Quality check procedure of products derived from DD-GNSS data analysis in CNRS-UGA

Document information Summary

Date	5 February 2018
Document title:	**Quality check procedure of products derived from DD-GNSS data analysis in CNRS-UGA**
WP name and no	WP10 TCS GNSS Data and Products
DDSS	WP10-DDSS-008 : Products.GNSS.DD.Position WP10-DDSS-010 : Products.GNSS.DD.TimeSeries WP10-DDSS-012 : Products.GNSS.DD.Velocity
Lead Partner	CNRS-UGA
Main Author(s)	Aline Deprez
Contributing author(s)	Anne Socquet

Table of Contents

1. Introduction

This document describes the different quality checks, which punctuated the genesis of the prototype solution in CNRS-UGA. This solution was implemented by using the GAMIT software (developed at MIT), which is based on a double difference method (DD).

The tested prototype solution consists in a set of daily SINEX files and its main contours are:

- A large number of stations (about 600). The RINEX observation files and the associated meta-data were downloaded through the existing GSAC nodes (RENAG, INGV, NOA) and through the EPN ftp.
- A 11 year period covered (2005-2015)
- An alignment on the ITRF08 reference frame.

Further information on the processing parameters are available in a dedicated document.

As for a traditional GNSS processing in double differences, each day was independently treated. The huge amount of data prevented from computing the double differences between all pairs of stations and satellites. To take up this up scaling challenge, each daily processing was spatially divided in station sub-networks (following the same strategy than the one developed in the framework of the PBO project in the USA). Given the fact that the processing was divided into a large number of small processes, which should run independently, the use of a highperformance computing platform became necessary. Each process with its own parameters was associated with a job launched on a cluster node. Then a daily combination allowed to obtain a daily Sinex file which was followed by a global combination over the whole period, giving access to the position time series.
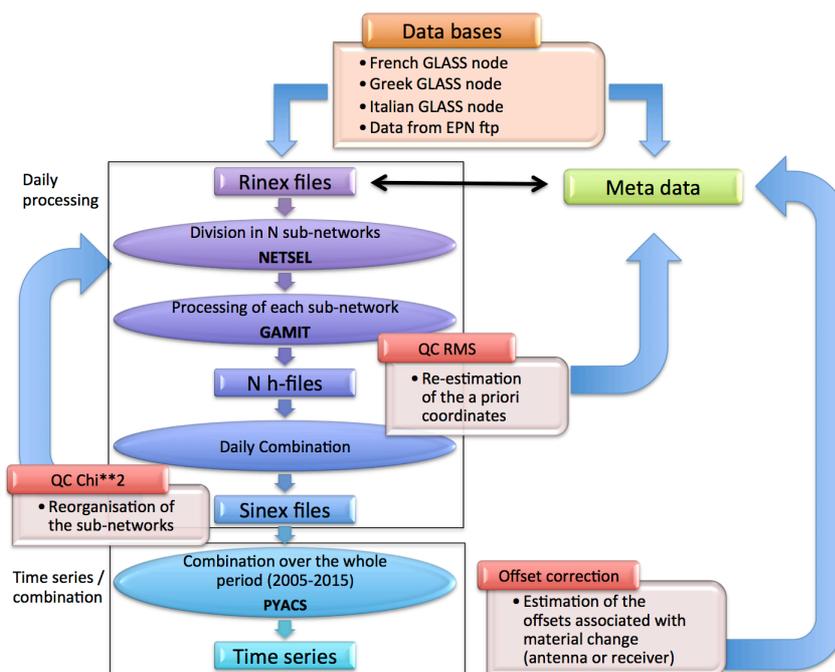

Figure 1: Processing chart at CNRS-UGA to produce time series

Since all the institutes involved in EPOS-WP10 were conducting their mission in parallel and at the same time, the data provided to the processing centres did not benefit from a careful and homogenous quality check. It was essential to punctuate these prototype solutions by different quality checks. Some of them should not be required anymore once the validated data and metadata from the EPOS data portal will be available. Moreover this prototype solution was also used to validate the method of velocity estimate. This validation will lead to a choice of the best-adapted method and to potential future improvements.

2. Check of individual processing jobs

The GAMIT processing of each pair day-subnetwork results in GAMIT h-files. Two checks are then performed:

- **All the expected data** must be **included in the resulting solution**. If one (or more) station doesn't appear in the resulting h-file, it usually means that the given **a priori coordinates** were too far from the real position of the station. In such a case the a priori position must be re-computed and updated in the meta-data base. The quality of the data file itself could be another cause of data rejection. These cases should not occur anymore once both the validated data files and associated metadata from the EPOS data portal will be available.
- **The normalized RMS** (square root of chi square per degree of freedom) for the loose solution with the ambiguities resolved is systematically checked. The expected value of this RMS is ~0.2. All the h-files fit such criterion but to refine the quality check and to take into account the temporal variation of the RMS value (due to an improvement of the data with time) a "warning" was associated with a threshold depending on the average annual value. This threshold was computed to point out the values out of a confidence interval of 2.576*standard deviation (i. e. confidence level of 99%).

3. Check of daily solutions

After this first step, every h-files for a given day must be combined in order to obtain a daily solution with the standard SINEX file format. For this purpose, GLOBK software is run and results into a single binary GLX solution file. The chosen parameters are associated with an estimation of an overall translation, rotation and scale (Helmert parameters) of the network and a set of 60 reliable sites are used for the stabilization. Since the previous GAMIT processing strategy used fixed the orbits, the obtained GLX file corresponds to a fixed-IGS solution. The conversion of the combined GLX file to SINEX of daily site position and covariance is straightforward using a script provided with the GAMIT/GLOBK distribution. The quality of these daily combinations is checked using the following criterion, based on the compatibility between the set of sub-networks:

- **The Chi**2 value** reflects the internal consistency of the daily solution. This value is incremented with the inclusion of each h-file and this increase is associated with the level of incompatibility between the newly added data and the solution from the data previously included. An anomalously large Chi**2 increment (typically ~ 1) points out an incompatibility problem for the newly added h-file, which probably encompasses a large number of low quality data and leads to a sub-networks reorganization (and therefore reprocessing).

4. Time-series quality check: statistics and visualization using Pyacs interactive chain

The computing and post-processing of time series were conducted using PYACS software (set of tools written in python language developed at Observatoire de Côte d'Azur, Nice, France by J.M. Nocquet). The quality of the combination was evaluated both along time with a check after each additional epoch and also station by station with a wrms check of each station position time series. In addition, the number of used epochs (i.e. number of days for which a position has been computed) and the time span of each station allows us to determine the percentage of working day for each station. This package also allows for the visualization of the position time series and for their post-processing:

- Removing of a linear trend,
- Removing of the outliers (0.005% of largest outliers),
- Estimating and correcting the jumps associated with documented changes of equipment (antenna or/and receiver).

At this step, all the time series were carefully visually examined in order to detect any anomalous changes in the position and/or in the position trend (i.e. changes not associated with geophysical phenomena and potentially associated with equipment failure).

For the prototype solution, the computed products associated with all the stations have been uploaded on the product portal without "station quality filter".

5. Comparison of prototype velocity fields estimated with GLOBK vs MIDAS softwares to validate the method of velocity estimate

The massive processing implementation requires innovative strategies in order satisfy the efficiency, robustness and automation criteria for the computation of another usual GNSS product: the velocity field. Given the number of data, a key issue is the time computation but taking into account the outlier and offset detection is also of crucial importance to obtain a velocity field of good quality. In order to test and validate the generation of a velocity field and to determine the more appropriate method, two independent velocity fields have been generated and compared: one using GLOBK (a well established method based on a Kalman Filter) and one using MIDAS software. This latter software was developed at Nevada Geodetic Laboratory, USA by G. Blewitt and allows to obtain GNSS velocities robust to outliers, steps and seasonality by estimating automatically the time series trend. On the contrary, Globk software (developed at MIT) based on Kalman filter needs a careful identification of the jumps (associated with equipment changes) and the velocity value is forced to be equal before and after each jump. However, with the current automated analysis, the velocity computation doesn't take into account the seismic jumps (this would require a synchronization with seismic catalog), the post-seismic acceleration and the seasonality. Then, if it can also be tested for this prototype solution, the limitations of such software with regard to the amount of data (stations number and time span) must be evaluated. Indeed the computation using GLOBK is very time consuming, this no splittable step was exposed to computer instability and MIDAS software may be a good and faster alternative.

In order to quantify the comparison, the estimated parameter were:
- Histogram of velocity difference
- Identification of the criteria in the origin of the largest differences.

6. Comparison with other EPOS solutions

We have applied the same method described previously to generate time series (with PYACS) and velocity field (with MIDAS) using the set of SINEX distributed by INGV and processed with Gipsy software based on a Precise Point Positioning (PPP) method.

EPOS PPP and DD derived products (time series & velocity fields) have been compared in order to point out any detectable systematic difference between both. The comparison between the two sets of products was based on a statistical approach, differentiating the zones that underwent high geophysical activity from the zones supposed to be stable. In order to quantify the comparison, the estimated parameters were:
- An histogram of repartition of differences in daily positions between PPP and DD solutions,
- An histogram of repartition of the scatter for the time series of the DD or PPP solutions
- The 2 velocity fields were compared (histogram of velocity difference, spatial variability of this difference)

References

- Blewitt, G., C. Kreemer, W.C. Hammond, J. Gazeaux, 2016, MIDAS robust trend estimator for accurate GPS station velocities without step detection, accepted for publication in the Journal of Geophysical Research, doi: 10.1002/2015JB012552
- Blewitt, G., C. Kreemer, W.C. Hammond, and J.M. Goldfarb, (2013). Terrestrial reference frame NA12 for crustal deformation studies in North America. J. of Geodynamics, 72, pp. 11-24, doi:10.1016/j. jog.2013.08.004.
- Gipsy-Oasis, JPL : https://gipsy-oasis.jpl.nasa.gov
- Herring, T. A., T. I. Melbourne, M. H. Murray, M. A. Floyd, W. M. Szeliga, R. W. King, D. A. Phillips, C. M. Puskas, M. Santillan, and L. Wang (2016), Plate Boundary Observatory and related networks: GPS data analysis methods and geodetic products, Rev. Geophys., 54, 759–808, doi:10.1002/2016RG000529
- King, R. W. et Bock, Y. (2015). Documentation for the GAMIT/GLOBK GPS software analysis, release 10.06. Massachusetts Institute of Technology, Cambridge.
- PYACS : PhD thesis Dinh Trong Tran supervised by J. M. Nocquet. Analyse rapide et robuste des solutions GPS pour la tectonique. Autre. Université Nice Sophia Antipolis, 2013. Français.