A data archive of GPS navigation messages

G. Beyerle, M. Ramatschi, R. Galas, T. Schmidt, J. Wickert,

M. Rothacher

GeoForschungsZentrum (GFZ) Potsdam, Germany

Abstract   Since 18 June 2007 navigation data messages transmitted by the GPS constellation are recorded by five receivers within GeoForschungsZentrum's global groundstation network. We describe the recording, processing, validation, analysis and archiving of the navigation data. During the 197 days between 18 June 2007 and 31 December 2007 a total of 125,723,666 subframes were collected. By taking into consideration that the same data set frequently is observed to two or more receivers concurrently, 65,153,955 unique subframes could be extracted from the observations. With an estimated 88,099,200 subframes transmitted by the constellation during this time period a data yield of about 74% was achieved. Simulation studies suggest that with two additional GPS receivers, which are scheduled for addition to the network in 2008, about 95% of the transmitted subframes will be retrieved. The message data archive is open to the scientific community for

non-commercial purposes and may be accessed through GFZ's Information System and Data Center (`http://isdc.gfz-potsdam.de`).

1 Introduction

Innovative utilizations of Global Navigation Satellite System (GNSS) signals in precise navigation, geodesy, atmospheric physics, meteorology, seismology and other scientific fields frequently rely on highly accurate data which cannot be extracted from the GNSS observation itself. E.g., precise Global Position System (GPS) reference data recorded by ground-based receivers as well as orbit and clock products are collected, archived and distributed by the International GNSS Service (`http://igscb.jpl.nasa.gov`).

Profiling of the neutral atmosphere and ionosphere by GPS radio occultation (RO) is ranked among these innovative techniques and is considered a valuable new data source for operational numerical weather prediction and climate change studies (*Kursinski et al.*, 1997; *Yunck et al.*, 2000; *Wickert et al.*, 2008). Furthermore, GPS signal reflections at the Earth's surface provide information on ocean sea states, altimetric heights or soil humidities (*Lowe et al.*, 2002; *Treuhaft et al.*, 2001; *Masters et al.*, 2004).

Data analysis and processing procedures of open-loop signal tracking techniques in reflectometry and occultation experiments require knowledge of the navigation message bit structure in order to fully reconstruct the carrier phase from the observed in-phase and quadrature-phase correlation sums (*Sokolovskiy et al.*, 2006; *Beyerle et al.*, 2006). When probing the

tropical troposphere the recorded RO signals are frequently impaired by multipath propagation leading to strong Doppler frequency variations and low signal-to-noise ratios (SNRs). Under these conditions the navigation bits cannot be reliably inferred from the measurement and have to be observed independently.

However, to the best of our knowledge no freely accessible data archive providing current navigation bit data exists. Within the GNSS remote sensing community two projects address the task of monitoring and archiving navigation data. Institut d'Estudis Espacials de Catalunya (IEEC), Barcelona, Spain conducted the Assisted Raw Sampling Project within the framework of the EUMETSAT-funded GRAS Meteorology Satellite Application Facility (`http://grassaf.org`). IEEC established a web site (`https://bond.ieec.uab.es/ars/index.php`) offering sample data recorded by a receiver located at Bellaterra, Spain for the time period between 29 September 2006 and 9 January 2007. Second, the COSMIC multi-satellite occultation mission is supported by a network of six 'bit grabber' GPS receivers (`http://www.cosmic.ucar.edu/~jjohnson/cosmic/BitGrabber/index.html`). The complete navigation data, though, appear not be distributed by the COSMIC data center (`http://cosmic-io.cosmic.ucar.edu/cdaac/index.html`).

To fill this gap, GFZ initiated an activity to record, validate and archive the complete navigation data set. When full coverage of the signal transmitters is achieved the data set will provide $n_{\mathrm{GPS}} \times 50 \times 86{,}400$ navigation bits

per day ($n_{\mathrm{GPS}} \times 540$ kilobytes per day) where n_{GPS} denotes the number of transmitting GPS satellites.

The paper is organized as follows: The first two sections recall the bit structure of the GPS navigation data and describe GFZ's ground receiver network as well as the navigation data processing on subframe level. In the third section we discuss results from verification tests, the validation criteria we implemented on the basis of the test results and recommend future extensions of the receiver network based on the current data set.

2 Navigation Data Structure

The GPS navigation message data is modulated onto the L1 (1.58 GHz) and L2 (1.23 GHz) carrier signal using binary phase shift keying. I.e., a 0 bit leaves the carrier unchanged, a 1 bit shifts the carrier phase by $180°$. The message is organized in frames, each frame consists of 1500 bits. With a transmission rate of about 50 bits per second (1 navigation bit per 20 C/A code sequences) the transmission of one full frame lasts about 30 s.

Each frame is subdivided in 5 equal-length subframes numbered from 1 to 5. In subframes 1 to 3 the satellite transmits information on its position and velocity, transmitter clock biases and drifts as well as ionospheric correction parameters. Subframes 4 and 5 contain almanac information for the full GPS constellation (e.g. *Misra and Enge*, 2002). To accommodate the almanac data for the complete satellite constellation, subframes 4 and 5 subcommutate 25 times. I.e., there are 25 different versions of these two

subframes and it takes about $25 \times 5 \times 6$ s $= 12.5$ min to complete a full navigation data cycle.

The 300 bits in a subframe are grouped into 10 words of 30 bits each. Within each word the first 24 bits are data bits followed by 6 parity bits. These parity bits are added in order to protect the data message against bit errors. The first two words in each subframe contain a characteristic marker bit pattern (Telemetry word, TLM) and a time tag (Handover word, HOW). Subframe words 3 to 10 encode the subframe-specific information (*Misra and Enge*, 2002).

We note that the navigation data exhibits a large degree of redundancy; there exists a high probability that the full navigation data cycle of 12.5 min repeat unchanged except for the updated time tag information encoded into the HOW word (see section 4).

3 Data Collection and Validation

As of this writing, GFZ has deployed 23 geodetic GPS receivers within its global 1 Hz ground-station network (*Galas et al.*, 2001; *Wickert et al.*, 2001). Table 1 lists the number of instruments as well as their type and manufacturer.

Four receivers within the network are 'PolaRx2' instruments manufactured be Septentrio, Leuven, Belgium; one additional 'PolaRx2' instrument (station identifier 'POTS') is placed at GFZ Potsdam for instrument trials and tests. The locations of the five Septentrio instruments are listed in

Table 1 The manufacturers, types and respective numbers of GPS receiver instruments deployed in GFZ's ground-station network. Currently, navigation data are recorded and transmitted by five 'PolaRx2' receivers. The other 19 instruments lack the option of complete navigation subframe recording.

type	manufacturer	number of receivers
PolaRx2	Septentrio	5
Benchmark ACT	Allan Osborne	14
SNR-8000 ACT	Allan Osborne	2
Z-XII3	Ashtech	3

Table 2 Current and intended locations of 'PolaRx2' receiver instruments within GFZ's ground-station network. The instruments at Salta, Argentina (station identifier 'UNSA') and Windhoek, South Africa ('WIND') are currently not capable of navigation data recording and will be replaced by 'PolaRx2' receivers in 2008.

station name	location	latitude / longitude
JOGJ	Jogjakarta, Indonesia	7.7737°S, 110.3765°E
MIZU	Mizusawa, Japan	39.1352°N, 141.1328°E
OUS2	Dunedin, New Zealand	45.8695°S, 170.5109°E
POTS	Potsdam, Germany	52.3793°N, 13.0661°E
ZWE2	Zwenigorod, Russia	55.6996°N, 36.7580°E
UNSA	Salta, Argentina	24.7275°S, 65.4076°W
WIND	Windhoek, South Africa	22.5749°S, 17.0894°E

Table 2. In the coming months two network receivers at Windhoek, South Africa (station identifier 'WIND') and Salta, Argentina (station identifier 'UNSA') will be replaced by 'PolaRx2' instruments.

3.1 Raw Data Blocks

On 18 June 2007 (day of year 169) the global network receivers at Jogjakarta, Indonesia (station identifier 'JOGJ'), Mizusawa, Japan ('MIZU'), Dunedin, New Zealand ('OUS2'), Zwenigorod, Russia ('ZWE2') and Potsdam, Germany ('POTS') were configured to transmit navigation data in addition to code and carrier phase data packets (*Septentrio*, 2004). The transmission of navigation bit data blocks, denoted by the manufacturer as data blocks 'GPSRaw' (Septentrio block identification number 5895), increases the total data bandwidth by about 130 bytes/s; as a consequence, the overall raw data volume transmitted be the five 'PolaRx2' receivers rose by about 30%.

Each 'GPSRaw' data block contains one navigation subframe extracted from the received C/A L1 signal channel together with time tag information and the observed satellites' pseudo-random code number (PRN). For convenience, we adopt the same data granularity and regard one subframe as the smallest unit within the navigation data processing system. I.e., if parity errors or other inconsistencies are detected in one word, the full subframe is discarded (see paragraph 3.2 below).

The instruments' data packets are transmitted via internet to the GPS data processing center at GFZ. More precisely, the data transmission is controlled by two monitor programs, one running on the local host at the receiver's location and the other running on the remote host at GFZ. The local monitor program performs two tasks. It assembles a raw data file from

all data packets delivered by the receiver instrument during a time period of 15 min. At the same time, the individual data packets are transferred to the remote host at GFZ using Transmission Control Protocol (TCP) streaming mode. The remote program at GFZ collects the packets and assembles another file from the streamed raw data. If the data file on the remote host shows a different byte count compared to the original data file stored on the local host, the latter file is transmitted to GFZ and replaces the remote copy; otherwise, the file on the local host is removed. Typically, the compressed data volume of 96 data files collected by one station during one day occupy about 15 megabytes.

3.2 Navigation Bit Processing

At this point of the general GPS data handling chain the specific navigation data processing starts. First, the data file archive is scanned for files originating from 'PolaRx2' instruments; the corresponding raw data files are read and the navigation packets (data block 'GPSRaw') are extracted.

Second, each data block and corresponding subframe is tested according to the following conditions and constraints.

1. The PRN value stored in the data packet is within the valid range between 1 and 32.
2. The subframe is decoded without parity errors.
3. The preamble in the TLM word (word 1) reads the bit pattern 10001011.
4. Bits 29 and 30 of word 2 (HOW word) and word 10 are zero.

5. The subframe number extracted from the HOW word (word 2) contains a value between one and five in the expected order.

6. The observations occur at a satellite elevations above 5 degrees.

Subframes failing one or more of these tests are discarded.

In the third step the time information is extracted from the HOW word and this value is used to tag the start of the subframe taking into account that the time-of-week (TOW) count message in bits 1–17 refer to the beginning of the following subframe. We do not use the time tag stored in the 'GPSRaw' data block since the latter gives the time of the subframe's reception and not necessarily the time corresponding to the TOW count. Indeed, in a few cases differences were observed.

With the current five receiver configuration the same subframe transmitted by a given GPS satellite is observed by more than one instrument with about 82% probability. In the following we denote the number of times a given subframe is observed as multiplicity M. E.g. $M = 3$ indicates that this subframes has been recorded by three receivers; for unobserved subframes we set $M = 0$.

If $M > 1$ the corresponding data blocks should produce identical subframe bit patterns. If bit pattern differences (in the following denoted as subframe deviations) are found at $M = 2$ both subframes are discarded. If $M > 2$ the subframe with the most prevalent bit pattern is used. We note that within the data set of 103,576,010 subframes with $M \geq 2$ occur-

rences of subframe deviations were not observed; see, however, the remark
in paragraph 4.2.

In the final processing step the validated subframes collected during one
day are assembled according to PRN number, stored in binary data files and
compressed to save disk space. Due to the subframes' high data redundancy
(see paragraph 4.3) compression ratios better than 1:6 are obtained. As
file format the NetCDF format is used, for details we refer to `http://`
`www.unidata.ucar.edu/software/netcdf`. The data files are labelled as
'NAVBIT-GPS-L1CA-*yyyy-ddd-pp-nnnnn*.nc' (after compression the suffix
'.gz' is added), with year *yyyy*, day of year *ddd*, satellite PRN *pp* and number
of validated subframes *nnnnn* ($1 \leq nnnnn \leq 14,400$).

Within the NetCDF data file the navigation data is represented as a
2-dimensional array of 32 bit integers. The least significant 30 bits of each
array element contains the GPS word, the most significant 2 bits are not
used and set to zero. The data array's first dimension describes the GPS
word count $(1, \ldots, 10)$; the second dimension denotes the time tag of each
subframe $(0, 6, 12, \ldots, 86394$ s). Missing subframes are removed from the
data and, therefore, the actual extent of the second dimension may be
smaller than the maximum value of $14,400$. In this case the array's sec-
ond dimension is adjusted accordingly. We note that the time tag refers to
transmitter clock time, not GPS coordinate time at the start of the first
navigation data bit. The NetCDF data file also contains information on the
transmitting satellite's PRN and the multiplicity of each subframe.

The archived data files are available for download from GFZ's Information System and Data Center (`http://isdc.gfz-potsdam.de`).

4 Discussion

In this section some aspects regarding subframe verification, parity checks with false positive outcomes, data redundancy, subframe multiplicity and geographical data coverage are briefly discussed.

During the first 197 days of navigation bit recording the network's five GPS receivers collected 132,288,085 subframes. Figure 1 shows the daily subframe yield of all five receivers combined. Hardware and data transmission problems occasionally cause reductions in the number of observed subframes. The overview of the observation statistics in Table 3 shows that 6,564,419 subframes were discarded leaving 125,723,666 validated subframes.

4.1 Subframe Verification

The closed architecture of the 'PolaRx2' receiver calls for an independent verification of the 'GPSRaw' data packets' content. The verification was performed at GFZ Potsdam with GFZ's OpenGPS receiver. The GFZ OpenGPS receiver is an open source single-frequency receiver based on the OpenSourceGPS hardware (*Kelley*, 2002). During the test period between day of year 305 and 334 (1 to 30 November 2007) the OpenGPS receiver was operating at about 200 m northwest of receiver antenna location of 'PolaRx2'
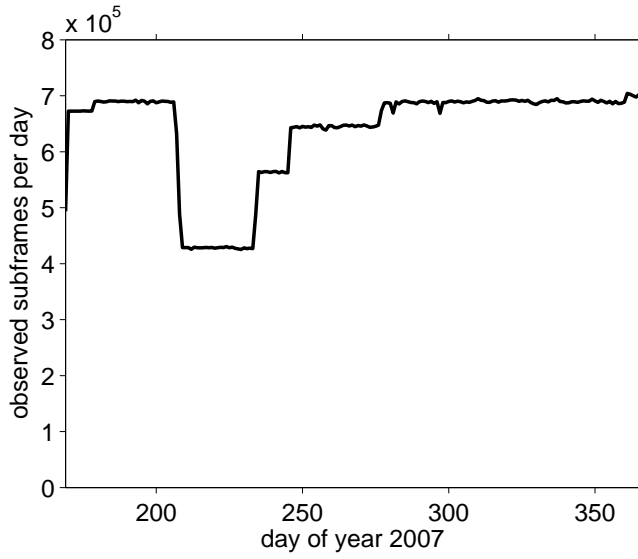
Fig. 1 Number of subframes per day observed between 18 June 2007 (day of year 169) and 31 December 2007 (day of year 365). Hardware and data transmission problems intermittently cause reductions in subframe yield.

instrument 'POTS'. Navigation bit data from both instruments were collected, processed and compared. We observed an almost perfect agreement between subframes concurrently observed by the two instruments, only one out of 1,807,765 subframes exhibits a different bit content.

4.2 False Positives in Parity Checks

It is well known that the parity bits (see section 3) do not provide a complete protection against multi-bit errors. In our data set we found 25 instances where subframes with $M > 1$ passed the first five of the six validation criteria, in particular the parity test, listed in section 3, but still exhibited different bit structures. In 21 of these 25 events the subframes were recorded

Table 3 Summary of subframe observation statistics based on 197 days of data recordings.

number of observation days	:	197
number of receivers	:	5
validated subframes from receiver 'JOGJ'	:	18,310,912
validated subframes from receiver 'MIZU'	:	27,182,780
validated subframes from receiver 'OUS2'	:	23,210,889
validated subframes from receiver 'POTS'	:	28,307,089
validated subframes from receiver 'ZWE2'	:	28,711,996
total number of validated subframes	:	125,723,666
number of failed parity checks	:	64
number of subframes below elevation mask	:	6,564,355
total number of discarded subframes	:	6,564,419

by three receivers ($M = 3$). In all 21 observations two subframes are identical and only one differs. Without exception the deviating subframe was observed at elevation angles below $3°$, whereas the other two subframes originated from observations at higher angles. The obvious interpretation is that subframes observed at very low elevation angles and, thus, at low SNR values are more likely to suffer multiple bit errors.

Similarly, in the remaining 4 cases with $M = 2$ one of two subframes was always recorded below $3°$. Without independent confirmation we can only assume that the low elevation observation produced the erroneous subframe.
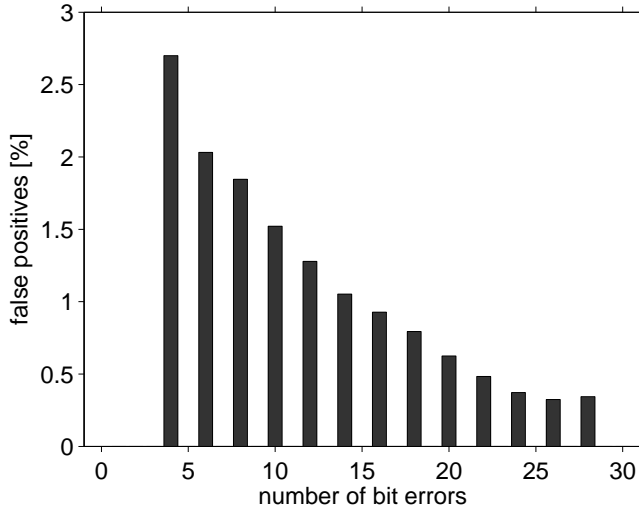
Fig. 2 Occurrence probability for false positive parity checks with n bits flips within the same data word. For odd n, $n = 2$ and $n = 30$ no false positives are observed.

Based on these results the 5° elevation angle threshold was added to the subframe validation criteria (section 3).

An elementary simulation study was performed to model parity checks with false positive outcome. In the simulation random n-bit errors ($1 \leq n \leq 30$) are added to the same (arbitrarily selected) data word and a parity test on the complete subframe is performed. The procedure is repeated 100,000 times for each value of n. The result is plotted in Fig. 2 and shows a probability for a false positive outcome between 0.3 and 2.7% provided n is even and $4 \leq n \leq 28$. Thus, if 1000 subframes suffer 4-bit errors within the same word on average 27 subframes will erroneously pass the parity check.

4.3 Subframe Data Redundancy

It is well known that the data content of GPS navigation messages are highly redundant; i.e., subframe contents can be correctly predicted from past subframes with high probability. Subframe prediction might be relevant for future radio occultation receivers, tracking the carrier signal with four quadrant phase extraction (*Beyerle et al.*, 2006). For a data set of 3,486,448 validated subframes recorded in 2007 between day of year 200 and 300 (19 July to 27 October 2007) predicted subframes were assembled taking into account the change in TOW count bits. The prediction was found to be correct in 31,197,659 out of 31,757,678 cases corresponding to a prediction probability of 98.3% in good agreement with *Beyerle et al.* (2006). Furthermore, our analysis confirm the well known fact that changes in subframe content usually occur every 2 hours at even hours GPS time.

4.4 Subframe Multiplicities

With the current configuration of five receivers derived multiplicities M vary between zero (subframe is not observed) and four (subframe is observed concurrently by four receivers) with a mean value of 1.43. In Fig. 3 the histogram distribution of multiplicity occurrence is plotted for 65,153,955 validated subframes (dark grey), the network did not record about 22,945,245 subframes ($M = 0$). I.e., during the observation period 74% of the transmitted subframes (65,153,955 / (65,153,955+22,945,245)) could successfully be recorded. 49% (43,006,299 out of 88,099,200 subframes) were observed con-
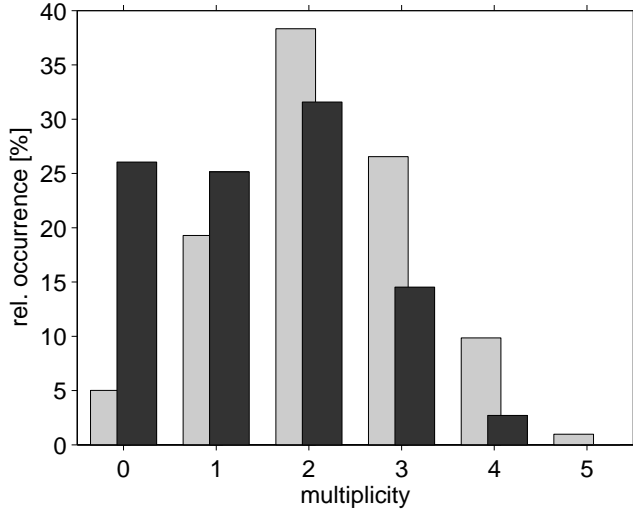
Fig. 3 Histogram distribution of observed subframe multiplicities for the time period between 19 July 2007 and 31 December 2007 (dark grey). A simulation result derived from a seven receiver configuration (five existing and two additional observation sites) for the time period between 1 to 31 July 2007 is shown in light grey.

currently with two receivers ($M = 2$). Clearly, values of $M \geq 3$ would be favourable since they allowed to determine the erroneous subframe if two or more subframe exhibited the same bit pattern. In the present data set only 17% (15,180,697 out of 88,099,200 subframes) of the subframes reach or exceed $M = 3$.

4.5 Subframe Data Coverage

In order to obtain a 100% subframe yield additional receivers capable of navigation data recording need to deployed in the network. To support the

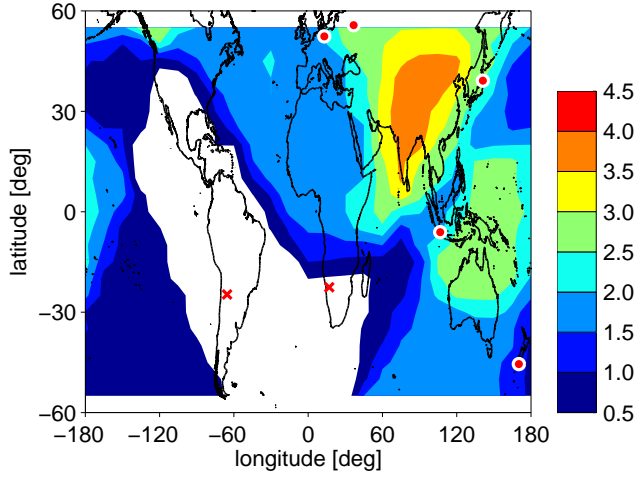**Fig. 4** Geographical distribution of subframe multiplicity averaged over $15° \times$ $5°$ latitude/longitude bins. The location of the current receivers are marked as red dots; two more 'PolaRx2' receivers will be set up at Salta, Argentina and Windhoek, South Africa (red crosses).

selection of new receiver locations we relate observed multiplicities with geographical locations. The dependence is determined by projecting the GPS satellite's location at the time of subframe transmission radially onto the Earth's surface. Then, all multiplicity values are sorted into bins extending over $5°$ latitude and $15°$ longitude and the mean value in each bin is calculated. The result obtained for the time period between day of year 169 and 365 (18 June and 31 December 2007) is plotted in Fig. 4. Since GPS orbit inclinations are about $55°$ the projected locations are confined to a latitude band between about $55°$N and about $55°$S. Fig. 4 shows that with five receiver stations a significant data gap exists on the South Atlantic, South and Central America, parts of North America and the West Pacific.

To fill this gap and to achieve a more complete data coverage two receivers at Salta, Argentina and Windhoek, South Africa will be added to the network in the near future. We have performed a simulation study using seven receivers located at the positions given in Table 2. The elevation angle mask is taken to be 5° and the simulated time period extends from 1 to 31 July 2007. The multiplicity distribution, derived from the simulated observations, is plotted in Fig. 3 (light grey). It shows that with these two additional receivers a yield of about 95% will be achieved and an expected fraction of 76% observation will exhibit multiplicities equal to or higher than 2. To further reduce the number of missed subframes another receiver station in North or Middle America would be needed.

5 Conclusions

Since 18 June 2007 five GPS receivers within GFZ's ground-station network are configured to monitor the L1 C/A navigation data messages. The messages are collected, processed and validated at GFZ Potsdam on a routine basis. The data set of validated message subframes is archived at GFZ's Information System and Data Center (ISDC) with a time delay of 7 days and open to the scientific community for non-commercial purposes. The ISDC may be accessed through the url `http://isdc.gfz-potsdam.de`.

Currently about 75% of the transmitted subframes are recorded and stored in the data base. Within the year 2008 another two receivers capable of navigation bit recording are scheduled for addition to the network allow-

ing for a 95% subframe yield. Further receiver instruments need to added to the network to achieve complete coverage and to guarantee at least threefold multiplicities.

Acknowledgments

References

Beyerle, G., T. Schmidt, J. Wickert, S. Heise, M. Rothacher, G. König-Langlo, and K. B. Lauritsen (2006), Observations and simulations of receiver-induced refractivity biases in GPS radio occultation, *J. Geophys. Res.*, *111*, D12101, doi:10.1029/2005JD006673.

Galas, R., J. Wickert, and W. Burghardt (2001), High rate low latency GPS ground tracking network for CHAMP, *Phys. Chem. Earth (A)*, *26*, 649–652.

Kelley, C. (2002), Internet-based open source software for learning about GPS, *GPS Solutions*, *6*(3), 201–205, doi:10.1007/s10291-002-0026-2.

Kursinski, E. R., G. A. Hajj, J. T. Schofield, R. P. Linfield, and K. R. Hardy (1997), Observing Earth's atmosphere with radio occultation measurements using Global Positioning System, *J. Geophys. Res.*, *19*(D19), 23,429–23,465.

Lowe, S. T., J. L. LaBrecque, C. Zuffada, L. J. Romans, L. E. Young, and G. A. Hajj (2002), First spaceborne observation of an Earth-reflected GPS signal, *Radio Sci.*, *37*(1), doi:10.1029/2000RS002539.

Masters, D., P. Axelrad, and S. Katzberg (2004), Initial results of land-reflected GPS bistatic radar measurements in SMEX02, *Remote Sensing of Environment*, *92*, 507–520.

Misra, P., and P. Enge (2002), *Global positioning system: Signals, measurements, and Performance*, Navtech Seminars and GPS Supply, Alexandria, VA, USA.

Septentrio (2004), *PolaRx2 User Manual*, Septentrio, Ubicenter, Philipssite 5, B-3001 Leuven, Belgium.

Sokolovskiy, S., C. Rocken, D. Hunt, W. Schreiner, J. Johnson, D. Masters, and S. Esterhuizen (2006), GPS profiling of the lower troposphere from space: Inversion and demodulation of the open-loop radio occultation signals, *Geophys. Res. Lett.*, *33*, L14816, doi:10.1029/2006GL026112.

Treuhaft, R. N., S. T. Lowe, C. Zuffada, and Y. Chao (2001), 2-cm GPS altimetry over Crater lake, *Geophys. Res. Lett.*, *22*(23), 4343–4346.

Wickert, J., R. Galas, G. Beyerle, R. König, and C. Reigber (2001), GPS ground station data for CHAMP radio occultation measurements, *Phys.*

Chem. Earth (A), *26*, 503–511.

Wickert, J., et al. (2008), GPS radio occultation: Results from CHAMP, GRACE and FORMOSAT-3/COSMIC, *Terrestrial, Atmospheric and Oceanic Sciences*, in print.

Yunck, T. P., C.-H. Liu, and R. Ware (2000), A history of GPS sounding, *Terrestrial, Atmospheric and Oceanic Sciences*, *11*(1), 1–20.