Minimal Detectable and Identifiable Biases for Quality Control

February 2, 2018

Abstract

The Minimal Detectable Bias (MDB) is an important diagnostic tool in data quality control. The MDB is traditionally computed for the case of testing the null hypothesis against a single alternative hypothesis. The actual practice of statistical testing and data quality control is however one where not one, but multiple alternative hypotheses are considered. We show that this has two important consequences for one's interpretation and use of the popular MDB. Firstly, we demonstrate, theoretically as well as empirically, that care should be exercised in using the single-hypothesis-based MDB for the multiple-hypotheses case. Secondly, as detection and identification of an alternative hypothesis do not coincide in the multiple-hypotheses case, we show that for identification purposes, not the MDB, but the Minimal Identifiable Bias (MIB) should be used as the proper diagnostic tool. We analyse the circumstances that drive the differences between the MDBs and MIBs, show how they can be computed using Monte Carlo simulation, and illustrate by means of examples the significant differences that one can experience between detectability and identifiability.

1 Introduction

Statistical testing for modelling errors and biases is an important component of any data quality control procedure (Alberda 1976; Amiri-Simkooei 2004; Klein et al. 2017, 2015; Yetkin, Berber and Inal 2014). Such a testing procedure then usually consists of a detection, identification and adaptation (DIA) step (Baarda 1968; Imparato 2016; Teunissen 2000). In the detection step one decides whether they can have enough confidence in the null-hypothesis $\mathcal{H}_0$, which is, at onset, the assumed working model. Once confidence in $\mathcal{H}_0$ has been declared, this model is eventually used to estimate the unknown parameters, usually in a least-squares sense. In case confidence in $\mathcal{H}_0$ is lacking, subsequently identification of the potential source of model error is carried out. This implies a search among the specified alternative hypotheses $\mathcal{H}_i$, $i = 1, ..., m$, for the most likely model misspecification. After successful identification of the suspected model error, adaptation is needed to perform a corrective action on earlier $\mathcal{H}_0$-based inferences. Thus, with the null hypothesis rejected, the identified alternative hypothesis becomes the new null hypothesis and this model is then effectively used to provide estimates for the unknown parameters. For the distributional theory of the DIA-method, including a discussion on how the first moment of the parameter estimators is impacted by the testing procedure, see (Teunissen 2017; Teunissen, Imparato and Tiberius 2017).

In this contribution we analyse what diagnostic quantities are actually needed to adequately describe the performance of the testing procedure. We thereby distinguish between detection and identification, and also between single alternative hypothesis testing and multiple alternative hypotheses testing. For detection, Baarda (Baarda 1967, 1968) introduced the important and well-known concept of the Minimal Detectable Bias (MDB). It describes, for each alternative hypothesis $\mathcal{H}_i$, the size of model error or bias that can be detected with a certain pre-set probability, when employing a detector test with user-defined false alarm rate (Salzmann 1991; Teunissen 1998; Van der Marel and Kosters 1990). Although the MDB is usually computed based on Baarda's normally distributed w-statistic, it can also be linked to the Chi-square distributed overall model test. The MDB is a general concept that applies in principle to any detection test for validating $\mathcal{H}_0$. Hence, when computing and reporting MDBs, one should always clarify which detection test is actually used

for validating the null hypothesis. In this contribution, we consider as the alternative hypotheses outliers in individual observations, and the testing procedure itself will be based on Baarda's w-test statistic for data snooping (Baarda 1967). Thus with m observations, there will also be m alternative hypotheses.

There is no difference between detection and identification in case only a single alternative hypothesis is in play. In that case, rejection of $\mathcal{H}_0$ automatically leads to acceptance of the only alternative, and the corresponding Minimal Detectable Bias (MDB) also describes the smallest bias that can be *identified*. As this equivalence disappears in multiple alternative hypotheses testing, one has to be aware of the following two consequences. Firstly, as a different test statistic will be used for *detection*, also the MDB of a hypothesis $\mathcal{H}_i$ will then change from the single alternative hypothesis $\mathrm{MDB}_1(i)$ into the m alternative hypotheses $\mathrm{MDB}_\mathrm{m}(i)$. Secondly, as *identification* differs from detection in the m alternative hypotheses case, the MDBs will not represent the minimal size of biases that one is able to correctly identify. To obtain such information (Teunissen 2017), one will have to consider — in analogy with the Minimal Detectable Bias (MDB) — the Minimal Identifiable Bias (MIB). Earlier it has been recognized already (Prószyński 2015) that the MDB does not cover identification. In this contribution we show that for the given testing procedure,

$$\mathrm{MDB}_\mathrm{m}(i) \leq \mathrm{MDB}_1(i) = \mathrm{MIB}_1(i) \leq \mathrm{MIB}_\mathrm{m}(i) \tag{1}$$

This relationship has two implied consequences. Firstly, it shows that the easy-to-compute Minimal Detectable Bias $\mathrm{MDB}_1(i)$ can be used as a safe upper bound for the MDB when m alternative hypotheses are considered, i.e. $\mathrm{MDB}_\mathrm{m}(i)$. We will analyse the circumstances under which such an approximation can be considered satisfactory. Secondly, the relationship implies that identification is more difficult than detection. This is something one has to be aware of when using identified hypotheses for adaptation purposes. The MDBs cannot be used as safe bounds for the MIBs. We will analyse the circumstances that drive the differences between the MDBs and MIBs, and in support of this analysis we also show how the m-hypotheses MDBs and MIBs can actually be computed using Monte Carlo simulation.

This contribution is organized as follows. We start in section 2 with a brief review of the necessary estimation and testing results of linear model theory. Then in section 3, we introduce our testing procedure, give a proof of relation (1) and show how the multivariate minimal biases $\mathrm{MDB}_\mathrm{m}(i)$ and $\mathrm{MIB}_\mathrm{m}(i)$ can be numerically computed by means of Monte Carlo simulation. In section 4 we provide a qualitative description, as a function of the normalized bias size and overall false alarm rate, of the behaviour of the univariate and multivariate probabilities of missed detection and correct identification. Then in section 5 we illustrate by means of two surveying examples the significant differences that one can indeed experience in practice between detectability and identifiability. The conclusions and findings are summarized in section 6.

2 Detection and the Minimal Detectable Bias

In this section we briefly review necessary estimation and testing results of linear model theory.

2.1 Detection

Before any start can be made with statistical model validation, one needs to have a clear idea of the models one wants to work with. The null-hypothesis $\mathcal{H}_0$, also referred to as working hypothesis, consists of the model one believes to be valid under nominal, or normal working conditions. We assume the null-hypothesis to be of the form

$$\mathcal{H}_0 : \quad \mathsf{E}(y) = Ax \,, \; \mathsf{D}(y) = Q_{yy} \tag{2}$$

with $\mathsf{E}(.)$ the expectation operator, $y \in \mathbb{R}^m$ the normally distributed random vector of observables, $A \in \mathbb{R}^{m \times n}$ the known design matrix of rank n, and $x \in \mathbb{R}^n$ the unknown parameter vector. $\mathsf{D}(.)$ is the dispersion operator and $Q_{yy} \in \mathbb{R}^{m \times m}$ the known positive-definite variance matrix of the observables. As A is assumed to be of full rank, the *redundancy* of the model is $r = m - n$. For example, in case of GNSS satellite positioning, the model (2) is formed by the linear(ized) pseudorange code and/or carrier-phase observation equations; see e.g. Hofmann-Wellenhof, Lichtenegger and Wasle (2008); Leick (2004); Teunissen and Montenbruck (Eds.).

The first step in testing usually consists of a check on the overall validity of the null-hypothesis $\mathcal{H}_0$. Such a detection-step provides information on whether one can have confidence in the assumed null hypothesis.

By choosing an acceptance region $\mathcal{A}_\alpha \subset \mathbb{R}^m$, the test takes the form

$$\text{Accept } \mathcal{H}_0 \text{ if } y \in \mathcal{A}_\alpha, \text{ reject otherwise} \tag{3}$$

The performance of such a test is described by its probabilities of false alarm P_{FA} and missed detection P_{MD},

$$
\begin{array}{rcl}
\mathsf{P}_{\mathrm{FA}} & = & P[y \notin \mathcal{A}_\alpha | \mathcal{H}_0 = \text{true}] = \alpha \\
\mathsf{P}_{\mathrm{MD}} & = & P[y \in \mathcal{A}_\alpha | \mathcal{H}_0 = \text{false}] = \beta
\end{array}
\tag{4}
$$

The probability of false alarm α, also known as significance level, is the probability of rejecting the null-hypothesis while it is true in fact (also known as a type-I error), while the probability of missed detection β is the probability of wrongfully accepting the null-hypothesis (also known as a type-II error). The complement to the probability of missed detection, $\mathsf{P}_{\mathrm{CD}} = 1 - \mathsf{P}_{\mathrm{MD}} = \gamma$, is the probability of correct detection, also known as the power of the test. The false alarm probability can be computed by means of a multivariate integration of the null-hypothesis' probability density function of y over the complement of $\mathcal{A}_\alpha$. For the computation of the missed detection, however, we first need to specify the alternative hypothesis $\mathcal{H}_a$ to which $\mathcal{H}_0$ is tested against. Here we consider the following type of alternative hypotheses,

$$\mathcal{H}_a : \quad \mathsf{E}(y) = Ax + Cb \,, \quad \mathsf{D}(y) = Q_{yy} \tag{5}$$

with $b_y = Cb$ the additional observation bias-vector, in which $b \in \mathbb{R}^q$ is unknown and matrix $C \in \mathbb{R}^{m \times q}$ known. We have $q \leq m - n = r$, as the design matrix under $\mathcal{H}_a$, $[A, C] \in \mathbb{R}^{m \times (n+q)}$, is assumed to be of full rank. Through the choice of matrix C one may model, for instance in GNSS applications, the presence of one or more blunders (outliers) in the code (pseudorange) data, cycle-slips in carrier-phase data, satellite failures, antenna-height errors, or incorrect neglectance of atmospheric delays.

For each possible alternative hypothesis $\mathcal{H}_a$, i.e. for each choice of matrix C and set values for the elements of vector b, one may compute the probability of missed detection P_{MD}. This boils down to a multivariate integration of the alternative hypothesis' probability density function of y over the acceptance region $\mathcal{A}_\alpha \subset \mathbb{R}^m$. Alternatively, as one is often interested in the 'size' of bias vector b that can be detected with a certain pre-set probability, one may instead follow the inverse route. In that case one assumes a certain probability of missed detection, $\mathsf{P}_{\mathrm{MD}} = \beta$ say, and then computes for a given C-matrix, the corresponding b-vector. This latter approach then leads to the concept of the Minimal Detectable Bias as introduced by Baarda (1967).

2.2 Minimal Detectable Bias (MDB)

No particular choice for the acceptance region $\mathcal{A}_\alpha$ of $\mathcal{H}_0$ has yet been made. Different choices lead to different probabilities of missed detection for the same $\mathcal{H}_a$. A commonly used choice is given by Baarda (1968) and by (Teunissen 2000),

$$\mathcal{A}_\alpha = \{ y \in \mathbb{R}^m \mid ||\hat{e}||^2_{Q_{yy}} \leq \chi^2_\alpha(r, 0) \} \tag{6}$$

with the least-squares residual vector $\hat{e} = \mathsf{P}_A^\perp y$, orthogonal projector $\mathsf{P}_A^\perp = I_m - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$, $||\hat{e}||^2_{Q_{yy}} = \hat{e}^T Q_{yy}^{-1} \hat{e}$, and critical value $\chi^2_\alpha(r, 0)$. The critical value is computed from the central Chi-square distribution with $r = m - n$ degrees of freedom and a chosen false alarm rate, or level of significance, α. The detection-test with the above acceptance region (6) is referred to as the overall-model test (OMT). It is known to be a uniformly-most-powerful-invariant (UMPI) detector test, from Arnold (1981); Lehmann and Voß-Böhme (2017); Teunissen (2000), when testing (2) against (5).

In order to determine the OMT's probability of missed detection, we need the distribution of the 'weighted-sum-of-squared-residuals' test statistic $||\hat{e}||^2_{Q_{yy}}$ under $\mathcal{H}_a$. It is given as

$$||\hat{e}||^2_{Q_{yy}} \sim \chi^2(r, \lambda_q) \tag{7}$$

with non-centrality parameter $\lambda_q = b^T Q_{\hat{b}\hat{b}}^{-1} b$, in which $Q_{\hat{b}\hat{b}} = (\bar{C}^T Q_{yy}^{-1} \bar{C})^{-1}$ is the variance matrix of the best linear unbiased estimator (BLUE) of b under $\mathcal{H}_a$, and $\bar{C} = \mathsf{P}_A^\perp C$. From combining (6) and (7), the probability of missed detection for hypothesis $\mathcal{H}_a$ follows as $\mathsf{P}_{\mathrm{MD}} = P[\chi^2(r, \lambda_q) \leq \chi^2_\alpha(r, 0)] = \beta$.

This relation can also be used as a starting point for computing the OMT's Minimal Detectable Biases. First, the non-centrality parameter λ_q is computed from this relation as a function of α, β and r, denoted as $\lambda_q(\alpha, \beta, r)$, the value of which provides then a yardstick for the detectable bias,

$$\lambda_q(\alpha, \beta, r) = b^T Q_{\hat{b}\hat{b}}^{-1} b = ||\bar{C}b||_{Q_{yy}}^2 \tag{8}$$

This quadratic equation describes a hyper-ellipsoid in $\mathbb{R}^q$ for the bias b. Biases outside this region will have a larger than $\gamma = 1 - \beta$ probability of being detected under $\mathcal{H}_a$ using the OMT, while biases inside it will have a smaller probability. A further inversion is possible to find the corresponding biases themselves. Let u be a unit vector ($||u|| = 1$) and parametrize the bias vector as $b = ||b||u$. Substitution into (8), followed by inversion gives then,

$$b = \sqrt{\frac{\lambda_q(\alpha, \beta, r)}{||\bar{C}u||_{Q_{yy}}^2}} u \quad \text{with} \ \ ||u|| = 1 \tag{9}$$

This is the vectorial form of Baarda's *Minimal Detectable Bias* (MDB) (Baarda 1967, 1968). Baarda referred to his MDBs as 'boundary values' (in Dutch: grenswaarden); nowadays more customary term MDB was introduced in Teunissen (1989). The length $||b||$ of the MDB vector is the smallest magnitude of the bias vector b that can be found with at least probability $\gamma = 1 - \beta$ in the direction u using the detection-test (6). By letting u vary over the unit sphere in $\mathbb{R}^q$, one obtains the whole range of MDBs that can be detected with probability γ. Applications can be found, for example, in Kok (1982); Van Mierlo (1980) for deformation analysis, in (Salzmann 1991; Tiberius 1998) for GNSS positioning and navigation and in Teunissen (1986) for trend testing.

Baarda, in his work on the strength analysis of general purpose geodetic networks, applied his general MDB-form to *data snooping* for outlier detection of DGCC (1982). In that case $q = 1$, and the m alternative hypotheses take the form

$$\mathcal{H}_i: \quad \mathsf{E}(y) = Ax + c_i b_i \ , \ i = 1, \ldots, m \ , \ \mathsf{D}(y) = Q_{yy} \tag{10}$$

with c_i being the canonical unit vector having 1 as its ith entry and zeroes otherwise. Since $\sigma_{\hat{b}_i}^2 = ||\bar{c}_i||_{Q_{yy}}^{-2}$ when $q = 1$, the corresponding scalar MDBs, $\text{MDB}(i) = |b_i|$, for each of these m hypotheses can then be computed as

$$\text{MDB}(i) = \sigma_{\hat{b}_i} \sqrt{\lambda_{q=1}(\alpha, \beta, r)} \ , \ i = 1, \ldots, m \tag{11}$$

The MDB gets smaller for larger α, larger β (smaller power γ) and smaller $\sigma_{\hat{b}_i}$ (more precisely estimated $\hat{b}_i$).

3 Detection and Identification

In the remainder of this contribution we restrict our attention to alternative hypotheses of the form (10), i.e. we restrict our attention to *data snooping*. And instead of detection only (accept or reject $\mathcal{H}_0$), we are now also interested in identification of the correct hypothesis.

3.1 Test procedure

There would be no difference between detection and identification in case there would be only a single alternative hypothesis, i.e. if one would assume that only one of two hypotheses could be true, $\mathcal{H}_0$ or $\mathcal{H}_a$. In that case rejection of $\mathcal{H}_0$ automatically implies acceptance of $\mathcal{H}_a$. In our case, however, we will be working with multiple alternatives. This implies that rejection of $\mathcal{H}_0$ still requires a further identification as to which of the m alternatives to accept.

For data snooping we make use of Baarda's w-test statistic, see Baarda (1968),

$$w_i = \frac{c_i^T Q_{yy}^{-1} \hat{e}}{\sqrt{c_i^T Q_{yy}^{-1} Q_{\hat{e}\hat{e}} Q_{yy}^{-1} c_i}} \sim \mathcal{N}\left(\frac{b_i}{\sigma_{\hat{b}_i}}, 1\right) \tag{12}$$

The UMPI-test for testing $\mathcal{H}_0$ against a *single* alternative $\mathcal{H}_i$ of (10) would then be to reject H_0 in favour of $\mathcal{H}_i$ if $y \notin A_{\alpha_1}^i$, with the acceptance region given as $\mathcal{A}_{\alpha_1}^i = \{y \in \mathbb{R}^m \mid |w_i| \leq \chi_{\alpha_1}(1,0)\}$, with $\chi_{\alpha_1}(1,0) = \sqrt{\chi_{\alpha_1}^2(1,0)}$. In our case however, we do not have a single alternative, but multiple alternative hypotheses. We therefore take as the acceptance region for $\mathcal{H}_0$ the intersection of the m individual acceptance regions $\mathcal{A}_{\alpha_1}^i$, $i = 1, \ldots, m$. We thus accept the null-hypothesis if $y \in \mathcal{A}_{\alpha_m} = \cap_{i=1}^m A_{\alpha_1}^i$ (i.e. if none of the m w-tests gets rejected), and reject otherwise. But as this is only the detection-part of the testing procedure, we still need to decide which $\mathcal{H}_i$ to choose once $\mathcal{H}_0$ is rejected. Here we choose for $\mathcal{H}_i$ if its w_i is largest in absolute value. The testing procedure can be summarized as

$$\text{Accept} \quad \mathcal{H}_0 \quad \text{if} \quad \max_{j \in \{1, \ldots, m\}} |w_j| \leq k = \chi_{\alpha_1}(1,0) \tag{13}$$

otherwise

$$\text{Accept} \quad \mathcal{H}_i \quad \text{if} \quad |w_i| = \max_{j \in \{1, \ldots, m\}} |w_j| > k = \chi_{\alpha_1}(1,0) \tag{14}$$

Hence, (13) is the detection-step, and (14) the identification-step. Note, since we are here solely working with the m w_i-statistics and alternative hypotheses of the form (10), that detector (13) is different from the one based on (6). The detector based on (6) is ellipsoidal shaped, while detector (13) is constructed from the intersections of m banded regions $\mathcal{A}_{\alpha_1}^i$, $i = 1, \ldots, m$. The two regions are shown in Figure 1, with (6) on the left, and (13) on the right (note: the region $\mathcal{A}_\alpha^i$ is shown to be spherical (circular) in the figure as the used metric is with respect to Q_{yy}).
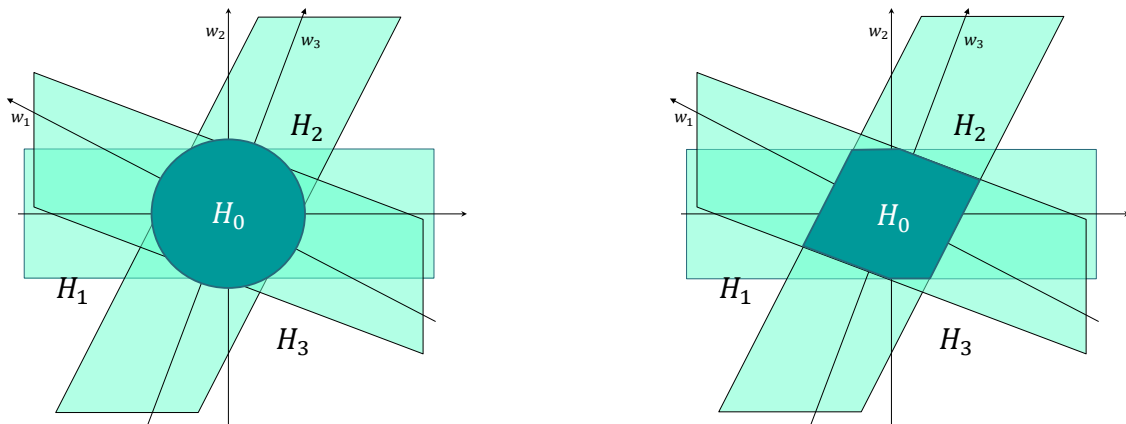

Figure 1: Two examples of acceptance regions for $\mathcal{H}_0$: ellipsoidal acceptance region $\mathcal{A}_\alpha$ (cf. 6) of the overall model test (left; dark green), and polygonal acceptance region $\mathcal{A}_{\alpha_m}$ (cf. 13) of the multiple w_i-test (right; dark green). The light-green banded regions are those of $\mathcal{A}_{\alpha_1}^i$, $i = 1, 2, 3$, for resp. $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ ($m = 3$).

As the two acceptance regions are different, their MDBs for the same alternative hypothesis $\mathcal{H}_i$ will be different as well, even when the same false alarm P_{FA} and the same missed detection probability P_{MD} are chosen. It is therefore important to explicitly describe the underlying testing procedure when presenting MDBs.

3.2 Minimal Detectable Bias revisited

As the MDB depends on the chosen detector, the MDBs for (13) will differ from the OMT-based ones of (11). For the probability of missed detection through (13) P_{MD_m} we have, since $\mathcal{A}_{\alpha_m} = \cap_{i=1}^m \mathcal{A}_{\alpha_1}^i$,

$$P_{\text{MD}_m}(i) = P[y \in \mathcal{A}_{\alpha_m} | \mathcal{H}_i] \leq P[y \in \mathcal{A}_{\alpha_1} | \mathcal{H}_i] = P_{\text{MD}_1}(i) \tag{15}$$

in which $P_{\text{MD}_1}(i)$ is the missed detection probability in case one would have only one alternative hypothesis $\mathcal{H}_i$ (P_{MD} with index $_1$ denotes the *univariate* probability of missed detection, and P_{MD} with index $_m$ the

multi-variate probability of missed detection). To contribute to the $\mathsf{P}_{\mathrm{MD_m}}(i)$ the observation y needs to lie in the intersection of all m individual acceptance regions, so that none of them gets rejected, while, in fact, $\mathcal{H}_i$ is true. The intersection is a smaller area than each of the individual ones; hence, the probability gets smaller, $\mathsf{P}_{\mathrm{MD_m}}(i) \leq \mathsf{P}_{\mathrm{MD_1}}(i)$. Thus, the missed detection probability gets smaller in case of multiple testing. The false alarm rate however, will generally get larger, $\alpha_m \geq \alpha_1$. Would the w_i-test-statistics all be independent, then $1 - \alpha_m = (1 - \alpha_1)^m$, which for small α_1 can be approximated as $1 - \alpha_m \approx 1 - m\alpha_1$ or $\alpha_m \approx m\alpha_1$. This approximation generally works quite well in practice, especially for small false alarm probabilities, unless the underlying model is too weak, having low redundancy and/or very correlated w_i-test-statistics. For the correlated case, we have the upper bound, $\alpha_m \leq m\alpha_1$, which is known as the Bonferroni inequality (Bonferroni 1936).

Although the right-hand side of the inequality (15) is usually easy to compute (with a one-dimensional normally distributed random variable, $\mathsf{P}_{\mathrm{MD_1}}(i) = \mathsf{P}[|w_i| < k|\mathcal{H}_i]$), the computation of the multivariate missed detection probability $\mathsf{P}_{\mathrm{MD_m}}(i)$ itself generally needs a Monte Carlo integration. This is due to the geometric complexity of the acceptance region $\mathcal{A}_{\alpha_m}$ (see Figure 1 at right). In the next section we will provide a qualitative analysis of the upper bound (15) together with an quantitative investigation of its closeness.

Similar to the P_{MD}-inequality (15), we have the MDB-inequality (with 11)

$$\mathrm{MDB_m}(i) \leq \mathrm{MDB_1}(i) = \sigma_{\hat{b}_i}\sqrt{\lambda_{q=1}(\alpha_1, \beta, 1)} \tag{16}$$

The right-hand side is the easy-to-use MDB-formula one often uses for computing the MDBs of data snooping. Hence, this result shows that these MDBs can be used as safe bounds for the actual values $\mathrm{MDB_m}(i)$. In a next section, the two will be compared for various examples.

3.3 The Minimal Identifiable Bias (MIB)

Next to detection, we have identification (cf. 14). The $\mathrm{MDB_m}$'s of (16) are minimal *detectable* biases, but not minimal *identifiable* biases (Teunissen 2017). To determine the MIBs, we start from the complement of P_{MD} being the probability of correct detection, $\mathsf{P}_{\mathrm{CD}} = 1 - \mathsf{P}_{\mathrm{MD}}$. As correct detection does not necessarily imply correct identification (this is only the case with $m = 1$), the probability of correct detection is the sum of the probability of correct identification (selecting the single correct alternative) and the probability of wrong identification (selecting one of the $m-1$ other alternatives): $\mathsf{P}_{\mathrm{CD}} = \mathsf{P}_{\mathrm{CI}} + \mathsf{P}_{\mathrm{WI}}$. Thus the probability of correct identification is smaller than that of correct detection. As a consequence the MIBs will be larger than their MDB counterparts,

$$\mathsf{P}_{\mathrm{CI}} \leq \mathsf{P}_{\mathrm{CD}} \implies \mathrm{MIB} \geq \mathrm{MDB} \tag{17}$$

This shows that one has to be careful in using MDBs as MIBs. It is one thing to be able to *detect* a bias, but an all together other thing is to also be able to *identify* it.

For the probability of correct identification, we have similarly to (15), the inequality

$$\mathsf{P}_{\mathrm{CI_m}}(i) = P[|w_i| = \max_j |w_j| > k|\mathcal{H}_i] \leq P[|w_i| > k|\mathcal{H}_i] = \mathsf{P}_{\mathrm{CI_1}}(i) = 1 - \mathsf{P}_{\mathrm{MD_1}}(i) \tag{18}$$

On the right-hand side, test-statistic w_i shall exceed, in absolute sense, the critical value k, and on the left-hand side, in addition, w_i needs to be the largest of all m w test-statistics, resulting in a smaller probability. The last equality follows from the fact that correct identification equals correct detection in case one would have only the one alternative hypothesis $\mathcal{H}_i$. From the probabilistic inequality (18) follows for the Minimal Identifiable Biases,

$$\mathrm{MIB_m}(i) \geq \mathrm{MIB_1}(i) = \mathrm{MDB_1}(i) \tag{19}$$

3.4 Computing the MDBs and MIBs

The univariate Minimal Detectable Bias $\mathrm{MDB_1}(i)$ is relatively easy to compute using the expression (16). The computation of $\mathrm{MDB_m}(i)$ and $\mathrm{MIB_m}(i)$, however, is more involved and needs to be based on simulation. First we need to be able to determine the acceptance region $\mathcal{A}_{\alpha_m}$ for a user-given value of the overall

false alarm rate α_m. Thus given α_m, we need to be able to compute α_1 or $k = \chi_{\alpha_1}(1,0)$, thereby giving $\mathcal{A}_{\alpha_m} = \cap_{i=1}^m \mathcal{A}_{\alpha_1}^i$. The procedure of computing α_1 from α_m is a problem already addressed in (Lehmann 2010, 2011) and can briefly be described as follows: simulate N samples from the $\mathcal{H}_0$-distribution of the m-vector of w-test-statistics $w = [w_1, w_2, \ldots w_m]^T \overset{\mathcal{H}_0}{\sim} \mathcal{N}(0, Q_{ww})$, with $(Q_{ww})_{ii} = 1$ and $(Q_{ww})_{ij} = \rho(w_i, w_j)$ being the correlation between w_i and w_j. The ith entry of the vth such sample is denoted as w_i^v. Then sort out in ascending order the maximum components $\tilde{w}^v = \max_i w_i^v$, with $v \in \{1, 2, \ldots, N\}$, and compute (Lehmann 2012; Lemeshko and Lemeshko 2005):

$$k = \frac{1}{2} \left(\tilde{w}_{\lfloor (1-\alpha_m)N \rceil} + \tilde{w}_{\lfloor (1-\alpha_m)N+1 \rceil} \right) \tag{20}$$

where the subscript of $\tilde{w}$ ($\lfloor (1-\alpha_m)N \rceil$ and $\lfloor (1-\alpha_m)N+1 \rceil$) is the rounded ordinal number that indicates the position of the selected elements in the ascendingly sorted list of $\tilde{w}^v$. From the above computed k, the corresponding univariate false alarm rate is computed as $\alpha_1 = \chi_k^{-1}(1,0)$.

With k (or α_1) known, $\mathcal{A}_{\alpha_m} = \cap_{i=1}^m A_{\alpha_1}^i$ is known, thus making it possible to compute the multivariate missed detection probability P_{MD} through simulation. Again N samples are drawn, but now from the $\mathcal{H}_j$-distribution of $w = [w_1, w_2, \ldots w_m]^T \overset{\mathcal{H}_j}{\sim} \mathcal{N}(b_w, Q_{ww})$, in which the ith entry of b_w is given as

$$\mathsf{E}(w_i | \mathcal{H}_j) = \rho(w_i, w_j) \frac{b_j}{\sigma_{\hat{b}_j}} \tag{21}$$

By counting the number of missed detections in the generated samples, one can determine the corresponding probability of missed detection. Thus if N is the total number of samples simulated, we first count the number of times $n_{\mathrm{MD}}(b_j)$ that $\cap_{i=1}^m |w_i^v| < k$ for $v = 1, \ldots, N$, and then approximate the probability of missed detection as

$$\mathsf{P}_{\mathrm{MD}_m}(b_j) \approx \frac{n_{\mathrm{MD}}(b_j)}{N} \tag{22}$$

The probability of correct identification $\mathsf{P}_{\mathrm{CI}_m}(b_j)$ is computed in a similar manner, although the counting is now based on the number of times that $\mathcal{H}_j$ is correctly identified, i.e. how often $|w_j^v| = \max_i |w_i^v| > k$ for $v = 1, \ldots, N$. As the approximation (22) gets better for larger N, the question is how large a sample should be chosen. Using the results of (Tanizaki 2004) we found that for our purposes $N = 10^5$ is sufficient to compute probabilities of order 10^{-3} or larger, with high accuracy.

The computed probabilities $\mathsf{P}_{\mathrm{MD}_m}(b_j)$ and $\mathsf{P}_{\mathrm{CI}_m}(b_j)$ depend on b_j, i.e. on the assumed size of the bias under $\mathcal{H}_j$. Thus in order to extract the MDB and MIB, the above simulations need to be repeated for a range of b_j-values, so that finally the minimal detectable and identifiable biases can be computed for a certain chosen fixed reference probability γ (e.g. $\gamma = 0.8$) as

$$\begin{aligned} \mathrm{MDB}_m &= \arg\min_{b_j} \mathsf{P}_{\mathrm{MD}_m}(b_j) < 1 - \gamma \\ \mathrm{MIB}_m &= \arg\min_{b_j} \mathsf{P}_{\mathrm{CI}_m}(b_j) > \gamma \end{aligned} \tag{23}$$

4 A qualitative description

In this section we provide a qualitative description of the behaviour of the univariate and multivariate probabilities of missed detection and correct identification.

4.1 Detection

We consider a simple model with $A = [1, \ldots, 1]^T$ and $\mathsf{D}(y) = I_m$. The data have been simulated with a bias in the first observable, i.e. the true hypothesis is $\mathcal{H}_{i=1}$. The bias size b was varied and the simulations were done for $m = 4$ and $m = 10$. Figure 2 (at left for $m = 4$, and at right for $m = 10$) shows the graphs of $\mathsf{P}_{\mathrm{MD}_m}$ (full line) and $\mathsf{P}_{\mathrm{MD}_1}$ (dashed line) as a function of the normalized bias size $b/\sigma_{\hat{b}}$ for five different values of α_m. The general characteristic of the probability of missed detection as a function of bias-size b, precision $\sigma_{\hat{b}}$, false alarm rate α_m or critical value k, is as follows:

$$\mathsf{P}_{\mathrm{MD}} \downarrow \quad \text{if} \quad b/\sigma_{\hat{b}} \uparrow \quad \text{and/or} \quad \alpha_m \uparrow \quad (k \downarrow) \tag{24}$$

Thus the P_{MD} gets smaller for larger biases, more precisely estimated biases, larger false alarm rates and smaller critical values. This is explained by the fact that in each of these cases the probability mass of the w-test-statistic over the acceptance region gets smaller. Either the PDF of the w-statistic moves away from the acceptance region when $b/\sigma_{\hat{b}}$ gets larger, or the acceptance region itself gets smaller when either α gets larger, or k gets smaller.

Also note that $\mathsf{P}_{\mathrm{MD}_1}$ (dashed line) gets larger when m gets larger (compare left and right graphs of Figure 2). This is due to the decrease in α_1. When the overall false alarm rate α_m is kept fixed, a smaller α_1 (and thus larger critical value k) needs to be chosen when the number of hypotheses m increases. Increasing k increases the acceptance region and thus P_{MD}.
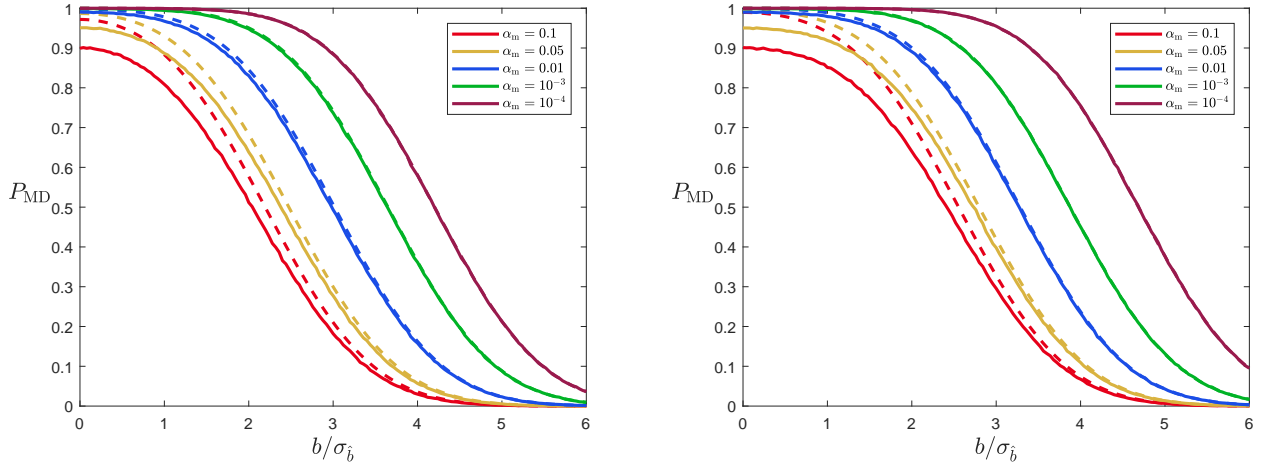

Figure 2: $P_{\mathrm{MD}_{\mathrm{m}}}$ (full line) as a function of the normalized bias size $b/\sigma_{\hat{b}}$, for different values of α_{m}, in comparison with the upperbound $\mathsf{P}_{\mathrm{MD}_1}$ (dashed lines). For the model $A = [1, \ldots, 1]^T \in \mathbb{R}^m$ and $\mathsf{D}(y) = I_m$ with on the left the case $m = 4$ ($\sigma_{\hat{b}} = 1.155\mathrm{m}$) and on the right the case $m = 10$ ($\sigma_{\hat{b}} = 1.054\mathrm{m}$.)

From Figure 2 one can also observe the difference $\mathsf{P}_{\mathrm{MD}_1} - \mathsf{P}_{\mathrm{MD}_{\mathrm{m}}}$. It follows that

$$\mathsf{P}_{\mathrm{MD}_1} - \mathsf{P}_{\mathrm{MD}_{\mathrm{m}}} \downarrow \quad \text{if} \quad b/\sigma_{\hat{b}} \uparrow \quad \text{and/or} \quad \alpha_m \downarrow \quad (k \uparrow) \tag{25}$$

This behaviour can be explained by means of the illustration given in Figure 3 at left. It shows for $m = 2$, the regions $\mathcal{A}_{\alpha_m}$ (dark green), $\mathcal{A}^1_{\alpha_1}$ (light green), and $\mathcal{A}^2_{\alpha_1}$ (light green) in case the two w-test-statistics are independent. When $b/\sigma_{\hat{b}}$ gets larger, the PDF of w_1 under $\mathcal{H}_1$ (illustrated by a circle) translates to the right and the probability mass over the region $\mathcal{A}^1_{\alpha_1}/\mathcal{A}_{\alpha_m}$ (indicated by the yellow area) gets smaller, therefore making the difference $\mathsf{P}_{\mathrm{MD}_1} - \mathsf{P}_{\mathrm{MD}_{\mathrm{m}}}$ smaller too. A similar explanation can be given for when k gets larger, or when α_m gets smaller.

4.2   Identification

We now consider identification. Figure 4 shows for the same model the graphs of $\mathsf{P}_{\mathrm{CI}_{\mathrm{m}}}$ (full line) and $\mathsf{P}_{\mathrm{CI}_1}$ (dashed line) as a function of the normalized bias size $b/\sigma_{\hat{b}}$ for five different values of α_{m} (at left for $m = 4$ and at right for $m = 10$). The general characteristic of the probability of correct identification as a function of bias-size b, precision $\sigma_{\hat{b}}$, false alarm rate α_m or critical value k is as follows:

$$\mathsf{P}_{\mathrm{CI}} \uparrow \quad \text{if} \quad b/\sigma_{\hat{b}} \uparrow \quad \text{and/or} \quad \alpha_m \uparrow \quad (k \downarrow) \tag{26}$$

Larger biases and/or more precisely estimated biases have a larger chance of being correctly identified. Also increasing the false alarm rate, thus reducing the acceptance region, obviously leads to larger probabilities of correct identification.

Also note that $\mathsf{P}_{\mathrm{CI}_1}$ (dashed line) gets smaller when m gets larger (compare left and right graphs of Figure 4). This is due to the decrease in α_1. When the overall false alarm rate α_m is kept fixed, a smaller α_1 (and thus larger critical value k) needs to be chosen when the number of hypotheses m increases. Increasing k increases the acceptance region and thus decreases P_{CI}.
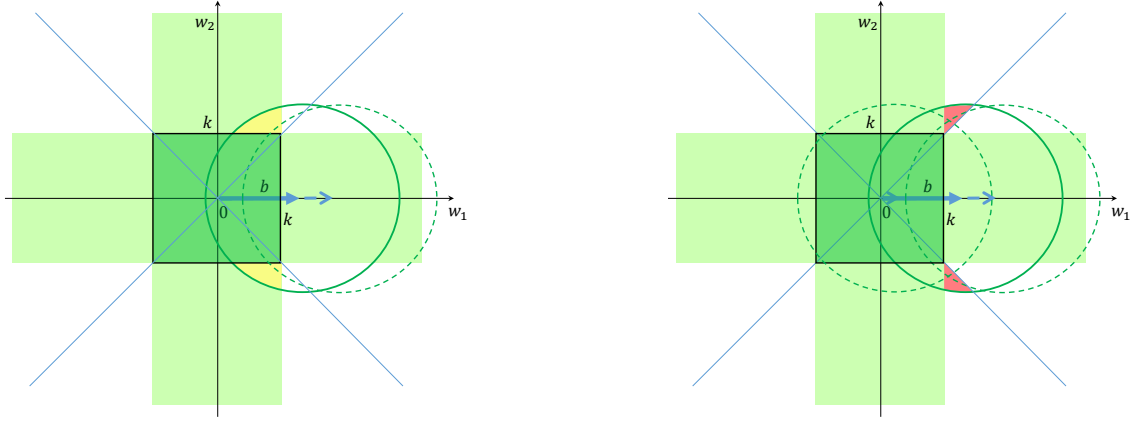
Figure 3: Left: $P_{MD_1} - P_{MD_2}$ (yellow area) gets smaller if $b/\sigma_{\hat{b}}$ gets larger. Right: $P_{CI_1} - P_{CI_m}$ (red area) gets smaller if a small $b/\sigma_{\hat{b}}$ gets even smaller and if a large $b/\sigma_{\hat{b}}$ gets even larger. The joint PDF of w_1 and w_2 is represented by a circle. The two diagonal lines in each figure denote the boundaries of identification between the two hypotheses, cf. (14)

We also see that $P_{CI} \neq 0$ for $b = 0$. In that case P_{CI} reduces to the false alarm rate for single alternative hypothesis testing, while for multiple alternative hypothesis testing it then reduces to a fraction of the false alarm rate, namely the probability by which the particular hypothesis will lead to wrongfull detection. For m hypotheses an approximation of this fraction is $\alpha_1 \approx \frac{1}{m}\alpha_m$.
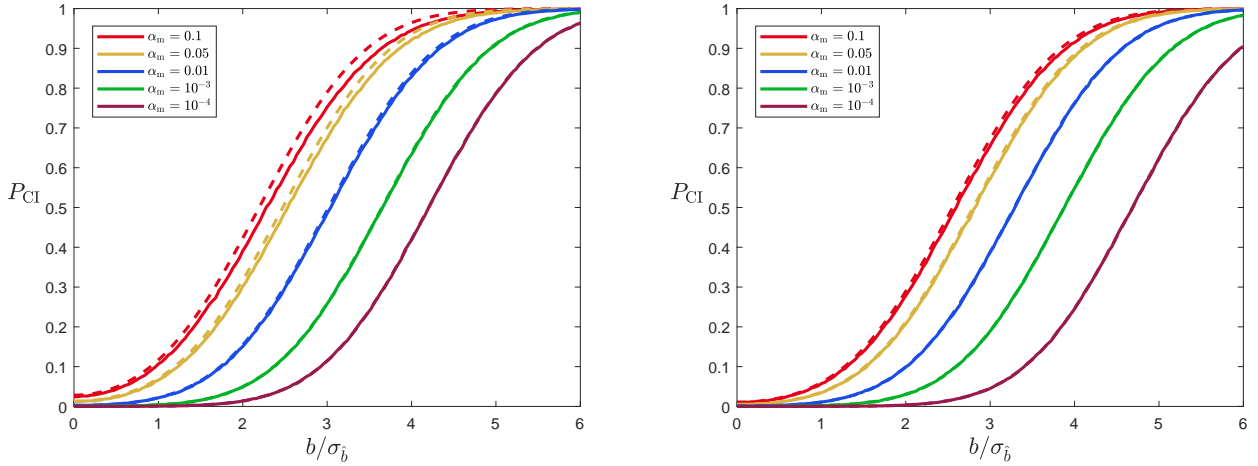

Figure 4: P_{CI_m} (full line) as a function of the normalized bias size $b/\sigma_{\hat{b}}$, for different values of α_m, in comparison with the upperbound P_{CI_1} (dashed lines). For the model $A = [1, \ldots, 1]^T \in \mathbb{R}^m$ and $D(y) = I_m$, with on the left the case $m = 4$ ($\sigma_{\hat{b}} = 1.155$m) and on the right the case $m = 10$ ($\sigma_{\hat{b}} = 1.054$m).

For the difference $P_{CI_1} - P_{CI_m}$ we have

$$P_{CI_1} - P_{CI_m} \downarrow \quad \text{if} \quad b/\sigma_{\hat{b}} \downarrow 0 \quad \text{or} \quad b/\sigma_{\hat{b}} \uparrow \infty \quad \text{and/or} \quad \alpha_m \downarrow \quad (k \uparrow) \tag{27}$$

This behaviour can be explained by means of the illustration given in Figure 3 at right. It shows that the difference $P_{CI_1} - P_{CI_2}$ gets smaller if an already small $b/\sigma_{\hat{b}}$ gets smaller (purple area) or when an already large $b/\sigma_{\hat{b}}$ gets larger (red area). But note that the purple area and red area both get larger when k gets smaller, thus explaining why the difference (27) gets smaller when k gets larger or α_m gets smaller.

4.3 MDBs and MIBs compared

It follows from (16) and (19) that

$$\mathrm{MDB_m}(i) \leq \mathrm{MDB_1}(i) = \mathrm{MIB_1}(i) \leq \mathrm{MIB_m}(i) \tag{28}$$

This clearly shows the impact of testing *multiple* alternative hypotheses. There is no difference between detection and identification in case of a single alternative hypothesis. With an increase in the number of hypotheses, however, the MDBs get smaller, while the MIBs get larger. It may seem strange that the MDBs get smaller by merely increasing the number of hypotheses. This however is a dimensionality effect and due to the increase in intersections of $\mathcal{A}_{\alpha_m} = \cap_{i=1}^m \mathcal{A}_{\alpha_1}^i$ when m gets larger. With this increase in m, also the overall false alarm rate α_m gets larger.

The relation (28) shows that the easy-to-compute $\mathrm{MDB_1}(i)$ can be used as a safe upper bound for its multivariate counterpart. This is *not* the case however for the MIBs. Figure 5 shows for the same earlier example, a comparison between $\mathrm{MDB_m}(1)$, $\mathrm{MDB_1}(1)$ and $\mathrm{MIB_m}(1)$ as a function of α_m for $\mathsf{P_{MD}} = 0.2$. First note that all get smaller if either α_m or m gets larger. This is explained as follows. If the false alarm rate gets larger, then k gets smaller and thus $\mathcal{A}_{\alpha_m}$ and $\mathcal{A}_{\alpha_1}$ get smaller. This implies, when $\mathsf{P_{MD}}$ is kept unchanged, that the MDBs and MIBs get smaller. The impact of m is due to the improved precision with which the bias can be estimated.

Figure 5 also shows that the *difference* between $\mathrm{MDB_m}(i)$ and $\mathrm{MIB_m}(i)$ is small for small α_m, but that it gets more pronounced when α_m gets larger. This is due to the difference between the probabilities of correct detection and correct identification, $\mathsf{P_{CD}} - \mathsf{P_{CI}}$. This difference gets more pronounced if α_m gets larger. If α_m gets larger, then the acceptance region gets smaller as a result of which $\mathsf{P_{CD}}$ increases more than an individual $\mathsf{P_{CI}}$.
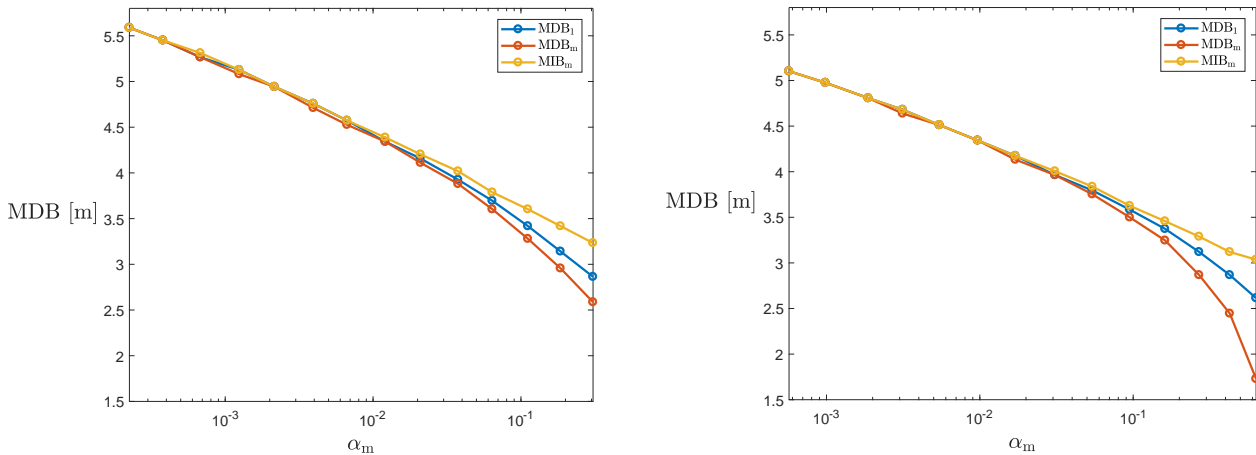

Figure 5: $\mathrm{MDB_m}(1)$, $\mathrm{MDB_1}(1)$ and $\mathrm{MIB_m}(1)$ as functions of α_m for $\mathsf{P_{CD}}$ and $\mathsf{P_{CI}}$ set at $\gamma = 0.8$. For the model $A = [1, \ldots, 1]^T \in \mathbb{R}^m$ and $\mathsf{D}(y) = I_m$, with on the left the case $m = 4$ ($\sigma_{\hat{b}} = 1.155\mathrm{m}$) and on the right the case $m = 10$ ($\sigma_{\hat{b}} = 1.054\mathrm{m}$).

5 Terrestrial and Satellite Positioning

In this section we illustrate by means of two practical surveying examples the significant differences one can experience between detectability and identifiability.

5.1 A planar surveying triangle

Consider the two-dimensional surveying triangle of Figure 6. We assume distance and azimuth measurements. The linearized observation equations for distance l_{ij} and azimuth a_{ij} between points i and j read
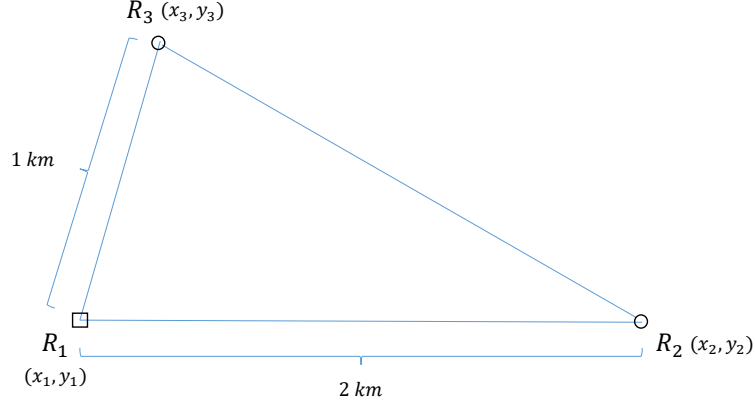
Figure 6: A two-dimensional planar surveying triangle

then

$$
\begin{aligned}
\Delta l_{ij} &= \frac{x_{ij}^0}{l_{ij}} \Delta x_{ij} + \frac{y_{ij}^0}{l_{ij}} \Delta y_{ij} \\
\Delta a_{ij} &= \frac{y_{ij}^0}{l_{ij}^2} \Delta x_{ij} - \frac{x_{ij}^0}{l_{ij}^2} \Delta y_{ij}
\end{aligned}
\tag{29}
$$

in which Δx_{ij} and Δy_{ij} are the increments of the unknown coordinate differences $x_{ij} = x_j - x_i$ and $y_{ij} = y_j - y_i$, respectively. The coordinates of point 1 are assumed to be known. In total, nine observations are available (three distances and six azimuth measurements). Hence, the redundancy equals 5 as there are 4 unknowns (the coordinates of two stations). The observations are assumed normally distributed, uncorrelated and with standard deviations given as

$$
\begin{aligned}
\sigma_{l_{ij}} &= 1\text{cm} \\
\frac{\sigma_{a_{ij}}}{l_{ij}} &= 1\text{cm}
\end{aligned}
\tag{30}
$$

Figure 7 shows the univariate and multivariate probabilities of missed detection ($\mathsf{P}_{\mathrm{MD_m}}, \mathsf{P}_{\mathrm{MD_1}}$) and correct identification ($\mathsf{P}_{\mathrm{CI_m}}, \mathsf{P}_{\mathrm{CI_1}}$), together with their minimal biases ($\mathrm{MDB_m}, \mathrm{MDB_1}, \mathrm{MIB_m}$), as functions of the normalized bias size $b/\sigma_{\hat{b}}$ for five different values of α_{m}. We observe a very similar behaviour as discussed for the averaging model of the previous section. The difference between the actual $\mathsf{P}_{\mathrm{MD_m}}$ and its upperbound $\mathsf{P}_{\mathrm{MD_1}}$ increases with decreasing bias size and increasing false alarm rate α_{m}. Also for identification, the behaviour of the difference between $\mathsf{P}_{\mathrm{CI_m}}$ and its upperbound $\mathsf{P}_{\mathrm{CI_1}}$, is very similar to what we saw earlier. However, note that this difference is now much larger. This can also be seen in the difference between $\mathrm{MDB_m}$ and $\mathrm{MIB_m}$. For instance for $\alpha_m = 0.1$, $\mathrm{MDB_m} = 4.2\text{m}$ and $\mathrm{MIB_m} = 5.3\text{m}$, an over 25% difference. The reason for these larger differences lies in the strength of the underlying model and the available redundancy. In the averaging model of the previous section, redundancy took a larger fraction of the number of observations, namely $\frac{3}{4} = 0.75$ and $\frac{9}{10} = 0.9$, respectively, than in the current surveying example, for which this fraction is $\frac{5}{9} = 0.56$. The lower this fraction is, the more difficult one can expect correct identification to be. In the following section we show a more extreme example of this effect.

5.2 GNSS single point positioning

We consider GNSS single point positioning for a GPS-only scenario and compare it to a GPS+Galileo scenario. The two different receiver-satellite geometries are depicted in Figure 8. In the GPS-only case, the $m \times 4$ design matrix A and pseudorange variance matrix Q_{yy} are structured as

$$
A = \begin{bmatrix} 1 & -u_1^T \\ \vdots & \vdots \\ 1 & -u_m^T \end{bmatrix} , \quad Q_{yy} = \mathrm{diag}(\sigma_1^2, \dots, \sigma_m^2)
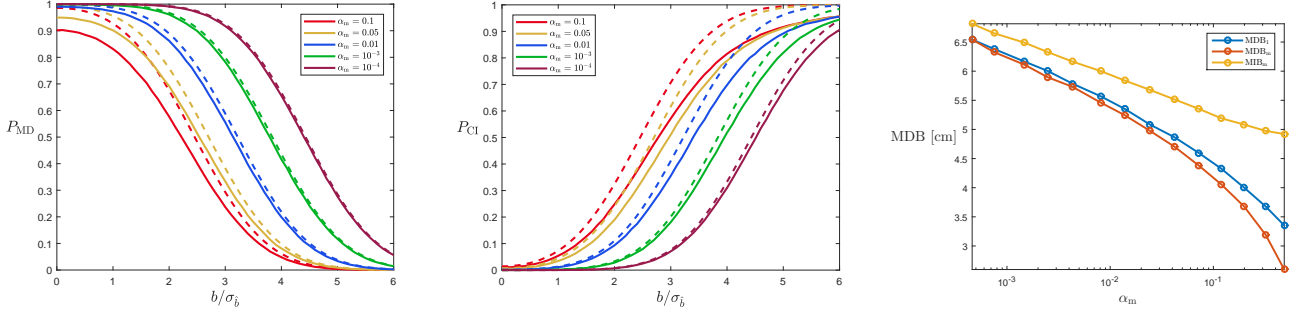\tag{31}
$$

11

Figure 7: Surveying triangle. On the left, P_{MD_m} (full) and P_{MD_1} (dashed), and in the middle, P_{CI_m} (full) and P_{CI_1} (dashed), as functions of the normalized bias $b/\sigma_{\hat{b}}$ (assuming outlier in distance observation l_{23}), for different values of α_m ($\sigma_{\hat{b}} = 1.4\,\mathrm{cm}$). At the right, MDBs and MIB as functions of α_m for $\gamma = 0.8$.
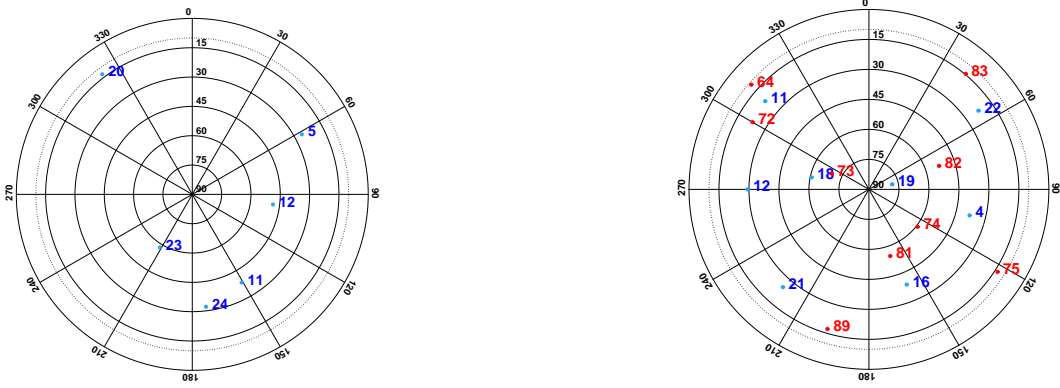

Figure 8: Two GNSS skyplots for Delft, the Netherlands. Left: GPS-only skyplot with $m = 6$ and redundancy $r = 2$; Right: GPS+Galileo skyplot with $m = 17$ and redundancy $r = 12$. GPS PRNs in blue and (planned) Galileo PRNs in red.

with u_i being the ith receiver-satellite unit direction vector. The unknown parameter vector consists of the receiver clock offset and the increments to the three receiver position coordinates. The stochastic model is based on ionosphere-free observations (from dual frequency L1 and L5), with the entries of the diagonal variance matrix constructed according to (Choi et al. 2011; Rippl et al. 2014), see also chapter 7 of (Imparato 2016). For the design matrix of the dual-system GPS+Galileo, an additional column is added to the above design matrix A so as to take care of the inter-system bias or system-specific receiver clock offset.

Figure 9 shows the univariate and multivariate probabilities of missed detection and correct identification, together with their minimal biases. Although again we observe globally a similar behaviour as discussed before, we now clearly see what impact the model strength has on the detection and identification capabilities of the testing procedure. In the weak GPS-only case, the MIB_m exceeds its lowerbound MDB_1 by more than a factor of 4. The differences between the GPS-only case and the GPS+Galileo case are very significant. The same is true for the fraction their redundancy takes of the number of observations, namely $\frac{2}{6} = 0.33$ versus $\frac{12}{17} = 0.71$. These results show how detection and identification fundamentally differ and how the capability of the latter can quickly be compromised when the strength of the model becomes too weak.

6   Conclusions

In this contribution we have discussed and analysed the performance of detection and identification of a DIA-testing procedure. We thereby discriminated between single alternative hypothesis testing and multiple alternative hypotheses testing. For detection, we have Baarda's well-known concept of the Minimal Detectable Bias (MDB). It describes for a specific alternative hypothesis $\mathcal{H}_i$, the smallest size of model error or bias that can be detected with a certain pre-set probability, when employing a detector test with user-
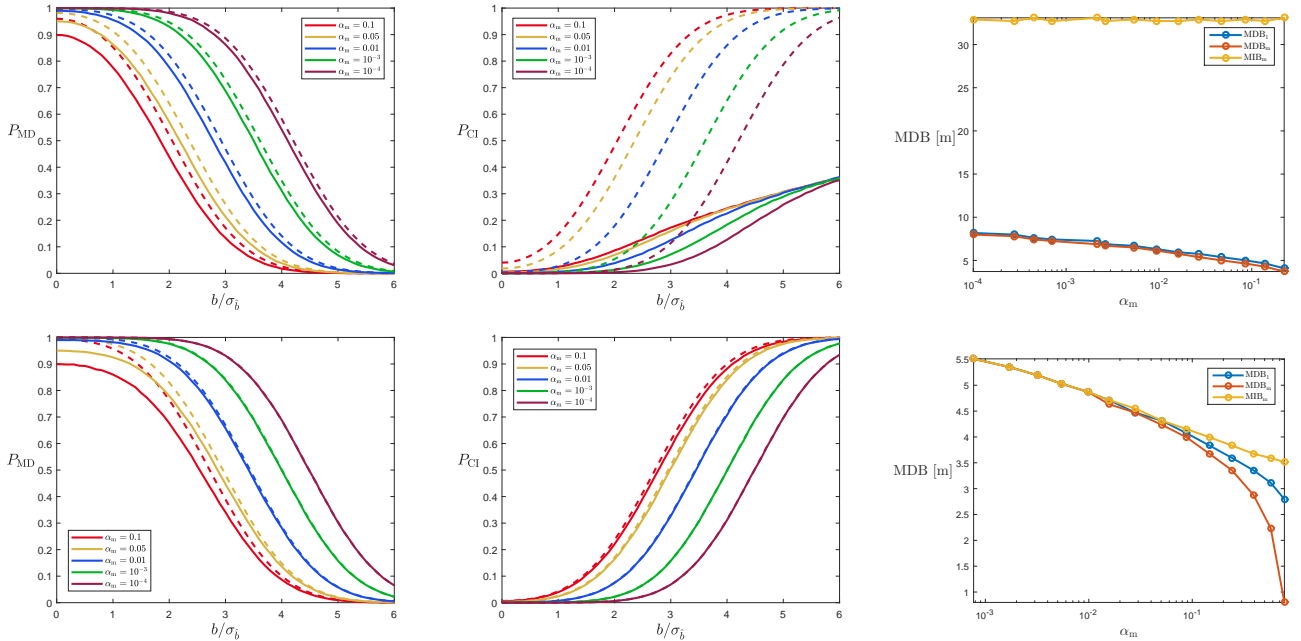
Figure 9: GNSS single point positioning. On the left, P_{MD_m} (full) and P_{MD_1} (dashed), and in the middle, P_{CI_m} (full) and P_{CI_1} (dashed), as functions of the normalized bias $b/\sigma_{\hat{b}}$, for different values of α_m. At the right, MDBs and MIB as functions of α_m for $\gamma = 0.8$. Top row, weak GPS-only case of Figure 8 at left, with outlier in PRN 5 ($\sigma_{\hat{b}} = 1.69$m); bottom row, strong GPS+Galileo case of Figure 8 at right, with outlier in PRN 12 ($\sigma_{\hat{b}} = 1.14$m).

defined false alarm rate. As the MDB depends on the chosen detection test, we pointed out that the MDB for single alternative hypothesis testing may differ from that of multiple alternative hypotheses testing. We analysed the difference and indicated the circumstances under which the easy-to-compute $MDB_1(i)$ can be considered to be a good upper bound for $MDB_m(i)$.

Next to detection, we have identification. The MDBs are Minimal *Detectable* Biases, but not Minimal *Identifiable* Biases. As the probability of correct identification is smaller than that of correct detection, the MIBs are larger than their MDB counterparts,

$$MIB \geq MDB \tag{32}$$

This shows that one has to be careful in using MDBs as MIBs. It is one thing to be able to detect a bias, but an all together other thing is to also be able to identify it. We have made this distinction clear by means of a detailed qualitative analysis and description of the probabilities of missed detection and correct identification as a function of the normalized bias size and overall false alarm rate. This was then further illustrated by means of two practical surveying examples with varying model strength.

As there are, in contrast to the single alternative hypothesis case, no easy-to-compute expressions available for the m-hypotheses MIBs and MDBs, we have also shown how they can be computed by means of Monte Carlo simulation. Such computation first determines the acceptance region from the user-defined overall false alarm rate α_m. Then, for any given bias-range of the hypothesis, the probabilities of correct detection and correct identification are simulated. The so obtained functions are 'inverted' to obtain the required MDB or MIB, respectively, for the given user-defined reference probability.

Future studies on the subject may want to address the numerical aspects of the MDB and MIB computation, in order to improve the efficiency of the Monte Carlo simulations, and explore ways to obtain easy-to-compute tight bounds for the MDBs/MIBs.

References

Alberda, J. E. 1976. "Quality control in surveying." *Chartered Surveyor* 4(2):23–28.

Amiri-Simkooei, A. 2004. "A new method for second order design of geodetic networks: aiming at high reliability." *Survey Review* 37(293):552–560.

Arnold, S.F. 1981. *The theory of linear models and multivariate analysis*. Vol. 2 Wiley, New York.

Baarda, W. 1967. Statistical Concepts in Geodesy. Technical Report 4 Netherlands Geodetic Commission, Publ. on Geodesy, New Series.

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

Bonferroni, C.E. 1936. "Teoria statistica delle classi e calcolo delle probabilità." *Pubblicazioni del R. Instituto Superiore de Scienze Economiche e Commerciali de Firenze* 8:3–62.

Choi, Myungjun, Juan Blanch, Dennis Akos, Liang Heng, Grace Gao, Todd Walter and Per Enge. 2011. Demonstrations of multi-constellation advanced RAIM for vertical guidance using GPS and GLONASS signals. In *Proceedings ION GNSS 2011, Portland, OR.* pp. 3227–3234.

Hofmann-Wellenhof, B., H. Lichtenegger and E. Wasle. 2008. *GNSS: Global Navigation Satellite Systems: GPS, Glonass, Galileo, and More.* Springer, New York.

Imparato, D. 2016. *GNSS based Receiver Autonomous Integrity Monitoring for Aircraft Navigation.* Ph.D. Thesis, TU Delft.

Klein, I., M. T. Matsuoka, M. P. Guzatto and F. G. Nievinski. 2017. "An approach to identify multiple outliers based on sequential likelihood ratio tests." *Survey Review* 49(357):449–457.

Klein, I., M. T. Matsuoka, M. P. Guzatto, S. F. de Souza and M. R. Veronez. 2015. "On evaluation of different methods for quality control of correlated observations." *Survey Review* 47(340):28–35.

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

Lehmann, R. 2010. "Normalized residuals — how large is too large?" *Allgemeine Vermessungsnachrichten* 2.

Lehmann, R. 2011. "Monte Carlo based data snooping with application to a geodetic network." *Journal of Applied Geodesy* 5(3–4):123–134.

Lehmann, R. 2012. "Improved critical values for extreme normalized and studentized residuals in Gauss-Markov models." *Journal of Geodesy* 86:1137–1146.

Lehmann, R and A Voß-Böhme. 2017. "On the statistical power of Baardas outlier test and some alternative." *Journal of Geodetic Science* 7(1):68–78.

Leick, A. 2004. *GPS Satellite Surveying.* John Wiley and Sons.

Lemeshko, B Y and S B Lemeshko. 2005. "Extending the application of Grubbs-type tests in rejecting anomalous measurements." *Meas Tech* 48(6):536–547.

of DGCC, Staff. 1982. "The Delft Approach for the Design and Computation of Geodetic Networks." *Forty Years of Thought..." Anniversary edition on the occasion of the 65th birthday of Professor W. Baarda* 1:202–274.

Prószyński, W. 2015. "Revisiting Baardas concept of minimal detectable bias with regard to outlier identifiability." *Journal of Geodesy* 89(10):993–1003.

Rippl, Markus, Ilaria Martini, Boubeker Belabbas and Michael Meurer. 2014. ARAIM operational performance tested in flight. In *Proceedings of the 2014 International Technical Meeting of The Institute of Navigation (ITM 2014).* pp. 601–615.

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

Tanizaki, H. 2004. *Computational methods in statistics and econometrics.* Marcel Dekker, New York.

Teunissen, P J G. 1986. "Adjusting and testing with the models of the affine and similarity transformation." *Manuscripta Geodaetica* 11:214–225.

Teunissen, P J G. 1989. "Quality control in integrated navigation systems." *IEEE Aerospace and Electronic Systems Magazine* 5(7):35–41.

Teunissen, P J G. 1998. "Minimal detectable biases of GPS data." *Journal of Geodesy* 72(4):236–244.

Teunissen, P J G. 2000. *Testing theory: an introduction.* Delft University Press. Series on Mathematical Geodesy and Positioning.

Teunissen, P. J. G. 2017. "Distributional theory for the DIA method." *J Geod* 91(DOI 10.1007/s00190-017-1045-7):1–22.

Teunissen, P. J. G. and O. Montenbruck (Eds.). 2017. *Handbook of Global Navigation Satellite Systems.* Springer Verlag.

Teunissen, Peter JG, Davide Imparato and Christian CJM Tiberius. 2017. "Does RAIM with correct exclusion produce unbiased positions?" *Sensors* 17(7):1508.

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

Van der Marel, H. and A.J.M. Kosters. 1990. "Statistical Testing and Quality Analysis in 3-D Networks: (part II) Application to GPS. Global Positioning System: An Overview. International Association of Geodesy Symposia, Volume 102." pp. 290–297.

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

Yetkin, M., M. Berber and C. Inal. 2014. "Importance of probability levels for robustness analysis of geodetic networks." *Survey Review* 46(335):132–141.