A Guide to QA Files for
Practical Quality Assessment of GPS Data

Geoffrey Blewitt
University of Nevada, Reno
gblewitt@unr.edu

Introduction

QA files provide statistics on individual GPS RINEX files. These statistics are computed from actual post-processing of the data using the high quality GPS data processing software with IGS standards. QA files thus provide hard measures for quality assessment that closely relate to the intended use of RINEX files and to the needs of the user, in a way that is not possible by the type of quality control provided by popular and very useful pre-processors, such as UNAVCO's "teqc." In fact, the process of generating QA files includes the use "teqc" as well as the pre-processing software that is an integral part of the data processing (including "turboedit").

The data processing engine used to generate QA files is GIPSY OASIS II from the Jet Propulsion Laboratory (JPL), under licence by the California Institute of Technology. The data processing strategy is "precise point positioning" (PPP), which uses precise satellite and orbit products from JPL to fit the parameters associated with the individual station associated with the RINEX file. Thus the QA statistics are not sensitive to problems that might occur with other stations in a regional GPS network. The QA records isolate the quality and possible problems of the data specific to each individual station on a specific day. Since the QA statistics are not derived from station coordinate time series analysis, it is the intention that relationships between the statistics and actual coordinate time series might be used to derive indices that can help identify problems such as failing antennas, failing receiver channels, or changes in multipath environment and sky obstructions. Such analysis might also help to assess and validate the swapping of station equipment, and rank the performance of station configurations.

In general, each QA file is a CSV file (ASCII format with comma separated values), with each line being a record, with each record referring to a specified daily RINEX file spanning up to 24 hours of a GPS day. Key data fields that identify the GPS data are the name of the RINEX file, the 4-character ID of the station, and the date. Statistical data fields are derived from modeling the GPS data. An optional single-line header starting with the character # provides a CSV list of data field names, which can be used either for automatic machine reading of the file, or as a reminder for human reading. This feature allows for possible future expansion of field types.

In principle, a QA file can have any number of records referring to a set of RINEX files. For internal database purposes at UNR, every RINEX file has a unique QA file containing a single record. For purposes of convenience, QA files are distributed in two flavors: (1) QA files "by date", which contain a single record from every RINEX file on a specific date, and (2) QA files "by site", which contain a single record for every daily RINEX file for a specific site.

Hence the "by site" files are a time series of statistics, thus "bad days" might be identified, or systematic time variation may show evidence of equipment changes or failures. In contrast, the "by date" files might be used to assure the user that the processing used to generate the QA files (including the JPL orbit products) falls within a certain tolerance on any specific day, and can be used to assess what constitutes a "normal" station, and stations that are performing poorly at any given time.

Obtaining QA Files

The University of Nevada, Reno (UNR) produces QA files for ~10,000 GPS stations around the world, of which over 6,000 stations produce RINEX files every day. QA files are available for RINEX files starting 1 January 1996. These files are planned to be made publicly available by UNAVCO. For Linux users, the following "tarball" file is typically updated every week on UNR's ftp server with 2 week latency:

```
ftp://gneiss.nbmg.unr.edu/qa.tar
```

After downloading is complete, the QA files can be extracted by the command:

```
tar -xf qa.tar
```

This produces the directory `qa/`, with two subdirectories `qa/by_date/`, and `qa/by_site/`, containing the two different flavors of QA files. The by_date directory has yearly subdirectories. Here are some examples of file names, with obvious naming conventions:

```
qa/by_date/1996/1996-01-01.001.qa
qa/by_date/1996/1996-01-02.002.qa
...
qa/by_date/1996/1996-12-31.365.qa
qa/by_date/1997/1997-01-01.001.qa
...
qa/by_date/2012/2012-06-16.168.qa

qa/by_site/ASHM.qa
qa/by_site/JPLM.qa
qa/by_site/SLID.qa
qa/by_site/TIVA.qa
```

QA File Records

The following example shows the first few records of `qa/by_date/1996/1996-01-01.001.qa`

```
#rinex,station,date,acceptedepochs[unit="hr"],dataintervalfromteqc[unit"s"],datainterval
assumed[unit="s"],nsatsrinex,nsatschecked,percentphasesaccepted,nphasebiases,phaseRMSall
[unit="mm"],phaseRMSmax[unit="mm"],codeRMSall[unit="m"],codeRMSmax[unit="m"],3dPPPsigmar
aw[unit="mm"],chi2perDOF,3dPPPsigmascaled[unit="mm"]
albh0010.96o,ALBH,1996-01-01,24.00,30,30,24,24,91.8,59,5.4,9.9,0.47,0.59,2.5,0.4,1.6
algo0010.96o,ALGO,1996-01-01,24.00,30,30,24,24,96.2,50,4.7,7.9,0.45,0.57,2.6,0.3,1.4
ankr0010.96o,ANKR,1996-01-01,24.00,30,30,25,24,89.3,57,6.0,7.8,0.77,0.95,3.3,0.6,2.6
areq0010.96o,AREQ,1996-01-01,24.00,30,30,24,24,93.4,67,8.1,11.1,0.50,0.63,3.3,1.0,3.3
arol0010.96o,AROL,1996-01-01,23.92,30,30,25,24,83.0,51,12.4,20.8,0.63,0.97,3.8,2.4,5.9
auck0010.96o,AUCK,1996-01-01,24.00,30,30,25,24,90.8,48,5.3,9.1,0.51,0.64,2.6,0.4,1.6
```

Each line is a record without any spaces, with data fields enclosed by commas. Any consecutive commas should be interpreted as empty fields (meaning that there is no available information on that field). The very first line starting with # is a comment line indicating the field types, and does not actually contain any line breaks until the end of the string `3dPPPsigmascaled[unit="mm"]`.

The first data record can be interpreted as follows:

#	Field Name	Unit	Data Example	Description
1	rinex		agmt0050.12o	RINEX file name
2	station		AGMT	4-character station ID
3	date		2012-01-05	date in format yyyy-mm-dd
4	acceptedepochs	hour	24.00	hours of data
5	dataintervalfromteqc	sec	15	data interval reported by teqc software
6	dataintervalassumed	sec	15	data interval assumed here
7	nsatsrinex		30	number of satellites in RINEX file
8	nsatschecked		29	number of satellites checked for QA
9	percentphasesaccepted		87.4	percent of epochs accepted

10	nphasebiases		57	number of phase bias parameters
11	phaseRMSall	mm	5.3	RMS post-fit residual of ion-free phase
12	phaseRMSmax	mm	6.9	as above but for satellite with max RMS
13	codeRMSall	m	0.59	RMS post-fit residual of ion-free pseudorange
14	codeRMSmax	m	0.81	as above but for satellite with max RMS
15	3dPPPsigmaraw	mm	2.0	formal error in 3-D position
16	chi2perDOF		0.4	chi-square/DOF (per degree of freedom)
17	3dPPPsigmascaled	mm	1.3	3-D formal error scaled by sqrt(chi-square/DOF)

Field-Specific Notes

4. Hours of data reflects the number of 5-minute epochs that contribute to the final position solution (which is based on 5-min decimated phase data). Ideally this number should be 24.00 for a daily RINEX file. This statistics could be smaller if there are data outages, or if the receiver was switched on or off part way into the GPS day. It could also be smaller if there are periods with too few satellites contributing to the solution (as detected by large formal errors in the white-noise station clock solution).
5. The teqc software tends to report the minimum data interval found in the file
6. Occasionally the RINEX file can have one or a few minimum data intervals, but a longer interval predominates. The algorithm used here infers the predominant data interval in seconds from the following set of assumed conventional values: {1,5,10,15,30,60}. Unconventional rates (if they ever exist), mixed data rates, or intervals shorter than 1 second will not be inferred correctly. If the data interval appears to be larger than 60 sec, then the teqc value is assumed. The reported data interval is then assumed in the computation of data field 9 (percentage of epochs accepted).
8. This is the number of satellites that have contributed to the final position solution. The number of satellites checked for QA is limited by the number of satellites with orbit and clock parameters provided by JPL. Satellite parameters might not be available due to poor orbit modeling. Moreover, some satellites may have parameters but have not contributed to the final solution due to poor common visibility (e.g., less than 4 satellites), or because of short phase-connected data arcs (<20 min) that were removed prior to computing the solution.
9. This is the percentage of 5-minute epochs that exist in the RINEX file used in the final position solution. This statistic mainly reflects the elimination of short data arcs (<20 min) that were removed prior to computing the solution. It might reflect deleted outliers, however the tolerance for outliers is set very high at 5 meters, so outlier deletion is extremely rare. Note that this statistic is not sensitive to data outages or the time span of the file ("short" files), which makes it independent of the statistic in field 4 (hours of data).
10. The number of phase biases is the final number that have been modeled in the positioning solution. This includes a minimum of one bias per satellite arc of data, plus cycle slips that could not be resolved in data pre-processing. Thus, unlike teqc, cycle slips that can easily be fixed are not included in this count. Also not included are slips associated with short (<20 minute) arcs that are deleted and do not contribute to the solution. Ideally this statistic should be close to twice the number of checked satellites. The global average is number is 63, though much larger numbers are observed at poorly performing sites that may have radio interference or significant sky obstructions.
11. This RMS residual phase statistic is computed over all phases that contributed to the position solution.
12. This statistic reflects the satellite with the largest RMS residual phase. This might indicate a problem with the satellite orbit parameters rather than a problem with the RINEX file.
13. Pseudoranges only weakly affect the final position solution as they have an *a posteriori* white noise sigma of 1 m, as compared 0.01 m for carrier phases. Nevertheless, this RMS statistic does reflect the multipath environment at a station.
15. The formal error in 3-D position is the RSS of formal error in X, Y, and Z. The formal error assumes an *a posteriori* white noise sigma of 1 cm for carrier phases, and 100 cm for pseudoranges.
16. The chi-square per degree of freedom is computed only from the phase residuals, and assumes the above *a posteriori*

phase noise of 1 cm. The degree of freedom assumes that the number of data are the number of contributing epochs N (with sufficient satellites to constrain the white-noise station clock), and that the number of parameters equals 3 coordinates, plus the number of phase biases, plus N station clocks, plus N zenith tropospheric delays. Because the real phase noise is often less than 1 cm, it is not unexpected that this statistic can be < 1. Much larger values, however, should be considered suspect.

17. This is the formal error assuming an *a posteriori* phase noise that normalizes the chi-square per degree of freedom to unity. This is computed by multiplying field 15 by the square-root of field 16.

General Notes

The statistics given in the example record above are fairly typical of a good RINEX file.

A missing QA record (say in a "by site" time series) might indicate a problem with a given station on that date. Therefore, the lack of a QA record should, by itself, make a RINEX file highly suspect.

A QA record will not be produced if the RINEX file is considered small (typically reflecting < 18 hours of data, depending on the network). Also a QA record will not be produced if a solution failed, or if an invalid position solution was detected. Solutions are defined as invalid if they have an extremely unlikely height (h < -1km or h >10 km), or if the formal error in height > 20 mm (an order of magnitude larger than typically expected). A solution might fail, for example, due to a badly formatted RINEX header, missing metadata in the header, bad approximate coordinates (e.g., from 2 stations having the same 4-character ID on different continents), or a corrupt file. Also, QA records will will not be compute unless the RINEX header has a valid IGS antenna type for which calibrations can be applied. (Much effort has been made to translate common mispellings and unconventional names into valid IGS antenna names, so this is not a common problem).

PPP Strategy Affecting QA statistics

The QA statistics are a function of the modeling and estimation strategy used for PPP. Here the PPP strategy is briefly documented:

Model

- IERS2010/IGS08 standard models as implemented at JPL's IGS Analysis Center
- satellite orbit and clocks fixed to final fiducial-free products from JPL's IGS Analysis Center
- GMF tropospheric mapping function
- ocean loading from Chalmers University, using FES2004 tidal model, including degree 1
- companion ocean loading constituents
- pole tide, and frequency dependent Love numbers for tidal model
- IGS antenna phase calibrations for stations and satellites
- no 2nd order ionosphere
- no non-tidal loading (e.g., atmospheric and hydrologic)

Estimation

- observation white noise sigma = 1 cm for ion-free phase, 1 m for ion-free pseudorange
- station XYZ as either
  - constant over 24 hours, or
  - random walk every 5 min as random walk at 1 m/sqrt(sec) = 17 m/sqrt(epoch)
    nominal = 0 +- 2000 km
- troposphere dry zenith delay fixed as function of station height:
  = 1.013 * 2.27 * exp(-.000116 * h) meters, where h is the

station height above the ellipsoid in meters
- troposphere wet zenith delay as random walk at 5 * 10**-8 km/sqrt(sec) ~ 1 mm/sqrt(epoch)
  nominal = 0 +- 0.1 m (which also absorbs residual dry bias)
- 2 troposphere gradients as random walk 5 x 10**-9 km/sqrt(sec) ~ 0.1 mm/sqrt(epoch)
- post-break cycle slips detected at 7 cm threshold
- ambiguities resolved using WLPB method from user station to IGS stations

Cuts on Data
- daily batches (GPS day)
- only use epochs with dual frequency phase available
- form ionosphere-free phase and ionosphere-free pseudorange combinations
- set elevation mask at 7 degrees
- pre-solution data editing using TurboEdit algorithm
- phase decimation to 300 sec, pseudorange carrier-smoothed to 300 sec
- minimum phase connected arc length = 20 min
- postfit residual outliers > 5 m