J Geod (2006) 80: 199–203
DOI 10.1007/s00190-006-0044-x

ORIGINAL ARTICLE

M. Ge · G. Gendt · G. Dick · F. P. Zhang
M. Rothacher

A new data processing strategy for huge GNSS global networks

Received: 7 October 2005 / Accepted: 20 March 2006 / Published online: 4 May 2006

Abstract In Global Positioning System (GPS) data analyses, large networks are usually divided into sub-networks to solve the conflict between increasing amounts of data and limited computer resources, although an integrated analysis would provide better results. This conflict becomes even more critical with the increasing number of stations, and low-Earth-orbiting satellites and the Galileo system coming into operation. The major reason is that a huge number of ambiguity parameters are kept in the normal equation for sequential integer ambiguity fixing. In this paper, the problem is solved by a special procedure of parameter elimination for both real-valued and ambiguity-fixed solutions, based on an adapted ambiguity-fixing approach where the covariance-matrix of ambiguity parameters is not required anymore. It is demonstrated that, with the new strategy, the required memory can be reduced to one-tenth and the computation time to at least one-third compared to the existing methods, and huge GPS networks with several hundred stations can be processed efficiently on a personal computer.

Keywords Global Navigation Satellite Systems (GNSS) · Integrated estimation · Huge global network · Parameter elimination

1 Introduction

The Global Positioning System (GPS) has demonstrated that Global Navigation Satellite Systems (GNSS) are extremely important to provide precise information for geodesy and geodynamics. Therefore, more and more stations will be deployed and an increasing number of low-Earth-orbiting (LEO) satellites will be equipped with GPS receivers. Additionally, the number of satellites in the Russian GLONASS system is growing to provide a better service, and the European Galileo system will come into operation around 2008. Hence, the International GNSS Service (IGS) has called for proper strategies for 'bigger, better and faster' precise orbit determination (POD) (Boomkamp and König 2004).

The well-developed and commonly used software packages for high precision GNSS data analysis, for example those used by the Analysis Centers (AC) of the IGS, have demonstrated their sophistication in modeling and strength in large-volume data handling. However, they are facing new challenges of dealing with the scaled-up networks and increased numbers of satellites in making more comprehensive, extremely precise and high-resolution products in a timely fashion using limited computing resources.

At present, only a subset of the whole IGS network is processed routinely. ACs providing products with more than 100 stations usually process the data in a sub-network mode (Dong et al. 2002; Fang 2003; Zhang et al. 2004; Steigenberger et al. 2004; Herring 2004), where the sub-network solutions are combined at the normal equation (NEQ) level. For the same reason, sub-network data processing is implemented in GPS meteorology to meet the near real-time and high-resolution requirements of weather forecasting (Gutman et al. 2004). Furthermore, fast LEO POD is sometimes divided into two steps: ground network analysis to obtain GPS orbits and clock parameters, and precise point positioning (PPP) for LEOs (e.g., Zhu et al. 2004).

The widely used sub-network strategy needs a certain number of common stations for the final combination. For example, 3–5 redundant stations are chosen for each sub-network for the double-difference (DD) approach (Dong et al. 2002). Obviously, the data from these stations are used more than once, so that the covariance matrix of the combined solution differs from that of the integrated solution. The accuracy in the connection of the sub-networks also depends on the number of common stations and their data quality. Therefore, the one-step integrated solution provides better products (Zhu et al. 2004).

M. Ge (✉) · G. Gendt · G. Dick · F. P. Zhang · M. Rothacher
Department of Geodesy and Remote Sensing,
GeoForschungsZentrum (GFZ) Potsdam,
Telegrafenberg A17,
14473 Potsdam, Germany
E-mail: maor@gfz-potsdam.de
Tel.: +49-331-2881116
Fax: +49-331-2881759

Table 1 Number of estimated parameters for a network of 100 stations and 30 GPS satellites

	All parameters		Active parameters	
Parameters	Formula	Number	Formula	Number
Station coordinates	nsta*3	300	nsta*3	300
Satellite orbits	nsat*15	450	nsat*15	450
Earth rotation		6		6
Receiver and satellite clocks	nsta+nsat	130	nsta+nsat	130
Zenith delays	nsta*12	1,200	nsta*1	100
Ambiguities	nsta*nsat*2	6,000	nsta*nobs	1,000
Total		8,086		1,986

nsta is the number of stations, nsat the number of satellites, nobs=10 the average number of active ambiguities at each station

In this study, it will be shown that the major challenges are the large number of ambiguity and zenith total delay (ZTD) parameters kept in the NEQ system. A new data processing strategy, especially applicable to global networks, is presented, which can significantly reduce the requirement on memory and computation time.

2 Review of current data processing

Taking the IGS data processing for network and orbit determination as an example, the parameters to be estimated include station coordinates, satellite orbits, station and satellite clocks, ZTD at stations and carrier-phase ambiguities. Table 1 shows the number of estimated parameters for a network of 100 stations observing 30 GPS satellites. As the clock parameters are usually eliminated epoch-by-epoch, only those at one single epoch are counted. The number of ambiguity parameters is 3–4 times larger than the sum of all the others. For 60 GNSS satellites, e.g., GPS plus Galileo, the number is even doubled.

The second largest set comprises the ZTD parameters, depending on the step-size for the piece-wise (PW) constant or linear functions. For special applications, e.g., GPS meteorology, more parameters might be required in order to provide high-resolution products. Similarly, Earth rotation parameters and even station positions can also be modeled with PW functions. Thus in practice, there could be many more PW parameters than those listed in Table 1.

Until now, PW and ambiguity parameters are often kept in the NEQ system, and are usually inverted together with all the other parameters. It requires computer hardware with both huge memory and high performance; a condition that can hardly be met for the ever-increasing number of stations. Therefore, ACs are selecting the sub-network data processing mode.

In fact, ambiguity or PW parameters are time-dependent, because they are only valid over a certain time interval. While generating the NEQ system, a time-dependent parameter has two possible states: active or inactive. It is active while processing data within its validity interval and is inactive otherwise. For example, at each station, only ambiguities of satellites being tracked continuously until the processed epoch are active. Thus, its number is approximately that of the observed satellites (nobs) at an epoch, which is about 10 on average for the current GPS constellation. The maximum number of active parameters is listed in Table 1 for the example network. Only about one-quarter of the full parameter set is active at any point in time.

Therefore, parameters should only be set up in the NEQ system when they are required (at the start of their validity interval), and should be immediately eliminated after the end of their validity interval. The eliminated parameters can be recovered by a resubstitution step as described in Sect. 3. In this way, only active parameters are kept in the NEQ system and its size as well as, consequently, the computation time for its manipulation and inversion are significantly reduced.

Usually, parameters are not eliminated but kept in the final NEQ system if their covariances are of interest. However, ambiguity parameters are often kept in the final NEQ only for integer ambiguity fixing, which is important for achieving the best quality products. Therefore, the implementation of the ambiguity fixing is crucial for the development of an efficient – fast and with small memory requirements – parameter estimation strategy.

In the following sections, we first present the algorithm for parameter elimination within a least-squares estimator and then a new ambiguity-fixing approach, which makes ambiguity elimination possible for both real-valued and fixed solutions.

3 Parameter elimination and recovery

Parameter elimination in GPS data processing is discussed in several publications (e.g., Schaffrin and Grafarend 1986; Boomkamp and Dow 2004). Here, we only describe how it is done in the context of our new strategy.

Assuming the observation equations at the first epoch are

$$\begin{aligned} \mathbf{v} &= \mathbf{Ax} + \mathbf{l}\ ,\ \mathbf{P} \\ \mathbf{v}_{\mathbf{x}_0} &= \mathbf{x} - \mathbf{x}_0\ ,\ \mathbf{P}_{\mathbf{x}_0} \end{aligned} \tag{1}$$

where $\mathbf{A}$ is the first design matrix, $\mathbf{x}$ vector of the active parameters, $\mathbf{v}$ the residual vector and $\mathbf{l}$ the observation minus computed (O−C) vector, $\mathbf{P}$ the weight matrix, $\mathbf{x}_0$ and $\mathbf{P}_{\mathbf{x}_0}$ are the a priori values of $\mathbf{x}$ and their weight matrix, respectively.

Then the corresponding normal equations read

$$\mathbf{Nx} = -\mathbf{w} := (\mathbf{A}^{\mathrm{T}}\mathbf{PA} + \mathbf{P}_{\mathbf{x}_0})\mathbf{x} = -\mathbf{A}^{\mathrm{T}}\mathbf{Pl} + \mathbf{P}_{\mathbf{x}_0}\mathbf{x}_0 \tag{2}$$

$$\mathrm{pll} = \mathbf{l}^{\mathrm{T}}\mathbf{Pl} + \mathbf{x}_0^{\mathrm{T}}\mathbf{P}_{\mathbf{x}_0}\mathbf{x}_0 \tag{3}$$

where pll is the sum of the weighted squares of O−C for calculating the standard deviation.

The random-walk constraint on a PW parameter set can be represented by the following pseudo-observations (Dixon and Wolf 1990)

$$v_x = x_{k,m} - x_{k,m-1}, \quad p_x := \frac{\sigma_0^2}{p_d^2 \Delta t} \tag{4}$$

where k is the position of the parameter in the NEQ system; $x_{k,m-1}$, $x_{k,m}$ are two consecutive PW parameters active during the $(m-1)$- and mth time intervals of length Δt, respectively; p_x is the weight of the constraint; p_d the power density of the process noise, for example the typical value for ZTD process is about $15\,\mathrm{mm}/\sqrt{\mathrm{h}}$; and σ_0^2 the a priori unit weight variance. The contribution of Eq. (4) to the NEQ is

$$\begin{bmatrix} p_x & -p_x \\ -p_x & p_x \end{bmatrix} \begin{bmatrix} x_{k,m-1} \\ x_{k,m} \end{bmatrix} = \mathbf{0} \tag{5}$$

As only the active one of each PW parameter set is kept in the NEQ, this contribution cannot be added to the NEQ directly. From the kth equation of Eq. (2),

$$\sum_{l=1;l\neq k}^{n_\mathrm{p}} n_{k,l} x_l + n_{k,k} x_{k,m-1} = -w_k,$$

the deactivated parameter $x_{k,m-1}$ can be expressed by considering Eq. (5)

$$x_{k,m-1} = -\frac{w_k - x_{k,m} p_x + \sum_{l=1;l\neq k}^{n_\mathrm{p}} n_{k,l} x_l}{n_{k,k} + p_x} \tag{6}$$

where n_p is the dimension and $n_{k,l}$ an element of the current NEQ matrix $\mathbf{N}$. Equation (6) must be stored for recovering the eliminated parameters if they are desired.

Inserting Eq. (6) into the NEQ and taking Eq. (5) into account, we have the equivalent NEQ with $x_{k,m}$ in place of $x_{k,m-1}$

$$\begin{aligned} n_{i,j} &:= n_{i,j} - \frac{n_{i,k} n_{j,k}}{(n_{k,k} + p_x)} \\ w_i &:= w_i - \frac{w_k n_{i,k}}{(n_{k,k} + p_x)} \\ \mathrm{pll} &:= \mathrm{pll} - \frac{w_k w_k}{(n_{k,k} + p_x)} \end{aligned} \tag{7}$$

$$\begin{aligned} n_{i,k} &:= \frac{n_{i,k} p_x}{(n_{k,k} + p_x)} \\ n_{k,k} &:= p_x - \frac{p_x p_x}{(n_{k,k} + p_x)} \\ w_k &:= \frac{w_k p_x}{(n_{k,k} + p_x)} \end{aligned} \tag{8}$$

where $i = 1, \ldots, n_\mathrm{p}; \quad j = 1, \ldots, i; i \neq k$ and $j \neq k$.

Using Eq. (8), the positions of the eliminated PW parameters in the NEQ are prepared for their successors. Obviously, Eqs. (7) and (8) hold also for ambiguity or PW parameters without random-walk constraint, i.e., if p_x is set to zero.

The above elimination procedure is performed for all deactivated parameters in the current NEQ system, so that only active parameters remain.

Subsequently, the existing NEQ system has to be extended to include the parameters becoming active at the next epoch. The contribution of all observations at that epoch and the constraints for newly introduced parameters, expressed by Eq. (2), is added to the extended NEQ. Then deactivated parameters are eliminated using Eqs. (5) to (8).

After having repeated the above-described procedure for all epochs, the resulting NEQ system is solved. The eliminated parameters are sequentially recovered by Eq. (6) in reversed order compared to the elimination procedure, i.e., from the last to the first eliminated parameter, because an eliminated parameter is expressed as a function of parameters that are eliminated later or remain in the final NEQ system. Parallel to the parameter recovery, observation residuals are calculated by inserting corresponding parameters into the observation equations.

It should be pointed out that the "one-by-one" elimination of parameters is not only easily implemented, but also faster than a block or batch elimination, because Eq. (6) contains many zero elements and the elimination can work on the non-zero elements only.

4 Ambiguity elimination for fixed solution

In most of the current software packages, the ambiguity-fixing approach works on the NEQ with DD ambiguities. The fixable ambiguities are inserted into the NEQ as known values and the estimates and the covariance matrix of the remaining parameters are updated for sequential fixing (Blewitt 1989; Dong and Bock 1989). Therefore, all ambiguity parameters have to be kept in the NEQ while performing real-valued solutions for ambiguity fixing.

An alternative approach works on the NEQ with zero-difference ambiguities (Ge et al. 2005). After the real-valued solution is obtained, the fixing decisions for all possible DD ambiguities are made according to their wide-lane and narrow-lane estimates and standard deviations. Estimates and standard deviations for the wide-lane linear combinations are derived directly from carrier-phase and code pseudorange observations using the Melbourne–Wübbena method (Melbourne 1985; Wübbena 1985), while those for the narrow-lane linear combinations are derived from the real-valued network solution. From all the fixable ambiguities, a maximum independent set is chosen.

The fixing of each DD ambiguity is realized by a functional constraint on the four related zero-difference ambiguities. It is represented by the following pseudo-observation equation:

$$v_\mathrm{fix} = \mathbf{dx} + l_\mathrm{fix}, \quad p_\mathrm{fix} \tag{9}$$

where $\mathbf{d}$ is the DD operator vector, $\mathbf{x}$ the parameter vector, l_fix the difference between the integer and real-valued ambiguity value, p_fix the weight, which must be large enough to keep the resulting residual v_fix as small as possible.

As these pseudo-observations can be added to the NEQ one by one, an ambiguity parameter can be eliminated when and only when all the pseudo-observations containing this parameter have already been added to the NEQ. Each DD ambiguity to be fixed has its start and end time, i.e., the latest start time and the earliest end time of the four related zero-difference ambiguities. A pseudo-observation can only be added to the NEQ after having processed the observation at the start time, otherwise not all of the four ambiguities are in the NEQ, and it must be added before the end time, where it will be eliminated. Thus, it is proper to add a pseudo-observation while processing the data at its end time. If we choose this end time as the observing epoch time for each fixed DD ambiguity and reorder all of them accordingly, then the corresponding pseudo-observations can be added sequentially as ordinary observations.

However, in the approach by Ge et al. (2005), standard deviations for narrow-lane ambiguities are used while making the fixing decision. This implies that ambiguities must also be kept in the NEQ in order to derive this information. In general, the standard deviations of ambiguities depend on the station and satellite geometry and the number of continuous observations. For large global networks, like those processed in the IGS, the station coordinates and orbits are well known compared to a campaign network. From our statistics, the standard deviations of DD ambiguities are rather small if their data segments are longer than 15 min. Therefore, the fixing decision relies on whether its estimate is close to an integer value.

Based on this fact, in the new strategy, the narrow-lane ambiguities are fixed according to their estimates only, assuming their standard deviation is small enough. From the decision region given by Blewitt (1989) and our experience, a narrow-lane ambiguity can be fixed to the nearest integer if it differs by less than 0.15 cycles from an integer value. The correctness of ambiguity-fixing is confirmed by testing whether the residuals of observations leading by each ambiguity are compatible before and afterwards. Fixing constraints must be removed if any one of the four undifferenced ambiguities cannot pass the test.

With this adaption of our ambiguity-fixing approach, inactive ambiguities can be eliminated immediately in both real-valued and fixed solutions.

Similar to the sequential fixing approaches, the above-mentioned ambiguity fixing and parameter estimation procedure can be repeated to resolve more ambiguities. Our experience has shown that more than 90% of the independent ambiguities can be fixed in the first iteration for IGS networks with about 100 stations or more, and additional 6% can be fixed in the next iterations. Therefore, two iterations are usually sufficient.

5 Experimental validation

The presented strategy has been implemented in the EPOS software (Gendt et al. 1999) developed at GeoForschungs-Zentrum (GFZ) for IGS data processing, which uses ionosphere-free zero-difference carrier-phase and code pseudorange observations.

The estimated parameters are listed in Table 1. The coordinates of IGS core stations are tightly constrained to their ITRF2000 values (Altamimi et al. 2002), using an a priori variance of about 5 mm for each component, in order to obtain stable estimates of ambiguities which can be fixed efficiently. The other parameters are constrained rather loosely according to the quality of their initial values, for example, using a priori variance of 50 m for an ambiguity derived from pseudorange observations.

To validate the results of the new strategy, data from about 100 IGS stations were processed both with our routine IGS procedure and the new strategy. The differences between the products from the two strategies are not significant, e.g., 0.2 mm RMS in the station coordinate differences.

To demonstrate the improved performance of the new strategy, networks with different numbers of stations were analyzed using three strategies on a Linux personal computer with one Pentium-4 3.0 GHz processor and 1 GB memory. In Strategy A (new strategy), only active parameters are kept, i.e., parameters are eliminated as soon as they get inactive. In Strategy B, all ambiguities and ZTD parameters are kept in the NEQ and are eliminated "one-by-one" after all observations are processed. In Strategy C, all parameters are kept and inverted together.

The memory required by Strategies B and C grows rapidly with the number of stations because of the fast increase of the number of ambiguity and ZTD parameters. For a network with 225 stations, it exceeds 1 GB. In Strategy A, the required memory is not larger than 300 MB, even for a network with 300 stations as the number of active parameters is less than 6,000 (see Table 1).

The computation times of the three strategies for one iteration of the least-squares adjustment with a 300-s data sampling rate are shown in Fig. 1. They increase gently and near-linearly with the number of stations for Strategy A, but very rapidly for the other two strategies. On average, Strategy A

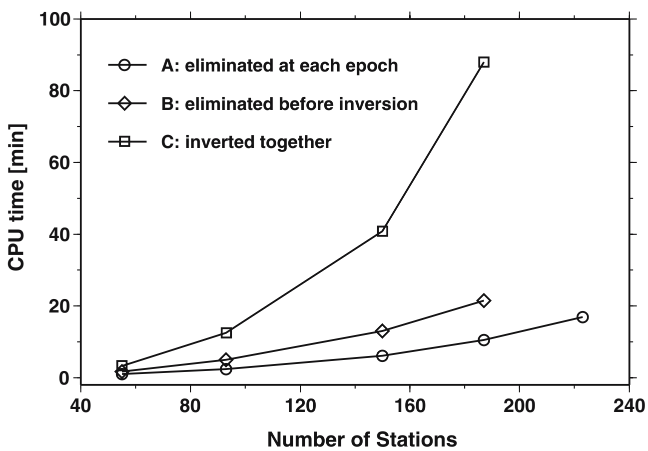

Fig. 1 Comparison of computation times of Strategy A, B and C for networks with different numbers of stations

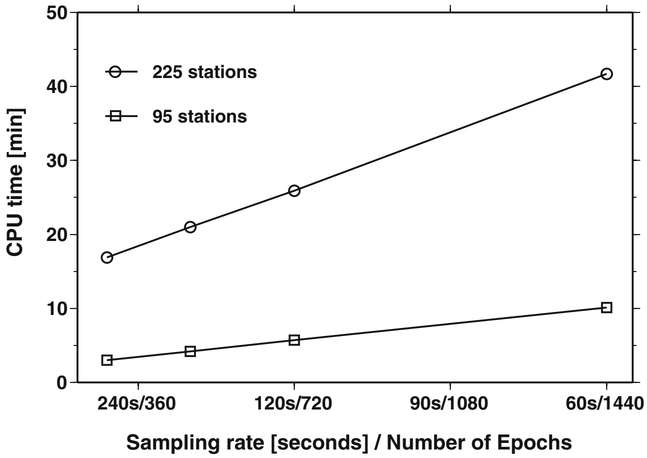

Fig. 2 Computation time of Strategy A (the new strategy) with respect to different data sampling rates for processing networks with 95 and 225 stations

reduces the time to half of that for Strategy B, while Strategy C needs 4–8 times more than Strategy A. A fixed solution takes a slightly longer computation time than a real-valued one because of adding the fixing constraints. There is no difference in the ambiguity-fixing results for all the strategies. On average, more than 95% of the independent ambiguities are fixed correctly after two iterations.

Figure 2 shows the computation time of the new strategy (A) for different sampling rates using 1 day of data for networks with 95 and 225 stations. The increase of time is linear with the sampling rate and is faster for larger networks. Thus the new strategy can also be used for applications where high sampling rates are desired.

It should be pointed out that the overall reduction of computation time for processing a network is several times of that shown in Figs. 1 and 2, as the estimation is carried out iteratively for both data cleaning and ambiguity fixing.

6 Conclusion

A new strategy for processing huge GNSS networks is presented, which solves the conflict between providing better and faster products and limited computer resources.

By adapting current ambiguity-fixing strategies, fixing decisions are made according to the ambiguity estimates without knowing their standard deviations. Based on the new fixing approach, the inversion of a huge NEQ system with all ambiguities is avoided, thus saving computation time. Furthermore, fixed ambiguities are represented by pseudo-observations, which are imposed on the NEQ system. This allows ambiguity parameters to be eliminated for both real-valued and fixed solutions as soon as there are no related observations to be processed anymore.

We have demonstrated with IGS data that computer memory is not a problem anymore for the new strategy and that the computation time is reduced to one-third or even to one-eighth of the current methods, depending on the number of stations. Huge GPS networks with several hundred stations can now be processed efficiently on a personal computer.

The strategy can easily be implemented into existing software packages that process GNSS data in zero as well as DD mode, and could replace the current methods for providing 'bigger, better and faster' products with an ever-increasing amount of data.

References

Altamimi Z, Sillard P, Boucher C (2002) ITRF2000: A new release of the international terrestrial reference frame for Earth science applications. J Geophys Res 107(B10):2214. DOI 10.1029/2001JB000561

Blewitt G (1989) Carrier phase ambiguity resolution for the global positioning system applied to geodetic baselines up to 2000 km. J Geophys Res 94(B8):10187–10203

Boomkamp H, König R (2004) Bigger, better and faster POD. In: Proceedings of IGS workshop and symposium, 1–6 March 2004, Berne, Switzerland

Boomkamp H, Dow J (2004) Large scale GPS processing at ESOC for LEO, GNSS and real-time applications. In: Proceedings of IGS workshop and symposium. 1–6 March 2004, Berne, Switzerland

Dixon T, Wolf K (1990) Some test of wet tropospheric calibration for the CASA UNO global positioning system experience. Geophys Res Lett 17(3):203–206

Dong D, Bock Y (1989) Global positioning system network analysis with phase ambiguity resolution applied to crustal deformation studies in California. J Geophys Res 94(B4):3949–3966

Dong D, Fang P, Bock Y, Cheng MK, Miyazaki S (2002) Anatomy of apparent seasonal variations from GPS-derived site position time series. J Geophys Res 107(B4):2075. DOI 10.1029/2001JB000573

Fang P (2003) Analysis of seasonal signals in GPS position time series, Paper presented at international workshop on GPS meteorology, 14–17 January 2003, Tsukuba, Japan

Ge M, Gendt G, Dick G, Zhang FP (2005) Improving carrier-phase ambiguity resolution in global GPS network solutions. J Geod 79 (1–3):103–110. DOI 20.2007/s00290-005-0447-0

Gendt G, Dick G, Soehne W (1999) GFZ analysis center of IGS – Annual Report 1998. IGS 1998 Technical Reports, pp 79–97

Gutman S, Sahm R, Benjamin G, Schwartz E, Hlub L, Stewart Q, Smith L (2004) Rapid retrieval and assimilation of ground based GPS-Met observations at the NOAA forecast systems laboratory: impact on weather forecasts. J Meteorol Soc Jpn 82:351–360

Herring T (2004) Aspects of large station networks for GPS orbits and clocks. In: Proceedings of IGS workshop and symposium, 1–6 March 2004, Berne, Switzerland

Melbourne WG (1985) The case for ranging in GPS-based geodetic systems. In: Proceedings 1st international symposium on precise positioning with the global positioning system. 15–19 April 1985, Rockville, pp 373–386

Schaffrin B, Grafarend E (1986) Generating classes of equivalent linear models by nuisance parameter elimination. manuscr geod 11(3):262–271

Steigenberger P, Rothacher M, Dietrich R, Rülke A, Fritsche M (2004) Reprocessing of a global GPS network. Geophys Res Abstr 6:07649

Wübbena G (1985) Software developments for geodetic positioning with GPS using TI-4100 code and carrier measurements. In: Proceedings of 1st international symposium on precise positioning with the global positioning system, 15–19 April 1985, Rockville, pp 403–412

Zhang FP, Gendt G, Ge M, Wuensch J (2004) Vertical rate of crustal motion at tide gauges derived by GPS. Geophys Res Abstr 6:03344

Zhu S, Reigber CH, König R (2004) Integrated adjustment of CHAMP GRACE and GPS data. J Geod 78(1–3):103–108. DOI 10.1007/s00190-004-0379-0