

# ILLUSTRATIVE PRESENTATION OF SOME BASIC CONCEPTS OF NETWORK INTERNAL RELIABILITY WITH COMMENTS AS REGARDS ENGINEERING SURVEYS

Witold Prószyński

Faculty of Geodesy and Cartography  
Warsaw University of Technology

## Abstract

*The paper integrates some earlier and the recent findings of the author in the area of network internal reliability and presents a consistent system of concepts in this respect. The concepts of outlier detection and outlier identification linked directly with the global model test and the outlier tests respectively, are shown as a basis for the concepts such as outlier detectability and outlier identifiability. Also, a four level classification of gross errors expressed in a form of a tree-diagram is presented including perceptible and imperceptible errors, detectable and undetectable errors and identifiable and unidentifiable errors. Their properties are given mainly in a descriptive way, deliberately limiting rigorous mathematical formulas to a necessary minimum. Understanding of different types of gross errors is useful in analyzing the results of the outlier detection and identification procedures as well as in designing the networks to make them duly robust to observation gross errors. It is of special importance for engineering surveys where quite often low-redundancy networks are used. Main objective of the paper is to demonstrate a clear and consistent system of basic concepts related to network internal reliability.*

**Keywords:** *unit redundancy, low-redundancy networks, detectability, identifiability, gross error classification*

## 1. Introduction

The subject-matter of internal reliability of observation systems is concerned with the behaviour of these systems in the presence of observation gross errors. It is of special importance for engineering surveys such as precise setting out or deformation monitoring in high risk structures. The knowledge of basic reliability concepts and the corresponding measures can be useful in the design of networks as well as in the interpretation of adjustment results.

It should be noted that the global index of internal reliability can be termed equivalently the index of unit redundancy, i.e. the redundancy per single observation. We get it immediately, denoting the number of observations by  $n$ , the number of model parameters by  $u$  and the network defect by  $d$ , and expressing the global index of network internal reliability  $\bar{h}$  ( $0 \leq \bar{h} < 1$ ) as:

$$\bar{h} = 1 - \frac{u - d}{n} = \frac{n - (u - d)}{n} = \frac{\text{redundancy}}{n} = \text{unit redundancy}$$

Hence, the networks of low or high internal reliability can be understood as networks of *low unit redundancy* or *high unit redundancy* respectively. To simplify terminology we will be calling them here just low-redundancy and high-redundancy networks.

First, the concepts of *detection* and *identification* are discussed as a basis for the concepts of *detectability* and *identifiability*. Further, the classification of possible types of gross errors is presented.

In the present paper use is made of some earlier and the recent publication of the author, devoted to gross errors (Prószyński, 2008, 2010, 2015). The presentation does not pretend to be an extensive study of all the properties of gross errors but is confined to those that are decisive for finding the outliers in the phase of data processing. Explanatory role of the paper comes to the fore.

## 2. Distinction between the concepts of detectability and identifiability

In the subsequent explanations we concentrate on a single outlier case in networks with uncorrelated or correlated observations.

In some publications dealing with data processing procedures the terms *detection* and *identification* are used interchangeably for describing the final effect of finding an observation contaminated with a gross error. Following rigorous distinction between these concepts made by Teunissen (2000), we emphasize here their direct assignment to the global model test and the outlier tests correspondingly. As shown in Fig. 1, when the test statistic  $T$  in a global model test is greater than its critical value  $T_{\text{crit}}$  and we may assume that the model itself is correct, we have a basis to state that we found (i.e. detected) the presence of a gross error in a network. Obviously, we do not know its location in a network. In the outlier tests [see (Baarda, 1968)] for uncorrelated observations and (Knight et al., 2010) for correlated observations), by finding the observations for which the absolute value of outlier test statistic i.e.  $|w_i|$  ( $i = 1, \dots, s$ ) exceeds the critical value  $|w|_{\text{crit}}$ , we obtain a set of  $s$  observations suspected of containing a gross error. The element of the set with the maximal  $|w|$ , is identified as the observation containing a gross error ( $k$ -th observation in Fig.1).

In the phase of a priori analysis of network internal reliability we deal with the corresponding concepts *detectability* and *identifiability*. *Detectability* is expressed by the minimal magnitude of gross error (MDB) in a particular observation, the presence of which in a system can be detected in a global model test. The concept of MDB initiated by Baarda (1968) was further analyzed by (Wang and Chen, 1994), Schaffrin 1997 and Teunissen, 1990, 1998, 2000). It should be noted that the MDB is a stochastic quantity that depends on the assumed probabilities  $\alpha$  and  $\beta$  of the I-type and II-type error respectively.

*Identifiability* relates to the chances that the observation found in the identification process is the observation actually contaminated with a gross error. The chances are expressed in a form of identifiability index  $ID$  for a particular observation as presented in (Prószyński, 2015), where coordination between the global test and the outlier tests was assumed with respect to non-centrality effect (Knight et al., 2010). The index is a probability that the  $i$ -th observation contaminated with a gross error of the  $MDB_i$  magnitude will have  $|w_i|$  dominating in a set of suspected observations. The value of  $ID_i$  can be determined on the basis of numerical simulation of observation random errors with the non-centrality of  $MDB_i$  magnitude for the  $i$ -th observation

Another measure of *identifiability* for the  $i$ -th observation can be the maximal of mis-identifiability indices  $MID_{j/i}$  ( $j \neq i$ ) for this observation, denoted as  $MID_{j/i,max}$  (see Fig.1). Each of  $MID_{j/i}$  indices is a probability of type III error  $\gamma_{j/i}$ , for  $j \neq i$ , i.e. a probability that with a gross error residing in the  $i$ -th observation,  $|w_j|$  will be dominating in a set of suspected observations. The concept of type III error was first presented in (Hawkins, 1980; Foerstner, 1983).

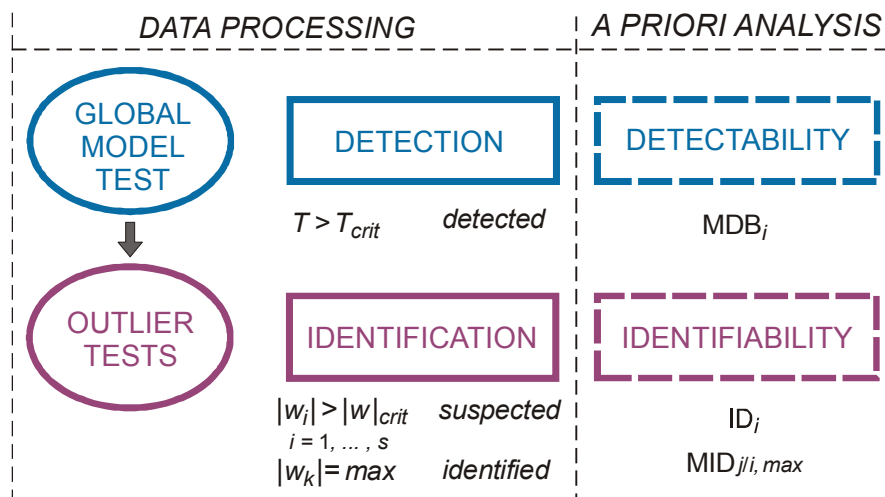


Fig. 1. Detection and identification as a basis for detectability and identifiability

The term *identifiability* has much in common with the more and more widely used term *separability* (eg. Wang and Knight, 2012; Yang et al., 2013). However, the latter term seems to have a broader usability especially for distinguishing between different types of alternative hypotheses.

It is necessary to mention also the concept of Minimal Separable Bias (MSB) (eg. Wang and Knight, 2012), which is the  $MDB$  magnified so as to be separable in the outlier test with a desired level of probability. For the error of  $MSB$  magnitude the  $ID$  index should satisfy the correspondingly high level of probability.

### 3. Classification of gross errors

Figure 2 shows classification of gross errors in a form of a tree-diagram. It has hierarchical 4-level structure which should be analyzed from top to bottom, i.e. to be detectable a gross error must be perceptible and to be identifiable the error must be detectable. Each classification level corresponds to a specified criterion as will be presented further in this Section.

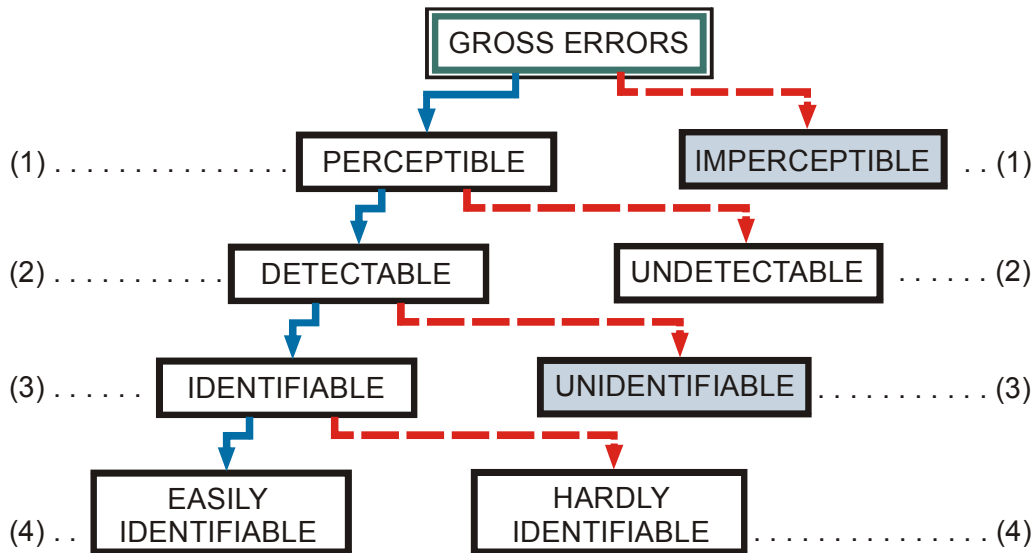


Fig. 2. Classification of gross errors

**Level 1.** A basic concept that decides about the distinction between *perceptible* and *imperceptible* gross errors for each individual adjustment model is its vector space of imperceptible observation errors (Prószyński, 2008).

The formal definition of this space is as follows:

$$U = \{\Delta y : \Delta y \in N(\mathbf{A})\}$$

where:

$\Delta y$  - the  $(n \times 1)$  vector of gross errors,

$\mathbf{A}$  - the  $(n \times u)$  design matrix,  $\text{rank } \mathbf{A} = u - d$  ( $d$  - network defect),

$N(\mathbf{A})$  - null space of  $\mathbf{A}$ ,  $\dim U = \dim N(\mathbf{A}) = \text{rank } \mathbf{A} = u - d$ .

The above definition can be formulated in a descriptive way as shown below:

- the space  $U$  is formed by such non-zero vectors of gross errors  $\Delta y$ , that do not effect the least squares (LS) residuals, or in short  $\Delta y \neq \mathbf{0} \Rightarrow \Delta v = \mathbf{0}$ .

Such errors are called imperceptible gross errors. Since they do not effect the LS residuals, they cannot be noticed in the tests based on the LS residuals. We can easily prove that their occurrence in the system results in distorted values of the model parameters.

Here are some chosen properties related with the space  $U$ :

- imperceptible* gross error - a gross error being any component of a vector belonging to the space  $U$ ,
- a single gross error is always perceptible,
- the relationship between the dimension of the space  $U$ , i.e.  $\dim U$ , and the global index of network internal reliability  $\bar{h}$  is the following:

$$\dim U = (1 - \bar{h})n$$

By increasing the network internal reliability we reduce the space of imperceptible errors, e.g.  $\bar{h} = 0$ ,  $\dim U = n$ ;  $\bar{h} < 0.5$ ,  $\dim U > 0.5n$ ;  $\bar{h} > 0.5$ ,  $\dim U < 0.5n$ .

d) the space of imperceptible errors does not depend on the assumed covariance matrix of observations, but on structural features of the design matrix **A**.

The corresponding box in a diagram has been shaded in order to emphasize that the network structures that allow the occurrence of imperceptible gross errors (i.e. the errors that are distortive with respect to parameter estimation) should be avoided.

**Level 2.** A single gross error in a particular observation is *detectable* when its magnitude is greater than the MDB for this observation. Otherwise the error is termed *undetectable*. To improve detectability of gross errors in a model we may redesign the model by strengthening it so that the MDBs become smaller quantities. The possibility of occurrence of gross errors smaller than MDB, and hence undetectable, is a natural feature of every network model.

**Level 3.** A basic factor that decides about the existence of *identifiable* and the *unidentifiable* gross errors in an individual model is whether the model contains the Region(s) of Unidentifiable Errors (RUE).

The RUE can be defined descriptively as follows:

- a subset of network observations  $y_1, \dots, y_s$  ( $2 \leq s \leq n$ ), such when contaminated with any combination of gross errors that do not form a vector belonging to the space  $U$ , always yields:

$$|w_1| = |w_2| = \dots = |w_s|$$

where  $w_i$  are the outlier test statistics either in Baarda w-test (for uncorrelated observations) or the test based on mean shift model (correlated observations). *Unidentifiable* gross error – a gross error residing in any of the observations which form RUE. Such an error is unidentifiable within a particular RUE.

The RUE takes place specially in leveling networks of low redundancy, and in any type of network with 1 redundant observation where it covers the whole network. The conditions for non-existence of RUE in a network are as follows:

- a)  $h_i > 0.5$  ( $i = 1, \dots, n$ ) uncorrelated observations
- b)  $h_{*i} > 0.5$ ;  $0 \leq k_i < 1$  ( $i = 1, \dots, n$ ) correlated observations

where  $h_i$  and  $h_{*i}, k_i$  are indices of response-based internal reliability for the  $i$ -th observation (Prószyński, 2010).

Therefore, we can see that in high-redundancy networks, being the networks of high internal reliability, the detectable gross errors are always identifiable.

To emphasize that the RUE regions that allow the occurrence of unidentifiable gross errors and should thus be avoided in a network, the corresponding box in a diagram has been deliberately shaded.

**Level 4.** The degree of identifiability can be characterized by the  $ID_i$  index (Prószyński, 2015). As a supporting quantity one can use the maximum value of misidentifiability index, i.e.  $MID_{j/i,max}$ . The criterion for distinguishing between *easily identifiable* and *hardly identifiable* gross errors has not yet been established. It is however known that the increase in network internal reliability implies the increase of  $ID_i$  and decrease of  $MID_{j/i,max}$ .

#### 4. Concluding remarks

In the light of internal reliability theory the low-redundancy networks ( $h < 0.5$ ) are risky measuring systems, where imperceptible gross errors as well as unidentifiable gross errors may occur. Hence, especially in engineering surveys of high importance such networks should as far as possible be avoided.

High redundancies ( $h > 0.5$ ) diminish the chances for occurrence of imperceptible gross errors and eliminate completely the unidentifiable errors.

The classification and its description need some upgrading to take into account recent achievements in constructing the testing procedures and specifying the MDBs for the case of multiple outliers. Obviously, the multiple outliers which contain imperceptible or unidentifiable gross errors are already covered in the proposed classification.

#### References

- Baarda, W. (1968). A testing procedure for use in geodetic networks. *Publications on Geodesy*, New Series, vol.2, no.5, Netherlands Geodetic Commission, Delft
- Förstner, W. (1983). Reliability and discernability of extended Gauss-Markov models. *Deutsche Geodätische Kommission*, Reihe A, No. 98, München
- Hawkins, D. M. (1980). *Identification of outliers*. Chapman and Hall, New York
- Knight, N. L., Wang, J. & Rizos, C. (2010). Generalised measures of reliability for multiple outliers. *J Geod.*, DOI. 10.1007/s00190-010-0392-4
- Prószyński, W. (2008). The vector space of imperceptible observation errors: a supplement to the theory of network reliability. *Geodesy and Cartography*; vol. 57, No.1, pp. 3-19
- Prószyński, W. (2010). Another approach to reliability measures for systems with correlated observations. *J Geod* 84: pp. 547-556
- Prószyński, W. (2015). Revisiting Baarda's concept of minimal detectable bias with regard to outlier identifiability. *J Geod* 89: pp. 993-1003
- Schaffrin, B. (1997). Reliability measures for correlated observations. *J. Surv. Eng.*, 123 (3), pp. 126-137
- Teunissen, P. J. G. (1990). Quality control in integrated navigation systems. *IEEE Aersp Electron Syst Mag* 5(7), pp. 35-41
- Teunissen, P. J. G. (1998). Minimal detectable biases of GPS data. *J Geod* 72: pp. 236-244.
- Teunissen, P. J. G. (2000). *Testing theory: an introduction*. Delft University Press, Delft
- Wang, J. & Chen, Y. (1994). On the reliability measure of observations. *Acta Geodaet. et Cartograph. Sinica*, English Edition, pp. 42-51
- Wang, J. & Knight, N. (2012). New Outlier Separability Test and Its Application in GNSS Positioning. *Journal of Global Positioning Systems*, 11(1), pp. 46-57
- Yang, L., Wang, J., Knight, N. & Shen, Y. (2013). Outlier Separability Analysis with a Multiple Alternative Hypotheses Test. *J Geod* 87(6), pp. 591-604

---

#### Authors:

Witold Prószyński <sup>1)</sup>, [w.proszynski@gik.pw.edu.pl](mailto:w.proszynski@gik.pw.edu.pl)

<sup>1)</sup> Warsaw University of Technology  
Faculty of Geodesy and Cartography  
Pl. Politechniki 1,  
00-661 Warsaw, Poland