An Empirical Stochastic Model for GPS

R.F. Leandro, M.C. Santos
Department of Geodesy and Geomatics Engineering,
University of New Brunswick, P.O. Box 4400, Fredericton, New Brunswick, Canada, E3B 5A3

Abstract. In addition to functional models, stochastic modeling of observations plays an important role in GPS data processing. The stochastic model has influence over several issues of coordinates determination with GPS, such as the covariance matrix of the observations (which leads to weighting scheme) and the estimated covariance matrix of the parameters. In this paper we present an empirical stochastic approach to create observation covariance matrices for GPS. Our approach aims to a more realistic and complete information about the stochastic behavior of GPS observations and an improvement in quality and quality control of estimated coordinates.
In the empirical approach the correlation functions and variances are computed using the observed data, instead. Therefore it is not necessary any a-priori assumption linking observables variances and correlations to elevation angle or time lags, usually given by formal models. The first step of this approach is a functional reduction of the observables, which is made according to the functional model used in the late adjustment. The objective of this first step is leading the observation time series to stationarity. An auto-regressive model is then adjusted, with the determination of its parameters and order. Parameters are estimated using least squares adjustment and the order determined by an assessment of residuals. Once all parameters of the stochastic model were determined empirically, it is used to create the observation covariance matrix to be used in the functional model.
In the present case study, GPS baselines were processed. Analyses were made in terms of obtained coordinates and their estimated covariance matrices, aiming at an improvement in GPS quality control. Improvements of at least 11 % in precision and accuracy were found when this empirical stochastic approach was used. Future research aims to the enhancement of the method until it is general enough to be used in any case of GPS data processing.

Keywords. Empirical, Stochastic, GPS.

1 Introduction

The covariance matrix of observations plays a fundamental role in GPS data adjustment. Usually the method of least squares is used to compute the coordinates of a given receiver. In the adjustment the covariance matrix of observations drives how each one of the observations will contribute for the update of the parameters. This matrix includes not just variances, but also a relation between all observations, the covariances. A usual approach to estimate the covariance matrix of observations is to set a weight for all observations and then, in case of baseline, propagate it using the double difference operator. It is not assumed any correlation between observations made in two different epochs, the so called autocorrelations. The weights can be set based on different types of information, such as elevation angle of the satellite or signal-to-noise ratio. Sometimes even the identity matrix (equally weighted observations) is used.
In this work, we are presenting an empirical approach to build the covariance matrix of observations. The main objective of such technique is populating the covariance matrix with realistic information, estimated by means of a stochastic analysis of the raw data. Trying to have as much as possible information inside the covariance matrix is a way to carry into the adjustment model a realistic picture of the quality and behavior of the observations. Eventually, the coordinates will be adjusted using this matrix. The ultimate goal of this work is to improve the quality of GPS data processing, as well as quality of the estimated precision for coordinates.
In order to analyze the data assuming a stochastic behavior a few requirements need to be satisfied, such as stationarity. Because of that, the GPS data need to undergo some modifications before it can be used. In our Empirical Stochastic model (which will be called herein with ESto model) all analysis is made based on raw data, without any external information, before the adjustment.
In next section a background explanation about stochastic processes and their analysis will be made. Next the treatment given to GPS data in order to

make its analysis as a stochastic process possible will be shown. A case study was carried out and its results are shown, as well as conclusions and recommendations for future work.

2 Stochastic Processes

A time series is a series of observations made through time. An important feature of a series is that, usually, the observations made at subsequent epochs are dependent on each other. The analysis of a time series is based on this dependency.

A stochastic process is a statistical phenomena which occurs through time according to probability laws, [Box et. al., 1994]. A time series can be considered as a realization of a stochastic process.

There is a particular process, called a stationary process, which has a particular statistic equilibrium state. A stochastic process is called widely stationary when its properties remain unaffected when the time origin is changed. This means that the joint probability function of a process with n observations $z_{t_1}, z_{t_2}, ..., z_{t_n}$, observed at time instants $t_1, t_2, ..., t_n$ is the same as a process with n observations $z_{t_1+k}, z_{t_2+k}, ..., z_{t_n+k}$, observed at time instants $t_1+k, t_2+k, ..., t_n+k$ for every integer k.

If a process is stationary, we can say that:

$$\rho(z_t) = \rho(z_{t+k}) \; \forall \; k \geq 0, t \geq 0 , \qquad (1)$$

where $\rho(z_t)$ is the probability density function, and because of that, we can conclude that the mean of a stationary process is constant:

$$\mu_t = E(z_t) = \int_{-\infty}^{\infty} z \cdot \rho(z) dz , \qquad (2)$$

and therefore:

$$\mu_t = \mu_{t+k} , \qquad (3)$$

for every integer k. The mean of a process can be estimated with the equation:

$$\bar{z} = \frac{1}{n} \cdot \sum_{t=1}^{n} z_t . \qquad (4)$$

And the variance of a process can be defined as:

$$s^2 = E[(z_t - \bar{z})^2] = \int_{-\infty}^{\infty} (z - \bar{z})^2 \rho(z) dz . \qquad (5)$$

The variance of a process can be estimated with the following equation:

$$s^2 = \frac{1}{n} \cdot \sum_{t=1}^{n} (z_t - \bar{z})^2 . \qquad (6)$$

The stationarity of a process implies that the joint probability function $\rho(z_{t1}, z_{t2})$ is the same for any instants $t1$ and $t2$, and for the same time interval between them. This means that we can estimate the joint probability function of a process for different time intervals k. The covariance between the values z_t and z_{t+k}, separated by a time interval k, which is constant for every t for a stationary process is called the auto-covariance function, and can be defined by:

$$\gamma_k = \mathrm{cov}(z_t, z_{t+k}) = E[(z_t - \bar{z})(z_{t+k} - \bar{z})] . \qquad (7)$$

It is possible to estimate the auto-correlation function of a process for a given lag k with the following equation:

$$\rho_k = \frac{E[(z_t - \bar{z})(z_{t+k} - \bar{z})]}{\sqrt{E[(z_t - \bar{z})^2] \cdot E[(z_{t+k} - \bar{z})^2]}} = \sum_{t=1}^{n-k} \frac{[(z_t - \bar{z})(z_{t+k} - \bar{z})]}{s^2} . \qquad (8)$$

Based on this equations and the assumption behind them, the variances for GPS observables are computed in the ESto model, which will be explained in next session.

3 ESto Model

In ESto model, the previously described approaches are used to estimate the correlations between different observations (made at different time and/or from different satellite pairs), as well as the variances for each one of the observed satellites. The input of the model is the raw data, and the output is the stochastic parameters.

As stated in the previous section, a stochastic process needs to have certain characteristics in order to be used for estimation of variances and covariances. Since the objective of this work is to analyze GPS data as a stochastic process, it is necessary to modify the original time series of observations to satisfy the assumptions made in stochastic analysis. The main assumption that needs to be satisfied is the stationarity of the process. It is clear that GPS observations are not stationary, because they are related to the distances between satellite and receivers (plus errors and biases) and always vary with time. Figure 1 shows an example of double differenced pseudoranges through time.

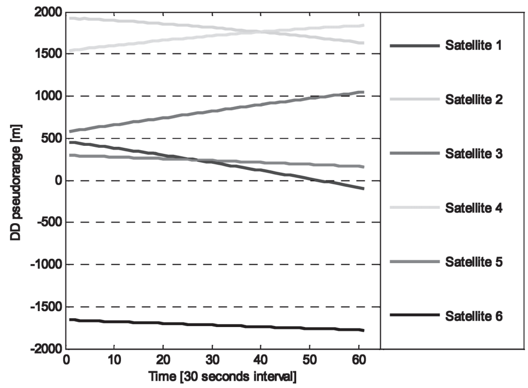

Fig. 1 Double differenced pseudoranges time series.

These observations could not be used directly in a stochastic analysis, and therefore need to be modified. In order to get a stationary series from the measurements, they are reduced by using the approximate geometric distances between receiver and satellite antennas, according to:

$$\Delta\nabla P_{red} = \Delta\nabla P_{obs} - \Delta\nabla\rho \, , \tag{9}$$

where $\Delta\nabla P_{obs}$ is the double differenced observed pseudorange, $\Delta\nabla\rho$ is the double differenced geometric distance between receiver and satellite antennas and $\Delta\nabla P_{red}$ is the double differenced reduced pseudorange. This simple reduction can be used in this case because, as it is going to be mentioned later, we are dealing with short length baselines, where the most of errors and biases are supposedly eliminated. Figure 2 shows the double differenced C/A pseudoranges before and after the reduction made in the ESto model, for each one of the satellites shown above.

In Figure 2 below, the light lines represent the original series, and the dark lines represent the reduced series. After this step the reduced series are used in the stochastic analysis. This part of the ESto model is based on the approaches mentioned in the previous section.

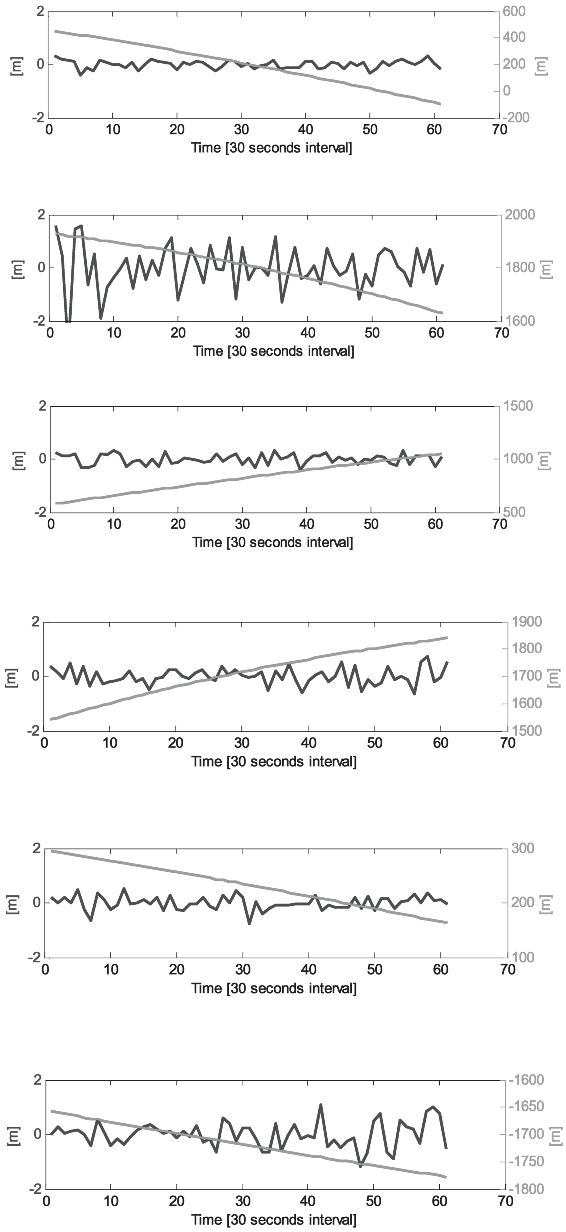

Fig. 2 Reduction of DD pseudoranges, in ESto model.

As an illustration of the stochastic part of the model, Figure 3 shows the auto-correlation function estimated with ESto model and also computed based on the adjustment residuals, for satellites with different elevation angles (65, 31 and 17 degrees). The observables are C/A pseudoranges, observed for a short baseline (5 km) during half hour. The sampling interval s 30 seconds, thus one epoch lag represents 30 seconds in time.

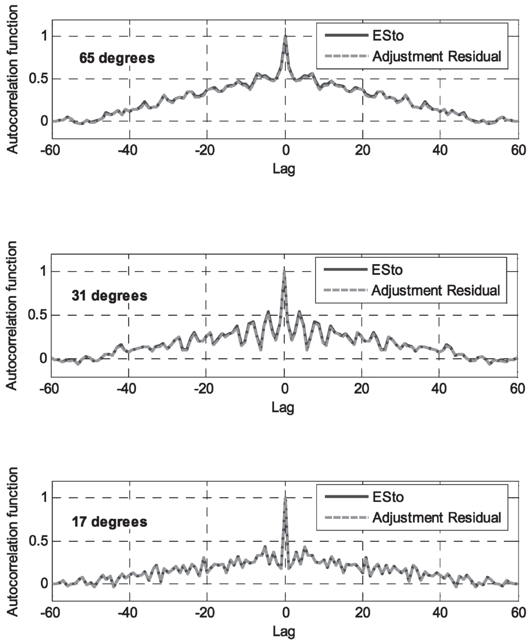

Fig 3 Auto-correlation function for satellites with different elevation angles.

ESto can provide a good approximation of the residuals auto-correlation and therefore for the observations. It can be also noticed how the correlations get smaller as the elevation angle gets lower, with the autocorrelation function getting closer to zero with smaller lags.

Another issue is the variance which is assigned to each satellite. In ESto model, as said previously, it is also computed by analyzing the raw data. Figure 4 shows the pseudorange standard deviations estimated by ESto for different elevation angles.

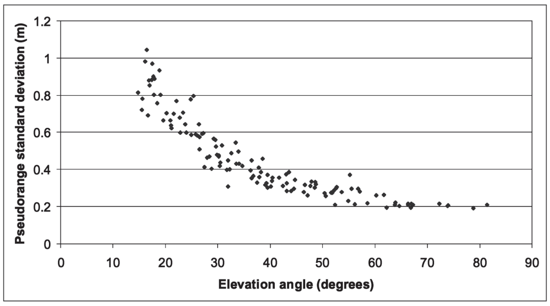

Fig. 4 Estimated standard deviation for different elevation angles.

The standard deviation varies with the elevation angle with values between around 0.2 m and 1 m. In Figure 5, it is shown the agreement between the estimated standard deviations and the residuals standard deviation after the adjustment.

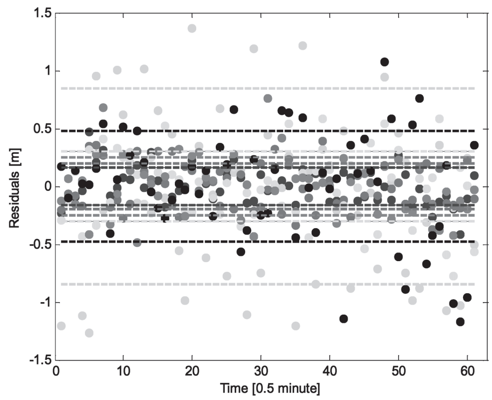

Fig. 5 Standard deviations estimated with ESto and adjustment residuals.

In Figure 5, the dots represent the residuals of each of the observations, during a half hour observation session with 30 seconds of sampling interval (7 satellites observed). The dashed lines represent the one sigma standard deviations estimated with ESto. For each satellite, the color of the lines and dots is the same. Table 1 shows the estimated values for the standard deviations.

Table 1. Comparison of residuals tandard deviation with ESto estimated standard deviations.

Satellite – elev. angle	Standard dev. [m]
1 (blue) – 65^o	0.15
2 (green) – 17^o	0.84
3 (red) – 52^o	0.19
4 (cyan) – 44^o	0.30
5 (magenta) – 45^o	0.25
6 (black) – 31^o	0.48

ESto model has been developed in order to allow a estimation of a fully populated observation covariance matrices for GPS. It also potentially allows a more complete analysis of the raw data before the adjustment, such as outliers detection. This type of pre-analysis is not explored in this work, however it is another potential contribution brought by the use of our empirical approach. In terms of validation, ESto has been compared with other techniques to build covariance matrices. The first one, called here Formal DD, the simplest of the ones explored in this work, is a matrix generated by propagating a identity matrix with the double difference operator, thus all satellites have the same weight. The second approach uses the elevation angle as a parameter to estimate the standard deviation of the observations (here this approach is being called Elevation based). After the standard deviations are estimated, they are propagated using the double difference operator as well. Figure 6 shows a representation of covariance matrices estimated using these three schemes.

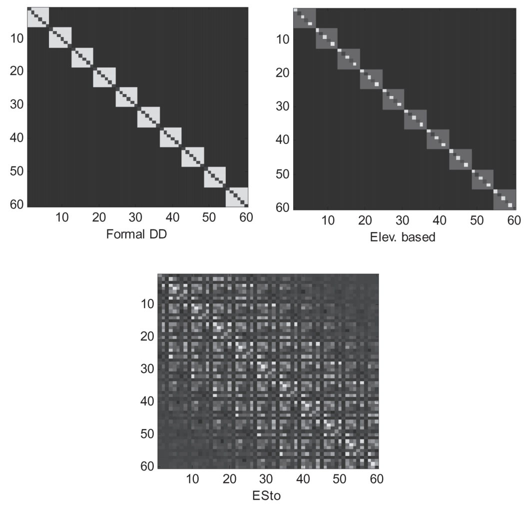

Fig 6 Representation of covariance matrices estimated using different schemes.

In the figure above, different colors represent different values within the matrices. Each one of them has its own scale, and they were generated from the same data set. In the case of Formal DD, it can be noticed that the elements on the diagonal have always the same color, what means the same variance. The green squares around the diagonal are the correlation between different satellite pairs within the same epoch. This correlation exists with this pattern due to the use of a common reference satellite. In the second case (Elev. based), the main difference from the previous in the different weights for different satellites (this can be noticed by the different colors in the diagonal). The blue squares are due to the same as the green ones in the previous. The dark blue means zero in both of them. When the ESto model is used, it can be noticed that a large amount of information is placed into the covariance matrix. These values were derived from raw data, and represent the correlation of each observation with respect to each other. The next step to validate the ESto model should be an analysis with real data processing, explored in next session.

4 Data Processing

In order to investigate the efficiency of this empirical stochastic approach with respect to the conventional techniques, it was used to process a short baseline using C/A pseudoranges. The choice of such data set is justified by the elimination of several errors in the measurements, such as clock errors and atmospheric refraction, which would need a special treatment in the stochastic analysis. It was used 24 hours of data, which were processed in half hour batch adjustments along the day. It was done for the baseline, UNB1-FRDN with approximately 5 km. It were used the C/A pseudoranges as observables. Figure 7 shows the results of the adjustment. In the plots below, zero is the published coordinate of FRDN station.

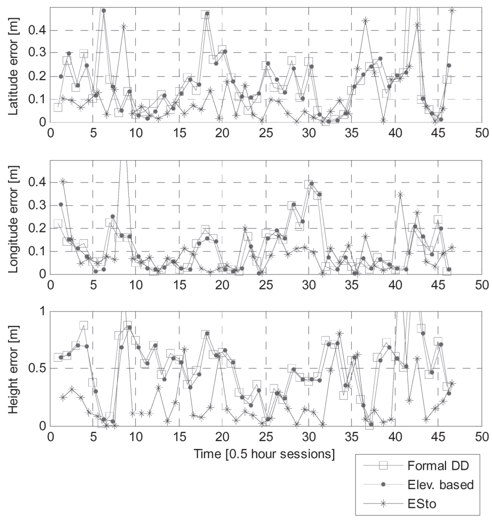

Fig. 7. Results for the adjustment of the baseline UNB1-FRDN.

Although all results are fairly good, when ESto was used, the mean bias was quite lower than other solutions. However the standard deviation didn't show a good improvement. The use of ESto shows some errors quite larger than their mean values, which can be affecting standard deviation values. Table 2 shows the results summary for this data processing.

Table 2. Comparison of results.

		Formal DD	Elev. based	ESto
Mean error (m)	Latitude	0.14	0.13	0.07
	Longitude	-0.08	-0.08	-0.04
	Height	0.52	0.50	0.15
RMS (m)	Latitude	0.24	0.23	0.16
	Longitude	0.14	0.15	0.16
	Height	0.60	0.57	0.39

Tables 3 and 4 show the absolute and relative improvements achieved by the use of ESto with respect to the other models.

Table 3. Absolute improvements when using ESto.

		w.r.t. Formal DD	w.r.t. Elev. based
Mean error (m)	Latitude	0.07	0.06
	Longitude	0.04	0.04
	Height	0.37	0.35
RMS (m)	Latitude	0.08	0.07
	Longitude	-0.02	-0.01
	Height	0.21	0.18

Table 4. Relative improvements when using ESto.

		w.r.t. Formal DD	w.r.t. Elev. based
Mean error	Latitude	50%	46%
	Longitude	50%	50%
	Height	71%	70%
RMS	Latitude	33%	30%
	Longitude	14%	7%
	Height	35%	46%

ESto model provided improvements ranging from 4cm to 37cm in bias and -2cm to 21cm in RMS. The greatest improvements were achieved for height determination. In terms of relative results, the improvements are very good, ranging from 46% to 71% for mean bias and -7% to 46% for RMS. These results show that the use of the empirical approach can definitively bring advantages in terms of quality of the estimated coordinates. Figure 8 shows the solutions in the horizontal position for the three models.

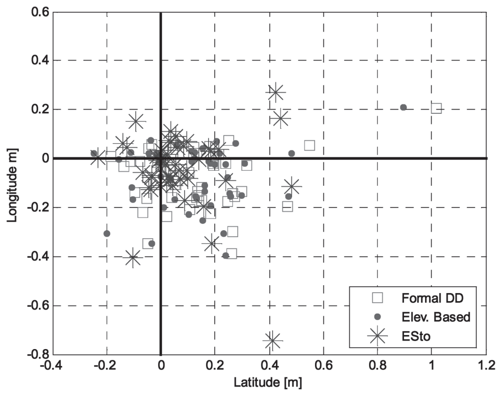

Fig. 8. Estimated horizontal coordinates for FRDN station.

It can be noticed that in the plot above the coordinates estimated with ESto are generally grouped closer to (0,0), than the solutions provided with the other techniques.

Further analysis in terms of estimated coordinate variances was also carried out. Figures 9, 10 and 11 show the error bars for the three models (latitude, longitude and height, respectively) at 1 sigma level.

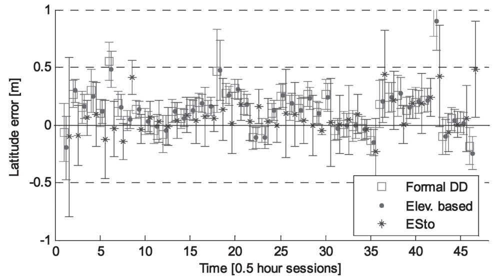

Fig. 9. Estimated standard deviations for latitude at 1 sigma level.

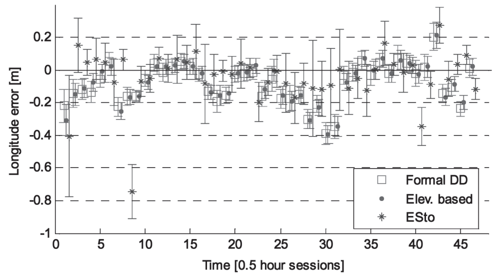

Fig. 10. Estimated standard deviations for longitude at 1 sigma level.

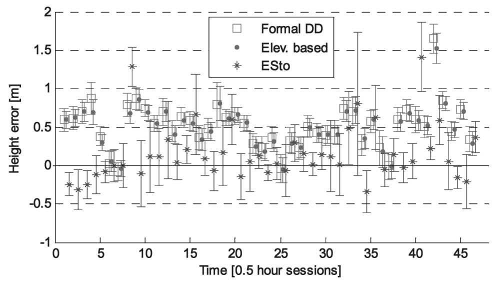

Fig. 11. Estimated standard deviations for height at 1 sigma level.

Clearly when ESto model is used, the estimated standard deviations are more realistic. Table 5 shows the percentage of the times when the real error was less than the estimated standard deviation (1 sigma) with the three weighting schemes:

Table 5. Comparison of estimated standard deviations (1 sigma) and real errors.

	Formal DD	Elev. based	ESto
Latitude	30 %	28 %	85 %
Longitude	46 %	39 %	78 %
Height	11 %	11 %	65 %

According to Table 5, the use of the empirical approach provides a better agreement between the estimated variances and the real errors, what means that in this case, the covariance matrix of the parameters computed in the adjustment is more trustful. The agreement for the two other models was quite lower than the supposed probability at 1 sigma level (68 %), while with ESto the agreement was higher than this value, except for height, when the agreement was of 65 %, which is still several times closer than for the other models (both 11 %).

5 Conclusions and Future Work

It was shown that the empirical model is capable of providing a good stochastic modeling of observables without any external source of information, before any adjustment.

The implementation of the model is computationally efficient, and can be easily implemented for GPS data processing. Although its use makes the process slower, it is not a great computational cost when compared with the gain in terms of the adjustment results. The greater time is due the computation of auto-correlation and cross-correlation functions that involve several combinations in order to have almost fully populated covariance matrices.

A baseline was processed using three different weighting schemes for observations. In general, when ESto was used a good improvement in the bias was obtained. The mean biases were at least 46 % smaller. The RMS values did not improve as well as the biases, with one case of degradation (-8 % for longitude), but still had relatively god improvements with values up to 46 %.

The greater improvements were achieved for height determination, with improvements in the order of 70% in mean bias for both comparisons.

In terms of standard deviation, the solutions obtained with ESto showed small or no improvements. There are still some outliers in the solution when ESto is used, what means that there is room for improvement in this sense. An attenuation of those outliers would help to make the standard deviations lower.

In terms of estimated standard deviation, ESto provided much more realistic estimates, with agreements (compared against the real errors) ranging between 65 % and 85 %. For the other two stochastic models the estimates had a lower agreement with the real errors, with values ranging from 11 % to 46 %.

The model certainly brings improvement to the solution, and future investigation is needed to investigate the source of solution outliers and improvements in the stochastic analysis for longer baselines.

Further research to apply ESto to point positioning and carrier phase measurements is also a desirable future step in the development and validation of the model.

6 References

Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. (1994). Time Series Analysis – Forecasting and Control. Prentice-Hall International, London.

Collins, J. P. and R. B. Langley (1999). Possible weighting schemes for GPS carrier phase observations in the presence of multipath. Final contract report for the U.S. Army Corps of engineers Topographic Engineering Center, No DAAH04-96-C-0086 / TCN 98151, March, 33pp.

El-Rabbany, A. (1994). The effect of physical correlations on the ambiguity resolution and accuracy estimation in GPS differential positioning. Ph.D. Thesis - University of New Brunswick, Department of Geodesy and Geomatics Engineering, Fredericton, NB.

Hofmann-Wellenhof, B., H. Lichteneger and J. Collins (2001). Global Positioning System: Theory and Practice. Springer-Verlag Wien New York.

Kim, D and R. Langley (2001). Estimation of the Stochastic Model for Long-Baseline Kinematic GPS Applications. ION National Technical Meeting, January 22-24, 2001, Long Beach-CA.

Langley, R. B. (1997). GPS observation noise. GPS World, Vol. 8, No. 6, April, pp 40-45.

Vanicek, P. and E. Krakiwsky (1982). Geodesy: The Concepts. North-Holland Publishing Company, Amsterdan – New York – Oxford.