{n-1}))
\]

in which \( W(t_n) \) and \( M(t_n) \) denote the soil water index and the SM 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

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

in which \( K_1 = 1 \), and parameter \( T \) is in \([1, 100]\), the time scale of SM 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 SM 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 active-passive combined SM product of ESA CCI SM [41], [66], [67]—see Table I for details. (ESA) CCI SM contains daily surface SM data at 0.25° 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 SM retrievals derived from four active and eight passive microwave sensors. The datasets are fused using resampling to 0.25° 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 SM time series data are statistically significant, suggesting both the active and the passive data are reliable at a given pixel [41].
TABLE I
SUMMARY OF PARENT, REFERENCE, ANCILLARY AND VALIDATION DATASETS USED IN THIS WORK

| Product name                                      | Resolutions (temporal/spatial) | Details                                                                 |
|--------------------------------------------------|--------------------------------|-------------------------------------------------------------------------|
| ASCAT (Advanced SCATiomer) Level 2 V5            | Daily/overpass (ascend/descend) at 9:30 PM/AM (local time) | 0.25°×0.25° Active C-band (5.3 GHz) radar backscatter from ASCAT on board MetOp-B |
| SMAP (Soil Moisture Active Passive) Level 3       | Daily/overpass (ascend/descend) at 6 PM/AM (local time)     | 36 km×36 km EASEv2-Grid Passive L-band (1.41 GHz)                       |
| LPRM (Land Parameter Retrieval Model) Level 3     | Daily SM/LST V1 [63]         | 0.25°×0.25° Passive X-band (10.65 GHz) from AMSR2 on GCOM-W            |
| ERA5-Land (European Centre for Medium-Range Weather Forecasts Re-Analysis 5 Land) [51] | Hourly                         | 0.1°×0.1° Volumetric soil water content in layer 1 (0–0.07 m), Skin temperature |
| CZ Updated Köppen–Geiger climate classification [64] | -                             | 0.25°×0.25° 5 primary classes: tropical, arid, temperate, cold, and polar regions |
| MCD12C1 (Moderate Resolution Spectroradiometer Terra+Aqu land cover climate modeling grid) V1 [65] | Yearly (2015)                | 0.05°×0.05° 6 primary classes: forest, shrublands, woodlands, grasslands, croplands, and unvegetated regions |
| ESA CCI SM (European Space Agency Climate Change Initiative Soil Moisture) V05.2 [41, 66, 67] | Daily                         | 0.25°×0.25° Active–passive combined surface SM product                 |
| ISMN (International Soil Moisture Network) [68, 69] and [70-82] | Hourly                        | 425 stations across 18 networks: SNOTEL [128], SCAN [111], PBO-HED [95], UCRN [112], RSMN [114], GOZeNT, SMOSSMAN, RISMA, SOILSCAPE, GROWN, REMEDHUSH, AMMA-CATCH, HOBET, PE2, HIEBRZA-SI, COSMOS, DAHRA, and TERENO |

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 to 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 use 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 $E(y^2) = 1$ is always assumed). Second, we estimate for every pixel the SNR matrix $N = E(\mathbf{e} \mathbf{e}^T)/E(y^2)$ and the scale factors $a$. We experiment with both the true and the estimated versions of $a$ and $N$, with the estimated ones obtained using the method described in Section III-C. We find the initial solution of (14) good enough as $a$. 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 SM data. This approximate value of $\beta$ is intended to show that the proposed SNR-opt outperforms the weighted average even with rough estimates of $a$. More careful estimates of $a$ can produce better outcomes—at worst, SNR-opt degenerates to a weighted averaging. Observe that weighted averaging (2) assumes $a = 1$ but SNR-opt (5) allows arbitrary $a$ for weight computation. To facilitate fairer comparisons between the two methods, we first estimate $a$ via (14) then rescale parent products $\mathbf{x} = \mathbf{y} + \mathbf{e}$ to $\mathbf{x} = \mathbf{y} + \mathbf{e}$ with $N = E(\mathbf{e} \mathbf{e}^T)/E(y^2)$ so that $N = E(\mathbf{e} \mathbf{e}^T)/E(y^2)$ 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 $a$ has been estimated, $\mathbf{u}^*$ can be obtained for $\mathbf{x}$ directly from (6) using $\mathbf{N} = \mathbf{C} - \mathbf{a} \mathbf{a}^T$, $\mathbf{C} = E(\mathbf{x} \mathbf{x}^T)/E(y^2)$. 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 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 since satellite observations result from a topsoil layer of a few centimeters rather than from the soil temperature estimated at the deeper layer, i.e., 7 cm. The two satellite LST datasets SMAP and LPRM are derived from the microwave brightness temperature observed through their descending overpasses, respectively at 6 A.M., and 1:30 A.M. local time. For this assume that the difference in LST at nighttime is not considerable compared to that of daytime. 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...
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Fig. 6. relRMSE of soil moisture prediction against the reference (ERA5-Land) over the five-year study period April 2015–March 2020, based on (Left) true and (Right) estimated scale factors and SNRs. The top and the middle plots show the RMSE of the averaged and our SNR-opt based merged products, 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.

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 $\beta = 0.6$.

D. Evaluation Metrics

We use two metrics for comparing predictions between the ground truth (or the reference) $y$ and $\hat{y}$ (prediction of $y$)—RMSE (or difference) and Pearson’s correlation $R$:

$$\text{RMSE}(\hat{y}, y) = \sqrt{\mathbb{E}((\hat{y} - y)^2)}$$

$$R(\hat{y}, y) = \frac{\mathbb{E}((y - \mathbb{E}(y))(\hat{y} - \mathbb{E}(\hat{y})))}{\sqrt{\mathbb{E}((y - \mathbb{E}(y))^2)\mathbb{E}((\hat{y} - \mathbb{E}(\hat{y}))^2)}}$$

We also use unitless, relative RMSE (relRMSE) defined as:

$$\text{relRMSE}(\hat{y}, y) = \sqrt{\frac{\mathbb{E}((\hat{y} - y)^2)}{\mathbb{E}((y - \mathbb{E}(y))^2)}}$$

that is, a standardized version of RMSE.

V. RESULTS AND DISCUSSION

A. SM Prediction

Fig. 6 shows global spatial distributions of relRMSE 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 Sahara 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 relRMSE of the predictions, unweighted averaging, CCI SM, and two parents (ASCAT and SMAP) are given in Fig. 7 (first row). We stratify all RMSEs on the CZ and LC classes as previously mentioned. 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 CZs; shrub and unvegetated LCs) 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 above-mentioned areas where SM retrieval is problematic. Unweighted averaging (green) shows the lowest overall performance.

As shown in the second (RMSE) and third (Pearson’s $R$) rows of Fig. 7, we further compare the errors of merged products, unweighted averaging, the active–passive combined product of CCI SM, ERA5-Land and parents (ASCAT and SMAP) against the ground measurements from the 425 ISMN stations as the truth. Again, CCI SM results have been included only as 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 with other well-established products. For RMSE (second row of Fig. 7), SNR-opt produces results that are consistently better than the others. It is noted the weighted averaging results are as good as or worse than CCI SM (itself based on a weighted averaging scheme and TC) and underperforms SMAP especially under the problematic retrieval conditions. Weighted averaging performs similar to unweighted averaging, except in the unvegetated regions. The third row of Fig. 7 verifies SNR-opt and weighted averaging indeed 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.

We also note that the data merged by SNR-opt often outperforms ERA5-Land across the various conditions and at least similar excepting for shrub regions. Although ERA5-Land with good overall performance as a single product [43] is taken as the reference, it does not mean that the reference is perfect, at least for now. As shown in [43] and other studies, see [89], various modeling and satellite-derived data compete with other data rather than perfect, and are improved through merging (or...
assimilation). In general, this is because land surface models focus on getting different fluxes correct rather than producing accurate SM products against direct observations [90], and various satellite data are based on microwave observations on the land surface but are limited by imperfect surface parameterization and retrieval algorithms [91].

When comparing these evaluation results with existing validation studies for various SM data, the following additional insights on data merging are presented. The early version of the CCI SM active-passive product is based on unweighted averaging [40]. From the validation of the early version of CCI SM using ground networks over the world, the correlation improvement for the merged data was not notable and exhibited similar interquartile ranges (0.50–0.80) with those of the parent products. Consequently, when the parents show similar performance in R, the merged product has been improved but in areas where either parent is significantly degraded, the performance of the merged product is comparable to those of the parent products. This tendency is also shown in Fig. 7 (third row), where unweighted averaging shows severe performance degradations across shrub and unvegetated regions where the functioning of ASCAT is problematic. The early version of CCI SM produces unbiased RMSE (ubRMSE) values around 0.05 (m$^3$/m$^3$) and interquartile ranges generally residing within 0.05–0.10 [40], [83]. The TC-based weighted averaging [41] started being applied for a later version of CCI SM by which the performance has been continuously improved [66], [67]. A direct comparison of our results with those for existing products is not available due to the differences in the CCI SM versions and the data processing. However, given that SNR-opt shows better performance than weighted averaging in this study, applying SNR-opt to weighted averaging based CCI SM has the potential to provide further improvements in the product. According to Ma et al. [92] where various satellite data (SMAP, AMSR2, and CCI SM) have been comprehensively evaluated using global ground data, CCI SM shows the lowest ubRMSE overall, while SMAP (passive) presents the highest correlation in general. This justifies the inclusion of SNR in merging SM data which will be synergetic with using SNR-opt. Zhou et al. [93] show the merged data is considerably improved in both RMSE and R by considering errors in time and space for weighted averaging.

Fig. 8 provides a comprehensive set of plots for the RMSE of SM 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.

It is noted 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 $a$, the weighted average merging of unscaled parents $x$ using weights $u^\dagger$ is identical to merging scaled parents $x'$ using $u'=a\cdot u^\dagger$. If by chance $a\approx s_1$ (9), $u'$ becomes closer to $u^\dagger$ so using $u^\dagger$ improves the RMSE performance of weighted averaging. Similarly, $u^\dagger$ reduces to $u^\ast=a\cdot u^\dagger$ but here the performance loss in SNR-opt due to using $u^\ast$ is not as dramatic as the improvement for weighted averaging. Since one does not generally have $a\approx s_1$, this type of ad hoc RMSE performance gain for weighted averaging is not always guaranteed.

B. LST 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 A.M.) and LPRM (1:30 A.M.). 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 [94], [95]. 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 [96], [97], radar reflectivity [98], and vegetation optical depth [99]–[101]. 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. Fourth, SNR-opt is a linear OLS regression method, and so is nonrobust against outliers [102]. It is noted the general solution of SNR-opt (6) is expressed with $N$ and $a$ as an alternative form of the general OLS solution (4). The nonrobustness of OLS originates from the quadratic (i.e., squares) penalty for the residuals, by which large residuals are more emphasized for the parameter estimation. However, this is not only a matter of SNR-opt but also a matter of weighted averaging, which is also linear OLS. Fifth, SNR-est improves the weaknesses of existing TC (inconvenience to apply to more than four products, high failure rate with high error cross correlation, and error covariance assumed to be 0), but SNR-est is disadvantageous compared to TC in terms of epistemic uncertainty. While TC reduces epistemic uncertainty by assuming zero ECC to solve the indeterminate system derived from variances and covariances of the parent data [19], SNR-est fully estimates $N$ (i.e., on- and off-diagonals) (13), which can lead to additional epistemic uncertainty. Nevertheless, this zero ECC assumption...
Fig. 10. relRMSE of merge predictions formed using SNR-opt (blue) and weighted averaging (red) with LST (ERA5-Land) as the prediction target (estimated scale factors and SNR used). We condition the RMSE on the (Left) CZ and the (Right) LC classes. We include the errors of SMAP and LPRM (first and second white of each class) for reference. Outliers of box plots not shown for clarity. See also Table S3 in the article supplement.

for TC is not often held in practice [21] and thus produces erroneous solutions. Sixth, theoretically, SNR-opt always produces a smaller MSE than weighted averaging. However, as presented in Fig. 2, imperfectly estimated parameters reduce their differences in merging results even though SNR-opt is still better (e.g., see solid boxes in Fig. 7 where estimated scale factors result in a loss in SNR-opt, but gain in weighted averaging). \( N \) and \( a \) could be poorly estimated when the assumption of SNR-est that the off-diagonal elements of \( N \) are small relative to the diagonal ones is considerably violated, and accordingly, the performance of SNR-opt may degrade. However, this limitation is not specific to SNR-opt and already exists for TC that explicitly assumes zero ECC (even worse for the failure rate, see Fig. 5). 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, suboptimality of weighted averaging has not fully been discussed by earlier studies. In this work, we first demonstrated the suboptimality of weighted averages and proposed an optimization-based merging method which derives its merging weights from the SNR of the parent products. For this reason, we referred to our merging method as SNR-opt. We validated SNR-opt using SM and LST, 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 SM and LST from satellite observations.

In the case of SM datasets, the merged product produced by SNR-opt showed performance comparable to well-established SM 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 article.

AUTHOR CONTRIBUTIONS

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

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, Jan. 2021, Art. no. 112126.
[2] R. M. Parinussa, A. G. Meesters, Y. Y. Liu, W. Dorigo, W. Wagner, and R. A. 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, Jul. 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, Jul. 2016.
[6] G. Elliott and A. Timmermann, Economic Forecasting. Princeton, NJ, USA: Princeton Univ. Press, 2016.
[7] M. Araujo and M. New, “Ensemble forecasting of species distributions,” Trends Ecol. Evol., vol. 22, no. 1, pp. 42–47, Jan. 2007.
[8] C. F. Dormann et al., “Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference,” Ecol. Monog., vol. 88, no. 4, pp. 485–504, Nov. 2018.
[9] H. E. Beck et al., “MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data,” *Hydrol. Earth Syst. Sci.*, 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. Climate*, vol. 30, no. 24, pp. 10139–10148, Dec. 2017.

[11] G. Masiar 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. Sci.*, vol. 24, no. 5, pp. 2687–2710, May 2020.

[12] S. Chowdthury and A. Sharma, “Multisite seasonal forecast of arid river flows using a dynamic model combination approach,” *Water Resour. Res.*, vol. 54, no. 10, pp. 1–16, Oct. 2018.

[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, Jun. 2018.

[14] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, “A survey on machine learning for data fusion,” *Inf. Fusion*, vol. 57, pp. 115–129, May 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, May 2020.

[16] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The m4 competition puzzle,” *J. Bus. Econ. Stat.*, vol. 34, no. 4, pp. 47–57, Jun. 2016.

[17] C. W. J. Granger and R. Ramanathan, “Improved methods of combining forecasts,” *J. Forecasting*, 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. Climate*, vol. 27, no. 10, pp. 3505–3515, May 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, Sep. 2014.

[20] P. Rachdenhofer, A. L. Vasneev, and W. Wang, “Too similar to combine? On negative weights in forecast combination,” *Bus. Anal.*, 2020, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3647603

[21] M. T. Yilmaz and W. T. Crow, “Evaluation of assumptions in soil moisture triple collocation analysis,” *J. Hydrometeorol.*, vol. 15, pp. 1293–1302, Jun. 2014.

[22] V. R. R. Jose and R. L. Winkler, “Simple robust averages of forecasts: Some empirical results,” *Int. J. Forecasting*, 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. Forecasting*, vol. 12, no. 1, pp. 205–231, 1993.

[24] W. Wagner, “Recent advances in (soil moisture) triple collocation and calibration using triple collocation,” *J. Geophys. Res.*, vol. 104, no. C4, pp. 7755–7766, Apr. 1998.

[25] 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.

[26] 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.

[27] 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.

[28] C. W. J. Granger and R. Ramanathan, “Improved methods of combining forecasts,” *J. Forecasting*, vol. 3, no. 2, pp. 197–204, 1984.

[29] M. Z. K. Khan, R. Mehrotra, A. Sharma, and A. Sankarasubramanian, “Global sea surface temperature forecasts using an improved multimodel approach,” *J. Climate*, vol. 27, no. 10, pp. 3505–3515, May 2014.
remote sensing observations,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 42–54, Jan. 2013.

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

[80] J. H. J. van Dijk, “Global long-term passive microwave satellite-based retrievals of vegetation optical depth,” *Geophys. Res. Lett.*, vol. 38, no. 18, Sep. 2011, Art. no. L18402.
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 currently a Research Associate with the 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 currently 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 datasets 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. 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. He is currently with the School of Civil and Environmental Engineering at UNSW. 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 the 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, NSW, Australia, in 2011 and 2018, respectively. In 2016, he was a Visiting Researcher at InterDigital Communications, San Diego, CA, USA. He is currently a Post-Doctoral Researcher at Stanford University, Stanford, CA. His research interests are large-scale optimization and inverse problems in image processing. Dr. Young received the APRS/IAPR Best Paper Award at Australian Pattern Recognition Society’s DICTA 2018 conference, together with David Taubman.