

COMPARING DIFFERENT GPS DATA PROCESSING TECHNIQUES FOR MODELLING RESIDUAL SYSTEMATIC ERRORS

Chalermchon Satirapod¹, Jinling Wang² and Chris Rizos³

Abstract: In the case of traditional GPS data processing algorithms, systematic errors in GPS measurements cannot be eliminated completely, nor accounted for satisfactorily. These systematic errors can have a significant effect on both the ambiguity resolution process and the GPS positioning results. Hence this is a potentially critical problem for high precision GPS positioning applications. It is therefore necessary to develop an appropriate data processing algorithm which can effectively deal with systematic errors in a non-deterministic manner. Recently several approaches have been suggested to mitigate the impact of systematic errors on GPS positioning results: the semi-parametric model, the use of wavelets and new stochastic modelling techniques. These approaches use different bases and have different implications for data processing. This paper aims to compare the above three methods, in both the theoretical and numerical sense.

¹Lecturer, Department of Survey Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, 10330, Thailand. Email: chalermchon.s@chula.ac.th

²Lecturer, School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, NSW, 2052, Australia. Email: jinling.wang@unsw.edu.au

³Professor, School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, NSW, 2052, Australia. Email: c.rizos@unsw.edu.au

KEYWORD TERMS: Global Positioning System (GPS), Iterative stochastic modelling procedure, Wavelet-based approach, Semi-parametric and penalised least squares technique, Systematic error.

INTRODUCTION

Since its introduction to civilian users in the early 1980's GPS has been playing an increasingly important role in high precision surveying and geodetic applications. Like traditional geodetic network adjustment, data processing for precise GPS static positioning is invariably performed using the least squares method. To employ the least squares method for GPS relative positioning, both the functional and stochastic models of the GPS measurements need to be defined. The functional model describes the mathematical relationship between the GPS observations and the unknown parameters, while the stochastic model describes the statistical characteristics of the GPS observations (see, for example, Leick, 1995). The stochastic model is dependent on the choice of the functional model. However, to achieve optimal results *both* the functional model and the stochastic model have to be correctly defined. A double-differencing technique is commonly used for constructing the functional model as it can eliminate many of the troublesome GPS biases, such as the atmospheric delays, the receiver and satellite clock biases, and orbit error. In current stochastic models it is usually assumed that all the one-way measurements have equal variance, and that they are statistically independent. Such functional and stochastic models form the basis of standard GPS data processing algorithms.

However, by the application of such GPS data processing algorithms, systematic errors in GPS measurements cannot be eliminated completely, nor accounted for satisfactorily. These systematic errors can have a significant effect on both the ambiguity resolution process and the final GPS positioning results. This is a potentially critical problem for high precision GPS positioning applications, and hence it is necessary to develop an appropriate data processing algorithm which can effectively deal with systematic errors in a non-deterministic manner. Several approaches have recently been proposed to mitigate the impact of systematic errors on GPS positioning results. Examples of such techniques include the semi-parametric and penalised least squares technique introduced by Jia et al. (2000), the iterative stochastic modelling procedure proposed by Wang et al. (2002), and the wavelet-based approach suggested by Satirapod et al. (2001). In this paper the authors will compare the above three techniques, from both the theoretical and numerical points of view. In the following sections a description of each data

processing technique is presented. Numerical examples are presented next, followed by an analysis of the results. Concluding remarks and some recommendations are given in the last section.

DATA PROCESSING TECHNIQUES

Data processing techniques considered in this section can be categorised as follows:

- i) changes made to the functional model, referred to as *the semi-parametric and penalised least squares technique*
- ii) changes made to the stochastic model, referred to as *the iterative stochastic modelling procedure*, and
- iii) changes made to both the functional and stochastic models, referred to as *the wavelet-based approach*.

Semi-parametric and penalised least squares techniques

Since parametric models do not always provide an adequate description of the observations there is a need to develop a model that can do so. In the field of physical geodesy, Moritz (1989) discussed the use of the semi-parametric model to fit the local irregularities of the terrestrial gravity field. Recently Jia et al. (2000) have proposed the semi-parametric model and penalised least squares technique to enhance the processing of GPS data. The basic idea of the semi-parametric model is to treat the estimated parameters in two parts: the parametric part and the non-parametric part. The parametric part is defined by a vector of parameters $x = [x_1, \dots, x_p]^T$ with p denoting the number of unknown parameters. The non-parametric part is defined by a set of smoothing functions (g_i) varying slowly with time. The semi-parametric model can be expressed as

$$l_i = A_i x + g_i + u_i, \quad i = 1, 2, \dots, n \quad (1)$$

with

$$l_i = [l_i(1), l_i(2), \dots, l_i(s)]^T;$$

$$A_i = [A_i(1), A_i(2), \dots, A_i(s)]^T;$$

$$g_i = [g_i(1), g_i(2), \dots, g_i(s)]^T; \text{ and}$$

$$u_i = [u_i(1), u_i(2), \dots, u_i(s)]^T.$$

where l_i is a $s \times 1$ vector of the observed-minus-computed double-differenced (DD) carrier phase values; A_i is the design matrix corresponding to the measurements l_i ; u_i is a $s \times 1$ vector of random errors; n is the number of satellite pairs forming the DD observables; and s is the number of observation epochs.

It can be seen from equation (1) that the total number of unknown parameters, in both the parametric and non-parametric parts, is larger than the number of the GPS observations. Hence these unknown parameters cannot be estimated using the standard least squares technique. For this reason the *penalised* least squares technique has been introduced. This technique uses a cubic spline to fit the non-parametric part, in order that the equation system becomes solvable. The estimated parameters are not only influenced by the goodness-of-fit to the data, but also by the roughness of the functions. The Generalised Cross Validation (GCV) term is used to estimate the best fit to the data as well as the smoothness of the functions. Figure 1 is an example of using the semi-parametric model to fit the DD residuals for one satellite pair. The solid line represents the DD residuals and the dashed line represents the estimated function. Details of this technique and some experimental results can be found in Jia et al. (2000) and Jia (2000). In the subsequent sections this technique is simply referred to as the *semi-parametric method*.

Iterative stochastic modelling procedure

In the iterative stochastic modelling procedure, the unmodelled systematic errors will be appropriately described in the stochastic model. However, because the inclusion of the systematic errors into the stochastic model will make the variance-covariance matrix become fully populated, the computational load can rapidly become excessive. The *iterative stochastic modelling procedure* has been proposed by Wang et al. (2002) in order to avoid this problem. The procedure is based on a linear classical least squares process, with the basic idea being to first transform the DD carrier phase measurements into a new set of measurements using estimated temporal correlation coefficients.

The standard observation equation for processing of GPS data is:

$$l_i = A_i x + e_i, \quad i = 1, 2, \dots, n \quad (2)$$

with $e_i = [e_i(1), e_i(2), \dots, e_i(s)]^T$.

Where e_i is a $s \times 1$ vector of the error terms for the measurements \bar{l}_i ; and A_i and l_i are the same as in equation (1).

According to Ibid (2002), equation (2) can be transformed to:

$$\bar{l}_i = \bar{A}_i x + u_i \quad (3)$$

where $\bar{l}_i = Gl_i$, $\bar{A}_i = GA_i$. The structure of the matrix G can be found in Ibid (2002).

After the measurements are transformed they can then be considered to be free of *temporal correlations*, and thus have a block diagonal variance-covariance matrix structure (Wang, 1998). Consequently, the immense memory requirements and computational load for the inversion of a fully populated variance-covariance matrix can be avoided, and the variance-covariance matrix for the transformed measurements can be estimated using a rigorous statistical technique such as the MINQUE method (Rao, 1971). An iterative process is performed until sufficient accuracy is achieved. Details of this technique and some experimental results can be found in Wang et al. (2002). In subsequent sections, for the sake of brevity, this technique is referred to as the *iterative procedure*.

Wavelet-based approach

The Wavelet Transform (WT) is a new tool for signal analysis that can provide simultaneously time *and* frequency information of a signal sequence. WT has many potential applications in filtering, subband coding, data compression and multi-resolution signal processing (see, for example, Chui, 1992; Wickerhauser, 1994). WT can also be used for the analysis of *non-stationary* signals. It is therefore very useful for analysing GPS measurements. Fu & Rizos (1997) have outlined some of the applications of wavelets to GPS data processing. GPS bias effects such as multipath and ionospheric delay behave like low-frequency noise, while random measurement errors are typically characterised as high-frequency noise. The Wavelet Transform can be used to achieve enough frequency resolution to discriminate these terms in the original GPS measurement. In view of this advantage, a wavelet-based approach has been proposed by Satirapod et al. (2001) to improve the efficiency of the conventional data processing technique. In the wavelet-based approach, the DD residuals are first decomposed into a low-frequency bias and

high-frequency noise terms using a wavelet decomposition technique. Figure 2 shows an example of signal extraction using wavelets. The solid line represents the DD residuals and the dashed line represents an extracted bias component. The extracted bias component is then applied directly to the GPS measurements to correct for the trend introduced by this bias. The remaining terms, largely characterised by the GPS range measurements and high-frequency measurement noise, are expected to give the best linear unbiased solutions from a least squares process. The simplified MINQUE (Satirapod et al, 2002) is applied in order to formulate the stochastic model. Details of this approach and some experimental results can be found in Satirapod et al. (2001).

NUMERICAL EXAMPLES

In this section results derived from two observed GPS data sets are presented to demonstrate the effectiveness of the three methods.

Data acquisition

The first data set was collected on 18 December 1996 using two Trimble 4000SSE receivers at a sampling interval of 1 second. A 40-minute span of data was extracted from the original data set and resampled every 15 seconds. Six satellites (PRNs 7, 14, 15, 16, 18, and 29) were selected, as they were visible during the entire span. The reference values of each baseline component derived from the 40-minute session are used to verify the accuracy of the results. The data were first processed using the standard GPS data processing technique in order to check the data quality. In the data processing step, PRN15 (the satellite with the highest elevation angle) was selected as the reference satellite to form the double-differenced observables. The DD residuals for various satellite pairs are shown in Figure 3. They indicate the presence of systematic errors for most of the satellite pairs.

The second data set was collected on 7 June 1999 using two Ashtech Z-XII receivers at a sampling interval of 1 second. The receivers were mounted on pillars that are part of a first-order terrestrial survey network. The known baseline length between the two pillars is $215.929 \pm 0.001\text{m}$. This was used as the ground truth to verify the accuracy of the results. A 30-minute span of data was extracted from the original data set and resampled every 15 seconds. Six satellites (PRNs 2, 7, 10, 13, 19, and 27) were selected, as they were visible during the entire observation period. Again, all data were first processed using the standard GPS data processing technique in order to check the data quality. PRN2 (the satellite with the highest elevation angle) was selected as the reference satellite in

the data processing step. The DD residuals are shown in Figure 4. The DD residuals indicate the presence of some significant multipath errors for satellite pairs PRN 2-7 and 2-19.

Data processing step

For convenience, the Trimble data set was divided into four batches and the Ashtech data set was divided into three batches, each of ten minutes length. Each batch was treated as an individual session and processed using the following four techniques:

- A. The standard GPS processing technique, with the simplified stochastic model (which includes only mathematical correlation), and assuming that temporal correlations are absent.
- B. The semi-parametric method.
- C. The iterative stochastic modelling method.
- D. The wavelet-based method.

It should be noted that only the L1 carrier phase data was used in these studies. The results obtained using the above data processing strategies are discussed in the next section.

ANALYSIS OF RESULTS

The results obtained from the different processing strategies described in the previous section have been analysed from four points of view: ambiguity resolution, accuracy of estimated baseline results, *a-posteriori* standard deviations, and computational efficiency.

Ambiguity resolution

For reliable ambiguity resolution the difference between the best and second-best ambiguity combination is crucial for the ambiguity discrimination step. The F-ratio is commonly used as the ambiguity discrimination statistic, and the larger the F-ratio value, the more reliable is assumed the ambiguity resolution. The critical value of the F-ratio is normally (arbitrarily) chosen to be 2.0 (e.g. Euler & Landau, 1992). The ambiguity validation test can also be based on an alternative statistic, the so-called W-ratio (Wang et al., 1998). In a similar fashion, the larger the W-ratio value, the more reliable the ambiguity resolution is assumed. The values of these statistics, as obtained from both sets of data, are shown in Figures 5 and 6. In both figures, each subplot represents a data set, and each group of

columns (I, II, III, IV - four in the case of the Trimble data, three in the case of the Ashtech data) represents the solution obtained from an individual observation session. The colours represent the results applying the different processing strategies.

Clearly from Figures 5 and 6, the F-ratio and W-ratio values obtained from the semi-parametric and the wavelet-based methods are larger compared to those obtained from the standard and iterative methods. This is mainly due to the removal of nonlinear trends from the data.

Accuracy of estimated baseline results

In the case of the estimated baseline components, the differences in each component between the estimated values and the reference values obtained from the Trimble data set are presented in Table 1. Table 2 presents the difference in baseline length between the estimated value and ground truth value obtained from the Ashtech data set. The results show that the iterative and the wavelet-based methods produce more reliable estimated baseline results than the standard and the semi-parametric methods. It can also be seen that the semi-parametric method produces results which are essentially identical to those obtained from the standard method.

In addition, it can be seen from Table 1 that the maximum difference in the height component between sessions for the Trimble data set is up to 27.6mm when the standard and the semi-parametric methods are used. This is reduced to 13.7mm when the iterative method is applied, and to 5.9mm in the case of the wavelet-based approach. Similarly, the maximum difference in the height component between sessions for the Ashtech data set is up to 19.3mm when the standard and the semi-parametric methods are used. This is reduced to 7.8mm when the iterative method is applied, and to 9.3mm in the case of the wavelet-based approach.

***A-posteriori* standard deviations**

In the case of the *a-posteriori* standard deviations of the baseline components, the results obtained from both data sets are summarised in Table 3. It can be seen that both the semi-parametric and the wavelet-based methods produced smaller standard deviations compared to those obtained from the standard method. In general the iterative method produced larger standard deviations than those obtained from the semi-parametric and the wavelet-based methods. In view of the fact that the iterative method tends to produce reliable estimated baseline results, this

method therefore gives a more meaningful quality measure of the baseline results compared to those obtained the semi-parametric and wavelet-based methods.

Computational efficiency

Table 4 is a summary of the computational time and memory requirements for processing a 10-min batch of GPS data (six satellites tracked, 15-second sampling rate). The computational time required for the semi-parametric method is significantly larger than those required for the other methods. This is mainly due to the process of determining the best GCV score, carried out using a grid search method. The computational time is also dependent on the volume of the search space. A practical procedure needs to be developed to reduce the volume of this search space in order to reduce the computational load. Note also that the computational time required especially for the semi-parametric method would dramatically increase as the number of tracked satellites increases.

CONCLUDING REMARKS AND SOME RECOMMENDATIONS

In this paper different GPS data processing algorithms that have been proposed to account for systematic errors in the observations have been reviewed. Two data sets were used to compare the results obtained from these different algorithms. Based on the numerical results obtained, the following tentative comments and recommendations can be made:

- The semi-parametric method has shown significant improvement in ambiguity resolution. However, there is no improvement in the accuracy of the estimated baseline results and the computational load is still a major limitation in implementing this method. The reader should note, however, that a quality measure such as the co-factor matrix of the estimated parameters (e.g. for use as a validation criterium in ambiguity resolution) is still available using this technique.
- The iterative stochastic modelling method has demonstrated an improvement in the accuracy of the estimated baseline results, and has produced the most realistic estimates of the accuracy of the baseline results. However, this method gave a relatively poor performance for ambiguity resolution. The method is therefore perhaps most suitable for high accuracy applications with long observation periods.
- Although the wavelet-based approach produced over-optimistic estimates of the accuracy of the baseline results, this method seems to give the best performance in terms of ambiguity resolution, the achievable baseline

accuracy and the computational efficiency. However, as in the case of the semi-parametric method, the derivation of the quality measure for the wavelet-based approach is a challenge.

Further investigations should therefore be carried out in order to derive the quality measure for the wavelet and semi-parametric approaches. In the opinion of the authors, based on experience in data processing, the performance of single-baseline GPS data processing is largely dependent on the magnitude of systematic errors and the satellite geometry. In practice it is impossible to *completely* eliminate the systematic errors from GPS data using any data processing strategy unless there is external information on the biases (for example, from the correction terms generated from a network of GPS reference stations, as in recent network-based positioning techniques).

ACKNOWLEDGEMENTS

Much gratitude to Mr. Minghai Jia for his suggestions on the implementation of the semi-parametric method. Special thanks also to Mr. Clement Ogaja and Mr. Michael Moore for helpful discussions with the authors. Part of the results in this paper are based on a paper presented at the IAG Scientific Assembly, Budapest, Hungary, 3-8 September 2001.

APPENDIX I. REFERENCES

- Chui C.K. (1992). *An Introduction to Wavelets*. Academic Press, Inc., Boston, 264pp.
- Euler H.J. & H. Landau (1992). Fast GPS ambiguity resolution on-the-fly for real-time applications. *6th Int. Symp. on Satellite Positioning*, Columbus, Ohio, 17-20 March, 650-659.
- Fu W.X. & C. Rizos (1997). The applications of wavelets to GPS signal processing. *10th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation*, Kansas City, Missouri, 16-19 September, 697-703.
- Jia M., M. Tsakiri & M.P. Stewart (2000). Mitigating multipath errors using semi-parametric models for high precision static positioning. *Geodesy Beyond 2000 - The Challenges of the First Decade, IAG Symposia*, 121, 393-398.
- Jia M. (2000). Mitigation of systematic errors of GPS positioning using vector semiparametric models. *13th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation*, Salt Lake City, Utah, 11-14 September, 1938-1947.

- Leick A. (1995). *GPS Satellite Surveying*, John Wiley & Sons, Inc., New York, N.Y., 560pp.
- Moritz H. (1989). *Advanced physical geodesy*. 2nd edition, Wichmann, Karlsruhe, 500pp.
- Rao C.R. (1971). Estimation of variance and covariance components - MINQUE. *Journal of Multivariate Analysis*, 1, 257-275.
- Satirapod C. (2001). Improving the accuracy of static GPS positioning with a new stochastic modelling procedure. Pres. *14th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation*, Salt Lake City, Utah, 11-14 September.
- Satirapod C., C. Ogaja, J. Wang & C. Rizos (2001). GPS analysis with the aid of wavelets. *5th Int. Symp. on Satellite Navigation Technology & Applications*, Canberra, Australia, 24-27 July, paper 39, CD-ROM Proc.
- Satirapod C., J. Wang & C. Rizos (2002). A simplified MINQUE procedure for the estimation of variance-covariance components of GPS observables. *Survey Review* (in press).
- Wang J. (1998). Stochastic assessment of the GPS measurements for precise positioning. *11th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation*, Nashville, Tennessee, 15-18 September, 81-89.
- Wang J., M.P. Stewart & M. Tsakiri (1998). A discrimination test procedure for ambiguity resolution on-the-fly. *Journal of Geodesy*, 72, 644-653.
- Wang J., C. Satirapod & C. Rizos (2002). Stochastic assessment of GPS carrier phase measurements for precise static relative positioning. *Journal of Geodesy*, 76(2), 95-104.
- Wickerhauser M. (1994). *Adapted Wavelet Analysis from Theory to Software*. AK Peters Ltd., Wellesley, Massachusetts, 486pp.

APPENDIX II. NOTATION

The following symbols are used in this paper:

- x = a vector of parameters;
- g_i = a set of smoothing functions;
- l_i = a $s \times 1$ vector of the observed-minus-computed double-differenced (DD) carrier phase values;
- A_i = the design matrix corresponding to the measurements l_i ;
- u_i = a $s \times 1$ vector of random errors;

n	=	the number of satellite pairs forming the DD observables;
s	=	the number of observation epochs;
e_i	=	a $s \times 1$ vector of the error terms for the measurements l_i ;
G	=	the transformation matrix;
\bar{l}_i	=	a $s \times 1$ vector of the transformed observed-minus-computed DD carrier phase values;
\bar{A}_i	=	the transformed design matrix.

Table captions

Table 1: The differences between estimated baseline components and the reference values for the Trimble data set.

Table 2: The differences between estimated baseline length and the ground truth value for the Ashtech data set.

Table 3: Comparison of standard deviations of the estimated baseline components.

Table 4: Comparison of computational time and memory usage.

Method	Session	Difference in each component		
		dN (mm)	dE(mm)	dH(mm)
A	I	-4.5	8.1	8.7
	II	0.5	0.0	4.5
	III	2.0	-5.4	-8.5
	IV	2.7	-8.6	-18.9
B	I	-4.5	8.1	8.7
	II	0.5	0.0	4.5
	III	2.0	-5.4	-8.5
	IV	2.7	-8.6	-18.9
C	I	0.3	2.3	-4.5
	II	0.5	-0.5	3.8
	III	1.7	-5.0	-7.7
	IV	5.2	-3.5	-9.9
D	I	-0.4	1.8	-2.0
	II	3.0	-2.1	0.6
	III	1.8	-2.6	-3.7
	IV	1.3	-1.2	-5.3

Method	Session	Different in baseline length (mm)
A	I	2.8
	II	2.5
	III	7.3
B	I	2.8
	II	2.5
	III	7.3
C	I	0.3
	II	1.4
	III	1.7
D	I	1.4
	II	1.2
	III	1.4

Data set	Method	Session	Standard deviations (mm)		
			North	East	Height
Trimble	A	I	2.5	0.8	3.2
		II	1.6	0.5	2.2
		III	1.5	0.5	2.4
		IV	2.3	0.7	4.0
	B	I	1.3	0.4	1.7
		II	0.9	0.3	1.2
		III	1.0	0.3	1.6
		IV	1.6	0.5	2.8
	C	I	1.8	1.3	2.4
		II	1.4	1.4	2.0
		III	2.7	0.9	4.8
		IV	5.5	2.3	9.2
	D	I	0.4	0.4	1.3
		II	0.2	0.3	0.6
		III	0.7	0.6	1.5
		IV	0.5	0.7	1.3
Ashtech	A	I	0.9	0.8	1.4
		II	0.9	0.8	1.3
		III	1.4	1.3	2.2
	B	I	0.6	0.5	0.9
		II	0.3	0.3	0.5
		III	0.8	0.7	1.2
	C	I	0.8	1.0	1.8
		II	0.3	0.4	0.7
		III	0.5	0.6	1.0
	D	I	0.3	0.4	0.4
		II	0.2	0.2	0.5
		III	0.5	0.6	0.8

Method	Computational time (seconds)	Memory Usage (Kbytes)
A	41.2	<i>None</i>
B	>12 hours	312.5
C	758.8	312.5
D	49.9	7.8

**All solutions computed using Matlab GPS baseline software on an Intel Celeron 466MHz Processor*

Figure captions

Fig. 1: Fitting the DD residuals using the semi-parametric model.

Fig. 2: Signal extraction using wavelets.

Fig. 3: DD residuals obtained for the Trimble data set of 18.12.1996.

Fig. 4: DD residuals obtained for the Ashtech data set of 7.6.1999.

Fig. 5: Comparison of the F-ratio value in ambiguity validation tests.

Fig. 6: Comparison of the W-ratio value in ambiguity validation tests.







