

**Improving rain/no-rain detection skill by merging precipitation estimates
from different sources**

Jianzhi Dong *

USDA Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA

Wade T. Crow

USDA Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA

Rolf Reichle

*Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt,
Maryland, USA*

¹⁰ *Corresponding author: Jianzhi Dong, jianzhi.dong@usda.gov

ABSTRACT

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

24 **1. Introduction**

25 Precipitation detection skill (i.e., the ability to accurately detect precipitation occurrence) is a
26 key metric for quantifying the accuracy of precipitation products (Dinku et al. 2010; Hamada and
27 Takayabu 2016). As demonstrated in gauge-based analyses (Tong et al. 2014; Yang and Luo 2014),
28 biases in remote-sensing precipitation are often attributable to their tendency to falsely detect and/or
29 miss precipitation events. Such detection errors are unavoidably propagated into hydrological and
30 off-line land surface model simulations. For example, streamflow estimation uncertainty during
31 low-flow periods is dominated by detection errors (e.g. Bitew and Gebremichael 2011; Bitew et al.
32 2012). Therefore, false precipitation events are a key error source in modeled steamflow estimates
33 acquired from remote sensing (RS) precipitation retrievals (Tobin and Bennett 2010; Behrangi
34 et al. 2011).

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

43 Precipitation occurrence can be determined using precipitation gauge observations. One gauge-
44 based approach is to spatially interpolate binary precipitation occurrence (e.g., -1 for no-rain and
45 +1 for rain) and then apply a threshold to the interpolated values to determine rain and no-rain
46 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-Garreton 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 over-predict 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 regards to precipitation occurrence. As an analogy to commonly used maximum likelihood estimation approaches (e.g., Yilmaz et al. 2012), multi-source (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 multi-source precipitation merging for several

71 reasons. First, relative to gauge-based correction, it leverages the detection skill of all products
72 without placing excessive confidence on gauge-based observations (particularly valuable for data-
73 poor regions). Second, it does not require additional ancillary data, as required, for example,
74 during the training of machine learning approaches. Third, statistical merging approaches are
75 not impacted by hydrological modeling uncertainties that afflict rain/no-rain correction techniques
76 based on data assimilation. Finally, it has the flexibility to ingest rain/no-rain estimates from all the
77 possible sources (e.g., from both cloud temperature and data assimilation based estimates) and to
78 effectively leverage such multi-source information for improving rain/no-rain time series estimates.

79 However, the application of any statistical merging approach requires estimates of detection
80 skill for each individual precipitation product - which are difficult to obtain without access to
81 high-quality reference datasets. Based on noisy, but independent, precipitation products, triple
82 collocation (TC; Stoffelen 1998) can solve for the error variance of each product using a set of
83 linear equations (Massari et al. 2017; Alemohammad et al. 2015; Li et al. 2018; Dong et al. 2019).
84 However, TC is not directly applicable to rain/no-rain skill estimates, since rain/no-rain estimates
85 are binary with bounded errors. As a consequence, the error of a binary dataset is cross-correlated
86 with the true signal. Such truth-non-orthogonal error, in turn, violates a key assumption of TC
87 analysis (McColl et al. 2016).

88 Recently, a categorical triple collocation algorithm (CTC; McColl et al. 2016) has been proposed
89 to evaluate the classification error of a geophysical dataset without requiring access to high-quality
90 reference datasets. Specifically, the covariance of two independent geophysical products is a
91 function of their classification error. Provided three independent products are available, CTC
92 can establish a set of linear equations that can be simultaneously solved to provide estimates of
93 classification error for each product. McColl et al. (2016) demonstrates that CTC can robustly

94 estimate both temporally stationary and non-stationary classification errors – making it particularly
95 valuable for evaluating geophysical classification errors (McColl et al. 2016; Lyu et al. 2018).

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

102 **2. Data and Method**

103 *a. The CTC algorithm*

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

$$\mathbf{x}_i = \mathbf{P} + \mathbf{e}_i \quad (1)$$

106 where $\mathbf{x}_i$ and $\mathbf{P}$ are the estimated and the true rain/no-rain time series, with +1 and -1 represent
107 rain and no-rain days, respectively, and $\mathbf{e}_i$ is the classification error.

108 The performance of $\mathbf{x}_i$ in correctly classifying a geophysical variable is quantified using the
109 balanced accuracy (π_i):

$$\pi_i = 0.5(A_i + D_i) \quad (2)$$

110 where A_i and D_i is the probability of $\mathbf{x}_i$ being correct when $\mathbf{P}$ is +1 and -1, respectively (Parisi
111 et al. 2014; McColl et al. 2016). The balanced accuracy simultaneously considers the skill of a
112 product in capturing both rain and no-rain days. For instance, a constant time series of +1 (i.e.,

113 events predicted at every time step) will lead to a balanced accuracy of 0.5 (i.e., $A_i = 1$ and $D_i = 0$)
 114 that is equivalent to the skill of a random binary time series (i.e., $A_i = 0.5$ and $D_i = 0.5$). Assuming
 115 that classification errors of different products are mutually independent, the covariance (Q) of the
 116 rain/no-rain time series provided by different products can then be expressed as:

$$Q_{12} \equiv Cov(\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{P})(2\pi_1 - 1)(2\pi_2 - 1) \quad (3)$$

$$Q_{13} \equiv Cov(\mathbf{x}_1, \mathbf{x}_3) = f(\mathbf{P})(2\pi_1 - 1)(2\pi_3 - 1) \quad (4)$$

$$Q_{23} \equiv Cov(\mathbf{x}_2, \mathbf{x}_3) = f(\mathbf{P})(2\pi_2 - 1)(2\pi_3 - 1) \quad (5)$$

119 where $f(\mathbf{P})$ equals the variance of $\mathbf{P}$ for stationary cases and more complex $\mathbf{P}$ statistics for non-
 120 stationary cases - see equations (9) and (12) of McColl et al. (2016).

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

$$v_1 = \sqrt{\frac{Q_{12}Q_{13}}{Q_{23}}} \quad (6)$$

$$v_2 = \sqrt{\frac{Q_{12}Q_{23}}{Q_{13}}} \quad (7)$$

$$v_3 = \sqrt{\frac{Q_{13}Q_{23}}{Q_{12}}} \quad (8)$$

125 where $v_i = (2\pi_i - 1)\sqrt{f(\mathbf{P})}$. Therefore, the relative size of π_i for the considered products can be
 126 inferred from v_i estimated by equations (6 - 8).

127 *b. CTC-based rain/no-rain merging*

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

$$\mathbf{x}_m = \text{sign}(w_1\mathbf{x}_1 + w_2\mathbf{x}_2 + w_3\mathbf{x}_3) \quad (9)$$

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:

$$w_i \propto (2\pi_i - 1)^n \quad (10)$$

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:

$$w_1 + w_2 + w_3 = 1. \quad (11)$$

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

$$w_i = \frac{(2\pi_i - 1)^n}{(2\pi_1 - 1)^n + (2\pi_2 - 1)^n + (2\pi_3 - 1)^n}. \quad (12)$$

The goal of CTC-based merging is to determine a value of n such that the merged product $\mathbf{x}_m$ is at least as accurate as the best original product, i.e. $\pi_m \geq \max(\pi_1, \pi_2, \pi_3)$. 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 A1). However, as shown in Appendix A1, our results are qualitatively insensitive to a range of n values, and consistently outperform the best parent product. Based on the results of Appendix A1, $n = 1.5$ 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 the discussion section.

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

The EOBS (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

149 products, e.g. the Climate Prediction Center (CPC) unified gauge-based daily precipitation (Chen
150 et al. 2008), EOBS is expected to have superior rain/no-rain detection skills. As a result, we limited
151 the test of our algorithms to Europe.

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

158 The first product used here was the SM2RAIN precipitation product (Brocca et al. 2015).
159 In the SM2RAIN algorithm, precipitation is estimated by inverting a soil water balance using
160 time series of Advanced Scatterometer surface soil moisture retrievals (Wagner et al. 1999).
161 This 0.25-degree daily precipitation product is available from January 2007 to June 2015
162 (<http://hydrology.irpi.cnr.it/download-area/sm2rain-data-sets/>) and is based on the inversion ap-
163 proach described in Brocca et al. (2017).

164 The L3 daily 0.25-degree precipitation product (TRMM_3B42RT_Daily) was considered as the
165 second independent precipitation product (Huffman et al. 2007). It is retrieved from a variety
166 of low-earth orbit passive microwave observations using the Goddard Profiling Algorithm and
167 produced by averaging the near real-time, 3-hourly TRMM Multi-Satellite Precipitation Analysis
168 (TMPA) 3B42RT product without gauge-based correction. Note that the most recent Integrated
169 Multi-satellitE Retrievals for GPM product (IMERG) is the state-of-the-art RS-based precipitation
170 product (Huffman et al. 2015). However, the skill of TRMM 3B42RT is generally similar to
171 that of other RS-based precipitation products (Dong et al. 2020b). Therefore, TRMM 3B42RT is

172 more representative for testing the general performance of CTC-M in merging multiple RS-based
173 precipitation products.

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

182 As mentioned above, our CTC analysis was based on binary rain/no-rain time series, and a
183 threshold value of 0.5 mm/day was used to distinguish rain and no-rain days (Dinku et al. 2010).
184 Based on this threshold value, the four raw precipitation datasets were converted into rain/no-rain
185 categorical measurements. Results presented below are qualitatively similar for a relatively wide
186 range of other daily threshold values (0.5 to 2 mm/day) - see Appendix A2. Given the common
187 data availability of the precipitation products, data collected from 2007 to 2015 were used for the
188 real-world tests.

189 *d. Precipitation evaluation metrics*

190 Although the balanced accuracy (π) can summarize the overall accuracy of rain/no-rain detection,
191 additional precipitation detection metrics were also considered, to be consistent with previous
192 evaluation analyses. To this end, we counted the number of hits (a), false alarms (b), misses (c)
193 and correct negatives (d) (see Table 1) with respect to the EOBS gauge-based precipitation. The
194 probability of detection (POD), which represents the fraction of actual rainfall events captured by

195 a given precipitation product, is then expressed as:

$$POD_i = \frac{a_i}{a_i + c_i}. \quad (13)$$

196 Likewise, the tendency of a product to falsely report precipitation events is captured by the false-
197 alarm-ratio (FAR):

$$FAR_i = \frac{b_i}{a_i + b_i}. \quad (14)$$

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

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

200 3. Results

201 a. Real world test of CTC

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

212 Next, Figure 2a and b compare relative EOBS and ERA performances (e.g., $v_{EOBS}/v_{ERA} =$
213 $(2\pi_{EOBS} - 1)/(2\pi_{ERA} - 1)$) estimated by different CTC triplets. Relative performance values

214 above/below one plotted in Figure 2a indicate that EOBS has higher/lower detection skills than
215 ERA. Therefore, results in Figure 2 demonstrate that EOBS generally outperforms ERA in western
216 Europe and underperforms ERA in eastern and southern Europe. This spatial tendency is further
217 confirmed in EOBS-ERA-SM2RAIN triplet-based estimates (Figure 2b), and qualitatively consis-
218 tent with the gauge distribution underlying the EOBS product (Figure 2c). Nevertheless, some
219 regional differences in estimates derived from the two triplets are observed (compare Figure 2a
220 and b), e.g., in areas of Germany. Such differences may be related to the low quality of SM2RAIN
221 estimates in this region, which can destabilize CTC estimates. Nonetheless, the spatial correla-
222 tions of EOBS-ERA relative skill estimated by the two triples is 0.80 [-] – confirming the general
223 robustness of CTC.

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

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

236 The spatial correlation of EOBS- and CTC-based weight estimates range from 0.38 for ERA (due
237 to limited spatial weight variability) to 0.88 for SM2RAIN - see Figure 3. However, inconsistencies

are still observed. For instance, EOBS 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 equations (6) to (8) is approximately zero, which destabilizes CTC estimates.

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

Based on the weights shown in Figure 3d-f, merged rain/no-rain time series are derived using the ERA, TRMM and SM2RAIN data on a grid-cell by grid-cell basis using the CTC-M approach outlined in equation (9). The performance of the merged and the original products is then evaluated in Figure 4. ERA, TRMM and SM2RAIN all have very different detection skills. As shown in Figure 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% to 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 (compare the first and the fourth columns of Figure 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 (Figure 5c).

EOBS data used as a reference is, 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 (Figure 2c). Nevertheless, the relative skill of ERA and CTC-M inside and outside of Germany are qualitatively consistent (Figure 5). Therefore, while EOBS 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, Figure 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 August 3 and 16. However, it accurately reports no-rain days that are falsely reported (as rainy days) by ERA on August 9, 14, 19 and September 4 and 9 - 11. 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.

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 A2).

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

284 optimized rain/no-rain detection skill (via a novel probabilistic procedure referred to as CTC-M).
285 We further demonstrate that CTC-M outperforms all of its constituent precipitation products in
286 detecting precipitation occurrence via an analytical discussion (Appendix A1).

287 Real-world tests based on ERA, TRMM and SM2RAIN precipitation products generally confirm
288 our analytical findings. We first performed a cross-validation analysis, which shows that EOBS
289 detection skills derived from various CTC triplets are mutually consistent. We further demonstrate
290 that EOBS- and CTC-based estimates of ERA, TRMM and SM2RAIN relative skill are highly
291 consistent, except for regions where SM2RAIN has especially low skill. These results suggest the
292 general robustness of the CTC algorithm for estimating precipitation detection skill. Based on the
293 CTC estimates, multi-source rain/no-rain time series are optimally merged into a new product that
294 outperforms all of its original parent products (Figure 2). As demonstrated in Dong et al. (2019),
295 ERA data has relatively higher performance in Europe, but reduced accuracy in South America,
296 Africa and central Asia. Hence, precipitation detection accuracy should benefit significantly from
297 CTC-M in these regions.

298 As mentioned above, CTC-M is based on the assumption that the parent products are unbiased
299 and independent. These assumptions may be violated in real world applications. However, based
300 on numerical experiments, we found that CTC-M is generally insensitive to a range of rain/no-rain
301 detection biases and improves the detection skill of the parent products, even when their errors are
302 strongly cross-correlated (Appendix A3). These characteristics make CTC-M particularly useful
303 for global multi-source precipitation merging.

304 Additionally, the performance of the merged product is proportional to the skill of its parent
305 products (Appendix A1). For instance, merging three time series that all consist of random
306 (binary) noise has no benefit for rain/no-rain estimation, while merging three error-free products
307 yields a perfect rain/no-rain time series. In practice, considering rain/no-rain estimates from

308 additional precipitation products and other sources (e.g. soil moisture- and cloud-based estimates),
309 may yield more than three independent rain/no-rain estimates. For these cases, the CTC analysis
310 should be applied for all candidate rain/no-rain products and merging should be based on the three
311 most accurate and most independent parent products.

312 In CTC-M, we employed a parameter n to mitigate the adverse impact of lower-skill products
313 on the merged rain/no-rain time series. This study employs an n value of 1.5, and higher n values
314 should be used when the balanced accuracy of the best parent product exceeds 0.9. Our analysis in
315 Appendix A1 shows that the balanced accuracy of ERA is approximately 0.81 in Europe - a region
316 where ERA has exceptionally high precipitation estimation precision (Dong et al. 2019). Therefore,
317 it appears reasonable to assume the balanced accuracy of reanalyzed and remotely sensed products
318 are universally lower than 0.9 and apply $n = 1.5$ for rain/no-rain merging at the global scale.

319 Relative to existing rain/no-rain detection approaches (e.g., Ebert and Weymouth 1999; Nasrollahi
320 et al. 2013; Cui et al. 2015; Ghajarnia et al. 2016), this study provides a more general framework
321 for improving the precipitation detection skill. First, rain/no-rain estimates obtained from all of
322 the above-mentioned methods can be used as independent CTC-M inputs for improved detection
323 accuracy. Additionally, CTC-M does not require high-quality training data, which makes it
324 applicable for rain/no-rain merging at global scales. Consequently, CTC-M can be employed within
325 other existing multi-source precipitation merging frameworks. For instance, the MSWEP version
326 2 product uses gauge-based precipitation observations for removing falsely reported precipitation
327 events in ERA (Beck et al. 2019). However, as mentioned above, even in densely instrumented
328 Europe, rain gauges are often insufficient for capturing grid-cell scale precipitation. Consequently,
329 adjusting all products to match gauge-based rain/no-rain discrimination is risky - particularly
330 for data-poor regions (e.g., Africa and high latitude regions of the Northern Hemisphere) with
331 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.

APPENDIX

A1. 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 $\mathbf{x}_1$ outperforms $\mathbf{x}_2$ and $\mathbf{x}_3$, i.e., $\pi_1 = \pi_b = \max\{\pi_1, \pi_2, \pi_3\}$. Additionally, we also assume all the products are unbiased (i.e., $A_i = D_i$) for simplicity. Under these assumptions, π_i directly represents the probability of $\mathbf{x}_i$ 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 π value. In this case, $\mathbf{x}_m$ is correct when at least two products are correct. Therefore, the balanced accuracy of $\mathbf{x}_m$ can be estimated as:

$$\pi_m = \pi_1\pi_2\pi_3 + \pi_1\pi_2(1 - \pi_3) + \pi_1(1 - \pi_2)\pi_3 + (1 - \pi_1)\pi_2\pi_3. \quad (\text{A1})$$

The detection skill differences of $\mathbf{x}_m$ and $\mathbf{x}_1$ (the best original product) is calculated as:

$$J \equiv \pi_m - \pi_1 = \pi_1\pi_2\pi_3 + \pi_1\pi_2(1 - \pi_3) + \pi_1(1 - \pi_2)\pi_3 + (1 - \pi_1)\pi_2\pi_3 - \pi_1 \quad (\text{A2})$$

where J is the difference between π_m and π_1 , and $J > 0$ means $\mathbf{x}_m$ outperforms $\mathbf{x}_1$. Taking the partial derivative of J with respect to π_2 yields:

$$\frac{\partial J}{\partial \pi_2} = \pi_1(1 - \pi_3) + \pi_3(1 - \pi_1) \quad (\text{A3})$$

As noted above, all the π values are within the range of 0.5 to 1. Therefore, equation (A3) is positive. Analogously, the partial derivative of J with respect of π_3 is also positive. Hence, J is a monotonically increasing function of both π_2 and π_3 . Here, we replace π_2 and π_3 values with y , where $y = \min\{\pi_2, \pi_3\}$, in equation (A2). This substitution yields:

$$J_m = (1 - 2\pi_1)y^2 + 2\pi_1y - \pi_1 \quad (\text{A4})$$

where J_m is the J value when both π_2 and π_3 equal y . As demonstrated in equation (A3), J is a monotonic function of both π_2 and π_3 . Therefore, $J \geq J_m$ – given π_2 (or π_3) is replaced by a smaller value y in equation (A2). Hence, a non-negative J_m guarantees a non-negative J , i.e., the skill of $\mathbf{x}_m$ is at least good as that of the original products.

According to equation (A4), $J_m \geq 0$ requires:

$$\beta - \sqrt{\beta^2 - \beta} \leq y \leq \beta + \sqrt{\beta^2 - \beta} \quad (\text{A5})$$

where $\beta = \pi_1/(2\pi_1 - 1)$. Since π_1 is in the range of $[0.5, 1]$, β is a positive number greater than 1.

Hence, equation (A5) can be simplified as:

$$\beta - \sqrt{\beta^2 - \beta} \leq y \leq 1 \quad (\text{A6})$$

According to equation (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 $\mathbf{x}_m$. However, if both $\mathbf{x}_2$ and $\mathbf{x}_3$ are low-skill (i.e., $y < \beta - \sqrt{\beta^2 - \beta}$), J_m is smaller than zero. Thus, the optimal $\mathbf{x}_m$ should be directly assigned as $\mathbf{x}_1$, and, hence, a n value that ensures $w_1 > w_2 + w_3$ should be used. Combining these two constrains and equation (12), yields:

$$2 \left[2 \left(\beta - \sqrt{\beta^2 - \beta} \right) - 1 \right]^n = (2\pi_1 - 1)^n. \quad (\text{A7})$$

366 Hence, the n value can be solve for as:

$$n = \frac{\ln(2)}{\ln(2\pi_1 - 1) - \ln\left[2\left(\beta - \sqrt{\beta^2 - \beta}\right) - 1\right]} \quad (\text{A8})$$

367 The solution in equation (A8) shows a monotonically increasing relationship between π_1 and n
368 (Figure A1).

369 Equation (A8) provides a threshold n value. For cases where all the parent products are skillful
370 (i.e., equation (A6) is satisfied), any n value below this threshold yields optimal rain/no-rain
371 merging. Otherwise, the optimal n value should be as large as this threshold. Since CTC cannot
372 directly solve for π , an optimal value of n cannot be determined for each grid cell. However,
373 as discussed above, a slightly overestimated n leads to a conservative merging and prevents the
374 merged product from being impacted by rain/no-rain detection errors in the low-skilled products.
375 Therefore, we recommend that threshold n value calculated in equation (A8) can be used for the
376 application of conservative rain/no-rain merging. As shown in Hutchinson et al. (2009), gauge-
377 interpolated rain/no-rain detection products tend to have 10% error. Given that reanalysis and
378 RS precipitation products are generally less accurate than such gauge-based estimates (Table A1),
379 $n = 1.5$ (corresponding to the threshold n value at $\pi = 0.9$, see Figure A1) appears to be a suitable
380 absolute value for conservative merging.

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

A2. 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. 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 A2). 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 A2).

A3. 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. $A = D$). 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 EOBS 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 (π) 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 (kc) of the rain/no-rain errors of the three synthetic products was assumed to contain the same binary error time series (i.e. $kc = 0$ represents independent products and $kc = 1$ captures products with identical errors). For each value of kc , 100 Monte Carlo replicates were generated to characterize the impact of sampling uncertainty on our conclusions.

We also considered a second set of experiment to examine the impact of rain/no-rain detection biases, i.e. cases with $A \neq D$. Likewise, synthetic products with $\pi = 0.72$ but $A - D$ 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 A2a 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 (Figure A2b and c). For cases where all products are simultaneously biased wet/dry, increased bias slightly reduces the performance of CTC-M (Figure A2b). Interestingly, increased bias slightly benefits the CTC-M, when the detection biases of the synthetic products are different in sign.

Acknowledgments. This paper was partially supported by the NASA SMAP mission through award NNH12ZDA001N-SMAP. 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 EOBS data are available on: <https://www.ecad.eu/download/ensembles/download.php>.

References

Alemohammad, S., K. McColl, A. Konings, D. Entekhabi, and A. Stoffelen, 2015: Characterization of precipitation product errors across the us using multiplicative triple collocation. *Hydrol. Earth Syst. Sci. Discuss*, **12** (2), 2527–2559.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, C.-H. Su, and D. R. Robertson, 2016:
Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce
catchments. *Water Resources Research*, **52** (7), 5357–5375.

Baez-Villanueva, O. M., and Coauthors, 2020: Rf-mep: A novel random forest method for
merging gridded precipitation products and ground-based measurements. *Remote Sensing of
Environment*, **239**, 111 606.

Beck, H. E., E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. Van Dijk, T. R. McVicar, and
R. F. Adler, 2019: Mswep v2 global 3-hourly 0.1 precipitation: methodology and quantitative
assessment. *Bulletin of the American Meteorological Society*, **100** (3), 473–500.

Behrangi, A., B. Khakbaz, T. C. Jaw, A. AghaKouchak, K. Hsu, and S. Sorooshian, 2011: Hydro-
logic evaluation of satellite precipitation products over a mid-size basin. *Journal of Hydrology*,
397 (3-4), 225–237.

Bhuiyan, E. M. A., E. I. Nikolopoulos, and E. N. Anagnostou, 2019: Machine learning–based
blending of satellite and reanalysis precipitation datasets: A multiregional tropical complex
terrain evaluation. *Journal of Hydrometeorology*, **20** (11), 2147–2161.

Bitew, M. M., and M. Gebremichael, 2011: Evaluation of satellite rainfall products through
hydrologic simulation in a fully distributed hydrologic model. *Water Resources Research*, **47** (6).

Bitew, M. M., M. Gebremichael, L. T. Ghebremichael, and Y. A. Bayissa, 2012: Evaluation
of high-resolution satellite rainfall products through streamflow simulation in a hydrological
modeling of a small mountainous watershed in ethiopia. *Journal of Hydrometeorology*, **13** (1),
338–350.

454 Brocca, L., and Coauthors, 2015: Rainfall estimation from in situ soil moisture observations
 455 at several sites in europe: an evaluation of the sm2rain algorithm. *Journal of Hydrology and*
 456 *Hydromechanics*, **63 (3)**, 201–209.

457 Brocca, L., and Coauthors, 2017: A review of the applications of ascats soil moisture products.
 458 *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10 (5)**,
 459 2285–2306.

460 Chen, M., W. Shi, P. Xie, V. B. Silva, V. E. Kousky, R. Wayne Higgins, and J. E. Janowiak, 2008:
 461 Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal*
 462 *of Geophysical Research: Atmospheres*, **113 (D4)**.

463 Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones, 2018: An ensemble
 464 version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research:*
 465 *Atmospheres*, **123 (17)**, 9391–9409.

466 Crow, W., and D. Ryu, 2009: A new data assimilation approach for improving runoff prediction
 467 using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences*, **13 (1)**,
 468 1–16.

469 Cui, Y., J. Shi, and Z. Wang, 2015: Multi-state adaptive bit false alarm reduction under degradation
 470 process. *IEEE Transactions on Instrumentation and Measurement*, **64 (3)**, 671–682.

471 Dee, D. P., and Coauthors, 2011: The era-interim reanalysis: Configuration and performance of
 472 the data assimilation system. *Quarterly Journal of the royal meteorological society*, **137 (656)**,
 473 553–597.

474 Dinku, T., P. Ceccato, K. Cressman, and S. J. Connor, 2010: Evaluating detection skills of satellite
 475 rainfall estimates over desert locust recession regions. *Journal of Applied Meteorology and*
 476 *Climatology*, **49** (6), 1322–1332.

477 Dong, J., W. T. Crow, Z. Duan, L. Wei, and Y. Lu, 2019: A double instrumental variable method
 478 for geophysical product error estimation. *Remote Sensing of Environment*, **225**, 217–228.

479 Dong, J., F. Lei, and L. Wei, 2020a: Triple collocation based multi-source precipitation merging.
 480 *Front. Water 2: 1. doi: 10.3389/frwa*.

481 Dong, J., L. Wei, X. Chen, Z. Duan, and Y. Lu, 2020b: An instrument variable based algorithm
 482 for estimating cross-correlated hydrological remote sensing errors. *Journal of Hydrology*, **581**,
 483 124413.

484 Ebert, E. E., and G. T. Weymouth, 1999: Incorporating satellite observations of rain
 485 in an australian daily rainfall analysis. *Journal of Applied Meteorology*, **38** (1), 44–56.

486 Ghajarnia, N., P. D. Arasteh, S. Araghinejad, and M. A. Liaghat, 2016: A hybrid bayesian-svd
 487 based method to detect false alarms in persiann precipitation estimation product using related
 488 physical parameters. *Journal of Hydrology*, **538**, 640–650.

489 Hamada, A., and Y. N. Takayabu, 2016: Improvements in detection of light precipitation with
 490 the global precipitation measurement dual-frequency precipitation radar (gpm dpr). *Journal of*
 491 *Atmospheric and Oceanic Technology*, **33** (4), 653–667.

492 Huffman, G. J., D. T. Bolvin, E. J. Nelkin, and Coauthors, 2015: Integrated multi-satellite retrievals
 493 for gpm (imerg) technical documentation. *NASA/GSFC Code*, **612** (2015), 47.

- 494 Huffman, G. J., and Coauthors, 2007: The trmm multisatellite precipitation analysis (tmpa):
495 Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of hy-*
496 *drometeorology*, **8** (1), 38–55.
- 497 Hutchinson, M. F., D. W. McKenney, K. Lawrence, J. H. Pedlar, R. F. Hopkinson, E. Milewska,
498 and P. Papadopol, 2009: Development and testing of canada-wide interpolated spatial models
499 of daily minimum–maximum temperature and precipitation for 1961–2003. *Journal of Applied*
500 *Meteorology and Climatology*, **48** (4), 725–741.
- 501 Li, C., G. Tang, and Y. Hong, 2018: Cross-evaluation of ground-based, multi-satellite and reanalysis
502 precipitation products: Applicability of the triple collocation method across mainland china.
503 *Journal of hydrology*, **562**, 71–83.
- 504 Lyu, H., and Coauthors, 2018: Validation of the smap freeze/thaw product using categorical triple
505 collocation. *Remote Sensing of Environment*, **205**, 329–337.
- 506 Massari, C., W. Crow, and L. Brocca, 2017: An assessment of the performance of global rainfall
507 estimates without ground-based observations. *Hydrology and Earth System Sciences*, **21** (9),
508 4347–4361.
- 509 McColl, K. A., A. Roy, C. Derksen, A. G. Konings, S. H. Alemohammed, and D. Entekhabi, 2016:
510 Triple collocation for binary and categorical variables: Application to validating landscape
511 freeze/thaw retrievals. *Remote Sensing of Environment*, **176**, 31–42.
- 512 Nasrollahi, N., K. Hsu, and S. Sorooshian, 2013: An artificial neural network model to reduce
513 false alarms in satellite precipitation products using modis and cloudsat observations. *Journal*
514 *of Hydrometeorology*, **14** (6), 1872–1883.

515 Parisi, F., F. Strino, B. Nadler, and Y. Kluger, 2014: Ranking and combining multiple predictors
516 without labeled data. *Proceedings of the National Academy of Sciences*, **111** (4), 1253–1258.

517 Reichle, R. H., and Coauthors, 2017: Global assessment of the smap level-4 surface and root-zone
518 soil moisture product using assimilation diagnostics. *Journal of hydrometeorology*, **18** (12),
519 3217–3237.

520 Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration
521 using triple collocation. *Journal of Geophysical Research: Oceans*, **103** (C4), 7755–7766.

522 Tobin, K. J., and M. E. Bennett, 2010: Adjusting satellite precipitation data to facilitate hydrologic
523 modeling. *Journal of Hydrometeorology*, **11** (4), 966–978.

524 Tong, K., F. Su, D. Yang, and Z. Hao, 2014: Evaluation of satellite precipitation retrievals and
525 their potential utilities in hydrologic modeling over the tibetan plateau. *Journal of Hydrology*,
526 **519**, 423–437.

527 Wagner, W., G. Lemoine, and H. Rott, 1999: A method for estimating soil moisture from ers
528 scatterometer and soil data. *Remote sensing of environment*, **70** (2), 191–207.

529 Yang, Y., and Y. Luo, 2014: Evaluating the performance of remote sensing precipitation products
530 cmorph, persiann, and tmpa, in the arid region of northwest china. *Theoretical and applied*
531 *climatology*, **118** (3), 429–445.

532 Yang, Z., K. Hsu, S. Sorooshian, X. Xu, D. Braithwaite, Y. Zhang, and K. M. Verbist, 2017: Merging
533 high-resolution satellite-based precipitation fields and point-scale rain gauge measurements—A
534 case study in chile. *Journal of Geophysical Research: Atmospheres*, **122** (10), 5267–5284.

535 Yilmaz, M., W. Crow, M. Anderson, and C. Hain, 2012: An objective methodology for merging
536 satellite-and model-based soil moisture products. *Water Resources Research*, **48** (11).

537 Zheng, L., C. Zhou, and Q. Liang, 2019: Variations in antarctic peninsula snow liquid water
538 during 1999–2017 revealed by merging radiometer, scatterometer and model estimations. *Remote*
539 *Sensing of Environment*, **232**, 111 219.

540	LIST OF TABLES	
541	Table 1. Contingency table for comparing the rain/no-rain detected by gauge and prod-	
542	ucts. Rain/no-rain threshold of 0.5 mm/day is used.	29
543	Table A1. The HSS score and balanced accuracy (π) of different precipitation products	
544	(using EOBS as the rain/no-rain) reference.	30
545	Table A2. The HSS score of different precipitation products (versus EOBS) for different	
546	rain/no-rain threshold values.	31

547 TABLE 1. Contingency table for comparing the rain/no-rain detected by gauge and products. Rain/no-rain
548 threshold of 0.5 mm/day is used.

Threshold	Gauge $\geq$ threshold	Gauge $<$ threshold
Product $\geq$ threshold	a (hits)	b (false alarms)
Product $<$ threshold	c (misses)	d (correct negatives)

549 Table A1. The HSS score and balanced accuracy (π) of different precipitation products (using EOBS as the
550 rain/no-rain) reference.

Product		HSS	π
ERA		0.593	0.817
TRMM		0.493	0.723
SM2RAIN		0.324	0.673
Zheng et al. (2019)		0.587	0.812
CTC-M	n = 1.00	0.640	0.826
	n = 1.25	0.637	0.826
	n = 1.50	0.632	0.825
	n = 1.75	0.626	0.824
	n = 2.00	0.621	0.823

552 Table A2. The HSS score of different precipitation products (versus EOBS) for different rain/no-rain threshold
553 values.

Threshold	ERA	SM2RAIN	TRMM	CTC-M
0.5 mm	0.593	0.324	0.493	0.632
1.0 mm	0.645	0.374	0.481	0.665
2.0 mm	0.651	0.415	0.475	0.654

554

LIST OF FIGURES

Fig. 1.	Relative detection skill of EOBS ($v_{EOBS} = (2\pi_{EOBS} - 1)\sqrt{f(P)}$) estimated from different triplets. Grey shading represents areas that lack adequate data for CTC analysis.	33
Fig. 2.	Relative performances of EOBS and ERA (i.e., $(2\pi_{EOBS} - 1)/(2\pi_{ERA} - 1)$) estimated using CTC triplets (a) EOBS-ERA-TRMM and (b) EOBS-ERA-SM2RAIN. Red shading in (a) and (b) indicate that the precipitation detection skill of EOBS exceeds that of ERA. (c) Number of gauges per 0.25 spatial grid-cell used in EOBS. Gauges that stopped providing data before 2007 are not shown. The black cross in (c) denotes the locations used for the rain/no-rain time series comparison in Figure 6.	34
Fig. 3.	Relative weights of ERA, TRMM and SM2RAIN products estimated using EOBS (first row) and CTC with no ground observations (second row). Grey shading represents grid cells that lack adequate temporal coverage for CTC analysis or where EOBS underperforms any of the three reanalysis and remote-sensing precipitation products.	35
Fig. 4.	Performance of ERA (first column), TRMM (second column), SM2RAIN (third column) and CTC-M (fourth column) with regards to precipitation detection skill for (a-d) POD, (e-h) FAR, and (i-l) HSS. Dark gray shading indicates grid cells where CTC cannot be performed due to lack of data, or where EOBS does not outperform the other three products.	36
Fig. 5.	Relative performance of the merged product versus 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.	37
Fig. 6.	Comparison of EOBS, ERA and the merged rain/no-rain time series for a grid cell near Nuremberg in southern Germany (see Figure 2c for location) during August-September 2010.	38
Fig. A1.	The solution of equation (A8). Since π_1 is assumed as the detection skill of the best parent product, $\pi_1 = \pi_b = \max\{\pi_1, \pi_2, \pi_3\}$.	39
Fig. A2.	Impact of cross-correlated detection errors (CEE, a) and detection biases (b and c) on the balanced accuracy (π) of CTC-M. The kc values in (a) captures 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. (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.	40

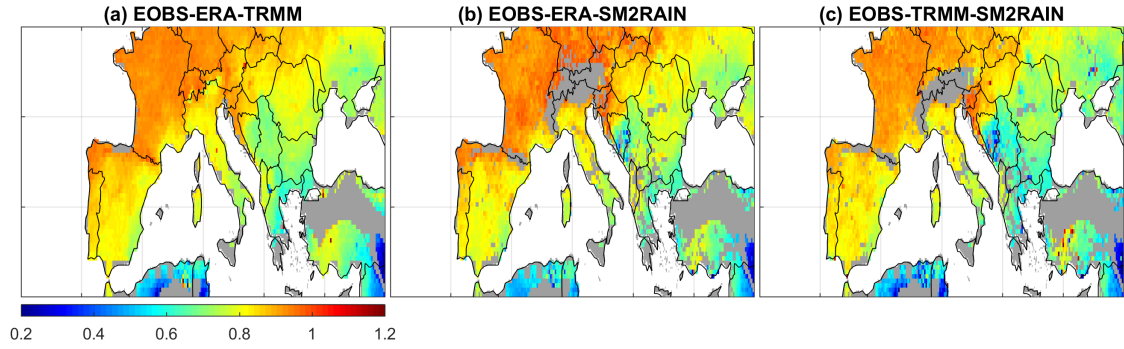


FIG. 1. Relative detection skill of EOBS ($v_{EOBS} = (2\pi_{EOBS} - 1)\sqrt{f(P)}$) estimated from different triplets.
 Grey shading represents areas that lack adequate data for CTC analysis.

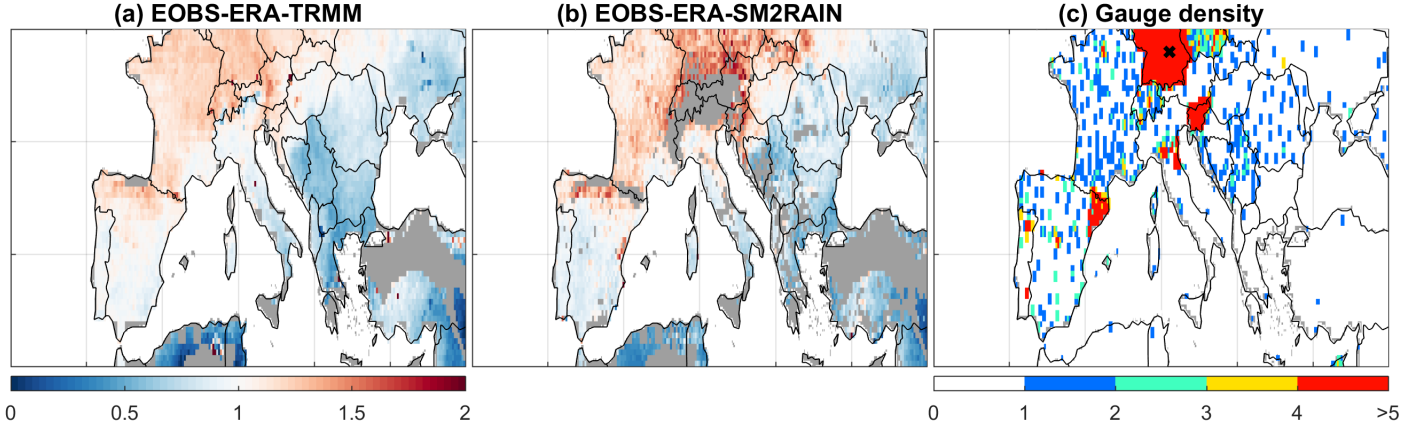


FIG. 2. Relative performances of EOBS and ERA (i.e., $(2\pi_{EOBS} - 1)/(2\pi_{ERA} - 1)$) estimated using CTC triplets (a) EOBS-ERA-TRMM and (b) EOBS-ERA-SM2RAIN. Red shading in (a) and (b) indicate that the precipitation detection skill of EOBS exceeds that of ERA. (c) Number of gauges per 0.25 spatial grid-cell used in EOBS. Gauges that stopped providing data before 2007 are not shown. The black cross in (c) denotes the locations used for the rain/no-rain time series comparison in Figure 6.

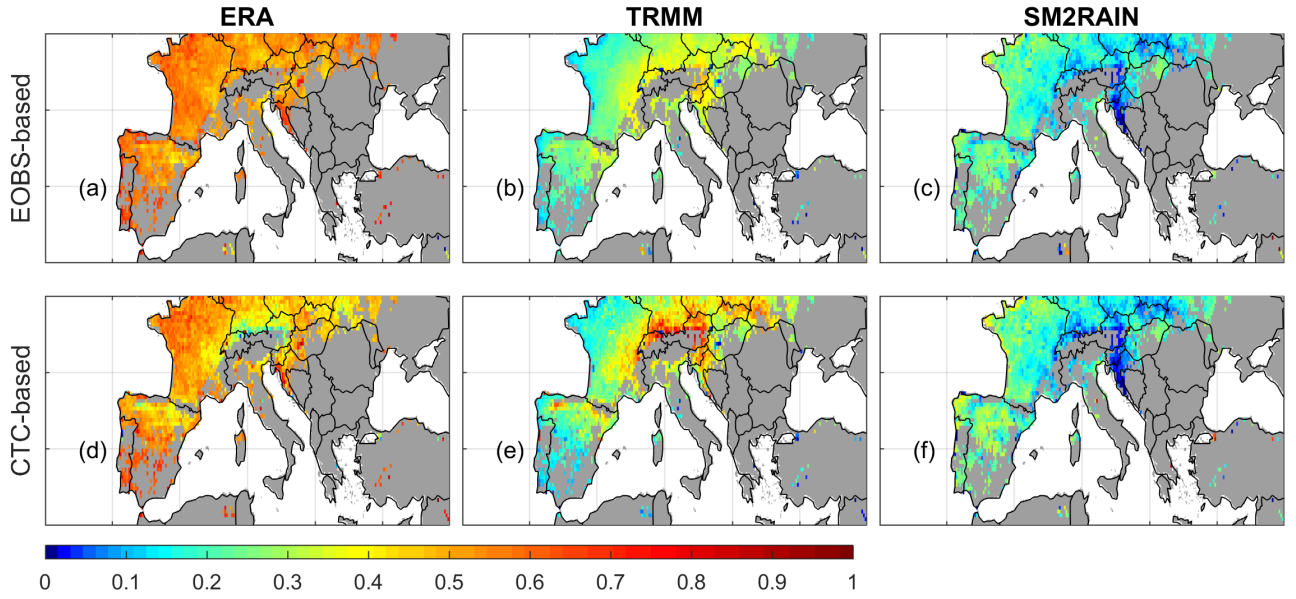


FIG. 3. Relative weights of ERA, TRMM and SM2RAIN products estimated using EOBS (first row) and CTC with no ground observations (second row). Grey shading represents grid cells that lack adequate temporal coverage for CTC analysis or where EOBS underperforms any of the three reanalysis and remote-sensing precipitation products.

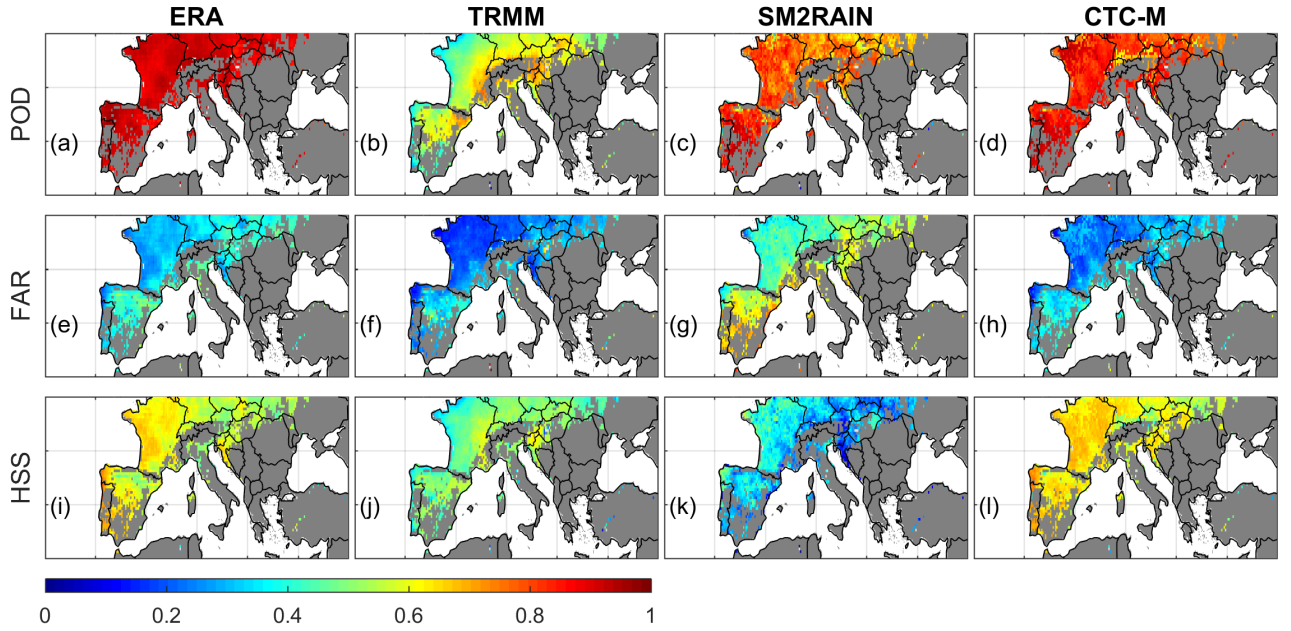


FIG. 4. Performance of ERA (first column), TRMM (second column), SM2RAIN (third column) and CTC-M (fourth column) with regards to precipitation detection skill for (a-d) POD, (e-h) FAR, and (i-l) HSS. Dark gray shading indicates grid cells where CTC cannot be performed due to lack of data, or where EOBS does not outperform the other three products.

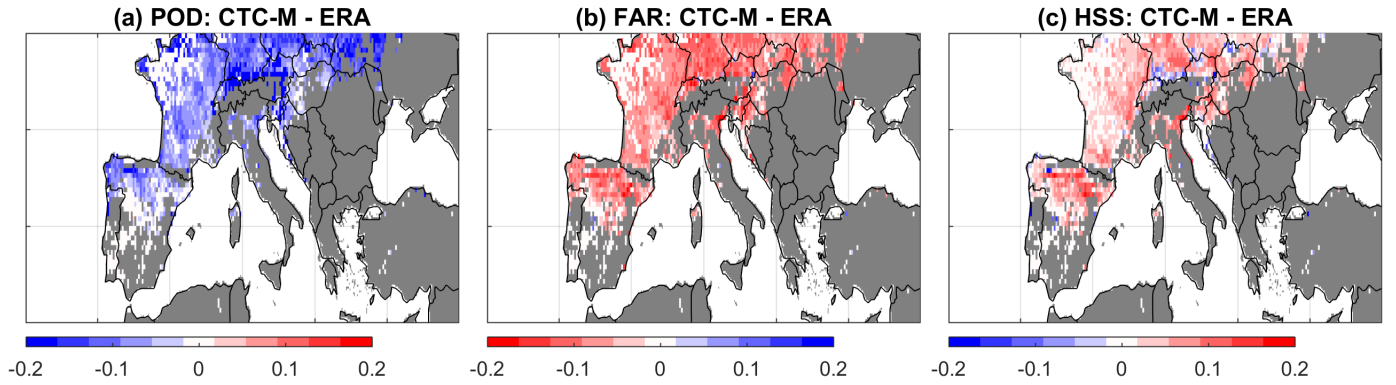


FIG. 5. Relative performance of the merged product versus the ERA product (MERGED minus ERA) with respect to the (a) POD, (b) FAR and (c) HSS categorical performance metrics. Red (blue) colors indicate that CTC-M is better (worse) than ERA.

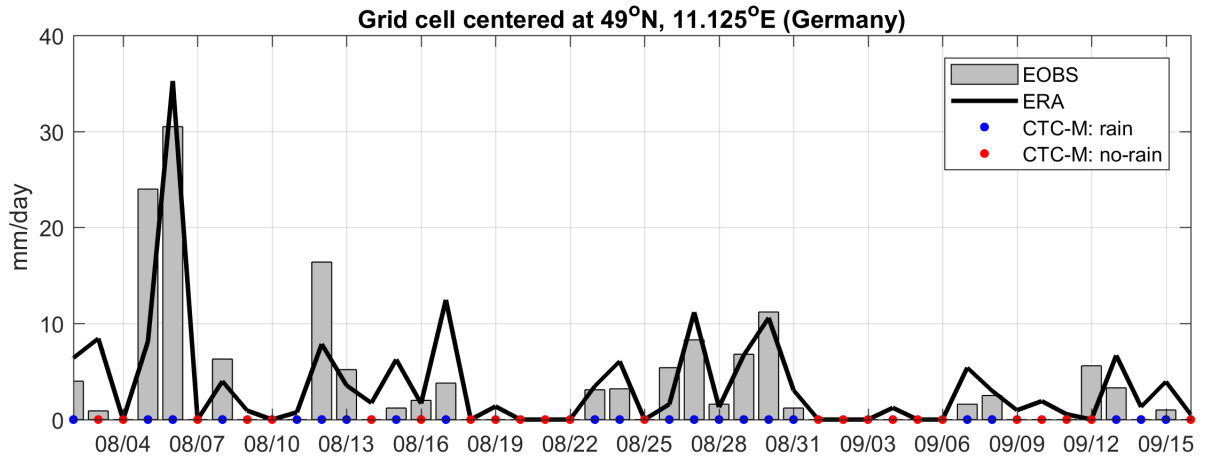


FIG. 6. Comparison of EOBS, ERA and the merged rain/no-rain time series for a grid cell near Nuremberg in
in southern Germany (see Figure 2c for location) during August-September 2010.

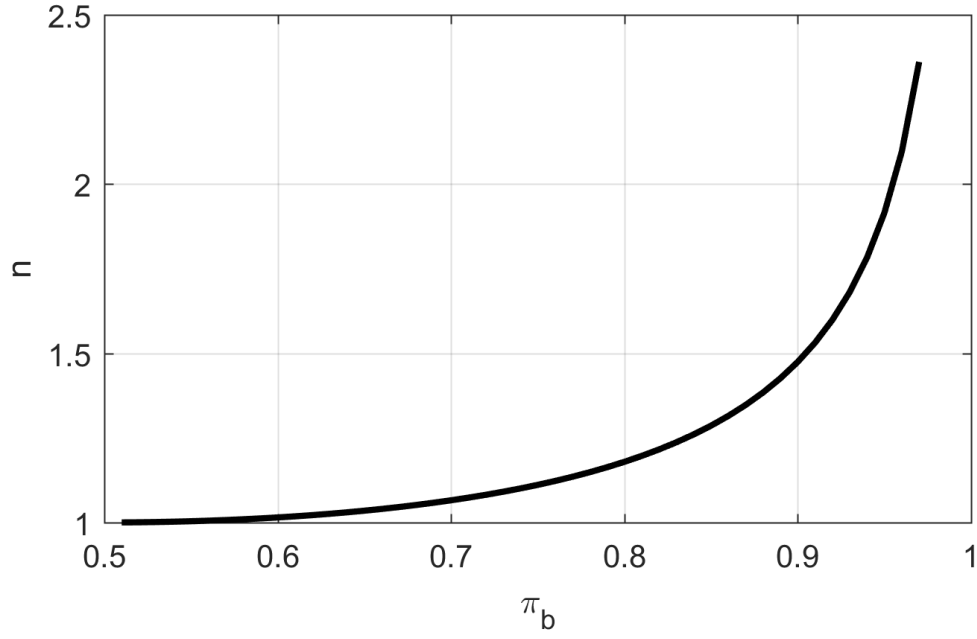


Fig. A1. The solution of equation (A8). Since π_1 is assumed as the detection skill of the best parent product,
 $\pi_1 = \pi_b = \max\{\pi_1, \pi_2, \pi_3\}$.

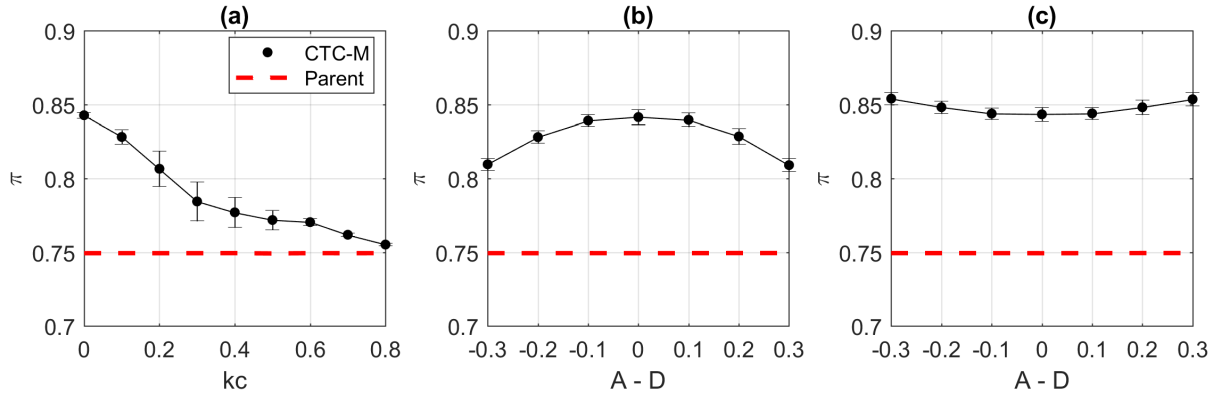


Fig. A2. Impact of cross-correlated detection errors (CEE, a) and detection biases (b and c) on the balanced accuracy (π) of CTC-M. The kc values in (a) captures the detection error cross-correlation strength of the three synthetic products, with $kc = 0$ for independent products and $kc = 1$ for products with identical errors. (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.