# A double instrumental variable method for geophysical product error estimation

Jianzhi Dong[1], Wade T. Crow[1], Zheng Duan[2], Lingna Wei[3] and Yang Lu[4]

*1. USDA Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA*

*2. Lancaster Environment Center, Lancaster University, Lancaster, UK*

*3. Nanjing University of Information Science & Technology, Nanjing, China*

*4. School of Geography and Environmental Science, University of Southampton, Southampton, UK*

**Abstract**

The global validation of remotely sensed and/or modeled geophysical products is often complicated by a lack of suitable ground observations for comparison. By cross-comparing three independent collocated observations, triple collocation (TC) can solve for geophysical product errors in error-prone systems. However, acquiring three independent products for a geophysical variable of interest can be challenging. Here, a double instrumental variable based algorithm (IVd) is proposed as an extension of the existing single instrumental variable (IVs) approach to estimate product error standard deviation ($\sigma$) and product-truth correlation ($R$) using only two independent products - an easier requirement to meet in practice. An analytical examination of the IVd method suggests that it is less prone to bias and has reduced sampling errors relative to IVs. Results from an example application of the IVd method to precipitation product error estimation show that IVd-based $\sigma$ and $R$ are good approximations of reference values obtained from TC at the global extent. In addition to their spatial consistency, IVd estimated error metrics also have only marginal (less than 5%) relative biases versus a TC baseline. Consistent with our earlier analytical analysis, these empirical results are shown to be superior to those obtained by IVs. However, several caveats for the IVd approach should be acknowledged. As with TC and IVs, IVd estimates are less robust when the signal-to-noise ratio of geophysical products is very low. Additionally, IVd may be significantly biased when geophysical products have strongly contrasting error auto-correlations.

*Keywords:* Error estimation, instrumental variable, triple collocation

## 1. Introduction

Remote-sensing retrieved and reanalyzed geophysical variables are increasingly available at the global scale (e.g. Huffman et al., 2007; Mu et al., 2007; Entekhabi et al., 2010; Kerr et al., 2010). However, these products are not completely consistent, as shown in comparison studies of global soil moisture (e.g. Chen et al., 2017; Burgin et al., 2017), evapotranspiration (ET) (e.g. Sörensson and Ruscica, 2018) and precipitation products (e.g. Dee et al., 2011; Gelaro et al., 2017). Validating these products, and interpreting inter-product differences, is often challenging due to the lack of intensive ground-based observations at the global scale (Chen et al., 2017). For example, a large fraction of the Western United States averages less than 0.5 rain gauges per 0.25-degree spatial grid (Massari et al., 2017). Spatial rain gauge density is further reduced for areas of Africa and South America (Koster et al., 2016; Dezfuli et al., 2017). Compared with these global rain gauge densities, ground-based soil moisture and ET observations are even more sparsely distributed (Crow et al., 2012; Lu et al., 2016). Consequently, neglecting ground observation errors leads to a high-bias in error estimates acquired for these products via comparisons against sparse ground-based observation networks (Massari et al., 2017; Chen et al., 2017).

Triple collocation (TC) analysis (Stoffelen, 1998) has proven to be a valuable tool for evaluating errors in uncertain measurement systems (McColl et al., 2014; Gruber et al., 2016a). TC essentially estimates the product error variances using a set of linear equations. To solve all the unknowns in the linear equation system, at least three independent products are required. While TC analysis is useful for the analysis of multiple land surface and atmospheric properties (Dong and Crow, 2017; McColl et al., 2014; Alemohammad et al., 2015; Chen et al., 2017), obtaining three independent estimates of a single variable can be challenging. For instance, due to the general similarity of land surface model structures, TC cannot use multiple model-based or reanalyzed products in a single triplet (Crow et al., 2015b). Likewise, remote-sensing products often share similar retrieval algorithms and therefore likely contain cross-correlated errors (Massari et al., 2017; Gruber et al., 2016b).

To address this issue, Su et al. (2014) proposed a single-instrumental variable (IVs) technique that enables a TC-type analysis using only two independent data products. In IVs, the lag-1 time-series of one product is used in lieu of a third independent product. If estimation errors are serially white, IVs is theoretically equivalent to TC. This is notable, since obtaining two independent products for a geophysical variable is generally straightforward. However, as discussed below, IVs is relatively sensitive to random sampling errors and biased in the presence of auto-correlated errors. Therefore, improving the tolerance of IVs to sampling errors and error auto-correlation would significantly benefit our ability to globally characterize geophysical data product errors.

Following Su et al. (2014), this study aims to provide a more robust global geophysical data error estimator requiring only two independent products. Specifically, by modifying the IVs formulation to include a second instrumental variable, we propose a new "double instrumental variable" algorithm (IVd) that significantly reduces sampling uncertainties and biases (associated with auto-correlated product errors) impacting IVs error estimates. We begin by analytically demonstrating the advantages of our new IVd approach and then comparing numerical TC, IVs and IVd results obtained via an example application to estimate the global error characteristics of reanalyzed and remotely sensed precipitation products.

## 2. Materials and Methods

### *2.1. Geophysical product error estimation algorithms*

#### *2.1.1. Triple collocation analysis (TC)*

In TC analysis (Stoffelen, 1998), three independent products ($x$, $y$ and $z$) are required. These products are typically assumed to be linearly related to the true signal ($P$). Taking $x$ for illustration, this linear model can be expressed as:

$$x = \alpha_x P + B_x + \epsilon_x \tag{1}$$

where $\alpha_x$ is a scaling factor; $B_x$ is a temporal constant bias and $\epsilon_x$ is zero-mean random error. In addition to linear error model shown in equation (1), a multiplicative error model can also be used in conjunction with a log transformation (Alemohammad et al., 2015).

In TC and instrumental variable based (see below) error analyses, the biases ($B$ in equation (1)) cannot be estimated, unless an unbiased reference is known. Instead, the goal of TC is to estimate the error variance of $x$ (noted as $\sigma_x^2$), and/or the correlation between $x$ and $P$ (product-truth correlation, denoted as $R$) - both of which are unaffected by the bias term. Assuming all product errors are mutually independent and orthogonal to the truth, the covariances between the products are expressed as:

$$C_{xy} = \alpha_x \alpha_y C_{PP} \tag{2}$$

$$C_{xz} = \alpha_x \alpha_z C_{PP} \tag{3}$$

$$C_{yz} = \alpha_y \alpha_z C_{PP} \tag{4}$$

$$C_{xx} = \alpha_x^2 C_{PP} + \sigma_x^2 \tag{5}$$

$$C_{yy} = \alpha_y^2 C_{PP} + \sigma_y^2 \tag{6}$$

$$C_{zz} = \alpha_z^2 C_{PP} + \sigma_z^2 \tag{7}$$

where $C$ represents the covariance of the subscript products. For instance, $C_{xy}$ represents the covariance of $x$ and $y$, and $C_{PP}$ is the variance of the true geophysical signal. Combining equations (2 - 7), the variances of the observation errors can be solved for as:

$$\sigma_x^2 = C_{xx} - \frac{C_{xy}C_{xz}}{C_{yz}} \tag{8}$$

$$\sigma_y^2 = C_{yy} - \frac{C_{xy}C_{yz}}{C_{xz}} \tag{9}$$

$$\sigma_z^2 = C_{zz} - \frac{C_{xz}C_{yz}}{C_{xy}}. \tag{10}$$

Likewise, the truth-product correlations ($R$) can be solved for as (McColl et al., 2014):

$$R_{Px}^2 = \frac{C_{xy}C_{xz}}{C_{yz}C_{xx}} \tag{11}$$

$$R_{Py}^2 = \frac{C_{xy}C_{yz}}{C_{xz}C_{yy}} \tag{12}$$

$$R_{Pz}^2 = \frac{C_{xz}C_{yz}}{C_{xy}C_{zz}}. \tag{13}$$

#### *2.1.2. Single instrumental variable based algorithm (IVs)*

As noted above, the goal of our study is to evaluate geophysical product errors using only two independent products. This can be achieved by introducing an instrumental variable ($I$) (Su et al., 2014). Provided that the product errors are serially white, $I$ can be directly taken from the lag-1 time series of one product. Here, for illustration, we take $I$ as lag-1 [day] time series of $x$, i.e., $I_t = \alpha_x P_{t-1} + B_x + \epsilon_{x_{t-1}}$. Following Su et al. (2014), the covariance between the original products and this instrumental variable can be expressed as:

$$C_{Ix} = \alpha_x^2 L_{PP} \tag{14}$$

$$C_{Iy} = \alpha_x \alpha_y L_{PP} \tag{15}$$

where $L_{PP}$ is the lag-1 auto-covariance of the true signal. Taking the ratio of equations (14) and (15) yields:

$$s_{ivs} \equiv \frac{C_{Ix}}{C_{Iy}} = \frac{\alpha_x}{\alpha_y} \tag{16}$$

where $s_{ivs}$ is the IVs-estimated scaling ratio of the two products. Combining equations (2), (5), (6) and (16), the error variances of $x$ and $y$ can be solved for as:

$$\sigma_x^2 = C_{xx} - C_{xy} s_{ivs} \tag{17}$$

$$\sigma_y^2 = C_{yy} - C_{xy} / s_{ivs} \tag{18}$$

and their correlation with truth can be estimated as:

$$R_{Px}^2 = \frac{C_{xy} s_{ivs}}{C_{xx}} \tag{19}$$

$$R_{Py}^2 = \frac{C_{xy}}{C_{xx} s_{ivs}}. \tag{20}$$

### *2.1.3. Double instrumental variable based algorithm (IVd)*

Here we modify the estimates of $s_{ivs}$ by including one additional instrumental variable, so that the method is now referred to as the double instrumental variable algorithm or IVd. As demonstrated below, this modification enhances the robustness and the accuracy of scaling ratio estimates made in equation (16) and, by extension, subsequent estimates of $\sigma$ and $R$.

In IVd, serially lag-1 geophysical observations from *both* products are used as instrumental variables, i.e., one additional instrumental variable ($J_t = \alpha_y P_{t-1} + B_y + \epsilon_{y_{t-1}}$ ) is used relative to IVs. Consequently, the covariance of $y$ and $J$ (i.e., $C_{Jy}$) is expressed as:

$$C_{Jy} = \alpha_y^2 L_{PP}. \tag{21}$$

Combining equations (14) and (21), the ratio of $\alpha_x$ and $\alpha_y$ can be solved for as:

$$s_{ivd} = \sqrt{\frac{C_{Ix}}{C_{Jy}}} \tag{22}$$

where $s_{ivd}$ is the scaling ratio estimated by IVd. Based on equation (22), the standard deviation of the product errors (i.e., $\sigma_x$ and $\sigma_y$) and their correlation with truth (i.e., $R_{Px}$ and $R_{Py}$) can be estimated using equations (17 - 20).

As demonstrated below, this modification reduces the impact of random sampling errors on scaling ratio estimates, which leads to reduced uncertainty in $\sigma$ and $R$ estimates. Additionally, this modification is more tolerant of auto-correlated errors (see below).

### *2.2. Analytical comparisons of IVs and IVd scaling ratios*

Here, we use analytical solutions to provide insight into IVs and IVd comparisons. As shown in equations (17 - 20), in both IV algorithms, error and/or bias in scaling ratio (i.e., $s_{ivs}$ and $s_{ivd}$) estimates is linearly propagated into $\sigma^2$ and $R^2$ estimates. Additionally, since IVd and IVs differ only in their scaling ratio calculation - compare equations (16) and (22) - sampling errors in other components (e.g., $C_{xy}$ and $C_{xx}$) will have the same impact on both IVs and IVd estimates. Therefore, the relative performance of IVs and IVd is determined by the relative accuracy of their scaling ratio estimates. This assumption is further confirmed by numerical synthetic experiments shown in Appendix A. Hence, this section focuses on an analytical description for the robustness of IVd- and IVs-estimated scaling ratios.

#### *2.2.1. Random sampling error impacts*

Since all the covariance terms are sampled with finite sample sizes, they are expected to be affected by random sampling errors. As shown above, $C_{Ix}$, $C_{Iy}$ and $C_{Jy}$ are linearly proportional to the same quantity, i.e., $L_{PP}$. Hence, we can assume for simplicity that sampling error has same impacts on the signal-to-noise ratios (SNR) for all covariance terms. Under such an assumption, the expression for covariance estimation given the presence of random sampling errors can be expressed as:

$$\tilde{C}_{Ix} = \alpha_x^2 L_{PP}(1 + v_{Ix}) \tag{23}$$

$$\tilde{C}_{Iy} = \alpha_x \alpha_y L_{PP}(1 + v_{Iy}) \tag{24}$$

$$\tilde{C}_{Jy} = \alpha_y^2 L_{PP}(1 + v_{Jy}) \tag{25}$$

where $v_{Ix}$, $v_{Iy}$ and $v_{Jy}$ are sampling errors with variance $V$. Clearly, extremely large sampling errors (i.e., $V$) can lead to unstable (or even negative) covariance estimates. These cases are filtered out by bootstrap sampling (see Section 2.4 to follow). Hence, we effectively assume that $V$ is relatively small, and $(1 + v_{Ix})$, $(1 + v_{Iy})$ and $(1 + v_{Jy})$ are all positive.

According to equations (23 - 24), the scaling ratio estimated by the IVs method, in the presence of sampling errors, should be modified as:

$$s_{ivs} = \frac{\alpha_x}{\alpha_y} \frac{1 + v_{Ix}}{1 + v_{Iy}}. \tag{26}$$

Based on the first-order term of Taylor's series expansion, equation (26) can be approximated as:

$$s_{ivs} \approx \frac{\alpha_x}{\alpha_y}(1 + v_{Ix} - v_{Iy}). \tag{27}$$

Hence, the mean-squared error of IVs estimated scaling ratio can be approximated as:

$$\overline{(s_{ivs} - s)^2} \approx 2\frac{\alpha_x^2}{\alpha_y^2} V\left(1 - \rho_{ii}\right) \tag{28}$$

where $s$ is the true scaling ratio (i.e., $\alpha_x/\alpha_y$), and $\rho_{ii}$ is the correlation coefficient of $v_{Ix}$ and $v_{Iy}$. Likewise, approximating the ratio of equations (23) and (25) using a Taylor's expansion yields the following IVd-estimated scaling ratio (in the presence of sampling errors):

$$s_{ivd} \approx \frac{\alpha_x}{\alpha_y}(1 + 0.5 v_{Ix} - 0.5 v_{Jy}). \tag{29}$$

Hence, the uncertainty of $s_{ivd}$ is:

$$\overline{(s_{ivd} - s)^2} \approx 0.5\frac{\alpha_x^2}{\alpha_y^2} V(1 - \rho_{ij}) \tag{30}$$

where $\rho_{ij}$ is the correlation coefficient of $v_{Ix}$ and $v_{Jy}$. If the sampling error cross-correlation (i.e., $\rho$ values) are negligible or similar in size, equations (28) and (30) reveal that the variance of $s$ estimation error for IVd is (approximately) 1/4 of the comparable IVs case. As shown in equations (17 - 20), error in $s$ is linearly propagated into $\sigma^2$ and $R^2$. Hence, relative to a IVs baseline, the application of IVd reduces $\sigma^2$ and $R^2$ sampling uncertainties by this same fraction.

As discussed above, this analysis assumes that sampling errors have the same impact on the SNR of all sampled covariance terms. While this assumption does not strictly hold in all cases, numerical results in Appendix A demonstrate that, for a wide variety of cases, the violation of this assumption does not alter our underlining conclusion regarding the sampling superiority of IVd versus IVs.

### *2.2.2. Impacts of auto-correlated errors*

To match unbiased TC algorithm estimates, both IVs and IVd require temporally white errors in all products. This section demonstrates the impact of temporally auto-correlated errors on both IVs and IVd. For the purpose of illustration, we take a lag-1 [day] time series of $x$ as the instrumental variable in IVs, which is assumed to have auto-correlated errors (i.e. $\overline{\epsilon_{x_t}\epsilon_{x_{t-1}}} \neq 0$). Given the presence of such errors, equation (16) has to be modified as:

$$s_{ivs} = \frac{\alpha_x}{\alpha_y} + \frac{L_{\epsilon_x}}{\alpha_x \alpha_y} \tag{31}$$

where $L_{\epsilon_x} = \overline{\epsilon_{x_t}\epsilon_{x_{t-1}}}/L_{PP}$. Clearly, equation (31) shows that auto-correlated errors lead to biased IVs scaling ratio estimates. Combining equations (17 - 20) reveals that subsequent IVs $\sigma^2$ and $R^2$ estimates are also biased by the same additive term of $L_{\epsilon_x}/\alpha_x\alpha_y$.

Likewise, in the case of auto-correlated error, the IVd-estimated scaling ratio (i.e., equation (22)) can be expressed as:

$$s_{ivd} = \sqrt{\frac{\alpha_x^2 + L_{\epsilon_x}}{\alpha_y^2 + L_{\epsilon_y}}} \tag{32}$$

where $L_{\epsilon_y} = \overline{\epsilon_{y_t}\epsilon_{y_{t-1}}}/L_{PP}$. Note that $L_{\epsilon_x}$ (or $L_{\epsilon_y}$) is zero if errors in $x$ (or $y$) are temporally white. Based on the first-order term of its Taylor's series expansion, equation (32) can be approximated as:

$$s_{ivd} \approx \frac{\alpha_x}{\alpha_y} + \frac{1}{2\alpha_x\alpha_y} L_{\epsilon_x} - \frac{1}{2\alpha_x\alpha_y}\frac{\alpha_x^2}{\alpha_y^2} L_{\epsilon_y}. \tag{33}$$

As demonstrated in both precipitation (see Section 3) and soil moisture (Dong and Crow, 2017) error analyses, error auto-correlation (i.e., the sign of $L_{\epsilon_x}$ and $L_{\epsilon_y}$) tends to be positive for most geophysical data products. Under such an assumption, equation (33) suggests that, in an IVd analysis, biases introduced by $L_{\epsilon_x}$ and $L_{\epsilon_y}$ will partly offset one another. Notably, the IVd net bias will be zero if the two products have the same error auto-correlation characteristics (i.e., $\alpha_y^2 L_{\epsilon_x} = \alpha_x^2 L_{\epsilon_y}$).

It should be noted that IVs uses only one single instrumental variable and hence requires only one product to contain temporally uncorrelated errors. On the contrary, IVd requires both products to contain serially white errors, or, as discussed above, that their error auto-correlation impacts are approximately equal and thus offset each other (see equation (33)). Hence, IVs is theoretically preferable when one of the geophysical product is known to have temporally white errors, or when the two products are known to have strongly contrasting error auto-correlations. However, this does not generally represent a practical advantage for an IVs analysis. When examining two sets of independent observations, it is generally impossible to determine which (if any) product has serially white errors. Furthermore, even small error auto-correlation will lead to large biases in IVs when the SNR of geophysical products is low (see Appendix A). Therefore, IVd results in a more conservative and robust strategy in response to the potential presence of auto-correlated errors in either input product. Although only the first-order term of the Taylor's series expansion are considered in the analytical discussion above, numerical results in Appendix A verify that the consideration of higher-order terms does not qualitatively change the conclusion that IVd is more robust to auto-correlated errors.

### *2.3. Precipitation data*

As discussed above, precipitation error analyses will be used as a case study for evaluating the relative performances of IVs and IVd. TC-based precipitation error analyses have previously been verified using intensive ground-based precipitation networks over the Eastern US and Southeastern China (Massari et al., 2017; Li et al., 2018). Hence, TC error analysis of daily precipitation products is arguably better validated than any other land surface variable.

For TC analysis, a variety of precipitation products were acquired. The SM2Rain precipitation product (Brocca et al., 2015) is based on Advanced Scatterometer (ASCAT) soil moisture retrievals (Wagner et al., 1999) and estimates precipitation by inverting the land surface water balance using ASCAT soil moisture time series. This 0.25-degree daily precipitation product is available from January 2007 to June 2015 (http://hydrology.irpi.cnr.it/download-area/sm2rain-data-sets/) and based on the approach described in Brocca et al. (2017).

The reanalyzed, daily, 0.5-degree ERA-Interim precipitation product (Dee et al., 2011) was collected from European Centre for Medium-Range Weather Forecasts (ECMWF, https://www.ecmwf.int/). The ERA-Interim is a data assimilation system based on ECMWF forecast model (Dee et al., 2011).

The L3 daily 0.25-degree precipitation product (TRMM_3B42_Daily) was provided by the Tropical Rainfall Measuring Mission (Huffman et al., 1997). It is generated by taking the daily average of the near real-time, 3-hourly TRMM Multi-Satellite Precipitation Analysis (TMPA) 3B42RT product. These estimates are retrieved from a variety of low-earth orbit passive microwave observations (e.g., Microwave Imager, Special Sensor Microwave Imager (SSM/I), Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E), and the Advanced Microwave Sounding Unit-B (AMSU-B)) using the Goddard Profiling Algorithm (Huffman et al., 2007).

Global, daily, ground-based observations from the 0.5-degree CPC precipitation product (Xie et al., 2007) were also collected from the NOAA Earth System Research Laboratory. To enable the comparison of precipitation products, both SM2Rain and TRMM precipitation estimates were linearly averaged onto a 0.5-degree global land grid (the native resolution of the ERA-Interim and CPC precipitation data). All TC, IVs and IVd results were based on a daily analysis conducted between January 2007 and June 2015 - a period in which all four products are available.

### *2.4. Implementation of precipitation error analysis*

As mentioned above, TC, IVd and IVs can be applied to a multiplicative error case via application of a log transform (Alemohammad et al., 2015), which may better capture the multiplicative nature of errors in short-term precipitation accumulation estimates (Tian et al., 2013). However, implementing the multiplicative error model in TC requires removing zero-precipitation days (Alemohammad et al., 2015), which is inappropriate for many climate regions (Massari et al., 2017). In addition, empirical TC results for precipitation errors (verified via comparisons against error estimates obtained from dense rain gauge networks) suggest that biases associated with the use of an additive error model are minimal (Li et al., 2018; Massari et al., 2017).

Since TC, IVs and IVd estimates can contain substantial uncertainties when applied to low-quality geophysical data products (Dong et al., 2018), a 1000-member bootstrap sampling was used to evaluate the uncertainty of TC-, IVs- and IVd-estimated $\sigma$ and $R$ for each grid cell over the globe. Unreliable estimates were then masked out according to the bootstrapped uncertainties (see below). Since the sampling length of the precipitation data sets was relatively long (approximately 8 years), the stability of error analyses is primarily determined by the geophysical product quality (Dong et al., 2018). Hence, this bootstrapping method was assumed to be sufficient for filtering out the unreliable estimates, and more sophisticated sampling techniques were not considered.

Each bootstrap sampling member was constructed by re-sampling the original observation time series with replacement to preserve the original sample size. For each randomly sampled time step, all products were paired to preserve their original correlation and/or auto-correlation strength. For example, at random time step $t$, $x_t$, $y_t$ and $z_t$ were simultaneously drawn for TC analysis, and $x_t$, $y_t$, $x_{t-1}$ and $y_{t-1}$ for IVd.

## 3. Results

### *3.1. Comparison of TC and IVs estimates*

Global TC and IVs estimates for the standard deviation of precipitation error ($\sigma$) in daily ERA-Interim and TRMM rainfall products are shown in Figure 1. A more detailed statistical comparison of the two methods is given in Figure 2. Due to the presumed error independence between the CPC, SM2Rain and ERA-Interim products (Massari et al., 2017), ERA-Interim errors were estimated by applying TC to this triplet (Figure 1a). Likewise, a CPC-SM2Rain-TRMM constructed triplet was used for the TRMM TC error analysis (Figure 1b). The lag-1 [day] time series of CPC-based precipitation was used as the single instrumental variable for both the CPC-ERA-Interim and the CPC-TRMM IVs analyses (Figures 1c and d).

Based on the TC results in Figure 1a and b, ERA-Interim generally demonstrates lower $\sigma$ than comparable TRMM estimates (particularly over North America, Europe, Australia and Central Asia). Both ERA-Interim and TRMM tend to present larger $\sigma$ over relatively wet regions, e.g., the Eastern United States, Amazon basin, Southeastern China. This tendency is consistent with earlier ground-based validation results presented in Chen et al. (2013) and previous findings that precipitation observation errors generally increase with areal mean precipitation (Huff, 1970).

The spatial distribution of IVs-based ERA-Interim and TRMM $\sigma$ (Figure 1c and d) is highly analogous to that obtained via TC. For example, the spatial correlation between TC and IVs $\sigma$ results is above 0.9 [-] for both ERA-Interim and TRMM (Figure 2). As shown above, the precipitation error primarily reflects patterns in mean annual precipitation. Therefore, the general consistency of TC- and IVs-based $\sigma$ values is not wholly unexpected. However, a subtle (but spatially persistent) bias is seen in IVs estimates of ERA-Interim and TRMM $\sigma$ relative to benchmark TC results (Figure 2). Larger differences are found for $R$ results. In particular, global $R$ patterns estimated by IVs are poorly correlated with TC results (Figure 3c and d), and a clear low bias in IVs estimates is evident relative to comparable TC results (Figure 4).

<Figure 1 here please >

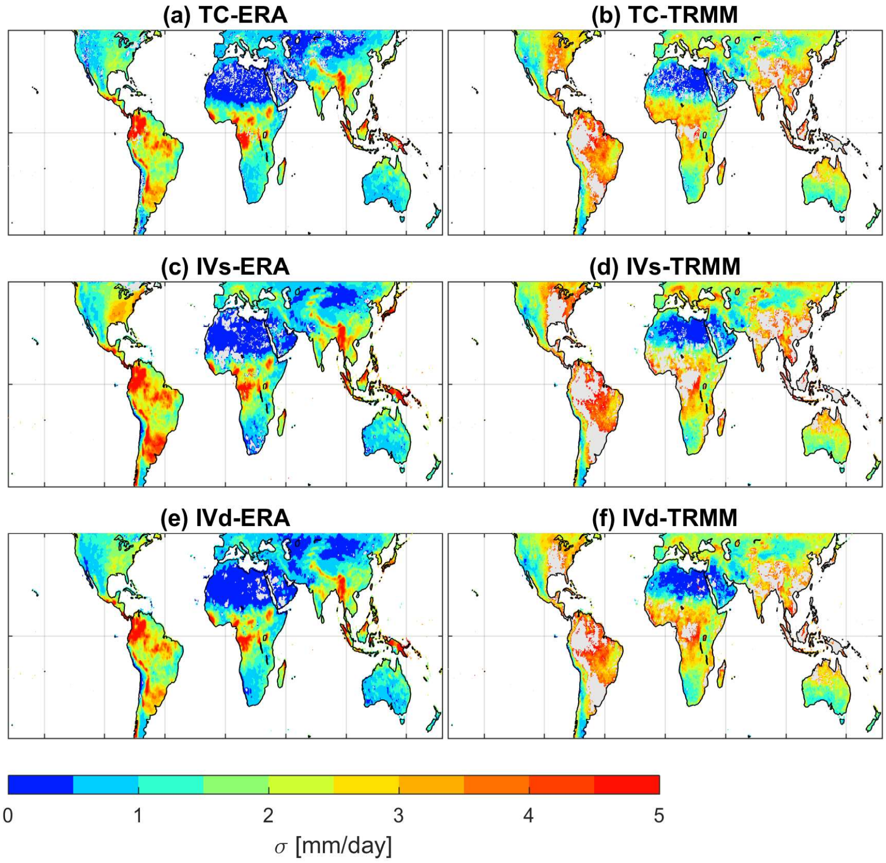


Figure 1: The standard deviation of daily precipitation error ($\sigma$, mm/day) for ERA-Interim (left column) and TRMM (right column) estimated using TC, IVs and IVd. The lag-1 [day] CPC precipitation time series was used as the instrumental variable for both IVs estimates. Grey shading indicates land areas where the bootstrapped uncertainty of $\sigma$ estimates is larger than 3 mm/day.

<Figure 2 here please >

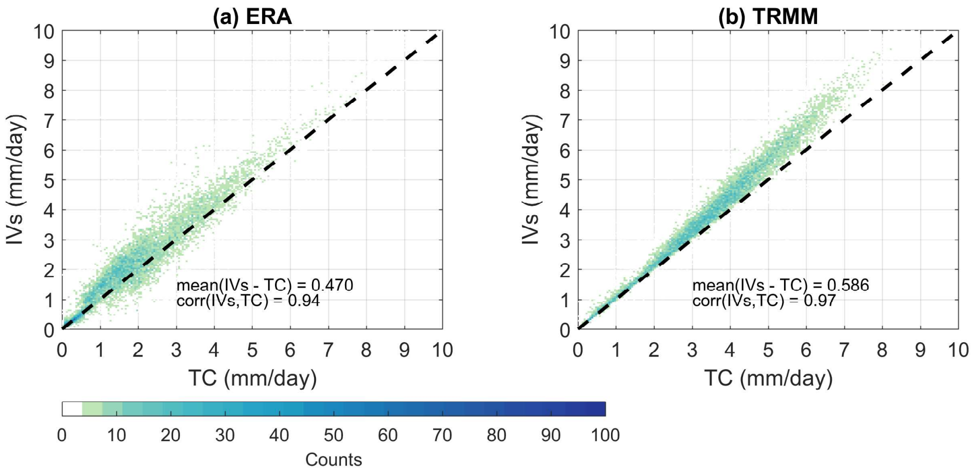


Figure 2: Density plot of TC- and IVs-estimated global daily ERA-interim (a) and TRMM (b) $\sigma$ as presented in Figure 1. Text provides the mean difference (mean(IVs - TC)) and consistency (spatial correlation, denoted as corr) between TC and IVs error estimates. Color shading captures the density of points within a 0.1 mm/day $\times$ 0.1 mm/day grid.

<Figure 3 here please >

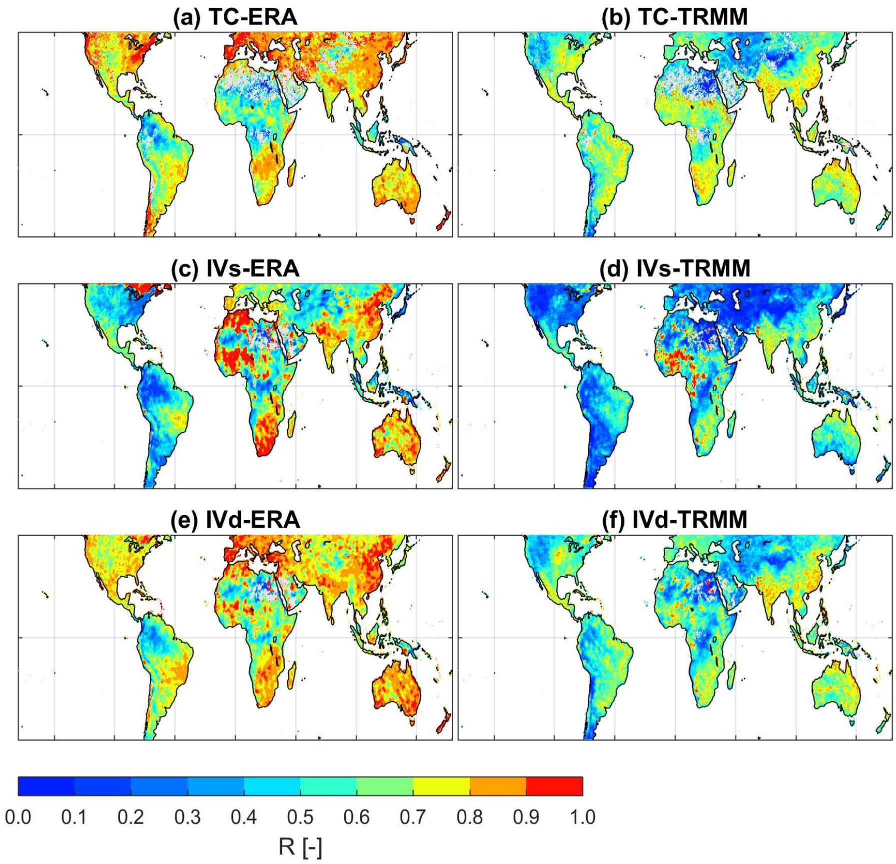


Figure 3: The product-truth correlation ($R$) for ERA-Interim (left column) and TRMM (right column) estimated using TC, IVs and IVd. The lag-1 [day] CPC precipitation time series was used as the instrumental variable for both IVs estimates. Grey shading indicates land areas where the bootstrapped uncertainty of the correlation estimate is larger than 0.3 [-].

<Figure 4 here please >

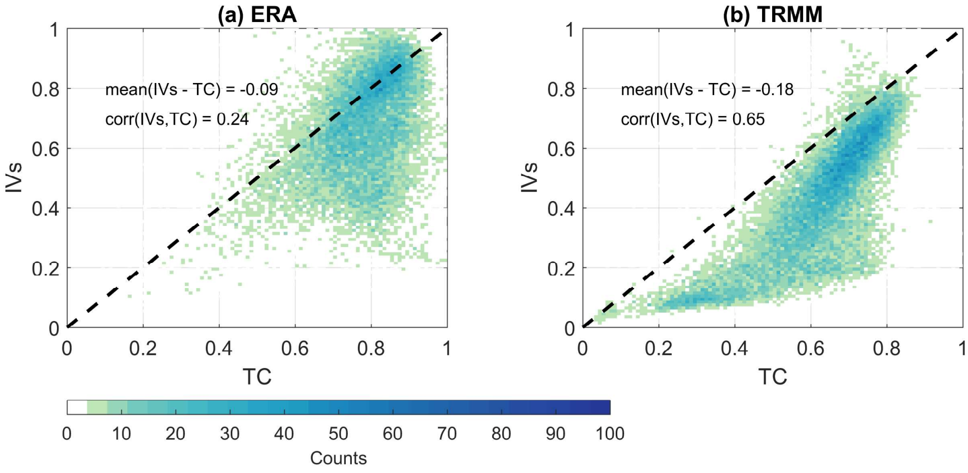


Figure 4: Density plot of TC- and IVs-based daily ERA-Interim (a) and TRMM (b) $R$ (i.e., correlation with truth) as presented in Figure 3. Text provides the mean difference (mean(IVs - TC)) and consistency (spatial correlation, denoted as corr) between TC and IVs error estimates. Color shading captures the density of points within a 0.1 [-] $\times$ 0.1 [-] grid.

The biases in Figures 2 and 4 suggest that the product selected as the instrumental variable (i.e., the CPC precipitation product) in the IVs analysis contains auto-correlated error (see Section 2.2.2). This bias cannot be detected if reference TC estimates are not available. Nonetheless, here we implement the alternative instrumental variable in the IVs analysis to investigate whether the biases shown above can be reduced. In this alternative case, lag-1 [day] ERA-Interim and TRMM data are used as instrumental variables for the CPC-ERA-Interim and CPC-TRMM IVs analyses, respectively.

By switching the instrumental variable (from CPC to ERA-Interim), biases in ERA-Interim estimates are slightly increased (Figure 5 a and c) relative to the previous IVs estimates (where CPC was used as instrumental variable - see Figures 2 and 4). Some bias reduction is observed in TRMM estimates when changing the instrumental variable from CPC to TRMM (Figure 5 b and d). However, residual $R$ and $\sigma$ biases are still evident in both cases, and the signs of the bias are opposite to the cases utilizing CPC as instrumental variable (comparing Figures 2, 4 and 5). This suggests that both ERA-Interim and TRMM precipitation errors are temporally auto-correlated and the sign of this error auto-correlation is the same as that for CPC errors (see Section 2.2.2). Additionally, utilizing non-CPC datasets as the instrumental variable leads to decreased correlation with

TC estimates (compare Figures 2 and 4, 5). Note that, the relative bias of IVs estimates derived from different instrumental variables depends on the strength and the sign of product error auto-correlation, which cannot be readily evaluated without a reliable reference.

<Figure 5 here please >

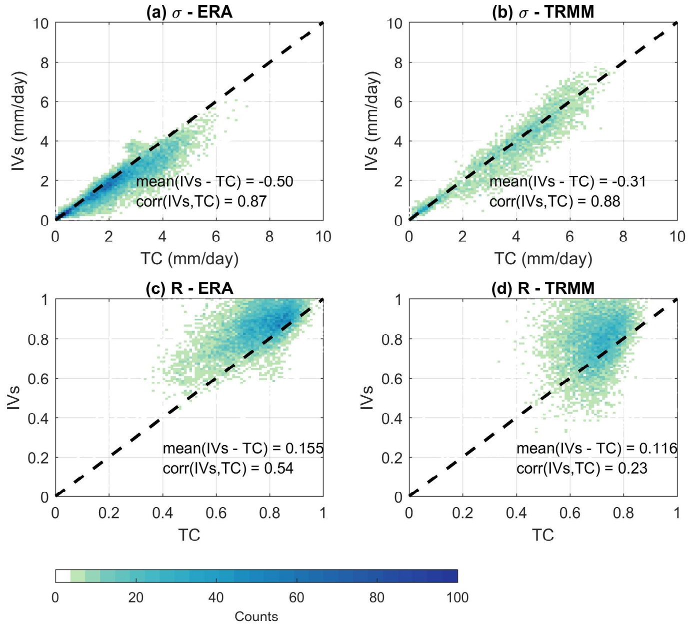


Figure 5: Same as Figures 2 and 4, but lag-1 [day] ERA-Interim and TRMM precipitation are used as instrumental variables for the CPC-ERA (a and c) and CPC-TRMM (b and d) IVs analyses, respectively. Text provides the mean difference (mean(IVs - TC)) and consistency (spatial correlation, denoted as corr) between TC and IVs error estimates. Color shading captures the density of points within a 0.1 [mm/day] × 0.1 [mm/day] (a and b) and 0.1 [-] × 0.1 [-] (c and d) grid.

### *3.2. Evaluation of the IVd algorithm*

The global pattern of ERA-Interim and TRMM errors estimated by IVd is presented in Figures 1 and 3. To be consistent with IVs results discussed above, CPC and ERA-Interim were used for the ERA-Interim IVd analysis (i.e., both lag-1 [day] CPC and ERA-Interim precipitation products are applied as instrumental variables (see Section 2.1.3)). Likewise, both CPC and TRMM were used as instrumental variables for IVd-based TRMM error estimation. Despite its use of only two products, IVd-based $\sigma$ and $R$ global patterns correspond closely to benchmark TC results (Figure 1 and 3). Improved consistency with the TC benchmark is reflected in the increased correlation between IVd and TC estimates in Figure 6 (relative to IVs and TC results shown earlier in Figures 2, 4 and 5). In addition, IVd demonstrates less global bias than IVs in $\sigma$ and $R$ estimates for both the ERA-Interim and TRMM cases. This is consistent with our analytical interpretation that temporal auto-correlated error impacts from different products tend to offset each other in IVd - see equation (33). Overall, the bias of IVd $\sigma$ and $R$ estimates is less than 5% of the nominal global mean for all cases considered.

<Figure 6 here please >

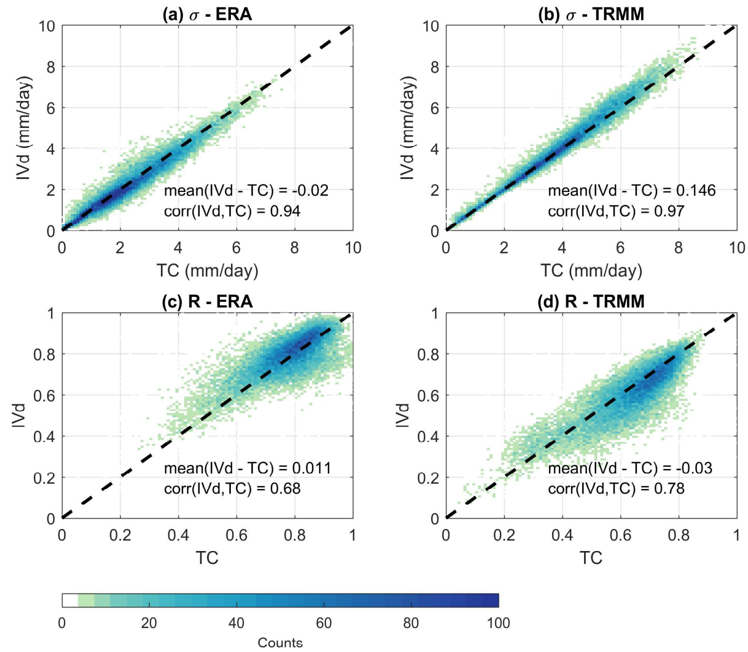


Figure 6: Density plot of global IVd and TC estimated $\sigma$ (top row) and $R$ (bottom row). Text provides the mean difference (mean(IVd - TC)) and consistency (spatial correlation, denoted as corr) between TC and IVd error estimates. Color shading captures the density of points within a 0.1 [mm/day] $\times$ 0.1 [mm/day] (a and b) and 0.1 [-] $\times$ 0.1 [-] (c and d) grid.

## 4. Discussion

Relative to IVs results based on the original formation of Su et al. (2014), IVd numerical results demonstrate increased robustness and reduced bias (Figure 6). These advantages can be explained via a straight-forward analytical comparison of error propagation within the IVd and IVs algorithms (see Section 2.2). In particular, including a second instrumental variable leads to the calculation of scaling ratio in a square root form which substantially reduces sampling error impacts on the scaling ratio estimation and, hence, increases the robustness of the precipitation error estimates (see Section 3.2 and Appendix A). Furthermore, given that geophysical data products generally contain positive error auto-correlation, error auto-correlation impacts in IVd tend to be reduced - see equation (33). This effectively reduces the net bias in IVd-based error estimates in the presence of temporally auto-correlated errors.

However, several caveats must be noted. As shown in Section 2.1, both IV approaches calculate the scaling ratios using the auto-covariance of the true geophysical signal (i.e., $L_{PP}$). Clearly, both IVs and IVd are less robust when the true signals have limited memories (auto-correlations), e.g., $L_{PP} \approx 0$. Likewise, if the geophysical products are overwhelmed by random errors, the sampled $L_{PP}$ will be extremely unstable and affect the subsequent accuracy of both IVs and IVd. Additionally, as shown in Section 2.2.2, IVd is most accurate when the geophysical product errors are temporally white or have comparable auto-correlation strengths. However, modeled state variables (e.g., soil moisture obtained from prognostic water balance calculations) are likely to have stronger temporal error auto-correlation than remote-sensing products (Dong and Crow, 2017). This contrast in auto-correlation strength between modeled and observed geophysical products could conceivably affect IVd estimates (see Appendix A).

It should also be acknowledged that precipitation accumulation error can be both multiplicative and non-orthogonal in form (Tian et al., 2013). Both of these characteristics would obviously violate the underlying orthogonal/additive form of equation (1). As a result, care should always be taken when applying either TC or IV to precipitation data sets (or any other geophysical variable). It is currently unclear how large a problem this poses for rainfall data sets in particular. For example, both Massari et al. (2017) and Li et al. (2018) found only minor biases, relative to a baseline of error quantification against high-quality rain gauges, when applying equation (1) for TC daily rainfall error estimation. Additionally, TC, IVd and IVs are all based on the same assumptions (except for the additional zero-error auto-correlation assumption required for IVd and IVs). Therefore, any factor that affects TC will also impact both IVd and IVs. Given this, there is no reason to suspect that errors in TC-based benchmark results will spuriously favor either IV technique over the other in our evaluation. While TC is certainly not error-free, it nevertheless provides an unbiased reference for a relative evaluation of IVs versus IVd.

## 5. Conclusion

Based on the single instrumental variable algorithm (IVs) introduced by Su et al. (2014), this study describes a “double instrumental variable” algorithm (IVd) which provides improved geophysical product error estimates using only two independent observations of a given geophysical variable. Furthermore, as demonstrated in the empirical precipitation case study, IVd shows strong consistency with TC. This suggests that IVd can significantly benefit general geophysical product error estimation for cases where only two independent products (of a single variable) are available.

Given the promising results shown in the precipitation error estimation case, a logical next step is testing and applying IVd to other geophysical products. IVd is likely to be particularly valuable for global evapotranspiration (ET) error estimation. In ET products, the line between modeled and remotely sensed retrievals is considerably blurred. As a result, available ET estimates are commonly derived from an over-lapping set of forcing datasets (e.g., solar radiation, land surface temperature, and/or air temperature). Hence, obtaining three independent products, and performing TC analysis, is particularly challenging for ET. As a result, IVd may prove to be especially useful for techniques utilizing TC to quantify (and compensate for) random errors in remotely sensed ET products (see e.g., Crow et al. (2015a) and Lei et al. (2018)).

## Acknowledgment

This paper was partially supported by the NASA SMAP mission through award NNH12ZDA001N-SMAP. The work of Dr. Lingna Wei was supported by National Science Foundation of China (grant no. 91747203) and the Natural Science Research Projects for Universities of Jiangsu Province (17KJB170016). We would like to thank the chief editor Dr. Menghua Wang and associate editor Dr. Tim R. McVicar for handling our manuscript, and Dr. Chun-Hsu Su, Dr. Rolf Reichle, Dr. Simon Zwieback and an anonymous reviewer for their constructive comments.

## Appendix A. Synthetic experiment for IVs and IVd comparison

Analytical comparisons of random sampling error and auto-correlated error impacts on IVs and IVd are shown in Section 2.2. Here, we describe a set of synthetic experiments to further demonstrate the superiority of IVd versus IVs.

To start, a true random precipitation time series is generated by adding random numbers (drawn from U(0,1000)) to a zero time series on randomly sampled time steps (denoted as $P$). Synthetic products ($x$ and $y$) are then generated by adding random zero-mean additive Gaussian errors to the truth according to equation (1):

$$x = P + \epsilon_x \tag{A.1}$$

$$y = P + \epsilon_y. \tag{A.2}$$

For simplicity, both $\alpha_x$ and $\alpha_y$ are assumed to be one. This assumption has no impact on results shown below.

In the first experiment, both $\epsilon_x$ and $\epsilon_y$ are assumed to be temporally white. Five experiment lengths were used (varying from 50 to 5000 daily time steps) to capture sample size impacts on the relative performance of IVd and IVs. For each sample size, the ratio of IVd and IVs mean-squared error was sampled from 1000 tests (Figure A.1). This set of experiment was repeated 3000 times to capture the uncertainty of the sampled IVd and IVs mean-squared error ratios (MSER) for each given sample size. Since only serially white errors are considered, both IVd and IVs have relatively small biases (typically less than 5%). Hence, comparisons of IVd and IVs biases are not presented.

As shown Section 2.2.1, provided the sampling error variances of the auto-covariances are proportional to the true signal atuo-covariance, the mean-squared error of IVd scaling factor should be 75% lower than that of IVs. However, Figure A.1 shows that this ratio varies with both the sample size and the assumed SNR of the synthetic products. Typically, the benefit of IVd is smaller than the analytical predictions for cases with relatively large sample sizes and high SNR (approximately 40% error reduction). On the contrary, under the scenario that the SNR of the synthetic products is low, the error reduction by IVd can be substantially higher than the analytically prediction of 75% (see cases with SNR = 0.1 [-] in Figure A.1 a). The difference between the analytical solution (see Section 2.2.1) and the numerical results are mainly due to the assumption that auto-covariance sampling errors are equal in magnitude (i.e., the variance of $v_{Ix}$, $v_{Iy}$ and $v_{Jy}$ are all equal to $V$). Nonetheless, none of the cases qualitatively change our central conclusion that IVd is more tolerant of sampling errors than the original IVs method. Since IVs and IVd estimate $R^2$ and $\sigma^2$ using the same procedure, their relative errors are simply proportional to the scaling factor estimation errors plotted in Figure A.1.

<Figure A.1 here please >

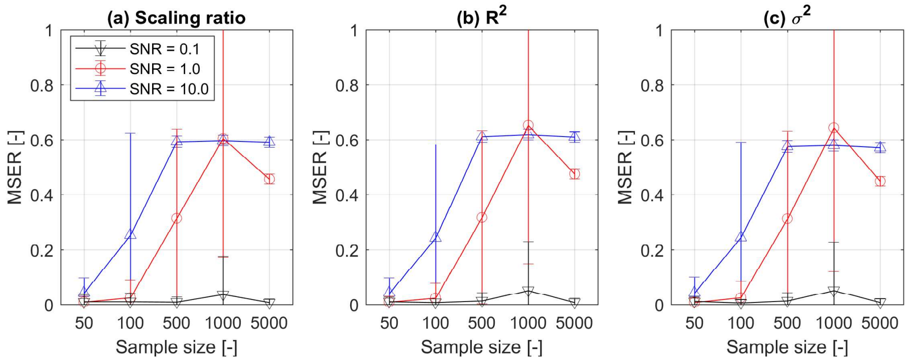


Figure A.1: The ratio of IVd and IVs mean-squared error as a function of sample size (experiment length) and product SNR. Error bar captures the standard deviation of the MSER values.

A second set of synthetic test repeats the previous experiment, but with temporally auto-correlated errors. We first examine the cases that both products (i.e., $x$ and $y$) have the same error auto-correlation coefficients (EAC). For simplicity, a fixed sample size of 1000 is used. As shown in Figure A.2, the bias of IVs increases with increased error auto-correlation strength. In contrast, IVd shows (approximately) zero-biases for most cases. This is consistent with the analytical solution shown in Section 2.2.2, which demonstrates that IVd is unbiased if the two products have the same EACs.

Next, the EAC of $x$ is taken as constant value of 0.1, but the EAC of $y$ is varied within the range of 0.1 to 0.9. Here, we assume that $x$ is known to have smaller EAC and therefore is selected as the instrumental variable in IVs. The biases presented in Figure A.3 are averaged across estimation errors of both $x$ and $y$. For relative high-SNR cases (Figure A.3 a to f), IVd outperforms IVs even when the EAC of $y$ is slightly higher than that of $x$ (e.g., when EAC of $y$ is below 0.3). As the difference between $x$ and $y$ EAC increases, IVd demonstrates larger biases than the IVs estimates, which is expected and wholly consistent with our analytical results shown in Section 2.2.2.

Interestingly, when the assumed SNR of $x$ and $y$ is low, IVd constantly outperforms IVs, regardless of the ECA differences between $x$ and $y$ (Figure A.3 g to i). Reduced SNR tend to increase $L_{\epsilon_x}$ (i.e., $\overline{\epsilon_{x_t}\epsilon_{x_{t-1}}}/L_{PP}$). For such cases, higher-order terms should be considered in the Taylor expansion analyses. Nonetheless, these synthetic experiment results confirm that IVd is more robust to auto-correlated errors than IVs.

<Figure A.2 here please >

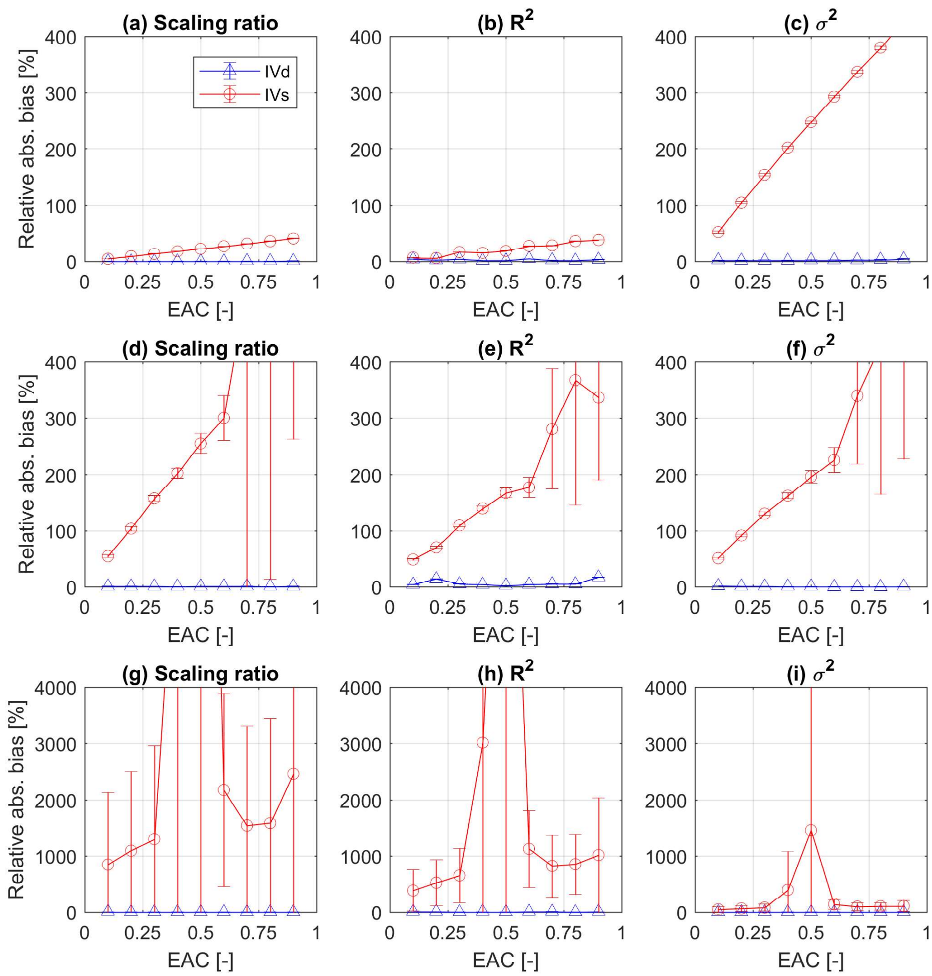


Figure A.2: The relative absolute bias of IVd and IVs estimates as a function of error auto-correlation coefficient (EAC). Both products (i.e., $x$ and $y$) are assumed to have same error auto-correlation coefficients. First row: SNR = 10; second row: SNR = 1; third row: SNR = 0.1. A sample size of 1000 is used in this experiment. Product $x$ is used as the instrumental variable for IVs estimates.

<Figure A.3 here please >

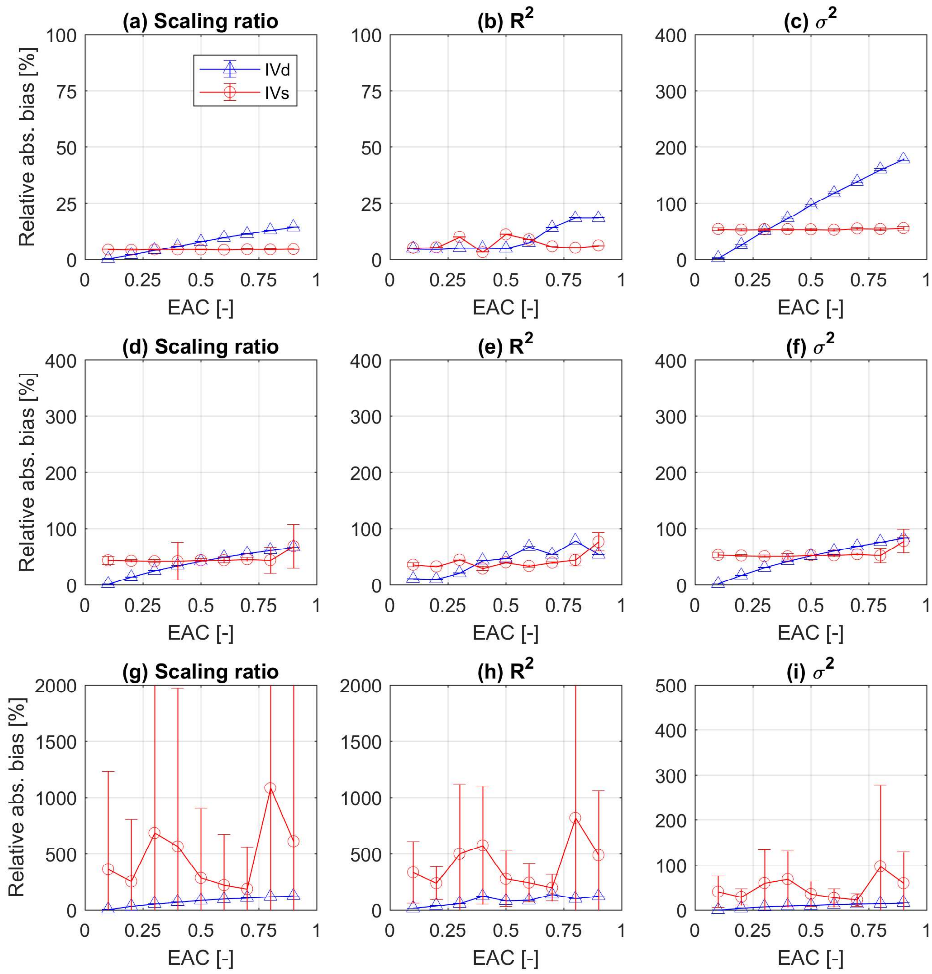


Figure A.3: Same as Figure A.2, but $x$ and $y$ are assumed to have different error auto-correlation strengths. Error auto-correlation coefficient (EAC) for $x$ is set to a constant value of 0.1 [-], and the EAC for $y$ varies from 0.1 to 0.9 [-]. Product $x$ is used as the instrumental variable in IVs.

## Reference

Alemohammad, S. H., McColl, K. A., Konings, A. G., Entekhabi, D., Stoffelen, A., 2015. Characterization of precipitation product errors across the United States using multiplicative triple collocation. Hydrology and Earth System Sciences 19 (8), 3489–3503.

Brocca, L., Crow, W. T., Ciabatta, L., Massari, C., De Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A., et al., 2017. A review of the applications of ASCAT soil moisture products. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (5), 2285–2306.

Brocca, L., Massari, C., Ciabatta, L., Moramarco, T., Penna, D., Zuecco, G., Pianezzola, L., Borga, M., Matgen, P., Martínez-Fernández, J., 2015. Rainfall estimation from in situ soil moisture observations at several sites in Europe: an evaluation of the SM2RAIN algorithm. Journal of Hydrology and Hydromechanics 63 (3), 201–209.

Burgin, M. S., Colliander, A., Njoku, E. G., Chan, S. K., Cabot, F., Kerr, Y. H., Bindlish, R., Jackson, T. J., Entekhabi, D., Yueh, S. H., 2017. A comparative study of the SMAP passive soil moisture product with existing satellite-based soil moisture products. IEEE Transactions on Geoscience and Remote Sensing 55 (5), 2959–2971.

Chen, F., Crow, W. T., Colliander, A., Cosh, M. H., Jackson, T. J., Bindlish, R., Reichle, R. H., Chan, S. K., Bosch, D. D., Starks, P. J., et al., 2017. Application of triple collocation in ground-based validation of Soil Moisture Active/Passive (SMAP) level 2 data products. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (2), 489–502.

Chen, S., Hong, Y., Cao, Q., Gourley, J. J., Kirstetter, P.-E., Yong, B., Tian, Y., Zhang, Z., Shen, Y., Hu, J., et al., 2013. Similarity and difference of the two successive V6 and V7 TRMM multisatellite precipitation analysis performance over China. Journal of Geophysical Research: Atmospheres 118 (23), 13–060.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., Walker, J. P., 2012. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. Reviews of Geophysics 50 (2), RG2002, doi:10.1029/2011RG000372.

Crow, W. T., Lei, F., Hain, C., Anderson, M. C., Scott, R. L., Billesbach, D., Arkebauer, T., 2015a. Robust estimates of soil moisture and latent heat flux coupling strength obtained from triple collocation. Geophysical Research Letters 42 (20), 8415–8423.

Crow, W. T., Su, C.-H., Ryu, D., Yilmaz, M. T., 2015b. Optimal averaging of soil moisture predictions from ensemble land surface model simulations. Water Resources Research 51 (11), 9273–9289.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137 (656), 553–597.

Dezfuli, A. K., Ichoku, C. M., Huffman, G. J., Mohr, K. I., Selker, J. S., van de Giesen, N., Hochreutener, R., Annor, F. O., 2017. Validation of IMERG precipitation in Africa. Journal of Hydrometeorology 18 (10), 2817–2825.

Dong, J., Crow, W., 2017. An improved triple collocation analysis algorithm for decomposing auto-correlated and white soil moisture retrieval errors. Journal of Geophysical Research: Atmospheres, 122, 13081-13094.

Dong, J., Crow, W. T., Bindlish, R., 2018. The error structure of the SMAP single and dual channel soil moisture retrievals. Geophysical Research Letters 45 (2), 758–765.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., et al., 2010. The Soil Moisture Active Passive (SMAP) mission. Proceedings of the IEEE 98 (5), 704–716.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). Journal of Climate 30 (14), 5419–5454.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., Wagner, W., 2016a. Recent advances in (soil moisture) triple collocation analysis. International Journal of Applied Earth Observation and Geoinformation 45, 200–211.

Gruber, A., Su, C.-H., Crow, W. T., Zwieback, S., Dorigo, W., Wagner, W., 2016b. Estimating error cross-correlations in soil moisture data sets using extended collocation analysis. Journal of Geophysical Research: Atmospheres 121 (3), 1208–1219.

Huff, F., 1970. Sampling errors in measurement of mean precipitation. Journal of Applied Meteorology 9 (1), 35–44.

Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., Schneider, U., 1997. The global precipitation climatology project (GPCP) combined precipitation dataset. Bulletin of the American Meteorological Society 78 (1), 5–20.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., Stocker, E. F., 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of hydrometeorology 8 (1), 38–55.

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., et al., 2010. The SMOS mission: New tool for monitoring key elements of the global water cycle. Proceedings of the IEEE 98 (5), 666–687.

Koster, R. D., Brocca, L., Crow, W. T., Burgin, M. S., De Lannoy, G. J., 2016. Precipitation estimation using l-band and c-band soil moisture retrievals. Water Resources Research 52 (9), 7213–7225.

Lei, F., Crow, W. T., Holmes, T. R., Hain, C., Anderson, M. C., 2018. Global investigation of soil moisture and latent heat flux coupling strength. Water Resources Research.

Li, C., Tang, G., Hong, Y., 2018. Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the triple collocation method across mainland china. Journal of Hydrology.

Lu, Y., Dong, J., Steele-Dunne, S. C., van de Giesen, N., 2016. Estimating surface turbulent heat fluxes from land surface temperature and soil moisture observations using the particle batch smoother. Water Resources Research 52 (11), 9086–9108.

Massari, C., Crow, W., Brocca, L., 2017. An assessment of the performance of global rainfall estimates without ground-based observations. Hydrology and Earth System Sciences 21 (9), 4347–4361.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., Stoffelen, A., 2014. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. Geophysical Research Letters 41 (17), 6229–6236.

Mu, Q., Heinsch, F. A., Zhao, M., Running, S. W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote sensing of Environment 111 (4), 519–536.

Sörensson, A. A., Ruscica, R. C., 2018. Intercomparison and uncertainty assessment of nine evapotranspiration estimates over south america. Water Resources Research 54 (4), 2891–2908.

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

Su, C.-H., Ryu, D., Crow, W. T., Western, A. W., 2014. Beyond triple collocation: Applications to soil moisture monitoring. Journal of Geophysical Research: Atmospheres 119 (11), 6419–6439.

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., Wu, H., 2013. Modeling errors in daily precipitation measurements: Additive or multiplicative? Geophysical Research Letters 40 (10), 2060–2065.

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

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., Liu, C., 2007. A gauge-based analysis of daily precipitation over East Asia. Journal of Hydrometeorology 8 (3), 607–626.