Impact of stochastic modelling on GPS height and zenith wet delay estimation

Johnny Lo
Curtin University of Technology, Australia
Phone: +61 (0) 8 9266 2704; Email: johnny.lo@postgrad.curtin.edu.au

Ahmed El-Mowafy
Curtin University of Technology, Australia
Phone: +61 (0) 8 9266 2704; Email: A.El-Mowafy@.curtin.edu.au

Nigel Penna
Newcastle University, UK
Phone: +44 (0) 191 222 8747; Email: nigel.penna@ncl.ac.uk

Will Featherstone
Curtin University of Technology, Australia
Phone: +61 (0) 8 9266 2734; Email: w.featherstone@curtin.edu.au

ABSTRACT

Most stochastic modelling techniques assume the physical correlations among the raw observations to be negligible when forming the variance-covariance matrix of the GPS observations. Such an assumption may, however, lead to significantly biased solutions. The Minimum Norm Quadratic Unbiased Estimation (MINQUE) method is an iterative technique that can be used to estimate spatial correlation among GPS measurements. Studies by previous authors have shown that MINQUE improves the accuracy and the reliability of the ambiguity resolution, and ultimately, the geodetic solution. However, its effect on the estimation of zenith wet delay (*ZWD*) is somewhat unknown. In this paper, an investigation into its impact on *ZWD*, as well as heighting, is carried out using simulated data. The results obtained from MINQUE for an observation window of five-days in static mode indicate an average improvement of 51% and 71% in the station height precision when compared against elevation-angle dependent and equal weighting models, respectively. This development, however, did not translate into better *ZWD* estimation, for which the differences between each respective stochastic model are generally at the sub-millimetre level.

KEYWORDS: Stochastic model; MINQUE; ZWD; height

1.0   INTRODUCTION

GPS data processing can be implemented via the least-squares (LS) principle. In the LS process, GPS measurements are characterised by a functional model and a stochastic model. The functional model represents the mathematical relationship among the GPS observables and the parameters of interest, whilst the stochastic model is defined by an appropriate covariance matrix describing the spatial and/or temporal correlation among the measurements. The functional model is usually well defined (e.g., Hofmann-Wellenhof *et al.*, 2001; Leick, 2004) and is not particularly controversial. On the other hand, stochastic modelling remains one of the more challenging aspects in precise GPS positioning (e.g., Wang *et al.*, 2002), and there does not seem to be a clear consensus on it.

In LS theory, a set of linearised GPS observations can be defined using a Gauss Markov model as follows:

$$Y = AX + v \tag{1}$$

where Y is an $n \times 1$ observed-computed vector; X is a $k \times 1$ vector of unknown parameters with A being the corresponding $n \times k$ design matrix; v is an $n \times 1$ residual vector, with $E(v) = 0$ and $Var(v) = \sigma^2$.

The simple form for the *weighted* linear LS estimate for the unknown set of parameters, *X*, is given by (Johnson and Wichern, 2007):

$$\hat{X} = \left(A^T W A\right)^{-1} A^T W Y \tag{2}$$

The variance property of $\hat{X}$ can be expressed via the variance-covariance (*VCV*) matrix:

$$Cov\left(\hat{X}\right) = C_{\hat{X}} = \hat{\sigma}^2 \left(A^T W A\right)^{-1} \tag{3}$$

$$\hat{\sigma}^2 = \frac{\hat{v}^T W \hat{v}}{n-k} \tag{4}$$

and the optimal choice for the weight matrix W is defined as the inverse of the variance-covariance matrix of the model residuals (e.g., Johnson and Wichern, 2007), i.e., $VCV = Cov(v)$. The quantity $\hat{\sigma}^2$, often referred to as the *a posteriori unit variance* or *variance factor*, is an unbiased estimate of σ^2 and is an indicator of the precision of the observations and the assigned weight matrix *W*.

From equation (2), the choice of stochastic model is an important factor in determining the final outcome of the LS parameter solution. LS possesses an attractive property in that the residual root mean square error (*RMSE*) is minimised. However, an inadequately defined covariance matrix will result in LS losing its optimality property (Johnson and Wichern, 2007). Many of the existing stochastic models implemented in GPS data processing are simplified for practical purposes. For real-time kinematic (RTK) data processing, for example, where results are needed almost instantaneously, a wrongly chosen stochastic model may result in faulty cycle slip detection and degrade ambiguity resolution success. The quality of the parameter estimates of interest, such as receiver coordinates, will also suffer as a result (Fuller *et al.*, 2005).

Han and Rizos (1995) concluded that the LS-solved parameter estimates are always over-optimistic when independence is assumed among the observations. Jin *et al.* (2005) reported an offset of over 2cm in the height components of both the Darwin-Tidbinbilla (3046km) and Townsville-Tidbinbilla (1792km) 24-h baseline solutions. Subsequently, height errors will also impact tropospheric delay estimates used in meteorological and climatological applications. Employing the standard stochastic model (SSM) carries the assumption that all raw observations have the same degree of uncertainty, i.e., the same variance. Such an assumption is unrealistic as studies have shown that systematic errors caused by the atmosphere and multipath have varying degrees of impact on GPS signals (e.g., Barnes and Cross, 1998). It was also demonstrated with statistical testing on the LS residuals that the assumption of constant variances can be inappropriate (Bischoff *et al.*, 2005).

The *VCV* can also be estimated using an elevation-angle-dependent model (e.g., Kim and Langley, 2001) and the signal-to-noise ratio model (e.g., Lau and Cross, 2007). Although these models do somewhat reflect the quality of the observed GPS signals, correlations among the raw measurements are again ignored. Nevertheless, the elevation-angle-dependent model (EADM) for example, has been shown to produce reliable tropospheric estimates (e.g., Steigenberger *et al.*, 2007). Although more rigorous stochastic modelling techniques are available, (e.g., Wang *et al.*, 1998; Teunissen and Amiri-Simkooei, 2007), the complexity of these models generally demands more processing time. Additionally, these models have predominantly been used to derive positional and integer ambiguity estimates, and the effects on *ZWD* estimates are still relatively unknown. Though one may hypothesise that better coordinates would lead to better *ZWD* estimates, the significance of the impact is still speculative.

The above issue leads to the objective of this investigation, which is to determine if the estimation of *ZWD* will benefit from a more sophisticated stochastic model, namely the Minimum Norm Quadratic Unbiased Estimation 'MINQUE' (Rao, 1970). MINQUE was successfully applied in GPS data processing, where it was shown to improve short baseline solutions, as well as ambiguity resolution (Wang *et al.*, 1998). However, it has not yet been for the purpose of *ZWD* recovery. In this paper, the performance of MINQUE for both height and *ZWD* estimation will be compared against conventional stochastic models, i.e., the SSM and EADM, as well simplified MINQUE (Satirapod *et al.*, 2002) and the non-negative definite MINQUE (Rao and Kleffe, 1988). Descriptions of the aforementioned stochastic models are provided next.

2.0 STOCHASTIC MODEL

2.1 Standard Stochastic Model

The standard stochastic model (SSM) refers to the simplest of all stochastic models. SSM is constructed with the assumption that all zero-differenced (as considered in this paper) GPS observations are independent (i.e., zero correlation) and have the same variance, σ^2.

2.2 Elevation Angle Dependent Model (EADM)

The dependence of measurement noise on satellite elevation can be attributed to the receiver antenna's gain pattern, atmospheric refraction and multipath (e.g., Kim and Langley, 2001).

Modelling the observational noise as a function of the satellite elevation can take on many forms. One of these elevation angle-based models has the general form (Wang *et al.*, 1998):

$$\sigma_{\Phi_i}^2 = a^2 + b^2 f\left(\theta_z^i\right) \tag{6}$$

where a^2 and b^2 are constant coefficients and $f\left(\theta_z^i\right)$ is the function that is defined with respect to the zenith angle θ_z^i for observation i. The cosine function can be utilized to define the variances of the zero difference measurements in the form (Jin *et al.*, 2005):

$$\sigma_{\Phi_i}^2 = a^2 + b^2 \cos^2\left(\theta_z^i\right) \tag{7}$$

The coefficients a and b are simply given as 0 and 1. The raw observations are also assumed to be spatially and temporally uncorrelated in the EADMs.

2.3   Minimum Norm Quadratic Unbiased Estimation (MINQUE)

Using the Gauss-Markov model given in equation (1), the $m \times m$ VCV matrix of Y, can be expressed as:

$$\Sigma = \sum_{i=1}^{q} C_i = \sum_{i=1}^{q} \phi_i V_i \text{ , where } q = \frac{m(m+1)}{2} \tag{8}$$

where $\left\{\phi_1, \phi_2, ..., \phi_q\right\} = \left\{\sigma_1^2, \sigma_2^2, ..., \sigma_m^2, \sigma_{12}, \sigma_{13}, ..., \sigma_{m(m-1)}\right\}$ are the VCV components to be estimated and $V_1, V_2, .., V_q$ are the so-called *accompanying matrices* (Wang *et al.*, 2002). The problem here is estimating the q unknown elements of Σ.

The MINQUE of the linear function ϕ_i ($i = 1, 2, .., q$), i.e., $p_1\phi_1 + p_2\phi_2 + .. + p_q\phi_q$, is the quadratic function $Y^T A Y$, where A is selected such that (Rao , 1971):

$$Trace(A\Sigma A\Sigma) \text{ is a minimum: subject to } AX = 0 \text{ and } Trace(AC_i) = p_i, \ i = 1, 2, .., q \tag{9}$$

The MINQUE of $\sum_{i=1}^{q} \phi_i V_i$ is then estimated from:

$$\gamma^T Q = \sum_{i=1}^{q} \gamma_i Q_i \ , \qquad Q_i = Y^T R V_i R Y \tag{10}$$

where the vector γ is a solution of

$$\sum_{i=1}^{q} \gamma_i Trace(RV_i RV_j) = p_j, \quad j = 1, 2, .., q \tag{11}$$

and

$$R = W (W^1 - X (X^T W X) X^T) W, \text{ where } W = \Sigma^{-1} \tag{12}$$

The symmetric R matrix can be partitioned as

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1s} \\ R_{21} & R_{22} & \ddots & R_{2s} \\ \vdots & \ddots & \ddots & \vdots \\ R_{s1} & R_{s2} & \cdots & R_{ss} \end{bmatrix} \tag{13}$$

where s is the number of epochs considered in a selected processing window session.

By expressing equation (11) as $S\gamma = p$, where

$$S_{ij} = Trace(RV_i RV_j) \tag{14}$$

this leads to $\gamma = S^{-1}p$.

Since the MINQUE of $\sum_{i=1}^{q} \phi_i V_i$ is

$$\gamma^T Q = p^T (S^{-1})^T Q = p^T S^{-1} Q = p^T \hat{\phi} \tag{15}$$

then $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, .., \hat{\phi}_q)$ is a solution of

$$S\phi = Q \tag{16}$$

Q can alternatively be defined as:

$$Q_i = Y^T RV_i RY = \varepsilon^T WV_i W\varepsilon \tag{17}$$

Given an initial estimate $\hat{\phi}_{(0)}$, the $(j+1)^{th}$ approximation can be generated using the following iterative procedure:

$$\hat{\phi}_{(j+1)} = S_{(j)}^{-1} Q_{(j)}, \, j = 0, 1, 2, \dots \tag{18}$$

2.4   Simplified MINQUE

The execution of MINQUE requires a computer processor with substantial power and memory as the number of observations becomes large. This is mostly due to the computation and storage of the R matrix, i.e., equation (13). The notion behind the simplified MINQUE (Satirapod *et al.*, 2002), which will be referred to here as SMINQUE, is to reduce the complexity of the R matrix, leading to the efficient computation of the MINQUE process. The proposed simplification of MINQUE disregards the off-diagonal block entries of the R

matrix and gives rise to a block-diagonal matrix $R*$ as its replacement in the procedure. The $R*$ matrix is expressed as

$$R* = \begin{bmatrix} R_{11} & 0 & \cdots & 0 \\ 0 & R_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & R_{ss} \end{bmatrix} \tag{19}$$

Subsequently, equation (14) can be simplified to

$$S_{ij} = \sum_{r=1}^{s} Trace\left(R_{rr}V_{ir}R_{rr}V_{jr}\right) \tag{20}$$

where V_{ir} is the block-diagonal element of V_i for epoch r and $V_{ir} = V_{jr}$.

2.5   Non-Negative Definite MINQUE (NND-MINQUE)

If all the matrices V_1, V_2, .., V_q are non-negative definite, an alternative iterative scheme can be applied (Rao and Kleffe,1988),

$$\hat{\phi}_{(j+1)}^i = \hat{\phi}_j^i \frac{Y^T R_{\hat{\phi}_i} V_i R_{\hat{\phi}_i}^T Y}{Trace\left(R_{\hat{\phi}_i} V_i\right)}, \quad i = 1, \ 2, \ .., \ q \tag{21}$$

where the $(j+1)^{th}$ approximation to the ith component of $\hat{\phi}$ can be computed. The non-negative definite MINQUE (NND_MINQUE) scheme of equation (21) ensures $\hat{\phi}^i$ will remain non-negative throughout the iterations when $\hat{\phi}_{(0)}^i \geq 0$. Using the NND_MINQUE is also computationally simpler than applying equation (18), which requires the calculation of the S matrix. In this study, only the variances are estimated.

3.0   SIMULATION DATA

To test the above MINQUE methods, a set of simulated data was processed. For a given location, a set of observations is generated by simulating the satellite coordinates for a specified session, using the Penna and Stewart (2003) 'perfect' GPS orbital simulator, i.e. with no perturbations. Therefore, for a particular location, the exact geometric range between the satellite and the receiver can be calculated, and the effect of different individual error sources readily assessed.

The GPS simulation software used can perform weighting by using the standard or the EADM models. However the user can easily incorporate other weighting schemes if needed. Tropospheric data, and/or any other error sources (e.g., multipath or random errors) determined through external functions, can also be accommodated. In this test, tropospheric delays were added to the Penna and Stewart (2003) line-of-sight- geometric ranges using simulated *ZWD* (*SZWD*) and the Niell mapping function (Niell, 1996). The S*ZWD* values

applied were those generated from 24 hour GIPSY version 2.6 software precise point positioning mode analyses, estimating them every 5 min together with horizontal gradients, whilst holding fixed 'legacy' JPL (Jet Propulsion Laboratory) 'fiducial-free' orbital and Earth rotation products, and using the Niell mapping function.

SZWD values were generated for the HOB2 IGS station from 1999 to 2004.. A five-day period in 2004 from June 15th to 19th was chosen for testing as there was a large *ZWD* variability over these five days, ranging from 5mm to 16mm. The elevation cut-off was selected as 15 degrees. Weighted one-hourly LS *ZWD*s estimates were retrieved in the analyses and compared to the simulated values, i.e., the S*ZWD*s, across three different processing windows, selected as 1h, 2h and 4h) and using the five different stochastic methods discussed in Sections 2.1-2.5.

In discussing the results, MINQUE and SMINQUE will be referred to collectively as (S)MINQUE. The most important aspect of this simulation analysis is that no errors were applied to the observations besides tropospheric delay, i.e. the only present "error source" is the variability within the simulated *ZWD*s themselves. If the stochastic model is correctly chosen, one would expect the coordinate correction estimates to be approximately zero and the output *ZWD* estimates to be similar to the (averaged) simulated values for the processing window considered.

4.0   RESULTS

The height component of the coordinates is the main positioning component that is affected by atmospheric delay (e.g., Bock *et al.*, 2001). As such, the height estimates resulting from the study were closely analysed. The *RMSE* values of the height estimates resulted from the above methods over the five-day simulation are presented in Table 1.

Window Size	SSM	EADM	MINQUE	SMINQUE	NND_MINQUE
1-h	5.35	3.90	1.84	1.94	8.57
2-h	2.59	1.35	0.86	0.94	2.90
4-h	1.17	0.66	0.23	0.25	1.03

Table 1 RMSE of the height estimates (mm) for HOB2 over the five-day data set using different stochastic methods.

Over the three different window sizes, both MINQUE and SMINQUE consistently produced the smallest height offsets. The NND_MINQUE was the worst performer over the five-day period. This is somewhat unexpected given that the variance factor, (given in equation (4)), of the linear model produced by the NND_MINQUE model is unity. Thus theoretically, the model should perform fairly well. Since this is not the case, the underlying notion here signifies the importance of proper modelling of the correlation among the observations.

Table 2 compares the MINQUE's *RMSE* in the height component to those of the other stochastic models. The relative improvement (*RI*) of MINQUE over the other models is calculated as follows:

$$RI = \frac{RMSE_i - RMSE_{MINQUE}}{RMSE_i} * 100,$$ (24)

where

i ={SSM, EADM, SMINQUE, NND_MINQUE}

The advantage of MINQUE over the other models that ignore spatial correlation among the raw observations is fairly substantial. The average improvements made by MINQUE are 71%, 51%, and 76% when compared to SSM, EADM and NND_MINQUE, respectively. However, the difference between MINQUE and SMINQUE is marginal.

Window Size	SSM	EADM	MINQUE	SMINQUE	NND_MINQUE
1-h	66%	53%	-	5%	79%
2-h	67%	36%	-	9%	70%
4-h	80%	65%	-	8%	78%

Table 2   Relative improvement in the height estimates for HOB2 over the five-day data set as a result of using MINQUE.

Unexpectedly however, better height recovery did not yield better *ZWD* estimates. Figure 1 illustrates that all models produced better *ZWD* estimates as the size of the processing window increases. The EADM, however, was the best stochastic model over all window sizes in the recovery of the wet delay estimates. The EADM recovered the *ZWD* with a better accuracy than SSM, MINQUE, SMINQUE and NND_MINQUE, by an average of 23%, 20%, 20%, and 30%, respectively.
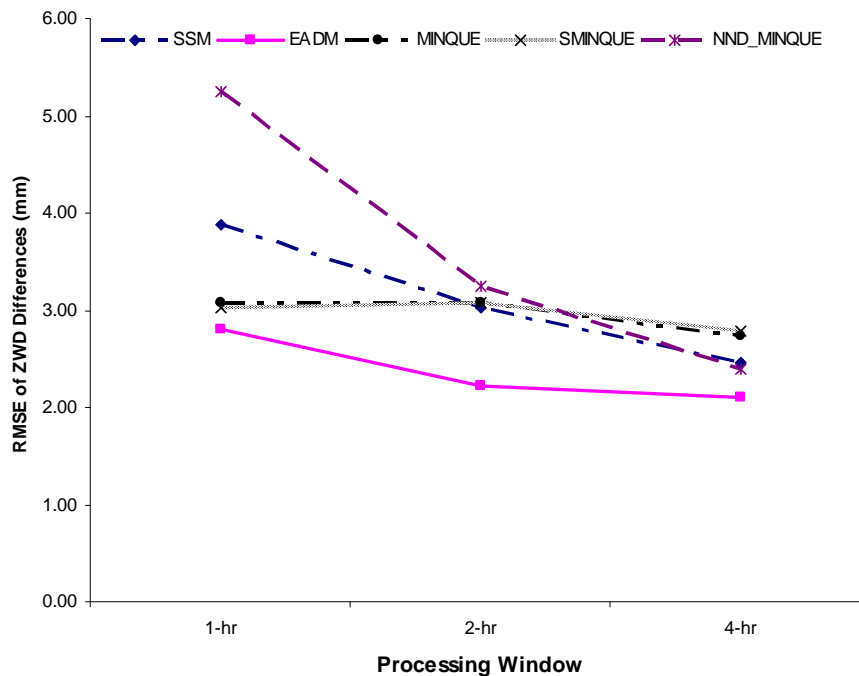

Figure 1   Comparison between the *RMSEs* of the (LS-simulated) *ZWD* differences for HOB2 for each of the stochastic model over various processing window sizes, for the five-day data set considered.

When estimating the zenith wet delay, GPS observations that are closer to the zenith are more appropriate than those at low elevations, and therefore should have greater weights. The EADM is reflective of this, and it may be a possible explanation as to why it had out-performed (S)MINQUE. Thus, the solutions may have been biased in favour of the EADM, as well as the other models where the correlations are ignored. Further investigation is planned.

5.0 CONCLUSIONS

Changing the stochastic model affects the GPS estimation of heights and *ZWD*. The results attained with MINQUE (via the modified approach) for the HOB2 IGS station from five-days of simulated data in static mode indicate an average improvement of 51% and 71% in the station height precision, when compared to the EADM and SSM, respectively. This superiority was not evident in the recovery of the *SZWD*. In fact, EADM recovered the *SZWD* better than the other models (SSM, MINQUE, SMINQUE, NND_MINQUE) by an average of 23%, 20%, 20% and 30%, respectively, across all window sizes. The dependence of Niell (1996) mapping function on the elevation angle could also flatter the results of the EADM. Nevertheless, further investigations are required. Future analysis may possibly involve independent *ZWD* estimates, e.g., from water vapour radiometers, and using real GPS data.

REFERENCES

Barnes, J.B., and P.A. Cross. 1998. Processing models for very high accuracy GPS positioning. *Journal of Navigation* 51 (2):180-193

Bischoff, W., B. Heck, J. Howind, and A. Teusch. 2005. A procedure for testing the assumption of homoscedasticity in least squares residuals: a case study of GPS carrier-phase observations. *Journal of Geodesy* 78:197-404, doi: 10.1007/s00190-004-0390-5.

Bock, O., J. Tarniewicz, C. Thom, J. Pelon, and M. Kasser. 2001. Study of external path delay correction techniques for high accuracy height determination with GPS. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy* 26:165-171, doi:10.1016/S1464-1895(01)00041-2

Fuller, S., P. Collier, and A. Kealy. 2005. Real time quality assessment for CORS networks. *Journal of Global Positioning Systems* 4 (1-2):223-229

Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins. 2001. *GPS theory and practice*. 5th ed. New York: Springer-Verlag Wien, 382pp.

Jin, S., J. Wang, and P.H. Park. 2005. An improvement of GPS height estimations: stochastic modeling. *Earth Planets Space* 57:253-259

Johnson, R.A., and D.W. Wichern. 2007. *Applied multivariate statistical analysis*. 6th ed. London, UK: Prentice-Hall, 800pp.

Kim, D., and R.B. Langley. 2001. Quality control techniques and issues in GPS applications: Stochastic modeling and reliability test. In *International Symposium on GPS/GNSS (the 8th GNSS Workshop)*. Jeju Island, Japan.

Lau, L., and P. Cross. 2007. Investigations into phase multipath mitigation techniques for high precision positioning in difficult environments. *Journal of Navigation* 60:457-482 doi:10.1017/S0373463307004341.

Leick, A. 2004. *GPS Satellite Surveying*. 3rd ed. New Jersey, USA: John Wiley & Sons, Inc., 435pp.

Niell, A.E. 1996. Global mapping functions for the atmosphere delay at radio wavelenghts. *Journal of Geophysical Research* 101 (B2):3227-3246

Penna, N,T,, and M.P. Stewart, 2003. Aliased tidal signatures in continuous GPS height time series, *Geophysical Research Letters* 30(23):2184, doi:10.1029/2003GL018828.

Rao, C.R. 1970. Estimation of heterogeneous variances in linear models. *Journal of American Statistical Association* 65:161-172

Rao, C.R., and J. Kleffe. 1988. *Estimation of variance components and applications*. Netherlands: Elsevier Science Publishers B.V., 361pp.

Satirapod, C., J. Wang, and C. Rizos. 2002. A simplified MINQUE procedure for the estimation of variance-covariance components of GPS observables. *Survey Review* 36 (286):582–590

Steigenberger, P, V. Tesmer, M. Krügel, D. Thaller, R. Schmid, S. Vey, and M. Rothacher. 2007. Comparisons of homogeneously reprocessed GPS and VLBI long time-series of troposphere zenith delays and gradients *Journal of Geodesy* 81 (6-8):503-514, doi: 10.1007/s00190-006-0124-y.

Teunissen, P.J.G., and A.R. Amiri-Simkooei. 2007. Least-squares variance component estimation *Journal of Geodesy* 82 (2):65-82, doi:10.1007/s00190-007-0157-x.

Wang, J., C. Satirapod, and C. Rizos. 2002. Stochastic assessment of GPS carrier phase measurements for precise static relative positioning. *Journal of Geodesy* 76 (2):95-104, doi:10.1007/s00190-001-0225-6.

Wang, J., M.P. Stewart, and M Tsakiri. 1998. Stochastic modeling for static GPS baseline data processing. *Journal of Surveying Engineering* 124 (4):171-181