Improving Rain/No-Rain Detection Skill by Merging Precipitation Estimates from Different Sources

JIANZHI DONG AND WADE T. CROW
Hydrology and Remote Sensing Laboratory, USDA, Beltsville, Maryland

ROLF REICHLE
Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 14 April 2020, in final form 28 July 2020)

ABSTRACT: Rain/no-rain detection error is a key source of uncertainty in regional and global precipitation products that propagates into offline hydrological and land surface modeling simulations. Such detection error is difficult to evaluate and/or filter without access to high-quality reference precipitation datasets. For cases where such access is not available, this study proposes a novel approach for improved rain/no-rain detection. Based on categorical triple collocation (CTC) and a probabilistic framework, a weighted merging algorithm (CTC-M) is developed to combine noisy, but independent, precipitation products into an optimal binary rain/no-rain time series. Compared with commonly used approaches that directly apply the best parent product for rain/no-rain detection, the superiority of CTC-M is demonstrated analytically and numerically using spatially dense precipitation measurements over Europe. Our analysis also suggests that CTC-M is tolerant to a range of cross-correlated rain/no-rain detection errors and detection biases of the parent products. As a result, CTC-M will benefit global precipitation estimation by improving the representation of precipitation occurrence in gauge-based and multisource merged precipitation products.

KEYWORDS: Precipitation; Precipitation

1. Introduction

Precipitation detection skill (i.e., the ability to accurately detect precipitation occurrence) is a key metric for quantifying the accuracy of precipitation products (Dinku et al. 2010; Hamada and Takayabu 2016). As demonstrated in gauge-based analyses (Tong et al. 2014; Yang and Luo 2014), biases in remote sensing precipitation are often attributable to their tendency to falsely detect and/or miss precipitation events. Such detection errors are unavoidably propagated into hydrological and offline land surface model simulations. For example, streamflow estimation uncertainty during low-flow periods is dominated by detection errors (e.g., Bitew and Gebremichael 2011; Bitew et al. 2012). Therefore, false precipitation events are a key error source in modeled streamflow estimates acquired from remote sensing (RS) precipitation retrievals (Tobin and Bennett 2010; Behrangi et al. 2011).

Although multisource precipitation merging can effectively reduce uncertainties in precipitation intensity estimates (Beck et al. 2019; Yang et al. 2017; Baez-Villanueva et al. 2020; Bhuiyan et al. 2019), it does not necessarily effectively remove rain/no-rain detection errors. Precipitation error is conditionally biased, since precipitation error on no-rain days is necessarily nonnegative. Consequently, the weighted averaging of multisource precipitation products accumulates false precipitation events contained in all parent products. This issue may be mitigated by high-density gauge observations (Yang et al. 2017), but is particularly acute over data-sparse regions (Dong et al. 2020a).

Precipitation occurrence can be determined using precipitation gauge observations. One gauge-based approach is to spatially interpolate binary precipitation occurrence (e.g., −1 for no-rain and +1 for rain) and then apply a threshold to the interpolated values to determine rain and no-rain grid cells (Cornes et al. 2018; Hutchinson et al. 2009). However, such gauge-based approaches are only effective when the gauge network is spatially dense. Unfortunately, large fractions of South America, Africa, central Asia, central Australia, and the high-latitude Northern Hemisphere contain few or no gauges (Chen et al. 2008), and consequently, gauge-based precipitation products are of poor quality in these regions (Reichle et al. 2017).

Machine learning based approaches have been proposed to detect rain/no-rain conditions without relying on precipitation gauge measurements (e.g., Nasrollahi et al. 2013; Cui et al. 2015; Ghajarnia et al. 2016). For instance, Nasrollahi et al. (2013) uses CloudSat observations and Moderate Resolution Imaging Spectroradiometer (MODIS) water vapor and infrared data to identify no-rain areas using an artificial neural network (ANN). They demonstrated that this framework can significantly reduce the false-alarm ratio for RS precipitation products. However, the performance of their approach is sensitive to changes in cloud condition and is not effective for detecting missed precipitation events.

Rain/no-rain errors can also be detected from a water balance perspective using satellite soil moisture retrievals in a data assimilation system (Crow and Ryu 2009; Alvarez-Garretón et al. 2016). However, applying such a data assimilation framework at continental or global scales can be challenging, particularly over regions/periods where soil moisture retrieval quality...
is relatively low. Additionally, this approach is prone to overpredict the occurrence of light precipitation events (Crow and Ryu 2009).

Alternatively, improved rain/no-rain detection skill may be achieved by statistically combining different products. Precipitation detection skill essentially reflects the probability of a product being correct with regard to precipitation occurrence. As an analogy to commonly used maximum likelihood estimation approaches (e.g., Yilmaz et al. 2012), multisource (independent) estimates can be used to constrain rain/no-rain estimates and generate new products with improved detection skill. Such a statistical approach is attractive for multisource precipitation merging for several reasons. First, relative to gauge-based correction, it leverages the detection skill of all products without placing excessive confidence on gauge-based observations (particularly valuable for data-poor regions). Second, it does not require additional ancillary data, as required, for example, during the training of machine learning approaches. Third, statistical merging approaches are not impacted by hydrological modeling uncertainties that afflict rain/no-rain correction techniques based on data assimilation. Finally, it has the flexibility to ingest rain/no-rain estimates from all the possible sources (e.g., from both cloud temperature and data assimilation based estimates) and to effectively leverage such multisource information for improving rain/no-rain time series estimates.

However, the application of any statistical merging approach requires estimates of detection skill for each individual precipitation product—which are difficult to obtain without access to high-quality reference datasets. Based on noisy, but independent, precipitation products, triple collocation (TC; Stoffelen 1998) can solve for the error variance of each product using a set of linear equations (Massari et al. 2017; Alemohammad et al. 2015; Li et al. 2018; Dong et al. 2019).

However, TC is not directly applicable to rain/no-rain skill estimates, since rain/no-rain estimates are binary with bounded errors. As a consequence, the error of a binary dataset is cross-correlated with the true signal. Such truth-nonorthogonal error, in turn, violates a key assumption of TC analysis (McColl et al. 2016).

Recently, a categorical triple collocation algorithm (CTC; McColl et al. 2016) has been proposed to evaluate the classification error of a geophysical dataset without requiring access to high-quality reference datasets. Specifically, the covariance of two independent geophysical products is a function of their classification error. Provided three independent products are available, CTC can establish a set of linear equations that can be simultaneously solved to provide estimates of classification error for each product. McColl et al. (2016) demonstrates that CTC can robustly estimate both temporally stationary and nonstationary classification errors—making it particularly valuable for evaluating geophysical classification errors (McColl et al. 2016; Lyu et al. 2018).

This study proposes a novel merging algorithm based on CTC (denoted as “CTC-M”) that maximizes the likelihood of correct rain/no-rain estimates at each time step. As such, CTC-M simultaneously leverages information from all available sources and should, in theory, outperform traditional approaches based on using only one source of information. Additionally, CTC-M does not require access to high-quality gauge observations, which makes it particularly suitable for improving precipitation detection in data-sparse regions.

2. Data and method

a. The CTC algorithm

Categorical triple collocation (CTC) requires three independent products. The rain/no-rain time series provided by product $i$ can be expressed as

$$x_i = P + e_i,$$

where $x_i$ and $P$ are the estimated and the true rain/no-rain time series, with $+1$ and $-1$ representing rain and no-rain days, respectively, and $e_i$ is the classification error.

The performance of $x_i$ in correctly classifying a geophysical variable is quantified using the balanced accuracy $\pi_i$,

$$\pi_i = 0.5(A_i + D_i),$$

where $A_i$ and $D_i$ are the probability of $x_i$ being correct when $P$ is $+1$ and $-1$, respectively (Parisi et al. 2014; McColl et al. 2016). The balanced accuracy simultaneously considers the skill of a product in capturing both rain and no-rain days. For instance, a constant time series of $+1$ (i.e., events predicted at every time step) will lead to a balanced accuracy of $0.5$ (i.e., $A_i = 1$ and $D_i = 0$) that is equivalent to the skill of a random binary time series (i.e., $A_i = 0.5$ and $D_i = 0.5$). Assuming that classification errors of different products are mutually independent, the covariance $Q$ of the rain/no-rain time series provided by different products can then be expressed as

$$Q_{12} = \text{Cov}(x_1, x_2) = f(P)(2\pi_1 - 1)(2\pi_2 - 1),$$
$$Q_{13} = \text{Cov}(x_1, x_3) = f(P)(2\pi_1 - 1)(2\pi_3 - 1),$$
$$Q_{23} = \text{Cov}(x_2, x_3) = f(P)(2\pi_2 - 1)(2\pi_3 - 1),$$

where $f(P)$ equals the variance of $P$ for stationary cases and more complex $P$ statistics for nonstationary cases—see Eqs. (9) and (12) of McColl et al. (2016).

However, $\pi_i$ cannot be directly solved using CTC because $f(P)$ is typically unknown. Instead, CTC solves for $v_i$, which is linearly proportional to $\pi_i$:

$$v_1 = \sqrt{Q_{12}Q_{13}} / Q_{23},$$
$$v_2 = \sqrt{Q_{12}Q_{23}} / Q_{13},$$
$$v_3 = \sqrt{Q_{13}Q_{23}} / Q_{12},$$

where $v_i = (2\pi_i - 1)\sqrt{f(P)}$. Therefore, the relative size of $\pi_i$ for the considered products can be inferred from $v_i$ estimated by Eqs. (6)–(8).
b. CTC-based rain/no-rain merging

As mentioned above, the goal of rain/no-rain merging is to derive estimates that have optimal rain/no-rain detection skill. This can be achieved by linearly combining all three products as

\[ x_m = \text{sign}(w_1 x_1 + w_2 x_2 + w_3 x_3), \]  

where \( w \) is the weight applied to different precipitation detection products. Clearly, \( w_i \) should monotonically increase with increased rain/no-rain detection skills, i.e., products with higher accuracy should receive more weight during merging. Without loss of generality, these weights can be expressed as

\[ w_i \propto (2\pi_i - 1)^n, \]  

where \( n \) is a positive constant to be determined. Additionally, for unbiased merging, the sum of the weights should be one. Assuming all product weights are positive, this yields the constraint

\[ w_1 + w_2 + w_3 = 1. \]  

Combining (10) and (11), these weights can be expressed as

\[ w_i = \frac{(2\pi_i - 1)^n}{(2\pi_1 - 1)^n + (2\pi_2 - 1)^n + (2\pi_3 - 1)^n}. \]  

The goal of CTC-based merging is to determine a value of \( n \) such that the merged product \( x_m \) is at least as accurate as the best original product, i.e., \( \sigma_m \geq \max(\sigma_1, \sigma_2, \sigma_3) \). For unbiased cases (i.e., \( A_i = D_i \)), this optimal \( n \) value depends on the relative detection skill of the original products, and larger \( n \) values should be used when the skill of the parent products varies significantly (appendix A). However, as shown in appendix A, our results are qualitatively insensitive to a range of \( n \) values, and consistently outperform the best parent product. Based on the results of appendix A, \( n = 1.5 \) is used in this study, which is expected to be adequate for a conservative merging in most of real-world applications. Nevertheless, this assumption is revisited in section 4.

c. Continental-scale validation of rain/no-rain detection and merging

The E-OBS (version 19.e) product was used as the reference for evaluating CTC and CTC-M. This daily product is constructed from gauge stations within Europe and considers rain/no-rain spatial interpolation in a binary manner (Cornes et al. 2018). Hence, relative to other gauge-based products, e.g., the Climate Prediction Center (CPC) unified gauge-based daily precipitation (Chen et al. 2008), E-OBS is expected to have superior rain/no-rain detection skills. As a result, we limited the test of our algorithms to Europe.

As mentioned above, CTC analysis requires a triplet of independent precipitation products. Many precipitation products share similar ancillary inputs or retrieval algorithms, and acquiring three products with independent errors is challenging. However, Massari et al. (2017) demonstrated that the error in soil moisture–inverted, remotely sensed, and reanalysis precipitation products are relatively independent. Therefore, products collected from these three separate categories were used for testing CTC.

The first product used here was the SM2RAIN precipitation product (Brocca et al. 2015). In the SM2RAIN algorithm, precipitation is estimated by inverting a soil water balance model using time series of Advanced Scatterometer surface soil moisture retrievals (Wagner et al. 1999). This 0.25° daily precipitation product is available from January 2007 to June 2015 (http://hydrology.irpi.cnr.it/download-area/sm2rain-datasets/) and is based on the inversion approach described in Brocca et al. (2017).

The L3 daily 0.25° precipitation product (TRMM_3B42RT_Daily) was considered as the second independent precipitation product (Huffman et al. 2007). It is retrieved from a variety of low-Earth-orbit passive microwave observations using the Goddard Profiling Algorithm and produced by averaging the near-real-time, 3-hourly TRMM Multisatellite Precipitation Analysis (TMPA) 3B42RT product without gauge-based correction. Note that the most recent Integrated Multisatellite Retrievals for GPM (IMERG) product is the state-of-the-art RS-based precipitation product (Huffman et al. 2015). However, the skill of TRMM 3B42RT is generally similar to that of other RS-based precipitation products (Dong et al. 2020b). Therefore, TRMM 3B42RT is more representative for testing the general performance of CTC-M in merging multiple RS-based precipitation products.

The third precipitation product used for CTC was the, daily, 0.5° ERA-Interim reanalysis precipitation product (denoted as ERA; Dee et al. 2011) produced by a data assimilation system based on the European Centre for Medium-Range Weather Forecasts (ECMWF) forecast model. To be consistent with other precipitation products, ERA precipitation was resampled onto a 0.25° spatial grid using the nearest-neighbor interpolation method implemented in Beck et al. (2019). The latest ERA5 product from ECMWF directly assimilates ASCAT soil moisture. Therefore, it is likely to contain cross-correlated errors with SM2RAIN and, as a result, was not considered here.

As mentioned above, our CTC analysis was based on binary rain/no-rain time series, and a threshold value of 0.5 mm day\(^{-1}\) was used to distinguish rain and no-rain days (Dinku et al. 2010). Based on this threshold value, the four raw precipitation datasets were converted into rain/no-rain categorical measurements. Results presented below are qualitatively similar for a relatively wide range of other rain/no-rain daily threshold values (0.5–2 mm day\(^{-1}\))—see appendix B. Given the common data availability of the precipitation products, data collected from 2007 to 2015 were used for the real-world tests.

d. Precipitation evaluation metrics

Although the balanced accuracy \( \pi \) can summarize the overall accuracy of rain/no-rain detection, additional precipitation detection metrics were also considered, to be consistent with previous evaluation analyses. To this end, we counted the number of hits \( a \), false alarms \( b \), misses \( c \), and correct negatives \( d \) (see Table 1) with respect to the E-OBS gauge-based precipitation. The probability of detection (POD), which represents the fraction of actual rainfall events captured by a given precipitation product, is then expressed as
Likewise, the tendency of a product to falsely report precipitation events is captured by the false alarm ratio (FAR):

$$\text{FAR}_i = \frac{b_i}{a_i + b_i}. \quad (14)$$

Finally, the overall detection skill for a binary sequence is assessed using the Heidke skill score (HSS):

$$\text{HSS}_i = \frac{2(a_i d_i - b_i c_i)}{(a_i + c_i)(c_i + d_i) + (a_i + b_i)(b_i + d_i)}. \quad (15)$$

3. Results

a. Real world test of CTC

For a real-data test, we evaluated the performance of CTC using a cross-validation technique. Specifically, we estimated CTC detection skills using three product triplets (E-OBS-ERA-TRMM, E-OBS-ERA-SM2Rain, and E-OBS-TRMM-SM2Rain). If the assumptions underlying CTC are respected, E-OBS detection skills derived from these three different triplets should be mutually consistent. As noted above, CTC estimates represent the relative performance of precipitation products within a specific triplet, i.e., \(v_i = (2\pi_i - 1)\sqrt{f(P)}\) [see Eqs. (6)–(8)] instead of the absolute values of \(\pi_i\). Therefore, the \(v_i\) values of the precipitation products are considered for evaluation/comparison, e.g., \(v_{E-OBS} = (2\pi_{E-OBS} - 1)\sqrt{f(P)}\). As shown in Fig. 1, the relative E-OBS rain/no-rain detection skills estimated by different triplets are consistent, and the spatial correlation of E-OBS detection skill estimated by any two different triplets is greater than 0.89.

Next, Figs. 2a and 2b compare relative E-OBS and ERA performances [e.g., \(v_{E-OBS/ERA} = [(2\pi_{E-OBS} - 1)/((2\pi_{ERA} - 1)]\)] estimated by different CTC triplets. Relative performance values above/below one plotted in Fig. 2a indicate that E-OBS has higher/lower detection skills than ERA. Therefore, results in Fig. 2 demonstrate that E-OBS generally outperforms ERA in western Europe and underperforms ERA in eastern and southern Europe. This spatial tendency is further confirmed in E-OBS-ERA-SM2RAIN triplet-based estimates (Fig. 2b), and qualitatively consistent with the gauge distribution underlying the E-OBS product (Fig. 2c). Nevertheless, some regional differences in estimates derived from the two triplets are observed (cf. Figs. 2a,b), e.g., in areas of Germany. Such differences may be related to the low quality of SM2RAIN estimates in this region, which can destabilize CTC estimates. Nonetheless, the spatial correlations of E-OBS-ERA relative skill estimated by the two triplets is 0.80—confirming the general robustness of CTC.

The detection skill of TRMM and SM2RAIN is relatively low (not shown). That is, these products are outperformed by E-OBS across the domain for any given triplet and are thus of limited value for evaluating the robustness of CTC, which is why only the relative skill of E-OBS and ERA are shown in Fig. 2.

The top row of Fig. 3 assumes that E-OBS reproduces a true precipitation occurrence time series and calculates the relative weights of ERA, TRMM and SM2RAIN using E-OBS-evaluated detection skills. These weights are then compared to CTC-based weights obtained from a triplet of the ERA, TRMM and SM2RAIN products (but no gauge-based observations) to examine the ability of CTC to reproduce gauge-based results using only RS products. Note that only land grid cells where E-OBS outperforms ERA in Fig. 2a are considered in Fig. 3. Both gauge- and CTC-based estimates demonstrate that ERA is generally assigned the largest weights (i.e., identified as having the highest relative accuracy), followed by TRMM and SM2RAIN, respectively.

The spatial correlation of E-OBS- and CTC-based weight estimates range from 0.38 for ERA (due to limited spatial

![Fig. 1. Relative detection skill of E-OBS \(v_{E-OBS} = (2\pi_{E-OBS} - 1)\sqrt{f(P)}\) estimated from different triplets. Gray shading represents areas that lack adequate data for CTC analysis.](image-url)
weight variability) to 0.88 for SM2RAIN (see Fig. 3). However, inconsistencies are still observed. For instance, E-OBS suggests that ERA should be assigned the largest weight in all grid cells (relative to TRMM and SM2RAIN). In contrast, CTC results suggest that TRMM should be assigned more weight than ERA over southern Germany and scattered areas of eastern Europe. These areas correspond to grid cells where SM2RAIN is assigned extremely low weight (<0.1) and the denominator of Eqs. (6)–(8) is approximately zero, which destabilizes CTC estimates.

b. Detection skills of CTC merged rain/no-rain days

Based on the weights shown in Figs. 3d–f, merged rain/no-rain time series are derived using the ERA, TRMM and SM2RAIN data on a gridcell by gridcell basis using the CTC-M approach outlined in Eq. (9). The performance of the merged and the original products is then evaluated in Fig. 4. ERA, TRMM, and SM2RAIN all have very different detection skills. As shown in Fig. 4, ERA has highest POD, and its FAR is only slightly higher than TRMM. Hence, ERA has the highest HSS score of the three products. Relative to TRMM, SM2RAIN demonstrates a higher POD. However, its FAR is 20%–30% higher than that of TRMM. As a result, its overall performance is still slightly inferior to TRMM. Given its strong detection skill, the ERA rain/no-rain time series is assigned the largest weighting in the merging process, and, hence, its individual performance most closely matches that of the merged product (cf. the first and the fourth columns of Fig. 4). However,
relative to ERA, clearly reduced FAR and increased HSS are still noticeable in the CTC-M estimates.

Figure 5 plots differences between CTC-M and ERA precipitation detection skill for a more direct comparison. Relative to ERA, CTC-M product has decreased POD—meaning that the merged product is slightly worse than ERA in terms of detecting precipitation events. However, the FAR of the merged product is up to 20% lower than ERA over certain grid cells. As a result, the net impact of weighted merging generally improves HSS precipitation detection skill compared to ERA—the best-performing parent product (Fig. 5c).

E-OBS data used as a reference are, of course, not error-free, which may affect our evaluation results. However, given the high gauge density found within Germany, rain/no-rain estimates there are expected to have particularly low error (Fig. 2c). Nevertheless, the relative skill of ERA and CTC-M inside and outside of Germany are qualitatively consistent (Fig. 5). Therefore, while E-OBS detection error will likely affect absolute error metrics, its impact on our conclusions regarding the relative performance of products appears to be limited.

To further illustrate differences between ERA and CTC-M, Fig. 6 compares an ERA precipitation time series and CTC-M merged rain/no-rain sequence for a grid-cell in southern Germany. Relative to ERA, the merged product misses precipitation events on 3 and 16 August. However, it accurately reports no-rain days that are falsely reported (as rainy days) by ERA on 9, 14, and 19 August and 4 and 9–11 September. Hence, although the merged product slightly reduces the POD, it also successfully removes a large fraction of falsely reported precipitation events, which leads to overall superior detection skill.

![Fig. 4. Performance of (a),(e),(h) ERA, (b),(f),(j) TRMM, (c),(g),(k) SM2RAIN, and (d),(h),(l) CTC-M with regard to precipitation detection skill for (top) POD, (middle) FAR, and (bottom) HSS. Dark gray shading indicates grid cells where CTC cannot be performed due to lack of data, or where E-OBS does not outperform the other three products.](image)

![Fig. 5. Relative performance of the merged product vs the ERA product (MERGED minus ERA) with respect to the (a) POD, (b) FAR, and (c) HSS categorical performance metrics. Red (blue) colors indicate that CTC-M is better (worse) than ERA.](image)
Note that ERA tends to overestimate the occurrence of low intensity precipitation events (Beck et al. 2019). Therefore, a higher rain/no-rain threshold yields increased ERA detection skills. However, employing a different rain/no-rain threshold has no qualitative impact on the relative detection skills of CTC-M versus ERA (see appendix B).

4. Discussion and conclusions

Using CTC, we investigate the feasibility of estimating the relative detection skill of reanalysis and remote sensing precipitation products without reliance on gauge-based observations. Estimates of this skill are then used to merge multiple precipitation products into a unified time series with optimized rain/no-rain detection skill (via a novel probabilistic procedure referred to as CTC-M). We further demonstrate that CTC-M outperforms all of its constituent precipitation products in detecting precipitation occurrence via an analytical discussion (appendix A).

Real-world tests based on ERA, TRMM, and SM2RAIN precipitation products generally confirm our analytical findings. We first performed a cross-validation analysis, which shows that E-OBS detection skills derived from various CTC triplets are mutually consistent. We further demonstrate that CTC-M outperforms all of its constituent precipitation products in detecting precipitation occurrence via an analytical discussion (appendix A).

Fig. 6. Comparison of E-OBS, ERA, and the merged rain/no-rain time series for a grid cell near Nuremberg in southern Germany (see Fig. 2c for location) during August–September 2010.

Relative to existing rain/no-rain detection approaches (e.g., Ebert and Weymouth 1999; Nasrollahi et al. 2013; Cui et al. 2015; Ghajarnia et al. 2016), this study provides a more general framework for improving the precipitation detection skill. First, rain/no-rain estimates obtained from all of the above-mentioned methods can be used as independent CTC-M inputs for improved detection accuracy. Additionally, CTC-M does not require high-quality training data, which makes it applicable for rain/no-rain merging at global scales. Consequently, CTC-M can be employed within other existing multisource precipitation merging frameworks. For instance, the MSWEP version 2 product uses gauge-based precipitation observations for removing falsely reported precipitation events in ERA (Beck et al. 2019). However, as mentioned above, even in densely instrumented Europe, rain gauges are often insufficient for capturing gridcell scale precipitation. Consequently, adjusting all products to match gauge-based rain/no-rain
discrimination is risky—particularly for data-poor regions (e.g., Africa and high-latitude regions of the Northern Hemisphere) with especially low rain gauge density. Since CTC-M explicitly considers the rain/no-rain detection error within all available data sources, it is expected to be particularly useful for improving MSWEP precipitation detection skill over such areas. Analogously, the CTC-M can also be incorporated in recent machine learning based precipitation merging frameworks (e.g., Bhuiyan et al. 2019; Baez-Villanueva et al. 2020) and benefit their precipitation detection skills over data-sparse regions.

Acknowledgments. This paper was partially supported by the NASA SMAP mission through Award NNH12ZDA001N-SMAP. R. Reichle was supported by the NASA Modeling, Analysis, and Prediction program. The CTC-M source code is available at https://github.com/dongjianzhi/Rain-no-rain-merging. The SM2Rain data used in this study are available on http://hydrology.irpi.cnr.it/download-area/sm2rain-data-sets/, the ERA-Interim data are available on https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim, TRMM data are available on https://disc.gsfc.nasa.gov/, and E-OBS data are available on https://www.ecad.eu/download/ensembles/download.php.

APPENDIX A
Determination of Merging Weights
This section provides analytical solutions for the optimal n value for CTC-M in section 2b. Given that the impact of relative product skills on CTC-M is symmetric, we can assume, without loss of generality, that \( x_1 \) outperforms \( x_2 \) and \( x_2 \), i.e., \( \pi_1 = \max \{ \pi_1, \pi_2, \pi_3 \} \). Additionally, we also assume all the products are unbiased (i.e., \( A_i = D_i \)) for simplicity. Under these assumptions, \( \pi_1 \) directly represents the probability of \( x_1 \) being correct. To start, we consider a hypothetical \( n \) value that leads to \( w_1 < w_2 + w_3 \). This corresponds to the case where a small \( n \) value is used. For instance, \( n = 0 \) leads to \( w_1 = w_2 = w_3 \) and, hence, \( w_1 < w_2 + w_3 \), for any given \( \pi \) value. In this case, \( x_m \) is correct when at least two products are correct. Therefore, the balanced accuracy of \( x_m \) can be estimated as

\[
\pi_m = \pi_1 \pi_2 (1 - \pi_3) + \pi_1 (1 - \pi_2) \pi_3 + (1 - \pi_1) \pi_2 \pi_3. \tag{A1}
\]

The detection skill differences of \( x_m \) and \( x_1 \) (the best original product) is calculated as

\[
J = \pi_m - \pi_1 = \pi_1 \pi_2 \pi_3 + \pi_1 \pi_2 (1 - \pi_3) + \pi_1 (1 - \pi_2) \pi_3 + (1 - \pi_1) \pi_2 \pi_3,
\tag{A2}
\]

where \( J \) is the difference between \( \pi_m \) and \( \pi_1 \), and \( J > 0 \) means \( x_m \) outperforms \( x_1 \). Taking the partial derivative of \( J \) with respect to \( \pi_2 \) yields

\[
\frac{\partial J}{\partial \pi_2} = \pi_1 (1 - \pi_3) + \pi_3 (1 - \pi_1). \tag{A3}
\]

As noted above, all the \( \pi \) values are within the range of 0.5 to 1. Therefore, Eq. (A3) is positive. Analogously, the partial derivative of \( J \) with respective to \( \pi_1 \) is also positive. Hence, \( J \) is a monotonically increasing function of both \( \pi_2 \) and \( \pi_3 \). Here, we replace \( \pi_2 \) and \( \pi_3 \) values with \( y \), where \( y = \min \{ \pi_2, \pi_3 \} \), in Eq. (A2). This substitution yields

\[
J_m = (1 - 2 \pi_1) y^2 + 2 \pi_1 y - \pi_1, \tag{A4}
\]

where \( J_m \) is the \( J \) value when both \( \pi_2 \) and \( \pi_3 \) equal \( y \). As demonstrated in Eq. (A3), \( J \) is a monotonic function of both \( \pi_2 \) and \( \pi_3 \). Therefore, \( J \approx J_m \) when \( \pi_2 \) (or \( \pi_3 \)) is replaced by a smaller value \( y \) in Eq. (A2). Hence, a nonnegative \( J_m \) guarantees a nonnegative \( J \); i.e., the skill of \( x_m \) is at least good as that of the original products.

According to Eq. (A4), \( J_m \geq 0 \) requires

\[
\beta - \sqrt{\beta^2 - \beta} \leq y \leq \beta + \sqrt{\beta^2 - \beta}, \tag{A5}
\]

where \( \beta = \pi_1 / (2 \pi_3 - 1) \). Since \( \pi_1 \) is in the range of \([0.5, 1]\), \( \beta \) is a positive number greater than 1. Hence, Eq. (A5) can be simplified as

\[
\beta - \sqrt{\beta^2 - \beta} \leq y \leq 1. \tag{A6}
\]

According to Eq. (A6), \( J_m \) is greater than zero only when \( y > \beta - \sqrt{\beta^2 - \beta} \). Therefore, assigning a \( n \) value that yields \( w_1 < w_2 + w_3 \) guarantees the superiority of \( x_m \). However, if both \( x_2 \) and \( x_3 \) are low skill (i.e., \( y < \beta - \sqrt{\beta^2 - \beta} \)), \( J_m \) is

<table>
<thead>
<tr>
<th>Product</th>
<th>HSS</th>
<th>( \pi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERA</td>
<td>0.593</td>
<td>0.817</td>
</tr>
<tr>
<td>TRMM</td>
<td>0.493</td>
<td>0.723</td>
</tr>
<tr>
<td>SM2RAIN</td>
<td>0.324</td>
<td>0.673</td>
</tr>
<tr>
<td>Zheng et al. (2019)</td>
<td>0.587</td>
<td>0.812</td>
</tr>
<tr>
<td>CTC-M</td>
<td>0.640</td>
<td>0.826</td>
</tr>
<tr>
<td></td>
<td>0.637</td>
<td>0.826</td>
</tr>
<tr>
<td></td>
<td>0.632</td>
<td>0.825</td>
</tr>
<tr>
<td></td>
<td>0.626</td>
<td>0.824</td>
</tr>
<tr>
<td></td>
<td>0.621</td>
<td>0.823</td>
</tr>
</tbody>
</table>

**FIG. A1.** The solution of Eq. (A8). Since \( \pi_1 \) is assumed as the detection skill of the best parent product, \( \pi_1 = \pi_0 = \max \{ \pi_1, \pi_2, \pi_3 \} \).
smaller than zero. Thus, the optimal $x_m$ should be directly assigned as $x_1$, and, hence, a $n$ value that ensures $w_1 > w_2 + w_3$ should be used. Combining these two constrains and Eq. (12), yields

$$2 \left[ 2 \left( \beta - \sqrt{\beta^2 - \beta} \right) - 1 \right]^n = (2\pi - 1)^n.$$  \hspace{1cm} (A7)

Hence, the $n$ value can be solved for as

$$n = \frac{\ln(2)}{\ln(2\pi - 1) - \ln \left[ 2 \left( \beta - \sqrt{\beta^2 - \beta} \right) - 1 \right]].$$  \hspace{1cm} (A8)

The solution in Eq. (A8) shows a monotonically increasing relationship between $\pi_1$ and $n$ (Fig. A1).

Equation (A8) provides a threshold $n$ value. For cases where all the parent products are skillful [i.e., Eq. (A6) is satisfied], any $n$ value below this threshold yields optimal rain/no-rain merging. Otherwise, the optimal $n$ value should be as large as this threshold. Since CTC cannot directly solve for $\pi$, an optimal value of $n$ cannot be determined for each grid cell. However, as discussed above, a slightly overestimated $n$ leads to a conservative merging and prevents the merged product from being impacted by rain/no-rain detection errors in the low-skilled products. Therefore, we recommend that threshold $n$ value calculated in Eq. (A8) can be used for the application of conservative rain/no-rain merging. As shown in Hutchinson et al. (2009), gauge-interpolated rain/no-rain detection products tend to have 10% error. Given that reanalysis and RS precipitation products are generally less accurate than such gauge-based estimates (Table A1), $n = 1.5$ (corresponding to the threshold $n$ value at $\pi = 0.9$, see Fig. A1) appears to be a suitable absolute value for conservative merging.

As demonstrated in Table A1, $n = 1.5$ provides similar results as the optimal case ($n = 1$). Additionally, all cases considered here (with $n$ ranging from 1 to 2) substantially outperform the single case that directly uses only the best parent product (e.g., ERA). The strategy of Zheng et al. (2019), which uses the CTC-estimated optimal product at each grid cell, will therefore underperform CTC-M for all values of $n$. Interestingly, the Zheng et al. (2019) strategy also has a lower skill than ERA. This is because CTC is degraded at grid cells where the skill in one of the input products is very low (see Fig. 3).

### APPENDIX B

**Impacts of Rain/No-Rain Threshold**

As mentioned above, all results are based on a (somewhat arbitrary) daily rain/no-rain threshold of 0.5 mm day$^{-1}$. The impact of other thresholds should therefore be considered. A different threshold value has no theoretical impact on the development of CTC and CTC-M. However, the skill of ERA rain/no-rain detection does increase with increased rain/no-rain threshold (Table B1). As a result, the added value of the relatively lower-skilled products is reduced in a CTC-M analysis. Nonetheless, none of the examined variations in threshold values quantitatively changes our core conclusions (see Table B1).

### APPENDIX C

**Impact of Cross-Correlated Errors and Rain/No-Rain Biases**

The CTC-M is based on the assumption that all the rain/no-rain parent products are independent and unbiased (i.e., $A = D$).

---

Table B1. The HSS score of different precipitation products (vs E-OBS) for different rain/no-rain threshold values.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>ERA</th>
<th>SM2RAIN</th>
<th>TRMM</th>
<th>CTC-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 mm</td>
<td>0.593</td>
<td>0.324</td>
<td>0.493</td>
<td>0.632</td>
</tr>
<tr>
<td>1.0 mm</td>
<td>0.645</td>
<td>0.374</td>
<td>0.481</td>
<td>0.665</td>
</tr>
<tr>
<td>2.0 mm</td>
<td>0.651</td>
<td>0.415</td>
<td>0.475</td>
<td>0.654</td>
</tr>
</tbody>
</table>

---

Fig. C1. Impact of (a) cross-correlated detection errors and (b),(c) detection biases on the balanced accuracy $\pi$ of CTC-M. The kc values in (a) capture the detection error cross-correlation strength of the three synthetic products, with kc = 0 for independent products and kc = 1 for products with identical errors. Panels (b) and (c) capture the cases that the rain/no-rain bias (quantified by $A - D$) of three synthetic products are same and different in sign, respectively. Reported values and error bars represent the mean and standard deviation of a 100-member Monte Carlo analysis, respectively. The synthetic products are assumed to have same and different signs in rain/no-rain detection bias in (b) and (c), respectively.
However, such assumptions may be violated in reality. Therefore, we performed synthetic experiments to test the impact of cross-correlated detection errors (CCE) and detection biases on CTC-M.

To start, we randomly sampled a grid-cell within Europe, and the corresponding E-OBS rain/no-rain time series is considered as the “synthetic truth.” Based on this truth, we generated three synthetic products for CTC-M analysis. For simplicity, all three synthetic products were assumed to have a balanced accuracy ($\pi$) of 0.72—a detection skill similar to that of TRMM (Table A1). This suggests that binary errors leading to 28% rain and no-rain detection errors should be used. To consider the CCE impacts, a fixed fraction ($k_c$) of the rain/no-rain errors of the three synthetic products was assumed to contain the same binary error time series (i.e., $k_c = 0$ represents independent products and $k_c = 1$ captures products with identical errors). For each value of $k_c$, 100 Monte Carlo replicates were generated to characterize the impact of sampling uncertainty on our findings.

We also considered a second set of experiment to examine the impact of rain/no-rain detection biases, i.e., cases with $A \neq \bar{D}$. Likewise, synthetic products with $\pi = 0.72$ but $A - \bar{D}$ ranging from $-0.3$ to $0.3$ were used. For simplicity, all three synthetic products were assumed to be independent, and a 100-member Monte Carlo ensemble was used to quantify the uncertainty of our findings.

Figure C1a shows that the benefit of CTC-M decreases with increased level of detection error cross correlation between the parent products. This is because cross-correlated products contain reduced information regarding the true rain/no-rain time series.

Relative to CCE, CTC-M is generally insensitive to the presence of detection bias within the synthetic products (Figs. C1b,c). For cases where all products are simultaneously biased wet/dry, increased bias slightly reduces the performance of CTC-M (Fig. C1b). Interestingly, increased bias slightly benefits the CTC-M, when the detection biases of the synthetic products are different in sign.

REFERENCES


Ebert, E. E., and G. T. Weymouth, 1999: Incorporating satellite observations of “no rain” in an Australian daily rainfall


