remote sensing

MDPI

Communication

The Effect of Error Non-Orthogonality on Triple Collocation Analyses

Jur Vogelzang , Ad Stoffelen * and Anton Verhoef

Royal Netherlands Meteorological Institute KNMI, Utrechtseweg 297, 3731 GA De Bilt, The Netherlands
* Correspondence: ad.stoffelen@knmi.nl

Abstract: Triple collocation analysis is an established technique for calculating the relative linear intercalibration coefficients and observation error variances for physical quantities measured simultaneously in space and time by three different observation systems. A simple parameterized error model is used. It relies on a few assumptions, one of which is that the observation errors are independent of the magnitude of the observed quantities. This is referred to as error orthogonality. Using an ocean surface vector winds data set of 44,948 collocations, this study shows that the violation of error orthogonality does affect the calibration coefficients but has only a small second-order effect on the observation error variances of the calibrated data.

Keywords: triple collocation; error orthogonality; ocean surface winds; OSCAT

check for updates

Citation: Vogelzang, J.; Stoffelen, A.; Verhoef, A. The Effect of Error Non-Orthogonality on Triple Collocation Analyses. *Remote Sens.* **2022**, *14*, 4268. https://doi.org/10.3390/rs14174268

Academic Editors: Bryan Stiles, Svetla Hristova-Veleva, Lucrezia Ricciardulli, Larry O'Neill, Zorana Jelenak and Joe Sapp

Received: 4 July 2022
Accepted: 26 August 2022
Published: 30 August 2022

1. Introduction

Triple collocation is a powerful technique for calculating the observation error variances and linear intercalibration coefficients of a physical quantity measured by three different observation systems at the same time and place. It was introduced by Stoffelen [1] for ocean surface vector winds and later applied by [2] to estimate the accuracy of high-resolution scatterometer wind products. In the past years the technique has been applied on an increasing number of geophysical quantities like ocean surface wind speed [3], ocean surface current [4], sea surface salinity [5], precipitation [6,7], soil moisture [8,9], groundwater storage [10], etc. The list of references is far from complete, and the interested reader is referred to the references and the references therein.

Triple collocation has been extended to include a generalization of the correlation coefficient known from linear regression analysis [11]. Extensions to more than three collocated measurement systems have also been proposed [4,12], but these fall outside the scope of this paper.

Triple collocation relies on the following assumptions:

(1) Linear calibration is sufficient to describe the interrelations between the observation systems;
(2) The observation errors are random with zero average after calibration;
(3) The observation errors are independent of the measured value (error orthogonality).

When deriving the triple collocation equations, one usually applies all approximations as soon as possible in order to simplify the algebra, arguing that the sufficiency of linear calibration can be judged from scatter plots and that error orthogonality is hard to check. To the best of the authors' knowledge, the effect of violation of error orthogonality (further referred to as error non-orthogonality) has never been tested. The aim of this short paper is to fill this gap. It will be shown that error non-orthogonality can be included in the triple collocation covariance equations in an elegant and concise manner. The effect of error non-orthogonality will be studied for a data set consisting of all buoy, OSCAT-25, and European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis ERA5 forecast collocations in 2013 [12,13].

The paper is organized as follows. Section 2 contains a brief description of the data used. In Section 3 error orthogonality is included in the covariance equations for triple collocation analysis. Representativeness errors are defined and included, and the method of solution is briefly described. The results are presented and discussed in Section 4. The paper ends with the conclusions in Section 5.

2. Data

The data set used in this work consists of 44,948 collocations of buoy surface winds data obtained from ECMWF's archive, called Meteorological Archival and Retrieval System (MARS), reprocessed OSCAT winds on a 25 km grid, and ERA5 forecast winds, all in 2013. OSCAT is a rotating pencil-beam Ku-band scatterometer carried by the Oceansat-2 satellite operated by the Indian Space Research Organization (ISRO). Oceansat-2 was launched 23 September 2009, and delivered data from 16 December 2009 until instrument breakdown on 14 February 2014. The data were reprocessed in 2021 at KNMI with the Pencil-beam Wind data Processor (PenWP) and collocated with the buoy and ERA-5 forecasts. The ERA5 forecasts were interpolated quadratically in time and bilinear in space to the time and position of the OSCAT measurements. The forecasts had a minimum lead time of three hours to avoid using forecasts in which the OSCAT data were assimilated. The collocation criteria were a maximum difference of 30 min in time and 17.7 km ($\sqrt{1/2}$ times the grid size) in position for the buoy and OSCAT measurements. See [14] for more information.

3. Methodology

3.1. Triple Collocation with Error Non-Orthogonality

The error model employed here is the canonical model

$$x_i = a_i(t + \varepsilon_i) + b_i \tag{1}$$

where x_i is the measurement by system i, $i = 1, 2, 3$, a_i the calibration scaling, b_i the calibration bias, t the common signal measured by all three systems (also referred to as "truth", but that term may be misleading as t is determined by the system with coarsest resolution), and ε_i a random measurement error with zero average and variance σ_i^2.

If ε_i consists of a constant part $\varepsilon_i^{(0)}$ and a t-dependent part $\varepsilon_i^{(1)}(t)$, $\varepsilon_i = \varepsilon_i^{(0)} + \varepsilon_i^{(1)}(t)$, the t-dependent part will be added to t in the error model (1). Compared with the other systems, the calibration of system i will appear different because the dependency x_i to t has changed. The error model thus implicitly removes error non-orthogonality, and it is not possible to retrieve $\varepsilon_i^{(1)}(t)$ from it. As a consequence, any information on error non-orthogonality must come from other sources.

Without loss of generality, system 1 can be chosen as the reference system relative to which systems 2 and 3 will be calibrated, so $a_1 = 1$ and $b_1 = 0$.

Taking first moments (averages) $M_i = \langle x_i \rangle$, where the brackets stand for taking the average (arithmetic mean) over all measurements, one obtains

$$M_i = a_i \langle t \rangle + b_i \tag{2}$$

since $\langle \varepsilon_i \rangle = 0$ because the random measurement errors have zero mean. Since $a_1 = 1$ and $b_1 = 0$, Equation (2) implies that

$$\langle t \rangle = M_1 \tag{3}$$

Substituting this back in Equation (2) yields

$$b_i = M_i - a_i M_1 \tag{4}$$

so that the calibration biases b_2 and b_3 are known once we know the calibration scalings.

To find the calibration scalings and the error variances σ_i^2 one must take second moments $M_{ij} = \langle x_i x_j \rangle$:

$$M_{ij} = a_i a_j \langle (t + \varepsilon_i)(t + \varepsilon_j) \rangle + a_i b_j \langle t + \varepsilon_i \rangle + a_j b_i \langle t + \varepsilon_j \rangle + b_i b_j \quad (5)$$

The terms with averaging brackets $\langle \ \rangle$ are handled as follows. From Equation (3) it follows that $\langle t + \varepsilon_i \rangle = M_1$. Further

$$\langle (t + \varepsilon_i)(t + \varepsilon_j) \rangle = \left\langle t^2 \right\rangle + \tau_j + \tau_i + e_{ij} \quad (6)$$

where $\tau_i = \langle t\varepsilon_i \rangle$ and $e_{ij} = \langle \varepsilon_i \varepsilon_j \rangle$. Note that $\tau_i = 0$ when the errors do not depend on the measured value, the usual assumption. In the literature this is called error orthogonality, so let's call τ_i the error non-orthogonality of system i. Further, note that $\sigma_i^2 = e_{ii}$. The calibration biases are removed by substituting Equation (4), and the result is

$$M_{ij} = a_i a_j \left\langle t^2 \right\rangle + a_i a_j (\tau_j + \tau_i + e_{ij}) + M_i M_j + a_i a_j M_1^2 \quad (7)$$

Introducing covariances $C_{ij} = \langle (x_i - M_i)(x_j - M_j) \rangle = M_{ij} - M_i M_j$ and common variance $T = \langle t^2 \rangle - M_1^2$, Equation (7) yields the covariance equations

$$C_{ij} = a_i a_j (T + \tau_i + \tau_j + e_{ij}) \quad (8)$$

The usual form of the covariance equations is $C_{ij} = a_i a_j (T + e_{ij})$. In Equation (8), two extra terms appear because the approximation of error orthogonality was not made. It should be noted here that the covariances C_{ij} are not unbiased; they can be made unbiased by multiplying them by a factor $n/(n-1)$, with n the number of collocations. When n is large, as is the case in this study with 44,948 collocations, the factor can safely be neglected.

For triple collocation, the error variances $\sigma_i^2 = e_{ii}$ follow from the three diagonal covariance equations as

$$\sigma_i^2 = \frac{C_{ii}}{a_i^2} - T - 2\tau_i \quad (9)$$

The three off-diagonal equations read

$$C_{12} = a_2 (T + \tau_1 + \tau_2 + e_{12}) \quad (10a)$$

$$C_{13} = a_3 (T + \tau_1 + \tau_3 + e_{13}) \quad (10b)$$

$$C_{23} = a_2 a_3 (T + \tau_2 + \tau_3 + e_{23}) \quad (10c)$$

Equations (10a)–(10c) are usually solved by assuming $\tau_i = e_{ij} = 0$, and the solution reads

$$T = \frac{C_{12} C_{13}}{C_{23}}, a_2 = \frac{C_{23}}{C_{13}}, a_3 = \frac{C_{23}}{C_{12}} \quad (11)$$

From Equations (10a)–(10c), it is clear that error non-orthogonality's and error covariances will affect the calibration scalings and common variance, and this may in turn affect the error variances.

3.2. Representativeness Errors

If the spatial, temporal or geophysical representation of the three systems is different, representativeness errors must be taken into account [1,2,15]. Suppose that the systems are ordered in decreasing spatial resolution with system 1 the buoys (calibration reference), system 2 the scatterometer, and system 3 the ERA-5 forecasts. Then systems 1 and 2 will measure a common signal r^2 that is not observed by system 3 because of its coarser resolution. This common signal is called the representativeness error or, sometimes, the representativeness signal. It will appear as an error correlation $e_{12} = r^2$ in Equation (10a).

The representativeness errors in this study are estimated from differences in spatial variances [15,16]. For the OSCAT and ERA5 data mentioned above, they are 0.181 $\mathrm{m}^2\,\mathrm{s}^{-2}$ for the zonal wind component u and 0.145 $\mathrm{m}^2\,\mathrm{s}^{-2}$ for the meridional wind component v.

3.3. Method of Solution

The triple collocation Equations (9) and (10) are solved in an iterative algorithm that starts by assuming that the systems are well intercalibrated. In each iteration the covariances C_{ij} are recalculated using the calibration coefficients from the previous iteration (or their starting values for the first iteration). Then the calibration coefficients are updated and outliers may be removed, see [15] for more details. The iteration converges in about 7 steps. The final values for the observation error variances σ_i^2 are for the calibrated data; those for the uncalibrated data can be obtained by dividing by a_i^2.

As representativeness errors and error non-orthogonalities must be obtained from other sources than the triple collocation analysis itself, they can be moved to the left-hand side of Equations (10a)–(10c), so the covariance Equation (8) become

$$C_{ij} - \frac{\tau_i + \tau_j + e_{ij}}{a_i a_j} = a_i a_j T \tag{12}$$

With the replacement

$$C_{ij} \to C_{ij} - \frac{\tau_i + \tau_j + e_{ij}}{a_i a_j} \tag{13}$$

the covariance equations can be solved as usual. In the solution scheme employed here, the calibration coefficients are known in each iteration, at least approximately, so the replacement Equation (13) can be applied without problems. Therefore, the effect of error non-orthogonalities can be studied in the same way as that of representativeness errors or error correlations by incorporating them in the covariances in each iteration step.

4. Results and Discussion

The effect of error non-orthogonality is studied by assuming that one of the observation systems contains an error non-orthogonality τ that ranges from -1 $\mathrm{m}^2\,\mathrm{s}^{-2}$ to +1 $\mathrm{m}^2\,\mathrm{s}^{-2}$, a range that is considerable with respect to the representativeness errors and the error variances. This is repeated for all three systems.

Figure 1a,b show the effect of error non-orthogonality in the buoys, τ_1, on the calibration biases, Figure 1c,d that on the calibration scalings, and Figure 1e,f that on the observation error variances for the zonal wind component u (Figure 1a,c,e) and the meridional wind component v (Figure 1b,d,f). Figures 2 and 3 are similar to Figure 1, but for error non-orthogonality in OSCAT and in the ECMWF forecast, τ_2 and τ_3, respectively. The figures show results for the observation error variances of the calibrated data.

In all figures, the calibration biases are only weakly affected by error non-orthogonality. Note that the calibration bias of the buoys equals 0 as they were selected as calibration reference. The effect on the bias depends on the dynamic range of t and hence differs for u and v.

More effect of error non-orthogonality can be seen in the calibration scalings, as expected. For error non-orthogonality in the buoys (Figure 1), the scaling of both scatterometer and ECMWF increases with increasing τ_1, while for error non-orthogonality in the scatterometer (Figure 2) or in the ECMWF forecasts (Figure 3), the calibration scaling of the system having error non-orthogonality decreases with τ, the other systems being unaffected.

This can be summarized as that error non-orthogonality only reduces the calibration scaling of the system having it. If that system happens to be the calibration reference, as in Figure 1, it will lift the other calibration scalings.

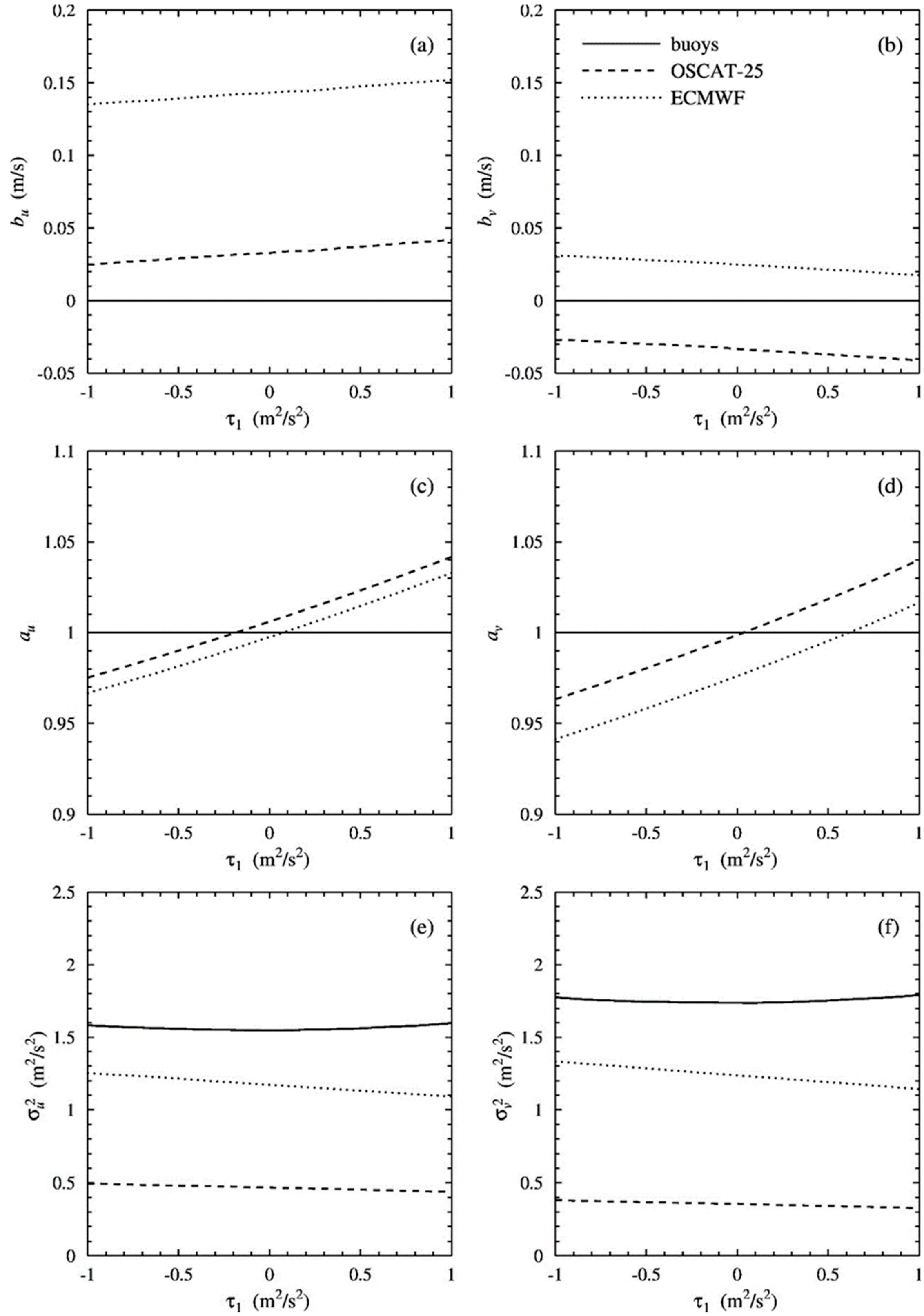

Figure 1. Effect of the error non-orthogonality τ_1 in the buoys (system 1) for the calibration bias in u (**a**), the calibration bias in v (**b**), the calibration scaling in u (**c**), the calibration scaling in v (**d**), the observation error variance in u (**e**), and the observation error variance in v (**f**).

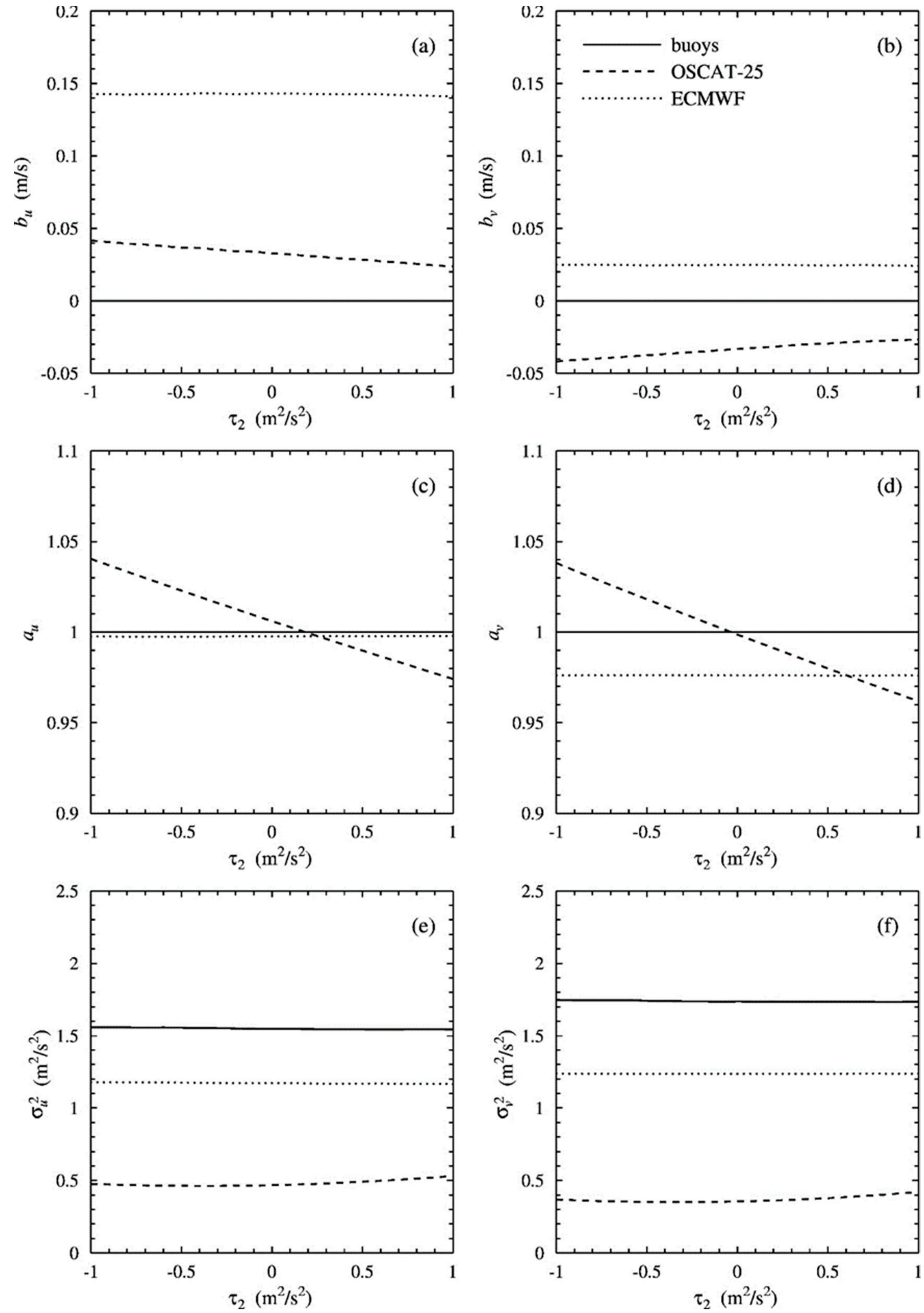

Figure 2. As Figure 1, but for error non-orthogonality τ_2 in OSCAT (system 2).

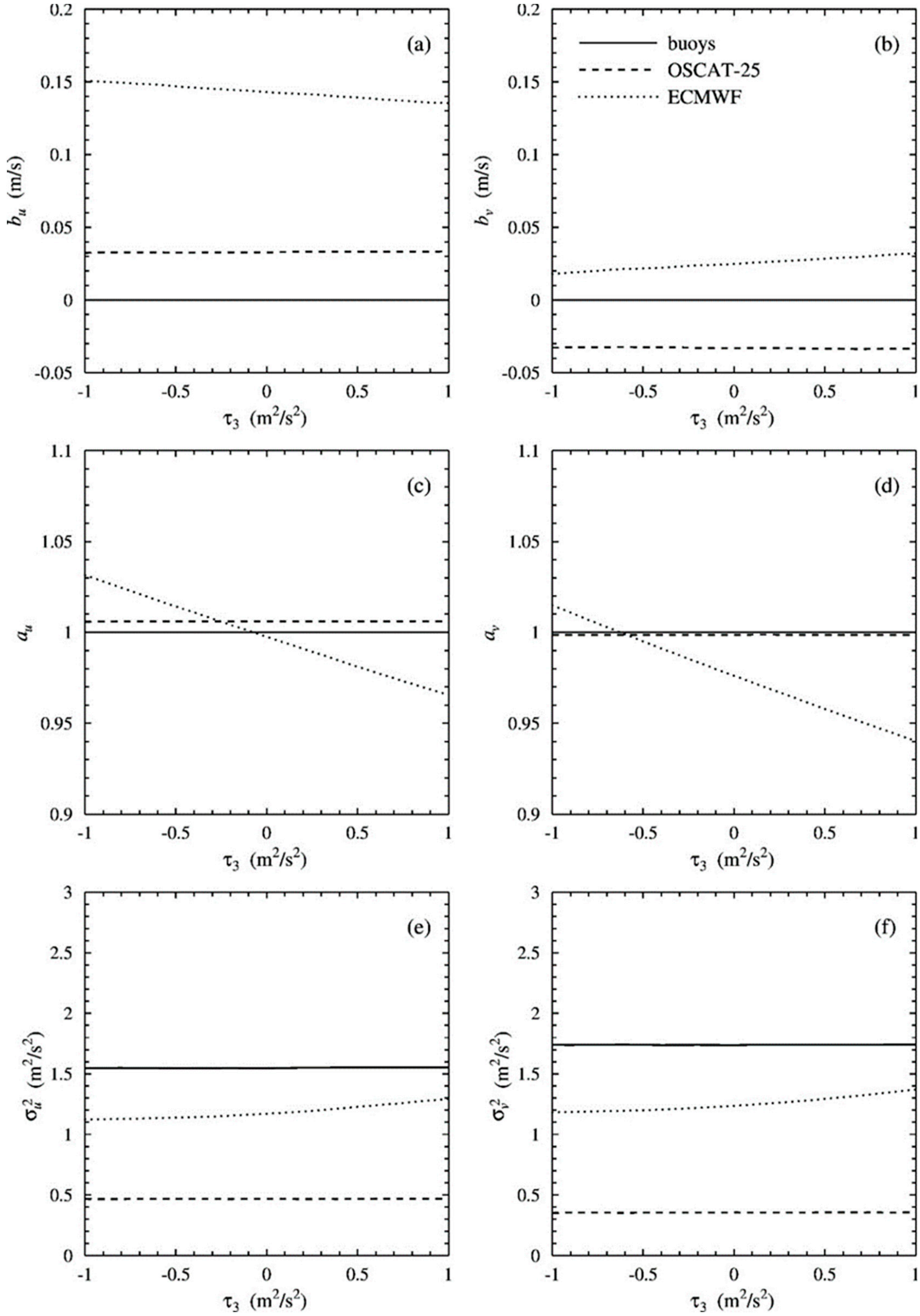

Figure 3. As Figure 1, but for error non-orthogonality τ_3 in the ECMWF forecasts (system 3).

The most surprising in Figures 1–3 is that the observation error variances are only very weakly affected by error non-orthogonality. Apparently, the effects of τ_j in Equations (9) and (10) largely cancel, and this can be understood by making a first-order Taylor expansion of the observation error variances σ_i^2 in τ_j,

$$\sigma_i^2(\tau_j) = \sigma_i^2(0) + \tau_j \left.\frac{d\sigma_i^2}{d\tau_j}\right|_{\tau_j=0} + O(\tau_j^2) \tag{14}$$

For example, let $j = 1$. From Equation (14) one readily finds that

$$\sigma_1^2(\tau_1) \approx \sigma_1^2(0) + \tau_1 \left(\frac{C_{12} + C_{13}}{C_{23}} - 2 \right) \tag{15}$$

Now the off-diagonal covariances are all equal to each other when the triple collocation algorithm has converged, because then the calibration scalings a_i are all equal to 1. Therefore σ_1^2 is independent of τ_1 to first order, and only dependencies of second and higher order remain. Similarly,

$$\sigma_2^2(\tau_1) \approx \sigma_2^2(0) + \tau_1 \frac{1}{C_{23}} \left(C_{12} + C_{13} - 2\frac{C_{13}C_{22}}{C_{23}} \right) \tag{16}$$

and the linear dependency of σ_2^2 on τ_1 vanishes because of the equality of the off-diagonal covariances. Finally,

$$\sigma_3^2(\tau_1) \approx \sigma_3^2(0) + \tau_1 \frac{1}{C_{23}} \left(C_{12} + C_{13} - 2\frac{C_{12}C_{33}}{C_{23}} \right) \tag{17}$$

because σ_3^2 follows from σ_2^2 by replacing C_{22} with C_{33} and interchanging C_{12} and C_{13}. The linear dependencies of σ_1^2, σ_2^2, and σ_3^2 on τ_2 and τ_3 can be shown to vanish in the same manner.

It must be stressed here that the observation error variances σ_i^2 are for the calibrated data. To retrieve the observation error variances in the uncalibrated data one must divide by a_i^2. As a result, the observation error variances of the uncalibrated data do depend on error non-orthogonality in a manner opposite to the calibration scalings.

5. Conclusions

Error non-orthogonality is included in an elegant and concise manner in the covariance equations for triple collocation. The formalism is applied on a data set consisting of all collocated wind velocities in 2013 from buoys, OSCAT-25, and ECMWF forecasts. Varying the error non-orthogonality of each system separately shows only weak dependence for the calibration biases and, in particular, the observation error variances. The observation error variances are sensitive to error non-orthogonality only in second and higher order. This implies that error non-orthogonality cannot be blamed for possible unexpected observation error variances from triple collocation analyses of ocean surface vector winds. Only the calibration scalings show a strong dependency on error non-orthogonality. The calibration scaling of the system with error non-orthogonality τ decreases with increasing τ, except for the reference system where the calibration scalings of all other systems will increase with increasing τ. This implies that error non-orthogonality is aliased with calibration error in scatter plots, at least qualitatively and for relatively large τ, where the calculated regression line deviates from the diagonal in the cloud of observed points for both cases.

Because of the generality of the covariance equations, as for higher-order collocations as elaborated in [15] as well, error non-orthogonality can also be studied for quadruple and higher-order collocations. However, the results are expected not to differ much from the results presented here.

Author Contributions: Conceptualization and methodology, J.V. and A.S.; software, J.V.; validation, J.V.; formal analysis, J.V.; investigation, J.V.; resources, A.S.; data curation, A.V.; writing—original draft preparation, J.V.; writing—review and editing, J.V., A.S., and A.V.; visualization, J.V.; supervision, A.S.; project administration, A.S. and A.V.; funding acquisition, A.S. and A.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by EUMETSAT within the framework of the Ocean and Sea Ice Satellite Application Facility (OSI SAF).

Data Availability Statement: The collocation data set can be obtained from the authors upon a request to scat@knmi.nl. The PenWP source code can be downloaded free of charge at https://nwp-saf.eumetsat.int/site/software/scatterometer/penwp/ (accessed on 25 August 2022). A simple free triple collocation code in Python can be found at https://github.com/knmiscat/triple_collocation (accessed on 25 August 2022).

Acknowledgments: The authors thank Jean Bidlot of ECMWF for his assistance in obtaining the buoy data and Weicheng Ni of the National University of Defense Technology in China for his interest in this work and stimulating discussions.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Stoffelen, A. Toward the True Near-surface Wind Speed: Error Modelling and Calibration using Triple Collocation. *J. Geophys. Res.* **1998**, *103*, 7755–7766. [CrossRef]
2. Vogelzang, J.; Stoffelen, A.; Verhoef, A.; Figa-Saldaña, J. On the Quality of High-resolution Scatterometer Winds. *J. Geophys. Res.* **2011**, *116*, C10033. [CrossRef]
3. Abdalla, S.; De Chiara, G. Estimating Random Errors of Scatterometer, Altimeter, and Model Wind Speed Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2406–2414. [CrossRef]
4. Danielson, R.E.; Johannessen, J.A.; Quartly, G.D.; Rio, M.-H.; Chapron, B.; Collard, F.; Donlon, C. Exploitation of Error Correlation in a Large Analysis Validation: GlobCurrent Case Study. *Rem. Sens. Env.* **2018**, *217*, 476–490. [CrossRef]
5. Hoareau, N.; Portabella, M.; Lin, W.; Ballabrera-Poy, J.; Turiel, A. Error Characterization of Sea Surface Salinity Products using Triple Collocation Analysis. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5160–5168. [CrossRef]
6. Roebeling, R.A.; Wolters, E.L.A.; Meirink, J.F.; Leijnse, H. Triple Collocation of Summer Precipitation Retrievals from SEVIRI over Europe with Gridded Rain Gauge and Weather Radar Data. *J. Hydrometeorol.* **2012**, *13*, 1552–1566. [CrossRef]
7. Wild, A.; Chua, Z.-W.; Kuleshov, Y. Triple Collocation Analysis of Satellite Precipitation Estimates over Australia. *Remote Sens.* **2022**, *14*, 2724. [CrossRef]
8. Gruber, A.; Su, C.-H.; Crow, W.T.; Zwieback, S.; Dorigo, W.A.; Wagner, W. Estimating Error Cross-correlations in Soil Moisture Data Sets using Extended Collocation Analysis. *J. Geophys. Res. Atmos.* **2016**, *121*, 1208–1219. [CrossRef]
9. Fan, X.; Lu, Y.; Liu, Y.; Li, T.; Xun, S.; Zhao, X. Validation of Multiple Soil Moisture Products over an Intensive Agricultural Region: Overall Accuracy and Diverse Responses to Precipitation and Irrigation Events. *Remote Sens.* **2022**, *14*, 3339. [CrossRef]
10. Su, K.; Zheng, W.; Yin, W.; Hu, L.; Shen, Y. Improving the Accuracy of Groundwater Storage Estimates Based on Groundwater Weighted Fusion Model. *Remote Sens.* **2022**, *14*, 202. [CrossRef]
11. McColl, K.A.; Vogelzang, J.; Konings, A.G.; Entekhabi, D.; Piles, M.; Stoffelen, A. Extended triple collocation: Estimating Errors and Correlation Coefficients with respect to an Unknown Target. *Geophys. Res. Lett.* **2014**, *41*, 6229–6236. [CrossRef]
12. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
13. Rivas, M.B.; Stoffelen, A. Characterizing ERA-Interim and ERA5 Surface Wind Biases using ASCAT. *Ocean Sci.* **2019**, *15*, 831–852. [CrossRef]
14. Verhoef, A.; Vogelzang, J.; Stoffelen, A. Scientific Validation Report (SVR) for the Ku-band Wind Data Records. In *OSI SAF Report SAF/OSI/CDOP3/KNMI/TEC/RP/415*; EUMETSAT: Darmstadt, Germany, 2022.
15. Vogelzang, J.; Stoffelen, A. Quadruple Collocation Analysis of In-situ, Scatterometer, and NWP Winds. *J. Geophys. Res. Ocean.* **2021**, *126*, e2021JC017189. [CrossRef]
16. Vogelzang, J.; King, G.P.; Stoffelen, A. Spatial Variances of Wind Fields and their Relation to Second-order Structure Functions and Spectra. *J. Geophys. Res. Ocean.* **2015**, *120*, 1048–1064. [CrossRef]