Characterizing Precipitation Product Errors across the United States using Triple Collocation

Seyed Hamed Alemohammad [1*], Kaighin A. McColl [1], Alexandra G. Konings [1], Dara Entekhabi [1], Ad Stoffelen [2]

[1] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology

[2] Koninklijk Nederlands Meteorologisch Instituut (KNMI), R&D Satellite Observations, De Bilt, the Netherlands

*hamed_al@mit.edu

Introduction

- Validation of precipitation estimates from various products is a challenging problem, since the true precipitation is unknown.
- In this study, triplets among NEXRAD-IV, TRMM 3B42, GPCP and GPI products are used to quantify the associated spatial error characteristics across a central part of the continental US using the Triple Collocation (TC) technique.
- A multiplicative (logarithmic) error model is incorporated in the original TC formulation to relate the precipitation estimates to the unknown truth.
- This study provides error estimates of the precipitation products that can be incorporated into hydrological and meteorological models, especially those used in data assimilation.
- The methodology presented in this study could be used to quantify the uncertainties associated with precipitation estimates from each of the constellation of GPM satellites. Such quantification is prerequisite to optimally merging these estimates.

Multiplicative Triple Collocation

TC takes advantage of three collocated data products of the same variable and estimates the Root-Mean-Square-Error (RMSE) of each, without requiring knowledge of the perfect (zero error) truth. TC assumes *zero error cross covariance between different products* and *zero covariance between the errors and truth*.

Denoting measurement i as R_i, based on multiplicative error model we have:

$$R_i = a_i\, T^{\beta_i}\, e^{\varepsilon_i}$$

Log-transforming this equations results in the classical TC formulation:

$$r_i = \alpha_i + \beta_i T + \varepsilon_i$$

Denoting C_{ij} as the $(i,j)^{th}$ element of the sample covariance matrix, TC gives:

$$\sigma_{r_1}^2 = C_{11} - \frac{C_{12}C_{13}}{C_{23}} \qquad \rho_{t,1}^2 = \frac{C_{12}C_{13}}{C_{11}C_{23}}$$

$$\sigma_{r_2}^2 = C_{22} - \frac{C_{12}C_{23}}{C_{13}} \qquad \rho_{t,2}^2 = \frac{C_{12}C_{23}}{C_{22}C_{13}}$$

$$\sigma_{r_3}^2 = C_{33} - \frac{C_{13}C_{23}}{C_{12}} \qquad \rho_{t,3}^2 = \frac{C_{13}C_{23}}{C_{33}C_{12}}$$

σ_{r_i} is the RMSE of product r_i and $\rho_{t,i}^2$ is the correlation coefficient between the underlying truth and product i.

Study Domain

- The study domain ranges from 30° to 40°N latitudes and 110° to 80°W longitudes. This region is selected to maximize the overlapping spatial coverage between the data sets that are used here.
- The temporal domain for this study is from Jan. 2002 until Apr. 2014, and the data from each product is temporally aggregated to biweekly values (to generate i.i.d. samples).

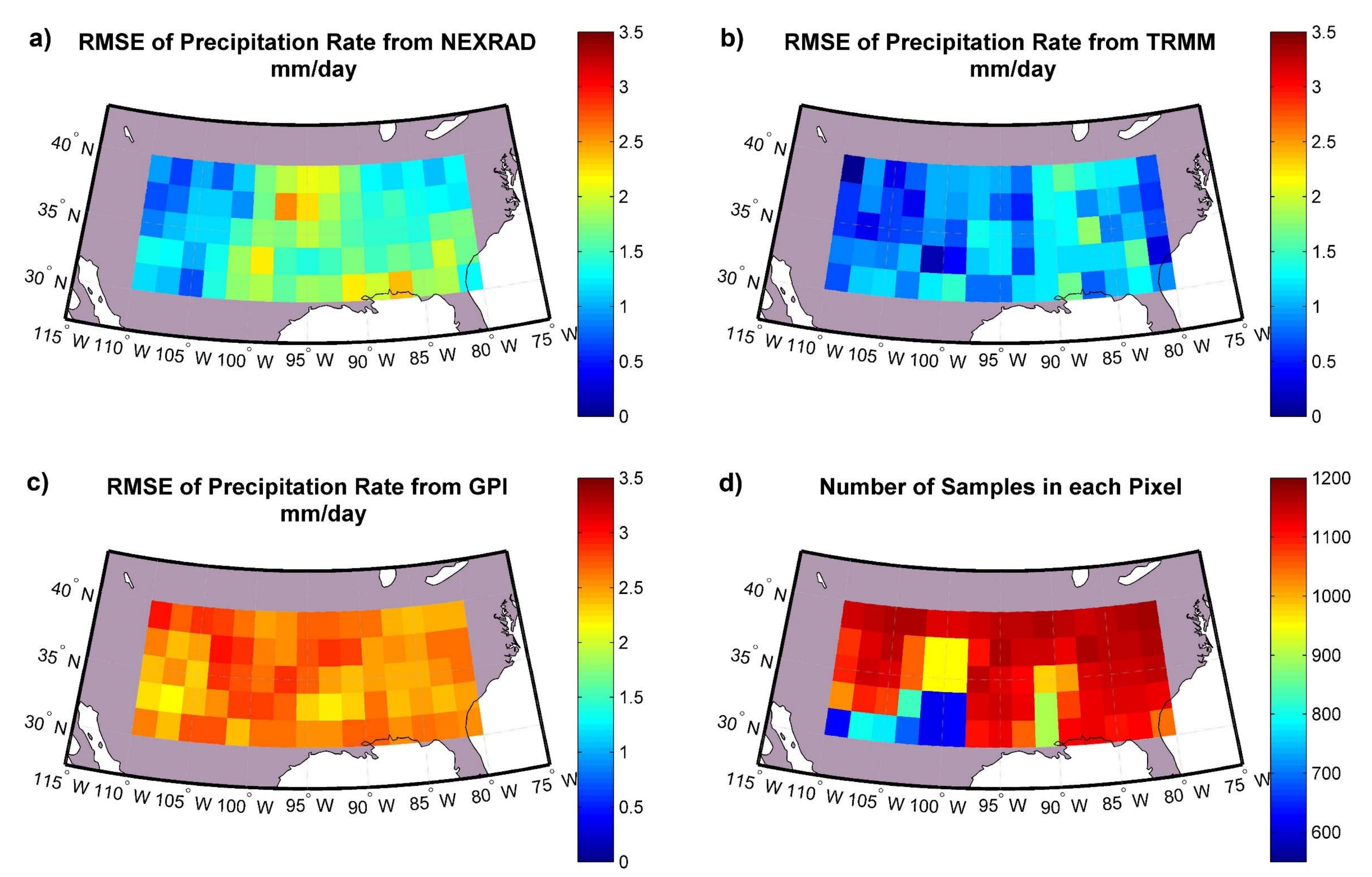


Figure 1: RMSE of the precipitation rate for products in group1; a)NEXRAD, b)TRMM, c)GPI. Panel d) shows the number of data points in each pixel.

Results

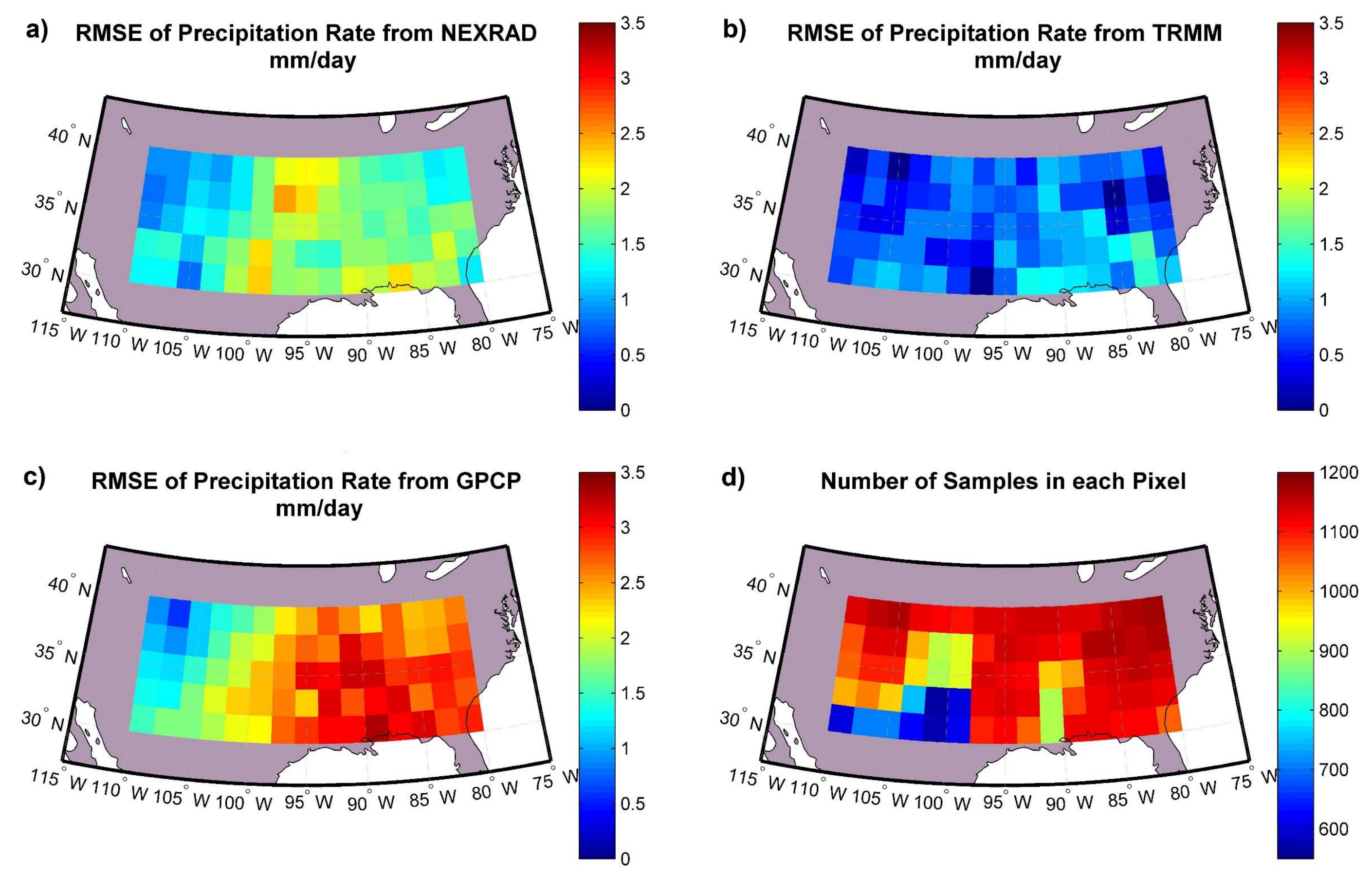


Figure 2: RMSE of the precipitation rate for products in group2; a)NEXRAD, b)TRMM, c)GPCP. Panel d) shows the number of data points in each pixel.

- RMSE and ρ^2 estimates for NEXRAD and TRMM are very similar in both groups. (Indicating that results are not dependent on the choice of triplets.)
- RMSE of the TRMM product in both of the triplets is small compared to the other two products and is also relatively small compared to the climatological mean. NEXRAD has higher RMSE compared to TRMM, but is considerably smaller than GPCP or GPI.
- The RMSE in TRMM, NEXRAD and GPCP products increases east to west. Comparing this pattern with the climatology maps, it is clear that in regions with higher mean precipitation rate, the RMSE is higher.
- Among the products analyzed here, the TRMM product has the highest correlation coefficient with the truth in almost all of the pixels. NEXRAD also has high correlation with the truth but there is a pattern that pixels toward the east of the region have higher correlation coefficients in the NEXRAD product. GPCP has less correlation with the truth (w/ similar east-west pattern). GPI exhibits very low correlation coefficients toward the west of the region.

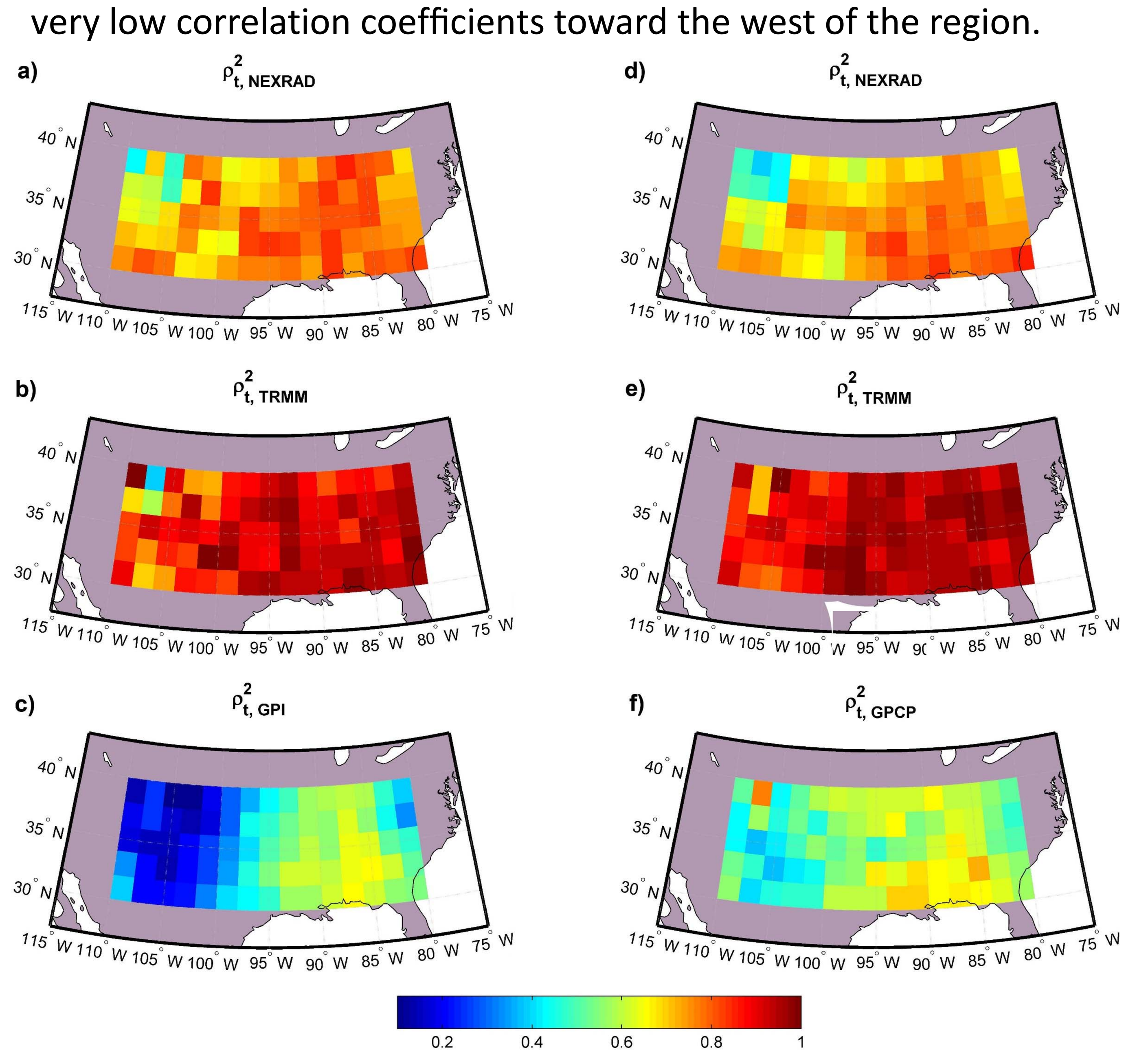


Figure 3: Correlation coefficient between the truth and each precipitation product. The left column shows the results for triplets in group 1, and the right column shows the results for triplets in group 2.

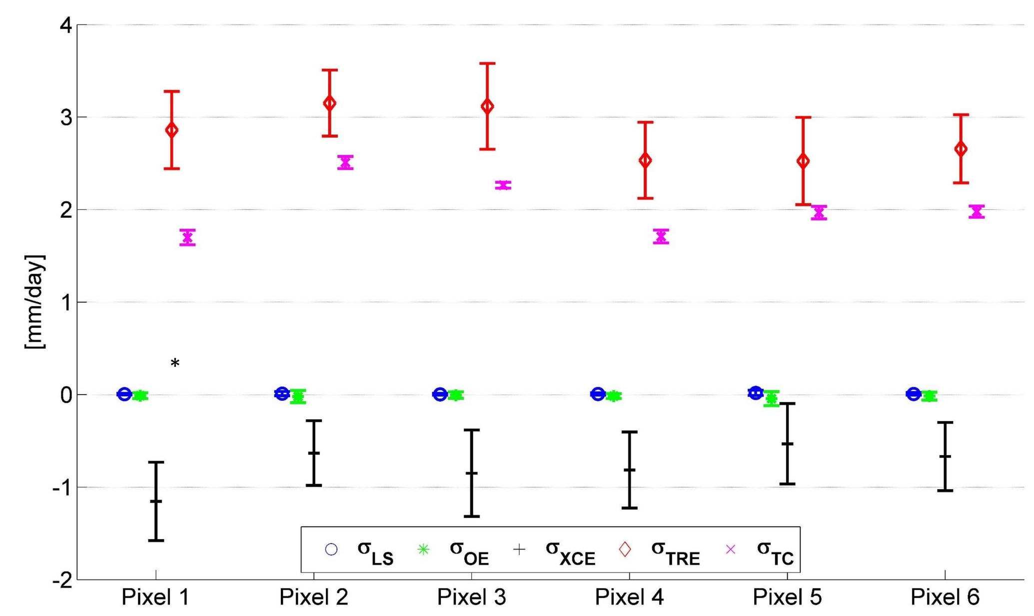
Error Decomposition

- The error estimates from TC analysis can be decomposed to four elements:

$$\sigma_{TC}^2 = \sigma_{TRE}^2 + \sigma_{LS}^2 + \sigma_{OE}^2 + \sigma_{XCE}^2$$

- The bias terms are caused due to violation of assumptions in TC. Here, we use in-situ gauge data as proxy fro the underlying truth to evaluate these bias terms.
- Six pixels over the state of Oklahoma were selected (dense in-situ gauge coverage) for this analysis. Figure 4 shows the contribution of each element to the total RMSE derived from TC.
- This figure shows that the bias caused by the leaked signal and error orthogonality assumption is almost zero in all of the cases. However, the zero error cross-covariance assumption is causing considerable underestimation in the RMSE estimated by TC. Therefore, the RMSE estimate from TC is a lower bound for the error.

Figure 4: Decomposition of RMSE estimate from TC across six pixels.



Conclusions

- This study presents, for the first time, error estimates of four precipitation products across a central part of the continental US using Triple Collocation (TC)
- The results show that the TRMM product is performing relatively better than the other three products. (lowest RMSE and the highest correlation coefficient with the underlying truth).
- Meanwhile, NEXRAD performs relatively poor in the west side of the study domain that is likely caused by the terrain beam blockage.
- The performance of the GPCP and GPI product were lower than that of TRMM and NEXRAD. GPI has significantly lower performance in the west side of the study domain that is likely caused by the simple retrieval algorithm used in this product. Meanwhile, GPI has a reasonably good correlation with the underlying truth in the east side of the domain.
- Unlike the RMSE, the correlation coefficients of the GPI have a distinct pattern that show the change in the quality of GPI estimates across the domain.
- The results of error decomposition reveal that the TC error estimates underestimate the true error in different products due to a violation of the assumption of zero error cross covariance. However, the result of RMSE estimates from TC have a lot of potential to be incorporated into data assimilation and data merging algorithms.

References

1) Alemohammad, S. H., McColl, K. A., Konings, A. G., Entekhabi, D., and Stoffelen, A., 2015, Characterization of Precipitation Product Errors across the United States using Multiplicative Triple Collocation, Hydrology and Earth System Sciences *(submitted).*
2) McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A., 2014, Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophysical Research Letters, 41, 6229–6236.
3) Konings, A. G., McColl, K. A., Alemohammad, S. H., Entekhabi, D., and Su, C.-H., 2015, Error Characterization of Similar Data Sets: Triple Collocation with Correlated Errors, Geophysical Research Letters *(In Revision).*