Reliability of GPS cycle slip and outlier detection[1]

P.J.G. Teunissen, C.D. de Jong
Department of Mathematical Geodesy and Positioning
Delft University of Technology
Thijsseweg 11, 2629 JA Delft, The Netherlands

ABSTRACT

The theory and application of statistical quality control is well established in surveying and geodesy. In recent years it is also gaining momentum in applications of marine positioning. The set of quality control recommendations of the United Kingdom Offshore Operators' Association (UKOOA) is a good example in this respect. Quality control is made up of various contributing factors, one of which is the concept of reliability. By means of the Minimal Detectable Biases (MDBs), the concept of reliability provides diagnostic tools to infer the strength with which positioning models can be validated. In this contribution we will present in analytical form the MDBs of single- and dual-frequency GPS code data (pseudoranges) and carrier phases. These MDBs describe the size of outliers in the code data and the size of slips in the phase data which can just be detected by the appropriate test statistics. These MDBs will be given for three different GPS models, the geometry-free model and two variants of the geometry-based model.

INTRODUCTION

Minimal Detectable Biases (MDBs) as introduced by Baarda (1967, 1968) are important diagnostic tools for inferring the strength of model validation. As such they can also be used to study the strength of the various GPS positioning models, in particular with respect to the detectability of outliers in the code data or slips in the carrier phase data. The fact that a whole suit of different GPS models exists, implies that there are different stages at which quality control can be excercized. Roughly speaking, one can discriminate between the following four levels:

- *Receiver-level*: in principle it is already possible to validate the time-series of undifferenced data of a single GPS receiver. Single-receiver quality control is very useful for reference receivers that are used in active GPS control networks or in DGPS.
- *Baseline-level*: in this case a pair of receivers is used. When the observation equations are parametrized in terms of the baseline vector, the strength of the model primarily stems from the presence in the design matrix of the relative receiver-satellite geometry. Additional redundancy enters when the baseline is considered stationary instead of moving.

[1] Proceedings INSMAP98, Melbourne, FL, USA, Nov. 30-Dec. 4, 1998

- *Network-level*: when sufficient (independent) baselines are used to form a network, redundancy enters by enforcing the closure of the 'baseline loops', similar as the loops of a traditional levelling network.
- *Connection-level*: additional redundancy enters again when a free GPS network is connected to points of an existing control. In this case the redundancy stems from the fact that the shape of the free network is compared with the shape of the existing control.

In this contribution we will consider the baseline level. Three different such models will be considered, the geometry-free model and two variants of the geometry-based model, the roving- and stationary variant. Attention will be restricted to the 'short' baseline case. The term 'short' refers to the assumption that double-differenced (DD) observables are sufficient insensitive to orbital uncertainties in the fixed orbits and to residual ionospheric and tropospheric delays.

INTERNAL RELIABILITY

In this section a brief review is given of internal reliability. For more details see e.g. (Baarda, 1968) or (Teunissen, 1985). Reliability in the context of GPS is also treated in (Teunissen and Kleusberg, 1998), (Tiberius, 1998) and (de Jong, 1998). Internal reliability as represented by the MDBs describes the size of the model errors which can just be detected using appropriate test statistics. Consider the following null-hypothesis H_o and alternative hypothesis H_a:

$$H_o : E\{y\} = Ax \; and \; H_a : E\{y\} = Ax + b$$

with *E{.}* the expectation operator, *y* the *m*-vector of normally distributed observables, *A* the $m \times n$ design matrix, *x* the *n*-vector of unknown parameters, and *b* the unknown bias vector. The bias vector is assumed to describe the model error. Hence it is absent under the null-hypothesis, but present under the alternative hypothesis. It is further assumed that the bias vector can be parameterized as

$$b = c\nabla \; with \; c = known, \nabla = unknown$$

The vector *c* specifies the type of model error, while ∇ describes its unknown size. The (uniformly) most powerful test statistic for testing H_o against H_a is given as

$$T = \frac{(c^T Q_y^{-1} \hat{e})^2}{c^T Q_y^{-1} (I_m - P_A) c}$$

with Q_y the vc-matrix of the observables, $\hat{e}$ the least-squares residual vector and P_A the projector that projects orthogonally onto the range space of the design matrix. This test statistic has a Chi-squared distribution which is central under the null-hypothesis, but noncentral under the alternative hypothesis. The noncentrality parameter of the distribution is a 'yardstick' which measures the distance between the two hypotheses.

The noncentrality parameter is known once reference values for the level of significance (probability of type-I error) and the detection power (1 minus the probability of type-II error) are chosen. The corresponding reference value for the noncentrality parameter will be denoted as λ_o (example: $\lambda_o \approx 17$ when the level of significance equals 10^{-3} and the detection power 80%). Once the noncentality parameter is known, the corresponding size of the bias can be obtained as

$$|\nabla| = \sqrt{\frac{\lambda_o}{c^T Q_y^{-1}(I_m - P_A)c}}$$

This is the celebrated MDB. It is the minimal size of the bias that can just be detected with the test statistic T for the chosen level of significance and the chosen detecetion power. Apart from the noncentrality parameter, the MDB depends on the vector c, the design matrix A and the vc-matrix Q_y. Since c specifies the type of model error, different model errors will have different MDBs. In this contribution we will consider outliers in the code data and cycle slips in the phase data. This will be done for three different GPS models.

THREE DIFFERENT MODELS

The three different GPS model that will be considered are: the geometry-free model, the roving-receiver geometry-based model and the stationary-receiver geometry-based model.

The geometry-free model

The geometry-free model is the simplest GPS model one can think of. In this model the observation equations are not parameterized in terms of the baseline components. Instead, they remain parametrized in terms of the unknown double-differenced (DD) receiver-satellite ranges. This implies that the observation equations remain linear and that the receiver-satellite geometry is not explicitly present in these equations. Hence the model permits both receivers to be either stationary or moving. The geometry-free model has particularly been studied in the context of carrier phase ambiguity resolution, see e.g. (Hatch, 1982), (Euler and Goad, 1991), (Dedes and Goad, 1994), (Euler and Hatch, 1994), (Teunissen, 1996). For a single epoch and a single pair of satellites, the dual-frequency DD observation equations of this model are given as

$$\begin{aligned}
p_1(t) &= \rho(t) + n_{p1}(t) \\
p_2(t) &= \rho(t) + n_{p2}(t) \\
\phi_1(t) &= \rho(t) + \lambda_1 N_1 + n_{\phi 1}(t) \\
\phi_2(t) &= \rho(t) + \lambda_2 N_2 + n_{\phi 2}(t)
\end{aligned}$$

where ρ denotes the DD receiver-satellite range, λ the wavelength, N the carrier phase ambiguity and n the measurement noise. The first pair consists of the DD code observation equations on L1 and L2, while the second pair consists of the DD phase observation equations on these two frequencies. Thus when m satellites are tracked there are $2(m-1)$ measurements per frequency for each epoch. The redundancy of the model equals $(m-1)(3k-2)$ for the dual-frequency case and $(m-1)(k-1)$ for the single-frequency case, where k denotes the number of observation epochs. Thus in order to have redundancy, we need to track two or more satellites while using at least one epoch for the dual-frequency case or at least two epochs for the single-frequency case.

The roving-receiver geometry-based model

In case of the geometry-based model, the above observation equations are further parametrized (and linearized) in terms of the baseline components. As a consequence the relative receiver-satellite geometry enters the model. For the roving-receiver case, one will have an unknown baseline for each epoch.

For the single-freqeuncy case, the redundancy equals $(m-1)(2k-1)-3k$. Thus in order to have redundancy for a single epoch, more than four satellites need to be tracked. For the dual-frequency case, redundancy already exists for $m \geq 2$ when $k=1$. This is due to the presence of the dual frequency code data. When all baseline components are estimable, the dual-frequency redundancy equals $2(m-1)(2k-1)-3k$.

The stationary-receiver geometry-based model

When the two receivers are stationary, the k baselines of the previous model all collapse to one single baseline. As a consequence the redundancy increases by $3(k-1)$. The use of the geometry-based model for ambiguity resolution has also been studied by many, see e.g. (Frei and Beutler, 1990), (Hatch, 1991), (Teunissen et al., 1995), or (Tiberius and de Jonge, 1995).

In order to compute the MDBs one needs apart from the functional model (the observation equations), also the stochastic model (the vc-matrix of the GPS observables). In (Teunissen, 1998) the MDBs were given for a rather general stochastic model. Included were cross-correlation, satellite elevation dependence and a difference in measurement precision between the two frequencies. In this contribution these results will be simplified somewhat. Cross-correlation will be assumed absent and the measurement precision will be assumed independent of the frequency.

MINIMAL DETECTABLE OUTLIERS

In this section the MDBs for outliers in the code data will be given. In order to obtain tractable expressions, a number of approximations have been made. The first approximation consists of neglecting the phase-code variance ratio. In case of GPS, this ratio is of the order 10^{-4}. The second approximation consists of assuming the satellite elevation dependence of the stochastic model to be time-invariant, while the third approximation consists of replacing the time-varying relative receiver-satellite geometry by its time average. Since the receiver-satellite GPS configuration changes only slowly with time, these last two approximations are permitted in case the observation time spans are not too long.

The MDBs will be given for the single-differenced (SD) code observable to satellite i ($i=1,...,m$) based on k epochs of data. For the three GPS models of the previous section, they follow from (ibid) as

$$Geometry-free: \qquad |\nabla_p(i)| = \sigma_p(i)\sqrt{\frac{\lambda_o}{[1-\delta/k]\left[1-w_i\Big/\sum_{j=1}^{m} w_j\right]}}$$

$$Roving-receiver: \qquad |\nabla_p(i)| = \sigma_p(i)\sqrt{\frac{\lambda_o}{[1-\delta/k]\left[1-w_i\Big/\sum_{j=1}^{m} w_j\right]+[\delta/k][1-c_i^T P_G c_i]}}$$

$$Stationary-receiver: \qquad |\nabla_p(i)| = idem$$

with w_i the satellite elevation dependent weight, $\sigma_p(i)$ the SD standard deviation of the code observable to satellite i (it is inversely proportional to w_i), c_i the canonical unit vector having a 1 as its ith entry, P_G the orthogonal projector $P_G=G(G^TG)^{-1}G^T$ and matrix G the SD design matrix of order $m\times4$ which contains the time-averaged receiver-satellite geometry. The scalar δ equals 1 in the single-frequency case and ½ in the dual-frequency case.

Discussion

The above result shows that the MDB of the geometry-free model is the largest. This is due to the absence of the receiver-satellite geometry. The geometry-free model has therefore less strength than the other two models. The MDB of the roving-variant equals the MDB of the stationary-variant (of course within the approximations used). This implies that one's ability to detect outliers in the code data is practically independent of whether the baseline is stationary or not. The three MDBs will coincide in the absence of satellite redundancy. In that case m equals 4 and G becomes a square matrix, implying that the projector P_G reduces to the identity matrix.

The MDBs depend on the precision of the code data ($\sigma_p(i)$), the number of satellites tracked (m), the satellite elevation dependent weights (w_i) and through δ on the presence of a second frequency. They do not depend (again within the approximations used) on the precision of the phase data. The MDBs will get smaller when the code precision improves, when more satellites are tracked, when the elevation dependent weights get larger or when the number of epochs increases.

In the case of the geometry-based model, the dependence on the relative receiver-satellite geometry is felt through the projector P_G. This geometry has no contribution to the lowering of the MDB in case $c_i^T P_G c_i=1$. This happens in general when satellite redundancy is absent. It also happens however for certain receiver-satellite configurations. For instance, when all satellites except satellite i lie on a cone with its top at the receiver's location, while satellite i lies in a plane perpendicular to the symmetry axis of this cone.

To get an idea of some of the characteristics of the above MDBs, figure 1 shows the single-frequency MDBs of both the geometry-free model (full curves) and the geometry-based model (dotted curves, roving and stationary variant). An average receiver-satellite geometry was taken, implying that $c_i^T P_G c_i$ was set equal to $4/m$. Satellite elevation dependence was assumed absent and the precision of the SD code data was set at $\sigma_p(i)=\sqrt{2}\times 30$ cm. These assumptions and parameters will also be used for the other figures in this contribution. The MDBs are shown as function of the number of satellites tracked (m), for three different values of k (2, 5 and 50). The MDBs of the geometry-based model are clearly smaller than their counterparts of the geometry-free model. This difference gets smaller though when k gets larger.

MINIMAL DETECTABLE CYCLE SLIPS

In this section the MDBs for slips in the carrier phase data will be given. In order to obtain tractable expressions, the same type of approximations were made as before. Since a slip is a different type of model error than the spike-like outlier, we need to consider apart from the number of epochs used, also the duration of the slip. If $l \leq k$ equals the epoch when the slip starts to occur, then $N=k-l+1$ equals the duration of the slip. For $N=1$, the slip occurred at the last epoch.

The MDBs will be given for the single-differenced (SD) phase observable to satellite i ($i=1,...,m$) based on k epochs of data. For the three different GPS models, they follow from (ibid) as

$$Geometry-free: \quad |\nabla_\phi(i)| = \frac{\sigma_\phi(i)\sqrt{2}}{\sqrt{N}}\sqrt{\frac{\lambda_o}{[1-N/k]\left[1-w_i \Big/ \sum_{j=1}^{m} w_j\right]}}$$

$$\{for\ L1-only: \sigma_\phi(i)\sqrt{2} \xrightarrow{replace} \sigma_p(i)\}$$

$$Roving-receiver: \quad |\nabla_\phi(i)| = \frac{\sigma_\phi(i)}{\sqrt{N}}\sqrt{\frac{\lambda_o}{[1-N/k]\left[\left(1-w_i \Big/ \sum_{j=1}^{m} w_j\right)(1-\delta)+\delta\left(1-c_i^T P_G c_i\right)\right]}}$$

$$Stationary-receiver: \quad |\nabla_\phi(i)| = \frac{\sigma_\phi(i)}{\sqrt{N}}\sqrt{\frac{\lambda_o}{[1-N/k]\left[1-w_i \Big/ \sum_{j=1}^{m} w_j\right]}}$$

Discussion

Note that the MDBs become infinite in case $N=k$. This reflects the situation that cycle slips cannot be found when they already commence with the first epoch. In this case the slip cannot be separated from the corresponding ambiguity itself. Also note that the MDBs, when considered as function of N, obtain their minimum for $N=k/2$ (k=even), see also Figure 2. This reflects the situation that slips are best detectable when they occur halfway the observation time span. In real-time applications one will usually work with the case that N=1. In that case the testing for the cycle slip is done at the current epoch, namely k.

In all cases except one, the MDBs depend on the precision of the phase data and not on the precision of the code data (of course again within the approximations used). The exception occurs when single-frequency data is used in the geometry-free model. In this case it is not the high precision of the phase data that counts, but rather the poor precision of the code data. As a consequence, slips in the geometry-free model will be very difficult to detect when only single-frequency data are used.

When the MDB of the geometry-free model is compared with the MDB of the roving-receiver variant, we note that the presence of the receiver-satellite geometry makes the difference. The two MDBs are only equal when $c_i^T P_G c_i=1$. Interestingly enough the third MDB, the one of the stationary-variant, is independent of the receiver-satellite geometry. This is of course a consequence of the fact that we approximated the receiver-satellite geometry over the observation time span by its time average. Hence in the stationary case, for short time spans, it is not so much the geometric distribution of the satellites that counts, but more the number of satellites that are tracked.

REFERENCES

Baarda W (1967): *Statistical concepts in geodesy*. Netherlands Geodetic Commission, Publications on Geodesy, New Series, vol. 2, no. 4.

Baarda W (1968): *A testing procedure for use in geodetic networks*. Netherlands Geodetic Commission, Publications on Geodesy, New Series, vol. 2, no. 5.

Dedes G, Goad C (1994): Real-time cm-level GPS positioning of cutting blade and earth movement equipment. In: *Proc ION-94*, pp 587-594.

Euler HJ, Goad C (1991): On optimal filtering of GPS dual-frequency observations without using orbit information. *Bull Geod* 65: 130-143.

Euler HJ, Hatch R (1994): Comparison of several AROF kinematic techniques. In: *Proc ION-94*, pp 363-370.

Frei E, Beutler G (1990): Rapid static positioning based on the Fast Ambiguity Resolution Approach FARA: theory and first results. *Manuscr Geod* 15: 325-356.

Hatch R (1982): The synergism of GPS code and carrier measurements. In: *Proc 3rd Int Geod Symp Satellite Positioning*, vol. 2, pp 1213-1231.

Hatch R (1991): Instantaneous ambiguity resolution. In: Schwarz KP and Lachapelle G (eds): *Kinematic Systems in Geodesy, Surveying and Remote Sensing.* Springer, New York.

Jong, de CD (1998): Generation of DGPS/DGlonass corrections including quality control. In: these *INSMAP proceedings*.

Teunissen, PJG (1985): Quality control in geodetic networks. In: Grafarend EW and Sanso F (eds): *Optimization and design of geodetic networks*. Springer, Berlin Heidelberg New York, Chapt. 17.

Teunissen, PJG (1998): Minimal detectable biases of GPS data. *Journal of Geodesy*, 72: 236-244.

Teunissen PJG, de Jonge PJ, Tiberius CCJM (1995): The LAMBDA method for fast GPS surveying. In: *Proc Int Symp GPS technology applications*, pp 203-210.

Teunissen, PJG and Kleusberg A (eds) (1998): *GPS for Geodesy* (2nd edition), Springer, Berlin Heidelberg New York.

Tiberius CCJM, de Jonge PJ (1995): Fast positioning using the LAMBDA method. In: *Proc DSNS95*.

Tiberius CCJM (1998): *Recursive data processing for kinematic GPS surveying*. Netherlands Geodetic Commission, Publications on Geodesy, No. 45.

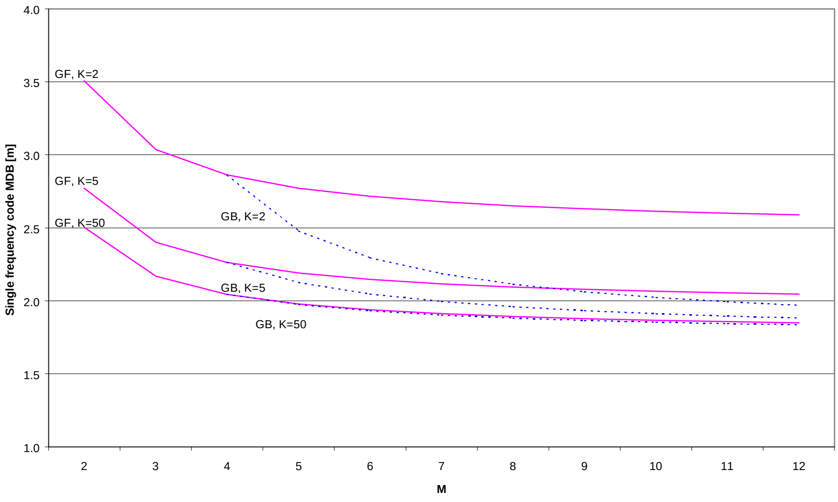

Figure 1: Single-frequency code outlier MDBs for the geometry-free (GF) model (full curves) and the geometry-based (GB) model (dotted curves) given as function of the number of satellites tracked (*m*). The three curves of each pair correspond with *k*=2, 5, 50.

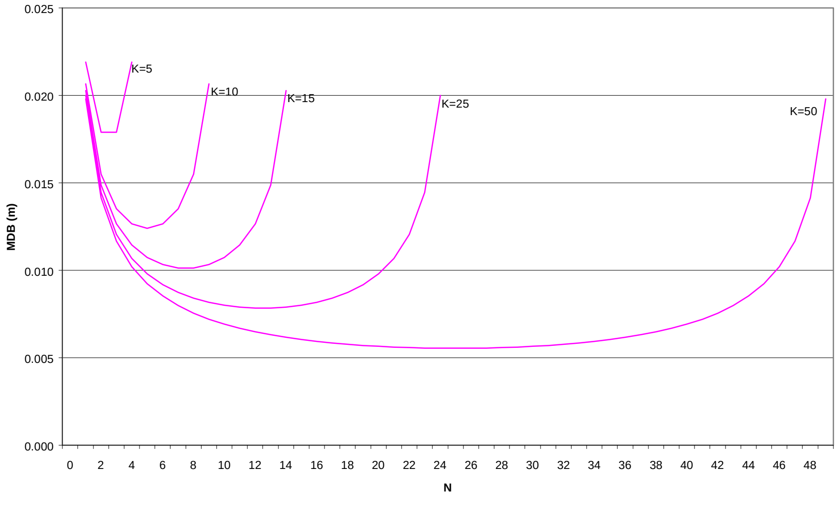

Figure 2: Carrier cycle slip MDBs for the stationary-receiver model given as function the window length *N* for *k*=5, 10, 15, 25, 50. The number of tracked satellites *m*=5.