Journal of Geodesy
https://doi.org/10.1007/s00190-018-1141-3

ORIGINAL ARTICLE

DIA-datasnooping and identifiability

S. Zaminpardaz[1] · **P. J. G. Teunissen**[1,2]

Received: 3 December 2017 / Accepted: 30 March 2018

Abstract

In this contribution, we present and analyze datasnooping in the context of the DIA method. As the DIA method for the detection, identification and adaptation of mismodelling errors is concerned with estimation and testing, it is the combination of both that needs to be considered. This combination is rigorously captured by the *DIA estimator*. We discuss and analyze the DIA-datasnooping decision probabilities and the construction of the corresponding partitioning of misclosure space. We also investigate the circumstances under which two or more hypotheses are *nonseparable* in the identification step. By means of a theorem on the equivalence between the nonseparability of hypotheses and the inestimability of parameters, we demonstrate that one can forget about adapting the parameter vector for hypotheses that are nonseparable. However, as this concerns the *complete* vector and not necessarily functions of it, we also show that parameter functions may exist for which adaptation is still possible. It is shown how this adaptation looks like and how it changes the structure of the DIA estimator. To demonstrate the performance of the various elements of DIA-datasnooping, we apply the theory to some selected examples. We analyze how geometry changes in the measurement setup affect the testing procedure, by studying their partitioning of misclosure space, the decision probabilities and the minimal detectable and identifiable biases. The difference between these two minimal biases is highlighted by showing the difference between their corresponding contributing factors. We also show that if two alternative hypotheses, say $\mathcal{H}_i$ and $\mathcal{H}_j$, are nonseparable, the testing procedure may have different levels of sensitivity to $\mathcal{H}_i$-biases compared to the same $\mathcal{H}_j$-biases.

Keywords Detection, identification and adaptation (DIA) · Datasnooping · Misclosure space partitioning · DIA estimator · Minimal detectable bias (MDB) · Minimal identifiable bias (MIB) · Probability of correct identification · Nonseparable hypotheses

1 Introduction

The DIA method for the detection, identification and adaptation of mismodelling errors combines estimation with testing. This combination of estimation and testing can be rigorously captured in the *DIA estimator* as introduced in (Teunissen 2017). The DIA method has already been widely employed in a variety of applications, such as the quality control of geodetic networks and the integrity monitoring of GNSS models, see, e.g., (DGCC 1982; Teunissen 1990; Salzmann 1995; Tiberius 1998; Perfetti 2006; Khodabandeh and Teunissen 2016; Zaminpardaz et al. 2015).

In this contribution, as an important example of multiple hypothesis testing, datasnooping (Baarda 1967, 1968; Teunissen 1985) is presented in the context of the DIA method. In doing so, we make use of the partitioning of misclosure space based on which we discuss the datasnooping decision probabilities and the construction of the corresponding DIA estimator. Through this partitioning, the distribution of the misclosure vector can be used to determine the correct detection (CD) and correct identification (CI) probabilities of each of the alternative hypotheses, as well as their corresponding minimal biases, the minimal detectable bias (MDB) and the minimal identifiable bias (MIB). We highlight their difference by showing the difference between their corresponding contributing factors. We also investigate the circumstances under which two or more hypotheses are *nonseparable* and discuss the relevant corrective actions

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

S. Zaminpardaz
safoora.zaminpardaz@postgrad.curtin.edu.au

[1] Department of Spatial Sciences, GNSS Research Centre, Curtin University, Perth, Australia

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

Published online: 09 April 2018

including 'remeasurement', 'adaptation' or stating that the solution is 'unavailable'. Of these, the adaptation step is the most involved and will be discussed in more detail.

This contribution is structured as follows. In Sect. 2, we briefly review the DIA method, describe the steps of DIA-datasnooping and define its corresponding DIA estimator. We hereby highlight the role played by the chosen partitioning of misclosure space. In Sect. 3, the decision probabilities of DIA-datasnooping are discussed, whereby between the following events are distinguished: correct acceptance (CA), false alarm (FA), correct/missed detection and correct/wrong identification. It is hereby highlighted that the MDB provides information about correct detection and *not* about correct identification. A high probability of correct detection does namely not necessarily imply a high probability of correct identification, unless one is dealing with the special case of having only one single alternative hypothesis.

As identification of hypotheses becomes problematic if the misclosure vector has the *same* distribution under *different* hypotheses, we study its consequences for the identification and adaptation steps in Sect. 4. We discuss the corrective actions one can choose from in terms of 'remeasurement', 'adaptation' or stating that the solution is 'unavailable'. Of these, the adaptation step is the most involved. By means of a theorem on the equivalence between the nonseparability of hypotheses and the inestimability of parameters, we demonstrate that one can forget about adapting the *complete* vector of unknowns for hypotheses that are nonseparable. However, it is demonstrated that there may exist parameter functions for which adaptation is still possible. It is shown how this adaptation looks like and how it changes the structure of the DIA estimator.

To illustrate and explain the performance of the various elements of DIA-datasnooping, the theory is applied to selected examples in Sect. 5. The following three different cases are treated: height-difference observations of a leveling network, distance measurements of a horizontal geodetic network and pseudorange measurements between a single ground station and GPS satellites. We analyze how geometry changes in the measurement setup affect the testing procedure, including its partitioning of the misclosure space, and the corresponding CD probabilities (MDB) and CI probabilities (MIB). We also demonstrate that for a given bias-to-noise ratio and a false alarm probability, the ordering of the CD probabilities of the alternative hypotheses is not necessarily the same as that of their CI probabilities. It is also shown if two alternative hypotheses, say $\mathcal{H}_i$ and $\mathcal{H}_j$, are not distinguishable, that the testing procedure may have different levels of sensitivity to $\mathcal{H}_i$-biases compared to the same $\mathcal{H}_j$-biases. Finally, a summary and conclusions are given in Sect. 6.

2 Detection, identification and adaptation (DIA)

2.1 DIA in brief

We first formulate the null- and alternative hypotheses, denoted as $\mathcal{H}_0$ and $\mathcal{H}_i$, respectively. Let the observational model under the null hypothesis be given as

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

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

The validity of the null hypothesis can be violated if the functional model and/or the stochastic model is misspecified. Here we assume that a misspecification is restricted to an underparametrization of the mean of y, which is the most common error that occurs when formulating the model. Thus, the alternative hypothesis $\mathcal{H}_i$ is formulated as

$$\mathcal{H}_i: \quad E(y) = Ax + C_i b_i; \quad D(y) = Q_{yy} \tag{2}$$

for some vector $C_i b_i \in \mathbb{R}^m/\{0\}$ such that $[A \;\; C_i]$ is a known matrix of full rank and rank $([A \;\; C_i]) < m$. C_i and b_i will further be specified in detail in Sect. 2.2. The best linear unbiased estimator (BLUE) of x under $\mathcal{H}_0$ and $\mathcal{H}_i$ is, respectively, denoted by $\hat{x}_0$ and $\hat{x}_i$ and given as

$$\hat{x}_0 = A^+ y \quad , \quad \hat{x}_i = \bar{A}_i^+ y \tag{3}$$

with $A^+ = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ the BLUE-inverse of A, $\bar{A}_i^+ = (\bar{A}_i^T Q_{yy}^{-1} \bar{A}_i)^{-1} \bar{A}_i^T Q_{yy}^{-1}$ the BLUE-inverse of $\bar{A}_i = P_{C_i}^{\perp} A$ 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, along the range space of C_i, onto the Q_{yy}^{-1}-orthogonal complement of the range space of C_i.

As one often will have to consider more than one single alternative hypothesis, the statistical model validation of $\mathcal{H}_0$ and k alternatives $\mathcal{H}_i$ $(i = 1, \ldots, k)$ usually goes along the following three steps of detection, identification and adaptation (DIA) (Baarda 1968; Teunissen 1990).

1. *Detection* The validity of the null hypothesis is checked by virtue of an overall model test, without the need of having to consider a particular set of alternative hypotheses. If $\mathcal{H}_0$ is accepted, $\hat{x}_0$ is provided as the estimate of x.

2. *Identification* In case $\mathcal{H}_0$ is rejected, a search is carried out among the specified alternative hypotheses $\mathcal{H}_i$ ($i = 1, \ldots, k$) with the purpose of pinpointing the potential source of model error. In doing so, two decisions can be made. Either one of the alternative hypotheses, say $\mathcal{H}_i$, is confidently identified, or none can be identified as such, in which case an 'undecided' decision is made.
3. *Adaptation* In case $\mathcal{H}_i$ is confidently identified, it is chosen as the new null hypothesis. The $\mathcal{H}_0$-based inferences are then accordingly corrected and $\hat{x}_i$ is provided as the estimate of x. However, in case the 'undecided' decision is made, then the solution for x is declared 'unavailable'.

All the information that is needed for the above three steps is contained in the *misclosure* vector $t \in \mathbb{R}^r$ given as

$$t = B^T y; \qquad Q_{tt} = B^T Q_{yy} B \tag{4}$$

where the $m \times r$ matrix B is a basis matrix of the null space of A^T (cf. 1), i.e., $A^T B = 0$ and $\text{rank}(B) = r$, and Q_{tt} is the variance matrix of t. Assuming that the observations are normally distributed as $y \overset{\mathcal{H}_i}{\sim} \mathcal{N}(Ax + C_i b_i, Q_{yy})$ for $i = 0, 1, \ldots, k$ and with $C_0 b_0 = 0$, the misclosure vector is then distributed as

$$t \overset{\mathcal{H}_i}{\sim} \mathcal{N}(\mu_{t_i} = B^T C_i b_i, Q_{tt}), \qquad \text{for} \quad i = 0, 1, \ldots k \tag{5}$$

As t is zero-mean under $\mathcal{H}_0$ and also independent of $\hat{x}_0$, it provides all the available information useful for validation of $\mathcal{H}_0$ (Teunissen 2017). Thus, an unambiguous testing procedure can be established through assigning the outcomes of t to the statistical hypotheses $\mathcal{H}_i$ for $i = 0, 1, \ldots, k$.

2.2 DIA-datasnooping

So far, no assumption was made about the structure of C_i in (2). As the problem of screening observations for possible outliers is an important example of multiple hypothesis testing (see, e.g., Baarda 1968; Van Mierlo 1980; Hawkins 1980; Teunissen 1985; Parkinson and Axelrad 1988; Sturza 1988; Van der Marel and Kosters 1990; Su et al. 2014), we will restrict our attention to this important case. We further assume that only one observation at a time is affected by an outlier. Thus, in (2), b_i is the scalar outlier and C_i takes the form of a canonical unit vector $c_i \in \mathbb{R}^m$ having 1 as its ith entry and zeros elsewhere. This leads to having as many alternative hypotheses as the observations, i.e., $k = m$. This procedure of screening each individual observation for the presence of an outlier is known as *datasnooping* (Baarda 1968; Kok 1984). The corresponding DIA steps are specified as follows:

1. *Detection* Accept $\mathcal{H}_0$ if $t \in \mathcal{P}_0$ with

$$\mathcal{P}_0 = \left\{ t \in \mathbb{R}^r \,|\, \|t\|^2_{Q_{tt}} \leq k_\alpha \right\} \tag{6}$$

in which $\|\cdot\|^2_{Q_{tt}} = (\cdot)^T Q_{tt}^{-1} (\cdot)$ and k_α is the user-chosen α-percentage of the central Chi-square distribution with r degrees of freedom. If $\mathcal{H}_0$ is accepted, then $\hat{x}_0$ is provided as the estimate of x. Otherwise, go to step 2.
2. *Identification* Form Baarda's test statistic as (Baarda 1967; Teunissen 2000)

$$w_i = \frac{c_{t_i}^T Q_{tt}^{-1} t}{\sqrt{c_{t_i}^T Q_{tt}^{-1} c_{t_i}}}, \quad i = 1, \ldots, k \tag{7}$$

in which $c_{t_i} = B^T c_i$. Since c_i is a canonical unit vector, c_{t_i} is then the ith column of matrix B^T. Select $\mathcal{H}_{i \neq 0}$ if $t \in \mathcal{P}_{i \neq 0}$ with

$$\mathcal{P}_{i \neq 0} = \left\{ t \in \mathbb{R}^r / \mathcal{P}_0 \,|\, |w_i| = \max_{j \in \{1, \ldots, k\}} |w_j| \right\} \tag{8}$$

3. *Adaptation* If $\mathcal{H}_i$ is selected, then $\hat{x}_i$ is provided as the estimate of x.

Note, since $t = B^T \hat{e}_0$, with $\hat{e}_0 = y - A\hat{x}_0$, that the above procedure can be formulated by means of the least-squares residual vector $\hat{e}_0$ as well, thus providing a perhaps more recognizable form of the testing procedure (Teunissen 2000). Also note that we assume the variance–covariance matrix Q_{yy} to be known. Variance-component estimation (Teunissen and Amiri-Simkooei 2008) with further modification of the partitioning of misclosure space would need to be included in case of unknown variance components. In the simplest case of a single unknown variance of unit weight, the datasnooping partitioning gets determined by only the w_j statistics, which then will have a studentized distribution instead of a standard normal one (Koch 1999; Teunissen 2000).

Finally note that the vector of misclosures t is not uniquely defined. This, however, does not affect the testing outcome as both the detector $\|t\|^2_{Q_{tt}}$ and Baarda's test statistic w_i remain invariant for any one-to-one transformation of the misclosure vector. Therefore, instead of t, one can for instance also work with

$$\bar{t} = Q_{tt}^{-\frac{1}{2}} t \tag{9}$$

which, given (5), is distributed as $\bar{t} \overset{\mathcal{H}_i}{\sim} \mathcal{N}(\mu_{\bar{t}_i} = Q_{tt}^{-\frac{1}{2}} \mu_{t_i}, I_r)$. The advantage of using $\bar{t}$ over t lies in the ease of visualizing certain effects due to the identity-variance matrix of $\bar{t}$. We will

make use of this in Sect. 5. The partitioning corresponding with $\bar{t}$ is then characterized through

$$\overline{\mathcal{P}}_0 = \left\{ \bar{t} \in \mathbb{R}^r \,|\, \|\bar{t}\|^2 \leq k_\alpha \right\} \tag{10}$$

$$\overline{\mathcal{P}}_{i\neq 0} = \left\{ \bar{t} \in \mathbb{R}^r / \overline{\mathcal{P}}_0 \,|\, |\bar{c}_i^T \bar{t}| = \max_{j\in\{1,\ldots,k\}} |\bar{c}_j^T \bar{t}| \right\} \tag{11}$$

with $\bar{c}_i = Q_{tt}^{-\frac{1}{2}} c_{t_i} / \|c_{t_i}\|_{Q_{tt}}$ being a unit vector and $\|\cdot\|^2 = (\cdot)^T(\cdot)$. As such, $\overline{\mathcal{P}}_0$ contains $\bar{t}$'s inside and on a zero-centered sphere with the radius of $\sqrt{k_\alpha}$ whereas $\overline{\mathcal{P}}_{i\neq 0}$ includes all $\bar{t}$'s outside the mentioned sphere which, among $\bar{c}_j$ for $j = 1, \ldots, k$, make the smallest angle with $\bar{c}_i$. The border between $\overline{\mathcal{P}}_{i\neq 0}$ and $\overline{\mathcal{P}}_{j\neq 0}$ is then the locus of the vectors $\bar{t} \in \mathbb{R}^r / \overline{\mathcal{P}}_0$ which make the same angle with $\bar{c}_i$ and $\bar{c}_j$. Therefore, the partitioning of $\mathbb{R}^r$ is driven by k_α and the relative orientation of $\bar{c}_j$ for $j = 1, \ldots, k$ with respect to each other.

2.3 DIA estimator

As the above three steps show, DIA-datasnooping combines estimation with testing. By using a canonical model formulation and a partitioning of misclosure space, a unifying framework to rigorously capture the probabilistic properties of this combination was presented in Teunissen (2017). It was there also shown how the combined estimation-testing scheme could be captured in one single DIA estimator. The DIA estimator is a function of $\hat{x}_j$ $(j = 0, 1, \ldots, k)$ and the misclosure vector t, and it is given as

$$\bar{x} = \sum_{j=0}^{k} \hat{x}_j \, p_j(t) \tag{12}$$

with $p_j(t)$ being the indicator function of region $\mathcal{P}_j$, i.e., $p_j(t) = 1$ for $t \in \mathcal{P}_j$ and $p_j(t) = 0$ elsewhere. As $\bar{x}$ is linear in $\hat{x}_j$, the DIA estimator of $\theta = F^T x$ with $F \in \mathbb{R}^{n\times p}$ is given as

$$\bar{\theta} = \sum_{j=0}^{k} \hat{\theta}_j \, p_j(t) \tag{13}$$

with $\hat{\theta}_j = F^T \hat{x}_j$. For a general probabilistic evaluation of the DIA estimator, we refer to Teunissen (2017), but see also Teunissen et al. (2017). Here we note, however, that expressions (12) and (13) are only valid under the assumption that the set of regions $\mathcal{P}_i$ $(i = 0, 1, \ldots, k)$ forms a *partitioning* of misclosure space, i.e., $\cup_{i=0}^{k} \mathcal{P}_i = \mathbb{R}^r$ and $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset$ for any $i \neq j$. Note the second condition is considered for the interior points of the distinct regions $\mathcal{P}_i$. The regions $\mathcal{P}_i$ are allowed to have common boundaries since we assume the probability of t lying on one of the boundaries to be zero. That the set of regions $\mathcal{P}_i$ $(i = 0, 1, \ldots, k)$ forms a *partitioning* of misclosure space requires that the canonical unit vectors of the individual hypotheses satisfy certain conditions.

Lemma 1 (Datasnooping partitioning) *The $m + 1$ regions $\mathcal{P}_i$ of (6) and (8) form a partitioning of misclosure space iff $c_{t_i} \nparallel c_{t_j}$ for any $i \neq j$.*

Proof See Appendix. □

It will be clear that the conditions of the above Lemma may not always be fulfilled. The question is then which strategy to follow to deal with such a situation. Should one decide for 'undecidedness' if $c_{t_i} \parallel c_{t_j}$ for some $i \neq j$, or should one re-measure all such involved observables, or would it still be possible to perform an adaptation? An answer to these questions is provided in Sect. 4, where we consider the more general case and not restrict C_i to be the canonical unit vector c_i. First, however, we discuss the testing probabilities that are involved in the detection and identification step.

3 Detection versus identification

3.1 The probabilities

As shown by (6), (7) and (8), the decisions of the testing procedure are driven by the outcome of the misclosure vector t. The probabilities of their occurrence depend on which of the hypotheses is true. If $\mathcal{H}_i$ is true, then the decision is correct if $t \in \mathcal{P}_i$, and wrong if $t \in \mathcal{P}_{j\neq i}$. We therefore discriminate between the following events

$$\begin{array}{lll} \mathrm{CA} &= (t \in \mathcal{P}_0 | \mathcal{H}_0) & = \text{correct acceptance} \\ \mathrm{FA} &= (t \notin \mathcal{P}_0 | \mathcal{H}_0) & = \text{false alarm} \\ \mathrm{MD}_i &= (t \in \mathcal{P}_0 | \mathcal{H}_i) & = \text{missed detection} \\ \mathrm{CD}_i &= (t \notin \mathcal{P}_0 | \mathcal{H}_i) & = \text{correct detection} \\ \mathrm{WI}_i &= (t \in \cup_{j\neq 0,i}^{k} \mathcal{P}_j | \mathcal{H}_i) & = \text{wrong identification} \\ \mathrm{CI}_i &= (t \in \mathcal{P}_i | \mathcal{H}_i) & = \text{correct identification} \end{array} \tag{14}$$

With $* = \{\mathrm{CA}, \mathrm{FA}, \mathrm{MD}_i, \mathrm{CD}_i, \mathrm{WI}_i, \mathrm{CI}_i\}$, we denote the probability of $*$ by P_* satisfying

$$\begin{aligned} &\mathrm{P}_{\mathrm{CA}} + \mathrm{P}_{\mathrm{FA}} = 1 \\ &\mathrm{P}_{\mathrm{MD}_i} + \mathrm{P}_{\mathrm{CD}_i} = 1 \\ &\mathrm{P}_{\mathrm{WI}_i} + \mathrm{P}_{\mathrm{CI}_i} = \mathrm{P}_{\mathrm{CD}_i} \end{aligned} \tag{15}$$

Computation of P_* requires information about the misclosures probability density function (PDF) which is given in (5). Here, it is important to note the difference between the CD and CI probabilities, i.e., $\mathrm{P}_{\mathrm{CD}_i} \geq \mathrm{P}_{\mathrm{CI}_i}$. They would be the same if there is only one alternative hypothesis, say $\mathcal{H}_i$, since then $\mathcal{P}_i = \mathbb{R}^r / \mathcal{P}_0$. Analogous to the CD and CI probabilities, we have the concepts of the minimal detectable bias (MDB)

(Baarda 1968) and the minimal identifiable bias (MIB) (Teunissen 2017). In the following, the difference between the MDB ($\mathrm{P}_{\mathrm{CD}_i}$) and the MIB ($\mathrm{P}_{\mathrm{CI}_i}$) is highlighted by showing the difference between their corresponding contributing factors.

3.2 Minimal detectable bias (MDB)

The MDB of the alternative hypothesis $\mathcal{H}_i$ is defined as the smallest value of $|b_i|$ that can be detected given a certain CD probability. Therefore, the MDB is an indicator of the sensitivity of the *detection* step. Under $\mathcal{H}_{i\neq 0}$ with the definition of $\mathcal{P}_0$ in (6), the probability of correct detection reads

$$\mathrm{P}_{\mathrm{CD}_i} = \mathrm{P}(t \notin \mathcal{P}_0|\mathcal{H}_i) = \mathrm{P}(\|t\|^2_{Q_{tt}} > k_\alpha|\mathcal{H}_i) \tag{16}$$

The MDB of $\mathcal{H}_i$ can then be computed by inverting the above equation for a certain CD probability. With (5), we have $\|t\|^2_{Q_{tt}} \overset{\mathcal{H}_i}{\sim} \chi^2(r, \lambda_i^2)$ with $\lambda_i^2 = \|\mu_{t_i}\|^2_{Q_{tt}}$. For certain $\mathrm{P}_{\mathrm{FA}} = \alpha$, $\mathrm{P}_{\mathrm{CD}_i} = \gamma_{\mathrm{CD}}$ and r, one can compute $\lambda_i^2 = \lambda^2(\alpha, \gamma_{\mathrm{CD}}, r)$ from the Chi-square distribution, and then the MDB is (Baarda 1968; Teunissen 2000)

$$|b_{i,\mathrm{MDB}}| = \frac{\lambda(\alpha, \gamma_{\mathrm{CD}}, r)}{\|c_{t_i}\|_{Q_{tt}}} \tag{17}$$

which shows that for a given set of $\{\alpha, \gamma_{\mathrm{CD}}, r\}$, the MDB depends on $\|c_{t_i}\|_{Q_{tt}}$. One can compare the MDBs of different alternative hypotheses for a given set of $\{\alpha, \gamma_{\mathrm{CD}}, r\}$, which provides information on how sensitive is the rejection of $\mathcal{H}_0$ for the biases the size of $|b_{i,\mathrm{MDB}}|$'s. The smaller the MDB $|b_{i,\mathrm{MDB}}|$ is, the more sensitive is the rejection of $\mathcal{H}_0$.

3.3 Minimal identifiable bias (MIB)

It is important to realize that the MDB provides information about correct detection and *not* correct identification. A high probability of correct detection does therefore not necessarily imply a high probability of correct identification (cf. 15), unless we have the special case of only a single alternative hypothesis. In case of multiple hypotheses, one can define the MIB of the alternative hypothesis $\mathcal{H}_i$ as the smallest value of $|b_i|$ that can be identified given a certain CI probability. It is an indicator of the sensitivity of the *identification* step. The MIB, denoted by $|b_{i,\mathrm{MIB}}|$, can be computed through inverting

$$\mathrm{P}_{\mathrm{CI}_i} = \mathrm{P}(t \in \mathcal{P}_i|\mathcal{H}_i) = \int_{\mathcal{P}_i} f_t(\tau|\mathcal{H}_i)\,\mathrm{d}\tau \tag{18}$$

for a given CI probability. The above probability is an r-fold integral over the complex region $\mathcal{P}_i$. Thus, the inversion of (18) is not as trivial as that of (16). The MIB needs then to be computed through numerical simulations, see, e.g., Teunissen (2017), pp. 73 and Robert and Casella (2013). From $\mathrm{P}_{\mathrm{CD}_i} \geq \mathrm{P}_{\mathrm{CI}_i}$, one can infer that $|b_{i,\mathrm{MDB}}| \leq |b_{i,\mathrm{MIB}}|$ given $\mathrm{P}_{\mathrm{CI}_i} = \gamma_{\mathrm{CD}}$. The identification of mismodeling errors is thus more difficult than their detection (Imparato et al. 2018).

Although computation of (18) is not trivial, we can still assess the behavior of CI probability in relation to the contributing factors. To simplify such assessment, we make use of $\bar{t}$ instead of t and present the CI probability as

$$\mathrm{P}_{\mathrm{CI}_i} = \mathrm{P}(\bar{t} \in \overline{\mathcal{P}}_i|\mathcal{H}_i) = \int_{\overline{\mathcal{P}}_i} f_{\bar{t}}(\tau|\mathcal{H}_i)\,\mathrm{d}\tau \tag{19}$$

With the definition of $\overline{\mathcal{P}}_i$ in (11) and $E(\bar{t}|\mathcal{H}_i) = (b_i\|c_{t_i}\|_{Q_{tt}})\bar{c}_i$, the CI probability, for a given value of b_i, is dependent on the following three factors

- $\overline{\mathcal{P}}_i$: As the integrand function in (19) is positive for all $\tau \in \mathbb{R}^r$, then the integral value will increase as $\overline{\mathcal{P}}_i$ expands.
- The orientation of $\bar{c}_i$ w.r.t. the borders of $\overline{\mathcal{P}}_i$: The unit vector $\bar{c}_i$, lying within the borders of $\overline{\mathcal{P}}_i$, determines the direction of $E(\bar{t}|\mathcal{H}_i)$ about which the PDF $f_{\bar{t}}(\tau|\mathcal{H}_i)$ is symmetric. The following lemma elaborates the role of the orientation of $\bar{c}_i$ in CI probability for $r = 2$. For this case, the regions $\overline{\mathcal{P}}_i$ in (11) are defined in $\mathbb{R}^2$. Each region has then three borders of which one is curved (with $\overline{\mathcal{P}}_0$) and two are straight lines on either sides of $\bar{c}_i$.

Lemma 2 *($\mathrm{P}_{\mathrm{CI}_i}$ as function of the orientation of $\bar{c}_i$) Let β_i be the angle between the two straight borders of $\overline{\mathcal{P}}_i$ and let $\beta_{i,1}$ be the angle between $\bar{c}_i$ and the closest straight border on its right side (see Fig. 2). For a given β_i, k_α and $f_{\bar{t}}(\tau|\mathcal{H}_i)$, the CI probability depends on $\beta_{i,1}$. We then have*

$$\frac{1}{2}\beta_i = \underset{\beta_{i,1}}{\mathrm{argmax}}\ \mathrm{P}_{\mathrm{CI}_i}(\beta_{i,1}) \tag{20}$$

Proof See the Appendix. □

Therefore, for $r = 2$, for a given β_i, k_α and $f_{\bar{t}}(\tau|\mathcal{H}_i)$, the CI probability reaches its maximum if $\bar{c}_i$ is parallel to the bisector line of the angle between the two straight borders of $\overline{\mathcal{P}}_i$.

- $\|c_{t_i}\|_{Q_{tt}}$: The scalar $\|c_{t_i}\|_{Q_{tt}}$ determines the magnitude of $E(\bar{t}|\mathcal{H}_i)$. Therefore, the larger the value of $\|c_{t_i}\|_{Q_{tt}}$, the further the center of $f_{\bar{t}}(\tau|\mathcal{H}_i)$ gets from the origin along $\bar{c}_i$, and the larger the probability mass of $f_{\bar{t}}(\tau|\mathcal{H}_i)$ inside $\overline{\mathcal{P}}_i$ will become.

We will use this insight in the contributing factors of the CI probability to explain some of the phenomena that we come across in our numerical analysis in Sect. 5.

4 Nonseparable hypotheses

4.1 Identifying nonseparable hypotheses

As any testing procedure is driven by the misclosure vector, identification of hypotheses becomes problematic if the misclosure vector has the *same* distribution under *different* hypotheses. According to (5) this happens when for two different hypotheses, say $\mathcal{H}_i$ and $\mathcal{H}_j$ $(i \neq j)$,

$$B^T C_i = B^T C_j X_{i,j} \quad \text{for some invertible } X_{i,j} \in \mathbb{R}^{q\times q} \tag{21}$$

In such a case, the misclosure vector t remains insensitive for the differences between $\mathcal{H}_i$ and $\mathcal{H}_j$, as a consequence of which we have $\mathcal{P}_i = \mathcal{P}_j$. One can then not distinguish between the two hypotheses $\mathcal{H}_i$ and $\mathcal{H}_j$ in the identification step. If this is the case and $t \in \mathcal{P}_i = \mathcal{P}_j$, one may consider the following:

1. *Remeasurement* If in case of datasnooping, $\mathcal{H}_i$ and $\mathcal{H}_j$ are singled out in the identification step, then it is one of the two observables, $y_i = c_i^T y$ or $y_j = c_j^T y$, that is suspected to contain a blunder or outlier. To remedy the situation, one may then decide to replace *both* y_i and y_j by their remeasured values.
2. *Adaptation* If remeasurement is not an option, one might think that adaptation of $\hat{x}_0$ would be an option by extending the design matrix to $[A\ C_i\ C_j]$, so as to cover both the hypotheses $\mathcal{H}_i$ and $\mathcal{H}_j$. But, as the theorem below shows, this is unfortunately not possible as x will then become inestimable. Also note, despite the nonseparability of the two hypotheses, that adaptation on either $[A\ C_i]$ or $[A\ C_j]$ should not be pursued. Such adaptation will still produce a biased result if done for the wrong hypothesis.
3. *Unavailability* Without remeasurement or adaptation, the remaining option is to declare a solution for x to be unavailable.

In the following theorem, we show an equivalence between the nonseparability of hypotheses and the inestimability of parameters.

Theorem 1 (Nonseparable hypotheses and inestimable parameters) *Let* $[A\ \ B]$ *be an invertible matrix, with* A *of order* $m \times n$ *and* B *of order* $m \times (m-n)$ *satisfying* $B^T A = 0$. *Furthermore, for any* $i \neq j$ *and* $i, j = 1, \ldots, l$, *let* C_i *be full-rank matrices of order* $m \times q$ *with* $m - n > q$ *such that rank* $([C_i\ \ C_j]) > q$ *and rank* $([A\ \ C_i]) = n + q$. *Then for any* $i \neq j$ *and* $i, j = 1, \ldots, l$, *for some invertible matrix* $X_{i,j} \in \mathbb{R}^{q\times q}$

$$B^T C_i = B^T C_j X_{i,j} \tag{22}$$

iff

$$\exists\, X \in \mathbb{R}^{n\times q}/\{0\} : [A\ \ C_i\ \ C_j] \begin{bmatrix} X \\ -I_q \\ X_{i,j} \end{bmatrix} = 0 \tag{23}$$

implying that the extended design matrix $[A\ \ C_i\ \ C_j]$ *is rank-deficient.*

Proof See the Appendix. □

The above theorem conveys that if the alternative hypotheses $\mathcal{H}_i$ with $i = 1, \ldots, l$ are not distinguishable, then extending the design matrix A by *any* two or more matrices C_i with $i = 1, \ldots, l$ will result in a rank-deficient design matrix and therefore make unbiased estimability of the parameter vector x impossible. The conclusion reads therefore that if remeasurement is not an option and x is the parameter vector for which a solution is sought, the issue of nonseparable hypotheses should already be tackled at the designing phase of the measurement experiment.

4.2 Adaptation for estimable functions

The above theorem has shown that one can forget about adapting $\hat{x}_0$ for hypotheses that are nonseparable. This concerns, however, the *complete* vector x and not necessarily functions of x. It could still be possible that some relevant components of x or some relevant functions of x remain estimable, despite the rank-deficiency of the extended design matrix. The following theorem specifies which parameters remain estimable after the mentioned extension of the design matrix as well as presents the corresponding adaptation step for these estimable parameters.

Theorem 2 (Adaptation for nonseparable hypotheses) (i) *Estimability: Let* C_i, *with* $i = 1, \ldots, l$, *be full-rank matrices of order* $m \times q$ *with* $m - n > q$ *satisfying* (22) *and* (23). *Also, let* $C \in \mathbb{R}^{m\times l_1 q}$ *be a matrix formed by putting* l_1 *matrices* C_i *column-wise next to each other. Then* $\theta = F^T x$, *with* $F \in \mathbb{R}^{n\times p}$, *is unbiased estimable under the extended model*

$$E(y) = [A\ \ C] \begin{bmatrix} x \\ b \end{bmatrix}; \quad D(y) = Q_{yy} \tag{24}$$

iff

$$F^T V = 0 \tag{25}$$

in which V *is a basis matrix of the null space of* $C^{\perp^T} A$, *i.e.,* $C^{\perp^T} A\, V = 0$, *and* $C^{\perp}$ *is a basis matrix of the orthogonal complement of the range space of* C.
(ii) *Adaptation: The BLUE of* $\theta = F^T x$ *under* (24) *and its variance matrix, denoted as* $\hat{\theta}$ *and* $Q_{\hat{\theta}\hat{\theta}}$, *respectively, can be written in adapted form as*

$$\begin{aligned} \hat{\theta} &= \hat{\theta}_0 \; + \; M\, y \\ Q_{\hat{\theta}\hat{\theta}} &= Q_{\hat{\theta}_0\hat{\theta}_0} \; + \; M\, Q_{yy}\, M^T \end{aligned} \tag{26}$$

with $\hat{\theta}_0 = F^T \hat{x}_0$, $Q_{\hat{\theta}_0\hat{\theta}_0} = F^T Q_{\hat{x}_0\hat{x}_0} F$ *and where* $M = F^T \bar{A}^- P_A^\perp$, *with* $P_A^\perp = I_m - AA^+$, $\bar{A} = P_C^\perp A$, $P_C^\perp = I_m - CC^+$ *with* $C^+ = (C^T Q_{yy}^{-1} C)^{-1} C^T Q_{yy}^{-1}$, $\bar{A}^- = S\,[(\bar{A}S)^T Q_{yy}^{-1} (\bar{A}S)]^{-1} (\bar{A}S)^T Q_{yy}^{-1}$, *with* S *a basis matrix of which the range space is complementary to that of* V.

Proof See the Appendix. □

Note if one opts for the adaptation of $\hat{\theta}_0$ as given above, that one cannot use the expression for the DIA estimator as given in (13) anymore. For example, if the hypotheses $\mathcal{H}_i$, with $i = 1, \cdots, l$, are indistinguishable, i.e., $\mathcal{P}_1 = \ldots = \mathcal{P}_l$, the adaptation according to (26) implies that the DIA estimator in (13) changes to

$$\bar{\theta} \;=\; \hat{\theta}_0\, p_0(t) + \hat{\theta}\, p_1(t) + \sum_{j=l+1}^{k} \hat{\theta}_j\, p_j(t) \tag{27}$$

Thus, the $k+1$ terms in the sum are now reduced to $k-l+2$, with $\hat{\theta}$ being the BLUE under (24).

5 Numerical analysis

In this section, we apply the theory of the previous sections to some selected examples so as to illustrate and explain the performance of the various decision elements in DIA-datasnooping. The insight so obtained will also help us appreciate some of the more complex intricacies of the theory. The following three different cases are considered: height-difference observations of a leveling network, distance measurements of a horizontal geodetic network and pseudorange measurements between a single ground station and GPS satellites. We analyze and illustrate how geometry changes in the measurement setup affect the testing procedure, including its partitioning of the misclosure space, and the corresponding CD probabilities (MDB) and CI probabilities (MIB). The CD probability under $\mathcal{H}_i$ $(i = 1, \ldots, k)$ is computed based on (16) from $\chi^2(r, \lambda_i^2)$, whereas the CI probability under $\mathcal{H}_i$ $(i = 1, \ldots, k)$ is computed as described in the Appendix.

5.1 Leveling network

Suppose that we have two leveling loops containing $n \geq 2$ height-difference observations each and sharing one observation with each other (see Fig. 1). For such leveling network, two misclosures can be formed stating that the sum of observations in each loop equals zero. Assuming that all the observations are uncorrelated and of the same precision σ, a misclosure vector t and its variance matrix Q_{tt} can be formed as

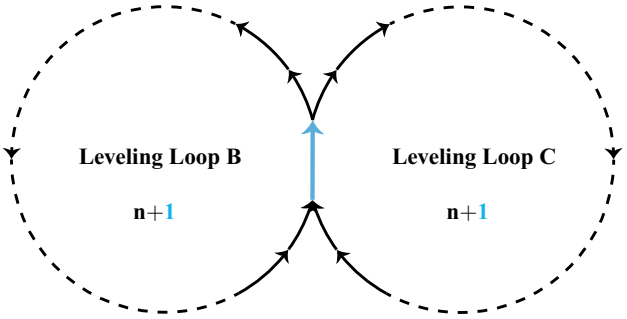

Fig. 1 A leveling network consisting of two leveling loops with n observations each and one shared observation (blue)

$$\begin{aligned} t &= \begin{bmatrix} 1 & \overbrace{1\,1\cdots 1}^{n} & \overbrace{0\,0\cdots 0}^{n} \\ 1 & 0\,0\cdots 0 & 1\,1\cdots 1 \end{bmatrix} \begin{bmatrix} y_{\mathsf{A}} \\ y_{\mathsf{B}} \\ y_{\mathsf{C}} \end{bmatrix} \\ Q_{tt} &= \sigma^2 \begin{bmatrix} n+1 & 1 \\ 1 & n+1 \end{bmatrix} \end{aligned} \tag{28}$$

where y_{A} is the observation shared between the two leveling loops, and y_{B} and y_{C} the n-vectors of observations of the leveling loops B and C, respectively. The number of datasnooping alternative hypotheses for the above model is equal to $2n+1$. But it will be clear of course that not all of them are separately identifiable. Looking at the structure of B^T in (28), it can be seen that out of $2n+1$ vectors c_{t_i} (columns of B^T), only the following three are nonparallel

$$c_{t_{\mathsf{A}}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\; c_{t_{\mathsf{B}}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\; c_{t_{\mathsf{C}}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{29}$$

which implies that in each leveling loop excluding the shared observation y_{A}, an outlier on each of the observations is sensed in the same way by the vector of misclosures. In other words, the testing procedure cannot distinguish between the outliers on the observations in y_{B}, and between those on the observations in y_{C}. Therefore, among the $2n+1$ alternative hypotheses, we retain three: $\mathcal{H}_{\mathsf{A}}$ corresponding with y_{A}, $\mathcal{H}_{\mathsf{B}}$ corresponding with one of the observations in y_{B} and $\mathcal{H}_{\mathsf{C}}$ corresponding with one of the observations in y_{C}.

5.1.1 Misclosure space partitioning

Given (29), the datasnooping partitioning of the misclosure space is formed by *four* distinct regions $\mathcal{P}_i$ with $i \in \{0, \mathsf{A}, \mathsf{B}, \mathsf{C}\}$. For the sake of visualization, instead of t, we work with $\bar{t}$ (cf. 9). The datasnooping partitioning, as mentioned earlier, is then driven by the relative orientation of $\bar{c}_{\mathsf{A}}$, $\bar{c}_{\mathsf{B}}$ and $\bar{c}_{\mathsf{C}}$ (cf. 11). The angles between these unit vectors are computed as

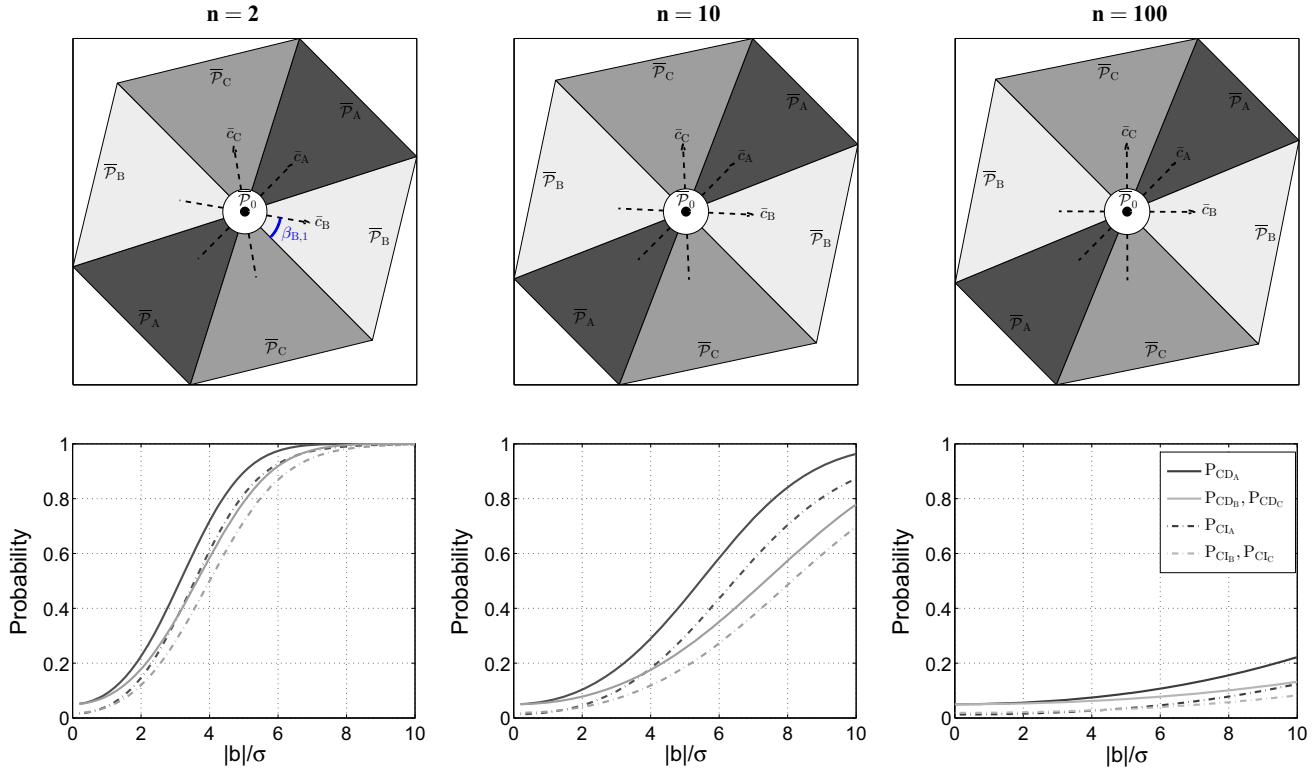

Fig. 2 Visualization of the datasnooping testing procedure defined in Sect. 2.2 for the leveling network shown in Fig. 1 assuming $\alpha = 0.05$ and $\sigma = 5$ mm. [Top] Datasnooping partitioning of the misclosure space $\mathbb{R}^2$ corresponding with $\bar{t}$ (cf. 9). [Bottom] The graphs of CD (solid lines) and CI probability (dashed lines) of different alternative hypotheses as function of bias-to-noise ratio

$$\angle(\bar{c}_{\mathsf{A}}, \bar{c}_{\mathsf{B}}) = \angle(\bar{c}_{\mathsf{A}}, \bar{c}_{\mathsf{C}}) = \cos^{-1}\sqrt{\frac{n}{2(n+1)}}$$
$$\angle(\bar{c}_{\mathsf{B}}, \bar{c}_{\mathsf{C}}) = \cos^{-1}\frac{-1}{n+1} \tag{30}$$

As (30) suggests, when $n \to \infty$, the angles $\angle(\bar{c}_{\mathsf{A}}, \bar{c}_{\mathsf{B}})$ and $\angle(\bar{c}_{\mathsf{A}}, \bar{c}_{\mathsf{C}})$ go to 45°, and the angle $\angle(\bar{c}_{\mathsf{B}}, \bar{c}_{\mathsf{C}})$ goes to 90°. Figure 2 demonstrates the impact of n on the misclosure space partitioning given $\alpha = 0.05$, $r = 2$ and $\sigma = 5$ mm. Using different shades of *gray* color, the *first row* of Fig. 2 shows, for $n = 2$, $n = 10$ and $n = 100$, the partitioning of the misclosure space formed by $\overline{\mathcal{P}}_i$ with $i \in \{0, \mathsf{A}, \mathsf{B}, \mathsf{C}\}$.

5.1.2 CD and CI probabilities

According to (17), for a given $\lambda(\alpha, \gamma_{\mathrm{CD}}, r)$, the MDB depends only on $\|c_{t_i}\|_{Q_{tt}}$. For the leveling network characterized in (28) and its corresponding vectors c_{t_i} in (29), we have

$$\|c_{t_{\mathsf{A}}}\|_{Q_{tt}} = \sigma^{-1}\sqrt{\frac{2}{n+2}}$$
$$\|c_{t_{\mathsf{B}}}\|_{Q_{tt}} = \|c_{t_{\mathsf{C}}}\|_{Q_{tt}} = \sigma^{-1}\sqrt{\frac{n+1}{n(n+2)}} \tag{31}$$

which clearly shows that for a given set of $\{\alpha, \gamma_{\mathrm{CD}}, r\}$, smaller $\mathcal{H}_{\mathsf{A}}$-biases can be detected compared to $\mathcal{H}_{\mathsf{B}}$ and $\mathcal{H}_{\mathsf{C}}$. Equivalently, it can be stated that for a given $\{\alpha, r\}$ and $b_i = b$, the CD probability of $\mathcal{H}_{\mathsf{A}}$ is larger than that of $\mathcal{H}_{\mathsf{B}}$ and $\mathcal{H}_{\mathsf{C}}$. That is because each observation in y_{B} and y_{C} contributes to only one leveling loop while y_{A} contributes to two leveling loops, thus being checked by the observations of both loops. The solid curves in Fig. 2 (second row) depict $\mathrm{P}_{\mathrm{CD}_i}$ as function of the bias-to-noise ratio $|b|/\sigma$. The *dark gray* graphs correspond with $\mathcal{H}_{\mathsf{A}}$, while the *light gray* graphs correspond with $\mathcal{H}_{\mathsf{B}}$ and $\mathcal{H}_{\mathsf{C}}$. These graphs can be used as follows. For a certain $b_i = b$, one can compare the corresponding $\mathrm{P}_{\mathrm{CD}_i}$ of different alternative hypotheses. One can also take the reverse route by comparing the MDB of different alternative hypotheses for a certain $\mathrm{P}_{\mathrm{CD}_i} = \gamma_{\mathrm{CD}}$. In agreement with (31), the solid *dark gray* graphs always lie above the solid *light gray* ones. As the number of observations increases in each loop ($n \uparrow$), the corresponding $\mathrm{P}_{\mathrm{CD}_i}$ decreases for a given $\mathcal{H}_i$-bias. This is due to the fact that the variance of the misclosure vector is an increasing function of the number of observations in each loop (see 28). The lower the precision of the misclosures, the lower is the sensitivity of the testing procedure to a given bias in an observation.

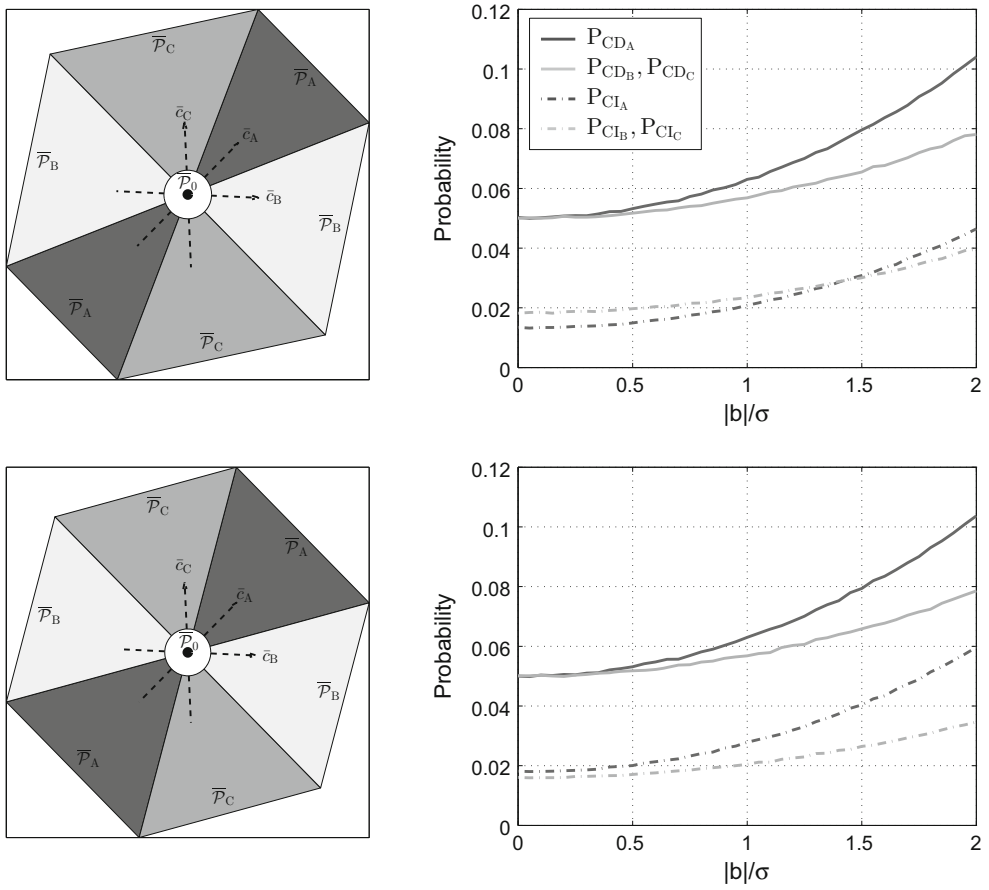

Fig. 3 Comparing two testing scheme for the leveling network in Fig. 1 assuming $n = 10, \alpha = 0.05$ and $\sigma = 5$ mm. [Top] Datasnooping testing procedure defined in Sect. 2.2. [Bottom] The testing procedure defined by (32). [Left] Partitioning of the misclosure space corresponding with $\bar{t}$. [Right] The graphs of CD (solid lines) and CI probabilities (dashed lines) of different alternative hypotheses as function of bias-to-noise ratio

The dashed curves in Fig. 2 (second row) depict $\mathrm{P}_{\mathrm{CI}_i}$ as function of $|b|/\sigma$. These curves ($\mathrm{P}_{\mathrm{CI}_i}$) always lie below their solid counterparts ($\mathrm{P}_{\mathrm{CD}_i}$). Like the solid graphs, these dashed graphs can be used either for comparing the MIB of different alternative hypotheses given a certain $\mathrm{P}_{\mathrm{CI}_i} = \gamma_{\mathrm{CI}}$, or for comparing the corresponding $\mathrm{P}_{\mathrm{CI}_i}$ of different alternative hypotheses given a certain $b_i = b$. We note that despite the CD probability of $\mathcal{H}_\mathsf{A}$ being *always* larger than that of $\mathcal{H}_\mathsf{B}$ and $\mathcal{H}_\mathsf{C}$, the CI probability of $\mathcal{H}_\mathsf{A}$ is *not always* larger than that of $\mathcal{H}_\mathsf{B}$ and $\mathcal{H}_\mathsf{C}$. Depending on the number of measurements in each loop n, if $|b|/\sigma$ is smaller than a certain value, then we have $\mathrm{P}_{\mathrm{CI}_\mathsf{A}} < \mathrm{P}_{\mathrm{CI}_\mathsf{B}} = \mathrm{P}_{\mathrm{CI}_\mathsf{C}}$. This discrepancy between the behavior of CD probability and that of CI probability as function of $|b|/\sigma$ for a given α is due to the fact that while $\mathrm{P}_{\mathrm{CD}_i}$ is driven only by $\|c_{t_i}\|_{Q_{tt}}$, $\mathrm{P}_{\mathrm{CI}_i}$ is in addition driven by $\overline{\mathcal{P}}_i$ and the orientation of $\bar{c}_i$ w.r.t. the straight borders of $\overline{\mathcal{P}}_i$ (cf. 19). Looking at the first row of Fig. 2, we note that $\overline{\mathcal{P}}_\mathsf{A}$ has smaller area compared to $\overline{\mathcal{P}}_\mathsf{B}$ and $\overline{\mathcal{P}}_\mathsf{C}$. Therefore, $|b|$ should be large enough such that $\|c_{t_\mathsf{A}}\|_{Q_{tt}} > \|c_{t_\mathsf{B}}\|_{Q_{tt}} = \|c_{t_\mathsf{C}}\|_{Q_{tt}}$ can compensate for $\overline{\mathcal{P}}_\mathsf{A}$ being smaller than $\overline{\mathcal{P}}_\mathsf{B}$ and $\overline{\mathcal{P}}_\mathsf{C}$.

5.1.3 Impact of partitioning on CI probability

As was mentioned, $\mathrm{P}_{\mathrm{CI}_i}$ depends on $\overline{\mathcal{P}}_i$, the orientation of $\bar{c}_i$ and the magnitude of $\|c_{t_i}\|_{Q_{tt}}$. While the last two factors are driven by the underlying model, the first one depends on the testing procedure. Our above conclusions about the CI probability will then change if we go for another testing scheme. For example, let $\overline{\mathcal{P}}_0$ be defined by (10) and

$$\overline{\mathcal{P}}_{i\neq 0} = \left\{ \bar{t} \in \mathbb{R}^2 / \overline{\mathcal{P}}_0 \,|\, |d_i^T \bar{t}| = \max_{k\in\{\mathsf{A},\mathsf{B},\mathsf{C}\}} |d_k^T \bar{t}| \right\} \tag{32}$$

where $d_\mathsf{A} = \bar{c}_\mathsf{A}$, $d_\mathsf{B} = R_{(-60^\circ)}\bar{c}_\mathsf{A}$ and $d_\mathsf{C} = R_{(60^\circ)}\bar{c}_\mathsf{A}$ with R_θ being the counterclockwise rotation matrix. This testing scheme leads to $\overline{\mathcal{P}}_\mathsf{A}$, $\overline{\mathcal{P}}_\mathsf{B}$ and $\overline{\mathcal{P}}_\mathsf{C}$ to be of the same shape. In addition, while $\bar{c}_\mathsf{A}$ is parallel to the bisector line of the angle between the two straight borders of $\overline{\mathcal{P}}_\mathsf{A}$, $\bar{c}_\mathsf{B}$ and $\bar{c}_\mathsf{C}$ are close to one of the straight borders of their corresponding region. This combined with the fact that $\|c_{t_\mathsf{A}}\|_{Q_{tt}} > \|c_{t_\mathsf{B}}\|_{Q_{tt}} = \|c_{t_\mathsf{C}}\|_{Q_{tt}}$ lead us to the conclusion that $P_{\mathrm{CI}_\mathsf{A}} > P_{\mathrm{CI}_\mathsf{B}} = P_{\mathrm{CI}_\mathsf{C}}$ holds for any given bias b. Figure 3 shows the difference

between the testing procedure based on (11) and (32), in terms of misclosure space partitioning [left] and CD and CI probability [right].

5.2 Horizontal geodetic network

Consider a horizontal geodetic network containing m reference points from which we measure distances toward an unknown point to determine its horizontal coordinates. Assuming that all the measurements are uncorrelated and of the same precision, the design matrix and the observations variance matrix of the linearized model under $\mathcal{H}_0$ read

$$A = \begin{bmatrix} -u_1^T \\ \vdots \\ -u_m^T \end{bmatrix}, \qquad Q_{yy} = \sigma^2 I_m \tag{33}$$

where the unit direction 2-vector from the unknown point to the reference point ($i = 1, \ldots, m$) is denoted by u_i. In this observational model, the redundancy is $r = m - 2$ revealing that the misclosure vector t is of dimension $m - 2$.

5.2.1 Misclosure space partitioning

For the model in (33), the angles between the corresponding $\bar{c}_i$ vectors are computed as

$$\cos \angle(\bar{c}_i, \bar{c}_j) = \frac{-u_i^T C_{xx}^{-1} u_j}{\sqrt{(1 - \|u_i\|^2_{C_{xx}}) \times (1 - \|u_j\|^2_{C_{xx}})}} \tag{34}$$

which is a consequence of $B\,Q_{tt}^{-1}\,B^T = Q_{yy}^{-1} - Q_{yy}^{-1} A\,Q_{xx}\,A^T\,Q_{yy}^{-1}$ with $Q_{xx} = \sigma^2\,C_{xx}^{-1}$ and $C_{xx} = \sum_{k=1}^{m} u_k u_k^T$. Assuming that the horizontal geodetic network comprises $m = 4$ reference points, Fig. 4 presents the same information as Fig. 2 but for geodetic networks corresponding with (33). The first row shows the orientation of vectors u_i. The standard deviation of the distance measurements is considered to be $\sigma = 5\,\text{mm}$, and the false alarm is set to $\alpha = 0.05$. In (a), the geometry of the measuring points leads to a cofactor matrix of $C_{xx} = 2\,I_2$ of which the substitution in (34) gives $\cos \angle(\bar{c}_i, \bar{c}_j) = -\cos \angle(u_i, u_j)$. Given that the angle between consecutive vectors u_i is 45°, the four regions $\overline{\mathcal{P}}_{i \neq 0}$ have then the same shape. Moving the reference point D to a new location such that $u_\mathsf{D} = -u_\mathsf{A}$ as illustrated in (b), the two regions $\overline{\mathcal{P}}_\mathsf{B}$ and $\overline{\mathcal{P}}_\mathsf{C}$, as Theorem 1 states, become identical. The proof is given as follows. Let $u_\mathsf{D} = p\,u_\mathsf{A}$ ($p = \pm 1$). As the vectors c_{t_i} are the columns of B^T and given that $B^T A = 0$, we have

$$(c_{t_\mathsf{A}} + p\,c_{t_\mathsf{D}})u_\mathsf{A}^T + c_{t_\mathsf{B}} u_\mathsf{B}^T + c_{t_\mathsf{C}} u_\mathsf{C}^T = 0 \tag{35}$$

Multiplying both sides of the above equation with $u_\mathsf{A}^\perp$ from the right, we get

$$c_{t_\mathsf{C}} = -\frac{u_\mathsf{B}^T u_\mathsf{A}^\perp}{u_\mathsf{C}^T u_\mathsf{A}^\perp} c_{t_\mathsf{B}} \tag{36}$$

which means that $c_{t_\mathsf{B}} \parallel c_{t_\mathsf{C}}$, thus $\bar{c}_\mathsf{B} \parallel \bar{c}_\mathsf{C}$ and $\overline{\mathcal{P}}_\mathsf{B} = \overline{\mathcal{P}}_\mathsf{C}$. If in addition, we have $u_\mathsf{C} = q\,u_\mathsf{B}$ ($q = \pm 1$), then (35) simplifies to

$$(c_{t_\mathsf{A}} + p\,c_{t_\mathsf{D}})u_\mathsf{A}^T + (c_{t_\mathsf{B}} + q\,c_{t_\mathsf{C}})u_\mathsf{B}^T = 0 \tag{37}$$

Multiplying the above once with $u_\mathsf{A}^\perp$ and once with $u_\mathsf{B}^\perp$ from the right, then we get $c_{t_\mathsf{A}} \parallel c_{t_\mathsf{D}}$ and $c_{t_\mathsf{B}} \parallel c_{t_\mathsf{C}}$, thus $\bar{c}_\mathsf{A} \parallel \bar{c}_\mathsf{D}$ and $\bar{c}_\mathsf{B} \parallel \bar{c}_\mathsf{C}$. From (b) to (c), as the angle between u_B and u_C decreases, the errors in the measurements A and D become less distinguishable from each other, but better separable from those in the measurements of B and C.

5.2.2 CD and CI probabilities

The illustrations on the third row of Fig. 4 show the graphs of $\mathrm{P}_{\mathrm{CD}_i}$ (solid lines) and $\mathrm{P}_{\mathrm{CI}_i}$ (dashed lines) under all the four alternative hypotheses $\mathcal{H}_i$ with $i \in \{\mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}\}$. The CD probability $\mathrm{P}_{\mathrm{CD}_i}$ corresponding with (33) for a given α, r and a bias value $|b|$ is driven by (cf. 17)

$$\|c_{t_i}\|_{Q_{tt}} = \sigma^{-1} \left[1 - \frac{1}{\det(C_{xx})} \sum_{j=1}^{m} \sin^2 \angle(u_i, u_j)\right]^{\frac{1}{2}} \tag{38}$$

with det(.) being the determinant operator. In (a), owing to 45° angle between the consecutive vectors u_i, we have $\|c_{t_i}\|_{Q_{tt}} = \|c_{t_j}\|_{Q_{tt}}$ for any $i \neq j$, hence $\mathrm{P}_{\mathrm{CD}_i} = \mathrm{P}_{\mathrm{CD}_j}$ for any given value of bias $|b|$ and $i \neq j$. Furthermore, as a consequence of having a symmetric partitioning, we also have $\mathrm{P}_{\mathrm{CI}_i} = \mathrm{P}_{\mathrm{CI}_j}$ for any given value of bias $|b|$ and $i \neq j$. In (b) and (c), given that $u_\mathsf{A} \parallel u_\mathsf{D}$ and $u_\mathsf{A} \perp u_\mathsf{C}$, we have $\|c_{t_\mathsf{A}}\|_{Q_{tt}} = \|c_{t_\mathsf{B}}\|_{Q_{tt}} = \|c_{t_\mathsf{D}}\|_{Q_{tt}}$ conveying that the hypotheses $\mathcal{H}_\mathsf{A}$, $\mathcal{H}_\mathsf{B}$ and $\mathcal{H}_\mathsf{D}$ have the same CD probability. $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{D}$ have, in addition, the same CI probability since $\overline{\mathcal{P}}_\mathsf{A}$ and $\overline{\mathcal{P}}_\mathsf{D}$ have the same shape and also the orientation of $\bar{c}_\mathsf{A}$ inside $\overline{\mathcal{P}}_\mathsf{A}$ is the same as that of $\bar{c}_\mathsf{D}$ inside $\overline{\mathcal{P}}_\mathsf{D}$.

In (b) and (c), $\mathcal{H}_\mathsf{B}$ is not distinguishable from $\mathcal{H}_\mathsf{C}$. For these hypotheses, although not identifiable from each other, we still define CI probability as $\mathrm{P}_{\mathrm{CI}_\mathsf{B}} = \mathrm{P}(\bar{t} \in \overline{\mathcal{P}}_\mathsf{B} | \mathcal{H}_\mathsf{B})$ and $\mathrm{P}_{\mathrm{CI}_\mathsf{C}} = \mathrm{P}(\bar{t} \in \overline{\mathcal{P}}_\mathsf{B} | \mathcal{H}_\mathsf{C})$. It can be seen that, although $\mathcal{H}_\mathsf{B}$ is not distinguishable from $\mathcal{H}_\mathsf{C}$, they are different in both the CD and CI probabilities. Also, the testing procedure is more sensitive to the biases in y_B compared to the same biases in y_C. This is due to the fact that the observation of C contributes to the misclosure vector less than the observation of B. The contribution of the measurement of C to the misclosure vector

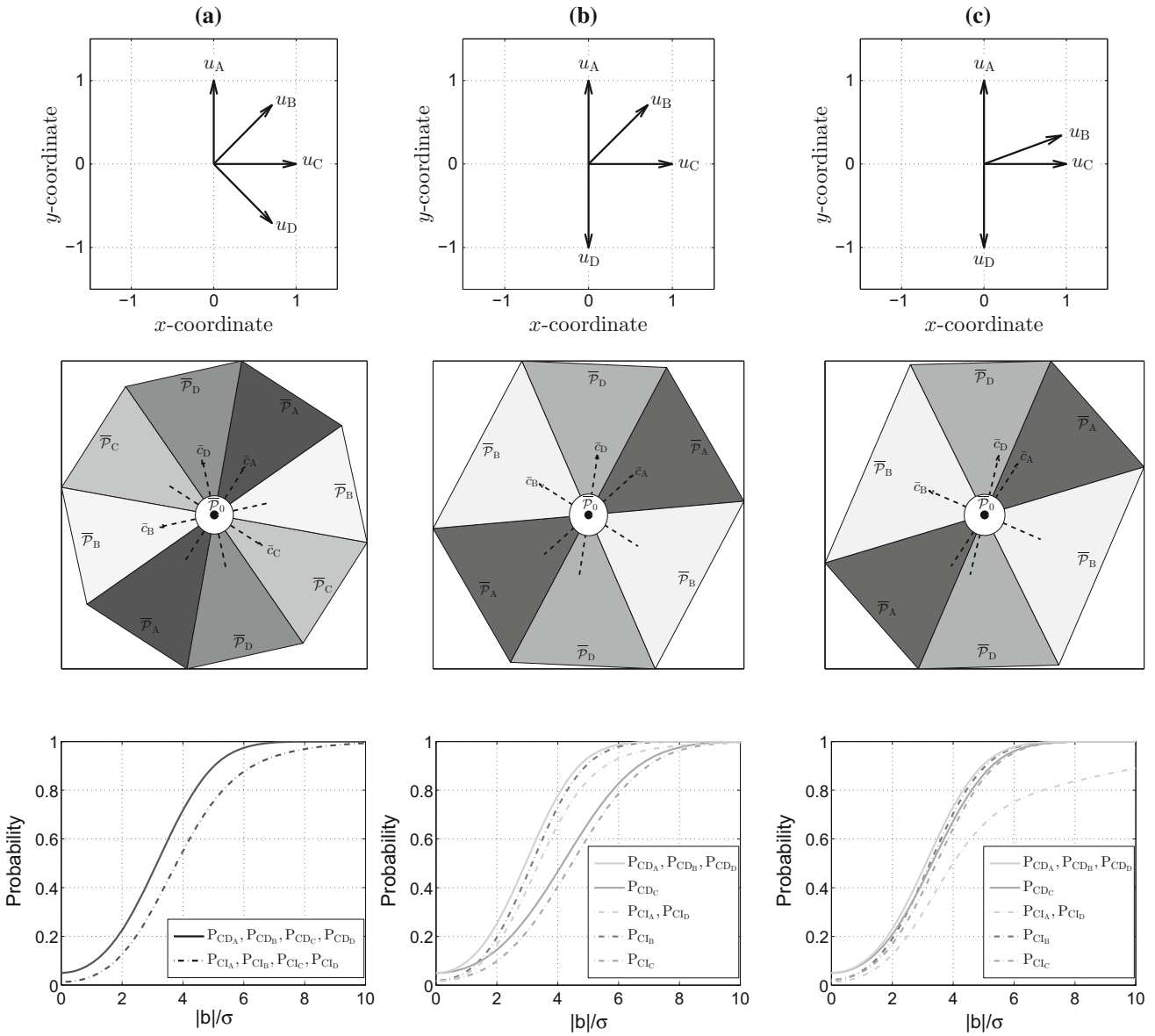

Fig. 4 Visualization of the datasnooping testing procedure defined in Sect. 2.2 for the horizontal geodetic networks shown in the first row assuming $\alpha = 0.05$ and $\sigma = 5\,\text{mm}$. [Top] Geometry of the four reference points w.r.t. the point of which the coordinates are to be estimated. [Middle] Datasnooping partitioning of the misclosure space $\mathbb{R}^2$ corresponding with $\bar{t}$ (cf. 9). [Bottom] The graphs of CD (solid lines) and CI probability (dashed lines) of different alternative hypotheses as function of bias-to-noise ratio

depends on the relative orientation of u_B w.r.t. u_C. In case u_B is parallel to u_A and u_D, the measurement of the point C would have zero contribution to the misclosure vector and cannot be screened at all. As the angle between u_B and u_C decreases, the mentioned contribution increases, so does the sensitivity of the testing procedure to the biases in the measurement of C.

Note that for the geometries shown in (b) and (c), if the misclosure vector lies in $\overline{\mathcal{P}}_B$, it cannot be inferred that whether y_B or y_C is biased. For adaptation, one may extend the design matrix A to $[A \;\; c_B \;\; c_C]$, which would be of relevance if the parameters of interest remain estimable (see Theorem 2). As c_B and c_C are canonical unit vectors, then $[c_B \;\; c_C]^{\perp^T} A$ is a matrix achieved by removing the rows of A corresponding with y_B and y_C as

$$[c_B \;\; c_C]^{\perp^T} A = \begin{bmatrix} -u_A^T \\ -u_D^T \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix} \tag{39}$$

which clearly shows that the x-coordinate is not estimable. However, the above adaptation strategy is still of relevance if one is interested in the y-coordinate.

A summary of the above qualitative findings in relation to the geometry of the measuring points is given as follows

- If $|\cos\angle(u_i, u_{i+1})| = \cos 45°$ for any $i = 1, 2, 3$, then
  - $|\cos\angle(\bar{c}_i, \bar{c}_{i+1})| = \cos 45°$
  - $\overline{\mathcal{P}}_i$ has the same shape of $\overline{\mathcal{P}}_j$ for any $i \neq j$
  - $\mathrm{P}_{\mathrm{CD}_i} = \mathrm{P}_{\mathrm{CD}_j}$ and $\mathrm{P}_{\mathrm{CI}_i} = \mathrm{P}_{\mathrm{CI}_j}$ for any $i \neq j$
- If $u_\mathsf{A} \parallel u_\mathsf{D}$, then
  - $\overline{\mathcal{P}}_\mathsf{B} = \overline{\mathcal{P}}_\mathsf{C}$
  - $\mathrm{P}_{\mathrm{CD}_\mathsf{A}} = \mathrm{P}_{\mathrm{CD}_\mathsf{D}}$ and $\mathrm{P}_{\mathrm{CI}_\mathsf{A}} = \mathrm{P}_{\mathrm{CI}_\mathsf{D}}$
- If $u_\mathsf{A} \parallel u_\mathsf{D}$ and $u_\mathsf{B} \parallel u_\mathsf{C}$, then
  - $\overline{\mathcal{P}}_\mathsf{A} = \overline{\mathcal{P}}_\mathsf{D}$ and $\overline{\mathcal{P}}_\mathsf{B} = \overline{\mathcal{P}}_\mathsf{C}$
  - $\overline{\mathcal{P}}_\mathsf{A}$ has the same shape of $\overline{\mathcal{P}}_\mathsf{B}$
  - $\mathrm{P}_{\mathrm{CD}_i} = \mathrm{P}_{\mathrm{CD}_j}$ and $\mathrm{P}_{\mathrm{CI}_i} = \mathrm{P}_{\mathrm{CI}_j}$ for any $i \neq j$
- If $u_\mathsf{A} \parallel u_\mathsf{D}$ and $u_\mathsf{C} \perp u_\mathsf{A}$, then
  - $\mathrm{P}_{\mathrm{CD}_\mathsf{A}} = \mathrm{P}_{\mathrm{CD}_\mathsf{B}} = \mathrm{P}_{\mathrm{CD}_\mathsf{D}}$
  - $\mathrm{P}_{\mathrm{CD}_\mathsf{B}} \geq \mathrm{P}_{\mathrm{CD}_\mathsf{C}}$ and $\mathrm{P}_{\mathrm{CI}_\mathsf{B}} \geq \mathrm{P}_{\mathrm{CI}_\mathsf{C}}$.
  - If $\angle(u_\mathsf{B}, u_\mathsf{C})$ decreases, so does the differences $\mathrm{P}_{\mathrm{CD}_\mathsf{B}} - \mathrm{P}_{\mathrm{CD}_\mathsf{C}}$ and $\mathrm{P}_{\mathrm{CI}_\mathsf{B}} - \mathrm{P}_{\mathrm{CI}_\mathsf{C}}$.
- If $u_\mathsf{A} \parallel u_\mathsf{B}, u_\mathsf{A} \parallel u_\mathsf{D}$ and $u_\mathsf{C} \perp u_\mathsf{A}$, then $\mathrm{P}_{\mathrm{CD}_\mathsf{C}} = \mathrm{P}_{\mathrm{CI}_\mathsf{C}} = 0$.

5.3 GPS single-point positioning

Let the pseudorange observations of m GPS satellites be collected by one single receiver to estimate its three-dimensional position coordinates and clock error. Assuming that all the code observations are uncorrelated and of the same precision σ, the corresponding linearized observational model, also known as the single-point positioning (SPP) model, under $\mathcal{H}_0$ is characterized through the following full-rank design matrix and the observations variance matrix

$$A = \begin{bmatrix} -u_1^T & 1 \\ \vdots & \vdots \\ -u_m^T & 1 \end{bmatrix}, \qquad Q_{yy} = \sigma^2 I_m \tag{40}$$

in which the 3-vectors u_i $(i = 1, \ldots, m)$ are the receiver-satellite unit direction vectors. The first three columns of A correspond with the receiver North-East-Up coordinate increments while the last one corresponds with the receiver clock error increment. Given that the design matrix A is of order $m \times 4$, the redundancy of the SPP model is $r = m - 4$.

5.3.1 Misclosure space partitioning

With the SPP model in (40), the angles between the vectors $\bar{c}_i$ are computed as

$$\cos\angle(\bar{c}_i, \bar{c}_j) = \frac{-\frac{1}{m} - (u_i - \bar{u})^T C_{xx}^{-1}(u_j - \bar{u})}{\sqrt{\left(\frac{m-1}{m} - \|u_i - \bar{u}\|^2_{C_{xx}}\right) \times \left(\frac{m-1}{m} - \|u_j - \bar{u}\|^2_{C_{xx}}\right)}} \tag{41}$$

in which $C_{xx} = \sum_{k=1}^m (u_k - \bar{u})(u_k - \bar{u})^T$ and $\bar{u} = \frac{1}{m}\sum_{k=1}^m u_k$. Assuming that six GPS satellites are transmitting signals to a single receiver ($m = 6$), two misclosures can be formed, i.e., $r = 2$. Figure 5, for three different geometries of these satellites (first row), shows the partitioning of the misclosure space (second row). The satellite geometries in (a) and (b) are artificial while that in (c), except for the name of satellites, is a real GPS geometry at Perth, Australia.

In (a), despite having six pseudorange observations, the partitioning is formed by five distinct regions. The regions corresponding with $\mathcal{H}_5$ and $\mathcal{H}_6$ coincide each other, i.e., $\overline{\mathcal{P}}_5 = \overline{\mathcal{P}}_6$, which can be explained as follows. The lines-of-sight of the four satellites G1, G2, G3, and G4 lie on a cone of which the symmetry axis is indicated as the *red* circle. Therefore, we have

$$u_i^T d = c; \qquad i = 1, \ldots, 4 \tag{42}$$

with d the unit 3-vector of the symmetry axis of the mentioned cone and c the cosine of the half the vertex angle of the cone. The extended SPP design matrix $[A \;\; c_5 \;\; c_6]$ will then satisfy

$$[A \;\; c_5 \;\; c_6] \begin{bmatrix} d \\ c \\ u_5^T d - c \\ u_6^T d - c \end{bmatrix} = 0 \tag{43}$$

Therefore, the 6×6 matrix $[A \;\; c_5 \;\; c_6]$ is rank-deficient which, according to Theorem 1, implies that the two alternative hypotheses $\mathcal{H}_5$ and $\mathcal{H}_6$ are not separable. If the misclosure vector lies in $\overline{\mathcal{P}}_5$, it cannot be inferred that whether observation y_5 or y_6 is biased. For adaptation, one may use the above-extended design matrix in case the parameters of interest remain estimable (see Theorem 2). As c_5 and c_6 are canonical unit vectors, then $[c_5 \;\; c_6]^{\perp^T} A$ is a matrix achieved by removing the last two rows of A. Based on such reduced design matrix, according to (42), the position solution in the direction of d is indeterminate. Since d is vertically oriented, the horizontal coordinates (East-North) remain estimable based on the first four rows of A.

In (b), all the alternative hypotheses are distinguishable. In (c), the two vectors $\bar{c}_3$ and $\bar{c}_5$ are almost parallel which is due to the satellites G1, G2, G4 and G6 forming a cone-like geometry of which the axis is indicated by a *red* circle.

5.3.2 CD and CI probabilities

The graphs of $\mathrm{P}_{\mathrm{CD}_i}$ and $\mathrm{P}_{\mathrm{CI}_i}$ for $i = 1, \ldots, 6$ as function of the bias-to-noise ratio are given in the third row of Fig. 5. One

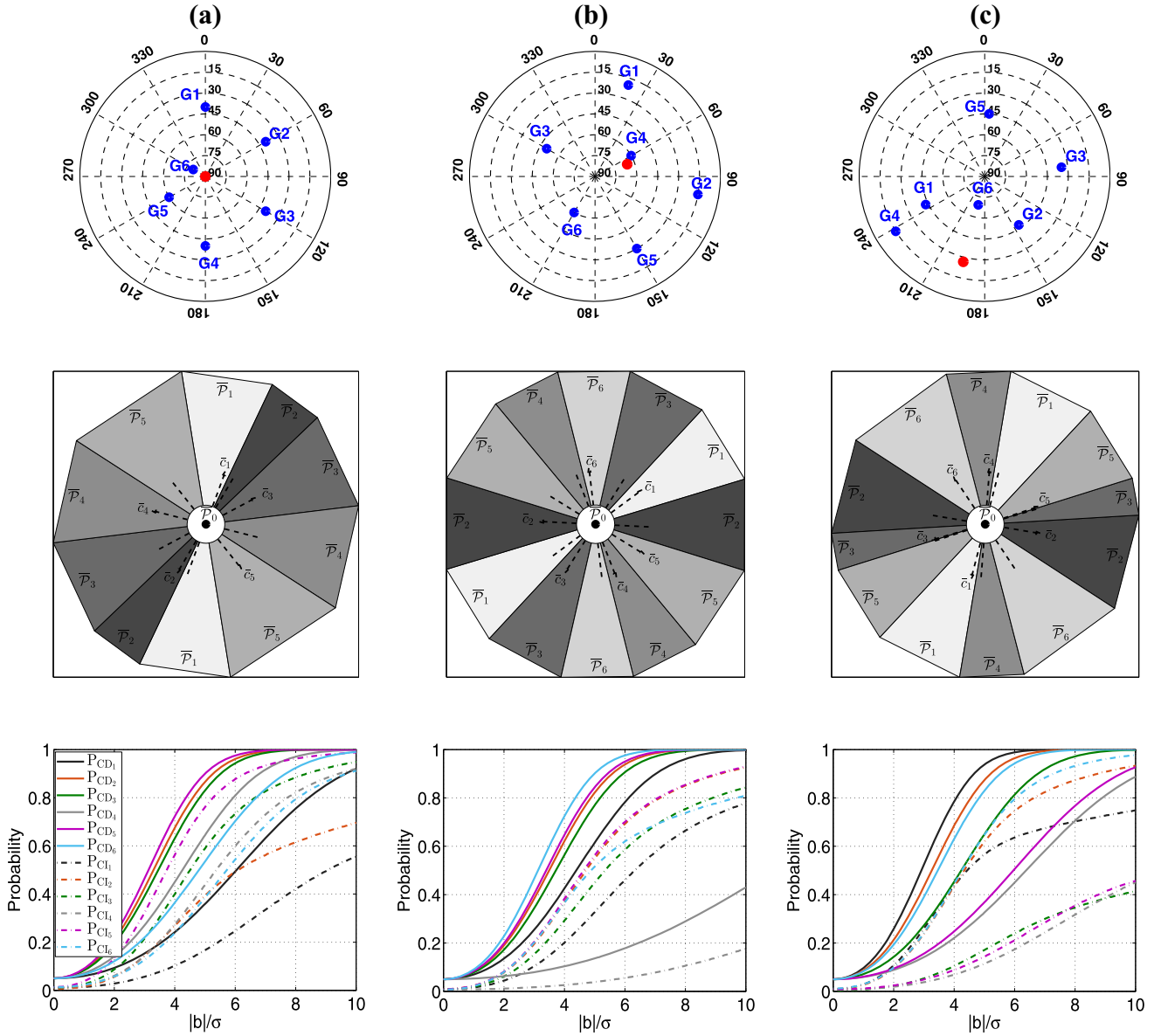

Fig. 5 Visualization of the datasnooping testing procedure defined in Sect. 2.2 for the SPP assuming $\alpha = 0.05$ and $\sigma = 30$ cm. [Top] Skyplot views of the satellite geometries. The six blue circles in each panel denote the skyplot position of the satellites. The red circle denotes the skyplot position of the symmetry axis of the cone formed by the satellites Gi with $i = 1, 2, 3, 4$ in **a**, $i = 1, 2, 3, 5, 6$ in **b** and $i = 1, 2, 4, 6$ in **c**. [Middle] Datasnooping partitioning of the misclosure space $\mathbb{R}^2$ corresponding with $\bar{t}$ (cf. 9). [Bottom] The graphs of CD (solid lines) and CI probabilities (dashed lines) of different alternative hypotheses as function of bias-to-noise ratio

notes that the signature of $\mathrm{P}_{\mathrm{CD}_i}$ is generally different from $\mathrm{P}_{\mathrm{CI}_i}$. For example, in (a), we have $\mathrm{P}_{\mathrm{CD}_2} > \mathrm{P}_{\mathrm{CD}_3}$ while $\mathrm{P}_{\mathrm{CI}_3} > \mathrm{P}_{\mathrm{CI}_2}$. That is because $\mathrm{P}_{\mathrm{CI}_i}$, in addition to $\|c_{t_i}\|_{Q_{tt}}$, is also driven by $\overline{\mathcal{P}}_i$ and the orientation of $\bar{c}_i$ within $\overline{\mathcal{P}}_i$. In (a), we also note that although $\mathcal{H}_5$ and $\mathcal{H}_6$ cannot be distinguished, the testing procedure has a different sensitivity to the $\mathcal{H}_5$- and $\mathcal{H}_6$-biases. For the same bias-to-noise ratios, we have $\mathrm{P}_{\mathrm{CD}_5} > \mathrm{P}_{\mathrm{CD}_6}$ and $\mathrm{P}_{\mathrm{CI}_5} > \mathrm{P}_{\mathrm{CI}_6}$, which can be explained as follows. The difference between $\mathrm{P}_{\mathrm{CD}_5}$ and $\mathrm{P}_{\mathrm{CD}_6}$ for a given bias-to-noise ratio lies in the difference between $\|c_{t_5}\|_{Q_{tt}}$ and $\|c_{t_6}\|_{Q_{tt}}$ (cf. 17). Given that c_{t_i} is the ith column of B^T and given (42), multiplying the corresponding SPP design matrix A with B^T from left and with $[d^T,\ c]^T$ from right, we arrive at

$$c_{t_5} = -\frac{u_6^T d - c}{u_5^T d - c}\, c_{t_6} \tag{44}$$

According to the skyplot in (a), $c = \cos 40°$ and $u_5^T d = \cos 60°$ and $u_6^T d = \cos 80°$, which means that $\|c_{t_5}\|_{Q_{tt}} >$

$\|c_{t_6}\|_{Q_{tt}}$, thus $\mathrm{P}_{\mathrm{CD}_5} > \mathrm{P}_{\mathrm{CD}_6}$. Since $\bar{c}_5 \parallel \bar{c}_6$ and $\overline{\mathcal{P}}_5 = \overline{\mathcal{P}}_6$, the difference between $\mathrm{P}_{\mathrm{CI}_5}$ and $\mathrm{P}_{\mathrm{CI}_6}$ for a given bias-to-noise ratio depends only on the difference between $\|c_{t_5}\|_{Q_{tt}}$ and $\|c_{t_6}\|_{Q_{tt}}$. Therefore, $\|c_{t_5}\|_{Q_{tt}} > \|c_{t_6}\|_{Q_{tt}}$ will also lead to $\mathrm{P}_{\mathrm{CI}_5} > \mathrm{P}_{\mathrm{CI}_6}$.

In (b), all the satellites except G4 locate *almost* on a cone with its axis shown as the *red* circle. If the satellites G1, G2, G3, G5 and G6 would have formed a *perfect* cone, then the contribution of the G4 observation to the misclosures would have been identically *zero*. This can be shown by proving that the fourth column of B^T, i.e., c_{t_4}, becomes a zero-vector. If the unit vectors u_i for $i \neq 4$ lie on a cone with d being its symmetry axis, then for some scalar $c \in \mathbb{R}$ we have $u_i^T\, d = c$ (cf. 42). Multiplying the corresponding SPP design matrix A with B^T from left and with $[d^T,\ c]^T$ from right, we arrive at

$$c_{t_4}\,(u_4^T\, d - c) = 0 \tag{45}$$

Since u_4 does not lie on the mentioned cone, then $u_4^T\, d \neq c$ implying that $c_{t_4} = 0$, thus $\mathrm{P}_{\mathrm{CD}_4} = \mathrm{P}_{\mathrm{CI}_4} = 0$. However, as the line-of-sights to the satellites G1, G2, G3, G5 and G6 do *not* form a perfect cone, i.e., $u_{i\neq 4}^T\, d \approx c$, the observation of satellite G_4 has a nonzero contribution to the misclosure vector resulting in nonzero values for $\mathrm{P}_{\mathrm{CD}_4}$ and $\mathrm{P}_{\mathrm{CI}_4}$. It can be seen that $\mathrm{P}_{\mathrm{CD}_4}$ and $\mathrm{P}_{\mathrm{CI}_4}$ are significantly smaller than, respectively, $\mathrm{P}_{\mathrm{CD}_{i\neq 4}}$ and $\mathrm{P}_{\mathrm{CI}_{i\neq 4}}$. To understand the distinct behavior of $\mathrm{P}_{\mathrm{CD}_4}$ compared to $\mathrm{P}_{\mathrm{CD}_{i\neq 4}}$, we look at $\|c_{t_i}\|_{Q_{tt}}$ given as

$$\|c_{t_i}\|_{Q_{tt}} \;=\; \sigma^{-1}\left[\frac{m}{m-1} + \|u_i - \bar{u}_{\neq i}\|^2_{C_{xx\neq i}}\right]^{-\frac{1}{2}} \tag{46}$$

where $C_{xx\neq i} = \sum_{k\neq i}^{m}(u_k - \bar{u}_{\neq i})(u_k - \bar{u}_{\neq i})^T$ and $\bar{u}_{\neq i} = \frac{1}{m-1}\sum_{k\neq i}^{m} u_k$. The quadratic expression within the brackets can be worked out using the eigenvalue decomposition of $C_{xx\neq i}$ as

$$\|u_i - \bar{u}_{\neq i}\|^2_{C_{xx\neq i}} = \sum_{j=1}^{3} \lambda_{j,i}^{-1}\left[(u_i - \bar{u}_{\neq i})^T\, v_{j,i}\right]^2 \tag{47}$$

in which $\lambda_{j,i}$ and $v_{j,i}$ for $j = 1, 2, 3$ are, respectively, the eigenvalues and the corresponding eigenvectors of $C_{xx\neq i}$. Assuming $\lambda_{1,i} \geq \lambda_{2,i} \geq \lambda_{3,i}$, for a given value of $\|u_i - \bar{u}_{\neq i}\|$, (47) achieves its maximum when $(u_i - \bar{u}_{\neq i}) \parallel v_{3,i}$. In the following, we check $\lambda_{3,i}$ (the minimum eigenvalue), the angle between $(u_i - \bar{u}_{\neq i})$ and $v_{3,i}$ (eigenvector corresponding with the minimum eigenvalue), and $\|u_i - \bar{u}_{\neq i}\|$ for $i = 1, \ldots, 6$.

- $\lambda_{3,i}$: For $i = 4$, since $u_{j\neq 4}^T\, d \approx c$, it can be concluded that $v_{3,4}$ is almost parallel to d and $\lambda_{3,4} \approx 0$. This implies that $\lambda_{3,4}^{-1}$ is extremely large. For $i \neq 4$, among the five remaining satellites, still there are four unit vectors which satisfy $u_{j\neq i,4}^T\, d \approx c$. Therefore, the eigenvector $v_{3,i\neq 4}$ does not deviate too much from the direction d. However, due to the presence of satellite G4 not lying on the mentioned cone, $\lambda_{3,i\neq 4}$ is much larger than zero, implying that $\lambda_{3,i\neq 4}^{-1}$ is much smaller than $\lambda_{3,4}^{-1}$.
- The angle between $(u_i - \bar{u}_{\neq i})$ and $v_{3,i}$: As shown in the skyplot in (b), while u_4 is almost parallel to $v_{3,4}$, $u_{i\neq 4}$ makes an almost 56° with $v_{3,i\neq 4}$ (almost parallel to d). For the geometry shown in (b), $\bar{u}_{\neq 4}$ is almost parallel to $v_{3,4}$, whereas this is not the case with $\bar{u}_{\neq i}$ $(i \neq 4)$. Therefore, we have $(u_4 - \bar{u}_{\neq 4}) \parallel v_{3,4}$.
- $\|u_i - \bar{u}_{\neq i}\|$: We can write $\|u_i - \bar{u}_{\neq i}\|^2 = 1 + \|\bar{u}_{\neq i}\|^2 - 2u_i^T\bar{u}_{\neq i}$. Since $\bar{u}_{\neq i}$ is computed based on five out of six unit direction vectors, its norm does not change too much for different i. Therefore, $\|u_i - \bar{u}_{\neq i}\|$ gets its minimum value for $i = 4$ as u_4 is almost parallel to $\bar{u}_{\neq 4}$. However, $\|u_4 - \bar{u}_{\neq 4}\| < \|u_i - \bar{u}_{\neq i}\|$ is overcompensated by $\lambda_{3,4}^{-1} > \lambda_{3,i}^{-1}$.

Given the above explanation, $\|u_4 - \bar{u}_{\neq 4}\|^2_{C_{xx\neq 4}}$ is much larger than $\|u_i - \bar{u}_{\neq i}\|^2_{C_{xx\neq i}}$, and $\|c_{t_4}\|_{Q_{tt}}$ is thus much smaller compared to $\|c_{t_i}\|_{Q_{tt}}$. This explains that the CD probability of $\mathcal{H}_4$ is much smaller than that of $\mathcal{H}_{i\neq 4}$. As $\overline{\mathcal{P}}_4$ and the orientation of $\bar{c}_4$ within it are similar to those of $\mathcal{H}_i$ with $i = 1, 3, 6$ and poorer than $\mathcal{H}_i$ with $i = 2, 5$, then $\|c_{t_{i\neq 4}}\|_{Q_{tt}} > \|c_{t_4}\|_{Q_{tt}}$ can also explain why $\mathrm{P}_{\mathrm{CI}_{i\neq 4}} > \mathrm{P}_{\mathrm{CI}_4}$.

6 Conclusion and summary

In this contribution, we presented datasnooping in the context of the DIA method, discussed its decision probabilities for detection and identification and showed what options one has available when two or more of the alternative hypotheses are nonseparable.

In our discussion, we emphasized the central role that is played by the partitioning of misclosure space, both in the formation of the decision probabilities and in the construction of the DIA estimator. In case of datasnooping, the partitioning is determined by the row vectors of the basis matrix of the null space of A^T. Through this partitioning, the distribution of the misclosure vector can be used to determine the correct detection (CD) and correct identification (CI) probabilities of each of the alternative hypotheses. These probabilities can be 'inverted' to determine their corresponding minimal biases, the minimal detectable bias (MDB) and the minimal identifiable bias (MIB). We highlighted their difference by showing the difference between their corresponding contributing factors. In particular, it should be realized that the MDB provides information about correct detection and *not* about correct identification. A high probability of correct detection does namely not necessarily imply a high probability of correct

identification, unless one is dealing with the special case of having only one single alternative hypothesis.

In the identification step, one has to ascertain whether or not all the hypotheses are identifiable. Identification of hypotheses becomes problematic if the misclosure vector has the *same* distribution under *different* hypotheses. We discussed the options one can choose from in terms of 'remeasurement', 'adaptation' or stating that the solution is 'unavailable'. Of these, the adaptation step is the most involved. By means of an equivalence between the nonseparability of hypotheses and the inestimability of parameters (cf. Theorem 1), we demonstrated that one can forget about adapting $\hat{x}_0$ for hypotheses that are nonseparable. However, as this concerns the *complete* vector x and not necessarily functions of x, we also demonstrated that functions of x may exist for which adaptation is still possible (cf. Theorem 2). It was shown how this adaptation looks like and how it changes the structure of the DIA estimator.

We applied the theory to selected examples so as to illustrate and explain the performance of the various elements of DIA-datasnooping. Three different cases were discussed in detail: height-difference observations of a leveling network, distance measurements of a horizontal geodetic network and pseudorange measurements between a single ground station and GPS satellites. We analyzed and illustrated how geometry changes in the measurement setup affect the testing procedure, including its partitioning of the misclosure space, and the corresponding CD probabilities (MDB) and CI probabilities (MIB). We also demonstrated that for a given bias-to-noise ratio and a false alarm probability, the ordering of the CD probabilities of the alternative hypotheses is not necessarily the same as that of their CI probabilities. And we showed, if two alternative hypotheses, say $\mathcal{H}_i$ and $\mathcal{H}_j$, are not distinguishable, that the testing procedure may have different levels of sensitivity to $\mathcal{H}_i$-biases compared to the same $\mathcal{H}_j$-biases.

Acknowledgements The second author is the recipient of an Australian Research Council (ARC) Federation Fellowship (Project Number FF0883188). This support is greatly acknowledged.

Appendix

Proof of Lemma 1 That $\cup_{i=1}^{k}\mathcal{P}_i = \mathbb{R}^r/\mathcal{P}_0$ follows from the fact that since any $t \in \mathbb{R}^r/\mathcal{P}_0$ produces a vector of k test statistics w_i $(i = 1, \ldots, k)$, for any such t there is a region $\mathcal{P}_i$ in which it lies for some i. 'if' part: If $c_{t_i} \nparallel c_{t_j}$ for any $i \neq j$, then only one of the test statistics w_i $(i = 1, \ldots, k)$ will produce the maximum absolute value. To prove this, let $|w_i| = |w_j| = \max_{k=\{1,\ldots,m\}} |w_k|$ for $i \neq j$ and for some $t \in \mathbb{R}^r/\mathcal{P}_0$. Then

$$(\|c_{t_i}\|_{Q_{tt}}^{-1} c_{t_i} - \|c_{t_j}\|_{Q_{tt}}^{-1} c_{t_j})^T \, Q_{tt}^{-1} \, t \;=\; 0 \tag{48}$$

from which follows that $t = 0$ cannot be a solution since $t \in \mathbb{R}^r/\mathcal{P}_0$, nor can t be orthogonal to $Q_{tt}^{-1}(\|c_{t_i}\|_{Q_{tt}}^{-1} c_{t_i} - \|c_{t_j}\|_{Q_{tt}}^{-1} c_{t_j})$ since the probability of that occurrence is zero, thus leaving as only possibility that $c_{t_i} \parallel c_{t_j}$, which contradicts our earlier assumption of $c_{t_i} \nparallel c_{t_j}$. Therefore, any $t \in \mathbb{R}^r/\mathcal{P}_0$ lies in only one of the regions $\mathcal{P}_i$ $(i = 1, \ldots, k)$, revealing that $\cup_{i=1}^{k}\mathcal{P}_i = \mathbb{R}^r/\mathcal{P}_0$ and $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset$ for any $i \neq j$. As a result, the $m + 1$ regions $\mathcal{P}_i$ form a partitioning of $\mathbb{R}^r$.

For the 'only if' part we have: Let us assume that $c_{t_i} \parallel c_{t_j}$ for $i \neq j$. Then, given (7), we have $|w_i| = |w_j|$ for any $t \in \mathbb{R}^r$. This reveals that $\mathcal{P}_i = \mathcal{P}_j$ which contradicts our earlier assumption of $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset$. Therefore, $c_{t_i} \nparallel c_{t_j}$ for any $i \neq j$. □

Proof of Lemma 2 Consider the 2-vector $\bar{t}$ with a length of $\bar{l}$ which makes an angle of $\bar{\beta}_i$ with $\bar{c}_i$ measured counterclockwise. Therefore, we have

$$\bar{t} = \bar{l} \begin{bmatrix} \cos\bar{\beta}_i & -\sin\bar{\beta}_i \\ \sin\bar{\beta}_i & \cos\bar{\beta}_i \end{bmatrix} \bar{c}_i \tag{49}$$

Given the above equation and the orientation of $\bar{c}_i$ w.r.t. the straight borders of $\overline{\mathcal{P}}_i$ in (11), one can write

$$\overline{\mathcal{P}}_{i\neq 0} = \left\{ \bar{t} \in \mathbb{R}^2 | \, \bar{t} = \bar{l} \begin{bmatrix} \cos\bar{\beta}_i & -\sin\bar{\beta}_i \\ \sin\bar{\beta}_i & \cos\bar{\beta}_i \end{bmatrix} \bar{c}_i, \; \bar{l} > \sqrt{k_\alpha}, \; \bar{\beta}_i \in \mathcal{L}_i \right\} \tag{50}$$

with

$$\mathcal{L}_i \;=\; \{\bar{\beta}_i \in \mathbb{R} | \, \bar{\beta}_i \in ([-\beta_{i,1}, \beta_i - \beta_{i,1}] \cup [\pi - \beta_{i,1}, \pi + \beta_i - \beta_{i,1}])\} \tag{51}$$

With (49), one can also obtain the joint PDF of $[\bar{l}, \bar{\beta}_i]^T$ from $f_{\bar{l}}(\tau|\mathcal{H}_i)$ through the PDF transformation rule as

$$\begin{aligned} f_{\bar{l},\bar{\beta}_i}(l, \beta|\mathcal{H}_i) &= l \; f_{\bar{l}}(\tau(l, \beta)|\mathcal{H}_i) \\ &= \frac{l}{2\pi} \exp\left\{ -\frac{1}{2} \left(l^2 + \|\mu_{\bar{t}_i}\|^2 - 2l\|\mu_{\bar{t}_i}\| \cos\beta \right) \right\} \end{aligned} \tag{52}$$

with $\|\mu_{\bar{t}_i}\| = |b_i| \, \|c_{t_i}\|_{Q_{tt}}$. Equations (50) and (52) enable us to express the CI probability in terms of $\bar{l}$ and $\bar{\beta}_i$ as

$$\begin{aligned} \mathrm{P}_{\mathrm{CI}_i} &= \int_{\overline{\mathcal{P}}_i} f_{\bar{l}}(\tau|\mathcal{H}_i) \, \mathrm{d}\tau \\ &= \int_{\mathcal{L}_i} \int_{\sqrt{k_\alpha}}^{\infty} f_{\bar{l},\bar{\beta}_i}(l, \beta) \, \mathrm{d}l\mathrm{d}\beta \end{aligned} \tag{53}$$

Substitution of (52) into (53), and then taking the derivative w.r.t. $\beta_{i,1}$, we achieve

$$\int_{\sqrt{k_\alpha}}^{\infty} \frac{l}{2\pi} \exp\left\{-\frac{1}{2}\left(l^2+\|\mu_{\bar{t}}\|^2\right)\right\} \times \left[\exp\left\{l\|\mu_{\bar{t}}\|\cos\beta_{i,1}\right\} - \exp\left\{l\|\mu_{\bar{t}}\|\cos(\beta_i-\beta_{i,1})\right\} + \exp\left\{-l\|\mu_{\bar{t}}\|\cos\beta_{i,1}\right\} - \exp\left\{-l\|\mu_{\bar{t}}\|\cos(\beta_i-\beta_{i,1})\right\}\right] \mathrm{d}l \tag{54}$$

Setting the above derivative equal to zero, a set of solutions for $\beta_{i,1}$ is given by

$$\beta_{i,1} = k\pi + \frac{1}{2}\beta_i, \qquad k \in \mathbb{Z} \tag{55}$$

Since $\beta_{i,1} < \pi$, then the only valid solution is $\beta_{i,1} = \frac{1}{2}\beta_i$. To check whether this critical point is the maximizer of (53), we compute the derivative of the expression in (54) at $\beta_{i,1} = \frac{1}{2}\beta_i$, which is

$$\int_{\sqrt{k_\alpha}}^{\infty} \frac{l^2\|\mu_{\bar{t}}\|}{2\pi} \exp\left\{-\frac{1}{2}\left(l^2+\|\mu_{\bar{t}}\|^2\right)\right\} \times \left[-2\sin\frac{\beta_i}{2}\exp\left\{l\|\mu_{\bar{t}}\|\cos\frac{\beta_i}{2}\right\} + 2\sin\frac{\beta_i}{2}\exp\left\{-l\|\mu_{\bar{t}}\|\cos\frac{\beta_i}{2}\right\}\right] \mathrm{d}l \tag{56}$$

Since $0 < \beta_i < \pi$, then $\sin\frac{\beta_i}{2} > 0$ and $\cos\frac{\beta_i}{2} > 0$. These, in tandem with $l\|\mu_{\bar{t}}\| > 0$ and the fact that $\exp\{\cdot\}$ is a 'positive' increasing function, imply that the expression in (56) is 'negative.' Thus, $\beta_{i,1} = \frac{1}{2}\beta_i$ is the maximizer of (53). □

Proof of Theorem 1 We start with the 'if' part: If there exists a nonzero matrix $X \in \mathbb{R}^{n\times q}$ such that $A\,X = C_i - C_j\,X_{i,j}$ for some invertible matrix $X_{i,j} \in \mathbb{R}^{q\times q}$, then multiplying both sides of the equation with B^T from left gives

$$B^T C_i \;=\; B^T C_j\, X_{i,j} \tag{57}$$

For the 'only if' part we have: If $B^T C_i = B^T C_j\, X_{i,j}$ for some invertible matrix $X_{i,j} \in \mathbb{R}^{q\times q}$, then two conclusions can be made: 1. $C_i = C_j\, X_{i,j}$ which is not possible as it contradicts our assumption that rank $([C_i\ C_j]) > q$; 2. $(C_i - C_j\, X_{i,j}) \in \mathcal{R}(A)$. Therefore, there exists a nonzero matrix $X \in \mathbb{R}^{n\times q}$ such that $A\,X = C_i - C_j\,X_{i,j}$ or equivalently

$$[A\ \ C_i\ \ C_j]\begin{bmatrix} X \\ -I_q \\ X_{i,j}\end{bmatrix} \;=\; 0 \tag{58}$$

□

Proof of Theorem 2 Since θ is a linear function of x and not b, we reduce the observational model in (24) to one containing only the unknown parameters x as

$$E(C^{\perp^T} y) \;=\; C^{\perp^T} A\, x; \quad D(y) \;=\; C^{\perp^T} Q_{yy} C^{\perp} \tag{59}$$

where $C^{\perp}$ is a basis matrix of the orthogonal complement of the range space of C. Let V be a basis matrix of the null space of $C^{\perp^T} A$. Furthermore, let S be a full-rank matrix of which the range space is complementary to that of V. Therefore, $[S\ \ V] \in \mathbb{R}^{n\times n}$ is an invertible matrix, thus a basis matrix of $\mathbb{R}^n$. This indicates that any vector $x \in \mathbb{R}^n$ can be parametrized as

$$x \;=\; S\,x_S \;+\; V\,x_V \tag{60}$$

(i) We now show that θ is unbiased estimable under (59) iff $F^T V = 0$. 'if' part: If $F^T V = 0$, then $\theta = F^T S\, x_S$. Substituting (60) into (59), as $C^{\perp^T} A\, V = 0$, gives us

$$E(C^{\perp^T} y) \;=\; C^{\perp^T} A\, S\, x_S; \quad D(y) \;=\; C^{\perp^T} Q_{yy} C^{\perp} \tag{61}$$

which shows that x_S as well as any linear function of it like θ are unbiased estimable under (59). 'only if' part: Equation (61) shows that x_V is *not* estimable. Therefore, for $\theta = F^T x$ to be estimable in (59), it should be a function of only x_S and not x_V, i.e., $F^T V = 0$ and $\theta = F^T S\, x_S$.

(ii) The BLUE of x_S based on (61) is as follows

$$\hat{x}_S \;=\; (S^T \bar{A}^T Q_{yy}^{-1} \bar{A} S)^{-1} S^T \bar{A}^T Q_{yy}^{-1} y \tag{62}$$

in which $\bar{A} = P_C^{\perp} A$ and $P_C^{\perp} = I_m - C(C^T Q_{yy}^{-1} C)^{-1} C^T Q_{yy}^{-1}$. It can be shown (Teunissen and Khodabandeh 2013) that the BLUE of x_S based on (1), $\hat{x}_{S_0}$, is linked to $\hat{x}_S$ as

$$\hat{x}_{S_0} \;=\; \hat{x}_S - Q_{\hat{x}_S,t} Q_{tt}^{-1} t \tag{63}$$

where t and Q_{tt} are given in (4) and $Q_{\hat{x}_S,t}$ is the covariance between $\hat{x}_S$ and t which can be computed given (4) and (62). Multiplying both sides of the above equation with $F^T S$ from left gives the link between $\hat{\theta} = F^T S\,\hat{x}_S$ and $\hat{\theta}_0 = F^T S\,\hat{x}_{S_0}$. The link between their variances is also obtained using the error propagation law. □

Numerical evaluation of P_{CI_i} in (18): In this study, given Eqs. (5)–(8), the CI probability under $\mathcal{H}_i$ (cf. 18) is computed through the following steps.

- Generate n samples of t from the normal distribution $\mathcal{N}(\mu_{t_i} = b_i c_{t_i}, Q_{tt})$.
- Compute $\|t\|^2_{Q_{tt}}$ for all the samples. Single out those samples satisfying $\|t\|^2_{Q_{tt}} > k_\alpha$ (i.e., $t \notin \mathcal{P}_0$) and collect them in a set denoted by Ω_{CD}.

– For each sample in Ω_{CD}, compute the w-tests w_j for $j = 1, \ldots, k$ in (7). Count the number of samples for which $|w_i| \geq |w_j|$ for any $j \neq i$ (i.e., $t \in \mathcal{P}_i$), and denote it by n_{CI}.
– Compute the CI probability under $\mathcal{H}_i$ as $P_{\mathrm{CI}_i} = \frac{n_{\mathrm{CI}}}{n}$. □

References

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

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

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 Prof W Baarda, By staff of the delft geodetic computing centre (DGCC), vol 2, pp 202–274

Hawkins DM (1980) Identification of outliers, vol 11. Chapman and Hall, London

Imparato D, Teunissen PJG, Tiberius CCJM (2018) Minimal detectable and identifiable biases for quality control. Surv Rev. https://doi.org/10.1080/00396265.2018.1437947

Khodabandeh A, Teunissen PJG (2016) Single-epoch GNSS array integrity: an analytical study. In: Sneeuw N, Novák P, Crespi M, Sansò F (eds) VIII Hotine-Marussi symposium on mathematical geodesy: proceedings of the symposium in Rome, 17–21 June, 2013, Springer International Publishing, pp 263–272

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

Kok J (1984) Statistical analysis of deformation problems using Baarda's testing procedure. In: "Forty years of thought" anniversary volume on the occasion of prof Baarda's 65th birthday, Delft, vol 2, pp 470–488

Parkinson BW, 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 (1995) Real-time adaptation for model errors in dynamic systems. Bull Geodesique 69(2):81–91

Sturza M (1988) Navigation system integrity monitoring using redundant measurements. Navigation 35(4):483–501

Su X, Zhan X, Niu M, Zhang Y (2014) Receiver autonomous integrity monitoring availability and fault detection capability comparison between BeiDou and GPS. J Shanghai Jiaotong Univ (Sci) 19:313–324

Teunissen PJG (1985) Generalized inverses, adjustment, the datum problem, and S-transformation. In: Grafarend E, Sanso F (eds) Optim Des Geod Netw. Springer, Berlin

Teunissen PJG (1990) An integrity and quality control procedure for use in multi-sensor integration. In: Proceedings of ION GPS-1990, ION, pp 513–522

Teunissen PJG (2000) Testing theory: an introduction. Series on mathematical geodesy and positioning. Delft University Press, Delft

Teunissen PJG (2017) Distributional theory for the DIA method. J Geodesy 92:59–80. https://doi.org/10.1007/s00190-017-1045-7

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(5):461–473

Teunissen PJG, Imparato D, Tiberius CCJM (2017) Does RAIM with correct exclusion produce unbiased positions? Sensors 17(7):1508

Tiberius C (1998) Recursive data processing for kinematic GPS surveying. Netherlands Geodetic Commission, Publ. on Geodesy, New series (45)

Van der Marel H, Kosters A (1990) Statistical testing and quality analysis in 3-d networks: (part ii) application to gps. global positioning system: an overview. In: International Association of Geodesy Symposia, Springer, vol 102, pp 290–297

Van Mierlo J (1980) A testing procedure for analytic deformation measurements. In: Proceedings of internationales symposium ueber Deformationsmessungen mit Geodaetischen Methoden, Verlag Konrad Wittwer, Stuttgart

Zaminpardaz S, Teunissen PJG, Nadarajah N, Khodabandeh A (2015) GNSS array-based ionospheric spatial gradient monitoring: precision and integrity analysis. In: Proceedings of the ION pacific PNT meeting, ION, USA, pp 799–814