On the integrity of deformation monitoring

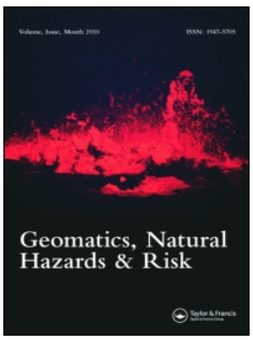
Teunissen, P.J.G.; Zaminpardaz, S.; Tiberius, C.C.J.M.

DOI
10.1080/19475705.2020.1716085

Publication date
2020

Document Version
Final published version

Published in
Geomatics, Natural Hazards and Risk



Taylor & Francis
Taylor & Francis Group

Geomatics, Natural Hazards and Risk

ISSN: 1947-5705 (Print) 1947-5713 (Online) Journal homepage: https://www.tandfonline.com/loi/tgnh20

On the integrity of deformation monitoring

P. J. G. Teunissen, S. Zaminpardaz & C. C. J. M. Tiberius

To cite this article: P. J. G. Teunissen, S. Zaminpardaz & C. C. J. M. Tiberius (2020) On the integrity of deformation monitoring, Geomatics, Natural Hazards and Risk, 11:1, 399-413, DOI: 10.1080/19475705.2020.1716085

To link to this article: https://doi.org/10.1080/19475705.2020.1716085

Published online: 04 Mar 2020.

GEOMATICS, NATURAL HAZARDS AND RISK
2020, VOL. 11, NO. 1, 399–413
https://doi.org/10.1080/19475705.2020.1716085

Taylor & Francis
Taylor & Francis Group

On the integrity of deformation monitoring

P. J. G. Teunissen[a,b], S. Zaminpardaz[c] and C. C. J. M. Tiberius[b]

[a]GNSS Research Centre, Curtin University, Perth, Western Australia, Australia; [b]Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands; [c]School of Science, College of Science, Engineering & Health, RMIT University, Melbourne, Victoria, Australia

ABSTRACT

In safety-critical applications, deformation monitoring systems are required to issue timely alerts when a deformation beyond a critical threshold occurs. Only a very small probability of failing to issue an alert when in fact one should have been given, is acceptable. This probability is referred to as *integrity risk*. In this contribution, we show how to evaluate this risk, thereby taking the intimate link between testing and estimation into account. Using a simple example, the basic integrity components of deformation monitoring are introduced and illustrated. The integrity risk is then formulated for the generalized case where multiple-hypothesis testing is involved. As monitoring systems, in addition to issuing timely alerts, are also required to provide deformation estimates, it is also crucial to assess their confidence levels. In doing so, the statistical testing, that preceded the estimation of the deformation parameters, needs to be accounted for. As this is not the customary procedure followed in practice, we show how the combined estimation and testing can be probabilistically accounted for, and thereby demonstrate that the customary practice can give a too optimistic outcome of the stated confidence levels. The presented methodology is worked out and numerically illustrated by means of two deformation examples.

ARTICLE HISTORY
Received 8 July 2019
Accepted 6 January 2020

KEYWORDS
Deformation; monitoring system; integrity risk; statistical testing; deformation estimation; confidence region

1. Introduction

There is a rich literature on the design and analysis of deformation monitoring systems for both man-made structures (such as a dam, a dike, or a bridge) and natural Earth structures (such as a volcano, a fault, or tectonic plates), see e.g. (Pelzer, 1971; van Mierlo, 1978; Niemeier, 1985; Caspary and Borutta, 1987; Chen et al., 1990; Alfaro et al., 2005; Devoti et al., 2011; Heunecke et al., 2013; Sabuncu and Ozener, 2014; Yigit et al., 2016; Scaioni et al., 2018; Yavaşoğlu et al., 2018). Although these studies developed important and sophisticated statistical procedures for the individual detection and estimation of deformations, no description has yet been given on how to risk-evaluate the overall performance of the system's alert-function. As monitoring

CONTACT P. J. G. Teunissen p.j.g.teunissen@tudelft.nl

systems are safety-critical, only a very small probability is acceptable of the system telling us that no change beyond a critical threshold has taken place (issuing no alert), while in reality it has. In this contribution we show how this probability, referred to as *integrity risk*, can be computed thereby taking all the multivariate aspects of the deformation monitoring into account. As we show, this requires that the statistical considerations of the deformation monitoring's estimation and testing cannot be treated separately anymore, but that their interactions need to be integrally accounted for in the probabilistic integrity evaluation.

This contribution is organized as follows. In Section 2 we first identify the various elements that contribute to the (lack of) integrity of a deformation monitoring system. We hereby make the case that for a proper integrity evaluation one needs to combine the probabilistic consequences of both estimation and testing in the statistical considerations. We then develop in Section 3 the integrity risk for the multiple hypothesis testing problem, which in Section 4 is further generalized to include the aspects of deformation estimation as well. Then, in Section 5, we show how, as a consequence of the interaction between estimation and testing, the confidence levels or confidence regions of the estimated deformation parameters need to be computed. This and the integrity consequences of the interplay between estimation and testing are then illustrated by means of numerical examples in Section 6. The contribution is finalized with our conclusions in Section 7.

2. Integrity elements of deformation monitoring

To illustrate the basic integrity elements of deformation monitoring, we start with the simplest case possible. We assume a deforming body of which the scalar deformation or displacement is directly measured. The *actual* displacement or deformation is denoted as δ, while the *measured* displacement is denoted as d. We assume the measurement to be unbiased, thus relating d to δ as $\mathrm{E}(d) = \delta$, with $\mathrm{E}(\cdot)$ the mathematical expectation operator. As a large deformation is considered unacceptable, an *alert* should be issued when δ becomes too large, say when $|\delta| > \varepsilon$. The measurement or observable d however is never exact and prone to random and other errors. As a consequence, there is a chance that the monitoring system fails to issue an alert, when in fact one should have been given. This is the *integrity risk*. It is the probability that the observed displacement is within limits, $|d| \leq \varepsilon$, while the actual displacement is not, $|\delta| > \varepsilon$

$$\mathbb{IR} = \mathrm{P}\Big(|d| \leq \varepsilon \;\Big|\; |\delta| > \varepsilon\Big) \tag{1}$$

If we assume the observed displacement to be normally distributed, $d \sim \mathrm{N}(\delta, \sigma_d^2)$, then $\mathbb{IR}$ is easily computed as a function of δ. This allows one to judge whether the integrity risk is acceptable or not for the likely occurring values of δ.

Although the above gives a clear picture of what the integrity risk entails, in reality the required derivation and computation of the integrity risk of a deformation monitoring system is far more complicated. As an example consider the deformation monitoring of a dam, whereby a network of reference and object points is used to

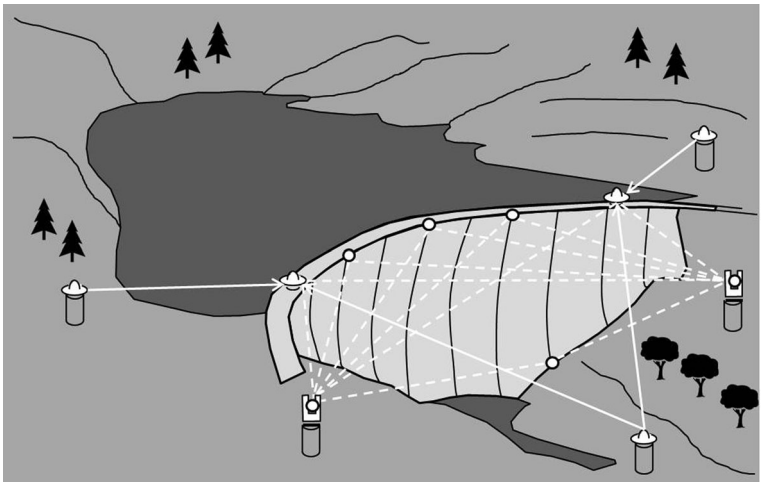


Figure 1. Deformation monitoring of a dam. The monitoring network consists of reference points on pillars around the dam and object points on the dam.

monitor the dam's stability (see Figure 1). In such and similar monitoring systems, one can discriminate between several different data-processing steps (Chen et al., 1990; Heunecke et al., 2013; Sabuncu and Ozener, 2014; Yigit et al., 2016; Scaioni et al., 2018; Yavaşoğlu et al., 2018): At each epoch, there is an epoch-wise determination of the network's geometry which implies a combined use of estimation and testing. The data are first validated by statistical means, after which the 'cleaned' data are used for the epoch-wise coordinate estimation of the network points. Then between-epoch statistical testing is applied to detect and identify possible changes in the network's geometry. This may include multiple tests, such as testing for the stability of reference and object points, as well as the testing for the identification of other hypothesized deformation models. Eventually, depending on the outcome of testing, alerts may be issued and deformation patterns may be estimated, and provided with their corresponding confidence levels.

It is clear that the integrity risk of the above monitoring systems cannot be captured by (1). The measurement scenarios, for instance, are much more involved than that assumed under (1). Often no direct measurements of the deformation are available, but only indirect measurements such as for instance the coordinate outputs of a monitoring network. Also, prior to the testing for deformation, the measurements undergo their own statistical testing for the purpose of quality control, e.g. checking for outliers. And in case multiple deformation models are in play, additional further statistical testing is done to identify the most likely deformation. The conclusion reads therefore that the complexity of deformation monitoring cannot be described as a simple univariate hypothesis testing problem, but instead involves multivariate challenges of both estimation and testing (Teunissen, 2018). In the following we will show how the multivariate combination of estimation and testing affects the composition of the integrity risk. First we consider the integrity risk for the multiple hypothesis testing problem, which we then generalize to include the aspects of deformation estimation as well.

3. Multiple hypothesis testing

As deformation monitoring involves statistical testing of multiple hypotheses, the whole chain of such complex decision making should be reflected in the computed integrity risk. Let the deformation monitoring multiple hypothesis testing problem therefore consist of $k+1$ hypotheses $H_i (i = 0, ..., k)$. The null-hypothesis H_0 is considered to model the all-stable, zero-deformation case, while the other hypotheses model departures from H_0. Then under H_0, and in the presence of redundancy, a vector-function of the observables can be formed that has a fixed and known probability distribution. This vector-function is an ancillary statistic and it is known as the misclosure vector $d \in \mathbb{R}^r$ (Teunissen, 2018). It is then by means of d, and its known probability density function (PDF), that the selection of the most likely hypothesis takes place. As such selection implies the partitioning of $\mathbb{R}^r$ in $k+1$ subsets $P_i \subset \mathbb{R}^r$ $(i = 0, ..., k)$, the testing procedure can be described as

$$\text{Select } H_i \text{ if and only if } d \in P_i \tag{2}$$

In the simple example of the previous section, the observed displacement takes the role of the misclosure for which we could use $P_0 = [-\varepsilon, +\varepsilon]$ and $P_1 = \mathbb{R}/P_0$.

If we now assume that an alert should be issued whenever the 'all-stable' null-hypothesis H_0 is rejected, or equivalently, whenever one of the alternative hypotheses is selected, we have

$$\text{Alert} : \{H_{j\neq 0} \text{ selected}\} \tag{3}$$

A missed alert happens therefore if H_0 is selected, i.e. $d \in P_0$, while $H_{j\neq 0}$ is true. The probability of this happening is given as

$$\mathbb{IR}|H_{j\neq 0} = \mathrm{P}\big(d \in P_0 | H_{j\neq 0}\big) \tag{4}$$

This is therefore the integrity risk conditioned on $H_{j\neq 0}$. To obtain the unconditional integrity risk, we need to consider (4) for all k alternative hypotheses. With $\mathrm{P}(H_i)$ being the probability of occurrence of H_i and $\sum_{i=0}^{k} \mathrm{P}(H_i) = 1$, the overall integrity risk is obtained as

$$\mathbb{IR} = \sum_{i=1}^{k} (\mathbb{IR}|H_i) \times \mathrm{P}(H_i) \tag{5}$$

It thus requires the computation of the probability that d resides in P_0 under each of the alternative hypotheses.

4. Deformation estimation

The integrity risk (5) is based on the assumption that an alert needs to be issued the very moment one of the alternatives $H_{j\neq 0}$ is selected. It does therefore not yet account for the severity, or lack thereof, in the alternative hypothesis. It could happen

namely that, even when a particular deformation as described by $H_{j\neq 0}$ is taking place, the threat of the deformation is still too small to warrant an alert. To build such tolerance into the integrity monitoring, we need to estimate the deformation parameters for each of the postulated deformation models.

We assume that each alternative hypothesis H_i, with $i = 1, ..., k$, has its own set of parameters δ_i. To build-in the mentioned tolerance, we assume that even if H_i is 'active', still no threat emanates from the deformation if δ_i is 'not too large yet' and resides in a zero-centred ball of radius ε, i.e. if $\delta_i \in B_\varepsilon$ with $B_\varepsilon = \{\delta \mid \|\delta\| \leq \varepsilon\}$ (note: for simplicity the radius is assumed here to be the same for all hypotheses; this can however be made hypothesis-specific too). The integrity risk under $H_{j\neq 0}$becomes then

$$\mathbb{IR}|H_{j\neq 0} = \mathrm{P}\big(\text{no alert}|H_j\big)\imath\big(\delta_j\big) \text{ for } j = 1, \ ..., \ k \tag{6}$$

where $\imath\big(\delta_j\big)$ is the indicator function of the region outside B_ε, defined as $\imath\big(\delta_j\big) = 0$ for $\delta_j \in B_\varepsilon$, and $\imath\big(\delta_j\big) = 1$ elsewhere. Thus now the integrity risk under $H_{j\neq 0}$ is still zero if $\delta_j \in B_\varepsilon$.

We need our data to operationalize the 'no alert' situation, i.e. to help verify whether or not all δ_i can be considered small enough. But since the deformation parameters δ_i are unknown, they need to be estimated from the observed data and this needs to be done for each of the postulated hypothesis $H_i (i = 1, ..., k)$. If we denote these estimators as $\hat{\delta}_i$ $(i = 1, ..., k)$, we can describe the 'alert' vs 'no-alert' situation as follows

$$\text{Alert} : \left\{H_{j\neq 0} \text{ selected and } \hat{\delta}_j \notin B_\varepsilon\right\}$$

$$\text{No Alert: } \{H_0 \text{ selected}\} \text{ or } \{H_{j\neq 0} \text{ selected and } \hat{\delta}_j \in B_\varepsilon\}$$

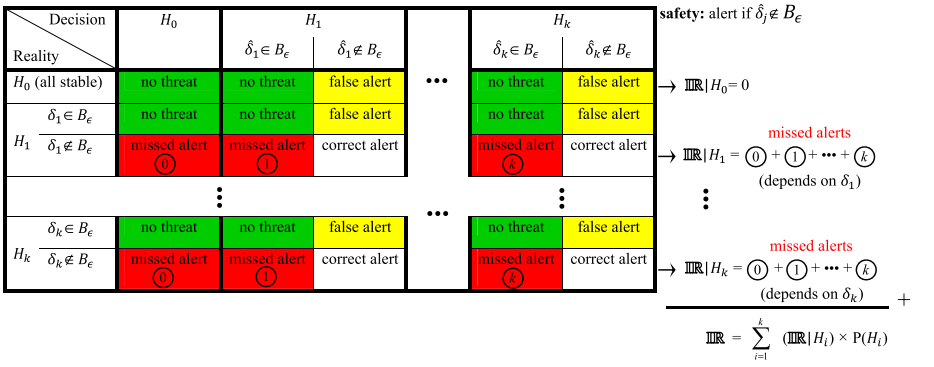
An alert is thus now, in contrast to (3), not immediately issued when one of the $H_{j\neq 0}$ is selected, but only when its estimated deformation parameter $\hat{\delta}_j$ has become too large as well. Note that although more than one of the hypotheses may have its estimated parameter (vector) $\hat{\delta}_j$ lying outside B_ε, the unambiguous selection (2) ensures that only one of the hypotheses will be identified.

With the above 'no-alert' situation, the integrity risk $\mathbb{IR}|H_{j\neq 0}$ can be worked out to give

$$\mathbb{IR}|H_{j\neq 0} = \left[\mathrm{P}\big(d \in P_0|H_j\big) + \mathrm{P}(\hat{\delta}_j \in B_\varepsilon, \ d \in P_j|H_j) + \sum_{i\neq 0,j}^{k} \mathrm{P}(\hat{\delta}_j \in B_\varepsilon, \ d \in P_i|H_j)\right]\imath\big(\delta_j\big) \tag{7}$$

This shows that the integrity risk is now built up from three different terms. The first term, $\mathrm{P}\big(d \in P_0|H_j\big)$, is the missed-alert contribution to the integrity risk. It can be compared with (4). The second term, $\mathrm{P}(\hat{\delta}_j \in \mathrm{B}_\varepsilon, \, d \in P_j|H_j)$, concerns the contribution from the underestimated, but correctly identified deformation, while the last term concerns the sum of all possible wrongfully identified hypotheses $(i \neq 0, \ i \neq j)$.

Table 1. An overview of integrity risk computation (cf. 7) for deformation monitoring application. Integrity risk under a specific hypothesis, say, is constructed from the probability of the corresponding missed alerts (in *red*). False alerts (in *yellow*) are inconvenient and typically imply costs or hassle (taking infrastructure out of service, or evacuating population, while this is not needed). The *white* areas indicate correct alerts, meaning that a critical movement occurs in reality, and the monitoring system issues indeed an alert (though this may be based on an incorrectly identified hypothesis); the word 'correct' is to be interpreted from a safety perspective. By 'no threat' (in green) we mean that the deformation threat in reality is still acceptable, and hence deemed not immediately dangerous. The decision for H_0, H_1, ..., H_k is driven by the misclosure vector d, see (2).



When computing the above integrity risk, it is important to realize that the probability of the joint events $\hat{\delta}_j \in B_\varepsilon,\ d \in P_j\ \ (j = 1, ..., k)$ cannot be computed as the product of the probabilities of the individual events. The random vectors $\hat{\delta}_j$ and d are namely not independent, since δ_j is estimated from d. With (7) substituted into (5) one obtains the overall integrity risk. An overview of the different contributing factors to the overall integrity risk is given in Table 1.

So far, it was assumed that all alternative hypotheses $H_i (i = 1, ..., k)$ can pose dangerous threats. For the case when only a subset of alternatives, say H_i for $i = 1, ..., q$ with $q \leq k$, is considered dangerous, then the event of 'no alert' contains the following events: 'H_0 is selected', 'H_i is selected and $\hat{\delta}_i \in B_\varepsilon\ \ (i = 1, ..., q)$', and '$H_i$ is selected $(i = q + 1, ..., k)$'. For this scenario, the integrity risk for $H_{j\neq 0}$ is no longer given by (7), but by

$$\mathbb{IR}|H_{j\neq 0} = \left[\mathrm{P}\left(d \in P_0|H_j\right) + \mathrm{P}\left(d \in \cup_{i=q+1}^{k} P_i|H_j\right) + \sum_{i=1}^{q} \mathrm{P}\left(\hat{\delta}_i \in B_\varepsilon,\ \ d \in P_i|H_j\right) \right] \iota(\delta_j) \qquad (8)$$

In the special case when only one alternative, say H_j, is considered dangerous $(q = 1)$ and we are only concerned with the threat $\delta_j \notin B_\varepsilon$ (single-threat scenario), the integrity risk (8) simplifies to

$$\mathbb{IR}|H_{j\neq 0} = \left[\mathrm{P}\left(d \notin P_j|H_j\right) + \mathrm{P}\left(\hat{\delta}_j \in B_\varepsilon,\ \ d \in P_j|H_j\right) \right] \iota(\delta_j)$$

which for $k = 1$, when $d = \hat{\delta}_1$ and $P_0 = B_\varepsilon$, with the vanishing of the second term, would give us (1) back again.

5. Presentation of identified deformation

Deformation monitoring not only has the task of issuing alerts when the situation is deemed too dangerous, but also of providing estimates of the deformation together with a computed confidence level, or confidence region of these estimates. To provide such a description of the identified deformation, we assume that the statistical hypothesis testing has done its job properly and identified the correct hypothesis, say $H_{j\neq 0}$. The estimate of the occurring deformation is then given by $\hat{\delta}_j$. The question is now what confidence to assign to this estimate.

To determine the confidence level, we have to define a δ_j-centered region, say $\delta_j + B_\varepsilon$ for a given ε, and compute the associated probability. Since $\hat{\delta}_j$ has been computed from d under $H_{j\neq 0}$, one may be inclined to compute the confidence level p for a given ε as

$$p = \mathrm{P}\left(\hat{\delta}_j \in \delta_j + B_\varepsilon | H_{j\neq 0}\right) \tag{9}$$

or alternatively, obtain the confidence region by computing ε for a given confidence level, say $p = 95\%$. This is indeed the approach that is usually followed in computing confidence regions for estimated deformations, see e.g. (Wieser, 2004; Alfaro et al., 2005; Shahar and Even-Tzur, 2005; Devoti et al., 2011; Dheenathayalan et al., 2016). Assuming the data to be normally distributed and the deformation models to be linear, one then usually presents the confidence regions as balls that are obtained by inverting (9) for a given value of p. In case δ_j is a scalar, the region becomes an interval.

Unfortunately however, the approach of using (9) to compute confidence levels or confidence regions is statistically incorrect. It does namely not do justice to the statistical testing that preceded the estimation of the deformation parameters. After all, the consideration of $\hat{\delta}_j$ in (9) is the result of a testing outcome, namely of having identified $H_{j\neq 0}$. And despite the fact that we assume this identification to be correct, one cannot do away with the fact that this identification to be correct, one cannot do away with the fact that this outcome is the result of the outcome of a random vector d lying in $P_{j\neq 0}$. Thus for a proper computation of the confidence level or confidence region, one has to take this into account as a condition.

The correct way of computing the confidence level or confidence region is therefore to use instead of (9) the following relation

$$p = \mathrm{P}\left(\hat{\delta}_j \in \delta_j + B_\varepsilon | d \in P_j,\ H_{j\neq 0}\right) \tag{10}$$

As was pointed out in the previous section, the conditioning in this relation cannot be nullified since $\hat{\delta}_j$ is estimated from d and is therefore not independent of it.

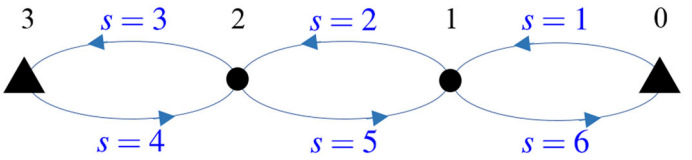


Figure 2. A leveling loop running through four equidistantly spaced points with two of them being reference points (black triangles), $n = 3$ and $r = 2$. The blue curves indicate the measured height differences and denotes the order of instrument set-ups. In this leveling loop, there are six instrument set-ups between successive points.

6. Numerical illustration

To numerically illustrate the integrity concept of the previous sections, we will give two basic examples in the context of vertical deformation monitoring. Let us assume that to monitor the vertical displacement of a deformable structure, e.g. a bridge, a leveling network of $n+1$ equidistantly-spaced control points, labelled as $0, 1, \ldots, n$, is established with $n+1-r$ points on the structure, as *object* points, and r points in a stable area close to this structure, as *reference* points. Comparing different campaigns of leveling data of such a network obtained at different times, we can then determine vertical deformations of the structure at hand. We assume that a leveling loop runs through the network points at two times (or epochs) $t = 1,\ 2$. In each leveling loop, we assume two instrument set-ups for every pair of successive points, implying $2n$ instrument set-ups for $n+1$ points. The leveling observation collected at the s^{th} instrument set-up at epoch t is denoted by $y_{s,t}$ which is expected to be equal to the height difference between the two points to which we take readings at the s^{th} instrument set-up. These observations are assumed to be normally distributed. Figure 2 illustrates an example of such a leveling loop for a network of *four* points of which two are reference points.

At epochs $t = 1, 2$, we observe the model $\mathrm{E}(y_t) = Ax_t$, $\mathrm{D}(y_t) = Q_{yy}$, with $\mathrm{E}(\cdot)$ and $\mathrm{D}(\cdot)$ the mathematical expectation and dispersion operators, $y_t = [y_{1,t},\ y_{2,t}, \ldots, y_{n,t}]^T \in \mathbb{R}^{2n}$ and $x_t \in \mathbb{R}^n$ denoting the height differences of the network points with respect to the first stable point (point 0). Assuming the observations made at different set-ups, i.e. $y_{s,t}$ (for $s = 1, \ldots, 2n$), are independent and of the same standard deviation σ, we have $Q_{yy} = \sigma^2\ I_{2n}$ with I_{2n} being the identity matrix of size $2n$. In case the leveling network contains more than one reference point, i.e. $r > 1$, then the model will extend to accommodate the known height differences between the stable points, i.e. the known height differences are included in the observational model as pseudo-observations with their standard deviations being zero. In the following, we present our analysis for two cases: point 0 is reference point ($r = 1$); points 0 and n are reference points ($r = 2$).

Under the null hypothesis H_0, where no deformation occurs, we assume

$$H_0 : x_2 = x_1 (\text{all stable}) \tag{11}$$

For simplicity of our analysis, we make the following assumptions about the alternative hypotheses that may occur. In case of deformation we assume that either only one point is unstable, or that all object points are unstable with their deformation

being simply linearly related in this example. Thus we have, in case only one point is unstable,

$$H_i : x_2 = x_1 + c_i\delta_i \ \ (\text{point } i = 1, ..., n \text{ is unstable}) \tag{12}$$

with $c_i \in \mathbb{R}^n$ being the canonical unit vector having the 1 as its i^{th} entry and $\delta_i \in \mathbb{R}$ a scalar unknown deformation parameter. Note that with the above alternatives, in addition to the stability of the object points, we also check the stability of the reference points other than point 0. In case all object points are unstable, we assume that the bridge is tilting about point 0, and then the vertical displacement of the object points, which are equidistantly spaced on the bridge, can be modelled as

$$H_{n+1} : x_2 = x_1 + c_{n+1}\delta_{n+1} \ \ (\text{all object points are unstable}) \tag{13}$$

in which $c_{n+1} = [1, 2, 3, ..., n+1-r, 0_{r-1}^T]^T$ and $\delta_{n+1} \in \mathbb{R}$ is a scalar unknown deformation parameter. Therefore, if at point 1 we have a height change of δ_{n+1}, at points $i = 2, 3, ..., n+1-r$ we have height changes of $i \times \delta_{n+1}$.

There is a total of $n+1$ alternative hypotheses. In the testing procedure to validate the above hypotheses, we use the overall model test and Baarda's w-test inducing the following partitioning of the misclosure space $\mathbb{R}^{3n-1+r}$ (Teunissen, 2018; Zaminpardaz and Teunissen, 2018; Zaminpardaz et al., 2019)

$$P_0 = \left\{ d \in \mathbb{R}^{3n-1+r} \;\middle|\; ||\hat{e}_0||_Q^2 \le k_{\alpha, 3n-1+r} \right\} \tag{14}$$

$$P_i = \left\{ d \in \mathbb{R}^{3n-1+r} \setminus P_0 \;\middle|\; |w_i| = \max_{j \in \{1, ..., n+1\}} |w_j| \right\}, \ \ i = 1, ..., n+1 \tag{15}$$

in which $\hat{e}_0$ is the least-squares residual vector under H_0 linked to the misclosure d as $\hat{e}_0 = QB(B^TQB)^{-1}d$ with Q the variance matrix of the observations, and B a basis matrix of the orthogonal complement of the range space of the design matrix (Teunissen, 2000). In (14), $||\cdot||_Q^2 = (\cdot)^T Q^{-1}(\cdot)$ and $k_{\alpha, 3n-1+r}$ is the α-percentage of the central Chi-square distribution with $3n-1+r$ degrees of freedom. α is the false alarm probability, which is usually set a priori by the user, unlike relating the threshold directly to ε, like we did in (1). In (15), w_i for $i = 1, ..., n+1$ are given by

$$w_i = \frac{c_{\hat{e}_{0i}}^T Q^{-1} \hat{e}_0}{||c_{\hat{e}_{0i}}||_Q}, \ \ i = 1, ..., n+1 \tag{16}$$

where $c_{\hat{e}_{0i}}$ characterizes the mean of $\hat{e}_0$ under H_i, i.e. $E(\hat{e}_0|H_i) = c_{\hat{e}_{0i}}\delta_i$.

6.1. Example 1: leveling network with one reference point

Here, we consider a leveling network of four points ($n+1 = 4$) with one, i.e. point 0, as reference ($r = 1$). Thus, there will be four alternative hypotheses; three of the form (12) and one of the form (13). Assuming that the movement of each point is bounded by ε, the integrity risk corresponding with the first three alternatives is then defined by the following zero-centered interval

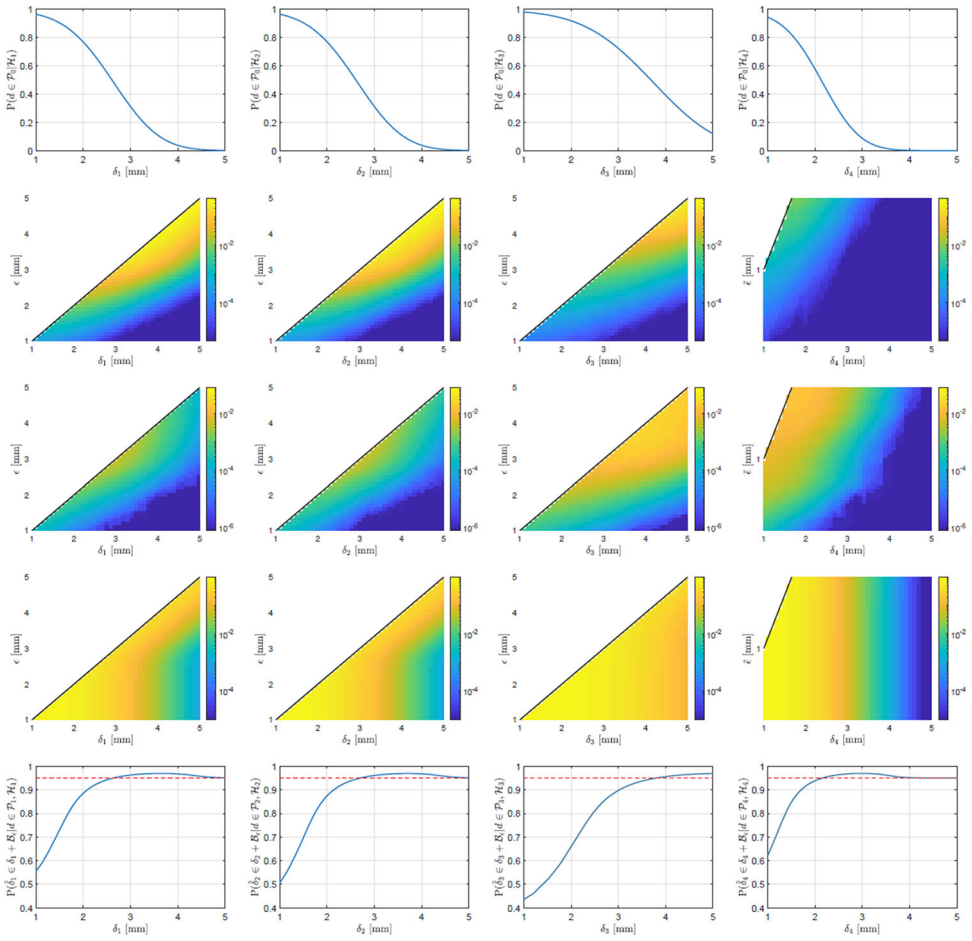


Figure 3. Illustration of integrity risk and confidence levels of deformation estimates corresponding with the leveling network in Figure 2, for, $n = 3$, $r = 1$, $\alpha = 0.01$ and $\sigma = 1\,\text{mm}$. The columns from left to right show the results under H_1, H_2, H_3 and H_4. [First row] Graphs of $\mathrm{P}\big(d \in P_0|H_j\big)$ as a function of [Second row] Colormaps of $\mathrm{P}(\hat{\delta}_j \in B_\varepsilon,\ d \in P_j|H_j)$ as a function of δ_j and ε. [Third row] Colormaps of $\sum_{i\neq 0,j}^{4} \mathrm{P}(\hat{\delta}_i \in B_\varepsilon,\ d \in P_i|H_j)$ as a function of δ_j and ε. [Fourth row] Colormaps of $\mathbb{IR}|H_j$ as a function of δ_j and ε. [Fifth row] Graphs of confidence level corresponding with $(\hat{\delta}_j \in \delta_j + B_\varepsilon|d \in P_j,\ H_{j\neq 0})$ as a function of δ_j according to (10), where B_ε is set to correspond to a confidence level of 0.95 according to (9), when the impact of testing is neglected; the red dashed line indicates the confidence level of 0.95. Note the difference in colorbar scale between colormaps of different rows.

$$B_\varepsilon = \left\{ \delta \in \mathbb{R} \,\middle|\, |\delta| \leq \varepsilon \right\} \tag{17}$$

The deformation described by the fourth alternative however, results in a displacement of $i \times \delta_{n+1}$ for point i. In this case, the maximum displacement that would

occur equals $(n+1-r)\times\delta_{n+1}$ for point $n+1-r$. To make sure that the movement of each point does not go beyond the required threshold ε, the maximum displacement needs to remain below ε, i.e. $\left|(n+1-r)\times\delta_{n+1}\right|\leq\varepsilon$. The integrity risk corresponding with the fourth alternative is then defined by $\mathrm{B}_{\tilde{\varepsilon}}$ with $\tilde{\varepsilon}=\frac{\varepsilon}{n+1-r}$.

Figure 3 depicts, for $\alpha=0.01$ and $\sigma=1$ mm, the behaviour of the integrity risk under each of the alternative hypotheses, each column represents an alternative hypothesis. Following the three terms in (7), the top row shows the graphs of $\mathrm{P}\left(d\in P_0|H_j\right)$ as a function of δ_j. With $\varepsilon_1=\varepsilon_2=\varepsilon_3=\varepsilon$ and $\varepsilon_4=\tilde{\varepsilon}$, the second row shows the colormaps of $\mathrm{P}\left(\hat{\delta}_j\in B_{\varepsilon_j},\ d\in P_j|H_j\right)$ as a function of δ_j and ε_j. The third row shows the colormaps of $\sum_{i\neq 0,j}^{4}\mathrm{P}\left(\hat{\delta}_i\in B_{\varepsilon_j},\ d\in P_i|H_j\right)$ as a function of δ_j and ε_j. The fourth row shows the colormaps of as a function of $\mathbb{IR}|H_j$ as a function of δ_j and ε_j. As integrity risk concerns those situations where the threat goes beyond the threshold ε_j, cf. (7), the part of the colormaps above the straight line $\varepsilon_j=\delta_j$ is left empty. Note that the scale of the colorbars is logarithmic and varies from one row to another.

The probability $\mathrm{P}\left(d\in P_0|H_j\right)$ describes the missed-detection probability of the testing procedure under H_j. Therefore, as the top panels in Figure 3 also show, $\mathrm{P}\left(d\in P_0|H_j\right)$ gets smaller when the deformation magnitude gets larger. It is observed that for a given deformation magnitude, the alternatives can be ordered in terms of deformation detectability as $H_4>H_2{=}H_1>H_3$. This behaviour can be explained as follows. The potential movement of point 1 or point 2 w.r.t. point 0 from $t=1$ to $t=2$, described by H_1 and H_2, will affect four leveling observations, i.e. $y_{1,2}$, $y_{2,2}$, $y_{5,2}$, $y_{6,2}$ for point 1 and $y_{2,2}$, $y_{3,2}$, $y_{4,2}$, $y_{5,2}$ for point 2 (see Figure 2). Whereas, movement of point 3 w.r.t. point 0 from $t=1$ to $t=2$, described by H_3, will affect only two leveling observations, i.e. $y_{3,2}$, $y_{4,2}$. And, finally, if all the points 1, 2 and 3 move according to H_4, then all the leveling observations $y_{s,2}$ ($s=1,\ \ldots,\ 6$) will sense these movements. Thus, it is expected that the testing detection step, cf. (14), has the same sensitivity to H_1- and H_2-deformations, greater than H_3-deformations but lower than H_4-deformations.

As the panels on the second row show, $\mathrm{P}(\hat{\delta}_j\in B_{\varepsilon_j},\ d\in P_j|H_j)$ only gets significant values around $\varepsilon_j=\delta_j$, particularly when δ_j gets larger than a specific value. For a given δ_j, when ε_j increases, the probability mass of the PDF of $\hat{\delta}_j$ inside B_{ε_j} increases while that of the PDF of d inside $P_j$remains unchanged. Therefore, for a given δ_j, $\mathrm{P}(\hat{\delta}_j\in B_{\varepsilon_j},\ d\in P_j|H_j)$ is an increasing function of ε_j.

The signature of $\mathrm{P}(\hat{\delta}_j\in B_{\varepsilon_j},\ d\in P_j|H_j)$ as a function of δ_j can be decreasing or increasing, which can be explained as follows. For a given ε_j, when δ_j increases, the probability mass of the PDF of $\hat{\delta}_j$ inside B_{ε_j} decreases. However, the probability mass of the PDF of d inside P_j increases when δ_j increases. Therefore, $\mathrm{P}(\hat{\delta}_j\in B_{\varepsilon_j},\ d\in P_j|H_j)$ may decrease or increase as a function of δ_j depending on the interaction between the probabilistic properties of $\hat{\delta}_j$ and d. The third part of the integrity risk, i.e. $\sum_{i\neq 0,j}^{4}\mathrm{P}(\hat{\delta}_i\in B_{\varepsilon_j},\ d\in P_i|H_j)$ shown on the third row, gets small values for almost all considered ranges of ε_j and δ_j. This can be attributed to low probabilities of wrong identification implying that under H_j, the PDF of d has low probability mass in $P_{i\neq 0,j}$. The colormaps on the fourth row are obtained by adding the

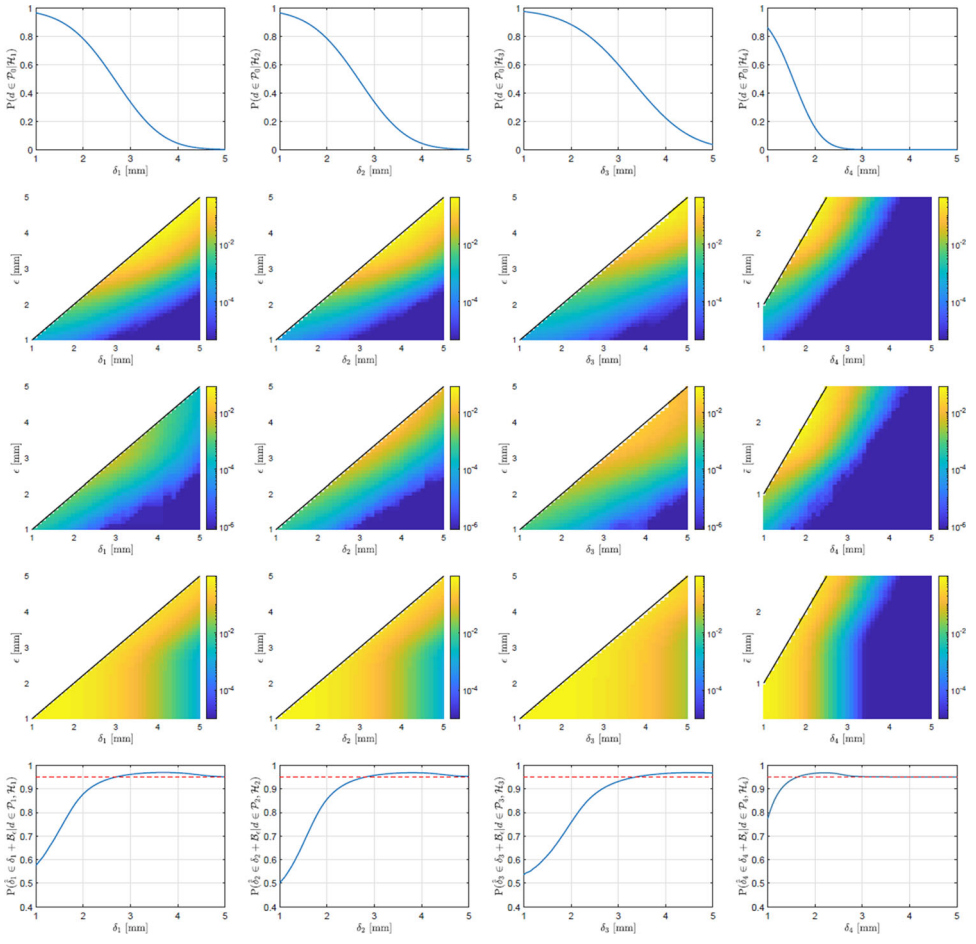


Figure 4. Illustration of integrity risk and confidence levels of deformation estimates corresponding with the leveling network in Figure 2, for $n = 3$, $r = 2$, $\alpha = 0.01$ and $\sigma = 1$ mm. The columns from left to right show the results under H_1, H_2, H_3 and H_4 . [First row] Graphs of $\mathrm{P}\big(d \in P_0|H_j\big)$ as a function of δ_j. [Second row] Colormaps of $\mathrm{P}(\hat{\delta}_j \in B_\varepsilon,\ d \in P_j|H_j)$ as a function of δ_j and ε. [Third row] Colormaps of $\sum_{i\neq 0,j}^{4} \mathrm{P}(\hat{\delta}_i \in B_\varepsilon,\ d \in P_i|H_j)$ as a function of δ_j and ε. [Fourth row] Colormaps of $\mathbb{IR}|H_j$ as a function of δ_j and ε. [Fifth row] Graphs of confidence level corresponding with $(\hat{\delta}_j \in \delta_j + B_\varepsilon|d \in P_j,\ H_{j\neq 0})$ as a function of δ_j according to (10), where B_ε is set to correspond to a confidence level of 0.95 according to (9), when the impact of testing is neglected; the red dashed line indicates the confidence level of 0.95. Note the difference in colorbar scale between colormaps of different rows.

corresponding integrity risk values on the first three rows, showing that in this case the missed detection delivers by far the largest contribution.

Assuming that the statistical hypothesis testing has identified the correct hypothesis and the corresponding deformation has been estimated, we now analyse the confidence level for this estimated deformation. To do so, for a given confidence level p in (9), where we set $p = 0.95$, we first determine the region B_ε. This region together with actual displacement δ_j, known from simulation, are then substituted into (10) to

compute the *correct* confidence level for the estimated deformation when preceded by statistical testing. The last row in Figure 3 shows the correct confidence levels corresponding with $(\hat{\delta}_j \in \delta_j + B_\varepsilon | d \in P_j,\ H_{j\neq 0})$ as a function of δ_j $(j = 1, ..., 4)$, where $B_\varepsilon$is set to correspond to a confidence level of 0.95 in (9), when the impact of testing is neglected. The red dashed lines in the last-row panels indicate this 0.95 confidence level. As can be seen, for $\delta_1 < 2.6$ mm, $\delta_2 < 2.6$ mm, $\delta_3 < 3.8$ mm and $\delta_4 < 2.1$ mm (typically displacement values of a few times the standard deviation of the observable), $\mathrm{P}(\hat{\delta}_j \in \delta_j + B_\varepsilon | d \in P_j,\ H_{j\neq 0})$ is *smaller* than $\mathrm{P}(\hat{\delta}_j \in \delta_j + B_\varepsilon | H_{j\neq 0})$ for $j = 1, ..., 4$, revealing that ignoring the conditioning on the testing decision results in a too optimistic description of the estimator's quality.

6.2. Example 2: leveling network with two reference points

Here, we again consider the leveling network of the previous example with the same set of observations, be it that in this example in addition to point 0, point 3 is also considered to be a reference point ($r = 2$). In this case, the leveling observation equation extends to accommodate the known height difference between the point 0 and point 3, i.e. the known height difference between the point 0 and point 3 is included in the observational model as a pseudo-observation with its standard deviation being zero. Figure 4 shows the same information as Figure 3, but for this second example where we have two reference points. In the following, we make comparisons between the results under $H_i$for $i = 1,\ 2,\ 3$ as these alternatives describe the movements of points $i = 1,\ 2,\ 3$ in both examples. However, one should note that, under each of these alternative hypotheses, the observational model is different for Example 1 and 2. In case of H_4, no comparison is made as the nature of this fourth alternative differs from Example 1 to this current example. Under H_4, the linear motion of the points on the bridge concerns points 1, 2 and 3 in Example 1, and only points 1 and 2 in Example 2.

As it can be seen from the first row of Figure 4, the detectability of points 1 and 2 deformation has remained almost unchanged compared to the previous example. However, the testing procedure is now a bit more sensitive to movement of point 3 which is due to the fact that the additional information in Example 2 directly constrains the height difference of point 3 with respect to point 0. The panels on the second row corresponding with $\mathrm{P}(\hat{\delta}_j \in B_{\varepsilon_j},\ d \in P_j | H_j)$ for $j = 1,\ 2$, show similar values as their counterparts did in Figure 3. Under H_3, the second part of integrity risk now has a bigger contribution as compared with the previous example particularly for the cases where both δ_3 and ε take large values. The third part of integrity risk, i.e. $\sum_{i\neq 0,j}^{4} \mathrm{P}(\hat{\delta}_i \in B_{\varepsilon_j},\ d \in P_i | H_j)$ has remained almost unchanged under H_1, increased under $H_2$and decreased under H_3. The overall integrity risk colormaps corresponding with points 1 and 2 deformations show almost the same signature as their counterparts in Example 1. In case of the deformation of point 3, which is now a reference point, the overall integrity risk takes smaller values. Comparing the confidence levels in Figure 4 with those in Figure 3, no significant difference can be found for the conditional deformation estimates under H_1 and H_2. Under H_3, a higher

confidence level is assigned to the corresponding conditional deformation estimate when point 3 is considered as a reference point.

7. Conclusions

Although statistical procedures for the individual detection and estimation of deformations have long been in place, risk evaluation of the overall performance of a monitoring system's alert-function has not yet been described. In this contribution, we presented integrity risk evaluation in the context of deformation monitoring applications. Integrity risk is referred to as the probability of a monitoring system failing to issue an alert, when in fact one should have been given.

Commencing with a simple deformation example with one alternative hypothesis, the basic integrity components of deformation monitoring were introduced and illustrated. As deformation monitoring involves statistical testing of multiple hypotheses, we then developed the integrity risk for the multiple hypothesis testing problem. Using the concept of misclosure space partitioning, it was shown how the integrity risk is constructed from the testing decisions under the alternative hypotheses in consideration. A further generalization was then introduced by having the alerts not solely dependent on the identified hypothesis, but also on the threat that the estimated size of deformations entails. It was thereby shown how the required probabilistic properties of both estimation and testing come together in the computation of the integrity risk.

Deformation monitoring, in addition to issuing timely alerts, is often also required to provide estimates of the deformation together with their associated confidence levels. These confidence levels are usually computed without taking the statistical testing that preceded the deformation estimation into account. We have shown however that it is the combined estimation and testing that needs to be probabilistically taken into account and that failing to do so can give a too optimistic outcome of the stated confidence levels.

To numerically illustrate the various aspects of the presented integrity concept, two simple examples of deformation monitoring were given. The contributions from different measurement setups, alternative hypotheses and testing decisions to the integrity risk were discussed and analysed. Also the confidence levels were computed with and without considering the intimate link between estimation and testing, thereby numerically confirming that ignoring the effect testing has on the confidence level, may lead to a too optimistic description of the deformation estimator's quality.

Acknowledgement

TU Delft Library is acknowledged for providing funding for Open Access publication of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Alfaro P, Estevez A, Blazquez E, Borque M, Garrido M, Gil A, Lacy M, Ruiz A, Gimenez J, Molina S, et al. 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. p. 2009–2016, Springer, Berlin, Heidelberg.

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

Chen Y, Chrzanowski A, Secord J. 1990. A strategy for the analysis of the stability of reference points in deformation surveys. CISM J. 44(2):39–46.

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(3–4):230–241.

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

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

Niemeier W. 1985. Deformationsanalyse (in German). In: Pelzer H, editor. Geodätische Netze in Landes- und In- genieursvermessung II: Vorträge des Kontaktstudiums Februar 1985 in Hannover, K. Stuttgart: Wittwer Verlag. Chapter 15; p. 559–623.

Pelzer H. 1971. Zur Analyse geodätischer Deformationsmessungen Deutsche Geodätische Kommission, Reihe C: Dissertationen - Heft [dissertation] Nr. 164, München, Germany (in German).

Sabuncu A, Ozener H. 2014. Monitoring vertical displacements by precise levelling: a case study along the Tuzla Fault, Izmir, Turkey. Geomat, Nat Hazards Risk. 5(4):320–333.

Scaioni M, Marsella M, Crosetto M, Tornatore V, Wang J. 2018. Geodetic and remote-sensing sensors for dam deformation monitoring. Sensors. 18(11):3682.

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, p. 138–145, Springer, Berlin, Heidelberg.

Teunissen PJG. 2018. Distributional theory for the DIA method. J Geod. 92(1):59–80.

Teunissen PJG. 2000. Testing theory: an introduction. Delft University Press, Series on Mathematical Geodesy and Positioning. Delft, The Netherlands

van Mierlo J. 1978. A Testing Procedure for Analysing Geodetic Deformation Measurements. In Proceedings of the II. International Symposium of Deformation Measurements by Geodetic Methods; Sep 25–28; Bonn, Germany: Konrad Wittwer, Stuttgart. p. 321–353.

Wieser A. 2004. Reliability checking for GNSS baseline and network processing. GPS Solut. 8(2):55–66.

Yavaşoğlu HH, Kalkan Y, Tiryakioğlu I, Yigit CO, ÖZbey V, Alkan MN, Bilgi S, Alkan RM. 2018. Monitoring the deformation and strain analysis on the Ataturk Dam, Turkey. Geomat, Nat Hazards Risk. 9(1):94–107.

Yigit C, Alcay S, Ceylan A. 2016. Displacement response of a concrete arch dam to seasonal temperature fluctuations and reservoir level rise during the first filling period: evidence from geodetic data. Geomat, Nat Hazards Risk. 7(4):1489–1505.

Zaminpardaz S, Teunissen PJG. 2018. DIA-datasnooping and identifiability. J Geod. 93:85–101.

Zaminpardaz S, Teunissen PJG, Tiberius C. 2019. Risking to underestimate the integrity risk. GPS Solut. 23(2):1–16.