MDBs Versus MIBs in Case of Multiple Hypotheses

A Study in Context of Deformation Analysis

Zaminpardaz, Safoora; Teunissen, Peter J.G.

DOI
10.1007/1345_2023_208

Publication date
2024

Document Version
Final published version

Published in
X Hotine-Marussi Symposium on Mathematical Geodesy

MDBs Versus MIBs in Case of Multiple Hypotheses: A Study in Context of Deformation Analysis

Safoora Zaminpardaz and Peter J. G. Teunissen

Abstract

Statistical testing procedures employed in geodetic quality control often consist of two steps: detection and identification. In the detection step, the null hypothesis (working model) $\mathcal{H}_0$ undergoes a validity check. If the outcome of the detection step is the rejection of $\mathcal{H}_0$, identification of potential source of model error is exercised through a search among the specified alternative hypotheses. The testing performance is thus not only led by its ability to detect biases but to correctly identify them as well. The detection capability of a testing regime is usually assessed by its Minimal Detectable Bias (MDB) given a certain correct detection probability. The information provided by the MDB only concerns correct detection and *not* correct identification. The testing identification performance should be evaluated by its Minimal Identifiable Bias (MIB) given a certain correct identification probability. In this contribution, we demonstrate the difference between MDB and MIB. It is hereby highlighted that a small MDB (or a high probability of correct detection) does not necessarily imply a small MIB (or a high probability of correct identification). The factors driving the difference between detection and identification performance are illustrated using a simple example. Our analysis is then continued in the framework of deformation monitoring.

Keywords

Deformation monitoring · Detection, identification and adaptation (DIA) · Minimal detectable bias (MDB) · Minimal identifiable bias (MIB) · Statistical testing

1 Introduction

In geodetic quality control, statistical testing procedures often consist of two steps: *detection* and *identification* (Baarda 1968; Teunissen 1985; Caspary and Borutta 1987; Kösters and Van der Marel 1990; Amiri Simkooei 2001; Perfetti 2006; Lehmann and Lösler 2017; Klein et al. 2019; Nowel 2020). In the detection step, the validity of the null hypothesis $\mathcal{H}_0$ is checked. If $\mathcal{H}_0$ is rejected in the detection step, an identification is carried out as to which of the alternative hypotheses to select. In case there is only one alternative hypothesis, say $\mathcal{H}_1$, the rejection of $\mathcal{H}_0$ is equivalent to the selection of $\mathcal{H}_1$. Thus, 'correct detection' of mismodelling error would be equivalent to

Authors Safoora Zaminpardaz and Peter J.G. Teunissen contributed equally to this work.

S. Zaminpardaz (✉)
School of Science, RMIT University, Melbourne, VIC, Australia
e-mail: safoora.zaminpardaz@rmit.edu.au

P. J. G. Teunissen
Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands

GNSS Research Centre, School of Earth and Planetary Sciences, Curtin University of Technology, Perth, WA, Australia

Department of Infrastructure Engineering, The University of Melbourne, Melbourne, VIC, Australia
e-mail: p.j.g.teunissen@tudelft.nl

J. T. Freymueller, L. Sánchez (eds.), *X Hotine-Marussi Symposium on Mathematical Geodesy*,
International Association of Geodesy Symposia 155, https://doi.org/10.1007/1345_2023_208

'correct identification' of it when working with a single alternative hypothesis. This is however not the case if one has to deal with multiple alternative hypotheses. In this contribution, for multiple-alternative testing, we study the performance of the detection and identification steps using the concepts of the minimal detectable bias (MDB) and the minimal identifiable bias (MIB), respectively, and highlight the factors driving the difference between them.

This contribution is structured as follows. In Sect. 2, we describe the null and alternative hypotheses, and highlight the role of the misclosure space partitioning in testing these hypotheses. The testing decisions and their probabilities are discussed, whereby the following events are defined: correct acceptance (CA), false alarm (FA), correct detection (CD), missed detection (MD), correct identification (CI) and wrong identification (WI). The concepts of MDB and MIB are discussed in Sect. 3 for a testing procedure comprising detection and identification steps. It is hereby highlighted that the MDB provides information about correct detection and not about correct identification. To provide insight into the difference between the MDB and the MIB, we compare them in Sect. 4, for a simple multiple-hypothesis testing example. It is demonstrated, in graphical form, that the MIB could be significantly larger than the MDB. The MDB-MIB comparison is then continued for actual deformation measurement system examples in Sect. 5. Finally a summary with conclusions is presented in Sect. 6.

We use the following notation: The n-dimensional space of real numbers is denoted as $\mathbb{R}^n$, and the set of points on the circumference of the n-dimensional zero-centered unit sphere as $\mathbb{S}^n$. Random vectors are indicated by use of the underlined symbol '$\underline{\cdot}$'. Thus $\underline{t} \in \mathbb{R}^n$ is a random vector, while t is not. The squared weighted norm of a vector, with respect to a positive-definite matrix Q, is defined as $\|\cdot\|^2_Q = (\cdot)^T Q^{-1}(\cdot)$. $\mathcal{H}$ is reserved for statistical hypotheses, $\mathcal{P}$ for regions partitioning the misclosure space, and $\mathcal{N}(x, Q)$ for the normal distribution with mean x and variance matrix Q. $\mathsf{P}(\cdot)$ denotes the probability of the occurrence of the event within parentheses. The symbol $\overset{\mathcal{H}}{\sim}$ should be read as 'distributed as ... under $\mathcal{H}$'. The superscripts ^{T} and $^{-1}$ are used to denote the transpose and the inverse of a matrix.

2 Statistical Hypothesis Testing

In any quality control procedure, a set of hypotheses, including a null and several alternative hypotheses, are postulated to explain the phenomenon in question. For example, in geodetic deformation monitoring, the null hypothesis describes the 'all-stable, no movement' model, while the alternative hypotheses capture different dynamic behaviors of the structure under consideration. Let the observational model under the null hypothesis $\mathcal{H}_0$, a.k.a. working hypothesis, be given as

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

with $\mathsf{E}(\cdot)$ the expectation operator, $\mathsf{D}(\cdot)$ the dispersion operator, $\underline{y} \in \mathbb{R}^m$ 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 $\text{rank}(A) = n$, and $Q_{yy} \in \mathbb{R}^{m\times m}$ the positive-definite variance matrix of $\underline{y}$. 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 are misspecified. Here we assume that a misspecification is restricted to an underparametrization of the mean of $\underline{y}$, which is the most common error that occurs when formulating the model (Teunissen 2017). Thus, the alternative hypothesis $\mathcal{H}_i$ is formulated as

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

for some vector $C_i b_i \in \mathbb{R}^m \setminus \{0\}$ such that $[A \;\; C_i]$ is a known matrix of full rank and b_i is an unknown vector.

2.1 Misclosure Space Partitioning

Let us assume that there are k types of mismodelling errors in the form of $C_i b_i$ (cf. 2) when parametrizing the mean of observations. The information required to validate the hypotheses at hand is contained in the *misclosure* vector $\underline{t} \in \mathbb{R}^r$ given as (Teunissen 2006)

$$\underline{t} = B^T \underline{y} \tag{3}$$

where $B \in \mathbb{R}^{m\times r}$ is a full-rank matrix, with $\text{rank}(B) = r$, such that $[A \;\; B] \in \mathbb{R}^{m\times m}$ is invertible and $A^T B = 0$. With $C_0 b_0 = 0$ and given that $\underline{y} \overset{\mathcal{H}_i}{\sim} \mathcal{N}(Ax + C_i b_i, Q_{yy})$ for $i = 0, 1, \ldots, k$, the misclosure vector is then distributed as

$$\underline{t} \overset{\mathcal{H}_i}{\sim} \mathcal{N}(C_{t_i} b_i, Q_{tt} = B^T Q_{yy} B), \quad \text{for} \quad i = 0, 1, \ldots k \tag{4}$$

with $C_{t_i} = B^T C_i$. As $\underline{t}$ has a known Probability Density Function (PDF) under $\mathcal{H}_0$, which is the PDF of $\mathcal{N}(0, Q_{tt})$, any statistical testing procedure is driven by the misclosure vector $\underline{t}$ and its known PDF under $\mathcal{H}_0$.

An unambiguous testing procedure can be established through assigning the outcomes of $\underline{t}$ to the statistical hypotheses $\mathcal{H}_i$ for $i = 0, 1, \ldots, k$, which can be realized through a *partitioning* of the misclosure space $\mathbb{R}^r$ (Teunissen 2018). Let $\mathcal{P}_i \subset \mathbb{R}^r$ $(i = 0, 1, \ldots, k)$ be a partitioning of the misclosure space, i.e. $\cup_{i=0}^k \mathcal{P}_i = \mathbb{R}^r$ and $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset$ for

$i \neq j$. The unambiguous testing procedure is then defined as

$$\text{select } \mathcal{H}_i \text{ if and only if } \underline{t} \in \mathcal{P}_i \quad \text{for } i = 0, 1 \ldots, k \tag{5}$$

We note, although in (5) the statistical testing is formulated in the misclosure vector $\underline{t}$, that one can equally well work with the least-squares residual vector $\hat{\underline{e}}_0 = \underline{y} - A\hat{\underline{x}}_0$ where $\hat{\underline{x}}_0 = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y}$. By using the relation $\underline{t} = B^T \hat{\underline{e}}_0$, there is no explicit need of having to compute $\underline{t}$ as testing can be expressed directly in $\hat{\underline{e}}_0$ (Teunissen 2006).

2.2 Testing Decisions

As (5) shows, the testing decisions are driven by the outcome of the misclosure vector $\underline{t}$. Under each hypothesis $\mathcal{H}_i$ ($i = 0, 1, \ldots, k$), the outcome of $\underline{t}$ can lead to $k + 1$ different decisions out of which only one is correct, i.e. when $\underline{t} \in \mathcal{P}_i$. With $k + 1$ hypotheses $\mathcal{H}_i$'s ($i = 0, 1, \ldots, k$), one can define different statistical events including Correct Acceptance (CA), False Alarm (FA), Missed Detection (MD), Correct Detection (CD), Correct Identification (CI) and Wrong Identification (WI). The definitions of these events together with their links are illustrated in Fig. 1. In this figure, the events under alternative hypotheses are given an identifying index, as they differ from alternative to alternative. In addition, the contributions of different alternative hypotheses to the events of false alarm and wrong identification are distinguished by means of an index.

Given the translational property of the PDF of $\underline{t}$ under the null and alternative hypotheses (cf. 4), the probabilities of the events in Fig. 1 can be computed based on the misclosure PDF under $\mathcal{H}_0$, denoted by $f_{\underline{t}}(\tau|\mathcal{H}_0)$, as

$$\begin{aligned}
\mathsf{P}_{\text{FA}} &= \mathsf{P}(\underline{t} \notin \mathcal{P}_0|\mathcal{H}_0) = \textstyle\int_{\mathbb{R}^r \setminus \mathcal{P}_0} f_{\underline{t}}(\tau|\mathcal{H}_0)\, d\tau \\
\mathsf{P}_{\text{CA}} &= 1 - \mathsf{P}_{\text{FA}} \\
\mathsf{P}_{\text{CD}_i} &= \mathsf{P}(\underline{t} \notin \mathcal{P}_0|\mathcal{H}_i) = \textstyle\int_{\mathbb{R}^r \setminus \mathcal{P}_0} f_{\underline{t}}(\tau - C_{t_i} b_i|\mathcal{H}_0)\, d\tau \\
\mathsf{P}_{\text{MD}_i} &= 1 - \mathsf{P}_{\text{CD}_i} \\
\mathsf{P}_{\text{CI}_i} &= \mathsf{P}(\underline{t} \in \mathcal{P}_i|\mathcal{H}_i) = \textstyle\int_{\mathcal{P}_i} f_{\underline{t}}(\tau - C_{t_i} b_i|\mathcal{H}_0)\, d\tau \\
\mathsf{P}_{\text{WI}_i} &= \mathsf{P}_{\text{CD}_i} - \mathsf{P}_{\text{CI}_i}
\end{aligned} \tag{6}$$

The probability of false alarm P_{FA} is usually set a priori by the user. We note that the last four probabilities all depend on the *unknown* b_i which one needs to set to evaluate the mentioned four probabilities.

Here, it is important to note the difference between the probabilities of correct detection and correct identification, i.e. $\mathsf{P}_{\text{CD}_i} \geq \mathsf{P}_{\text{CI}_i}$. These two probabilities would be identical if there is only one alternative hypothesis, say $\mathcal{H}_i$, since then $\mathcal{P}_i = \mathbb{R}^r \setminus \mathcal{P}_0$. Similar to the CD- and CI-probability, we have the concepts of the minimal detectable bias (MDB) (Baarda 1968) and the minimal identifiable bias (MIB) (Teunissen

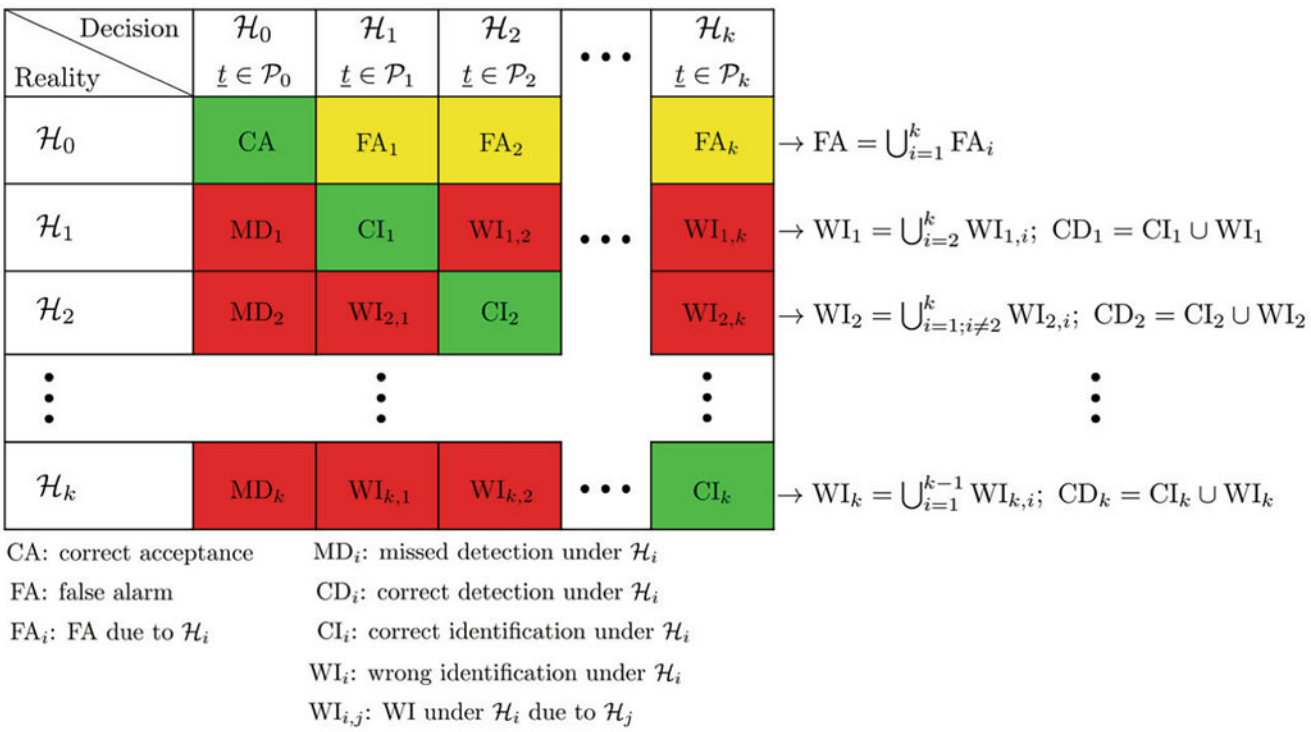

Fig. 1 An overview of testing decisions, driven by the misclosure vector $\underline{t}$, under null and alternative hypotheses

2018). In the following sections, we highlight the difference between the MDB ($\mathsf{P}_{\mathrm{CD}_i}$) and the MIB ($\mathsf{P}_{\mathrm{CI}_i}$).

3 Testing Performance

Statistical testing procedures employed in quality control often comprises two steps (Baarda 1968; Teunissen 1985; Caspary and Borutta 1987; Kösters and Van der Marel 1990; Amiri Simkooei 2001; Perfetti 2006; Lehmann and Lösler 2017; Nowel 2020), as follows

- *Detection*: The null hypothesis $\mathcal{H}_0$ undergoes a validity check, without considering a particular set of alternatives.
- *Identification*: If $\mathcal{H}_0$ is rejected in the detection step, i.e. $\underline{t} \notin \mathcal{P}_0$, a search is carried out among the specified alternatives $\mathcal{H}_i$ $(i = 1, \ldots, k)$ to pinpoint the potential source of model error.

The testing performance is thus not only led by its ability to detect biases but to correctly identify them as well. While the former is measured by means of the MDB (or alternatively CD-probability), the latter should be measured using the MIB (or alternatively CI-probability) (Teunissen 2018; Zaminpardaz and Teunissen 2019; Imparato et al. 2019). Note, in single-redundancy case $r = 1$, that $\mathcal{P}_1 = \ldots = \mathcal{P}_k = \mathbb{R}^r \setminus \mathcal{P}_0$, implying that the alternative hypotheses are not distinguishable from one another, and thus identification would not be possible.

3.1 Minimal Detectable Bias (MDB)

The concept of the MDB was introduced in Baarda (1967, 1968) as a diagnostic tool for measuring the ability of the testing procedure to *detect* misspecifications of the model. The MDB, for each alternative hypothesis $\mathcal{H}_i$, is defined as the smallest size of b_i that can be detected given a certain CD- and FA-probability. As the third equality in (6) shows, $\mathsf{P}_{\mathrm{CD}_i}$ depends, in addition to the PDF of $\underline{t}$ under $\mathcal{H}_0$ and b_i, also on $\mathcal{P}_0$ which is commonly defined as (Baarda 1968; Teunissen 2006)

$$\mathcal{P}_0 = \left\{ t \in \mathbb{R}^r \,|\, \|t\|^2_{Q_{tt}} \leq \chi^2_{1-\mathsf{P}_{\mathrm{FA}}}(r, 0) \right\} \tag{7}$$

where $\chi^2_{1-\mathsf{P}_{\mathrm{FA}}}(r, 0)$ is the $(1 - \mathsf{P}_{\mathrm{FA}})$ quantile of the central Chi-square distribution with r degrees of freedom. Using (7), one in fact compares the test statistic $\|\underline{t}\|^2_{Q_{tt}}$ against the critical value $\chi^2_{1-\mathsf{P}_{\mathrm{FA}}}(r, 0)$, with user-defined P_{FA}, to decide whether $\mathcal{H}_0$ is valid or not. This testing process is called the *overall model test*, which would be a Uniformly Most Powerful Invariant (UMPI) detector test in case of dealing with a single alternative hypothesis (Arnold 1981; Teunissen 2006; Lehmann and Voß-Böhme 2017).

With (7), the CD-probability of $\mathcal{H}_i$ is given by

$$\mathsf{P}_{\mathrm{CD}_i} = \mathsf{P}\left(\|\underline{t}\|^2_{Q_{tt}} > \chi^2_{1-\mathsf{P}_{\mathrm{FA}}}(r, 0) | \mathcal{H}_i \right) \tag{8}$$

where, according to (4), $\|\underline{t}\|^2_{Q_{tt}}$ under $\mathcal{H}_i$ has a non-central Chi-square distribution with r degrees of freedom and the non-centrality parameter $\lambda_i^2 = \|C_{t_i} b_i\|^2_{Q_{tt}}$. One can compute $\lambda_i^2 = \lambda^2(\mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{CD}_i}, r)$ from the Chi-square distribution for a given model redundancy r, CD-probability $\mathsf{P}_{\mathrm{CD}_i}$ and FA-probability P_{FA}. If $b_i \in \mathbb{R}$ is a scalar, then C_{t_i} takes the form of a vector c_{t_i}, and the MDB is given by (Baarda 1968; Teunissen 2006)

$$b_i \in \mathbb{R} : \; |b_{i,\mathrm{MDB}}| = \frac{\lambda(\mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{CD}_i}, r)}{\|c_{t_i}\|_{Q_{tt}}} \tag{9}$$

which shows that for a given set of $\{\mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{CD}_i}, r\}$, the MDB depends on $\|c_{t_i}\|_{Q_{tt}}$. For the higher-dimensional case when $b_i \in \mathbb{R}^{q>1}$ is a vector instead of a scalar, a similar expression can be obtained. Let the bias vector be parametrized, in terms of its magnitude $\|b_i\|$ and its unit direction vector d, as $b_i = \|b_i\| \, d$. Then the MDB along the direction $d \in \mathbb{S}^{q-1}$ is given by (Teunissen 2006)

$$b_i \in \mathbb{R}^{q>1} : \; \|b_{i,\mathrm{MDB}}(d)\| = \frac{\lambda(\mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{CD}_i}, r)}{\|C_{t_i} d\|_{Q_{tt}}}; \; d \in \mathbb{S}^{q-1} \tag{10}$$

If the unit vector d sweeps the surface of the unit sphere $\mathbb{S}^{q-1}$, an ellipsoidal region is obtained of which the boundary defines the MDBs in different directions. The shape and the orientation of this ellipsoidal region is governed by the variance matrix $Q_{\hat{b}_i \hat{b}_i} = (C_{t_i}^T Q_{tt}^{-1} C_{t_i})^{-1}$, and its size is determined by $\lambda(\mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{CD}_i}, r)$ (Zaminpardaz et al. 2015; Zaminpardaz 2016).

The MDB concept expresses the sensitivity of the *detection* step of the testing procedure. One can compare the MDBs of different alternative hypotheses for a given set of $\{\mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{CD}}, r\}$, which provides information on how sensitive is the rejection of $\mathcal{H}_0$ for the $\mathcal{H}_i$-biases the size of their MDBs. The smaller the MDB is, the more sensitive is the rejection of $\mathcal{H}_0$.

3.2 Minimal Identifiable Bias (MIB)

As the last equality in (6) shows, a high CD-probability $\mathsf{P}_{\mathrm{CD}_i}$ does not necessarily imply a high CI-probability $\mathsf{P}_{\mathrm{CI}_i}$ unless we have the special case of only a single alternative hypothesis. Therefore, in case of multiple hypotheses, the MDB

does *not* provide information about correct identification. To assess the sensitivity of the identification step, one can analyse the MIBs of the alternative hypotheses. The MIB of the alternative hypothesis $\mathcal{H}_i$ is defined as the smallest size of b_i that can be identified given a certain CI probability (Teunissen 2018).

The MIB corresponding with $\mathcal{H}_i$ can be found from inverting the fifth equality in (6). This inversion is, however, not trivial as $\mathsf{P}_{\mathrm{CI}_i}$ is an r-fold integral over the complex region $\mathcal{P}_i$. One can take resort to numerical evaluation techniques. For example, the MIBs in Sect. 4 are numerically computed as follows. The probability $\mathsf{P}_{\mathrm{CI}_i}$ is computed, by means of Monte Carlo simulation, see e.g. Teunissen (2018), at discrete biases b_i and then the bias at which $\mathsf{P}_{\mathrm{CI}_i}$ gets close enough to the pre-set CI-probability is the MIB sought.

According to the fifth equality in (6), the MIB for a given $\mathsf{P}_{\mathrm{CI}_i}$ depends on the probability mass of the PDF of $\underline{t}$ under $\mathcal{H}_i$ over $\mathcal{P}_i$. This probability mass is driven by the shape and size of $\mathcal{P}_i$, magnitude of $\mathsf{E}(\underline{t}|\mathcal{H}_i)$ and its direction with respect to the borders of $\mathcal{P}_i$. Note, if $b_i \in \mathbb{R}^{q>1}$ is a vector, then, a given CI-probability yields different MIBs along different directions in $\mathbb{R}^q$. In this case, a pre-set CI-probability defines a region in $\mathbb{R}^q$ the boundary of which defines the MIBs in different directions. The MIB of $\mathcal{H}_i$ for a given CI-probability is denoted by $|b_{i,\mathrm{MIB}}|$ if $b_i \in \mathbb{R}$, and $\|b_{i,\mathrm{MIB}}(d)\|$ along the unit direction $d \in \mathbb{S}^{q-1}$ if $b_i \in \mathbb{R}^{q>1}$.

4 MDB Versus MIB

As for a given bias b_i, the CD-probability exceeds the CI-probability, i.e. $\mathsf{P}_{\mathrm{CD}_i} \geq \mathsf{P}_{\mathrm{CI}_i}$, then for a given $\mathsf{P}_{\mathrm{CD}_i} = \mathsf{P}_{\mathrm{CI}_i}$, we have

$$\begin{aligned} b_i \in \mathbb{R} \quad &: |b_{i,\mathrm{MIB}}| \geq |b_{i,\mathrm{MDB}}| \\ b_i \in \mathbb{R}^{q>1} &: \|b_{i,\mathrm{MIB}}(d)\| \geq \|b_{i,\mathrm{MDB}}(d)\| \text{ for any } d \in \mathbb{S}^{q-1} \end{aligned} \tag{11}$$

The following example elaborates more on the above link between the MDB and the MIB.

Example Let $\underline{y} \in \mathbb{R}^4$ contain two pairs of observations of an unknown distance $x \in \mathbb{R}$ made using two different instruments, e.g., two different tape measures. The observations are assumed uncorrelated and equally precise with the same standard deviation σ. Under the null hypothesis $\mathcal{H}_0$, the observations are assumed to be bias-free, whereas under the alternative hypotheses $\mathcal{H}_i$ $(i = 1, 2)$, it is assumed that the observation pair made by one of the instruments are biased by $C_i b_i$ $(i = 1, 2)$ with $C_i \in \mathbb{R}^{4\times 2}$ and $b_i \in \mathbb{R}^2$. These hypotheses are formulated as

$$\begin{aligned} \mathcal{H}_0 &: \mathsf{E}(\underline{y}) = e_4\, x, & \mathsf{D}(\underline{y}) &= \sigma^2 I_4 \\ \mathcal{H}_i &: \mathsf{E}(\underline{y}) = e_4\, x + \left(u_i^2 \otimes I_2\right) b_i, & \mathsf{D}(\underline{y}) &= \sigma^2 I_4 \end{aligned} \tag{12}$$

where $\otimes$ shows the Kronecker product (Henderson and Pukelsheim 1983), $e_* \in \mathbb{R}^*$ the vector of ones, $I_* \in \mathbb{R}^{*\times *}$ the identity matrix, and $u_i^2 \in \mathbb{R}^2$ the canonical unit vector having one as its ith element and zeros otherwise.

The redundancy of $\mathcal{H}_0$-model is $r = 4 - 1 = 3 > 1$, which means, upon the rejection of $\mathcal{H}_0$, that the identification of potential source of error would be possible. Under $\mathcal{H}_1$, it is assumed that the mean-difference of the observables of the second instrument is zero, while under $\mathcal{H}_2$, this is assumed for the first instrument. To test the three hypotheses in consideration, the following detection and identification steps are exercised:

- Detection: The null hypothesis $\mathcal{H}_0$ is accepted if $\underline{t} \in \mathcal{P}_0$ with $\mathcal{P}_0$ given by (7).
- Identification: If $\mathcal{H}_0$ is rejected in the detection step, then $\mathcal{H}_i$ $(i = 1, 2)$ is selected if $\underline{t} \in \mathcal{P}_i$ with

$$\mathcal{P}_i = \left\{ t \in \mathbb{R}^r \setminus \mathcal{P}_0 \,\middle|\, T_i = \max_{j\in\{1,\ldots,k\}} T_j \right\} \tag{13}$$

where

$$T_i = t^T Q_{tt}^{-1} C_{t_i} \left(C_{t_i}^T Q_{tt}^{-1} C_{t_i}\right)^{-1} C_{t_i}^T Q_{tt}^{-1} t \tag{14}$$

would be a realization of the Generalized Likelihood Ratio (GLR) test statistic in case there is only one single alternative hypothesis (Teunissen 2006).

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

$$\bar{\underline{t}} = \mathcal{G}^{-T}\underline{t} \begin{cases} \overset{\mathcal{H}_0}{\sim} \mathcal{N}(0,\ I_r) \\ \overset{\mathcal{H}_i}{\sim} \mathcal{N}(\bar{C}_{t_i} b_i,\ I_r) \end{cases} \tag{15}$$

with $\bar{C}_{t_i} = \mathcal{G}^{-T} C_{t_i}$ and the Cholesky-factor $\mathcal{G}^T$ of the Cholesky-factorisation $Q_{tt} = \mathcal{G}^T\mathcal{G}$. The advantage of using $\bar{\underline{t}}$ over $\underline{t}$ lies in the ease of visualizing certain effects due to

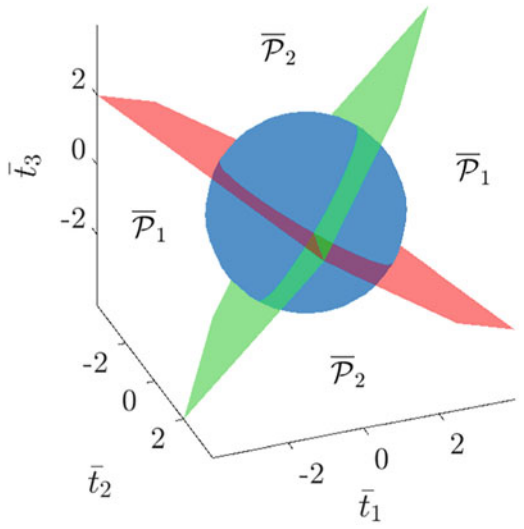

Fig. 2 Partitioning of the misclosure space $\mathbb{R}^3$ corresponding with $\bar{t}$ (15) using (7) and (13). The blue sphere shows the boundary of $\overline{\mathcal{P}}_0$ with $\mathsf{P}_{\mathrm{FA}} = 0.1$, while the orthogonal green and red planes separate $\overline{\mathcal{P}}_1$ from $\overline{\mathcal{P}}_2$

the identity-variance matrix of $\bar{t}$ (Zaminpardaz and Teunissen 2019). The partitioning corresponding with $\bar{t}$ is denoted by $\overline{\mathcal{P}}_i$ for $i = 0, 1, 2$.

The misclosure space ($\mathbb{R}^3$) partitioning corresponding with (7) and (13) is shown in Fig. 2. For the sake of visualization, instead of $\underline{t}$, we work with $\bar{t}$ defined in (15). The blue sphere shows the boundary of $\overline{\mathcal{P}}_0$ choosing $\mathsf{P}_{\mathrm{FA}} = 0.1$, while the green and red planes separate $\overline{\mathcal{P}}_1$ from $\overline{\mathcal{P}}_2$. The two planes are orthogonal to each other implying that $\overline{\mathcal{P}}_1$ and $\overline{\mathcal{P}}_2$ are the same in shape and size.

As b_i in (12) is a 2-vector, i.e. $b_i = [b_{i,1},\ b_{i,2}]^T$, the MDBs and the MIBs of the alternative hypotheses are dependent not only on the pre-set CD- and CI-probability, but also on the bias direction in $\mathbb{R}^2$. Figure 3 shows the MDB and MIB curves for $\mathcal{H}_i$ $(i = 1, 2)$ given $\sigma = 0.1$, $\mathsf{P}_{\mathrm{FA}} = 0.1$ and for different values of $\mathsf{P}_{\mathrm{CD}_i} = \mathsf{P}_{\mathrm{CI}_i}$. In each panel, in agreement with (11), it can be seen that the MIB curve encompasses the MDB curve.

Note, if $\mathsf{E}(\bar{t}|\mathcal{H}_i) = \bar{C}_{t_i} b_i$ lies on the border of $\overline{\mathcal{P}}_1$ and $\overline{\mathcal{P}}_2$, that the CI-probability of $\mathcal{H}_i$ cannot reach above 0.5. As shown in Fig. 2, the regions $\overline{\mathcal{P}}_1$ and $\overline{\mathcal{P}}_2$ are separated from each other by the following two planes

$$\bar{\tau}^T \left(\frac{\bar{C}_{t_1}^{\perp}}{\|\bar{C}_{t_1}^{\perp}\|} \pm \frac{\bar{C}_{t_2}^{\perp}}{\|\bar{C}_{t_2}^{\perp}\|} \right) = 0;\ \bar{\tau} \in \mathbb{R}^3 \tag{16}$$

with $\bar{C}_{t_i}^{\perp} \in \mathbb{R}^3$ being a vector of which the range space is the orthogonal complement of the range space of $\bar{C}_{t_i}$. It can be easily verified, if b_i is parallel to $[1,\ 1]^T$, that $\mathsf{E}(\bar{t}|\mathcal{H}_i)$ will lie on the intersection of the above planes. This explains the bands around the direction of $[1,\ 1]^T$ in Fig. 3 when $\mathsf{P}_{\mathrm{CI}_i}$ is set to be larger than 0.5. On the other hand, when b_i is parallel to $[1,\ -1]^T$, the MDB and the MIB are very close to each other. A bias along the direction of $[1,\ -1]^T$ makes $\mathsf{E}(\bar{t}|\mathcal{H}_i)$ lie at its farthest position from the planar borders of $\overline{\mathcal{P}}_1$ and $\overline{\mathcal{P}}_2$. Thus, under $\mathcal{H}_i$ $(i = 1, 2)$, most of the probability mass of the PDF of $\bar{t}$ that lies outside $\overline{\mathcal{P}}_0$ falls into the region $\overline{\mathcal{P}}_i$. As a result $\mathsf{P}_{\mathrm{CD}_i}$ and $\mathsf{P}_{\mathrm{CI}_i}$ are very close to each other for a given bias along $[1,\ -1]^T$, or alternatively the MDB and the MIB are very close to each other along $[1,\ -1]^T$ for a pre-set $\mathsf{P}_{\mathrm{CD}_i} = \mathsf{P}_{\mathrm{CI}_i}$. □

The above example clearly shows that the detection and identification performance of a testing procedure could be completely different from each other.

5 Deformation Monitoring

In this section, we continue our MDB-MIB comparison for a dam deformation monitoring case, inspired by an example in Heunecke et al. (2013, p. 227), see also (Zaminpardaz et al. 2020). Figure 4 [top] shows a top view of a dam over a lake,

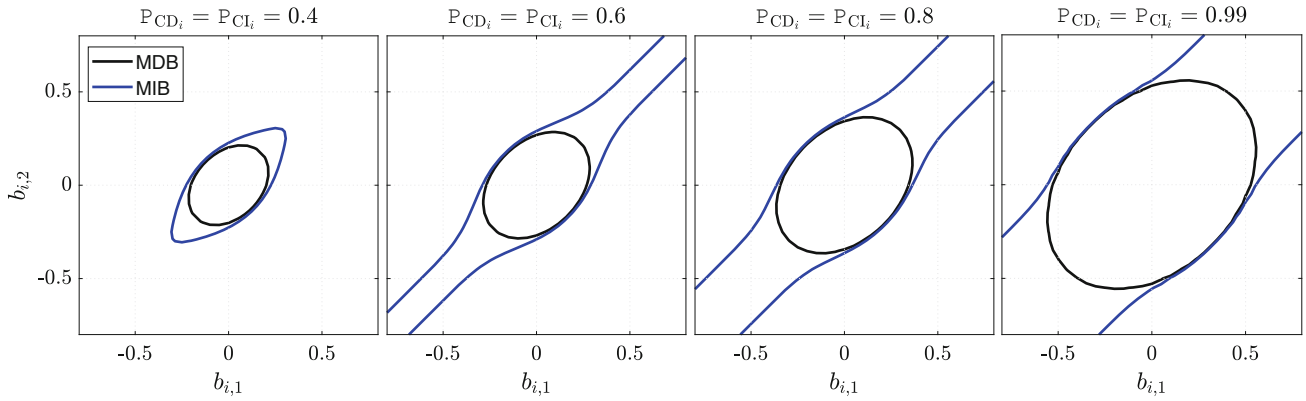

Fig. 3 Illustration of the MDB versus the MIB curves for testing the hypotheses in (12) using (7) and (13), given $\sigma = 0.1$ and $\mathsf{P}_{\mathrm{FA}} = 0.1$. The panels from left to right correspond to $\mathsf{P}_{\mathrm{CD}_i} = \mathsf{P}_{\mathrm{CI}_i}$ of 0.4, 0.6, 0.8 and 0.99, respectively

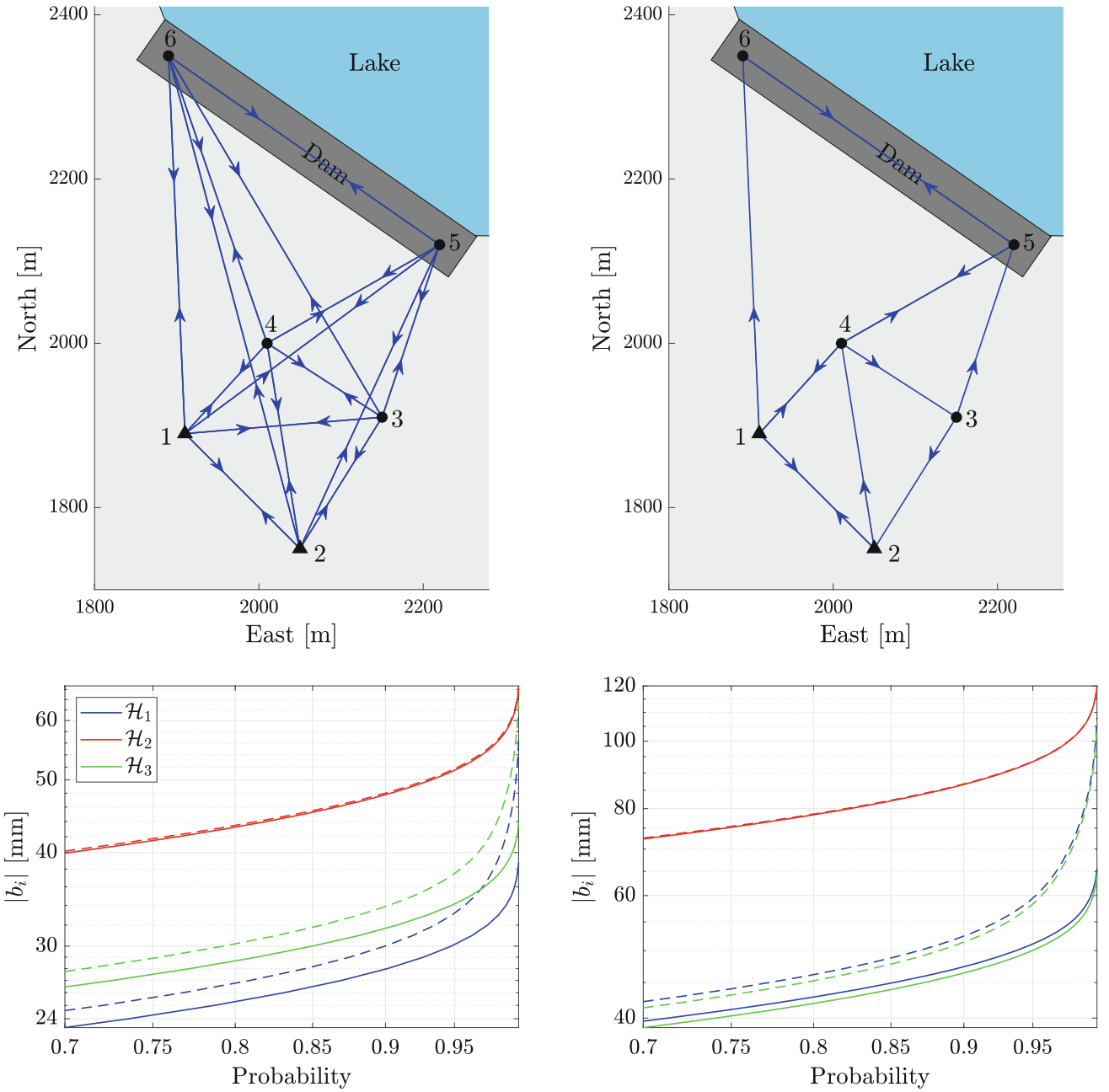

Fig. 4 Deformation monitoring of a dam (Zaminpardaz et al. 2020). [Top] The horizontal monitoring network consists of *four* reference points around the dam and *two* object points on the dam (points 5 and 6). The blue lines indicate the distance+direction measurements between their ending points, and the arrows point from total station to target. [Bottom] The graphs of MDB (solid lines) and MIB (dashed lines) of different alternative hypotheses in (18) as function of the pre-set probability. The results correspond with the testing procedure in (7) and (13), given $\mathsf{P}_{\mathrm{FA}} = 0.01$

together with two different 2-D terrestrial survey networks designed to monitor the dam's horizontal displacement. For simplicity, it is assumed that the dam is vertically stable. The survey networks consist of two *object* points on the dam subject to displacement (points 5, 6), and four *reference* points in a stable area close to the dam (points 1, 2, 3, 4). To determine horizontal deformations of the dam, two sets of measurements are collected at two times (or epochs), $l = 1, 2$.

In the survey network shown in Fig. 4 [top-left], each measurement set contains 60 measurements; five distance measurements and five direction measurements taken from each of the six points to the rest of the points by a total station. The distance and direction measurements are assumed

to be normally distributed with standard deviations of 1 cm and 10 s of arc, respectively. The measurements are assumed to be all uncorrelated. To make the scale, orientation and location of the 2-D survey network estimable, the coordinates of the reference points 1 and 2 (black triangles in Fig. 4 [top]) are assumed given. The 60 distance and direction observations at epoch l are then used to estimate the Easting and Northing of points $i = 3,\ldots,6$, together with the unknown instrument scale factor (one for the whole network) and six unknown orientations (one per instrument set-up).

To analyse the dam's horizontal displacement, we make use of the epoch-wise estimated coordinatesof points $i = 3,\ldots,6$ and their corresponding variance matrices. Let $x_{i,l} \in \mathbb{R}^2$ (for $i = 3,\ldots,6$ and $l = 1,2$) be the coordinate vector of point i at epoch l, and let $x_l = [x_{3,l}^T,\ x_{4,l}^T,\ x_{5,l}^T,\ x_{6,l}^T]^T \in \mathbb{R}^8$ for $l = 1,2$. Under the null hypothesis $\mathcal{H}_0$, where deformation is absent, we assume

$$\mathcal{H}_0 : x_2 = x_1 \quad \text{(all stable)} \tag{17}$$

The redundancy under $\mathcal{H}_0$ is $r = 8$. The dam is supposed to be subject to load of the water in the lake, and hence it is assumed that either only one or both of the dam points may be pushed back in the direction perpendicular to the dam. Thus we have three alternative hypotheses as

$$\begin{aligned} &\mathcal{H}_i : x_2 = x_1 + (u_{i+2}^4 \otimes d)\, b_i \quad \text{(point } i+4 \text{ is unstable, } i = 1,2) \\ &\mathcal{H}_3 : x_2 = x_1 + (u \otimes d)\, b_3 \quad \text{(points 5 and 6 are unstable)} \end{aligned} \tag{18}$$

with $u_{i+2}^4 \in \mathbb{R}^4$ the canonical unit vector having one as its $(i+2)$th element and zeros otherwise, $u = u_3^4 + u_4^4$, $d \in \mathbb{S}$ the known unit vector in the direction perpendicular to the dam, and $b_i \in \mathbb{R}$ the unknown scalar deformation size parameter. Note, under $\mathcal{H}_3$, that we assume that the object points 5 and 6 deform with the same amount.

We note that since $r = 8 > 1$, our testing procedure involves both the detection and identification step (7) and (13). Assuming $\mathsf{P}_{\text{FA}} = 0.01$, Fig. 4 [bottom-left] shows the MDB as a function of the CD-probability in solid curves, and the MIB as a function of the CI-probability in dashed curves for the three hypotheses in (18). For each hypothesis, its MIB graph lies above its MDB graph corroborating the first inequality in (11). For example, for a given pre-set probability of $\mathsf{P}_{\text{CD}_i} = \mathsf{P}_{\text{CI}_i} = 0.98$, there is an offset of almost 6mm between the MIB and the MDB in case of $\mathcal{H}_1$ and $\mathcal{H}_3$, while the $\mathcal{H}_2$'s MDB and MIB difference is at sub-mm level.

The MIB-MDB difference will change if the survey network measurement set-up changes. Figure 4 [top-right] shows a survey network obtained by removing 17 pairs of distance/direction measurements from the top-left network. As a result of loosing 34 measurements compared to the previous survey network, both the MDBs and the MIBs increase as shown in Fig. 4 [bottom-right]. It is observed that the MIB and the MDB can differ significantly from each other. For example, for a given pre-set probability of $\mathsf{P}_{\text{CD}_i} = \mathsf{P}_{\text{CI}_i} = 0.98$, there is an offset of almost 16mm between the MIB and the MDB in case of $\mathcal{H}_1$ and $\mathcal{H}_3$.

As shown in Fig. 4 [bottom], the MDB and the MIB, for a pre-set probability, differ from hypothesis to hypothesis. For example, for the range of probabilities shown in Fig. 4 [bottom-left], it is observed that

$$\begin{aligned} &|b_{2,\text{MDB}}| > |b_{3,\text{MDB}}| > |b_{1,\text{MDB}}| \\ &|b_{2,\text{MIB}}| > |b_{3,\text{MIB}}| > |b_{1,\text{MIB}}| \end{aligned} \tag{19}$$

As the MDB, for a given set of $\{\mathsf{P}_{\text{FA}}, \mathsf{P}_{\text{CD}_i}, r\}$, is driven by $\|c_{t_i}\|_{Q_{tt}}$, the first expression in the above equation can be explained by comparing $\|c_{t_i}\|_{Q_{tt}}$ for $i = 1,2,3$. The larger the value of $\|c_{t_i}\|_{Q_{tt}}$, the smaller the MDB is expected to be. For example, for the survey network shown in Fig. 4 [top-left], we have

$$\|c_{t_1}\|_{Q_{tt}} \approx 180;\ \|c_{t_2}\|_{Q_{tt}} \approx 105;\ \|c_{t_3}\|_{Q_{tt}} \approx 158 \tag{20}$$

which are driven by the network geometry, measurement precision and the direction of displacement. The above equation implies that $\mathcal{H}_1$ and $\mathcal{H}_2$ should, respectively, have the smallest and the largest MDBs among the three alternatives for a pre-set CD-probability. The MIB inequalities in (19) are due to a combination of (20), the shape and size of $\mathcal{P}_i$, magnitude of $\mathsf{E}(\underline{t}|\mathcal{H}_i)$ and its direction with respect to the borders of $\mathcal{P}_i$.

6 Summary and Concluding Remarks

In this contribution, a comparative analysis was provided of the detection and identification steps of statistical testing procedures. The *detection* step aims to validate the null hypothesis $\mathcal{H}_0$, while the identification step, upon the rejection of $\mathcal{H}_0$, aims to select the most likely alternative hypothesis among those in consideration.In case there is only

one alternative hypothesis, say $\mathcal{H}_1$, the rejection of $\mathcal{H}_0$ is equivalent to the identification of $\mathcal{H}_1$. This is however not the case when working with multiple alternatives. Having different functionalities, the detection and identification performance of the testing procedure should then be assessed using two different diagnostic tools. The detection capability of a testing regime is usually assessed by its Minimal Detectable Bias (MDB), whereas the testing identification performance should be evaluated by its Minimal Identifiable Bias (MIB).

Using the concept of misclosure space partitioning, we discussed testing decisions and their probabilities. Through this partitioning, it was shown that 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. One can then 'invert' these probabilities to determine their corresponding minimal biases, i.e. the MDB and the MIB. It was highlighted that a small MDB (or high probability of correct detection) does not necessarily imply a small MIB (or a high probability of correct identification), unless one is dealing with the special case of having only one single alternative hypothesis. The factors driving the difference between detection and identification performance were illustrated using a simple multiple-alternative testing example. Our evaluations were extended to basic deformation measurement system examples with multiple alternative hypotheses, where monitoring measurements were provided by a 2D terrestrial survey network.

References

Amiri Simkooei A (2001) Comparison of reliability and geometrical strength criteria in geodetic networks. J Geodesy 75(4):227–233

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

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

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

Caspary W, Borutta H (1987) Robust estimation in deformation models. Surv Rev 29(223):29–45

Henderson HV, Pukelsheim F, Searle SR (1983) On the history of the Kronecker product. Linear Multilinear Algebra 14:113–120

Heunecke O, Kuhlmann H, Welsch W, Eichhorn A, Neuner H (2013) Handbuch Ingenieurgeodäsie: Auswertung Geodätischer Überwachungsmessungen (in German). Wichmann, Berlin

Imparato D, Teunissen PJG, Tiberius CCJM (2019) Minimal detectable and identifiable biases for quality control. Surv Rev 51(367):289–299

Klein I, Matsuoka MT, Guzatto MP, Nievinski FG, Veronez MR, Rofatto VF (2019) A new relationship between the quality criteria for geodetic networks. J Geodesy 93(4):529–544

Kösters A, Van der Marel H (1990) Statistical testing and quality analysis of 3-D networks. In: Global positioning system: an overview, pp 282–289

Lehmann R, Lösler M (2017) Congruence analysis of geodetic networks–hypothesis tests versus model selection by information criteria. J Appl Geodesy 11(4):271–283

Lehmann R, Voß-Böhme A (2017) On the statistical power of Baarda's outlier test and somealternative. J Geodetic Sci 7(1):68—78

Nowel K (2020) Specification of deformation congruence models using combinatorial iterative dia testing procedure. J Geodesy 94(12):1–23

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

Teunissen PJG (1985) Quality control in geodetic networks. In: Grafarend E, Sanso F (eds) Optimization and design of geodetic networks, pp 526–547

Teunissen PJG (2006) Testing theory: an introduction, 2nd edn. Series on mathematical geodesy and positioning. Delft University Press, Delft

Teunissen PJG (2017) Batch and recursive model validation. In: Teunissen PJG, Montenbruck O (eds) Springer handbook of global navigation satellite systems, pp 727–757, Chap. 24

Teunissen PJG (2018) Distributional theory for the DIA method. J Geodesy 92(1):59–80. https://doi.org/10.1007/s00190-017-1045-7

Zaminpardaz S (2016) Horizon-to-elevation mask: a potential benefit to ionospheric gradient monitoring. In: Proceedings of the 29th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2016), pp 1764–1779

Zaminpardaz S, Teunissen PJG (2019) DIAdatasnooping and identifiability. J Geodesy 93(1):85–101

Zaminpardaz S, Teunissen P, Nadarajah N, Khodabandeh A (2015) GNSS array-based ionospheric spatial gradient monitoring: precision and integrity analyses. In: Proceedings of the ION 2015 Pacific PNT Meeting, pp 799–814

Zaminpardaz S, Teunissen P, Tiberius C (2020) A risk evaluation method for deformation monitoring systems. J Geodesy 94(3):1-Ű15