J Geod (2018) 92:59–80
https://doi.org/10.1007/s00190-017-1045-7

ORIGINAL ARTICLE

Distributional theory for the DIA method

P. J. G. Teunissen[1,2]

In memory of Willem Baarda on his 100th birthday

Received: 25 February 2017 / Accepted: 19 June 2017 / Published online: 6 July 2017
© The Author(s) 2017. This article is an open access publication

Abstract The DIA method for the detection, identification and adaptation of model misspecifications combines estimation with testing. The aim of the present contribution is to introduce a unifying framework for the rigorous capture of this combination. By using a canonical model formulation and a partitioning of misclosure space, we show that the whole estimation–testing scheme can be captured in one single DIA estimator. We study the characteristics of this estimator and discuss some of its distributional properties. With the distribution of the DIA estimator provided, one can then study all the characteristics of the combined estimation and testing scheme, as well as analyse how they propagate into final outcomes. Examples are given, as well as a discussion on how the distributional properties compare with their usage in practice.

Keywords Detection, Identification and Adaptation (DIA) · Tienstra transformation · Baarda test statistic · Misclosure partitioning · Voronoi-partitioning unit sphere · DIA estimator · Best linear unbiased estimation (BLUE) · Best linear unbiased prediction (BLUP) · Hazardous probability · Bias · Missed detection (MD) · Correct detection (CD) · Correct identification (CI)

✉ P. J. G. Teunissen
p.teunissen@curtin.edu.au

1 GNSS Research Centre, Curtin University of Technology, Perth, Australia

2 Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands

1 Introduction

The DIA method for the detection, identification and adaptation of model misspecifications, together with its associated internal and external reliability measures, finds its origin in the pioneering work of Baarda (1967, 1968a), see also e.g. Alberda (1976), Kok (1984), Teunissen (1985). For its generalization to recursive quality control in dynamic systems, see Teunissen and Salzmann (1989), Teunissen (1990). The DIA method has found its use in a wide range of applications, for example, for the quality control of geodetic networks (DGCC 1982), for geophysical and structural deformation analyses (Van Mierlo 1980; Kok 1982), for different GPS applications (Van der Marel and Kosters 1990; Teunissen 1998b; Tiberius 1998; Hewitson et al. 2004; Perfetti 2006; Drevelle and Bonnifait 2011; Fan et al. 2011) and for various configurations of integrated navigation systems (Teunissen 1989; Salzmann 1993; Gillissen and Elema 1996).

The DIA method combines estimation with testing. Parameter estimation is conducted to find estimates of the parameters one is interested in and testing is conducted to validate these results with the aim to remove any biases that may be present. The consequence of this practice is that the method is not one of estimation only, nor one of testing only, but actually one where estimation and testing are combined.

The aim of the present contribution is to introduce a unifying framework that captures the combined estimation and testing scheme of the DIA method. This implies that one has to take the intricacies of this combination into consideration when evaluating the contributions of the various decisions and estimators involved. By using a canonical model formulation and a partitioning of misclosure space, we show that the whole estimation–testing scheme can be captured in one single estimator $\bar{x}$. We study the characteristics of this estimator and discuss some of its distributional properties. With

the distribution of $\bar{x}$ provided, one can then study all the characteristics of the combined estimation and testing scheme, as well as analyse how they propagate into the final outcome $\bar{x}$.

This contribution is organized as follows. After a description of the null and alternative hypotheses considered, we derive the DIA estimator $\bar{x}$ in Sect. 2. We discuss its structure and identify the contributions from testing and estimation, respectively. We also discuss some of its variants, namely when adaptation is combined with a remeasurement of rejected data or when adaptation is only carried out for a subset of misclosure space. In Sect. 3, we derive the distribution of the estimator $\bar{x}$. As one of its characteristics, we prove that the estimator is unbiased under $\mathcal{H}_0$, but not under any of the alternative hypotheses. We not only prove this for estimation, but also for methods of prediction, such as collocation, Kriging and the BLUP. Thus although testing has the intention of removing biases from the solution, we prove that this is not strictly achieved. We show how the bias of $\bar{x}$ can be evaluated and on what contributing factors it depends.

In Sect. 4, we decompose the distribution conditionally over the events of missed detection, correct identification and wrong identification, thus providing insight into the conditional biases as well. We also discuss in this context the well-known concept of the minimal detectable bias (MDB). In order to avoid a potential pitfall, we highlight here that the MDB is about detection and not about identification. By using the same probability of correct detection for all alternative hypotheses $\mathcal{H}_a$, the MDBs can be compared and provide information on the sensitivity of rejecting the null hypothesis for $\mathcal{H}_a$-biases the size of their MDBs. The MDBs are therefore about correct detection and not about correct identification. This would only be true in the binary case, when next to the null hypothesis only a single alternative hypothesis is considered. Because of this difference between the univariate and multivariate case, we also discuss the probability of correct identification and associated minimal bias sizes.

In Sect. 5, we discuss ways of evaluating the DIA estimator. We make a distinction between unconditional and conditional evaluations and show how they relate to the procedures followed in practice. We point out, although the procedures followed in practice are usually a conditional one, that the conditional distribution itself is not strictly used in practice. In practice, any follow-on processing for which the outcome of $\bar{x}$ is used as input, the distribution of the estimator under the identified hypothesis is used without regards to the conditioning process that led to the identified hypothesis. We discuss this difference and show how it can be evaluated. Finally, a summary with conclusions is provided in Sect. 6. We emphasize that our development will be nonBayesian throughout. Hence, the only random vectors considered are the vector of observables y and functions thereof, while the unknown to-be-estimated parameter vector x and unknown bias vectors b_i are assumed to be deterministic.

2 DIA method and principles

2.1 Null and alternative hypotheses

Before any start can be made with statistical model validation, one needs to have a clear idea of the null and alternative hypotheses $\mathcal{H}_0$ and $\mathcal{H}_i$, respectively. The null hypothesis, also referred to as working hypothesis, consists of the model that one believes to be valid under normal working conditions. We assume the null hypothesis to be of the form

$$\mathcal{H}_0: \quad \mathsf{E}(y) = Ax \ , \ \mathsf{D}(y) = Q_{yy} \tag{1}$$

with $\mathsf{E}(.)$ the expectation operator, $y \in \mathbb{R}^m$ the normally distributed random vector of observables, $A \in \mathbb{R}^{m\times n}$ the given design matrix of rank n, $x \in \mathbb{R}^n$ the to-be-estimated unknown parameter vector, $\mathsf{D}(.)$ the dispersion operator and $Q_{yy} \in \mathbb{R}^{m\times m}$ the given positive-definite variance matrix of y. The *redundancy* of $\mathcal{H}_0$ is $r = m - \text{rank}(A) = m - n$.

Under $\mathcal{H}_0$, the best linear unbiased estimator (BLUE) of x is given as

$$\hat{x}_0 = A^+ y \tag{2}$$

with $A^+ = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ being the BLUE-inverse of A. As the quality of $\hat{x}_0$ depends on the validity of the null hypothesis, it is important that one has sufficient confidence in $\mathcal{H}_0$. Although every part of the null hypothesis can be wrong of course, we assume here that if a misspecification occurred it is confined to an underparametrization of the mean of y, in which case the alternative hypothesis is of the form

$$\mathcal{H}_i: \quad \mathsf{E}(y) = Ax + C_i b_i \ , \ \mathsf{D}(y) = Q_{yy} \tag{3}$$

for some vector $C_i b_i \in \mathbb{R}^m/\{0\}$, with $[A, C_i]$ a known matrix of full rank. Experience has shown that these types of misspecifications are by large the most common errors that occur when formulating the model. Through $C_i b_i$, one may model, for instance, the presence of one or more blunders (outliers) in the data, cycle slips in GNSS phase data, satellite failures, antenna-height errors, erroneous neglectance of atmospheric delays, or any other systematic effect that one failed to take into account under $\mathcal{H}_0$. As $\hat{x}_0$ (cf. 2) looses its BLUE-property under $\mathcal{H}_i$, the BLUE of x under $\mathcal{H}_i$ becomes

$$\hat{x}_i = \bar{A}^+ y \tag{4}$$

with $\bar{A}^+ = (\bar{A}^T Q_{yy}^{-1} \bar{A})^{-1} \bar{A}^T Q_{yy}^{-1}$ the BLUE-inverse of $\bar{A} = P_{C_i}^{\perp} A$ and $P_{C_i}^{\perp} = I_m - C_i (C_i^T Q_{yy}^{-1} C_i)^{-1} C_i^T Q_{yy}^{-1}$ being the orthogonal projector that projects onto the orthogonal complement of the range space of C_i. Note that orthogonality is here with respect to the metric induced by Q_{yy}.

As it is usually not only one single mismodelling error $C_i b_i$ one is potentially concerned about, but quite often many more than one, a testing procedure needs to be devised for handling the various, say k, alternative hypotheses $\mathcal{H}_i$, $i = 1, \ldots, k$. Such a procedure then usually consists of the following three steps of detection, identification and adaptation (DIA), (Baarda 1968a; Teunissen 1990; Imparato 2016):

Detection: An overall model test on $\mathcal{H}_0$ is performed to diagnose whether an unspecified model error has occurred. It provides information on whether one can have sufficient confidence in the assumed null hypothesis, without the explicit need to specify and test any particular alternative hypothesis. Once confidence in $\mathcal{H}_0$ has been declared, $\hat{x}_0$ is provided as the estimate of x.

Identification: In case confidence in the null hypothesis is lacking, identification of the potential source of model error is carried out. It implies the execution of a search among the specified alternatives $\mathcal{H}_i$, $i = 1, \ldots, k$, for the most likely model misspecification.

Adaptation: After identification of the suspected model error, a corrective action is undertaken on the $\mathcal{H}_0$-based inferences. With the null hypothesis rejected, the identified hypothesis, $\mathcal{H}_i$ say, becomes the new null hypothesis and $\hat{x}_i$ is provided as the estimate of x.

As the above steps illustrate, the outcome of testing determines how the parameter vector x will be estimated. Thus although estimation and testing are often treated separately and independently, in actual practice when testing is involved the two are intimately connected. This implies that one has to take the intricacies of this combination into consideration when evaluating the properties of the estimators involved. In order to help facilitate such rigorous propagation of the uncertainties, we first formulate our two working principles.

2.2 DIA principles

In the development of the distributional theory, we make use of the following two principles:

1. Canonical form: as validating inferences should remain invariant for one-to-one model transformations, use will be made of a canonical version of $\mathcal{H}_0$, thereby simplifying some of the derivations.
2. Partitioning: to have an unambiguous testing procedure, the $k+1$ hypotheses $\mathcal{H}_i$ are assumed to induce an unambiguous partitioning of the observation space.

2.2.1 Canonical model

We start by bringing (1) in canonical form. This is achieved by means of the *Tienstra*-transformation (Tienstra 1956)

$$\mathcal{T} = [A^{+T}, B]^T \in \mathbb{R}^{m \times m} \tag{5}$$

in which B is an $m \times r$ basis matrix of the null space of A^T, i.e. $B^T A = 0$ and $\text{rank}(B) = r$. The Tienstra transformation is a one-to-one transformation, having the inverse $\mathcal{T}^{-1} = [A, B^{+T}]$, with $B^+ = (B^T Q_{yy} B)^{-1} B^T Q_{yy}$. We have used the $\mathcal{T}$-transformation to canonical form also for LS-VCE (Teunissen and Amiri-Simkooei 2008) and for the recursive BLUE-BLUP (Teunissen and Khodabandeh 2013). Application of $\mathcal{T}$ to y gives under the null hypothesis (1),

$$\begin{bmatrix} \hat{x}_0 \\ t \end{bmatrix} = \mathcal{T} y \overset{\mathcal{H}_0}{\sim} \mathcal{N}\left(\begin{bmatrix} x \\ 0 \end{bmatrix}, \begin{bmatrix} Q_{\hat{x}_0 \hat{x}_0} & 0 \\ 0 & Q_{tt} \end{bmatrix} \right) \tag{6}$$

in which $\hat{x}_0 \in \mathbb{R}^n$ is the BLUE of x under $\mathcal{H}_0$ and $t \in \mathbb{R}^r$ is the vector of misclosures (the usage of the letter t for misclosure follows from the Dutch word '$t$egenspraak'). Their variance matrices are given as

$$\begin{aligned} Q_{\hat{x}_0 \hat{x}_0} &= (A^T Q_{yy}^{-1} A)^{-1} \\ Q_{tt} &= B^T Q_{yy} B \end{aligned} \tag{7}$$

As the vector of misclosures t is zero mean and stochastically independent of $\hat{x}_0$, it contains all the available information useful for testing the validity of $\mathcal{H}_0$. Note, since a basis matrix is not uniquely defined, that also the vector of misclosures is not uniquely defined. This is, however, of no consequence as the testing will only make use of the intrinsic information that is contained in the misclosures and hence will be invariant for any one-to-one transformation of t.

Under the alternative hypothesis (3), $\mathcal{T} y$ becomes distributed as

$$\begin{bmatrix} \hat{x}_0 \\ t \end{bmatrix} = \mathcal{T} y \overset{\mathcal{H}_i}{\sim} \mathcal{N}\left(\begin{bmatrix} I_n & A^+ C_i \\ 0 & B^T C_i \end{bmatrix} \begin{bmatrix} x \\ b_i \end{bmatrix}, \begin{bmatrix} Q_{\hat{x}_0 \hat{x}_0} & 0 \\ 0 & Q_{tt} \end{bmatrix} \right) \tag{8}$$

Thus, $\hat{x}_0$ and t are still independent, but now have different means than under $\mathcal{H}_0$. Note that $C_i b_i$ gets propagated differently into the means of $\hat{x}_0$ and t. We call

$$\begin{aligned} b_y &= C_i b_i = \text{observation bias} \\ b_{\hat{x}_0} &= A^+ b_y = \text{influential bias} \\ b_t &= B^T b_y = \text{testable bias} \end{aligned} \tag{9}$$

These biases are related to the observation space $\mathbb{R}^m$ as

$$\begin{aligned} b_y &= P_A b_y + P_A^{\perp} b_y \\ &= A b_{\hat{x}_0} + B^{+T} b_t \\ &= \mathcal{T}^{-1}[b_{\hat{x}_0}^T, b_t^T]^T \end{aligned} \tag{10}$$

with the orthogonal projectors $P_A = AA^+$ and $P_A^{\perp} = I_m - P_A$.

As the misclosure vector t has zero expectation under $\mathcal{H}_0$ (cf. 6), it is the component b_t of b_y that is testable. The component $b_{\hat{x}_0}$ on the other hand cannot be tested. As it will be directly absorbed by the parameter vector, it is this component of the observation bias b_y that directly influences the parameter solution $\hat{x}_0$. The *bias-to-noise ratios* (BNRs) of (9),

$$\lambda_y = ||b_y||_{Q_{yy}}, \lambda_{\hat{x}_0} = ||b_{\hat{x}_0}||_{Q_{\hat{x}_0\hat{x}_0}}, \lambda_t = ||b_t||_{Q_{tt}} \tag{11}$$

are related through the Pythagorian decomposition $\lambda_y^2 = \lambda_{\hat{x}_0}^2 + \lambda_t^2$ (see Fig. 1; note: here and in the following we use the notation $||.||_M^2 = (.)^T M^{-1}(.)$ for a weighted squared norm). As the angle ϕ between b_y and $P_A b_y$ determines how much of the observation bias is testable and influential, respectively, it determines the ratio of the testable BNR to the influential BNR: $\lambda_t = \lambda_{\hat{x}_0} \tan(\phi)$. The smaller the angle ϕ, the more b_y and $\mathcal{R}(A)$ are aligned and the more influential the bias will be. The angle itself is determined by the type of bias (i.e. matrix C_i) and the strength of the underlying model (i.e. matrices A and Q_{yy}). The BNRs (11) were introduced by Baarda in his reliability theory to determine measures of *internal* and *external* reliability, respectively (Baarda 1967, 1968b, 1976; Teunissen 2000). We discuss this further in Sect. 5.2.

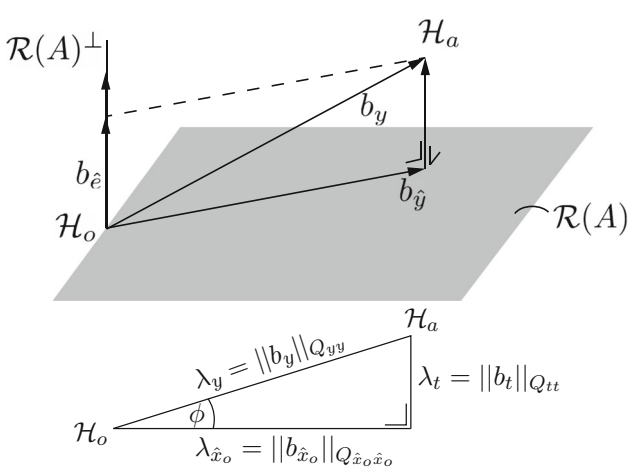

Fig. 1 The Pythagorian BNR decomposition $\lambda_y^2 = \lambda_{\hat{x}_0}^2 + \lambda_t^2$, with $\lambda_y = ||b_y||_{Q_{yy}}$, $\lambda_{\hat{x}_0} = ||b_{\hat{x}_0}||_{Q_{\hat{x}_0\hat{x}_0}} = ||Ab_{\hat{x}_0}||_{Q_{yy}}^2 = ||P_A b_y||_{Q_{yy}} = ||b_{\hat{y}}||_{Q_{yy}}^2$ and $\lambda_t = ||b_t||_{Q_{tt}} = ||P_A^{\perp} b_y||_{Q_{yy}}$ (Teunissen 2000)

Due to the canonical structure of (8), it now becomes rather straightforward to infer the BLUEs of x and b_i under $\mathcal{H}_i$. The estimator $\hat{x}_0$ will not contribute to the determination of the BLUE of b_i as $\hat{x}_0$ and t are independent and the mean of $\hat{x}_0$ now depends on more parameters than only those of x. Thus, it is t that is solely reserved for the determination of the BLUE of b_i, which then on its turn can be used in the determination of the BLUE of x under $\mathcal{H}_i$. The BLUEs of x and b_i under $\mathcal{H}_i$ are therefore given as

$$\begin{aligned} \hat{x}_i &= \hat{x}_0 - A^+ C_i \hat{b}_i \\ \hat{b}_i &= (B^T C_i)^+ t \end{aligned} \tag{12}$$

in which $(B^T C_i)^+ = (C_i^T B Q_{tt}^{-1} B^T C_i)^{-1} C_i^T B Q_{tt}^{-1}$ denotes the BLUE-inverse of $B^T C_i$. The result (12) shows how $\hat{x}_0$ is to be adapted when switching from the BLUE of $\mathcal{H}_0$ to that of $\mathcal{H}_i$. With it, we can now establish the following useful transformation between the BLUEs of x under $\mathcal{H}_i$ and $\mathcal{H}_0$,

$$\begin{bmatrix} \hat{x}_i \\ t \end{bmatrix} = \begin{bmatrix} I_n & -L_i \\ 0 & I_r \end{bmatrix} \begin{bmatrix} \hat{x}_0 \\ t \end{bmatrix} \tag{13}$$

in which

$$L_i = \begin{cases} 0 & \text{for} \quad i = 0 \\ A^+ C_i (B^T C_i)^+ & \text{for } i \neq 0 \end{cases} \tag{14}$$

Note that transformation (13) is in block-triangular form and that its inverse can be obtained by simply replacing $-L_i$ by $+L_i$.

The distribution of (13) under $\mathcal{H}_a$ is given as

$$\begin{aligned} \begin{bmatrix} \hat{x}_i \\ t \end{bmatrix} &\overset{\mathcal{H}_a}{\sim} \mathcal{N}\left(\begin{bmatrix} x + A^+ R_i C_a b_a \\ B^T C_a b_a \end{bmatrix}, \right. \\ &\left. \begin{bmatrix} Q_{\hat{x}_0\hat{x}_0} + L_i Q_{tt} L_i^T & -L_i Q_{tt} \\ -Q_{tt} L_i^T & Q_{tt} \end{bmatrix} \right) \end{aligned} \tag{15}$$

with the projector

$$R_i = I_m - C_i (B^T C_i)^+ B^T \tag{16}$$

This projector projects along $\mathcal{R}(C_i)$ and onto $\mathcal{R}(A, Q_{yy} B (B^T C_i)^{\perp})$, with $(B^T C_i)^{\perp}$ a basis matrix of the null space of $C_i^T B$. Note that $\hat{x}_{i \neq a}$ is biased under $\mathcal{H}_a$ with bias $\mathsf{E}(\hat{x}_{i \neq a} - x | \mathcal{H}_a) = A^+ R_{i \neq a} C_a b_a$. Hence, any part of $C_a b_a$ in $\mathcal{R}(A)$ gets directly passed on to the parameters and any part in $\mathcal{R}(C_{i \neq a})$ or in $\mathcal{R}(Q_{yy} B (B^T C_i)^{\perp})$ gets nullified.

2.2.2 Partitioning of misclosure space

The one-to-one transformation (13) clearly shows how the vector of misclosures t plays its role in linking the BLUEs of the different hypotheses. This relation does, however, not

yet incorporate the outcome of testing. To do so, we now apply our partitioning principle to the space of misclosures $\mathbb{R}^r$ and unambiguously assign outcomes of t to the estimators $\hat{x}_i$. Therefore, let $\mathcal{P}_i \subset \mathbb{R}^r$, $i = 0, 1, \ldots, k$, be a *partitioning* of the r-dimensional misclosure space, i.e. $\cup_{i=0}^{k} \mathcal{P}_i = \mathbb{R}^r$ and $\mathcal{P}_i \cap \mathcal{P}_j = \{0\}$ for $i \neq j$. Then, the unambiguous relation between t and $\hat{x}_i$ is established by defining the testing procedure such that $\mathcal{H}_i$ is selected if and only if $t \in \mathcal{P}_i$. An alternative way of seeing this is as follows. Let the unambiguous testing procedure be captured by the mapping $H : \mathbb{R}^r \mapsto \{0, 1, \ldots, k\}$, then the regions $\mathcal{P}_i = \{t \in \mathbb{R}^r \mid i = H(t)\}$, $i = 0, \ldots, k$, form a partition of misclosure space.

As the testing procedure is defined by the partitioning $\mathcal{P}_i \subset \mathbb{R}^r$, $i = 0, \ldots, k$, any change in the partitioning will change the outcome of testing and thus the quality of the testing procedure. The choice of partitioning depends on different aspects, such as the null and alternative hypotheses considered and the required detection and identification probabilities. In the next sections, we develop our distributional results such that it holds for any chosen partitioning of the misclosure space. However, to better illustrate the various concepts involved, we first discuss a few partitioning examples.

Example 1 (Partitioning: detection only) Let $\mathcal{H}_0$ be accepted if $t \in \mathcal{P}_0$, with

$$\mathcal{P}_0 = \{t \in \mathbb{R}^r \mid ||t||_{Q_{tt}} \leq \tau \in \mathbb{R}^+\} \tag{17}$$

If testing is restricted to detection only, then the complement of $\mathcal{P}_0$, $\mathcal{P}_1 = \mathbb{R}^r / \mathcal{P}_0$ becomes the region for which no parameter solution is provided. Thus, $\mathcal{P}_0$ and $\mathcal{P}_1$ partition misclosure space with the outcomes: $\hat{x}_0$ if $t \in \mathcal{P}_0$ and 'solution unavailable' if $t \in \mathcal{P}_1$. The probabilities of these outcomes, under $\mathcal{H}_0$ resp. $\mathcal{H}_a$, can be computed using the Chi-square distribution $||t||^2_{Q_{tt}} \overset{\mathcal{H}_a}{\sim} \chi^2(r, \lambda_t^2)$, with $\lambda_t^2 = ||P_A^{\perp} C_a b_a||^2_{Q_{yy}}$ (Arnold 1981; Koch 1999; Teunissen 2000). For $\mathcal{H}_0$, b_a is set to zero. □

Example 2 (Partitioning: one-dimensional identification only) Let the design matrices $[A, C_i]$ of the k hypotheses $\mathcal{H}_i$ (cf. 3) be of order $m \times (n+1)$, denote $C_i = c_i$ and $B^T c_i = c_{t_i}$, and write Baarda's test statistic (Baarda 1967, 1968b; Teunissen 2000) as

$$|w_i| = ||P_{c_{t_i}} t||_{Q_{tt}} \tag{18}$$

in which $P_{c_{t_i}} = c_{t_i}(c_{t_i}^T Q_{tt}^{-1} c_{t_i})^{-1} c_{t_i}^T Q_{tt}^{-1}$ is the orthogonal projector that projects onto c_{t_i}. If testing is restricted to identification only, then $\mathcal{H}_{i \neq 0}$ is identified if $t \in \mathcal{P}_{i \neq 0}$, with

$$\mathcal{P}_{i \neq 0} = \{t \in \mathbb{R}^r \mid |w_i| = \max_{j \in \{1, \ldots, k\}} |w_j|\}\ ,\ i = 1, \ldots, k \tag{19}$$

Also this set forms a partitioning of misclosure space, provided not two or more of the vectors c_{t_i} are the same. When projected onto the unit sphere, this partitioning can be shown to become a Voronoi partitioning of $\mathbb{S}^{r-1} \subset \mathbb{R}^r$. The Voronoi partitioning of a set of unit vectors $\bar{c}_j$, $j = 1, \ldots, k$, on the unit sphere $\mathbb{S}^{r-1}$ is defined as

$$V_i = \{\bar{t} \in \mathbb{S}^{r-1} \mid\ d(\bar{t}, \bar{c}_i) \leq d(\bar{t}, \bar{c}_j),\ j = 1, \ldots, k\} \tag{20}$$

in which the metric $d(u, v) = \cos^{-1}(u^T v)$ is the *geodesic distance* (great arc circle) between the unit vectors u and v (unicity is here defined in the standard Euclidean metric). If we now define the unit vectors

$$\bar{t} = Q_{tt}^{-1/2} t / ||t||_{Q_{tt}} \quad \text{and} \quad \bar{c}_i = Q_{tt}^{-1/2} c_{t_i} / ||c_{t_i}||_{Q_{tt}} \tag{21}$$

we have $|w_i| = ||t||_{Q_{tt}} \bar{c}_i^T \bar{t}$ and therefore $\max_i |w_i| \Leftrightarrow \min_i d(\bar{t}, \bar{c}_i)$, thus showing that

$$V_i = \{\bar{t} \in \mathbb{S}^{r-1} \mid\ \bar{t} = Q_{tt}^{-\frac{1}{2}} t / ||t||_{Q_{tt}}, t \in \mathcal{P}_i\} \tag{22}$$

For the distributions of t and $\bar{t}$ under $\mathcal{H}_a$, we have

$$t \overset{\mathcal{H}_a}{\sim} \mathcal{N}(\mu_t, Q_{tt}) \quad \text{and} \quad \bar{t} \overset{\mathcal{H}_a}{\sim} \mathcal{PN}(\mu_{\bar{t}}, I_r) \tag{23}$$

with $\mu_t = B^T C_a b_a$ and $\mu_{\bar{t}} = Q_{tt}^{-1/2} B^T C_a b_a$. Thus, $\bar{t}$ has a *projected normal* distribution, which is unimodal and rotationally symmetric with respect to its mean direction $\mu_{\bar{t}} / ||\mu_{\bar{t}}||_{I_r}$. The scalar $||\mu_{\bar{t}}||_{I_r} = \lambda_t$ is a measure of the peakedness of the PDF. Thus, the larger the testable BNR λ_t is, the more peaked the PDF of $\bar{t}$ becomes. The density of $\bar{t}$ is given in, e.g. Watson (1983). Under $\mathcal{H}_0$, when $\mu_{\bar{t}} = 0$, the PDF of $\bar{t}$ reduces to

$$f_{\bar{t}}(\tau | \mathcal{H}_0) = \frac{\Gamma(\frac{r}{2})}{2\pi^{\frac{r}{2}}} \tag{24}$$

which is one over the surface area of the unit sphere $\mathbb{S}^{r-1}$. Hence, under the null hypothesis, the PDF of $\bar{t}$ is the *uniform* distribution on the unit sphere. The selection probabilities under $\mathcal{H}_0$ are therefore given as $\mathsf{P}(t \in \mathcal{P}_i | \mathcal{H}_0) = |V_i| \frac{\Gamma(\frac{r}{2})}{2\pi^{\frac{r}{2}}}$, in which $|V_i|$ denotes the surface area covered by V_i. □

An important practical application of one-dimensional identification is *datasnooping*, i.e. the procedure in which the individual observations are screened for possible outliers (Baarda 1968a; DGCC 1982; Kok 1984). If we restrict ourselves to the case of one outlier at a time, then the c_i-vector of the alternative hypothesis takes the form of a canonical unit vector having 1 as its ith entry and zeros elsewhere. This would then lead to $k = m$ regions $\mathcal{P}_i$ provided not two or more of the vectors $c_{t_i} = B^T c_i$ are the same. If the latter happens only one of them should be retained, as no difference

between such hypotheses can then be made. This would, for instance, be the case when datasnooping is applied to levelled height differences of a single levelling loop.

Example 3 (Partitioning: datasnooping) To illustrate the datasnooping partitioning on the unit sphere, we consider the GPS single-point positioning pseudorange model,

$$A = \begin{bmatrix} u_1^T & 1 \\ \vdots & \vdots \\ u_m^T & 1 \end{bmatrix} \quad , \quad Q_{yy} = \sigma^2 I_m \tag{25}$$

in which the u_i, $i = 1, \ldots, m$, are the receiver-satellite unit direction vectors of the m satellites. First we assume that the receiver position is known, thus reducing the design matrix to $A = [1, \ldots, 1]^T$. Because of the symmetry in this model, one can expect the unit vectors $\bar{c}_i$ (cf. 21), $i = 1, \ldots, m$, to have a symmetric distribution over the unit sphere and thus the datasnooping partitioning to be symmetric as well. Indeed, for the correlation between the w_i-statistics and therefore for the angle between the $\bar{c}_i$ vectors, we have

$$\rho_{w_i w_j} = \cos \angle(\bar{c}_i, \bar{c}_j) = \frac{1}{m-1} \ , \ i \neq j. \tag{26}$$

Thus for $m = 3$, we have $\cos^{-1}(\frac{1}{2}) = 60°$ and for $m = 4$, we have $\cos^{-1}(\frac{1}{3}) = 70.53°$. The partitioning of this latter case is shown in Fig. 2 (top). In case the receiver position is unknown, the receiver-satellite geometry comes into play through the variations in the unit direction vectors u_i of (25). In that case, we expect a varying partitioning on the unit sphere. This is illustrated in Fig. 2 (bottom) for the case $m = 7, n = 4$. □

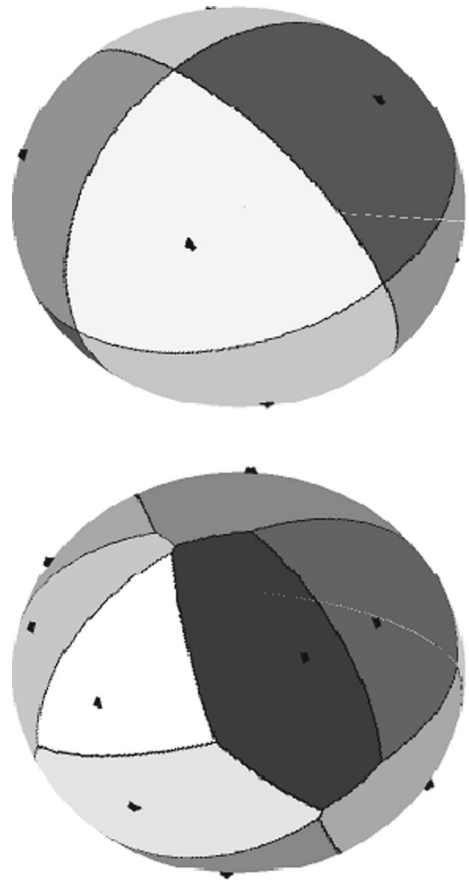

Fig. 2 Datasnooping partitioning on the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^{r=3}$ for two cases of the GPS single-point positioning model: (*top*) position constrained; (*bottom*) position unconstrained

Example 4 (Partitioning: detection and identification) Again let the design matrices $[A, C_i]$ of the k hypotheses $\mathcal{H}_i$ (cf. 3) be of order $m \times (n+1)$ and now consider detection and identification. Then

$$\mathcal{P}_0 = \{t \in \mathbb{R}^r \mid ||t||_{Q_{tt}} \leq \tau \in \mathbb{R}^+\} \tag{27}$$

and

$$\mathcal{P}_{i \neq 0} = \{t \in \mathbb{R}^r \mid ||t||_{Q_{tt}} > \tau; |w_i| = \max_{j \in \{1,\ldots,k\}} |w_j|\} \tag{28}$$

form a partitioning of the misclosure space, provided not two or more of the vectors c_{t_i} are the same. An example of such partitioning is given in Fig. 3 for $r = 2$, $k = 3$, and $\tau^2 = \chi^2_\alpha(r = 2, 0)$. The inference procedure induced by this partitioning is thus that the null hypothesis gets accepted (not rejected) if $||t||_{Q_{tt}} \leq \tau$, while in case of rejection, the largest value of the statistics $|w_j|$, $j = 1, \ldots, k$, is used for identifying the alternative hypothesis. In the first case, $\hat{x}_0$ is provided as the estimate of x, while in the second case, $\hat{x}_i = \hat{x}_0 - L_i t$ (cf. 13) is provided for the identified alternative. The false- alarm selection probabilities under $\mathcal{H}_0$ are now given as $\mathsf{P}(t \in \mathcal{P}_{i \neq 0} | \mathcal{H}_0) = \alpha |V_i| \frac{\Gamma(\frac{r}{2})}{2\pi^{\frac{r}{2}}}$, in which α is the overall level of significance. Note, the more correlated two w-test statistics w_i and w_j are, the smaller the angle $\angle(c_{t_i}, c_{t_j})$ (see Fig. 3) and the more difficult it will be to discern between the two hypotheses $\mathcal{H}_i$ and $\mathcal{H}_j$, especially for small biases, see e.g. Foerstner (1983), Tiberius (1998), Yang et al. (2013b). □

In the above examples, the partitioning is formulated in terms of the misclosure vector t. It can, however, also be formulated by means of the least-squares residual vector $\hat{e}_0 = y - A\hat{x}_0$, thus providing a perhaps more recognizable form of testing. As $t \in \mathbb{R}^r$ and $\hat{e}_0 \in \mathbb{R}^m$ are related as $t = B^T \hat{e}_0$, we have (Teunissen 2000)

$$||t||^2_{Q_{tt}} = ||\hat{e}_0||^2_{Q_{yy}} \tag{29}$$

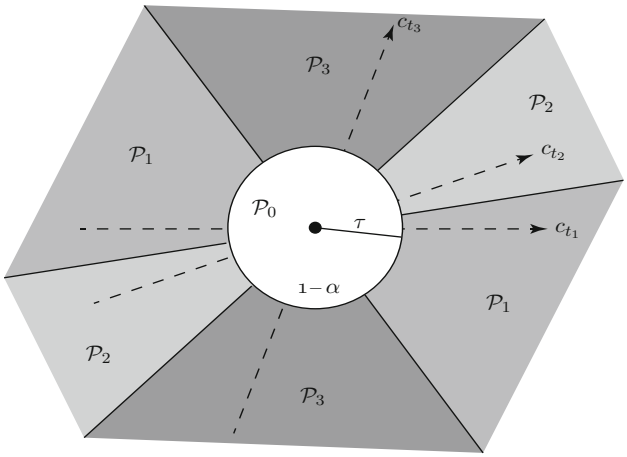

Fig. 3 Example of a partitioning of misclosure space $\mathbb{R}^{r=2}$ for $k = 3$. The *circular region* centred at the origin is $\mathcal{P}_0$, while the remaining three sectors, having their borders at the bisector axes of adjacent c_{t_i}'s, represent the regions $\mathcal{P}_i$ for $i = 1, 2, 3$. Note, since the figure has been drawn in the metric of Q_{tt}, that the detection region $\mathcal{P}_0$ is shown as a circle instead of an ellipse

and

$$|w_i| = ||P_{c_{t_i}} t||_{Q_{tt}} = \frac{|c_i^T Q_{yy}^{-1} \hat{e}_0|}{\sqrt{c_i^T Q_{yy}^{-1} Q_{\hat{e}_0 \hat{e}_0} Q_{yy}^{-1} c_i}} \tag{30}$$

Thus, the actual verification in which of the regions $\mathcal{P}_i$ the misclosure vector t lies can be done using the least-squares residual vector $\hat{e}_0$ obtained under $\mathcal{H}_0$, without the explicit need of having to compute t.

Note that in case of uncorrelated observations, i.e. $Q_{yy} = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$, the adapted design matrix $\bar{A}_{(i)} = P_{c_i}^{\perp} A$ for computing the BLUE under $\mathcal{H}_i$ is the original design matrix with its ith row replaced by zeros. Hence, in case of datasnooping, the BLUE $\hat{x}_i = \bar{A}_{(i)}^{+} y = \hat{x}_0 - A^{+} c_i \hat{b}_i$ is then the estimator with the ith observable *excluded*. Such is, for instance, done in exclusion-based RAIM (Kelly 1998; Yang et al. 2013a).

As a final note of this subsection, we remark that also other than the above partitioning examples can be given. For instance, in case of $\mathcal{H}_a$'s with one-dimensional biases, one can also find examples in which the ellipsoidal detection region (27) has been replaced by the intersection of half spaces, $\{t \in \mathbb{R}^r \,|\, |w_i| \leq c, i = 1, \ldots, k\}$, see e.g. Joerger and Pervan (2014), Imparato (2016), Teunissen et al. (2017). In fact, it is important to recognize that in the multivariate case of multihypotheses testing, no clear optimality results are available. Baarda's test statistic w_i and the detector $||t||_{Q_{tt}}$ are known to provide for uniformly most powerful invariant (UMPI) testing when $\mathcal{H}_0$ is tested against a *single* alternative (Arnold 1981; Teunissen 2000; Kargoll 2007), but not necessarily when multiple alternatives are in play. To be able to infer the quality of the various possible partitionings, one should therefore be able to diagnose their impact on the actual output of the DIA method.

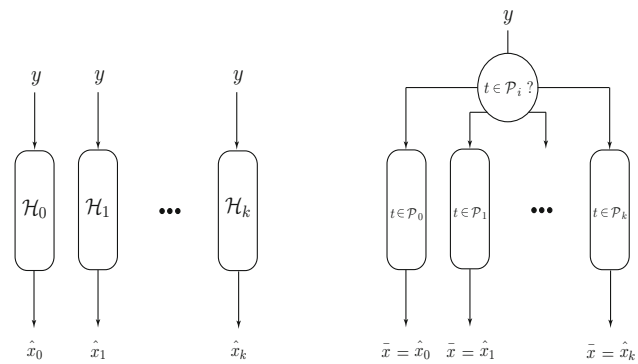

Fig. 4 The DIA estimator $\bar{x} = \sum_{i=0}^{k} \hat{x}_i \, p_i(t)$: (*left*) BLUEs $\hat{x}_i$ of individual $\mathcal{H}_i$'s; (*right*) outcomes of $\bar{x}$ in dependence on testing

2.3 The DIA estimator

With the DIA method, the outcome of testing determines how the parameter vector x gets estimated. As the outcome of such estimation is influenced by the testing procedure, one cannot simply assign the properties of $\hat{x}_0$ or $\hat{x}_i$ to the actual DIA estimator computed. That is, the actual estimator that is produced is not $\hat{x}_0$ nor $\hat{x}_i$, but in fact (see Fig. 4)

$$\bar{x} = \begin{cases} \hat{x}_0 & \text{if } t \in \mathcal{P}_0 \\ \hat{x}_i & \text{if } t \in \mathcal{P}_{i \neq 0} \end{cases} \tag{31}$$

By making use of the *indicator* functions $p_i(t)$ of the regions $\mathcal{P}_i$ (i.e. $p_i(t) = 1$ for $t \in \mathcal{P}_i$ and $p_i(t) = 0$ elsewhere), the DIA estimator $\bar{x}$ can be written in the compact form

$$\bar{x} = \sum_{i=0}^{k} \hat{x}_i \, p_i(t). \tag{32}$$

This expression shows how the $\hat{x}_i$s and t contribute to the estimator $\bar{x}$. If we now make use of the transformation (13), we can obtain its counterpart for $\bar{x}$ as

$$\begin{bmatrix} \bar{x} \\ t \end{bmatrix} = \begin{bmatrix} I_n & -\bar{L}(t) \\ 0 & I_r \end{bmatrix} \begin{bmatrix} \hat{x}_0 \\ t \end{bmatrix} \tag{33}$$

with $\bar{L}(t) = \sum_{i=1}^{k} L_i \, p_i(t)$. This important result shows how $[\bar{x}^T, t^T]^T$ stands in a one-to-one relation to $[\hat{x}_0^T, t^T]^T = \mathcal{T} y$ and thus to the original vector of observables y. Although the structure of (33) resembles that of the linear transformation (13), note that (33) is a *nonlinear* transformation due to the presence of t in $\bar{L}(t)$. Hence, this implies that the DIA estimator $\bar{x}$ will *not* have a normal distribution, even if all

the individual estimators $\hat{x}_i$, $i = 0, \ldots, k$, are normally distributed. In the next section, we derive the probability density function (PDF) of $\bar{x}$. But before we continue, we first briefly describe two variations on the above-defined DIA procedure.

Remeasurement included In case of datasnooping, with $Q_{yy} = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$, one may take the standpoint in certain applications that rejected observations should be remeasured. After all, one may argue, the measurement setup was designed assuming the rejected observation included. If one remeasures and replaces the rejected observation, say y_i, with the remeasured one, say $\bar{y}_i$, one is actually not producing $\hat{x}_i$ as output, but instead the solution for x based on the following extension of the model under $\mathcal{H}_i$,

$$\mathsf{E}\begin{bmatrix}\hat{x}_i\\ \bar{y}_i\end{bmatrix} = \begin{bmatrix} I_n \\ a_i^T\end{bmatrix} x \ , \ \mathsf{D}\begin{bmatrix}\hat{x}_i\\ \bar{y}_i\end{bmatrix} = \begin{bmatrix} Q_{\hat{x}_i\hat{x}_i} & 0 \\ 0 & \sigma_i^2\end{bmatrix} \tag{34}$$

with a_i^T the ith row of A and σ_i^2 the variance of the rejected and remeasured observations. The BLUE of x under this model is

$$\hat{\hat{x}}_i = \hat{x}_i + A^+ c_i(\bar{y}_i - a_i^T \hat{x}_i). \tag{35}$$

Hence, in case the remeasurement step is included, the computed DIA estimator is the one obtained by replacing $\hat{x}_i$ by $\hat{\hat{x}}_i$ in (32).

Undecided included Above, each hypothesis $\mathcal{H}_i$ was given its own region $\mathcal{P}_i$, such that together these regions cover the whole misclosure space, $\cup_{i=0}^k \mathcal{P}_i = \mathbb{R}^r$. This implies, whatever the outcome of t, that one always will be producing one of the estimates $\hat{x}_i$, $i = 0, \ldots, k$, even, for instance, if it would be hard to discriminate between some of the hypotheses or when selection is unconvincing. To accommodate such situations, one can generalize the procedure and introduce an undecided region $\Omega \subset \mathbb{R}^r$ for which no estimator of x is produced at all when $t \in \Omega$. Thus when that happens, the decision is made that a solution for x is unavailable,

$$\bar{x} = \text{unavailable} \ \text{ if } \ t \in \Omega \subset \mathbb{R}^r \tag{36}$$

This is similar in spirit to the undecided regions of the theory of *integer aperture estimation* (Teunissen 2003a). As a consequence of (36), the regions $\mathcal{P}_i$ of each of the hypotheses have become smaller and now only partition a part of the misclosure space,

$$\cup_{i=0}^k \mathcal{P}_i = \mathbb{R}^r / \Omega \tag{37}$$

As an example, consider the following generalization of (28),

$$\bar{\mathcal{P}}_{i\neq 0} = \mathcal{P}_{i\neq 0} \cap \{t \in \mathbb{R}^r \mid ||t||^2_{Q_{tt}} - w_i^2 \leq \bar{\tau}^2\} \tag{38}$$

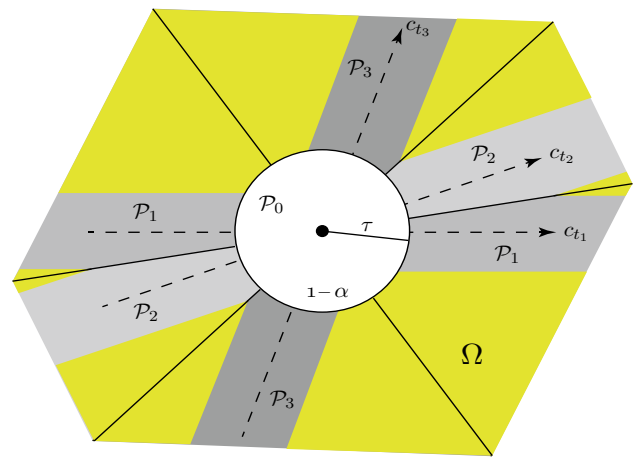

Fig. 5 Example of undecided region Ω (*yellow*) for t far away from $\mathcal{H}_0$ and all $\mathcal{H}_i$s. Compare with Fig. 3

Now it is also tested whether the maximum w-test statistic achieves a sufficient reduction in $||t||^2_{Q_{tt}}$, i.e. whether $||P^{\perp}_{c_{t_i}} t||^2_{Q_{tt}} = ||t||^2_{Q_{tt}} - w_i^2$ is small enough. If this is not the case, then the undecided decision is made, see Fig. 5.

The choice made for the undecided region Ω may of course affect the regions $\mathcal{P}_{i\neq 0}$ of some hypotheses more than others. For instance, if one or more of the alternative hypotheses turn out to be too poorly identifiable, one may choose to have their regions $\mathcal{P}_i$ completely assigned to the undecided region Ω. In that case, one would only proceed with identification for a subset of the k alternative hypotheses. In the limiting special case when all alternative hypotheses are considered too poorly identifiable, the undecided strategy would become one for which $\Omega = \mathbb{R}^r / \mathcal{P}_0$. In this case, one thus computes $\hat{x}_0$ if $\mathcal{H}_0$ gets accepted, but states that the solution is unavailable otherwise. As a result, the testing procedure is confined to the detection step. This was, for instance, the case with earlier versions of RAIM which had detection but no exclusion functionality, see e.g. Parkinson and Axelrad (1988), Sturza (1988).

To conclude this section, we pause a moment to further highlight some of the intricacies of the estimator (32). As the construction of $\bar{x}$ has been based on a few principles only, it is important to understand that the estimator describes the outcome of *any* DIA method. In it, we recognize the separate contributions of $p_i(t)$ and $\hat{x}_i$. Both contribute to the uncertainty or randomness of $\bar{x}$. The uncertainty of testing, i.e. of detection and identification, is channelled through $p_i(t)$, while the uncertainty of estimation is channelled through $\hat{x}_i$. Their combined outcome provides $\bar{x}$ as an estimator of x. It is hereby important to realize, however, that there are for now no a priori 'optimality' properties that one can assign to $\bar{x}$, despite the fact that its constituents do have some of such properties. The estimator $\hat{x}_0$, for instance, is optimal under $\mathcal{H}_0$ as it is then the BLUE of x. And in case of a single

alternative hypothesis ($k = 1$), also the testing can be done in an optimal way, namely by using uniformly most powerful invariant tests. These properties, however, are individual properties that do not necessarily carry over to $\bar{x}$. One may ask oneself, for instance, why use $\hat{x}_i$ when $\mathcal{H}_i$ is selected. Why not use, instead of $\hat{x}_i$, an estimator that takes the knowledge of $t \in \mathcal{P}_i$ into account. Also note, as the testing itself gives discrete outcomes, that the DIA estimator is a *binary* weighted average of all of the $k+1$ $\hat{x}_i$s. But one may wonder whether this binary weighting is the best one can do if the ultimate goal is the construction of a good estimator of x. For instance, although the weights $p_i(t)$ are binary in case of the DIA estimator, smoother weighting functions of the misclosures will provide for a larger class of estimators that may contain estimators with better performances for certain defined criteria. This in analogy with integer (I) and integer-equivariant (IE) estimation, for which the latter provides a larger class containing the optimal BIE estimator (Teunissen 2003b). And just like the nonBayesian BIE estimator was shown to have a Bayesian counterpart (Teunissen 2003b; Betti et al. 1993), the nonBayesian DIA estimator with smoother weights may find its counterpart in methods of Bayesian and information-theoretic multimodel inference (Burnham and Anderson 2002).

Answering these and similar questions on 'optimizing' the estimator $\bar{x}$ is complex and not the goal of the present contribution. The aim of the present contribution is to provide a general framework that captures the testing and estimation characteristics in a combined way through the single estimator $\bar{x}$. For that purpose, we present distributional properties of the DIA estimator $\bar{x}$ in the next and following sections, thus making a rigorous quality evaluation of any estimator of the form of (32) possible.

3 The distribution of the DIA estimator

3.1 The joint, conditional and marginal PDFs

In order to be able to study the properties of the DIA estimator, we need its probability distribution. As its performance is driven for a large part by the misclosure vector t, we determine the joint PDF $f_{\bar{x},t}(x,t)$ and the conditional PDF $f_{\bar{x}|t}(x|t)$, next to the marginal PDF $f_{\bar{x}}(x)$. We express their PDFs in the PDFs $f_{\hat{x}_0}(x)$ and $f_t(t)$ of $\hat{x}_0$ and t, respectively. We have the following result.

Theorem 1 (PDFs of $\bar{x}$ and t) *Let $\bar{x}$ be given as (32), with the $\hat{x}_i$s related to $\hat{x}_0$ and t as in (13). Then, the joint, conditional and marginal PDFs of the DIA estimator $\bar{x}$ and misclosure vector t can be expressed in the PDFs $f_{\hat{x}_0}(x)$ and $f_t(t)$ as*

$$f_{\bar{x},t}(x,t) = \sum_{i=0}^{k} f_{\hat{x}_0}(x + L_i t) f_t(t) p_i(t)$$

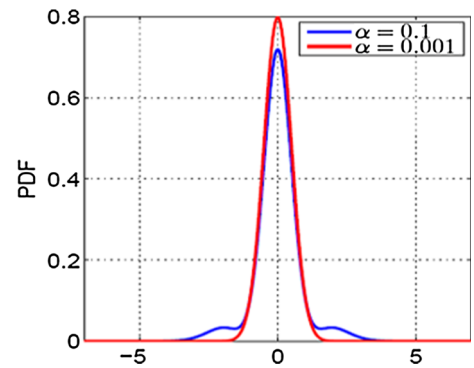

Fig. 6 The PDF $f_{\bar{x}}(x|\mathcal{H}_0)$ of the DIA estimator $\bar{x}$ under $\mathcal{H}_0$ (Example 5) for two values of α

$$f_{\bar{x}|t}(x|t) = \sum_{i=0}^{k} f_{\hat{x}_0}(x + L_i t) p_i(t)$$

$$f_{\bar{x}}(x) = \sum_{i=0}^{k} \int_{\mathcal{P}_i} f_{\hat{x}_0}(x + L_i \tau) f_t(\tau) d\tau \tag{39}$$

Proof See Appendix. □

This result shows that the PDF of $\bar{x}$ is constructed from a sum with shifted versions $f_{\hat{x}_0}(x + L_i t)$ of the PDF of $\hat{x}_0$. In fact, it is a weighted average of these shifted functions,

$$f_{\bar{x}}(x) = \sum_{i=0}^{k} \mathsf{E}(f_{\hat{x}_0}(x + L_i t) p_i(t)), \tag{40}$$

thus showing how it is driven by the distribution of the misclosure vector t. It is thus indeed a nonnormal distribution, which only will approach one when the PDF of the misclosures is sufficiently peaked. For instance, when $f_t(t) = \delta(t - \tau)$ and $\tau \in \mathcal{P}_j$, then $f_{\bar{x}}(x) = f_{\hat{x}_0}(x + L_j \tau)$.

Example 5 (PDF for $k = n = r = 1$) It follows from (39) that in case of only one alternative hypothesis ($k = 1$), the PDF of $\bar{x}$ under $\mathcal{H}_0$ can be written as

$$f_{\bar{x}}(x|\mathcal{H}_0) = f_{\hat{x}_0}(x|\mathcal{H}_0) + \int_{\mathbb{R}/\mathcal{P}_0} [f_{\hat{x}_0}(x + L_1 \tau|\mathcal{H}_0) - f_{\hat{x}_0}(x|\mathcal{H}_0)] f_t(\tau|\mathcal{H}_0) \mathrm{d}\tau \tag{41}$$

Let $\hat{x}_0 \overset{\mathcal{H}_0}{\sim} \mathcal{N}(0, \sigma^2_{\hat{x}_0} = 0.5)$, $t \overset{\mathcal{H}_0}{\sim} \mathcal{N}(0, \sigma^2_t = 2)$ (thus $n = r = 1$), and $L_1 = \frac{1}{2}$, with acceptance interval $\mathcal{P}_0 = \{\tau \in \mathbb{R} | (\tau/\sigma_t)^2 \leq \chi^2_\alpha(1,0)\}$. Figure 6 shows this PDF for $\alpha = 0.1$ and $\alpha = 0.001$. Note that $f_{\bar{x}}(x|\mathcal{H}_0)$, like $f_{\hat{x}_0}(x|\mathcal{H}_0)$, is symmetric about 0, but that its tails become heavier the larger α gets. □

Quite often in practice one is not interested in the complete parameter vector $x \in \mathbb{R}^n$, but rather only in certain functions

of it, say $\theta = F^T x \in \mathbb{R}^p$. As its DIA estimator is then computed as $\bar{\theta} = F^T \bar{x}$, we need its distribution to evaluate its performance.

Corollary 1 *Let $\bar{\theta} = F^T \bar{x}$ and $\hat{\theta}_0 = F^T \hat{x}_0$. Then, the PDF of $\bar{\theta}$ is given as*

$$f_{\bar{\theta}}(\theta) = \sum_{i=0}^{k} \int_{\mathcal{P}_i} f_{\hat{\theta}_0}(\theta + F^T L_i \tau) f_t(\tau) \mathrm{d}\tau. \tag{42}$$

Although we will be working with $\bar{x}$, instead of $\bar{\theta}$, in the remaining of this contribution, it should be understood that the results provided can similarly be given for $\bar{\theta} = F^T \bar{x}$ as well.

We also remark that here and in the remaining of this contribution, a $k+1$ partitioning covering the whole misclosure space, $\mathbb{R}^r = \cup_{i=0}^{k} \mathcal{P}_i$ is used. Thus, no use is made of an undecided region. The results that we present are, however, easily adapted to include such event as well. For instance, in case the $\mathcal{P}_i$ do not cover the whole misclosure space and $\Omega = \mathbb{R}^r / \cup_{i=0}^{k} \mathcal{P}_i$ becomes the undecided region, then the computed estimator is conditioned on $t \in \Omega^c = \cup_{i=0}^{k} \mathcal{P}_i$ and its PDF reads as,

$$f_{\bar{x}|t\in\Omega^c}(x) = \frac{\sum_{i=0}^{k} \int_{\mathcal{P}_i} f_{\hat{x}_0}(x + L_i \tau) f_t(\tau) \mathrm{d}\tau}{\mathsf{P}(t \in \Omega^c)}. \tag{43}$$

The approach for obtaining such undecided-based results is thus to consider the undecided region $\Omega \subset \mathbb{R}^r$ as the $(k+2)$th region that completes the partitioning of $\mathbb{R}^r$, followed by the analysis of the DIA estimator $\bar{x}$ conditioned on Ω^c, the complement of Ω.

3.2 The mean of $\bar{x}$ under $\mathcal{H}_0$ and $\mathcal{H}_a$

The estimators $\hat{x}_i$, $i = 0, 1, \ldots, k$, (cf. 2, 4) are BLUEs and therefore unbiased under their respective hypotheses, e.g. $\mathsf{E}(\hat{x}_0|\mathcal{H}_0) = x$ and $\mathsf{E}(\hat{x}_a|\mathcal{H}_a) = x$. However, as shown earlier, these are not the estimators that are actually computed when testing is involved. In that case, it is the DIA estimator $\bar{x}$ that is produced. As unbiasedness is generally a valued property of an estimator, it is important to know the mean of $\bar{x}$. It is given in the following theorem.

Theorem 2 (Mean of DIA estimator) *The mean of $\bar{x}$ under $\mathcal{H}_0$ and $\mathcal{H}_a$ is given as*

$$\begin{aligned} \mathsf{E}(\bar{x}|\mathcal{H}_0) &= x \\ \mathsf{E}(\bar{x}|\mathcal{H}_a) &= x + A^+ \bar{b}_{y_a} \end{aligned} \tag{44}$$

with

$$\begin{aligned} \bar{b}_{y_a} &= C_a b_a - \sum_{i=1}^{k} C_i \beta_i(b_a) \\ \beta_i(b_a) &= (B^T C_i)^+ \mathsf{E}(t p_i(t)|\mathcal{H}_a) \end{aligned} \tag{45}$$

where

$$\begin{aligned} \mathsf{E}(t p_i(t)|\mathcal{H}_a) &= \int_{\mathcal{P}_i} \tau f_t(\tau|\mathcal{H}_a) \mathrm{d}\tau \\ &\propto \int_{\mathcal{P}_i} \tau \exp\{-\frac{1}{2}||\tau - B^T C_a b_a||^2_{Q_{tt}}\} \mathrm{d}\tau \end{aligned} \tag{46}$$

Proof See Appendix. □

The theorem generalizes the results of Teunissen et al. (2017). It shows that $\bar{x}$, like $\hat{x}_0$, is unbiased under $\mathcal{H}_0$, but that, contrary to $\hat{x}_a$, $\bar{x}$ is *always* biased under the alternative,

$$\mathsf{E}(\bar{x}|\mathcal{H}_0) = \mathsf{E}(\hat{x}_0|\mathcal{H}_0) \quad \text{and} \quad \mathsf{E}(\bar{x}|\mathcal{H}_a) \neq \mathsf{E}(\hat{x}_a|\mathcal{H}_a) \tag{47}$$

It is thus important to realize that the result of adaptation will always produce an estimator that is biased under the alternative. Moreover, this is true for any form the adaptation may take. It holds true for, for instance, exclusion-based RAIM (Brown 1996; Kelly 1998; Hewitson and Wang 2006; Yang et al. 2013a), but also when adaptation is combined with remeasurement (cf. 35).

To evaluate the bias, one should note the difference between $\mathsf{E}(\bar{x}|\mathcal{H}_a)$ and $\mathsf{E}(\hat{x}_0|\mathcal{H}_a) = x + A^+ C_a b_a$,

$$\mathsf{E}(\bar{x}|\mathcal{H}_a) = \mathsf{E}(\hat{x}_0|\mathcal{H}_a) - A^+ \sum_{i=1}^{k} C_i \beta_i(b_a). \tag{48}$$

This expression shows the positive impact of testing. Without testing, one would produce $\hat{x}_0$, even when $\mathcal{H}_a$ would be true, thus giving the influential bias $A^+ C_a b_a$ (cf. 9). However, with testing included, the second term in (48) is present as a correction for the bias of $\hat{x}_0$ under $\mathcal{H}_a$. The extent to which the bias will be corrected for depends very much on the distribution of the misclosure vector t. In the limit, when $f_t(t|\mathcal{H}_a)$ is sufficiently peaked at the testable bias $B^T C_a b_a$ (cf. Theorem 2), the second term in (48) reduces to $A^+ C_a b_a$, thereby removing all the bias from $\hat{x}_0$ under $\mathcal{H}_a$. A summary of the means of $\hat{x}_0$, $\bar{x}$ and $\hat{x}_a$ is given in Table 1.

Table 1 The mean of $\hat{x}_0$, $\bar{x}$ and $\hat{x}_a$ under $\mathcal{H}_0$ and $\mathcal{H}_a$, with $b_{y_a} = C_a b_a$ and $\bar{b}_{y_a} = b_{y_a} - \sum_{i=1}^{k} C_i \beta_i(b_a)$

	$\mathcal{H}_0$	$\mathcal{H}_a$
$\hat{x}_0$	$\mathsf{E}(\hat{x}_0\|\mathcal{H}_0) = x$	$\mathsf{E}(\hat{x}_0\|\mathcal{H}_a) = x + A^+ b_{y_a}$
$\bar{x}$	$\mathsf{E}(\bar{x}\|\mathcal{H}_0) = x$	$\mathsf{E}(\bar{x}\|\mathcal{H}_a) = x + A^+ \bar{b}_{y_a}$
$\hat{x}_a$	$\mathsf{E}(\hat{x}_a\|\mathcal{H}_0) = x$	$\mathsf{E}(\hat{x}_a\|\mathcal{H}_a) = x$

Example 6 (Bias in $\bar{x}$) Let $\mathcal{H}_0$ have a single redundancy ($r = 1$) and let there be only a single alternative $\mathcal{H}_a$ ($k = 1$). Then, (45) simplifies to

$$\bar{b}_{y_a} = c_a \bar{b}_a \quad \text{with} \quad \bar{b}_a = b_a - \beta_a(b_a). \tag{49}$$

For the partitioning, we have: $\mathcal{P}_0$ being an origin-centred interval and $\mathcal{P}_a = \mathbb{R}/\mathcal{P}_0$. The size of $\mathcal{P}_0$ is determined through the level of significance α as $\mathsf{P}(t \in \mathcal{P}_0|\mathcal{H}_0) = 1-\alpha$. In the absence of testing (i.e. $\mathcal{P}_0 = \mathbb{R}$), the bias would be $\bar{b}_a = b_a$. In the presence of testing however, we have $\bar{b}_a \leq b_a$ for every value of $b_a > 0$, thus showing the benefit of testing: the bias that remains after testing is always smaller than the original bias. This benefit, i.e. the difference between $\bar{b}_a$ and b_a, will kick in after the bias b_a has become large enough. The difference is small, when b_a is small, and it gets larger for larger b_a, with $\bar{b}_a$ approaching zero again in the limit. For smaller levels of significance α, the difference between $\bar{b}_a$ and b_a will stay small for a larger range of b_a values. This is understandable as a smaller α corresponds with a larger acceptance interval $\mathcal{P}_0$, as a consequence of which one would have for a larger range of b_a values an outcome of testing that does not differ from the no-testing scenario. □

One can also compare the weighted mean squared errors of the three estimators $\hat{x}_0$, $\bar{x}$ and $\hat{x}_a$ under $\mathcal{H}_a$. By making use of the fact that $\mathsf{E}(||u||^2_Q) = \text{trace}(Q^{-1}Q_{uu}) + \mu^T Q^{-1}\mu$ if $\mathsf{E}(u) = \mu$ and $\mathsf{D}(u) = Q_{uu}$ (Koch 1999), we have the following result.

Corollary 2 (Mean squared error) *Let $b_{\hat{x}_0} = A^+ b_y$ and $b_{\bar{x}} = A^+ \bar{b}_y$ be the bias in $\hat{x}_0$ and $\bar{x}$, respectively, and $Q_{\hat{x}_a\hat{x}_a} = Q_{\hat{x}_0\hat{x}_0} + L_a Q_{tt} L_a^T$ the variance matrix of $\hat{x}_a$. Then*

$$\begin{aligned}
&\mathsf{E}(||\hat{x}_0 - x||^2_{Q_{\hat{x}_0\hat{x}_0}}|\mathcal{H}_a) = n + ||b_{\hat{x}_0}||^2_{Q_{\hat{x}_0\hat{x}_0}} \\
&\mathsf{E}(||\bar{x} - x||^2_{Q_{\hat{x}_0\hat{x}_0}}|\mathcal{H}_a) = \text{trace}(Q^{-1}_{\hat{x}_0\hat{x}_0} Q_{\bar{x}\bar{x}}) + ||b_{\bar{x}}||^2_{Q_{\hat{x}_0\hat{x}_0}} \\
&\mathsf{E}(||\hat{x}_a - x||^2_{Q_{\hat{x}_0\hat{x}_0}}|\mathcal{H}_a) = n + ||Q_{yy}^{-\frac{1}{2}} P_A C_a Q^{\frac{1}{2}}_{\hat{b}_a\hat{b}_a}||^2_F
\end{aligned} \tag{50}$$

with $||.||^2_F = \text{trace}(.)^T(.)$ denoting the squared Frobenius norm.

Note that for $C_a = c_a$,

$$\frac{\mathsf{E}(||\hat{x}_0 - x||^2_{Q_{\hat{x}_0\hat{x}_0}}|\mathcal{H}_a) - n}{\mathsf{E}(||\hat{x}_a - x||^2_{Q_{\hat{x}_0\hat{x}_0}}|\mathcal{H}_a) - n} = \left(\frac{b_a}{\sigma_{\hat{b}_a}}\right)^2 = \lambda_t^2. \tag{51}$$

3.3 Bias in BLUP, Kriging and collocation

We now make a brief side step and show that also methods of prediction, like collocation, Kriging and best linear unbiased prediction (BLUP), are affected by the bias $\bar{b}_{y_a}$ of Theorem 2. Let our prediction–aim be to predict the random vector $z = A_z x + e_z$, in which e_z is zero mean and matrices A_z, $Q_{zy} = \mathsf{Cov}(z, y)$ and $Q_{zz} = \mathsf{Cov}(z, z) = \mathsf{D}(z)$ are known. Then, we may extend our $\mathcal{H}_i$-model (3) as

$$\mathcal{H}_i : \mathsf{E}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} A & C_i \\ A_z & 0 \end{bmatrix}\begin{bmatrix} x \\ b_i \end{bmatrix}, \ \mathsf{D}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} Q_{yy} & Q_{yz} \\ Q_{zy} & Q_{zz} \end{bmatrix} \tag{52}$$

With the $\mathcal{H}_i$-BLUEs $\hat{x}_i$ and $\hat{b}_i$, the BLUP of z under $\mathcal{H}_i$ is then given as

$$\hat{z}_i = A_z\hat{x}_i + Q_{zy}Q_{yy}^{-1}(y - A\hat{x}_i - C_i\hat{b}_i) \tag{53}$$

But with the DIA method in place to validate the different hypotheses, the actual output of the combined estimation–testing–prediction process is then given as

$$\begin{aligned}
\bar{z} &= \sum_{j=0}^{k} \hat{z}_j p_j(t) \\
&= A_z\bar{x} + Q_{zy}Q_{yy}^{-1}\left(y - A\bar{x} - \sum_{j=1}^{k} C_j\hat{b}_j p_j(t)\right)
\end{aligned} \tag{54}$$

As the aim is to predict $z = A_z x + e_z$, it is of interest to know the difference $\mathsf{E}(\bar{z} - A_z x|\mathcal{H}_a)$. The following result shows how the bias in $\bar{z}$ can be expressed in the vector $\bar{b}_{y_a}$ of (45).

Corollary 3 (Bias in DIA predictor) *The bias in the DIA predictor $\bar{z} = \sum_{i=0}^{k} \hat{z}_i p_i(t)$ is given as*

$$b_{\bar{z}} = \mathsf{E}(\bar{z} - A_z x|\mathcal{H}_a) = A_z A^+ \bar{b}_{y_a} + Q_{zy}Q_{yy}^{-1}P_A^{\perp}\bar{b}_{y_a} \tag{55}$$

with $P_A^{\perp} = I_m - AA^+$.

4 Decomposition of probabilities

4.1 Correct and incorrect decisions

The decisions in the testing procedure are driven by the outcome of the misclosure vector t, i.e. choose $\mathcal{H}_i$ if $t \in \mathcal{P}_i$. Such decision is correct if $\mathcal{H}_i$ is true, and it is incorrect otherwise. The probabilities of such occurrences can be put into a *probability matrix*:

$$P_{ij} = \mathsf{P}(t \in \mathcal{P}_i|H_j) \quad , i, j = 0, 1, \ldots, k \tag{56}$$

As we are working with a partitioning of misclosure space, each of the columns of this matrix sums up to one, $\sum_{i=0}^{k} P_{ij} = 1, \forall j$. Ideally one would like this matrix to be as diagonally dominant as possible, i.e. P_{ij} large for $i = j$

(correct decision) and small for $i \neq j$ (wrong decision). By making use of the translational property of the PDF of t under $\mathcal{H}_j$ and $\mathcal{H}_0$ (cf. 8), the entries of the probability matrix can be computed as

$$P_{ij} = \int_{\mathcal{P}_i} f_t(\tau - B^T C_j b_j | \mathcal{H}_0) \mathrm{d}\tau. \tag{57}$$

Instead of addressing all entries of this probability matrix, we concentrate attention to those of the first column, i.e. of the null hypothesis $\mathcal{H}_0$, and of an arbitrary second column $a \in \{1, \ldots, k\}$, i.e. of alternative hypothesis $\mathcal{H}_a$. We then discriminate between two sets of events, namely under $\mathcal{H}_0$ and under $\mathcal{H}_a$. Under $\mathcal{H}_0$ we consider,

$$\begin{aligned} \mathrm{CA} &= (t \in \mathcal{P}_0 | \mathcal{H}_0) = \text{correct acceptance} \\ \mathrm{FA} &= (t \notin \mathcal{P}_0 | \mathcal{H}_0) = \text{false alarm.} \end{aligned} \tag{58}$$

Their probabilities are given as

$$\begin{aligned} &\mathsf{P}_{\mathrm{CA}} = \mathsf{P}(t \in \mathcal{P}_0 | \mathcal{H}_0) = \textstyle\int_{\mathcal{P}_0} f_t(\tau | \mathcal{H}_0) \mathrm{d}\tau \\ &\mathsf{P}_{\mathrm{FA}} = \mathsf{P}(t \notin \mathcal{P}_0 | \mathcal{H}_0) = \textstyle\int_{\mathbb{R}^r / \mathcal{P}_0} f_t(\tau | \mathcal{H}_0) \mathrm{d}\tau \\ &1 = \mathsf{P}_{\mathrm{CA}} + \mathsf{P}_{\mathrm{FA}} \end{aligned} \tag{59}$$

These probabilities are usually user defined, for instance by setting the probability of false alarm (i.e. level of significance) to a certain required small value $\mathsf{P}_{\mathrm{FA}} = \alpha$. In Example 1, this would, for instance, mean choosing τ^2 as the $(1-\alpha)$-percentile point of the central Chi-square distribution with r degrees of freedom, $\tau^2 = \chi^2_\alpha(r, 0)$.

For the events under $\mathcal{H}_a$, we make a distinction between detection and identification. For detection, we have

$$\begin{aligned} \mathrm{MD} &= (t \in \mathcal{P}_0 | \mathcal{H}_a) = \text{missed detection} \\ \mathrm{CD} &= (t \notin \mathcal{P}_0 | \mathcal{H}_a) = \text{correct detection} \end{aligned} \tag{60}$$

with their probabilities given as

$$\begin{aligned} &\mathsf{P}_{\mathrm{MD}} = \mathsf{P}(t \in \mathcal{P}_0 | \mathcal{H}_a) = \textstyle\int_{\mathcal{P}_0} f_t(\tau | \mathcal{H}_a) \mathrm{d}\tau \\ &\mathsf{P}_{\mathrm{CD}} = \mathsf{P}(t \notin \mathcal{P}_0 | \mathcal{H}_a) = \textstyle\int_{\mathbb{R}^r / \mathcal{P}_0} f_t(\tau | \mathcal{H}_a) \mathrm{d}\tau \\ &1 = \mathsf{P}_{\mathrm{MD}} + \mathsf{P}_{\mathrm{CD}} \end{aligned} \tag{61}$$

Although we refrained, for notational convenience, to give these probabilities an additional identifying index a, it should be understood that they differ from alternative to alternative. For each such $\mathcal{H}_a$, the probability of missed detection can be evaluated as

$$\mathsf{P}_{\mathrm{MD}} = \int_{\mathcal{P}_0} f_t(\tau - B^T C_a b_a | \mathcal{H}_0) \mathrm{d}\tau \tag{62}$$

Note, as b_a is unknown, that this requires assuming the likely range of values b_a can take.

With identification following detection, we have

$$\begin{aligned} \mathrm{WI} &= (t \in \cup_{j \neq 0,a}^k \mathcal{P}_j | \mathcal{H}_a) = \text{wrong identification} \\ \mathrm{CI} &= (t \in \mathcal{P}_a | \mathcal{H}_a) = \text{correct identification} \end{aligned} \tag{63}$$

with their probabilities given as

$$\begin{aligned} &\mathsf{P}_{\mathrm{WI}} = \mathsf{P}(t \in \cup_{j \neq 0,a}^k \mathcal{P}_j | \mathcal{H}_a) = \textstyle\sum_{j \neq 0,a}^k \int_{\mathcal{P}_j} f_t(\tau | \mathcal{H}_a) \mathrm{d}\tau \\ &\mathsf{P}_{\mathrm{CI}} = \mathsf{P}(t \in \mathcal{P}_a | \mathcal{H}_a) = \textstyle\int_{\mathcal{P}_a} f_t(\tau | \mathcal{H}_a) \mathrm{d}\tau \\ &\mathsf{P}_{\mathrm{CD}} = \mathsf{P}_{\mathrm{WI}} + \mathsf{P}_{\mathrm{CI}} \end{aligned} \tag{64}$$

Note, as the three probabilities of missed detection, wrong identification and correct identification sum up to unity,

$$1 = \mathsf{P}_{\mathrm{MD}} + \mathsf{P}_{\mathrm{WI}} + \mathsf{P}_{\mathrm{CI}} \tag{65}$$

that a small missed detection probability does not necessarily imply a large probability of correct identification. This would only be the case if there would be a *single* alternative hypothesis ($k = 1$). In that case, the probability of wrong identification is identically zero, $\mathsf{P}_{\mathrm{WI}} = 0$, and we have $\mathsf{P}_{\mathrm{CI}} = 1 - \mathsf{P}_{\mathrm{MD}}$. In the general case of multiple hypotheses ($k > 1$), the available probability of correct detection is spread out over all alternative hypotheses, whether correct or wrong, thus diminishing the probability of correct identification. It is up to the designer of the testing system to ensure that P_{CI} remains large enough for the application at hand. We have more to say about $\mathsf{P}_{\mathrm{CI}} \neq 1 - \mathsf{P}_{\mathrm{MD}}$ when discussing the concept of minimal detectable biases in Sect. 4.4.

4.2 The bias decomposition of $\bar{x}$

We will now use the events of missed detection (MD), correct detection (CD), wrong identification (WI) and correct identification (CI), to further study the mean of $\bar{x}$. We have already shown that the DIA estimator is biased under $\mathcal{H}_a$ (cf. 44). To study how this bias gets distributed over the different events of testing, we define

$$\begin{aligned} b_{\bar{x}} &= \mathsf{E}(\bar{x} - x | \mathcal{H}_a) \\ b_{\bar{x}|\mathrm{MD}} &= \mathsf{E}(\bar{x} - x | t \in \mathcal{P}_0, \mathcal{H}_a) \\ b_{\bar{x}|\mathrm{CD}} &= \mathsf{E}(\bar{x} - x | t \notin \mathcal{P}_0, \mathcal{H}_a) \\ b_{\bar{x}|\mathrm{WI}} &= \mathsf{E}(\bar{x} - x | t \in \cup_{j \neq 0,a}^k \mathcal{P}_j, \mathcal{H}_a) \\ b_{\bar{x}|\mathrm{CI}} &= \mathsf{E}(\bar{x} - x | t \in \mathcal{P}_a, \mathcal{H}_a) \end{aligned} \tag{66}$$

By application of the total probability rule for expectation, $\mathsf{E}(\bar{x}) = \sum_{j=0}^k \mathsf{E}(\bar{x} | t \in \mathcal{P}_j) P(t \in \mathcal{P}_j)$, we may write the unconditional bias $b_{\bar{x}}$ as the 'weighted average',

$$b_{\bar{x}} = b_{\bar{x}|\mathrm{MD}} \mathsf{P}_{\mathrm{MD}} + \underbrace{b_{\bar{x}|\mathrm{WI}} \mathsf{P}_{\mathrm{WI}} + b_{\bar{x}|\mathrm{CI}} \mathsf{P}_{\mathrm{CI}}}_{b_{\bar{x}|\mathrm{CD}} \mathsf{P}_{\mathrm{CD}}} \tag{67}$$

This expression shows how the unconditional bias is formed from its conditional counterparts. These conditional biases of the DIA estimator are given in the following theorem.

Theorem 3 (Conditional biases) *The conditional counterparts of the unconditional bias* $b_{\bar{x}} = A^+[C_a b_a - \sum_{i=1}^k C_i \beta_i(b_a)]$ *are given as*

$$\begin{aligned} b_{\bar{x}|\mathrm{MD}} &= A^+ C_a b_a \\ b_{\bar{x}|\mathrm{CD}} &= A^+ \left[C_a b_a - \sum_{i=1}^{k} C_i \beta_i(b_a)/\mathsf{P}_{\mathrm{CD}} \right] \\ b_{\bar{x}|\mathrm{WI}} &= A^+ \left[C_a b_a - \sum_{i=1,\neq a}^{k} C_i \beta_i(b_a)/\mathsf{P}_{\mathrm{WI}} \right] \\ b_{\bar{x}|\mathrm{CI}} &= A^+ C_a \left[b_a - \beta_a(b_a)/\mathsf{P}_{\mathrm{CI}} \right] \end{aligned} \tag{68}$$

Proof See Appendix. □

The above result shows that the DIA estimator $\bar{x}$ is not only unconditionally biased under $\mathcal{H}_a$, $b_{\bar{x}} \neq 0$, but even also when it is conditioned on *correct identification*, $b_{\bar{x}|\mathrm{CI}} \neq 0$. Thus if one would repeat the measurement–estimation–testing experiment of the DIA estimator a sufficient number of times under $\mathcal{H}_a$ being true and one would then be only collecting the correctly adapted outcomes of $\bar{x}$, then still their expectation would not coincide with x. This fundamental result puts the often argued importance of unbiasedness (e.g. in BLUEs) in a somewhat different light.

Note that in case of *datasnooping*, when the vectors $C_i = c_i$ take the form of canonical unit vectors, the conditional bias under correct identification is given as

$$b_{\bar{x}|\mathrm{CI}} = A^+ \bar{b}_{y_a|\mathrm{CI}}, \quad \bar{b}_{y_a|\mathrm{CI}} = c_a[b_a - \beta_a(b_a)/\mathsf{P}_{\mathrm{CI}}] \tag{69}$$

while the corresponding unconditional bias is given as

$$\begin{aligned} b_{\bar{x}} &= A^+ \bar{b}_{y_a}, \\ \bar{b}_{y_a} &= \left[c_a(b_a - \beta_a(b_a)) - \sum_{i=1,\neq a}^{k} c_i \beta_i(b_a) \right] \end{aligned} \tag{70}$$

Comparison of these two expressions shows that where in the CI case the bias contribution to $b_{\bar{x}|\mathrm{CI}}$ stays confined to the correctly identified ath observation, this is not true anymore for the bias contribution in the unconditional case. In the unconditional case, the bias in $\bar{x}$ also receives contributions from the entries of $\bar{b}_{y_a}$ other than its ath, i.e. instead of $\bar{b}_{y_a|\mathrm{CI}}$, which only has its ath entry being nonzero, the vector $\bar{b}_{y_a}$ has next to its ath entry also its other entries being filled up with nonzero values.

4.3 The PDF decomposition of $\bar{x}$ under $\mathcal{H}_0$ and $\mathcal{H}_a$

Similar to the above bias decomposition, we can decompose the unconditional PDF of $\bar{x}$ into its conditional constituents under $\mathcal{H}_0$ and $\mathcal{H}_a$. We have the following result.

Theorem 4 (PDF decomposition) *The PDF of* $\bar{x}$ *can be decomposed under* $\mathcal{H}_0$ *and* $\mathcal{H}_a$ *as,*

$$\begin{aligned} f_{\bar{x}}(x|\mathcal{H}_0) &= f_{\bar{x}|\mathrm{CA}}(x|\mathrm{CA})\mathsf{P}_{\mathrm{CA}} + f_{\bar{x}|\mathrm{FA}}(x|\mathrm{FA})\mathsf{P}_{\mathrm{FA}} \\ f_{\bar{x}}(x|\mathcal{H}_a) &= f_{\bar{x}|\mathrm{MD}}(x|\mathrm{MD})\mathsf{P}_{\mathrm{MD}} + f_{\bar{x}|\mathrm{CD}}(x|\mathrm{CD})\mathsf{P}_{\mathrm{CD}} \end{aligned} \tag{71}$$

with $f_{\bar{x}|\mathrm{CA}}(x|\mathrm{CA}) = f_{\hat{x}_0}(x|\mathcal{H}_0)$, $f_{\bar{x}|\mathrm{MD}}(x|\mathrm{MD}) = f_{\hat{x}_0}(x|\mathcal{H}_a)$, *and where*

$$f_{\bar{x}|\mathrm{CD}}(x|\mathrm{CD}) = f_{\bar{x}|\mathrm{WI}}(x|\mathrm{WI})\frac{\mathsf{P}_{\mathrm{WI}}}{\mathsf{P}_{\mathrm{CD}}} + f_{\bar{x}|\mathrm{CI}}(x|\mathrm{CI})\frac{\mathsf{P}_{\mathrm{CI}}}{\mathsf{P}_{\mathrm{CD}}} \tag{72}$$

with

$$f_{\bar{x}|\mathrm{CI}}(x|\mathrm{CI}) = \int_{\mathcal{P}_a} f_{\hat{x}_a,t}(x,\tau|\mathcal{H}_a)d\tau/\mathsf{P}_{\mathrm{CI}} \tag{73}$$

Proof See Appendix. □

The above decomposition can be used to evaluate the performance of the DIA estimator for the various different occurrences of the testing outcomes. Note that the PDF of $\bar{x}$, when conditioned on correct acceptance (CA) or missed detection (MD), is simply that of $\hat{x}_0$ under $\mathcal{H}_0$ and $\mathcal{H}_a$, respectively. Such simplification does, however, not occur for the PDF when conditioned on correct identification (CI cf. 73). This is due to the fact that $\hat{x}_a$, as opposed to $\hat{x}_0$, is *not* independent from the misclosure vector t. Thus

$$\begin{aligned} &f_{\bar{x}|\mathrm{CA}}(x|\mathrm{CA}) = f_{\hat{x}_0}(x|\mathcal{H}_0), \text{ but} \\ &f_{\bar{x}|\mathrm{CI}}(x|\mathrm{CI}) \neq f_{\hat{x}_a}(x|\mathcal{H}_a) \end{aligned} \tag{74}$$

Also note, since $\mathcal{H}_a$ depends on the unknown bias b_a, that one needs to make assumptions on its range of values when evaluating $f_{\bar{x}}(x|\mathcal{H}_a)$. One such approach is based on the well-known concept of minimal detectable biases (Baarda 1968a; Teunissen 1998a; Salzmann 1991).

4.4 On the minimal detectable bias

The minimal detectable bias (MDB) of an alternative hypothesis $\mathcal{H}_a$ is defined as the (in absolute value) smallest bias that leads to rejection of $\mathcal{H}_0$ for a given CD probability. It can be computed by 'inverting'

$$\mathsf{P}_{\mathrm{CD}} = \int_{\mathbb{R}^r/\mathcal{P}_0} f_t(\tau|\mathcal{H}_a)\mathrm{d}\tau = \int_{\mathbb{R}^r/\mathcal{P}_0} f_t(\tau - B^T C_a b_a|\mathcal{H}_0)\mathrm{d}\tau \tag{75}$$

for a certain CD probability, say $\mathsf{P}_{\mathrm{CD}} = \gamma_{\mathrm{CD}}$. For the $\mathcal{P}_0$ of example 1 (cf. 27), with $\tau^2 = \chi^2_\alpha(r, 0)$, it can be computed from 'inverting'

$$\mathsf{P}(||t||^2_{Q_{tt}} > \chi^2_\alpha(r, 0)|\mathcal{H}_a) = \gamma_{\mathrm{CD}} \tag{76}$$

using the fact that under $\mathcal{H}_a$, $||t||^2_{Q_{tt}} \sim \chi^2(r, \lambda^2_t)$, with $\lambda^2_t = ||P^\perp_A C_a b_a||^2_{Q_{yy}}$ (Arnold 1981; Koch 1999). For the case $C_a = c_a$, inversion leads then to the well-known MDB expression (Baarda 1968a; Teunissen 2000)

$$|b_{a,\mathrm{MDB}}| = \lambda_t(\alpha, \gamma_{\mathrm{CD}}, r)/||P^\perp_A c_a||_{Q_{yy}} \tag{77}$$

in which $\lambda_t(\alpha, \gamma_{\mathrm{CD}}, r)$ denotes the function that captures the dependency on the chosen false-alarm probability, $\mathsf{P}_{\mathrm{FA}} = \alpha$, the chosen correct detection probability, $\mathsf{P}_{\mathrm{CD}} = \gamma_{\mathrm{CD}}$, and the model redundancy r. For the higher-dimensional case when b_a is a vector instead of a scalar, a similar expression can be obtained, see Teunissen (2000).

The importance of the MDB concept is that it expresses the sensitivity of the *detection* step of the DIA method in terms of minimal bias sizes of the respective hypotheses. By using the same $\mathsf{P}_{\mathrm{CD}} = 1 - \mathsf{P}_{\mathrm{MD}} = \gamma_{\mathrm{CD}}$ for all $\mathcal{H}_a$, the MDBs can be compared and provide information on the sensitivity of rejecting the null hypothesis for $\mathcal{H}_a$-biases the size of their MDBs. For instance, in case of datasnooping of a surveyor's trilateration network, the MDBs and their mutual comparison would reveal for which observed distance in the trilateration network the rejection of $\mathcal{H}_0$ would be least sensitive (namely distance with largest MDB), as well as how large its distance bias needs to be to have rejection occur with probability $\mathsf{P}_{\mathrm{CD}} = \gamma_{\mathrm{CD}}$.

It is important to understand, however, that the MDB is about correct detection and not about correct identification. This would only be true in the binary case, when next to the null hypothesis only a *single* alternative hypothesis is considered. For identification in the multiple hypotheses case ($k > 1$), one can, however, pose a somewhat similar question as the one that led to the MDB: what is the smallest bias of an alternative hypothesis $\mathcal{H}_a$ that leads to its identification for a given CI probability. Thus similar to (75), such *minimal identifiable bias* is found from 'inverting'

$$\mathsf{P}_{\mathrm{CI}} = \mathsf{P}(t \in \mathcal{P}_a|\mathcal{H}_a) = \int_{\mathcal{P}_a} f_t(\tau - B^T C_a b_a|\mathcal{H}_0)\mathrm{d}\tau \tag{78}$$

for a given CI probability. Compare this with (75) and note the difference.

Example 7 (Minimal Identifiable Bias) Let $t \overset{\mathcal{H}_a}{\sim} \mathcal{N}(c_{t_a} b_a, Q_{tt})$ and define $[\hat{b}^T_a, \bar{t}^T_a]^T = \mathcal{T}_a t$, using the Tienstra transformation $\mathcal{T}_a = [c^{+T}_{t_a}, d_{t_a}]^T$, in which $c^+_{t_a}$ is the BLUE-inverse of c_{t_a} and d_{t_a} is a basis matrix of the null space of $c^T_{t_a}$. Then $\bar{t}_a$ and w_a are independent and have the distributions $\bar{t}_a \overset{\mathcal{H}_a}{\sim} \mathcal{N}(0, Q_{\bar{t}_a\bar{t}_a})$ and $w_a \overset{\mathcal{H}_a}{\sim} \mathcal{N}(b_a/\sigma_{\hat{b}_a}, 1)$, with the decomposition $||t - c_{t_a} b_a||^2_{Q_{tt}} = ||\bar{t}_a||^2_{Q_{\bar{t}_a\bar{t}_a}} + (w_a - \frac{b_a}{\sigma_{\hat{b}_a}})^2$. As the identification regions, we take the ones of (38). For a given b_a, the probability of correct identification can then be computed as

$$\mathsf{P}(t \in \mathcal{P}_a|\mathcal{H}_a) = \int_{\mathcal{P}_a} \frac{\exp\left\{-\frac{1}{2}||\bar{t}_a||^2_{Q_{\bar{t}_a\bar{t}_a}}\right\}}{\sqrt{|2\pi\, Q_{\bar{t}_a\bar{t}_a}|}} \frac{\exp\left\{-\frac{1}{2}(w_a - b_a/\sigma_{\hat{b}_a})^2\right\}}{\sqrt{2\pi}} d\bar{t}_a dw_a \tag{79}$$

with the $\mathcal{P}_a$-defining constraints given as $\tau^2 - w^2_a < ||\bar{t}_a||^2_{Q_{\bar{t}_a\bar{t}_a}} \leq \bar{\tau}^2$ and $|w_a| = \max_j |w_j|$. Reversely, one can compute, for a given CI probability, the minimal identifiable bias by solving b_a from $P_{\mathrm{CI}} = \mathsf{P}(t \in \mathcal{P}_a|\mathcal{H}_a)$. As a useful approximation of the CI probability from above, one may use

$$\begin{aligned}\mathsf{P}(t \in \mathcal{P}_a|\mathcal{H}_a) &\leq \mathsf{P}(\tau^2 - w^2_a < ||\bar{t}_a||^2_{Q_{\bar{t}_a\bar{t}_a}} \leq \bar{\tau}^2|\mathcal{H}_a) \\ &\leq \mathsf{P}(||\bar{t}_a||^2_{Q_{\bar{t}_a\bar{t}_a}} \leq \bar{\tau}^2|\mathcal{H}_a)\mathsf{P}(w^2_a > \tau^2 - \bar{\tau}^2|\mathcal{H}_a)\end{aligned} \tag{80}$$

The first inequality follows from discarding the constraint $|w_a| = \max_j |w_j|$, the second from discarding the curvature of $\mathcal{P}_0$ and the fact that $\bar{t}_a$ and w_a are independent. A too small value of the easy-to-compute upper bound (80) indicates then that identifiability of $\mathcal{H}_a$ is problematic. □

Since $\mathsf{P}_{\mathrm{CD}} \geq \mathsf{P}_{\mathrm{CI}}$, one can expect the minimal identifiable bias to be larger than the MDB when $\mathsf{P}_{\mathrm{CI}} = \gamma_{\mathrm{CD}}$. Correct identification is thus more difficult than correct detection. Their difference depends on the probability of wrongful identification. The smaller P_{WI} is, the closer P_{CI} gets to P_{CD}.

We note that in the simulation studies of Koch (2016, 2017) discrepancies were reported between the MDB and simulated values. Such discrepancies could perhaps have been the consequence of not taking the difference between the MDB and the minimal identifiable bias into account.

As the 'inversion' of (78) is more difficult than that of (75), one may also consider the 'forward' computation. For instance, if all alternative hypotheses are of the same type (e.g. distance outliers in a trilateration network), then (78) could be used to compare the P_{CI}'s of the different hypotheses for the same size of bias. Alternatively, if one wants to

infer the P_{CI} for hypotheses that all have the same correct detection probability $\mathsf{P}_{\mathrm{CD}} = \gamma_{\mathrm{CD}}$, the easy-to-compute MDB can be used in the 'forward' computation as

$$\mathsf{P}_{\mathrm{CI,MDB}} = \int_{\mathcal{P}_a} f_t(\tau - B^T C_a b_{a,\mathrm{MDB}}|\mathcal{H}_0)\mathrm{d}\tau \tag{81}$$

This is thus the probability of identifying a bias the size of the MDB. As the computation of (81) is based on the same P_{CD} for each $\mathcal{H}_a$, it again makes comparison between the hypotheses possible. It allows for a ranking of the alternative hypotheses in terms of their identifiability under the same correct detection probability. Would the probability $\mathsf{P}_{\mathrm{CI,MDB}}$ then turn out to be too small for certain hypotheses (given the requirements of the application at hand), a designer of the measurement–estimation–testing system could then either try to improve the strength of the model (by improving the design matrix A and/or the variance matrix Q_{yy}) or decide to have the regions $\mathcal{P}_{i\neq 0}$ of the poorly identifiable hypotheses added to the undecided region. In the latter case, one would allow such hypotheses to contribute to the rejection of $\mathcal{H}_0$, but not to its adaptation.

4.5 On the computation of P_{CI}

Many of the relevant probabilities in this contribution are multivariate integrals over complex regions. They therefore need to be computed by means of numerical simulation, the principles of which are as follows: as a probability can always be written as an expectation and an expectation can be approximated by taking the average of a sufficient number of samples from the distribution, the random generation of such samples is used for the required approximation. Thus, the approximation of the probability of correct identification,

$$\mathsf{P}_{\mathrm{CI}} = \mathsf{P}(t \in \mathcal{P}_a|\mathcal{H}_a) = \mathsf{E}(p_a(t)|\mathcal{H}_a) \tag{82}$$

for $b_a = b_{a,\mathrm{MDB}}$, is then given by

$$\hat{\mathsf{P}}_{\mathrm{CI,MDB}} = \frac{1}{N}\sum_{i=1}^{N} p_a(\tau_i) \tag{83}$$

(cf. 81), with N the total number of samples and τ_i, $i = 1, \ldots, N$, being the samples drawn from $f_t(\tau|\mathcal{H}_a) = f_t(\tau - B^T C_a b_{a,\mathrm{MDB}}|\mathcal{H}_0)$ and thus from the multivariate normal distribution $\mathcal{N}(B^T C_a b_{a,\mathrm{MDB}}, Q_{tt})$. Whether or not the drawn sample $\tau = \tau_i$ contributes to the average is regulated by the indicator function $p_a(\tau)$ and thus the actual test. The approximation (83) constitutes the simplest of simulation-based numerical integration. For more advanced methods, see Robert and Casella (2013).

5 How to evaluate and use the DIA estimator?

5.1 Unconditional evaluation

The unconditional PDF $f_{\bar{x}}(x)$ (cf. 39) contains all the necessary information about $\bar{x}$. As the DIA estimator $\bar{x}$ is an estimator of x, it is likely to be considered a good estimator if the probability of $\bar{x}$ being close to x is sufficiently large. Although the quantification of such terms as 'close' and 'sufficiently large' is application dependent, we assume that a (convex) shape and size of an x-centred region $\mathcal{B}_x \subset \mathbb{R}^n$, as well as the required probability $1-\epsilon$ of the estimator residing in it, is given. Thus if $\mathcal{H}$ would be the true hypothesis, the estimator $\bar{x}$ would be considered an acceptable estimator of x if the inequality $\mathsf{P}(\bar{x} \in \mathcal{B}_x|\mathcal{H}) \geq 1 - \epsilon$, or

$$\mathsf{P}(\bar{x} \notin \mathcal{B}_x|\mathcal{H}) \leq \epsilon \tag{84}$$

is satisfied for (very) small ϵ. With reference to safety-critical situations, for which occurrences of $\bar{x}$ falling outside $\mathcal{B}_x$ are considered hazardous, the probability (84) will be referred to as the *hazardous* probability. Using (33) and (39) of Theorem 1, we have the following result.

Corollary 4 (Hazardous probability) *The hazardous probability can be computed as*

$$\mathsf{P}(\bar{x} \notin \mathcal{B}_x|\mathcal{H}) = \int_{\mathbb{R}^n/\mathcal{B}_x} f_{\bar{x}}(x|\mathcal{H})dx \tag{85}$$

with the PDF of $\bar{x}$ given as

$$f_{\bar{x}}(x|\mathcal{H}) = \sum_{i=0}^{k}\int_{\mathcal{P}_i} f_{\hat{x}_0}(x + L_i\tau|\mathcal{H}) f_t(\tau|\mathcal{H})d\tau \tag{86}$$

for which, when $\mathcal{H} = \mathcal{H}_a$,

$$\begin{aligned} f_{\hat{x}_0}(x + L_i\tau|\mathcal{H}_a) &= \frac{\exp\left\{-\frac{1}{2}||x - \mu(\tau)||^2_{Q_{\hat{x}_0\hat{x}_0}}\right\}}{\sqrt{|2\pi Q_{\hat{x}_0\hat{x}_0}|}} \\ f_t(\tau|\mathcal{H}_a) &= \frac{\exp\left\{-\frac{1}{2}||\tau - B^T C_a b_a||^2_{Q_{tt}}\right\}}{\sqrt{|2\pi Q_{tt}|}} \end{aligned} \tag{87}$$

with $\mu(\tau) = A^+(\mathsf{E}(y|\mathcal{H}_0) + C_a b_a - C_i(B^T C_i)^+\tau)$. To get the PDF under $\mathcal{H} = \mathcal{H}_0$, set $b_a = 0$ in the above.

By computing (85) for all $\mathcal{H}_i$s, one gets a complete picture of how the DIA estimator performs under each of the hypotheses. Such can then be used for various designing purposes, such as of finding a measurement–design (i.e. choice of A and Q_{yy}) and/or a testing–design (i.e. choice of $\mathcal{P}_i$'s and their partitioning) that realizes sufficiently small hazardous

probabilities. One may then also take advantage of the difference between *influential* and *testable* biases, by testing less stringent for biases that are less influential. Under certain circumstances, one may even try to optimize the DIA estimator, by minimizing an (ordinary or weighted) average hazardous probability

$$\mathsf{P}(\bar{x} \notin \mathcal{B}_x) = \sum_{i=0}^{k} \omega_i \mathsf{P}(\bar{x} \notin \mathcal{B}_x | \mathcal{H}_i) \tag{88}$$

as function of certain estimator defining properties, such as, for instance, the chosen partitioning of the misclosure space. In the event that one has information about the frequency of occurrence of the hypotheses, the weights ω_i would then be given by these a priori probabilities.

Example 8 (Continuation of Example 5) It follows from (39) that in case of only one alternative hypothesis ($k = 1$), the PDF of $\bar{x}$ under $\mathcal{H}_a$ can be written as

$$f_{\bar{x}}(x|\mathcal{H}_a) = f_{\hat{x}_0}(x|\mathcal{H}_a) + \int_{\mathbb{R}/\mathcal{P}_0} [f_{\hat{x}_0}(x + L_1\tau|\mathcal{H}_a) - f_{\hat{x}_0}(x|\mathcal{H}_a)] f_t(\tau|\mathcal{H}_a) \mathrm{d}\tau \tag{89}$$

Let $\hat{x}_0 \overset{\mathcal{H}_a}{\sim} \mathcal{N}(b_{\hat{x}_0} = \frac{1}{2}b_a, \sigma^2_{\hat{x}_0} = 0.5)$, $t \overset{\mathcal{H}_0}{\sim} \mathcal{N}(b_t = b_a, \sigma_t^2 = 2)$ (thus $n = r = 1$), and $L_1 = \frac{1}{2}$, with acceptance interval $\mathcal{P}_0 = \{\tau \in \mathbb{R} | (\tau/\sigma_t)^2 \leq \chi^2_\alpha(1, 0)\}$. Figure 7 shows this PDF for the three cases $b_a = 1, 3, 5$ using $\alpha = 0.1$ and $\alpha = 0.001$, while Fig. 8 shows these same three cases but now with twice improved variance. Note the positive effects that increasing α (cf. Fig. 7) and improving precision (cf. Fig. 8) have on the estimator $\bar{x}$. The larger α gets, the more probability mass of the PDF of $\bar{x}$ becomes centred again around the correct value 0, this of course at the expense of heavier tails in $f_{\bar{x}}(x|\mathcal{H}_0)$, see Fig. 6. For small values of α, thus when the acceptance interval of $\mathcal{H}_0$ is large, the estimator $\bar{x}$ is more biased as its PDF $f_{\bar{x}}(x|\mathcal{H}_a)$ becomes more centred around $b_{\hat{x}_0}$. Also the effect of precision improvement is clearly visible, with ultimately, for $b_a = 5$, a PDF of $\bar{x}$ that has most of its mass centred around 0 again. □

To provide further insight in the factors that contribute to (85), we make use of the decomposition

$$\mathsf{P}(\bar{x} \notin \mathcal{B}_x|\mathcal{H}) = \mathsf{P}(\bar{x} \notin \mathcal{B}_x, t \in \mathcal{P}_0|\mathcal{H}) + \mathsf{P}(\bar{x} \notin \mathcal{B}_x, t \notin \mathcal{P}_0|\mathcal{H}) \tag{90}$$

The first term of the sum is relevant for detection, the second when identification is included as well. We first consider the detection-only case.

5.2 Detection only: precision and reliability

In the detection-only case, the solution $\bar{x}$ is declared unavailable when the null hypothesis is rejected, i.e. when $t \notin \mathcal{P}_0$. The probability of such outcome is under $\mathcal{H}_0$, the false alarm

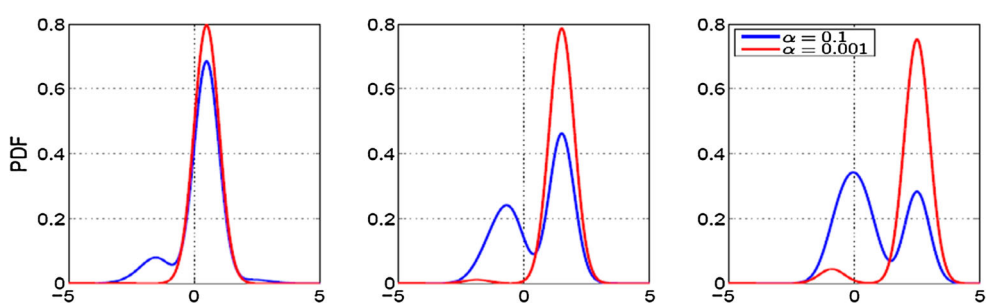

Fig. 7 PDF $f_{\bar{x}}(x|\mathcal{H}_a)$ of estimator $\bar{x}$ under $\mathcal{H}_a$ (Example 8, $\sigma^2_{\hat{x}_0} = 0.5, \sigma_t^2 = 2$). *Left* $b_a = 1$ ($b_{\hat{x}_0} = 0.5$), *middle* $b_a = 3$ ($b_{\hat{x}_0} = 1.5$), *right* $b_a = 5$ ($b_{\hat{x}_0} = 2.5$)

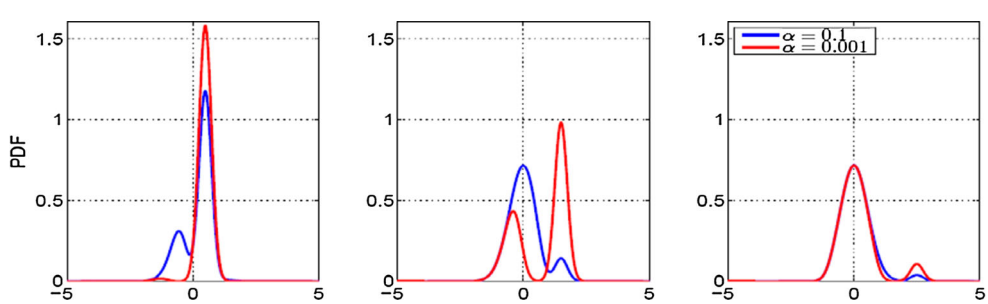

Fig. 8 PDF $f_{\bar{x}}(x|\mathcal{H}_a)$ of DIA estimator $\bar{x}$ under $\mathcal{H}_a$ (Example 8, $\sigma^2_{\hat{x}_0} = 0.25, \sigma_t^2 = 1$). *Left* $b_a = 1$ ($b_{\hat{x}_0} = 0.5$), *middle* $b_a = 3$ ($b_{\hat{x}_0} = 1.5$), *right* $b_a = 5$ ($b_{\hat{x}_0} = 2.5$)

$\mathsf{P}(t \notin \mathcal{P}_0|\mathcal{H}_0) = \mathsf{P}_{\text{FA}}$, and under $\mathcal{H}_a$, the probability of correct detection $\mathsf{P}(t \notin \mathcal{P}_0|\mathcal{H}_a) = \mathsf{P}_{\text{CD}}$. The false alarm is often user controlled by setting the appropriate size of $\mathcal{P}_0$. The probability of correct detection, however, depends on which $\mathcal{H}_a$ is considered and on the size of its bias b_a. We have $\mathsf{P}_{\text{CD}} = \mathsf{P}_{\text{FA}}$ for $b_a = 0$, but $\mathsf{P}_{\text{CD}} > \mathsf{P}_{\text{FA}}$ otherwise.

The estimator $\bar{x}$ is computed as $\hat{x}_0$ when the null hypothesis is accepted, i.e. when $t \in \mathcal{P}_0$. Its hazardous probability is then given for $\mathcal{H} = \mathcal{H}_0, \mathcal{H}_a$ as

$$\mathsf{P}(\bar{x} \notin \mathcal{B}_x, t \in \mathcal{P}_0|\mathcal{H}) = \begin{cases} \mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_0)\mathsf{P}_{\text{CA}} \\ \mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_a)\mathsf{P}_{\text{MD}} \end{cases} \tag{91}$$

where use has been made of the independence of $\hat{x}_0$ and t. As P_{CA} is user fixed and $\hat{x}_0 \overset{\mathcal{H}_0}{\sim} \mathcal{N}(x, Q_{\hat{x}_0\hat{x}_0})$, the hazardous probability under $\mathcal{H} = \mathcal{H}_0$ is completely driven by the variance matrix $Q_{\hat{x}_0\hat{x}_0}$ and thus by the *precision* of $\hat{x}_0$. This is not the case under $\mathcal{H}_a$, however, since the hazardous probability, also referred to as hazardous missed detection (HMD) probability, becomes then dependent on b_a as well,

$$\mathsf{P}_{\text{HMD}}(b_a) = \mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_a)\mathsf{P}_{\text{MD}} \tag{92}$$

Note, since $\hat{x}_0 \overset{\mathcal{H}_a}{\sim} \mathcal{N}(x+b_{\hat{x}_0}, Q_{\hat{x}_0\hat{x}_0})$ and $t \overset{\mathcal{H}_a}{\sim} \mathcal{N}(b_t, Q_{tt})$ (cf. 8), that the two probabilities in the product (92) are affected differently by b_a. The probability $\mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_a)$ is driven by the *influential* bias $b_{\hat{x}_0} = A^+C_ab_a$, while the missed detection probability is driven by the *testable* bias $b_t = B^TC_ab_a$ (cf. 9).

To compute and evaluate the probability of hazardous missed detection P_{HMD}, one can follow different routes. The first approach, due to Baarda (1967, 1968a), goes as follows. For each $\mathcal{H}_a$, the MDB is computed from inverting (75). Then, each MDB is taken as reference value for b_a and propagated to obtain the corresponding $b_{\hat{x}_0}$, thus allowing for the computation of P_{HMD} for each $\mathcal{H}_a$. The set of MDBs is said to describe the internal reliability, while their propagation into the parameters is said to describe the external reliability (Baarda 1968a; Teunissen 2000). The computations simplify, if one is only interested in P_{HMD}, while $\mathcal{B}_x$ and $\mathcal{P}_0$ are defined as $\mathcal{B}_x = \{u \in \mathbb{R}^n | \, ||u - x||^2_{Q_{\hat{x}_0\hat{x}_0}} \leq d^2\}$ and $\mathcal{P}_0 = \{t \in \mathbb{R}^r | \, ||t||^2_{Q_{tt}} \leq \chi^2_\alpha(r, 0)\}$. Then, since $||\hat{x}_0 - x||^2_{Q_{\hat{x}_0\hat{x}_0}} \overset{\mathcal{H}_a}{\sim} \chi^2(n, \lambda^2_{\hat{x}_0})$ and $||t||^2_{Q_{tt}} \overset{\mathcal{H}_a}{\sim} \chi^2(r, \lambda^2_t)$, one can make a direct use of the Pythagorian relation (see Fig. 1) and use

$$\lambda_{\hat{x}_0} = \frac{||P_Ac_a||_{Q_{yy}}}{||P_A^\perp c_a||_{Q_{yy}}}\lambda_t(\alpha, \gamma_{\text{CD}}, r) \tag{93}$$

to compute the hazardous missed detection probability as $\mathsf{P}_{\text{HMD}} = \mathsf{P}(\chi^2(n, \lambda^2_{\hat{x}_0}) > d^2) \times (1 - \gamma_{\text{CD}})$ for each $\mathcal{H}_a$. By comparing the P_{HMD}s between hypotheses, one would know which of the $\mathcal{H}_a$s would be the most hazardous to miss detection of and what its hazardous probability would be for a given P_{MD}.

An alternative, more conservative approach would be to directly evaluate (92) as function of the bias b_a. As P_{MD} gets smaller, but $\mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_a)$ larger, for larger biases, the probability (92) will have a maximum for a certain bias. With this approach, one can thus evaluate whether the 'worst-case' scenario $\max_{b_a} \mathsf{P}_{\text{HMD}}(b_a)$ for each of the hypotheses still satisfies ones criterion (Ober 2003; Teunissen 2017).

As the above computations can be done without the need of having the actual measurements available, they are very useful for design verification purposes. Starting from a certain assumed design or measurement setup as described by A and Q_{yy}, one can then infer how well the design can be expected to protect against biases in the event that one of the alternative hypotheses is true.

5.3 Detection and identification

The above analysis is not enough when next to detection, identification is involved as well. With identification included one eliminates (or reduces) the unavailability, but this goes at the expense of an increased hazardous probability. That is, now also the second term in the sum of (90) needs to be accounted for. This extra contribution to the hazardous probability is given for $\mathcal{H} = \mathcal{H}_0, \mathcal{H}_a$ as

$$\begin{aligned} &\mathsf{P}(\bar{x} \notin \mathcal{B}_x, t \notin \mathcal{P}_0|\mathcal{H}) \\ &= \begin{cases} \mathsf{P}(\bar{x} \notin \mathcal{B}_x, t \notin \mathcal{P}_0|\mathcal{H}_0) \leq \mathsf{P}_{\text{FA}} \\ \mathsf{P}(\hat{x}_a \notin \mathcal{B}_x|\text{CI})\mathsf{P}_{\text{CI}} + \mathsf{P}(\bar{x} \notin \mathcal{B}_x|\text{WI})\mathsf{P}_{\text{WI}} \end{cases} \end{aligned} \tag{94}$$

Thus in order to limit the increase in hazardous probability, the design–aim should be to keep the hazardous probability under correct identification, $\mathsf{P}(\hat{x}_a \notin \mathcal{B}_x|\text{CI})$, small, as well as the probability of correct identification, P_{CI}, close to that of correct detection, P_{CD}, thus giving a small probability of wrong identification, P_{WI}.

Note that the computation of (94) is more complex than that of (91), since it involves all $\hat{x}_{i\neq 0}$'s as well. Under $\mathcal{H}_0$, one can get rid of this dependency, when one is willing to work with the, usually small, upper bound P_{FA}. In that case, by adding (94) to (91), the unconditional hazardous probability under $\mathcal{H}_0$ can be bounded from above as

$$\begin{aligned} \mathsf{P}(\bar{x} \notin \mathcal{B}_x|\mathcal{H}_0) &\leq \mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_0)\mathsf{P}_{\text{CA}} + \mathsf{P}_{\text{FA}} \\ &= \mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_0) + \mathsf{P}(\hat{x}_0 \in \mathcal{B}_x|\mathcal{H}_0)\mathsf{P}_{\text{FA}} \end{aligned} \tag{95}$$

With this upper bound, the computational complexity is the same as that of the detection-only case. The situation becomes a bit more complicated under $\mathcal{H}_a$ however. One can then still get rid of a large part of the complexity if one is

willing to use the upper bound $\mathsf{P}(\bar{x} \notin \mathcal{B}_x|\text{WI})\mathsf{P}_{\text{WI}} \leq \mathsf{P}_{\text{WI}} = 1 - \mathsf{P}_{\text{CI}} - \mathsf{P}_{\text{MD}}$. In that case the unconditional hazardous probability can be bounded from above as

$$\begin{aligned}&\mathsf{P}(\bar{x} \notin \mathcal{B}_x|\mathcal{H}_a)\\ &\quad\leq \mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_a)\mathsf{P}_{\text{MD}} + \mathsf{P}(\hat{x}_a \notin \mathcal{B}_x|\text{CI})\mathsf{P}_{\text{CI}} + \mathsf{P}_{\text{WI}}\\ &\quad= 1 - \mathsf{P}(\hat{x}_0 \in \mathcal{B}_x|\mathcal{H}_a)\mathsf{P}_{\text{MD}} - \mathsf{P}(\hat{x}_a \in \mathcal{B}_x|\text{CI})\mathsf{P}_{\text{CI}}\end{aligned} \tag{96}$$

But to assure identification, the computation remains of course still more complex than that of the detection-only case. As the upper bound depends on b_a, its computation can be done (using e.g. the MDB) in a similar way as discussed for (92).

Note that the upper bound of (96) can be sharpened by using

$$\mathsf{P}(\bar{x} \notin \mathcal{B}_x|\mathcal{H}_a) \leq 1 - \sum_{i\in\mathcal{I}} \mathsf{P}(\hat{x}_i \in \mathcal{B}_x|t \in \mathcal{P}_i, \mathcal{H}_a)P_{ia} \tag{97}$$

in which $\mathcal{I}$ denotes the set of hypothesis identifiers for which the probabilities $P_{ia} = \mathsf{P}(t \in \mathcal{P}_i|\mathcal{H}_a)$ are considered significant.

5.4 Conditional evaluation

In practice, the outcome produced by an estimation–testing scheme is often not the end product but just an intermediate step of a whole processing chain. Such follow-on processing, for which the outcome of the DIA estimator $\bar{x}$ is used as input, then also requires the associated quality description. This is not difficult in principle and only requires the proper forward propagation of the unconditional PDF of $\bar{x}$. This is, however, not how it is done in practice. In practice, often two approximations are made when using the DIA estimator. The first approximation is that $\bar{x}$ is not evaluated unconditionally, but rather conditionally on the outcome of testing. Thus instead of working with the PDF of $\bar{x}$, one works with the PDF of $(\bar{x}|t \in \mathcal{P}_i)$, i.e. the one conditioned on the outcome of testing, $t \in \mathcal{P}_i$. The second approximation is that one then neglects this conditioning and uses the unconditional PDF of $\hat{x}_i$ instead for the evaluation. But the fact that the random vector $(\bar{x}|t \in \mathcal{P}_i)$ has the outcome $\hat{x}_i$ does not mean that the two random vectors $(\bar{x}|t \in \mathcal{P}_i)$ and $\hat{x}_i$ have the same distribution. This would only be the case if $\hat{x}_i$ and t are independent, which is true for $\hat{x}_0$ and t, but not for $\hat{x}_{i\neq 0}$ and t.

From a practical point of view, of course, it would be easiest if indeed it would suffice to work with the relatively simple normal PDFs of $\hat{x}_i$. In all subsequent processing, one could then work with the PDF of $\hat{x}_0$, $\mathcal{N}(x, Q_{\hat{x}_0,\hat{x}_0})$, if the null hypothesis gets accepted, and with the PDF of $\hat{x}_a$, $\mathcal{N}(x, Q_{\hat{x}_a,\hat{x}_a})$, if the corresponding alternative hypothesis gets identified. To show what approximations are involved when doing so, we start with the case the null hypothesis gets accepted.

We then have, as $\hat{x}_0$ and t are independent,

$$(\bar{x}|t \in \mathcal{P}_0) = (\hat{x}_0|t \in \mathcal{P}_0) = \hat{x}_0 \tag{98}$$

Thus in this case, it is correct to use the PDF $f_{\hat{x}_0}(x)$ to describe the quality of the conditional random vector $(\bar{x}|t \in \mathcal{P}_0)$. This PDF may thus then indeed be used in the computation of

$$\mathsf{P}(\bar{x} \notin \mathcal{B}_x, t \in \mathcal{P}_0) = \mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x)\mathsf{P}(t \in \mathcal{P}_0) \tag{99}$$

However, since in contrast to (98),

$$(\bar{x}|t \in \mathcal{P}_a) = (\hat{x}_a|t \in \mathcal{P}_a) \neq \hat{x}_a \tag{100}$$

it is not correct to use the PDF $f_{\hat{x}_a}(x)$ for describing the uncertainty when $t \in \mathcal{P}_a$. It depends then on the difference between the two PDFs, $f_{\hat{x}_a}(x)$ and $f_{\hat{x}_a|t\in\mathcal{P}_a}(x|t \in \mathcal{P}_a)$, whether the approximation made is acceptable. The relation between the two PDFs is given in the following.

Corollary 5 *The PDF of $\hat{x}_a$ can be expressed in that of $\hat{x}_a|t \in \mathcal{P}_a$ as*

$$f_{\hat{x}_a}(x) = f_{\hat{x}_a|t\in\mathcal{P}_a}(x|t \in \mathcal{P}_a)\mathsf{P}(t \in \mathcal{P}_a) + R(x) \tag{101}$$

in which

$$f_{\hat{x}_a|t\in\mathcal{P}_a}(x|t \in \mathcal{P}_a) = \int_{\mathcal{P}_a} f_{\hat{x}_0}(x + L_a\tau)\frac{f_t(\tau)}{\mathsf{P}(t \in \mathcal{P}_a)}d\tau \tag{102}$$

and $R(x) = \sum_{i=0,\neq a}^{k} f_{\hat{x}_a|t\in\mathcal{P}_i}(x|t \in \mathcal{P}_i)\mathsf{P}(t \in \mathcal{P}_i)$.

Proof Follows from an application of the PDF total probability rule and (39). □

The difference between the two PDFs gets smaller the larger the probability $\mathsf{P}(t \in \mathcal{P}_a)$, thus illustrating, when under $\mathcal{H}_a$, the importance of having a large enough probability of correct identification, $\mathsf{P}_{\text{CI}} = \mathsf{P}(t \in \mathcal{P}_a|\mathcal{H}_a)$. Note that the conditional PDF is not normal (Gaussian), but can be seen as a 'weighted sum' of shifted versions of $f_{\hat{x}_0}(x)$. The more peaked $f_t(\tau)$ is, the fewer terms in this 'sum'.

Although it is thus not correct to use the PDF of $\hat{x}_a$ to describe the uncertainty in $\bar{x}$ when $t \in \mathcal{P}_a$, it is perhaps comforting to know that the associated probability when using $f_{\hat{x}_a}(x)$ will at least allow for giving a conservative conclusion, since

$$\mathsf{P}(\bar{x} \notin \mathcal{B}_x, t \in \mathcal{P}_a) \leq \mathsf{P}(\hat{x}_a \notin \mathcal{B}_x) \tag{103}$$

Depending on how $\mathcal{P}_a$ is defined, one may obtain a sharper bound by making use of the property that $\hat{x}_a$ and $\bar{t}_a$ are independent (just like $\hat{x}_0$ and t are).

Example 9 By using the same considerations and same $\mathcal{P}_a$ as in Example 5 and using the fact that $\hat{x}_a$ and $\bar{t}_a$ are independent, we have the following sharper upper bound,

$$\begin{aligned}&\mathsf{P}(\bar{x} \notin \mathcal{B}_x, t \in \mathcal{P}_a)\\ &\quad \leq \mathsf{P}(\hat{x}_a \notin \mathcal{B}_x, w_a^2 > \tau^2 - \bar{\tau}^2)\mathsf{P}(||\bar{t}_a||^2_{Q_{\bar{t}_a\bar{t}_a}} \leq \bar{\tau}^2)\\ &\quad \leq \mathsf{P}(\hat{x}_a \notin \mathcal{B}_x)\mathsf{P}(||\bar{t}_a||^2_{Q_{\bar{t}_a\bar{t}_a}} \leq \bar{\tau}^2)\end{aligned} \tag{104}$$

Note that using the upper bound $\mathsf{P}(\hat{x}_a \notin \mathcal{B}_x|\mathrm{CI})\mathsf{P}_{\mathrm{CI}} \leq \mathsf{P}(\hat{x}_a \notin \mathcal{B}_x|\mathcal{H}_a)$ in (96) also allows for an easier-to-compute, albeit looser, upper bound of the hazardous probability,

$$\begin{aligned}&\mathsf{P}(\bar{x} \notin \mathcal{B}_x|\mathcal{H}_a)\\ &\quad \leq \mathsf{P}(\hat{x}_0 \notin \mathcal{B}_x|\mathcal{H}_a)\mathsf{P}_{\mathrm{MD}} + \mathsf{P}(\hat{x}_a \notin \mathcal{B}_x|\mathcal{H}_a) + \mathsf{P}_{\mathrm{WI}}\\ &\quad = 1 - \mathsf{P}(\hat{x}_0 \in \mathcal{B}_x|\mathcal{H}_a)\mathsf{P}_{\mathrm{MD}} + \mathsf{P}(\hat{x}_a \notin \mathcal{B}_x|\mathcal{H}_a) - \mathsf{P}_{\mathrm{CI}}.\end{aligned} \tag{105}$$

As $\hat{x}_0 \overset{\mathcal{H}_a}{\sim} \mathcal{N}(x + b_{\hat{x}_0}, Q_{\hat{x}_0\hat{x}_0})$ and $\hat{x}_a \overset{\mathcal{H}_a}{\sim} \mathcal{N}(x, Q_{\hat{x}_a\hat{x}_a})$, the complexity of the computation is now essentially concentrated in P_{CI}, the probability of correct identification. The more precise $\hat{x}_a$ is the smaller, the upper bound becomes.

6 Summary and conclusions

This contribution introduced a unifying framework for the combined estimation–testing scheme by capturing the overall problem of detection, identification and adaptation (DIA) as one of estimation. The motivation for this approach stems from the fact that although estimation and testing are often treated separatedly and independently, in actual practice when testing is involved the two are intimately connected. This implies that one has to take the intricacies of this combination into consideration when evaluating the contributions of the various decisions and estimators involved. By using a canonical model formulation and a partitioning of the misclosure space, it was shown that the whole estimation–testing scheme can be captured in one single estimator as

$$\bar{x} = \sum_{i=0}^{k} \hat{x}_i\, p_i(t) \tag{106}$$

with the $\hat{x}_i$'s being the BLUEs of each of the hypotheses and the $p_i(t)$'s their indicator functions. All the characteristics of estimation and testing can be studied by studying the properties of the DIA estimator $\bar{x}$.

It was shown that although all the $\hat{x}_i$'s are unbiased under their respective hypotheses, the estimator $\bar{x}$ itself is *not* unbiased, except under the null hypothesis. The presence of such bias was also proven to exist in results of prediction. The nature of the bias was studied, and it was shown that a nonzero bias remains even in case of correct identification. Thus, all successful adaptations still produce results that are biased. This implies, for instance, that any successful outlier detection and exclusion method will always produce parameter outcomes that are still biased. It was shown how this bias can be evaluated and on what contributing factors it depends.

The unconditional PDF of the DIA estimator was derived and formulated in the PDFs of the BLUE $\hat{x}_0$ and the misclosure t, as

$$f_{\bar{x}}(x|\mathcal{H}) = \sum_{i=0}^{k} \int_{\mathcal{P}_i} f_{\hat{x}_0}(x + L_i\tau|\mathcal{H})\, f_t(\tau|\mathcal{H})\mathrm{d}\tau. \tag{107}$$

It forms the basis for any probabilistic evaluation of the estimator, such as, for instance, the evaluation of the hazardous probability $\mathsf{P}(\bar{x} \notin \mathcal{B}_x|\mathcal{H})$. As the PDF gives a complete picture of how the DIA estimator performs under each of the hypotheses, it can be used for various designing purposes, such as of finding a measurement–design and/or a testing–design that realizes sufficiently small hazardous probabilities. One may even try to optimize the DIA estimator by minimizing the hazardous probability as function of certain estimator defining properties. Although optimizing the DIA estimator $\bar{x}$ was not the goal of the present contribution, we did reveal its ingredients that can be varied to influence its performance. The degrees of freedom that one has available for improving its performance are the $\hat{x}_i$'s, i.e. the choice of estimators under the respective hypotheses; the $\mathcal{P}_i$'s, i.e. the choice of partitioning in misclosure space; and/or the p_i's, i.e. the $\bar{x}$-defining choice of weights.

For the evaluation of the PDF of $\bar{x}$ under $\mathcal{H} = \mathcal{H}_a$, one needs to make assumptions on the range of values the unknown bias b_a can take. This can be done by making use of Baarda's well-known concept of minimal detectable biases (MDBs). As was pointed out however, one should understand that in the multiple hypotheses case ($k > 1$) the MDB itself is about correct detection and not about correct identification. As the MDB $b_{a,\mathrm{MDB}}$ is defined to satisfy

$$\mathsf{P}_{\mathrm{CD}} = \int_{\mathbb{R}^r/\mathcal{P}_0} f_t(\tau - B^T C_a b_{a,\mathrm{MDB}}|\mathcal{H}_0)\mathrm{d}\tau \tag{108}$$

for a certain chosen correct detection probability, say $\mathsf{P}_{\mathrm{CD}} = \gamma_{\mathrm{CD}}$, the MDB concept expresses the sensitivity of the *detection* step in terms of minimal bias sizes of the respective hypotheses. Thus by using the same value γ_{CD} for all $\mathcal{H}_a$'s, the MDBs can be compared and provide information on the sensitivity of rejecting the null hypothesis for $\mathcal{H}_a$-biases the

size of their MDBs. This does, however, not necessarily provide information on the *identifiability* of the hypotheses. We therefore introduced, in analogy of (108), a minimal identifiable bias (MIB) as the smallest bias of an alternative hypothesis that leads to its identification for a given CI probability. Would one want to compare the identifiability of alternative hypotheses for the same probability of correct detection however, then the 'forward' computed probability

$$\mathsf{P}_{\mathrm{CI,MDB}} = \int_{\mathcal{P}_a} f_t(\tau - B^T C_a b_{a,\mathrm{MDB}}|\mathcal{H}_0)\mathrm{d}\tau \tag{109}$$

can be used as the diagnostic. It provides a direct link with the MDB, and it is also somewhat easier to compute than the 'inversion' needed for obtaining MIBs.

We also considered conditional performances of the estimator $\bar{x}$, thus enabling evaluation of its performance under different occurrences of the testing outcomes. The increase in the hazardous probability due to identification was described by discriminating between the detection-only case and the detection-and-identification case. We also pointed out that for any follow-on processing for which the outcome of the estimator $\bar{x}$ is used as input, a rigorous forward propagation of its nonnormal PDF is often missing in practice. Instead one works with simpler to evaluate normal PDFs, namely $f_{\hat{x}_0}(x)$ when the output of $\bar{x}$ is $\hat{x}_0$ and $f_{\hat{x}_a}(x)$ if the output is $\hat{x}_a$. This is correct for $\hat{x}_0$, but not for $\hat{x}_a$, even not in case of correct identification, since

$$f_{\bar{x}|\mathrm{CI}}(x|\mathrm{CI}) \neq f_{\hat{x}_a}(x|\mathcal{H}_a) \tag{110}$$

From the given relation between these two distributions, one can infer whether the approximation made is acceptable for the application at hand.

Acknowledgements This contribution benefited from author's discussions with Drs Davide Imparato and Christian Tiberius. Mrs Safoora Zaminpardaz contributed the figures. The author is the recipient of an Australian Research Council (ARC) Federation Fellowship (Project Number FF0883188). This support is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Proof of Theorem 1 By using the total probability rule and the fact that $\bar{x} = \hat{x}_i$ if $t \in \mathcal{P}_i$, we have for any $\Theta \subset \mathbb{R}^{n+r}$,

$$\mathsf{P}\left(\begin{bmatrix}\bar{x}\\ t\end{bmatrix} \in \Theta\right) = \sum_{i=0}^{k} \mathsf{P}\left(\begin{bmatrix}\hat{x}_i\\ t\end{bmatrix} \in \Theta | t \in \mathcal{P}_i\right) P(t \in \mathcal{P}_i) \tag{111}$$

from which it follows that

$$\begin{aligned} f_{\bar{x},t}(x,t) &= \sum_{i=0}^{k} f_{\hat{x}_i,t}(x,t|t \in \mathcal{P}_i) P(t \in \mathcal{P}_i) \\ &= \sum_{i=0}^{k} f_{\hat{x}_i,t}(x,t) p_i(t) \end{aligned} \tag{112}$$

in which the last equality was obtained from using

$$f_{\hat{x}_i,t}(x,t|t \in \mathcal{P}_i) = \frac{f_{\hat{x}_i,t}(x,t)}{P(t \in \mathcal{P}_i)} p_i(t) \tag{113}$$

To express the joint PDF $f_{\hat{x}_i,t}(x,t)$ of (112) in that of $\hat{x}_0$ and t, we apply the PDF transformation rule to (13) to obtain

$$\begin{aligned} f_{\hat{x}_i,t}(x,t) &= f_{\hat{x}_0,t}(x + L_i t, t) \\ &= f_{\hat{x}_0}(x + L_i t) f_t(t) \end{aligned} \tag{114}$$

in which the last equation follows from the independence of $\hat{x}_0$ and t. Substitution of (114) into (112) gives the first equation of (39). The second equation of (39) follows from the first by invoking the definition $f_{\bar{x}|t}(x|t) = f_{\bar{x},t}(x,t)/f_t(t)$. Similarly, the third equation of (39) follows from the first by invoking the definition $f_{\bar{x}}(x) = \int_{\mathbb{R}^r} f_{\bar{x},t}(x,t)dt$. □

Proof of Theorem 2 From (32) and (13), we obtain

$$\begin{aligned} \bar{x} &= \hat{x}_0 - \textstyle\sum_{i=1}^{k} L_i t p_i(t) \\ &= \hat{x}_0 - A^+ \textstyle\sum_{i=1}^{k} C_i (B^T C_i)^+ t p_i(t) \end{aligned} \tag{115}$$

By taking the expectation, we get

$$\mathsf{E}(\bar{x}|\mathcal{H}) = \mathsf{E}(\hat{x}_0|\mathcal{H}) - A^+ \sum_{i=1}^{k} C_i (B^T C_i)^+ \mathsf{E}(t p_i(t)|\mathcal{H}) \tag{116}$$

In this expression, we have for the mean of $\hat{x}_0$, $\mathsf{E}(\hat{x}_0|\mathcal{H}_0) = x$ and $\mathsf{E}(\hat{x}_0|\mathcal{H}_a) = x + A^+ C_a b_a$. For the mean of $t p_i(t)$, $i = 1, \ldots, k$, we have

$$\mathsf{E}(t p_i(t)|\mathcal{H}) = \begin{cases} 0 & \text{if} \quad \mathcal{H} = \mathcal{H}_0 \\ \int_{\mathcal{P}_i} \tau f_t(\tau|\mathcal{H}_a)\mathrm{d}\tau & \text{if} \quad \mathcal{H} = \mathcal{H}_a \end{cases} \tag{117}$$

The zero result under $\mathcal{H}_0$ follows from the symmetry about the origin of the PDF $f_t(\tau|\mathcal{H}_0)$ and the regions $\mathcal{P}_i$. Substitution of (117) into (116) proves the result. □

Proof of Theorem 3 Since

$$\mathsf{E}(tp_i(t)|t \in \Omega, \mathcal{H}_a) = \int_{\mathcal{P}_i \cap \Omega} \tau \frac{f_t(\tau|\mathcal{H}_a)}{\mathsf{P}(t \in \Omega|\mathcal{H}_a)} d\tau \tag{118}$$

we have, with $\hat{b}_i = (B^T C_i)^+ t$, for $i = 1, \ldots, k$,

$$\mathsf{E}(\hat{b}_i p_i(t)|t \in \Omega, \mathcal{H}_a)$$

$$= \begin{cases} 0 & \text{if } \Omega = \mathcal{P}_0 \quad \text{(MD)} \\ \frac{\beta_i(b_a)}{\mathsf{P}_{\mathrm{CD}}} & \text{if } \Omega = \mathbb{R}^r/\mathcal{P}_0 \quad \text{(CD)} \\ \left\{ \begin{array}{ll} 0 & (i = a) \\ \frac{\beta_i(b_a)}{\mathsf{P}_{\mathrm{WI}}} & (i \neq a) \end{array} \right\} & \text{if } \Omega = \mathbb{R}^r/\{\mathcal{P}_0, \mathcal{P}_a\} \quad \text{(WI)} \\ \left\{ \begin{array}{ll} \frac{\beta_i(b_a)}{\mathsf{P}_{\mathrm{CI}}} & (i = a) \\ 0 & (i \neq a) \end{array} \right\} & \text{if } \Omega = \mathcal{P}_a \quad \text{(CI)} \end{cases} \tag{119}$$

where $\beta_i(b_a) = \mathsf{E}(\hat{b}_i p_i(t)|\mathcal{H}_a)$. Substitution of (119) into

$$\begin{aligned} \mathsf{E}(\bar{x} - x|t \in \Omega, \mathcal{H}_a) &= \mathsf{E}(\hat{x}_0 - x|t \in \Omega, \mathcal{H}_a) \\ &\quad - A^+ \sum_{i=1}^{k} C_i \mathsf{E}(\hat{b}_i p_i(t)|t \in \Omega, \mathcal{H}_a) \end{aligned} \tag{120}$$

proves (68). □

Proof of Theorem 4 The decompositions (71) and (72) follow from applying the PDF total probability rule to the PDF of $\bar{x}$, i.e. that $f_{\bar{x}}(x) = f_{\bar{x}|t\in\Omega}(x|t \in \Omega)\mathsf{P}(t \in \Omega) + f_{\bar{x}|t\in\Omega^c}(x|t \in \Omega^c)\mathsf{P}(t \in \Omega^c)$ for a partitioning Ω and $\Omega^c = \mathbb{R}^r/\Omega$. The conditional PDFs $f_{\bar{x}|\mathrm{CA}}(x|\mathrm{CA}) = f_{\hat{x}_0}(x|\mathcal{H}_0)$, $f_{\bar{x}|\mathrm{MD}}(x|\mathrm{MD}) = f_{\hat{x}_0}(x|\mathcal{H}_a)$ and (73) follow similarly from the general expression

$$f_{\bar{x}|t\in\Omega}(x|t \in \Omega) = \int_{\Omega} f_{\bar{x},t}(x, \tau)\mathrm{d}\tau/\mathsf{P}(t \in \Omega) \tag{121}$$

where, in addition, use has been made of the fact that $\hat{x}_0$ and t are independent. □

References

Alberda JE (1976) Quality control in surveying. Chart Surv 4(2):23–28

Arnold S (1981) The theory of linear models and multivariate analysis, vol 2. Wiley, New York

Baarda W (1967) Statistical concepts in geodesy. Netherlands Geodetic Commission, Publ. on geodesy, New series 2(4)

Baarda W (1968a) A testing procedure for use in geodetic networks. Netherlands Geodetic Commission, Publ. on geodesy, New Series 2(5)

Baarda W (1968b) Enkele inleidende beschouwingen tot de B-methode van toetsen. Tech. rep, Laboratorium voor Geodesie, Delft

Baarda W (1976) Reliability and precision of geodetic networks. Tech. rep., VII Int Kurs fur Ingenieursmessungen hoher Prazision, Band I, T.H. Darmstadt, Inst. fur Geodaesie

Betti B, Crespi M, Sanso F (1993) A geometric illustration of ambiguity resolution in GPS theory and a Bayesian approach. Manuscr Geod 18:317–330

Brown R (1996) Receiver autonomous integrity monitoring. Glob Position Syst: Theory Appl 2:143–165

Burnham K, Anderson D (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, Berlin

DGCC (1982) The Delft Approach for the Design and Computation of Geodetic Networks. In: "Forty Years of Thought" Anniversary edition on the occasion of the 65th birthday of Professor W Baarda By staff of the Delft Geodetic Computing Centre (DGCC) vol 1, pp 202–274

Drevelle V, Bonnifait P (2011) A set-membership approach for high integrity height- aided satellite positioning. GPS Solut 15(4):357–368

Fan L, Zhai G, Chai H (2011) Study on the processing method of cycle slips under kinematic mode. Theor Math Found Comput Sci 164:175–183

Foerstner W (1983) Reliability and discernability of extended Gauss-Markov models. Deutsche Geod Komm 98:79–103

Gillissen I, Elema I (1996) Test results of DIA: a real-time adaptive integrity monitoring procedure, used in an integrated navigation system. Int Hydrogr Rev 73(1):75–103

Hewitson S, Wang J (2006) GNSS receiver autonomous integrity monitoring (RAIM) performance analysis. GPS Solut 10(3):155–170

Hewitson S, Kyu Lee H, Wang J (2004) Localizability analysis for GPS/Galileo receiver autonomous integrity monitoring. J Navig 57(02):245–259

Imparato D (2016) GNSS-based receiver autonomous integrity monitoring for aircraft navigation. Delft University of Technology

Joerger M, Pervan B (2014) Solution separation and Chi-squared ARAIM for fault detection and exclusion. In: Position, location and navigation symposium-PLANS 2014, 2014 IEEE/ION, IEEE, pp 294–307

Kargoll B (2007) On the theory and application of model misspecification tests in geodesy. Universitats und Landesbibliothek Bonn

Kelly R (1998) The linear model, RNP, and the near-optimum fault detection and exclusion algorithm. Glob Position Syst 5:227–259

Koch K (1999) Parameter estimation and hypothesis testing in linear models. Springer, Berlin

Koch KR (2016) Minimal detectable outliers as measures of reliability. J Geodesy 89:483–490

Koch KR (2017) Expectation maximization algorithm and its minimal detectable outliers. Stud Geophys Geod 61

Kok J (1982) Statistical analysis of deformation problems using Baarda's testing procedures. In: "Forty Years of Thought" Anniversary Volume on the Occasion of Prof Baarda's 65th Birthday, Delft vol 2, pp 470–488

Kok JJ (1984) On data snooping and multiple outlier testing. US Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Charting and Geodetic Services

Ober P (2003) Integrity prediction and monitoring of navigation systems, vol 1. Integricom Publishers Leiden

Parkinson B, Axelrad P (1988) Autonomous GPS integrity monitoring using the pseudorange residual. Navigation 35(2):255–274

Perfetti N (2006) Detection of station coordinate discontinuities within the Italian GPS fiducial network. J Geodesy 80(7):381–396

Robert C, Casella G (2013) Monte Carlo statistical methods. Springer, Berlin

Salzmann M (1991) MDB: a design tool for integrated navigation systems. Bull Geodesique 65(2):109–115

Salzmann M (1993) Least squares filtering and testing for geodetic navigation applications. Netherlands Geodetic Commission, Publ. on geodesy, New series (37)

Sturza M (1988) Navigation system integrity monitoring using redundant measurements. Navigation 35(4):483–501
Teunissen PJG (1985) Quality control in geodetic networks. In: Grafarend EW, Sanso F (eds) Optimization and design of geodetic networks, pp 526–547
Teunissen PJG (1989) Quality control in integrated navigation systems. IEEE Aerosp Electron Syst Mag 5(7):35–41
Teunissen PJG (1990) An integrity and quality control procedure for use in multi sensor integration. In: Proceedings ION GPS (republished in ION Red Book Series, vol. 7, 2010)
Teunissen PJG (1998a) Minimal detectable biases of GPS data. J Geodesy 72(4):236–244
Teunissen PJG (1998b) Quality Control and GPS. Chap. 7 in GPS for Geodesy, 2nd ed, pp 187–229
Teunissen PJG (2000) Testing theory: an introduction. Delft University Press, Series on Mathematical Geodesy and Positioning
Teunissen PJG (2003a) Integer aperture GNSS ambiguity resolution. Artif Satel 38(3):79–88
Teunissen PJG (2003b) Theory of integer equivariant estimation with application to GNSS. J Geodesy 77:402–410
Teunissen PJG (2017) Batch and Recursive Model Validation. In: Teunissen PJG, Montenbruck O (eds) Chapter 24 in handbook of global navigation satellite systems, pp 187–229
Teunissen PJG, Amiri-Simkooei AR (2008) Least-squares variance component estimation. J Geodesy 82(2):65–82
Teunissen PJG, Khodabandeh A (2013) BLUE, BLUP and the Kalman filter: some new results. J Geodesy 87:461–473
Teunissen PJG, Salzmann MA (1989) A recursive slippage test for use in state-space filtering. Manuscr Geod 14(6):383–390
Teunissen PJG, Imparato D, Tiberius CCJM (2017) Does RAIM with correct exclusion produce unbiased positions? Sensors 17(7):1508. doi:10.3390/s17071508
Tiberius C (1998) Recursive data processing for kinematic GPS surveying. Netherlands Geodetic Commission, Publ. on Geodesy, New series (45)
Tienstra J (1956) Theory of the adjustment of normally distributed observation. Argus, Amsterdam
Van der Marel H, Kosters A (1990) Statistical testing and quality analysis in 3-d networks: (part II) application to GPS. Int Assoc Geod Symp 102:290–297
Van Mierlo J (1980) A testing procedure for analytic deformation measurements. Proceedings of Internationales Symposium Ueber Deformationsmessungen mit Geodaetischen Methoden. Verlag Konrad Wittwer, Stuttgart
Watson GS (1983) Statistics on spheres, vol 3. Wiley, New York
Yang L, Knight N, Li Y, Rizos C (2013a) Optimal fault detection and exclusion applied in GNSS positioning. J Navig 66:683–700
Yang L, Wang J, Knight N, Shen Y (2013b) Outlier separability analysis with a multiple alternative hypotheses test. J Geodesy 87:591–604