On the Evaluation of Confidence Levels with Application to GNSS

Zaminpardaz, Safoora; Teunissen, Peter J.G.; Tiberius, Christiaan

DOI
10.33012/2020.17561

Publication date
2020

Document Version
Final published version

Published in
Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2020)

On the Evaluation of Confidence Levels with Application to GNSS

Safoora Zaminpardaz, *School of Science, College of Science, Engineering & Health, RMIT University, Melbourne, Victoria, Australia*
Peter J.G. Teunissen, *GNSS Research Centre, Curtin University, Perth, Western Australia, Australia; Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands*
Christiaan C.J.M. Tiberius, *Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands*

BIOGRAPHIES

Safoora Zaminpardaz is a lecturer in Geospatial Sciences at RMIT University, Melbourne, Australia. Her research interests include multi-GNSS positioning, integrity monitoring and ionosphere sensing.

Peter J.G. Teunissen is Professor of Geodesy and Navigation at Curtin University, Perth, Australia, and Delft University of Technology, Delft, The Netherlands. His current research focuses on multi-GNSS and the modeling of next-generation GNSS for high-precision positioning, navigation and timing applications.

Christiaan C.J.M. Tiberius is an associate professor at Delft University of Technology. He has been involved in GNSS positioning and navigation research since 1991, currently with emphasis on data quality control, satellite-based augmentation systems and precise point positioning.

ABSTRACT

The goal of this contribution is to assess the impact statistical model selection has on confidence levels of parameter estimators in linear(ized) GNSS models. In the processing of observational data, parameter estimation and statistical testing are often combined. A testing procedure is exercised to select the most likely observational model among the hypothesized ones, which is then followed by the estimation of the identified model parameters. The resulting estimator will inherit the uncertainties involved in both estimation and testing which need to be properly taken into account when computing the corresponding confidence level. The approach that is usually followed in practice to determine the confidence level is to compute the probability of the estimator lying in a region around its true value conditioned on the identified hypothesis. Therefore, use is made of the estimator's distribution under the identified hypothesis without regard to the conditioning process that led to the decision of accepting this hypothesis. In this contribution, it will be shown that for a proper computation of the confidence level in combined estimation-testing procedures, the associated probability should be conditioned not only on the identified hypothesis, but also on the testing outcome that led to the decision of accepting this hypothesis. Therefore, use need to be made of the conditional distribution of the estimator. We will provide numerical analysis of confidence levels with and without accounting for conditioning on testing decision using a number of examples in the context of GNSS single point positioning. It will be demonstrated that the customary practice which makes use of unconditional distributions to evaluate the confidence level, may give a too optimistic description of the estimator's quality.

INTRODUCTION

In the processing of observational data, parameter estimation and statistical testing are often combined. Usually, a set of candidate observational models, say $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_k$ with $\mathcal{H}_0$ being the working (null) hypothesis and $\mathcal{H}_i$ for $i = 1, \ldots, k$ the alternative hypotheses, are put forward and a testing procedure is exercised to select the most likely one. The parameters of interest, denoted by x, then get estimated according to the identified model. For example, in GNSS positioning, first GNSS data undergoes statistical testing to detect and identify potential biases like outliers or

 https://doi.org/10.33012/2020.17561

cycle-slips and then, depending on the outcome of testing, the positioning parameters get estimated either based on the working hypothesis where no bias is modelled or based on one of the alternatives in which the identified biases are modelled. The resulting estimator will inherit the uncertainties involved in both estimation and testing [8] which need to be properly taken into account when assessing the estimator's quality.

In this contribution we analyse the impact of statistical model selection on the quality of the estimators generated through combined testing-estimation procedures, and specifically concentrate on the *confidence level*. The approach that is usually followed in practice to determine the confidence level is to compute the probability that the estimator of the unknown parameters, directed by the testing outcome, is inside a region around its true value, *without* taking into account the statistical testing that preceded the estimation [1, 4–6, 9]. Therefore, if $\mathcal{H}_i$ is the identified hypothesis, then use is made of the distribution of the estimator under $\mathcal{H}_i$, say $\hat{x}_i$, without regard to the conditioning process that led to the decision of accepting this hypothesis. Assuming the data to be normally distributed and the observational model to be linear, the estimator $\hat{x}_i$ will also be normally distributed, and thus the confidence level is computed on the basis of the normal distribution.

The approach of using the estimator's unconditional distribution for computing the corresponding confidence levels neglects the statistical testing that preceded the estimation of the model parameters, which will result in an incorrect description of the estimator's quality. In this contribution, it will be shown that for a proper computation of the confidence level in combined estimation-testing procedures, the associated probability should be conditioned on the testing outcome that led to the decision of accepting this hypothesis. Therefore, use need to be made of the conditional distribution of the estimator under $\mathcal{H}_i$ which, as will be shown, is not normal anymore.

This contribution is organized as follows. We first describe the null and alternative hypotheses, highlight the role of the misclosure space partitioning in testing these hypotheses, and present the unknown parameters estimator capturing the contributions from both testing and estimation. Next, assuming that statistical hypothesis testing has done its job properly and identified the correct hypothesis, the actual conditional confidence level is formulated using the estimator's conditional distribution. The various factors that contribute to the difference between the conditional and unconditional confidence levels are identified and discussed. We then demonstrate in graphical form, using a simple observational model with one unknown parameter, both the unconditional and conditional distributions of the parameter estimator, so that the different contributions to the confidence level, as well as the differences between the two approaches, are understood. The confidence level comparison is then continued for a number of examples in the context of GPS single point positioning. Finally, a summary with conclusions are presented.

INTEGRATED TESTING AND ESTIMATION

To illustrate the interaction between testing and estimation, we first specify the null and alternative hypotheses. Let the observational model under the null hypothesis $\mathcal{H}_0$ be given as

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

with $\mathsf{E}(\cdot)$ and $\mathsf{D}(\cdot)$ the expectation and dispersion operator, respectively, $y \in \mathbb{R}^m$ the normally distributed random observable vector, $x \in \mathbb{R}^n$ the estimable unknown parameter vector, $A \in \mathbb{R}^{m\times n}$ the design matrix of $\text{rank}(A) = n$, and $Q_{yy} \in \mathbb{R}^{m\times m}$ the positive-definite variance matrix of y. As, in practice, there are several different sources that can make the observations deviate from $\mathcal{H}_0$-model, multiple alternative hypotheses usually need to be considered to capture the corresponding deviations (the alternative hypotheses here imply extensions to the null hypothesis; they present additional unknown parameters). For example when modeling GNSS data, we may need to take into account alternative hypotheses describing pseudorange outliers and carrier-phase cycle slips. Here, we assume that there are k alternative hypotheses $\mathcal{H}_i$ (for $i = 1,\ldots,k$) of the form

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

for some vector $C_i\delta_i \in \mathbb{R}^m \setminus \{0\}$ such that $[A \;\; C_i]$ is a known matrix of full rank and $\delta_i \in \mathbb{R}^q$ is unknown (representing the outliers and/or cycle slips). We assume that the hypotheses at hand do not occur simultaneously, indicating that only one hypothesis is true at a time.

Hypothesis Testing

To make statistical model validation of $\mathcal{H}_i$ $(i = 0, 1, \ldots, k)$ feasible, it is necessary to have redundant measurements under $\mathcal{H}_0$, i.e. $r = m - n \neq 0$. In that case, an ancillary statistic, known as the misclosure vector $t \in \mathbb{R}^r$, can be obtained as

$$t \;=\; B^T y; \qquad Q_{tt} \;=\; B^T Q_{yy} B \tag{3}$$

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 $y \overset{\mathcal{H}_i}{\sim} \mathcal{N}(Ax + C_i \delta_i,\ Q_{yy})$ for $i = 0, 1, \ldots, k$ and $C_0 \delta_0 = 0$, the misclosure vector is then distributed as

$$t \overset{\mathcal{H}_i}{\sim} \mathcal{N}(\mu_{t_i} = C_{t_i} \delta_i,\ Q_{tt}), \qquad \text{for} \quad i = 0, 1, \ldots, k \tag{4}$$

with $C_{t_i} = B^T C_i$. Therefore, since $\mathsf{E}(t|\mathcal{H}_0) = 0$, the misclosure vector t has a *known* probability density function (PDF) under $\mathcal{H}_0$; it captures all the redundancy in the model. Any statistical model selection mechanism is then driven by the misclosure vector $t \in \mathbb{R}^r$ and its known PDF under $\mathcal{H}_0$. Such model selection mechanism can be established through unambiguously assigning the outcomes of 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 $\mathbb{R}^r$ in $k+1$ subsets $\mathcal{P}_i \subset \mathbb{R}^r$ $(i = 0, 1, \ldots, k)$. The testing procedure is then unambiguously defined as [8]

$$\text{select } \mathcal{H}_i \quad \Longleftrightarrow \quad t \in \mathcal{P}_i, \qquad \text{for} \quad i = 0, 1, \ldots, k \tag{5}$$

Therefore the decisions of the testing procedure are driven by the outcome of the misclosure vector t. If $\mathcal{H}_i$ is true, then the decision is correct if $t \in \mathcal{P}_i$, and wrong if $t \in \mathcal{P}_{j \neq i}$. The probability $\text{P}_{\text{FA}} = \text{P}(t \notin \mathcal{P}_0 | \mathcal{H}_0)$ is called *false alarm* probability, and usually user defined by setting the appropriate size of $\mathcal{P}_0$. Note that in case $r = 1$ (single redundancy), then $\mathcal{P}_1 = \cdots = \mathcal{P}_k = \mathcal{P}_0^c$, implying that no identification can be exercised if $\mathcal{H}_0$ gets rejected.

Parameter Estimation

Statistical model selection is usually followed by the estimation of the parameters of interest x. Assuming that the testing outcome be the selection of the hypothesis $\mathcal{H}_j$ $(t \in \mathcal{P}_j)$, then the parameters get estimated according to the $\mathcal{H}_j$-model; (1) if $j = 0$, and (2) if $j \neq 0$. Therefore, the outcome of testing determines how the parameters get estimated. The probabilistic properties of such an estimation-testing combination can be captured through a unifying framework presented in [8]. As such, the estimator of x is given as

$$\bar{x} \;=\; \sum_{i=0}^{k} \hat{x}_i \, p_i(t) \tag{6}$$

with $p_i(t)$ being the indicator function of region $\mathcal{P}_i$ (cf. 5), i.e. $p_i(t) = 1$ for $t \in \mathcal{P}_i$ and $p_i(t) = 0$ for t elsewhere, and $\hat{x}_i$ the estimator of x under the $\mathcal{H}_i$-model. In practice, with one specific sample of observations y, one of the hypotheses, say $\mathcal{H}_j$, is selected and thus $\hat{x}_j$ will be the numerical final outcome. However, for a correct assessment of the estimator's statistical properties, we need to consider instead $\bar{x}$; if the observation values would have been slightly different namely, due to noise, it could be that statistical testing would have led to deciding that hypothesis $\mathcal{H}_i$ should be used, resulting in $\hat{x}_i$ instead, and all these possibilities need to be properly accounted for through (6).

In this contribution, we make use of Best Linear Unbiased Estimation (BLUE) from which the estimators $\hat{x}_0$ and $\hat{x}_{i \neq 0}$ follow as

$$\hat{x}_0 = A^+ y \qquad , \qquad \hat{x}_i \;=\; \hat{x}_0 \;-\; L_i t \qquad \text{for} \quad i = 1, \ldots k \tag{7}$$

where $A^+ = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ is the BLUE-inverse of A, and $L_i = A^+ C_i C_{t_i}^+$ with $C_{t_i}^+ = (C_{t_i}^T \, Q_{tt}^{-1} C_{t_i})^{-1} C_{t_i}^T \, Q_{tt}^{-1}$ being the BLUE-inverse of $C_{t_i} = B^T C_i$. As $\hat{x}_0$, t, and thus $\hat{x}_i$, are linear functions of the normally-distributed observables y, they are normally distributed as well. Also, from $A^T B = 0$ follows that $\hat{x}_0$ and t are independent from each other. It is however important to note that $\bar{x}$ is *not* normally distributed as it is *non-linearly* dependent on the misclosure t through the indicator functions $p_i(t)$.

CONFIDENCE LEVEL

When providing estimates of the parameters of interest x, it is crucial to provide a quality description of these estimates as well. Therefore, parameter estimates are usually accompanied by their corresponding *confidence levels*. Let us assume that statistical hypothesis testing is successful and *correctly* selects the hypothesis $\mathcal{H}_i$, i.e. $t \in \mathcal{P}_i$. Then, the estimate of x is given by $\hat{x}_i$. With $\mathcal{B}_{x_i} \subset \mathbb{R}^n$ being a x-centered region, the approach that is usually followed in practice to determine the confidence level (CL) is to compute the probability of $\hat{x}_i \in \mathcal{B}_{x_i}$ under the identified hypothesis as follows [1, 4–6, 9]

$$\mathrm{CL}_{x_i} = \mathrm{P}(\hat{x}_i \in \mathcal{B}_{x_i} | \mathcal{H}_i) \qquad \text{for } i = 0, 1, \ldots, k \tag{8}$$

The probability in (8) is computed based on the PDF of $\hat{x}_i$ under the identified hypothesis $\mathcal{H}_i$. Assuming the data to be normally distributed and the observational model to be linear, the estimator $\hat{x}_i$ will also be normally distributed, and thus the confidence level is computed on the basis of normal distribution. This approach, however, neglects the statistical testing that preceded the estimation of the model parameters, resulting in an *incorrect* description of the estimator's quality. That $\hat{x}_i$ is provided as the estimate of the parameters is the result of a testing outcome, namely of having identified $\mathcal{H}_i$ ($t \in \mathcal{P}_i$). Thus, for a proper computation of the confidence level, one has to take this into account as a condition. The correct confidence level is then given by

$$\mathrm{CL}_{x_i} | (t \in \mathcal{P}_i) = \mathrm{P}(\hat{x}_i \in \mathcal{B}_{x_i} | t \in \mathcal{P}_i,\ \mathcal{H}_i) \qquad \text{for } i = 0, 1, \ldots, k \tag{9}$$

The probability in (9) is computed based on the conditional PDF of $\hat{x}_i | t \in \mathcal{P}_i$ under $\mathcal{H}_i$, which is *not* normal anymore.

The difference between $\mathrm{CL}_{x_i} | (t \in \mathcal{P}_i,\ \mathcal{H}_i)$ and CL_{x_i} reads

$$\mathrm{CL}_{x_i} | (t \in \mathcal{P}_i) - \mathrm{CL}_{x_i} = \int_{\mathcal{B}_{x_i}} \left[f_{\hat{x}_i | t \in \mathcal{P}_i}(\theta | t \in \mathcal{P}_i,\ \mathcal{H}_i) - f_{\hat{x}_i}(\theta | \mathcal{H}_i) \right] d\theta \tag{10}$$

with $f_{\hat{x}_i}(\theta|\mathcal{H}_j)$ and $f_{\hat{x}_i | t \in \mathcal{P}_i}(\theta | t \in \mathcal{P}_i,\ \mathcal{H}_i)$ being the PDFs of $\hat{x}_i$ and $\hat{x}_i | t \in \mathcal{P}_i$ under $\mathcal{H}_i$, respectively. The above equation shows, for a given $\mathcal{B}_{x_i}$, that the difference between the correct and incorrect confidence levels depends on the difference between $f_{\hat{x}_i | t \in \mathcal{P}_i}(\theta | t \in \mathcal{P}_i,\ \mathcal{H}_i)$ and $f_{\hat{x}_i}(\theta | \mathcal{H}_j)$. The conditional PDF $f_{\hat{x}_i | t \in \mathcal{P}_i}(\theta | t \in \mathcal{P}_i,\ \mathcal{H}_i)$ is given as [8]

$$f_{\hat{x}_i | t \in \mathcal{P}_i}(\theta | t \in \mathcal{P}_i,\ \mathcal{H}_i) \ = \ \frac{1}{\mathrm{P}(t \in \mathcal{P}_i | \mathcal{H}_i)} \int_{\mathcal{P}_i} f_{\hat{x}_i, t}(\theta, \tau | \mathcal{H}_i)\, d\tau \tag{11}$$

where $f_{\hat{x}_i, t}(\theta, \tau | \mathcal{H}_i)$ is the joint PDF of $\hat{x}_i$ and t under $\mathcal{H}_i$ which given the relations in (7) and the fact that $\hat{x}_0$ and t are independent can be expressed in the PDFs of $\hat{x}_0$ and t as

$$f_{\hat{x}_i, t}(\theta, \tau | \mathcal{H}_i) \ = \ f_{\hat{x}_0}(\theta + L_i \tau | \mathcal{H}_i)\, f_t(\tau | \mathcal{H}_i) \tag{12}$$

Unconditional vs. Conditional CL

Under $\mathcal{H}_0$, as the correlation between $\hat{x}_0$ and t is zero, we have $f_{\hat{x}_0 | t \in \mathcal{P}_i}(\theta | t \in \mathcal{P}_0, \mathcal{H}_0) = f_{\hat{x}_0}(\theta | \mathcal{H}_0)$, thus

$$\mathrm{CL}_{x_0} | (t \in \mathcal{P}_0) \ = \ \mathrm{CL}_{x_0} \tag{13}$$

Under $\mathcal{H}_{i \neq 0}$, the conditioning in $f_{\hat{x}_i | t \in \mathcal{P}_i}(\theta | t \in \mathcal{P}_i,\ \mathcal{H}_i)$ can be nullified, i.e. $f_{\hat{x}_i | t \in \mathcal{P}_i}(\theta | t \in \mathcal{P}_i,\ \mathcal{H}_i) = f_{\hat{x}_i}(\theta | \mathcal{H}_i)$, if the correlation between $\hat{x}_i$ and t would be zero and/or when the event $t \in \mathcal{P}_i$ takes place with 100% certainty, i.e. $\mathrm{P}(t \in \mathcal{P}_i | \mathcal{H}_i) = 1$. As (7) shows, the correlation between $\hat{x}_i$ and t becomes equal to zero if $Q_{tt} = 0$ and/or $L_i = 0$. The former cannot happen as t is a random vector and not deterministic, while the latter can happen if $A^T Q_{yy}^{-1} C_i = 0$ (C_i being Q_{yy}-orthogonal to the columns of A). In addition, if $L_i t$ be far more precise than $\hat{x}_0$, then the correlation between $\hat{x}_i$ and t would be close to zero. The probability of the occurrence of $t \in \mathcal{P}_i$ never becomes identical to 1. However, this probability will approach 1 when $\|\delta_i\| \to \infty$, i.e. the model error or bias gets very big, and /or $\mathcal{P}_i \to \mathbb{R}^r$, i.e. $\mathcal{H}_i$

is almost always accepted. Also, in case the mean of the misclosure lies in $\mathcal{P}_i$, i.e. $C_{t_i}\delta_i \in \mathcal{P}_i$, then $\mathrm{P}(t \in \mathcal{P}_i)$ becomes close to 1 if the PDF of t, i.e. $f_t(\tau|\mathcal{H}_i)$, becomes highly peaked at $C_{t_i}\delta_i \in \mathcal{P}_i$.

Therefore, under $\mathcal{H}_{i\neq 0}$, the circumstances under which the confidence level using unconditional distribution provides a reasonable approximation to the correct one using conditional distribution can be summarized as follows

$$\begin{array}{lll} & A^T Q_{yy}^{-1} C_i = 0 & \Longrightarrow \quad \mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) = \mathrm{CL}_{x_i} \\ & L_i \to 0 & \Longrightarrow \quad \mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) \to \mathrm{CL}_{x_i} \\ \text{For } i \neq 0 \qquad & \|\delta_i\| \to \infty & \Longrightarrow \quad \mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) \to \mathrm{CL}_{x_i} \\ & f_t(\tau|\mathcal{H}_i) \text{ becomes highly peaked at } C_{t_i}\delta_i \in \mathcal{P}_i & \Longrightarrow \quad \mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) \to \mathrm{CL}_{x_i} \\ & \mathcal{P}_i \to \mathbb{R}^r & \Longrightarrow \quad \mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) \to \mathrm{CL}_{x_i} \end{array} \tag{14}$$

Simple Model

Here, using a simple observational model, we provide insight into the confidence levels' characteristics. Consider a linear observational model with only a single alternative hypothesis ($k = 1$) as follows

$$\begin{aligned} \mathcal{H}_0: \quad \mathsf{E}(y) &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} x, & \mathsf{D}(y) &= \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \\ \mathcal{H}_1: \quad \mathsf{E}(y) &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + c_1\delta_1, & \mathsf{D}(y) &= \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \end{aligned} \tag{15}$$

With $y \in \mathbb{R}^2$ being the 2-vector of observations ($m = 2$) and $x \in \mathbb{R}$ being the unknown parameter ($n = 1$), the redundancy of $\mathcal{H}_0$ is $r = 1$, implying that $t \in \mathbb{R}$. The observations are assumed to be uncorrelated and have the same variance σ_y^2. An extra parameter, namely $\delta_1 \in \mathbb{R}$, is introduced in the alternative hypothesis $\mathcal{H}_1$ with respect to the null hypothesis $\mathcal{H}_0$, for instance, to accommodate a bias. To test the validity of $\mathcal{H}_0$, we make use of an overall model test implying the partitioning of the misclosure space $\mathbb{R}$ in two intervals, i.e.

$$\mathcal{P}_0 = [-\sqrt{\kappa_{\mathrm{P_{FA},1}}}\sigma_t,\ \sqrt{\kappa_{\mathrm{P_{FA},1}}}\sigma_t] \quad , \quad \mathcal{P}_1 = \mathcal{P}_0^c \tag{16}$$

where σ_t is the misclosure standard deviation and $\kappa_{\mathrm{P_{FA},1}}$ the ordinate value of the $\chi^2(1,0)$-distribution above which we find an area of size $\mathrm{P_{FA}}$.

We compare $\mathrm{CL}_{x_1}|(t \in \mathcal{P}_1)$ with CL_{x_1} over the x-centered interval

$$\mathcal{B}_{x_1} = [x - \varepsilon\ ,\ x + \varepsilon], \quad \varepsilon \in \mathbb{R}^+ \tag{17}$$

Figure 1 illustrates the PDFs $f_{\hat{x}_1}(\theta|\mathcal{H}_1)$ and $f_{\hat{x}_1|t\in\mathcal{P}_1}(\theta|t \in \mathcal{P}_1,\ \mathcal{H}_1)$ on the top, and the graphs of $\mathrm{CL}_{x_1}|(t \in \mathcal{P}_1) - \mathrm{CL}_{x_1}$ at the bottom, for different sets of values of contributing factors including σ_y, $\mathrm{P_{FA}}$, δ_1 and the following two $c_1 \in \mathbb{R}^2$ vectors

$$\text{Case 1}: \quad c_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad , \qquad \text{Case 2}: \quad c_1 = \begin{bmatrix} -0.8 \\ 1 \end{bmatrix} \tag{18}$$

On the top, in each panel, the normal PDF $f_{\hat{x}_1}(\theta|\mathcal{H}_1)$, which does not depend on the bias value δ_1, is shown in black, and the non-normal PDF $f_{\hat{x}_1|t\in\mathcal{P}_1}(\theta|t \in \mathcal{P}_1,\ \mathcal{H}_1)$, which does depend on the bias value δ_1, is shown for two values of δ_1 in gray and blue. The blue graph, $f_{\hat{x}_1|t\in\mathcal{P}_1}(\theta|t \in \mathcal{P}_1,\ \mathcal{H}_1)$ for $\delta_1 = 5$, almost coincides with the black one, $f_{\hat{x}_1}(\theta|\mathcal{H}_1)$, in panels (b), (c) and (d), and slightly deviates from the black one in panel (a). At the bottom, the corresponding graphs of $\mathrm{CL}_{x_1}|(t \in \mathcal{P}_1) - \mathrm{CL}_{x_1}$ are shown as a function of CL_{x_1} in the same colors. It is observed that when the bias value δ_1 gets sufficiently large, shown in blue, the difference $f_{\hat{x}_1|t\in\mathcal{P}_1}(\theta|t \in \mathcal{P}_1,\ \mathcal{H}_1) - f_{\hat{x}_1}(\theta|\mathcal{H}_1)$ gets small, so does $\mathrm{CL}_{x_1}|(t \in \mathcal{P}_1) - \mathrm{CL}_{x_1}$. However, for relatively small bias values, in gray, there is a significant difference between $\mathrm{CL}_{x_1}|(t \in \mathcal{P}_1)$ and CL_{x_1}. We note that, depending on the bias value δ_1, $\mathrm{CL}_{x_1}|(t \in \mathcal{P}_1)$ could be

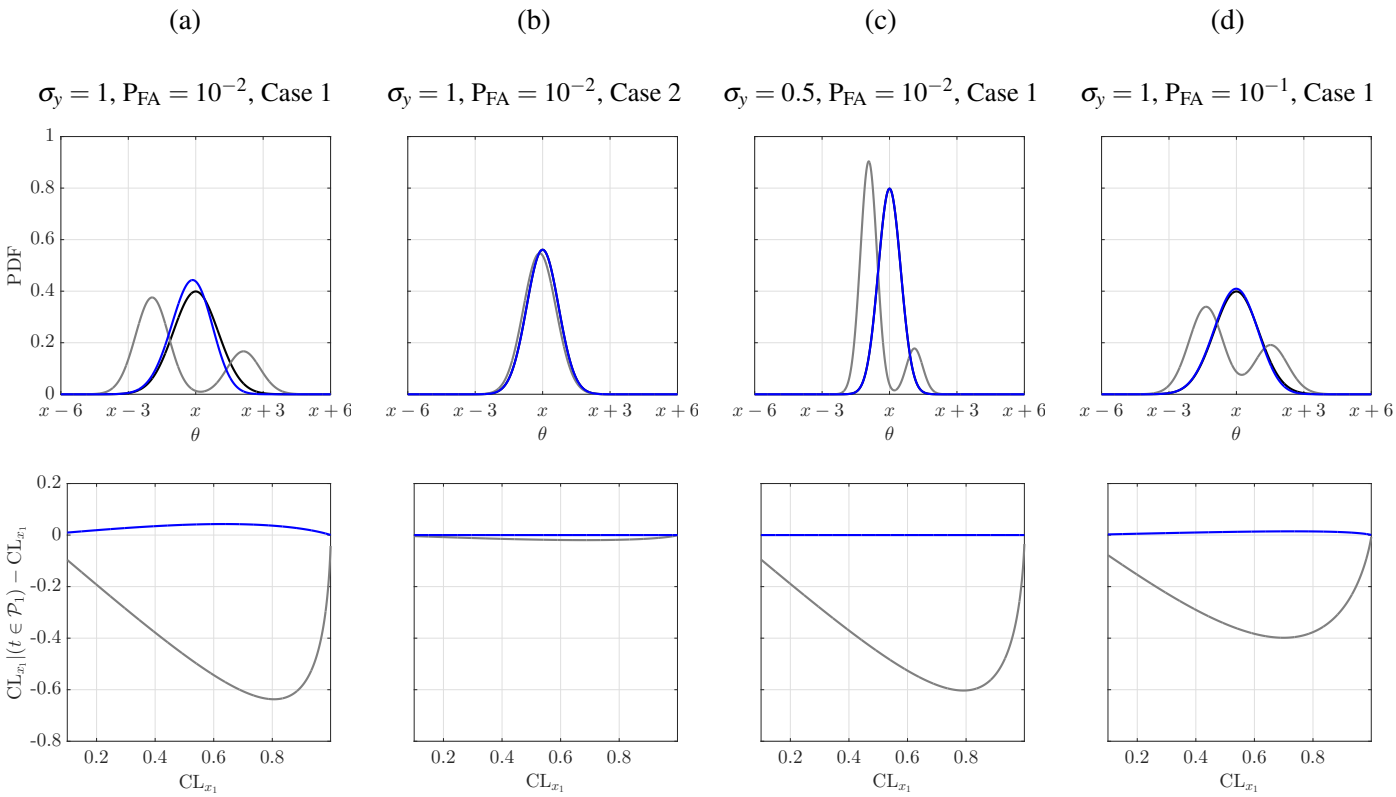

Figure 1: Illustration of the non-normal distribution $f_{\hat{x}_1|t\in\mathcal{P}_1}(\theta|t \in \mathcal{P}_1,\ \mathcal{H}_1)$ in gray ($\delta_1 = 0.2$) and blue ($\delta_1 = 5$) and the normal distribution $f_{\hat{x}_1}(\theta|\mathcal{H}_1)$ in black. The settings are: **(a)** $\sigma_y = 1$, $P_{FA} = 10^{-2}$ and Case 1 in (18); **(b)** $\sigma_y = 1$, $P_{FA} = 10^{-2}$ and Case 2 in (18); **(c)** $\sigma_y = 0.5$, $P_{FA} = 10^{-2}$ and Case 1 in (18); **(d)** $\sigma_y = 1$, $P_{FA} = 10^{-1}$ and Case 1 in (18).

larger or smaller than $CL_{x_1}|(t \in \mathcal{P}_1)$, implying that ignoring the conditioning on the testing decision may result in a *too optimistic* description of the estimator's quality, as demonstrated in Figure 1 at the bottom for the curves in gray, where $CL_{x_1}|(t \in \mathcal{P}_1) < CL_{x_1}$. Comparing columns (a) and (b), the conditional distribution almost coincides with the unconditional one when $c_1 = [-0.8,\ 1]^T$ replaces $c_1 = [0,\ 1]^T$. The value of L_1 (cf. 7) corresponding with 'Case 1' is $L_1 = 0.5$, and with 'Case 2' is $L_1 = 0.1$. Thus, for 'Case 2' compared to 'Case 1', there is a much smaller correlation between $\hat{x}_1$ and t which explains the closeness of the conditional and unconditional PDFs and the small values for $CL_{x_1}|(t \in \mathcal{P}_1) - CL_{x_1}$.

Decreasing σ_y by a specific factor will also decrease $\sigma_{\hat{x}_0}$ (standard deviation of $\hat{x}_0$), σ_t (standard deviation of t) and $\sigma_{\hat{x}_1}$ (standard deviation of $\hat{x}_1$) by the same factor. This explains why the unconditional PDF is more peaked in column (c) in comparison with that in column (a). In column (c), it can also be seen that while $f_{\hat{x}_1|t\in\mathcal{P}_1}(\theta|t \in \mathcal{P}_1,\ \mathcal{H}_1)$ and $f_{\hat{x}_1}(\theta|\mathcal{H}_1)$ coincide for $\delta_1 = 5$, still a significant difference exists between them for $\delta_1 = 0.2$. This can be understood as follows. Taking $B = [-1,\ 1]^T$ (cf. 3), then we have $\sigma_t = \sqrt{2}$ and $\mathcal{P}_0 = [-3.64,\ 3.64]$ in column (a), while $\sigma_t = 1/\sqrt{2}$ and $\mathcal{P}_0 = [-1.82,\ 1.82]$ in column (c). Therefore, from column (a) to (c), the peakedness of $f_t(\tau|\mathcal{H}_1)$ increases at $c_{t_1}\delta_1$, and also $\mathcal{P}_1$ gets larger. In case $c_{t_1}\delta_1 \in \mathcal{P}_1$, then both of the mentioned changes will increase $P(t \in \mathcal{P}_1|\mathcal{H}_1)$. Otherwise, these two changes will have opposite impacts on $P(t \in \mathcal{P}_1|\mathcal{H}_1)$ and thus this probability may decrease or increase. With $c_{t_1}\delta_1 = \delta_1$, in both columns, $\delta_1 = 5$ (blue graph) lies in $\mathcal{P}_1$ whereas $\delta_1 = 0.2$ (gray) graph lies in $\mathcal{P}_0$. In column (d), there is an increase in P_{FA} compared to column (a), which will result in $\mathcal{P}_1$ getting expanded, which leads to $P(t \in \mathcal{P}_1|\mathcal{H}_1)$ getting larger. This explains the smaller differences between $f_{\hat{x}_1}(\theta|\mathcal{H}_1)$ and $f_{\hat{x}_1|t\in\mathcal{P}_1}(\theta|t \in \mathcal{P}_1,\ \mathcal{H}_1)$, and also between $CL_{x_1}|(t \in \mathcal{P}_1)$ and CL_{x_1}, in column (d) in comparison to column (a).

NUMERICAL ANALYSIS

In this section, we compare the incorrect unconditional confidence levels (8) with the correct conditional ones (9) in the context of GPS single-point positioning (SPP) application. Assuming that m GPS satellites are tracked by a single receiver, the SPP observational model under $\mathcal{H}_o$ reads

$$\mathcal{H}_0 \quad : \quad \mathsf{E}(y) = \underbrace{[G \quad e_m]}_{A} \begin{bmatrix} x \\ dt \end{bmatrix} \qquad , \qquad Q_{yy} = \sigma_y^2 I_m \tag{19}$$

where $y \in \mathbb{R}^m$ is the vector of pseudo-range observables, $G = [-u_1^T, \ldots, -u_m^T]^T \in \mathbb{R}^{m\times 3}$ is the geometry matrix containing the receiver-satellite unit direction vectors u_i as its rows, $e_m \in \mathbb{R}^m$ is the vector of ones and $I_m \in \mathbb{R}^{m\times m}$ is the identity matrix. The unknown receiver coordinate components and clock error are, respectively, denoted by $x \in \mathbb{R}^3$ and $dt \in \mathbb{R}$. The observables are assumed to be all uncorrelated and of the same standard deviation σ_y. At this stage, in order to simplify our analysis, we do not consider a satellite elevation-dependent variance matrix.

As alternative hypotheses, we consider those describing outliers in individual observations. Here we restrict ourselves to the case of one outlier at a time. In that case there are as many alternative hypotheses as there are observations, i.e. $k = m$. Therefore, the observational model under $\mathcal{H}_0$ and $\mathcal{H}_i$ is given as

$$\mathcal{H}_i \quad : \quad \mathsf{E}(y) = [G \quad e_m] \begin{bmatrix} x \\ dt \end{bmatrix} + c_i \delta_i \qquad , \qquad Q_{yy} = \sigma_y^2 I_m \qquad \text{for } \; i = 1, \ldots, m \tag{20}$$

with $c_i \in \mathbb{R}^m$ the canonical unit vector having one as its i^{th} entry and zeros elsewhere, and $\delta_i \in \mathbb{R}$ the scalar bias. Note that $[A \;\; c_i]$ is a known matrix of full rank.

Hereafter we concentrate on receiver coordinates x as our parameters of interest and analyse the difference between $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i)$ and CL_{x_i} for $i = 1, \ldots, m$. For SPP application, this difference is driven by the misclosure space partitioning regions ($\mathcal{P}_j$ for $j = 0, 1, \ldots, m$), the satellites geometry G, pseudo-range precision σ_y, the confidence region $\mathcal{B}_{x_i}$, the true hypothesis and δ_i. In the following, for two satellite geometries, we illustrate $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ ($i = 1, \ldots, m$) as a function of the mentioned contributing factors, with $\mathcal{B}_{x_i}$ being defined as

$$\mathcal{B}_{x_i} = \{\theta \in \mathbb{R}^3 | \; \|\theta - x\|^2_{Q_{\hat{x}_i\hat{x}_i}} \leq \kappa_{\alpha,3}\} \qquad \text{for } \; i = 1, \ldots, m \tag{21}$$

in which $\|.\|^2_{Q_{\hat{x}_i\hat{x}_i}} = (.)^T Q^{-1}_{\hat{x}_i\hat{x}_i}(.)$ and $\kappa_{\alpha,3}$ is the ordinate value of the $\chi^2(3,0)$-distribution above which we find an area of size α. With $Q_{\hat{x}_i\hat{x}_i}$ being the variance matrix of $\hat{x}_i$, $\mathcal{B}_{x_i}$ is the $100(1-\alpha)\%$ confidence ellipsoid of $\hat{x}_i$.

Testing Procedure

With $\mathcal{H}_0$ in (19) and $\mathcal{H}_i$ ($i = 1, \ldots, m$) in (20), our testing strategy comprises two steps of detection and identification, respectively, and is specified as follows

- *Detection*: The validity of the null hypothesis is checked through an overall model test (the redundancy needs to be $r > 0$). The null hypothesis $\mathcal{H}_0$ is accepted if $t \in \mathcal{P}_0$ with

$$\mathcal{P}_0 \quad = \quad \left\{ t \in \mathbb{R}^r \,\middle|\, \|t\|^2_{Q_{tt}} \leq \kappa_{\mathrm{P_{FA}},r} \right\} \tag{22}$$

  in which $\|.\|^2_{Q_{tt}} = (.)^T Q_{tt}^{-1}(.)$ and $\kappa_{\mathrm{P_{FA}},r}$ is the ordinate value of the $\chi^2(r,0)$-distribution above which we find an area of size $\mathrm{P_{FA}}$.

- *Identification*: If $\mathcal{H}_0$ is rejected in the detection step, a search is carried out among the specified alternatives $\mathcal{H}_i$ ($i = 1, \ldots, m$) to select the potential source of model error (note that with $r = 1$ identification is not possible). The alternative hypothesis $\mathcal{H}_{i\neq 0}$ is selected if $t \in \mathcal{P}_{i\neq 0}$ with

$$\mathcal{P}_i \quad = \quad \left\{ t \in \mathbb{R}^r \setminus \mathcal{P}_0 \,\middle|\, |w_i| = \max_{j\in\{1,\ldots,m\}} |w_j| \right\}, \quad i = 1, \ldots, m \tag{23}$$

in which w_i is Baarda's test statistic computed as [2, 7]

$$w_i \;=\; \frac{c_{t_i}^T Q_{tt}^{-1} t}{\sqrt{c_{t_i}^T Q_{tt}^{-1} c_{t_i}}}, \quad i = 1, \ldots, m \tag{24}$$

Example 1: Figure 2

Figure 2 (left) shows the skyplot of six satellites for which six alternative hypotheses ($k = 6$) of the form of (20) can be considered. With $m = 6$ and $n = 4$, the redundancy under $\mathcal{H}_0$ is $r = 2$, hence $t \in \mathbb{R}^2$ (so that we can conveniently visualize the misclosure space). Without loss of generality, we choose B (cf. 3) in such a way that $Q_{tt} = I_2$. Figure 2 (right) shows the corresponding misclosure space ($\mathbb{R}^2$) partitioning in seven regions $\mathcal{P}_0$ (cf. 22) and $\mathcal{P}_i$ for $i = 1, \ldots, 6$ (cf. 23). The shown vectors $\bar{c}_i$ are the unit vectors $\bar{c}_i = c_{t_i}/\|c_{t_i}\|$. Given (23) and (24), region $\mathcal{P}_{i\neq 0}$ contains samples of the misclosure $t \in \mathbb{R}^r \setminus \mathcal{P}_0$ which have the largest projection onto the unit vector $\bar{c}_i$ compared to other unit vectors $\bar{c}_{j\neq i}$ ($j = 1 \ldots, 6$). As $\mathsf{E}(t|\mathcal{H}_i) = \|c_{t_i}\|\delta_i \bar{c}_i$ and $Q_{tt} = I_2$, $\|c_{t_i}\|$ is the indicator of minimal detectable bias (MDB) under $\mathcal{H}_i$ [3, 7]; the larger the value of $\|c_{t_i}\|$, the smaller the MDB, and thus the better the detectability under $\mathcal{H}_i$. For the model at hand, we have

$$\|c_{t_1}\| \approx 0.68,\ \|c_{t_2}\| \approx 0.72,\ \|c_{t_3}\| \approx 0.55,\ \|c_{t_4}\| \approx 0.53,\ \|c_{t_5}\| \approx 0.64,\ \|c_{t_6}\| \approx 0.12 \tag{25}$$

implying that the detectability under $\mathcal{H}_6$ is much poorer compared to other alternatives. In other words, for the same bias value δ_i, $\mathrm{P}(t \notin \mathcal{P}_0|\mathcal{H}_6)$ is much smaller than $\mathrm{P}(t \notin \mathcal{P}_0|\mathcal{H}_{i\neq 0,6})$.

Figure 3 shows the corresponding colormaps of $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ for $i = 1, \ldots, 6$ as a function of CL_{x_i} horizontally and δ_i vertically, where each column represents an alternative hypothesis. From top to bottom, the underlying settings are $\sigma_y = 1\mathrm{m},\ \mathrm{P_{FA}} = 10^{-1}$, $\sigma_y = 1\mathrm{m},\ \mathrm{P_{FA}} = 10^{-2}$ and $\sigma_y = 0.5\mathrm{m},\ \mathrm{P_{FA}} = 10^{-1}$, respectively. It is observed that depending on the underlying settings and the bias value δ_i, the conditional confidence level could be much *lower* than the unconditional one. For example, under $\mathcal{H}_6$ when $\sigma_y = 1\mathrm{m}$, $\mathrm{P_{FA}} = 10^{-2}$, $\delta_6 = 5\mathrm{m}$ and $\mathrm{CL}_{x_6} = 0.95$, the conditional confidence level $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_6)$ is smaller than CL_{x_6} by an amount of 0.65, implying that CL_{x_6} is *too optimistic* by almost a factor of 3.

From Figure 3, we note that the difference $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ for a given δ_i and $\mathrm{CL_{x_i}}$ shows larger magnitudes for $\mathcal{H}_6$ compared to the other alternatives. To explain this behavior, we consider (7) which describes the link between $\hat{x}_i$ and t, established through L_i. For the observational model at hand, in which the redundancy is $r = 2$ and also $x \in \mathbb{R}^3$, L_i is a 3×2 matrix given as

$$L_i \;=\; \frac{1}{\|c_{t_i}\|} Q_{\hat{x}_0\hat{x}_0}(\bar{u} - u_i)\, \bar{c}_i^T \tag{26}$$

with $Q_{\hat{x}_0\hat{x}_0}$ the variance matrix of $\hat{x}_0$ and $\bar{u} = \frac{1}{6}\sum_{j=1}^{6} u_j$ the average receiver-satellite unit direction vector. Table 1 gives the components of the 3-vector $\frac{1}{\|c_{t_i}\|} Q_{\hat{x}_0\hat{x}_0}(\bar{u} - u_i)$ for all $i = 1, \ldots, 6$. As can be seen, the components of L_6 are almost 6 times larger than those of $L_{i\neq 6}$. This has two implications: 1) the correlation between $\hat{x}_6$ and t is by far larger than the correlation between $\hat{x}_{i\neq 6}$ and t; 2) $\hat{x}_6$ has much poorer precision compared to $\hat{x}_{i\neq 6}$. In addition, for $i \neq 0$, since $\mathsf{E}(t|\mathcal{H}_i) = \|c_{t_i}\|\delta_i \bar{c}_i$ and $Q_{tt} = I_2$, the probability $\mathrm{P}(t \in \mathcal{P}_i|\mathcal{H}_i)$ for a given bias value δ_i is driven by several factors including the region $\mathcal{P}_i$ and the magnitude of $\|c_{t_i}\|$. The larger the region $\mathcal{P}_i$ and the absolute value of $\mathsf{E}(t|\mathcal{H}_i)$, the larger the probability $\mathrm{P}(t \in \mathcal{P}_i|\mathcal{H}_i)$. Given that $\|c_{t_i}\|$ (cf. 25) and $\mathcal{P}_i$ (Figure 2) are much smaller for $i = 6$ than $i \neq 6$, $\mathrm{P}(t \in \mathcal{P}_6|\mathcal{H}_6)$ is also smaller than $\mathrm{P}(t \in \mathcal{P}_{i\neq 0,6}|\mathcal{H}_{i\neq 0,6})$ for a given bias value δ_i. All these factors lead to relatively large differences between $f_{\hat{x}_6|t\in\mathcal{P}_6}(\theta|t \in \mathcal{P}_6,\ \mathcal{H}_6)$ and $f_{\hat{x}_6}(\theta|\mathcal{H}_6)$, thus between $\mathrm{CL}_{x_6}|(t \in \mathcal{P}_6)$ and CL_{x_6}.

For a given value of δ_i, the absolute value of $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ as a function of CL_{x_i} first increases and then decreases to zero. To understand this behavior, we note that increasing CL_{x_i} is the result of expanding $\mathcal{B}_{x_i}$ (cf. 21) or

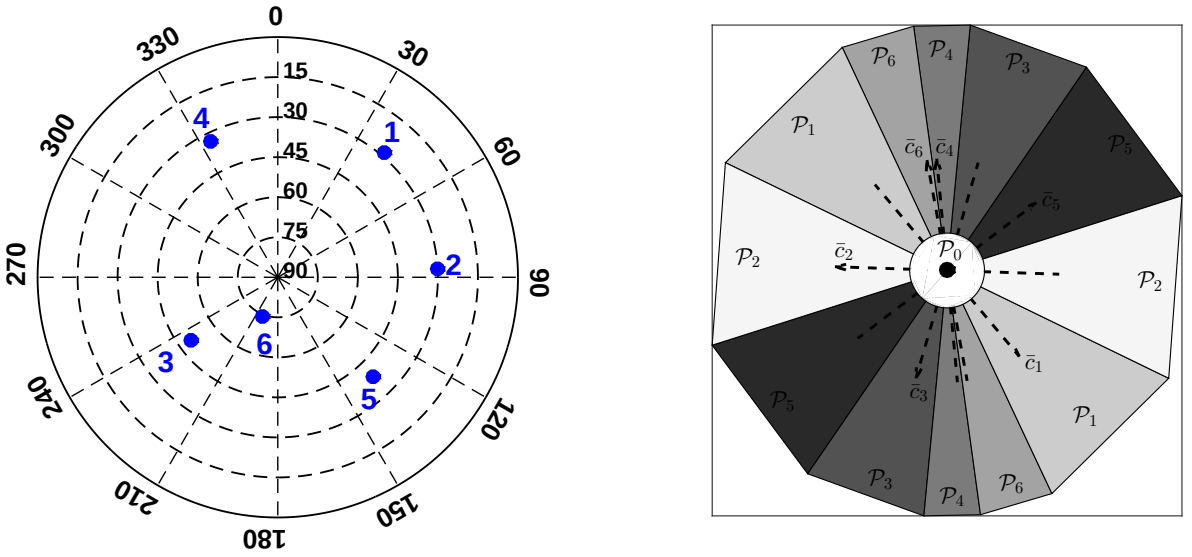

Figure 2: [Left] Skyplot of six satellites. [Right] The corresponding datasnooping misclosure space partitioning with $\mathcal{P}_0$ and $\mathcal{P}_i$ $(i = 1,\ldots,6)$ formulated in (22) and (23), respectively, for $Q_{tt} = I_2$.

Table 1: Evaluation of (26) for the satellite geometry in Figure 2 and the hypotheses given in (19) and (20), assuming $\sigma_y = 1$m.

$\frac{1}{\|c_t\|} Q_{\hat{x}_0\hat{x}_0}(\bar{u} - u_i)$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\mathcal{H}_4$	$\mathcal{H}_5$	$\mathcal{H}_6$
first component	7.11	5.95	6.91	8.80	5.86	40.24
second component	6.53	6.24	6.28	6.90	6.54	38.16
third component	-32.92	-30.07	-37.37	-40.04	-32.80	-211.82

alternatively decreasing α. Given (8) and (9), we have

$$\begin{aligned} \alpha \to 1 \quad &\Longrightarrow \quad \begin{cases} \mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) & \to \ 0 \\ \mathrm{CL}_{x_i} & \to \ 0 \end{cases} \\ \alpha \to 0 \quad &\Longrightarrow \quad \begin{cases} \mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) & \to \ 1 \\ \mathrm{CL}_{x_i} & \to \ 1 \end{cases} \end{aligned} \tag{27}$$

As Figure 1 shows, depending on δ_i, either $f_{\hat{x}_i|t\in\mathcal{P}_i}(\theta|t \in \mathcal{P}_i,\ \mathcal{H}_i)$ or $f_{\hat{x}_i}(\theta|\mathcal{H}_i)$ is more peaked around x, implying that one of the confidence levels increases more rapidly as α decreases. This, together with (27) and the fact that the confidence levels $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i)$ and CL_{x_i} are continuous functions of α, results in an increasing and then decreasing behavior for the absolute value of $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ as a function of CL_{x_i}. We note that, for δ_i larger than a particular value varying from alternative to alternative, the dependency of $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ on CL_{x_i} almost vanishes and is almost equal to zero. This is due to the fact that when $\delta_i \to \infty$, we have $\mathrm{P}(t \in \mathcal{P}_i|\mathcal{H}_i) \to 1$ and thus $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) \to \mathrm{CL}_{x_i}$.

Comparing the panels on the first and second rows in Figure 3, it can be seen that decreasing P_{FA} from 10^{-1} to 10^{-2} makes the absolute value of $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ larger for a given δ_i. That is because when P_{FA} decreases, the region $\mathcal{P}_0$ gets expanded, while the other regions $\mathcal{P}_{i\neq 0}$ shrink. Therefore, as P_{FA} decreases, $\mathrm{P}(t \in \mathcal{P}_i|\mathcal{H}_i)$ for a given value of δ_i decreases resulting in larger differences between the conditional and unconditional confidence levels. Decreasing

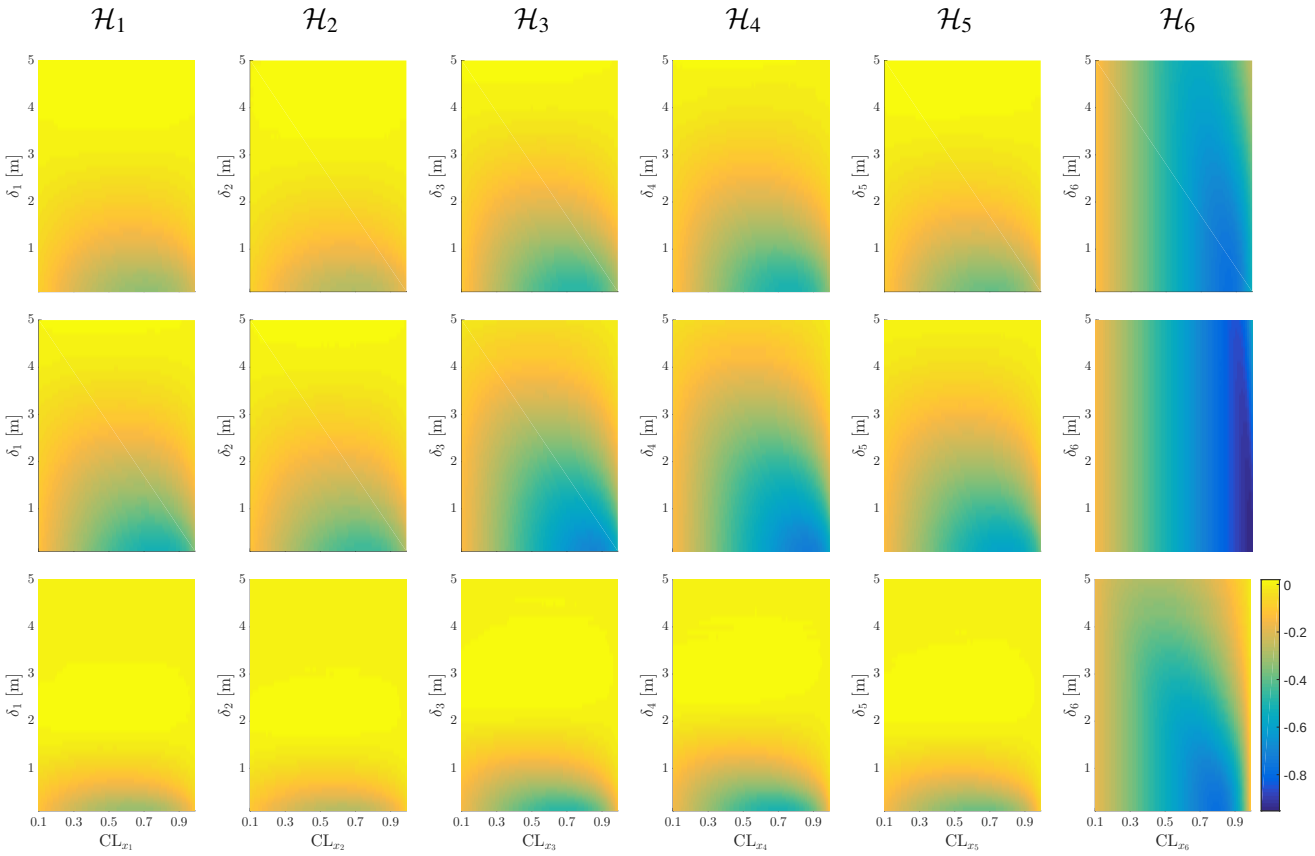

Figure 3: Colormaps of confidence level differences $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ as a function of CL_{x_i} horizontally and δ_i vertically, corresponding with the satellite geometry and misclosure space partitioning in Figure 2. The columns from left to right show the results under $\mathcal{H}_i$ for $i = 1,\ldots,6$. The settings from top to bottom are $\sigma_y = 1\mathrm{m}$, $\mathrm{P_{FA}} = 10^{-1}$, $\sigma_y = 1\mathrm{m}$, $\mathrm{P_{FA}} = 10^{-2}$ and $\sigma_y = 0.5\mathrm{m}$, $\mathrm{P_{FA}} = 10^{-1}$.

σ_y from 1m to 0.5m, the results on the first row change to those on the third row. As was explained before with the simple model, decreasing σ_y by a specific factor will also decrease $Q_{\hat{x}_0\hat{x}_0}$, Q_{tt} and $Q_{\hat{x}_i\hat{x}_i}$ by the same factor, making $f_{\hat{x}_i}(\theta|\mathcal{H}_i)$ more peaked around x, and $f_t(\tau|\mathcal{H}_i)$ more peaked around $c_{t_i}\delta_i$. It also makes $\mathcal{P}_0$ shrink while the regions $\mathcal{P}_{i\neq 0}$ expand. Therefore, if $c_{t_i}\delta_i$ lies in the expanded $\mathcal{P}_i$, then $\mathrm{P}(t \in \mathcal{P}_i|\mathcal{H}_i)$ increases, giving rise to smaller differences between $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i)$ and CL_{x_i}. However, if $c_{t_i}\delta_i$ lies in the shrunk $\mathcal{P}_0$, then $f_t(\tau|\mathcal{H}_i)$ getting more peaked and $\mathcal{P}_i$ getting expanded have opposite impacts on $\mathrm{P}(t \in \mathcal{P}_i|\mathcal{H}_i)$, and thus this probability may decrease or increase.

Example 2: Figure 4

To see the impact of satellite geometry on the confidence levels, Figure 5 presents the same type of information as Figure 3, but for the geometry of six satellites shown in Figure 4. The difference in confidence levels (conditional minus unconditional) in general behaves similar to the earlier example. For this example, again we note the 'negative' values for $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ under different hypotheses. For example, under $\mathcal{H}_3$ when $\sigma_y = 1\mathrm{m}$, $\mathrm{P_{FA}} = 10^{-2}$, $\delta_3 = 1\mathrm{m}$ and $\mathrm{CL}_{x_3} = 0.90$, the conditional confidence level is $\mathrm{CL}_{x_3}|(t \in \mathcal{P}_i) = 0.02$, implying that CL_{x_3} is *too optimistic* by almost a factor of 40.

SUMMARY AND CONCLUSION

In this contribution we analysed the impact of statistical model selection on the quality of the estimators generated through combined testing-estimation procedures. As in such procedures it is the testing outcome which determines

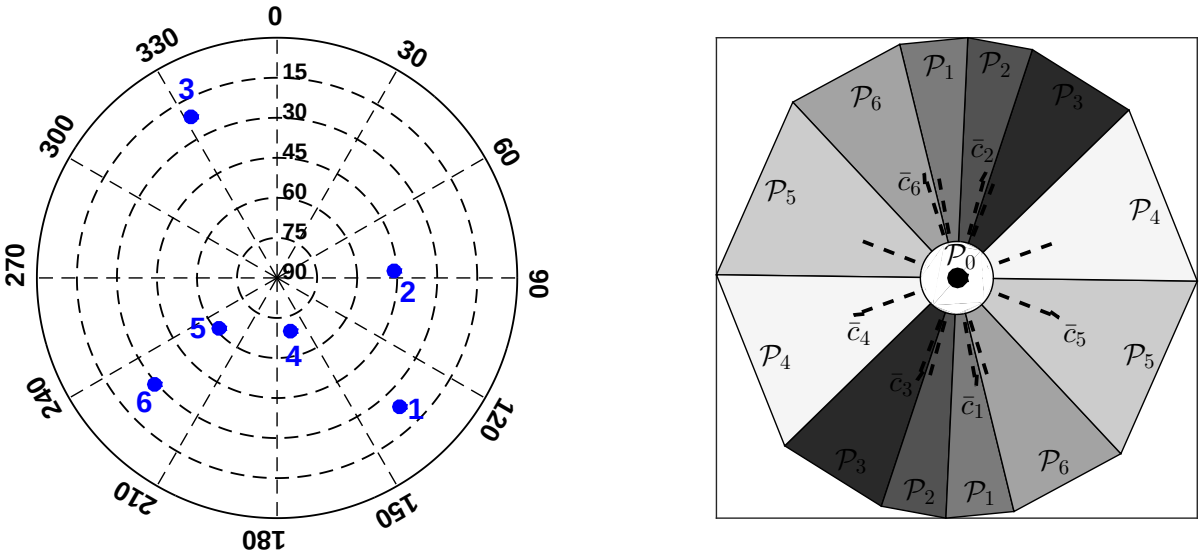

Figure 4: [Left] Skyplot of six satellites. [Right] The corresponding datasnooping misclosure space partitioning with $\mathcal{P}_0$ and $\mathcal{P}_i$ ($i = 1,\ldots,6$) formulated in (22) and (23), respectively, for $Q_{tt} = I_2$.

how the parameters get estimated, the eventual estimator's quality is driven by the characteristics of not only estimation but testing as well. As an important estimator's quality indicator, we considered the *confidence level* which is often computed as the probability of the estimator lying in a region around its true value *without* taking into account the statistical testing that preceded the estimation. Therefore, once one of the hypothesized models is identified through the testing procedure, the customary approach followed in practice to determine the confidence level is to make use of the estimator's distribution under the identified hypothesis without regard to the conditioning process that led to the decision to accept this hypothesis. It was demonstrated that this conditioning process needs to be taken into account for a proper computation of the confidence level by using the *conditional* distribution of the estimator under the identified hypothesis conditioned on the testing outcome that led to the selection of this hypothesis.

Assuming that statistical hypothesis testing has done its job properly and identified the correct hypothesis, say $\mathcal{H}_i$, we formulated the actual conditional confidence level using the concept of misclosure space partitioning. It was shown that under the null hypothesis, there would be no difference between the conditional and unconditional confidence levels. Under alternative hypotheses however, the actual confidence level would be different from the unconditional one. The factors contributing to this difference were identified and discussed.

Considering a binary-hypothesis testing applied to a simple observational model, the unconditional and conditional distribution of the parameter estimator were demonstrated under the alternative hypothesis. It was shown that, with normally distributed observables and linear models, the distributions of the estimators conditioned on the testing outcome turn out to be no longer normal. For this simple observational model, it was shown that the actual confidence level, computed based on the non-normal estimator's conditional distribution, could be much smaller than the unconditional confidence level which is computed on the normal estimator's unconditional distribution. Therefore the customary approach may provide a too optimistic description of the estimator's quality. We further continued confidence level comparison using two examples in the context of GPS single point positioning. In our analyses, we considered the contribution of several various factors to the difference between conditional and unconditional confidence levels including satellite geometry, testing procedure, pseudo-range precision, confidence region and bias value under the identified hypotheses. These examples corroborated the previous finding that the unconditional confidence level can be much larger than its conditional version, thus providing a too optimistic description of the quality of the estimator.

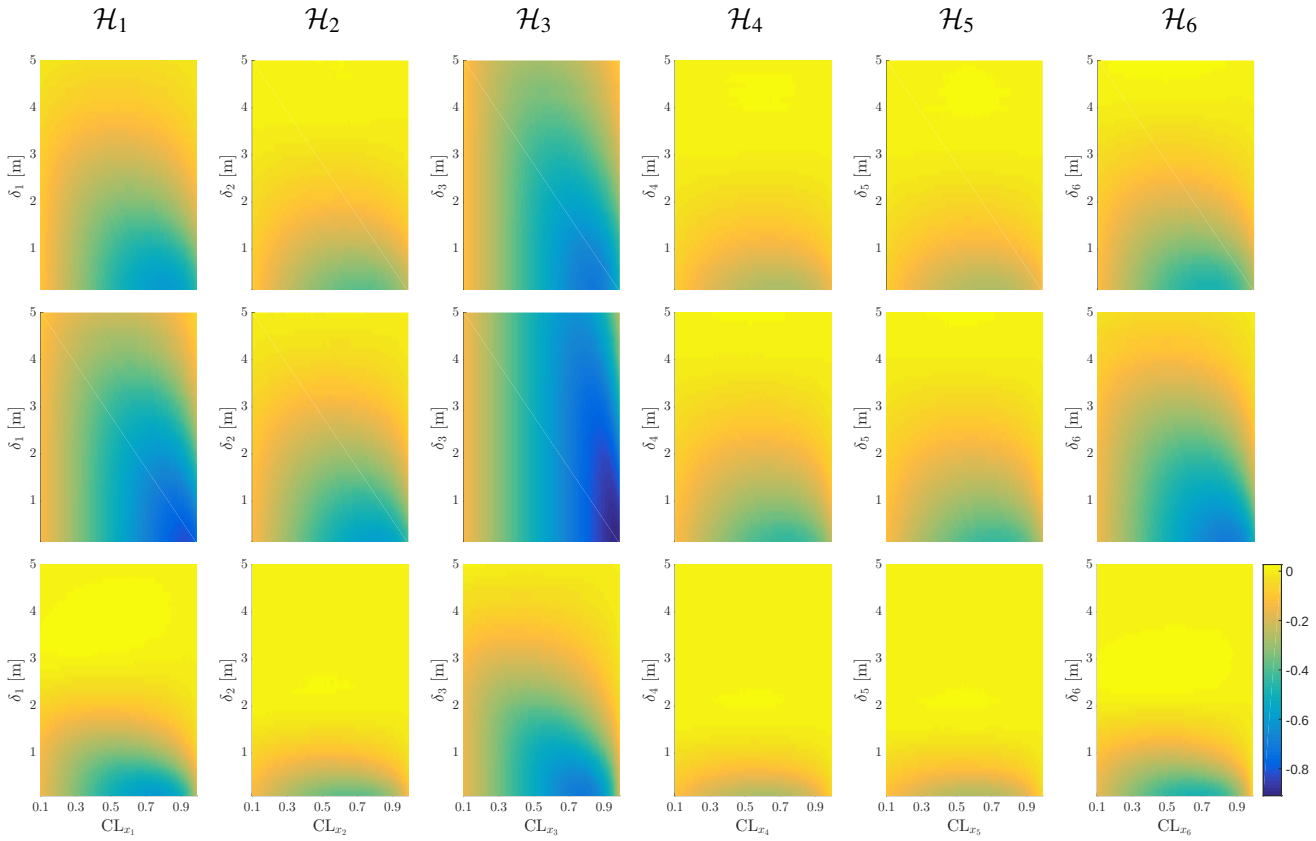

Figure 5: Colormaps of confidence level differences $\mathrm{CL}_{x_i}|(t \in \mathcal{P}_i) - \mathrm{CL}_{x_i}$ as a function of CL_{x_i} horizontally and δ_i vertically, corresponding with the satellite geometry and misclosure space partitioning in Figure 4. The columns from left to right show the results under $\mathcal{H}_i$ for $i = 1,\ldots,6$. The settings from top to bottom are $\sigma_y = 1\mathrm{m},\ \mathrm{P_{FA}} = 10^{-1}$, $\sigma_y = 1\mathrm{m},\ \mathrm{P_{FA}} = 10^{-2}$ and $\sigma_y = 0.5\mathrm{m},\ \mathrm{P_{FA}} = 10^{-1}$.

REFERENCES

[1] Alfaro P, Estevez A, Blazquez E, Borque M, Garrido M, Gil A, Lacy M, Ruiz A, Gimenez J, Molina S, Rodriguez-Caderot G, Ruiz-Morales M, Sanz de Galdeano C (2005) Geodetic Control of the Present Tectonic Deformation of the Betic Cordillera (Spain). In: Geodetic Deformation Monitoring: From Geophysical to Engineering Roles, IAG Proc Vol 131, pp 2009–2016

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

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

[4] Devoti R, Esposito A, Pietrantonio G, Pisani A, Riguzz iF (2011) Evidence of large scale deformation patterns from GPS data in the Italian subduction boundary. Earth Planet Sci Lett (311):230–241

[5] Dheenathayalan P, Small D, Schubert A, Hanssen R (2016) High-precision positioning of radar scatterers. Journal of Geodesy 90(5):403–422

[6] Shahar L, Even-Tzur G (2005) Deformation Monitoring in Northern Israel between the Years 1996 and 2002. In: Geodetic Deformation Monitoring: From Geophysical to Engineering Roles, IAG Proc Vol 131, pp 138–145

[7] Teunissen PJG (2000) Testing theory: an introduction. Delft University Press, Series on Mathematical Geodesy and Positioning

[8] Teunissen PJG (2018) Distributional theory for the DIA method. Journal of Geodesy 92(1):59–80, DOI 10.1007/s00190-017-1045-7

[9] Wieser A (2004) Reliability checking for GNSS baseline and network processing. GPS Solutions 8(3):55–66