Environnement Canada
Centre météorologique canadien

Environment Canada
Canadian Meteorological Centre

Processing of Ground-based GPS Observations at the Canadian Meteorological Centre

by
Stephen Macpherson
ARMA

Version 3.0

May 2021

Revision history		
Version	**Date**	**Remarks**
1.0	07/24/14	S. Macpherson (ARMA). First version.
1.1	09/09/14	S. Macpherson. Updates related to E-GVAP data and thinning1.
1.2	11/20/14, 06/09/15	S. Macpherson. Updated for new version of GEM model vertical coordinate where lowest level (2m above surface) pressure is actually P_{2m} instead of P_{sfc} (P0). New GPS site maps and some minor edits (June 2015).
2.0	07/29/19	S. Macpherson. Major update including replacement of NOAA GPS network with UCAR SuomiNet and changes to EnVar (MIDAS source modules) that impact GB-GPS observations.
3.0	05/12/21	S. Macpherson. Updated to include new ZTD bias correction scheme implemented in Innovation Cycle 3 (IC-3). Miscellaneous other updates including updates to site maps.

Table of Contents

1. Introduction

This document describes the processing of ground-based GNSS (GPS) observations at CMC. Ground-based GPS (GB-GPS) observations were introduced into the CMC data assimilation system in November 2014 as part of version 4.0 GDPS and RDPS deliveries to Operations.

As in space-based GPS radio occultation (GPS-RO) meteorology, atmospheric properties are remotely sensed by measuring the effect of atmospheric refractivity on the reception of GPS microwave signals using a high-precision GPS receiver. In GPS-RO, the receiver is located on a polar orbiting low-earth-orbit (LEO) satellite. In the case of ground-based GPS, the receiver is located at a fixed position on the surface of the earth. (It is also possible to have a receiver on a moving platform such as an oil rig or ship.) Only integrated atmospheric properties are sensed with GB-GPS, while profiles of atmospheric properties (e.g., temperature) can be retrieved with GPS-RO.

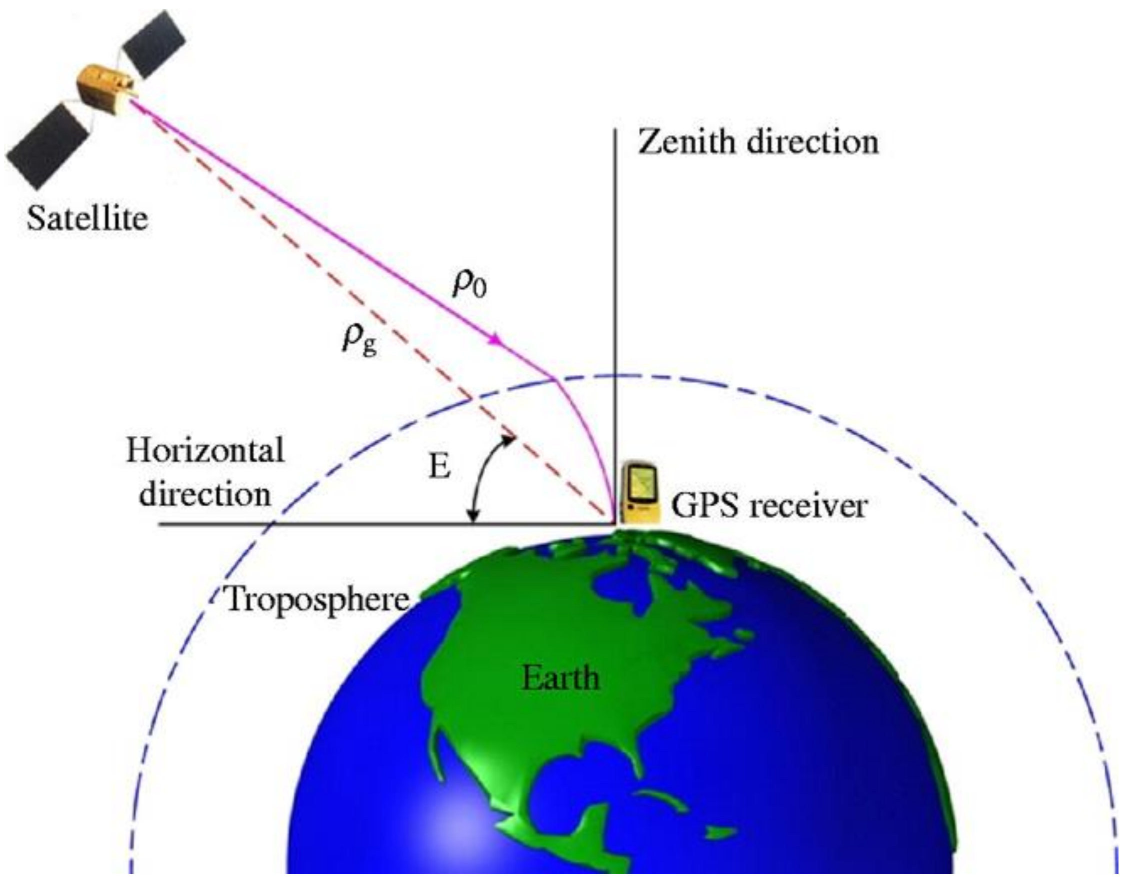


Figure 1. Principles of GB-GPS measurement

There are thousands of high-precision GPS receivers located at GPS sites all over the globe, installed and maintained primarily for the purpose of precise positioning applications. Special geodetic software is used to process the GPS signals received at GPS sites in order to determine precise site X-Y-Z coordinates. As the GPS signals travel from satellite to receiver, they are "delayed" due to electric charge in the ionosphere and refractivity in the neutral atmosphere (Figure 1). This delay is considered a "nuisance parameter" in the computation of precise GPS site coordinates, seriously reducing the accuracy of position measurements if not accounted for. However, geodetic GPS signal processing software accurately estimates the delay so site coordinates can be obtained with millimetre-level accuracy.

The ionospheric delay can be accurately determined and removed using dual-frequency GPS receivers. This leaves the neutral atmospheric delay, which is due to the integrated effects of atmospheric mass (density) and the presence of water vapour in the troposphere, more specifically the electric dipole of H_2O molecules. At any given time, a GPS receiver receives signals from multiple GPS satellites at different elevation angles (Figure 2), and there is a "slant delay" associated with each satellite. These multiple slant delays can be mapped to the zenith direction and averaged using special mapping functions, giving a single zenith delay value. A cut-off angle is typically applied to eliminate data from satellites with very low elevation angles.

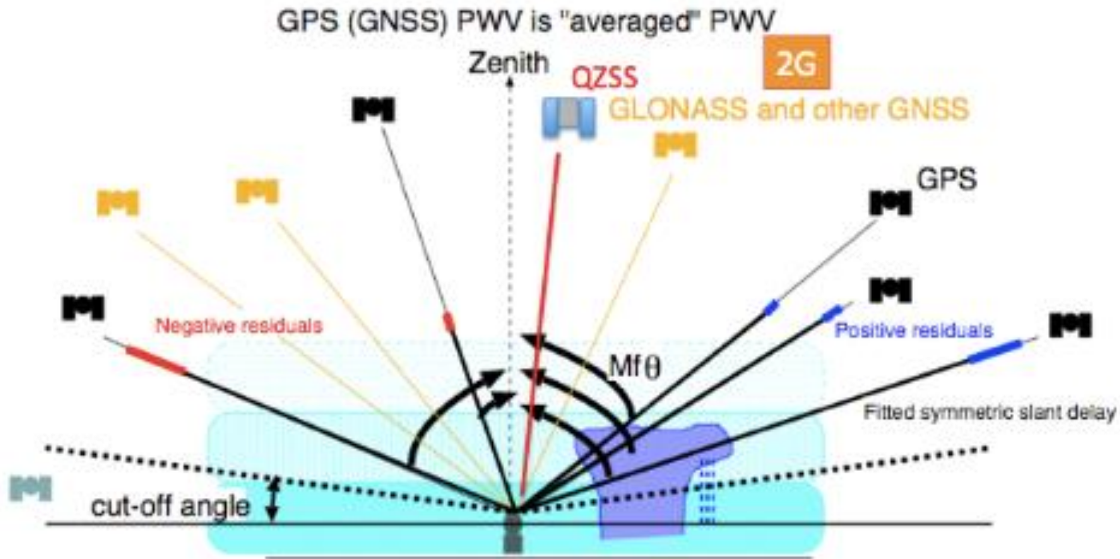


Figure 2. GB-GPS geometry

The total neutral atmospheric zenith delay is called the **zenith total delay** (ZTD). Typical ZTD values at sea-level are on the order of 2.5 metres excess path length. The contribution to ZTD from total atmospheric mass is directly related to the surface pressure at the GPS site (at the receiver antenna height) and is called the zenith hydrostatic or dry delay (ZHD), while the delay due to integrated water vapour above the site is called the zenith wet delay (ZWD). Typical sea-level ZHD is ~2.3 metres and ZWD ranges from 0.0 to 0.6 metres depending on integrated water vapour. In GB-GPS meteorology, integrated water vapour or IWV (kg m^{-2}) is normally expressed as precipitable water PW (mm). While surface pressure, and hence ZHD, can be accurately determined using a collocated or nearby barometer, accurate measurements or estimates of PW with high temporal frequency are generally not available.

GPS data processing software estimates the ZTD using GPS receiver data, GPS satellite orbits and clocks, previously determined "first-guess" precise site coordinates, and climatological surface pressure values. A high degree of accuracy for the GPS satellite orbital parameters (orbits and clocks) is needed to obtain precise site coordinates and accurate ZTD estimates from raw GPS receiver data processing. The International GNSS Service (IGS) is the main supplier of such orbit and clock products for the geodetic community and has a slightly less accurate near-real-time orbit and clock product available for near-real-time applications such as GPS meteorology.

As stated before, ZTD estimates are generally quite accurate so it is possible to use ZTD to retrieve PW at a GPS site provided surface pressure (Psfc) and temperature (Tsfc) measurements are available. The Psfc measurements are needed to compute the dry (ZHD) part of ZTD so that the wet delay ZWD can be determined accurately (ZWD = ZTD − ZHD(Psfc)). PW with accuracy on the order of 2 mm can then be retrieved from ZWD and Tsfc. Such PW observations are largely unaffected by weather except in extreme cases such as deep/severe convective storms. This is the basis of ground-based GPS meteorology.

In the 1990's in the USA, it was recognized that a GB-GPS PW observation network could be established at relatively low cost by simply installing Automatic Weather Stations (AWS) at existing geodetic network GPS sites. ZTD could be estimated at each site using freely available geodetic precise positioning software (e.g. GAMIT, GIPSY or Bernese), with Psfc and Tsfc observations from collocated AWS allowing PW to be retrieved at potentially high temporal frequency. The network could be augmented by installing additional dedicated GPS receivers and weather stations at strategically located sites.

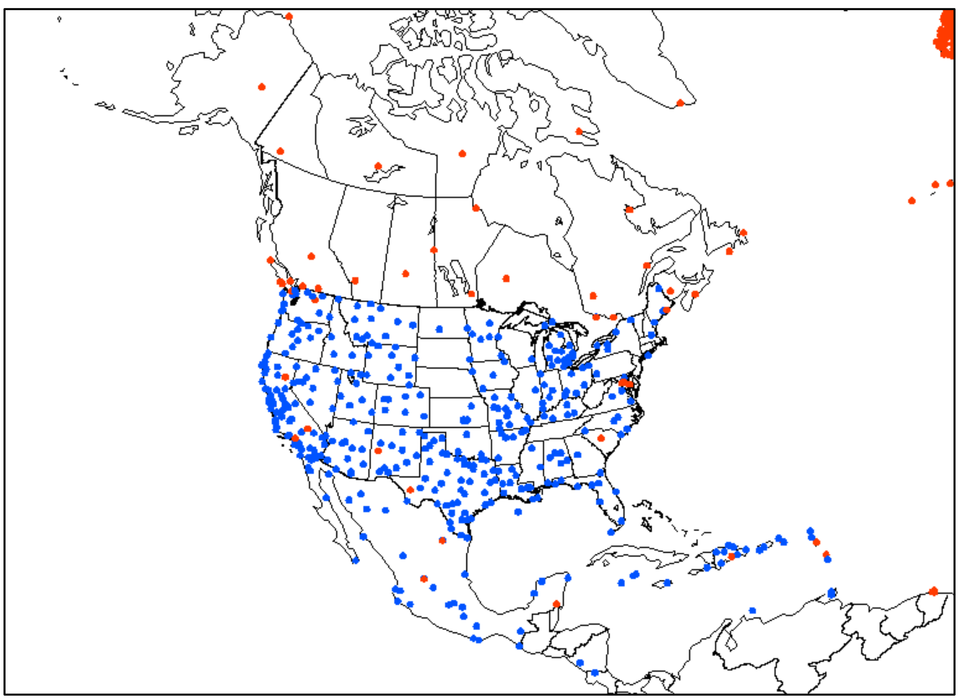


Figure 3: North American GB-GPS Sites (Feb 2019). **UCAR** sites in **blue**. **E-GVAP** in **red**.

The NOAA Forecast Systems Laboratory (FSL) used this strategy to create the real-time NOAA GPS-PW network in 1994, while a similar network, the Suomi-Net, was created by NCAR for research purposes around the same time. NOAA determined that dedicated collocated weather stations (AWS) were not needed at a GPS site if sub-hourly Psfc and Tsfc observations from a nearby (within 50 km) AWS, such as those located at a METAR or SYNO station, were available. Assimilation of the NOAA network ZTD and surface met data at CMC commenced in November

2014. In 2016, NOAA made the decision to contract out their GPS data processing to a private company and hence the free NOAA GB-GPS data became unavailable in November 2016.

In January 2019, CMC acquired operational access to GB-GPS data from the Suomi-Net processed by UCAR. The data are very similar to the NOAA data in terms of number of sites, reported elements, observation frequency and data quality. Operational assimilation of the UCAR GB-GPS data commenced on 23 July 2019. The UCAR network includes over 400 active GPS receiver sites located mostly in the continental USA (Figure 3), providing 30-minute reports of ZTD, ZWD and PW as well as Psfc, Tsfc and RHsfc observations from collocated or nearby AWS. The Psfc data are barometer measurements hydrostatically adjusted to the GPS antenna height. The UCAR data are retrieved as NetCDF files from an FTP server and the observations are encoded into in-house BURP ADE dbase files.

EIG EUMETNET is a grouping of over 30 European National Meteorological Services that provides a framework to organise co-operative programmes between its members in the various fields of basic meteorological activities. The EIG EUMETNET GNSS water vapour (E-GVAP) network was set up in April 2005 to provide its members with European GNSS ZTD estimates for operational meteorology in near real-time. As is the case for the NOAA and UCAR networks, the vast majority of GNSS sites were installed for positioning purposes by geodetic institutions and private companies. Raw receiver data from the GNSS sites are collected by 10 GNSS analysis centres (ACs), mainly geodetic institutions and universities, which process the data to estimate ZTDs and other parameters (see Table 1). The data are then forwarded to a data server at the UK Met Office for distribution to meteorological institutes. Most ACs process data from many receivers at once using a network approach, but some ACs have used the precise point positioning (PPP) approach. The PPP approach uses precise GNSS clocks and orbits calculated from a global network to calculate precise receiver positions and estimate the ZTD at a site equipped with a single or dual frequency GPS receiver.

Currently the E-GVAP network consists of ~3200 GNSS sites and ~7000 site-AC combinations (as more than one AC can process a given site). The sites are mainly in Europe (Figure 4b) with some global GNSS sites to meet the needs of global data assimilation at NWP centres. There are about 30 sites located in Canada (Figure 3). Average site density is high over Europe compared to the UCAR network (Figure 4a, 4c). The data are received in BUFR files on the GTS. Unlike the UCAR network, only a very small number of sites report surface weather data (Psfc, Tsfc, RHsfc). However, some ACs do provide ZWD and PW observations in addition to ZTD, using surface met data from the closest SYNO or METAR station, but unlike UCAR do not include the actual AWS data in the reports. Observation frequency is 15 minutes, 30 minutes or 60 minutes. One AC (LPT) also provides data every 5 minutes for ~45 sites. See Table 1 for details. Around 1200 sites in the E-GVAP network are processed by two or more ACs, giving multiple ZTD estimates for the same site/time. Some ACs also provide two solutions for the same site so there can be as many as 17 different ZTD observations for a given site/time. There are also ~35 “super-sites” that are processed by more than 5 ACs. Data quality for each AC/station assessed from innovation monitoring statistics is used to select the “best” ZTD observations for assimilation for sites with multiple observations (see section 4(b)).

Analysis Centre (AC)	Country (Location of AC)	Approximate Number of Sites	Frequency (minutes)	Number of Obs/ 6h
GFZ	Germany	580 / 450	15	25
ASI	Italy	560 / 270	15, 60	25, 5
IGE *	Spain	330 / 110	15	30
LPT	Switzerland	200 / 45	30, 5	12, 66
NGA	Sweden, Norway, Finland	700 / 360	15	24
MET *	United Kingdom (UK Met Office)	220/200/240/200/60	15	25,30
ROB*	Belgium	620 / 260 / 210	15	25, 30, 46
SGN *	France	380/240/200/30/20	15	25
WUEL*	Poland	160	30	15
WUHN	China	300	15	24
UCAR	United States (CONUS)	400	30	12

Table 1: E-GVAP GNSS Network Analysis Centres (ACs) with AC identifier, main country sites processed, number of sites processed, observation frequency, and the total number of observations per site in a 6h assimilation window. UCAR is also shown for comparison purposes. The asterisk (*) indicates that the centre issues "duplicate" reports at one minute before the hour and on the hour. Some ACs process multiple groups of stations independently. For these ACs, the number of sites in each group are given, separated by slashes (/). See the E-GVAP web site for more details: http://egvap.dmi.dk/.

All but two analysis centres process up to 5 different groups of sites under more than one AC "name". Examples are MET (METG, METO, MTRS,…) , GFZ (GF1G, and GF1R) and NGA (NGA1, NGA2). The groups of sites may contain common sites. Different processing strategies may be applied for the groups, or the site coverage may differ (e.g. regional vs global where "G" and "R" are often added respectively to the 3-character base AC identifier). The 4-character AC name forms part of each station name, (see Section 3).

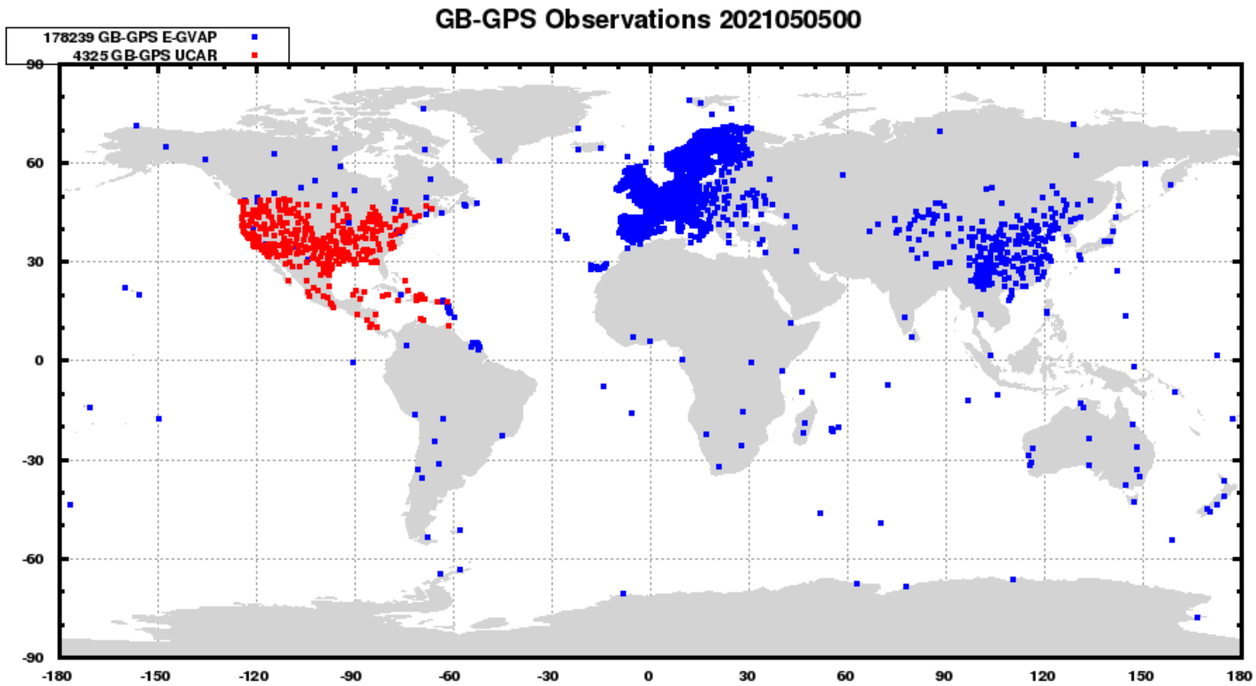


Figure 4a:. Map of GB-GPS sites for which ZTD data are received at CMC (May 2021). **UCAR** network sites are plotted in **red** and **E-GVAP** network sites in **blue**.

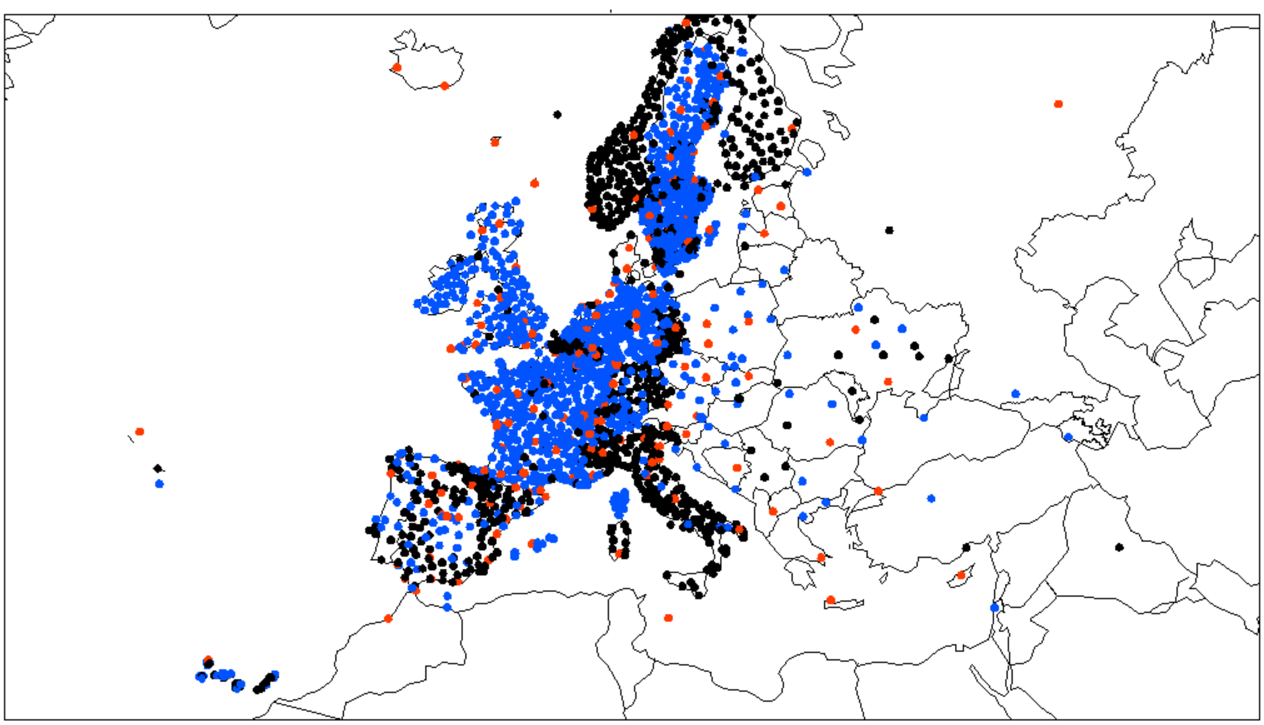


Figure 4b: E-GVAP network GPS sites in Europe (May 2021). Color code: **black = site processed by 1 AC, blue = 2−5 ACs, red = 6 or more ACs,** where "AC" refers to the unique 4-character AC part of the site-ac station name, e.g. "GF1R", "METG", "NGA1").

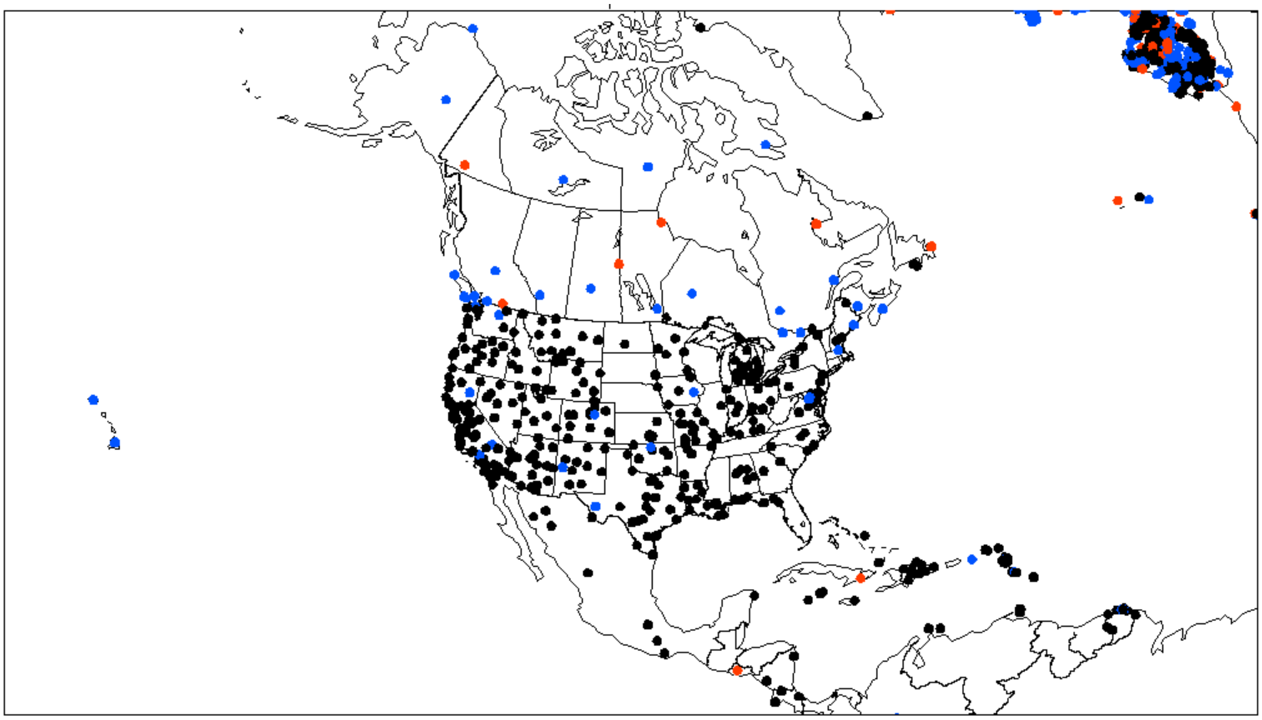
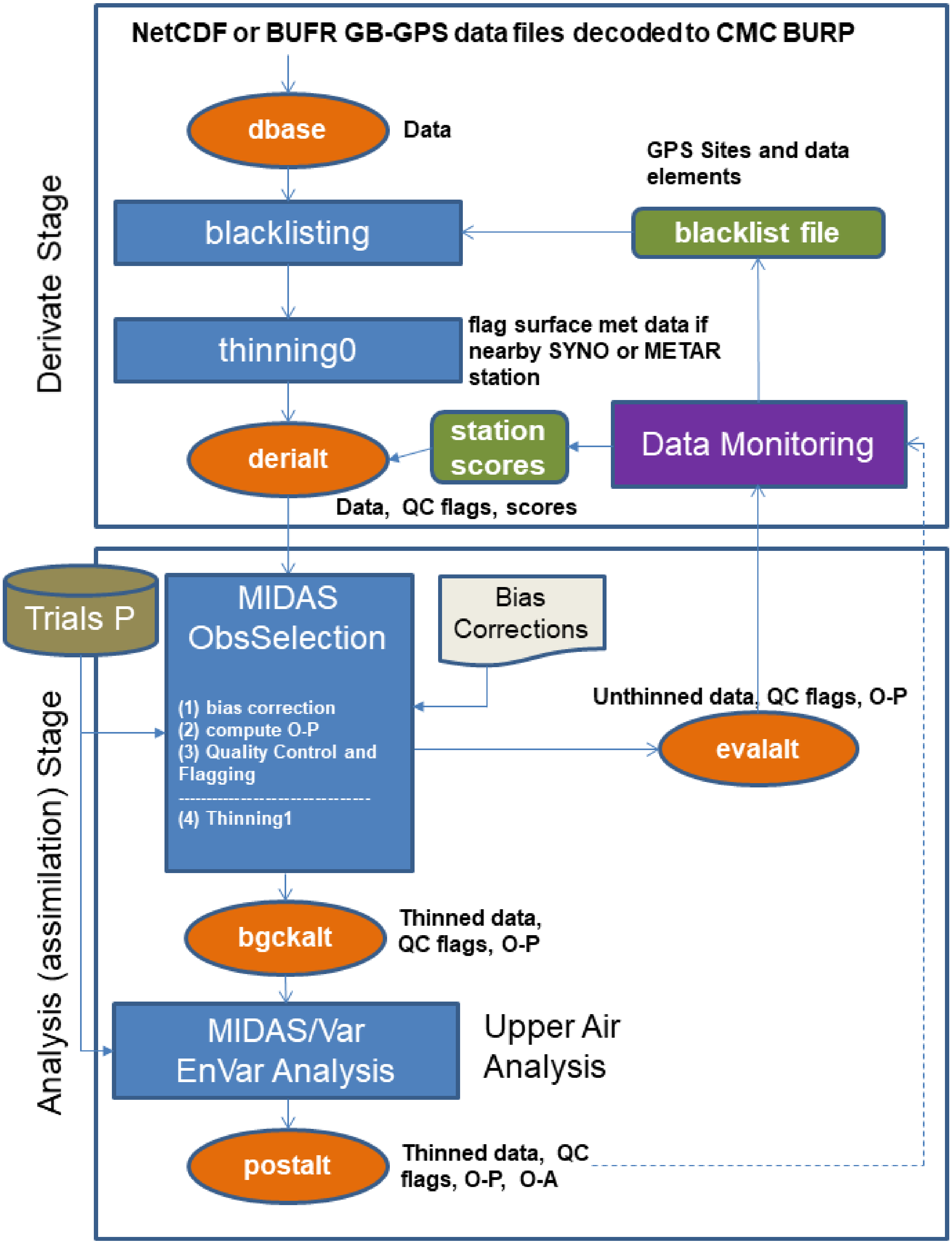
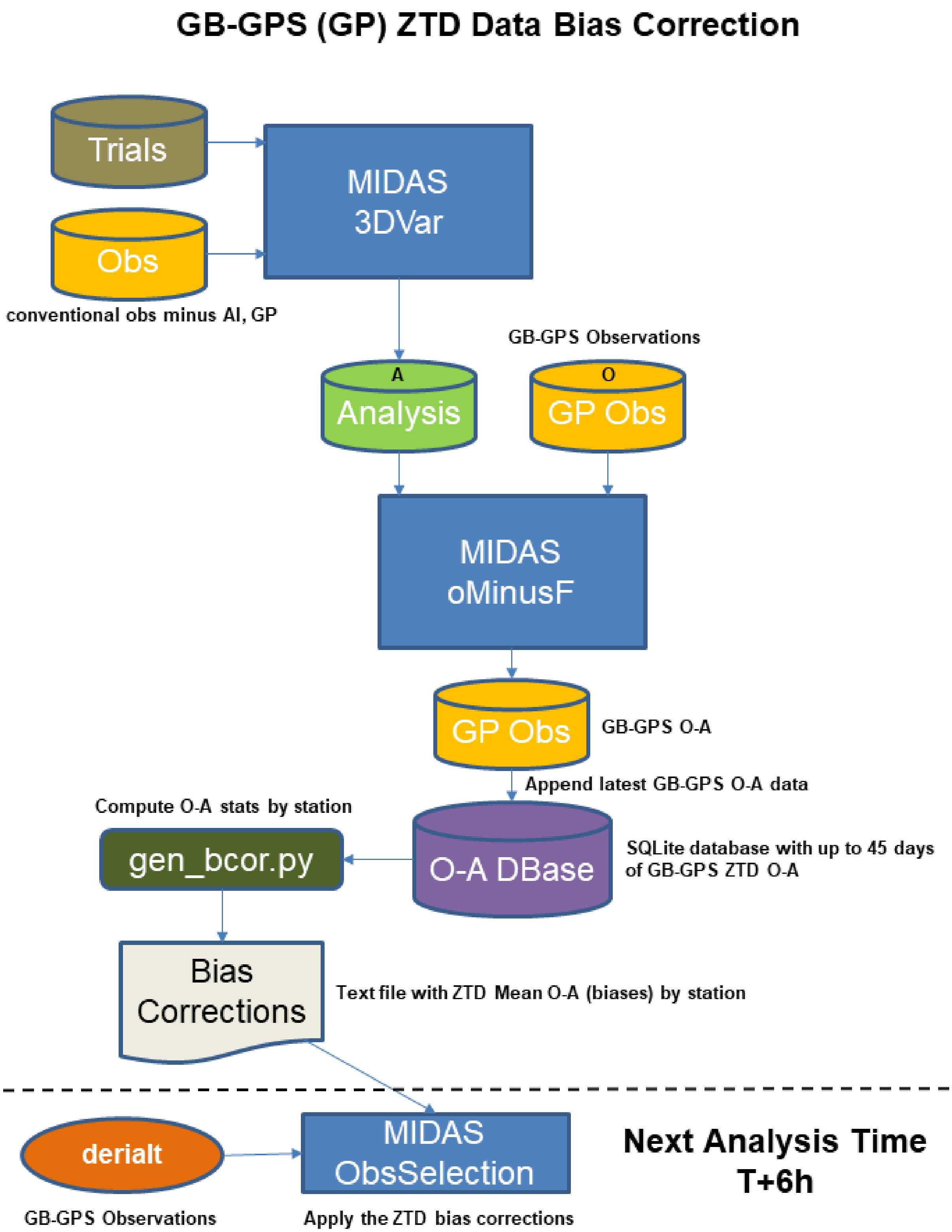


Figure 4c: As Fig. 4b but for North America (UCAR and E-GVAP sites)

2. Flowcharts of ground-based GPS observation processing




GB-GPS (GP) ZTD Data Bias Correction
Trials
Obs
MIDAS
3DVar
conventional obs minus AI, GP
GB-GPS Observations
A
Analysis
O
GP Obs
MIDAS
oMinusF
GP Obs
GB-GPS O-A
Append latest GB-GPS O-A data
Compute O-A stats by station
gen_bcor.py
O-A DBase
SQLite database with up to 45 days
of GB-GPS ZTD O-A
Bias
Corrections
Text file with ZTD Mean O-A (biases) by station
derialt
MIDAS
ObsSelection
Next Analysis Time
T+6h
GB-GPS Observations
Apply the ZTD bias corrections

3. Observation details

Code type 189 GB-GPS (GP Family)

E-GVAP GB-GPS data are available on the GTS as BUFR files. The UCAR network data are available as NetCDF files on a UCAR server. Station elevations in the files are the GPS receiver antenna heights. The BUFR header for GB-GPS data is of the form ISX*i*14 EGRR *ddhhhh* BUFR…. where *i* = [A-L], N, S, T, X and *ddhhhh* is the day of the month *dd* and hour *hhhh* UTC.

All GB-GPS sites have 9 character station IDs that consist of a 4-character site name {*nnnn*} and a 4-character analysis centre {*aacc*} separated by a dash (-). The 9-character station ID is *nnnn-aacc*. The analysis centres (ACs) are given in Table 1. In the table, only 3-character AC identifiers are given for most ACs. However, the AC names in the station IDs (*aacc*) are always 4 characters. The fourth character is either the underscore "_" or a number or a letter. Examples of station IDs for ACs MET and SGN are HERS-METO, CHUR-METG, and ALAC-SGN_. Sites are often processed by more than one AC (e.g. CHUR-METG, CHUR-ROBG). Also, some ACs process the same site twice (e.g., AC LPT and site LECH with station IDs LECH-LPT_ and LECH-LPTX). In some cases, there are two or more sites (with different site names) at the same location (e.g. ZIMM, ZIM2).

A list of the elements found in a GP family BURP file is given in Table 2.

Element	**Description**	**Units**
13016	PW	mm
10004	**Surface Pressure**	Pa
12001	***Surface Temperature***	K
13003	Surface RH	%
15031	**ZTD**	m
15035	ZWD	m
15032	ZTD error (FERR)	m
12203	***Dew point Depression***	K
8021	Time significance	code
4025	Time period	min
40026	ZTD Quality Score	n/a
15234	ZTD Bias Correction	m

Table 2: Details and BUFR/BURP codes. Assimilated elements are in **bold**. Elements in *italics* were also assimilated in the RDPS until the NOAA/FSL network data were lost in 2016. These data (primarily from the replacement UCAR network) may once again be assimilated in the future.

- **ZTD** (element 15031) is the primary GB-GPS observable for meteorology and is the total zenith delay estimate for the site obtained from geodetic software during raw GPS data processing. ZTD accuracy depends on many factors such as the particular data processing method applied, the accuracy of the a-priori site coordinates and other "first-guess" data, the quality of the GPS satellite orbits and clocks used, the characteristics of the GPS receiver and antenna, siting issues that could affect signal reception (e.g. signal multipath, surface reflections), and the characteristics of the GPS satellite constellation in view during the ZTD estimation epoch (number and positions of satellites). The GPS network topology is also a factor for the "network" data processing method. Finally, significant disturbances in the ionosphere and severe storms in the troposphere may affect the accuracy of ZTD estimates.

- **Surface Pressure (Psfc), Temperature (Tsfc)** and **RH** are observations from collocated automatic surface weather stations. For UCAR files, the data are mostly taken from an Automated Surface Observing Systems (ASOS) installed at METAR stations within 50 km of the site. Surface Pressure is the barometer measurement adjusted hydrostatically to the GPS antenna height (station height in the BURP file). Currently, only the UCAR sites and a few E-GVAP sites report surface weather observations. To avoid potential data duplication, surface weather observations are not assimilated if the GPS site is close to a surface (SF) SYNO site (GDPS) or SYNO/METAR site (RDPS).

- Surface Pressure and Temperature are used to retrieve **ZWD** and **PW** from ZTD. PW is only reported to the nearest millimetre. ZWD and PW are reported for most UCAR sites and some E-GVAP sites. The Psfc and Tsfc observations used to compute PW at these sites are taken from the nearest surface weather station (SYNO/METAR) but the Psfc and Tsfc data are generally not included in the reports from the E-GVAP network. PW data are useful for NWP verifications as in the EMET system and for assessing biases in radiosonde and aircraft humidity observations.

- The **ZTD error** (element 15032) or formal error (FERR) is a measure of ZTD estimate accuracy/quality. For UCAR data, the normal range of values is 4−10 mm so large values (> 15 mm for example) generally indicate a problem with the ZTD estimate for the site. FERR is determined somewhat differently for the E-GVAP stations. The normal range (1−4 mm) is lower than for UCAR.

- **Dew point Depression** (12203) is computed locally from the surface met observations and added as a new element in the BURP file in the conversion of the GB-GPS source files to BURP dbase files. It is referred to as **DPDsfc** in the rest of the document. It is currently the conventional assimilated humidity "observation" at CMC.

- **ZTD Quality Score** (40026): From monthly monitoring *zdscores* file. Added in the THINNING0 (derivate) stage. See section 4 for details (THINNING0 and THINNING1 subsections).

- **ZTD Bias Correction** (15234): This element, added to the BURP file during the analysis stage of the assimilation cycle, is the value of the bias correction (if any) applied to the ZTD value. This value is added to the raw ZTD to get the corrected ZTD. To retrieve the original raw ZTD value simply subtract the bias correction value from the "corrected" ZTD.

GB-GPS (GP family) data assimilated at CMC (as of May 2021):

- ZTD
- Surface Pressure (if no nearby SYNO/METAR station)

4. Processing details (refer to the flowcharts in Section 2)

(a) Derivate file generation stage:

BLACKLISTING

The data are first blacklisted according to a list of GPS sites and elements in a blacklist file. Zeros (0) and ones (1) are used to indicate "good" and "blacklisted" elements respectively for each site in the file. Data quality control (QC) flag bit 8 is set in the BURP file for all blacklisted data. The blacklist file is updated monthly by CMDA according to GPS site O−P statistics from monitoring of the data. The current blacklisting criteria are given in Table 3.

Element	**Mean O−P (max)**	**StdDev O−P (max)**
ZTD	12 mm	45 mm
Psfc	1 hPa	6 hPa
Tsfc	4 K	6 K
DPDsfc	4K	6K

Table 3: Blacklisting criteria for GB-GPS data with the site O−P stats taken from CMDA monthly monitoring.

In addition, data are blacklisted for a site/element if the percentage of gross errors exceeds 25%, where gross error is determined by a very large Abs(O−P). Data are also blacklisted if there have not been enough reports to generate reliable O−P statistics. Finally, ZTD data are blacklisted if the ZTD formal error (element 15032 in the BURP files) exceeds 15 mm.

The blacklisting program source code can be found here:
https://gitlab.science.gc.ca/derivate-programs/deriv.blacklistgbgps/blob/master/src/deriv.blacklist_gbgps.f90

THINNING0

Here thinning0 refers to an initial rejection of data based on a pre-determined data selection process. All surface met data (Psfc, Tsfc, DPDsfc) for a GPS site are flagged (rejected) if there is a surface SYNO station within 50 km of the site (data QC flag bit 11 set). This is done to avoid potential duplication of surface met data in the assimilation system, as surface met data for some GPS sites come from AWS installed at nearby METAR stations collocated with SYNO stations. As METAR data are also assimilated in the RDPS, a check for nearby METAR stations is also done when dealing with GB-GPS observations destined for the RDPS. At this stage, the station ZTD quality scores based on the previous month monitoring (from the *zdscores* file) are added to each GB-GPS observation for use later on in the THINNING1 step.

The thinning0 program source code can be found here:
https://gitlab.science.gc.ca/derivate-programs/deriv.thinning0gbgps/blob/master/src/deriv.thinning0_gbgps.f90

(b) In an assimilation cycle:

OBSERVATION ERRORS

Observation errors for GPS surface met data Psfc, Tsfc and DPDsfc are the same as specified for equivalent surface family SF (SYNO) station data.

An error model in the *EnVar* code is applied to assign the ZTD observation error dynamically at analysis time. The errors as a function of observed ZWD are determined using the Desroziers technique, which uses O−P and O−A statistics over summer and winter seasons to estimate the errors. The actual error for each observation at analysis time is computed from the observed ZWD using the coefficients from a linear regression of ZTD error with ZWD (with the ZTD errors obtained from the Desroziers method). Observations from sites in warm humid regions (high ZWD/PW) will have higher ZTD errors than those in drier regions (low ZWD/PW). For sites that do not report ZWD, it is estimated from ZWD = ZTD – ZHD(Psfc) where Psfc is taken from the background state. Error values are restricted to the range 4−30 mm. If the ZTD formal error is greater than 4 mm, then it is used as the lower limit of the range. With this ZTD error model, the ratio of observation error to model error tends to remain relatively constant over a wide range of PW, as model ZTD (humidity) error also increases with increasing model PW.

Other options for ZTD error specification exist such as:

- constant error
- error based on the formal ZTD error in the observation file

The error options are set in the NAMGPSGB section of the EnVar namelist file. See Appendix C for details.

Spatial and temporal observation error correlations are not considered using the current spatial thinning parameter (50 km) and temporal thinning parameter (2 hours). However, studies suggest that error correlations may not be negligible with these thinning parameters and should certainly be considered if the parameters are reduced.

BIAS CORRECTION

Site ZTD O−P biases are significant in some cases but the magnitudes of the biases are generally less than 20 mm. Biases over 12 mm are considered "too high" for blacklisting purposes. The cause of ZTD biases is not known. They may be related to the ZTD observation O (GPS data processing strategy or site location issues such as signal multipath from nearby obstructions), the trial P (incorrect GPS antenna height or biases in the trial pressure and/or humidity fields), or a combination of both. Dynamic bias correction of ZTD observations by GPS site was implemented as part of Innovation Cycle 3 in the fall of 2021. The bias correction system is illustrated in the second flowchart of Section 2.

At each analysis time (00, 06, 12, 18 UTC) a MIDAS 3D-Var analysis using conventional data only (minus AI and GP families) is done to obtain an "unbiased" estimate of the 3D atmospheric state for the bias correction of AI and GP data. The (uncorrected) GPS ZTD observations (O) are compared with the analysis ZTD (A) to get the ZTD O−A at each GPS site, which are then appended to a master GB-GPS O−A SQLite database after 2-hour temporal thinning is applied. The database contains ZTD O−A data for the last 45 days. The bias at each site is taken as the Mean O−A over the 45 day (max) period and is written to a GB-GPS (GP) bias correction text file. At the next analysis time 6 hours later, corrections are read from this file and applied to the GP ZTD data as part of the MIDAS ObsSelection program. Corrections for a site are set to a "missing value" (−999.00) in the text file under certain conditions, in which case bias correction is not applied to observations from that site. Updating the O−A database and generating the bias correction file are done by shell and Python scripts within the assimilation cycle suites. See Appendix B for more details.

BACKGROUND CHECK

The conventional (non-satellite-radiance) data "background check" consists of

- an **O−P check** to identify and remove observations with large (gross) errors, and
- topography filtering (**TOPOFILT**).

The background check is done as part of the MIDAS ObsSelection program. GB-GPS data are included in the "conventional data" background check (**BgckConv** in MAESTRO suite) along with other surface data, upper air data, GPS-RO data, aircraft data, etc..

In the ZTD **O-P check**, the variable

```
ZBGCHK = (ZOMP)**2/(StdOMP(ZWD))**2
```

is used to determine rejection, where ZOMP = O−P and StdOMP is an estimate of the standard deviation (Std) of the O−P as a function of ZWD. Statistical analysis of ZTD O−P data over summer and winter seasons show that the Std O−P increases linearly with the mean ZWD. This is due to the increase in both the observation and background ZTD errors with ZWD(PW). The statistical relationship is used to obtain StdOMP as a function of the observed ZWD for the background check. If ZBGCHK > 16, the ZTD observation is flagged for rejection. Data QC flag bits 9 and 16 are set for all rejected data.

In the **TOPOFILT**, observations are rejected where differences between station height and model surface height (topography) exceed specified limits. The current height difference limits are 1000 m for ZTD, 800 m for Psfc and Tsfc, and 50 m for DPDsfc. Data QC flag bit 18 is set for the rejected data. Currently, TOPOFILT is applied during the initial stages of the *EnVar* analysis.

The trial values (P) for the GB-GPS surface met data O−P check are generated by *EnVar* using the same conventional surface data observation operators used for SYNO and other SF family data (basically using simple extrapolation/interpolation of model thermodynamic fields). For GB-GPS ZTD observations, a special ZTD observation operator in the code is used to generate trial ZTD (P) at the GB-GPS site locations for ZTD O−P determination. The ZTD observation operator is part of the MIDAS Fortran90 source code repository. See Appendices A and C for details.

THINNING1

Here thinning1 refers to spatial and temporal thinning of the GB-GPS data. The thinning is done as part of the MIDAS ObsSelection program. A special "fast" brute thinning algorithm is applied that ensures all assimilated observations are separated by a minimum distance of 50 km and minimum time difference of 2 hours. The algorithm is currently configured such that the majority of observations are at (or close to) the times T-2h, T, and T+2h where T is the "central" analysis time (i.e., 00, 06, 12, or 18 UTC). The thinning parameters are specified in the suite's *bgckalt.cfg* file as:

```
BGCKALT_gp_thinning_delta=50
BGCKALT_gp_thinning_temporal=2H_v2
```

For GPS sites processed by multiple analysis centres (ACs), there will be multiple ZTD observations to choose from for the same location (site) and time. In this case, the site-AC combination with the best ZTD data quality score will be preferentially selected by the thinning algorithm. A text file containing ZTD data quality scores (**ZDScore**) for each GPS station is generated from CMDA monthly data monitoring statistics. The formula for computing station **ZDScore**s is

```
ZDScore = 100 * ( 0.15*((MAXreps-Nreps)/MAXreps)) +
                  0.35*((Std(O-P)-5.0)/25) +
                  0.25*(Abs(Mean(O-P))/12) +
                  0.10*((Nout/Nreps)/0.05) +
                  0.15*(Mean(ZTDferr)/norm) )
```

MAXreps = the maximum number of reports from any station for the last monitoring month.
Nreps = the number of reports for the given station for the last monitoring month
Nout = the number of "bad" or outlier data as determined from Abs(O−P) values
norm = normalizing factor for mean ZTDferr; equals 10 mm for UCAR data and 3 mm for E-GVAP data..

The **ZDScore** is largely determined by the ZTD Std(O−P) and Mean(O−P) for the station for the last monitoring month. Note that lower score values are better. If there are not enough reports for a station to determine data quality (Nreps < 70), the value of ZDScore is set high to 9999. The site ZDScores for each station in the monitoring text file are added to the GP family derivate file observations by adding a new BURP element (element 40026) in the data blocks (in the THINNING0 program).

ANALYSIS (EnVar)

In the EnVar code, GB-GPS observations (ZTD and surface met data) comprise the 'GP' family. First, all GP family data flagged for rejection by the data processing and QC steps described above are filtered out. Tangent linear and adjoints of the non-linear observation operators described in the BACKGROUND CHECK section above are used during the analysis to minimize the cost function. In addition, Variational Quality Control (QCVar) is activated during the analysis procedure to reduce the weight of observations that do not fit the current "analysis" state. Observations are considered "rejected" when their weight is reduced from 1.0 to 0.25. With the current set of ZTD QCVar parameters, ZTD observations are rejected when $|O-A| > 3.42*\sigma_o$ where σ_o is the ZTD observation error and A is ZTD of the current analysis state. Note that, for GPS-RO and GB-GPS ZTD, there are only non-linear observation operators which are contained in special GPS Fortran modules. Fortran90 GPS structures/types, operator overloading and automatic differentiation are applied to obtain model background ZTD and Jacobians $\partial ZTD/\partial x$ (x = control vector [P0, TT, LQ]) at the same time. For efficiency, the Jacobians are computed once at the start of the minimization rather than re-computed at each iteration.

Details of the EnVar assimilation and the GPS ZTD observation operator are found in Appendix A and Appendix C.

5. APPENDIX A: ZTD Observation Operator Details

A description of the GB-GPS ZTD observation operator as applied in the MIDAS code is given in this appendix. The actual MIDAS modules and subroutines are described in Appendix C.

GPS ZTD can be expressed as a function of atmospheric refractivity for microwaves (*N*) as follows:

$$\frac{\partial ZTD}{\partial z} = -10^{-6} N \qquad \textbf{Equation 1}$$

$$ZTD = \int_{zsite}^{ztoa} 10^{-6} N \, dz$$

where *zsite* = height of GPS receiver antenna and *ztoa* = height at "top of atmosphere"

Refractivity *N* can be expressed as a function of atmospheric *P*, *T*, *q*

$$\boxed{N = \frac{k_1 P}{T_v} + (k_2 - \varepsilon k_1)\frac{Pq}{\varepsilon T} + k_3 \frac{Pq}{\varepsilon T^2}} \qquad \textbf{Equation 2}$$

P = pressure (Pa)
T = temperature (K)
T_v = virtual temperature (K) = $T(1 + \kappa q)$
q = specific humidity (kg/kg)

$$\varepsilon = \frac{R_d}{R_v} = 0.621948$$

$$\kappa = \frac{1}{\varepsilon} - 1 = 0.60777$$

and the refractivity *k* constants (Bevis, 1994) are

$k_1 = 0.776$ K/Pa, $k_2 = 0.704$ K/Pa, $k_3 = 3.739 \times 10^3$ K^2/Pa

Using the hydrostatic equation, Equation 1 can be expressed in terms of a pressure vertical coordinate *P*:

$$\boxed{\frac{\partial ZTD}{\partial P} = 10^{-6} N^* \frac{R_d T_v}{g(\varphi, z)}} \qquad \textbf{Equation 3}$$

where *g* is gravity as a function of latitude φ and height z and

$$\boxed{N^* = \frac{N}{P} = \frac{k_1}{T_v} + k_2' \frac{q}{\varepsilon T} + k_3 \frac{q}{\varepsilon T^2}} \qquad \textbf{Equation 4}$$

where $k'_2 = k_2 - \varepsilon k_1 = 0.221$ K/Pa.

The change in ZTD over a discrete layer bounded by 2 pressure levels is then:

$$\Delta ZTD = \left(\frac{\partial ZTD}{\partial P} \right)_{AVG} \Delta P$$ **Equation 5**

where AVG = average over the ΔP layer.

In EnVar, model profiles of control variables P(k), T(k), and q(k) at each observation location are computed from horizontal interpolation of the 2D fields PP, TT and HU respectively. The GB-GPS ZTD operator computes $\frac{\partial ZTD}{\partial P}$ at each model thermodynamic level k using the model T(k), q(k) profiles and Equations 3 and 4. The operator then builds model ZTD(k) profiles from the model top (k=1) down to the model surface (k=*nlev*) using the stored $\frac{\partial ZTD}{\partial P}$(k) profiles and Equation 5 as follows:

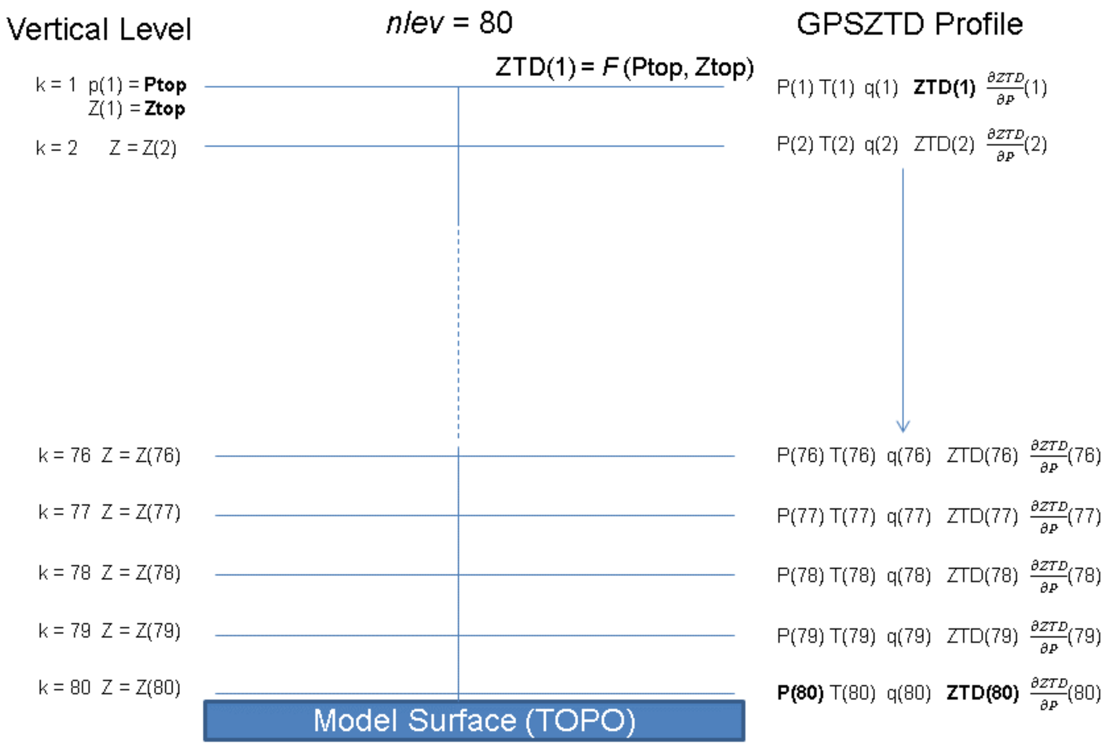
$$ZTD(1) = ZTD_{TOP} = 10^{-6} \frac{k_1 R_d P_{TOP}}{g_{TOP}(\varphi, z_{TOP})}$$ where $g_{TOP} \approx 9.6$ m/s^2 at $z_{TOP} = 65$ km ($P_{TOP} = 0.1$ hPa)

$$ZTD(k) = ZTD(k-1) + \left(\frac{\partial ZTD}{\partial P} \right)_{AVG} \left(P(k) - P(k-1) \right), \quad k = 2, nlev$$ **Equation 6**

where

$$\left(\frac{\partial ZTD}{\partial P} \right)_{AVG} = \frac{1}{2} \left(\left(\frac{\partial ZTD}{\partial P} \right)_k + \left(\frac{\partial ZTD}{\partial P} \right)_{k-1} \right)$$

Figure A1 illustrates how model profiles of thermodynamic variables are used to compute ZTD on each model level.



$$ZTD(k) = ZTD(k-1) + \frac{1}{2}\left(\frac{\partial ZTD}{\partial P}(k) + \frac{\partial ZTD}{\partial P}(k-1)\right)(P(k) - P(k-1))$$

Figure A1: Construction of a GPS ZTD profile on 80 model levels

The final step is to compute model *ZTD* at the actual GPS observation (antenna) height using the model ZTD(k) profile. There is generally a difference between the model lowest level height (*Z(nlev)*) and the observation (GPS antenna) height (*Zgps*) due to the use of a smoothed interpolated grid-point topography field in the model. Also, the GPS antenna is generally mounted at some fixed height above the ground. The height difference can range from a few metres or less to well over 1000 metres in mountainous or complex terrain. The method of computing model ZTD at the observation height depends on the sign and magnitude of the height difference $\Delta z = Zgps - Z(nlev)$.

- When abs (Δz) > **1000 m** [1], do not assimilate ZTD
- When Δz = 0 m,
  - *ZTD* = ZTD(*nlev*) (profile lowest level ZTD "as-is")
- For Δz > 0 (GPS antenna **above** lowest model level (surface))
  - Do linear-log interpolation in height of ZTD(k) between the model levels above and below the observation height *Zgps*.
- For Δz < 0 (GPS antenna **below** lowest model level (surface))
  - if abs(Δz) < **2 m** [2],
    - *ZTD* = ZTD(*nlev*) (profile lowest level ZTD "as-is")
  - if **2 m** < abs(Δz) < **100 m** [3],

- $ZTD = \mathrm{ZTD}(nlev) + \left(\frac{\partial ZTD}{\partial P}\right)_{NLEV} \big(P(Zgps) - P(nlev)\big)$

- if abs(Δz) > **100 m**
  - $ZTD = \mathrm{ZTD}(nlev) + \left(\frac{\partial ZTD}{\partial P}\right)_{AVG} \big(P(Zgps) - P(nlev)\big)$
- $(\partial ZTD/\partial P)_{AVG}$ is Δz layer average of lowest level (*nlev*) value and value at *Zgps*. *P*, *T*, *q* at *Zgps* (for *N(Zgps)*) are determined from lowest level (*nlev*) data. A fixed lapse rate is assumed for *T*, *Tv* ($\gamma = -6.5$K/100 m) and *q(Zgps)* is fixed at the lowest level value *q(nlev)*. *P(Zgps)* is then determined from *P(nlev)* using **CMC hydrostatic method**. Then *N**(*Zgps*) = *N* / *P* is computed from *N*(*Zgps*) and *P*(*Zgps*) to compute $(\partial ZTD/\partial P)_{\mathrm{Zgps}}$ using Equation 3.

  CMC hydrostatic method to get pressure at observation height:

  $$P(Zgps) = P(nlev)\left(\frac{T_v(Zgps)}{T_v(nlev)}\right)^{\frac{g_0}{R\gamma}} \qquad \textbf{Equation 7}$$

  where assumed T_v lapse rate $\gamma = 0.0065$ K/m

  g_0 = gravity at surface (*Zsfc*)

  $$T_v(nlev) = T(nlev)(1 + \kappa q(nlev))$$

  $$T_v(Zgps) = T_v(nlev) - \gamma \Delta z$$

  $$\Delta z = Zgps - Zsfc$$

[1] EnVar namelist parameter `DZMAX` (overrides value in EnVar code = 1000 m)
[2] EnVar namelist parameter `DZMIN`
[3] Hard-coded value dzmax in observation operator subroutine *gpsZTDopv*

The computation of model ZTD at observation height *Zgps* is illustrated in Figure A2.

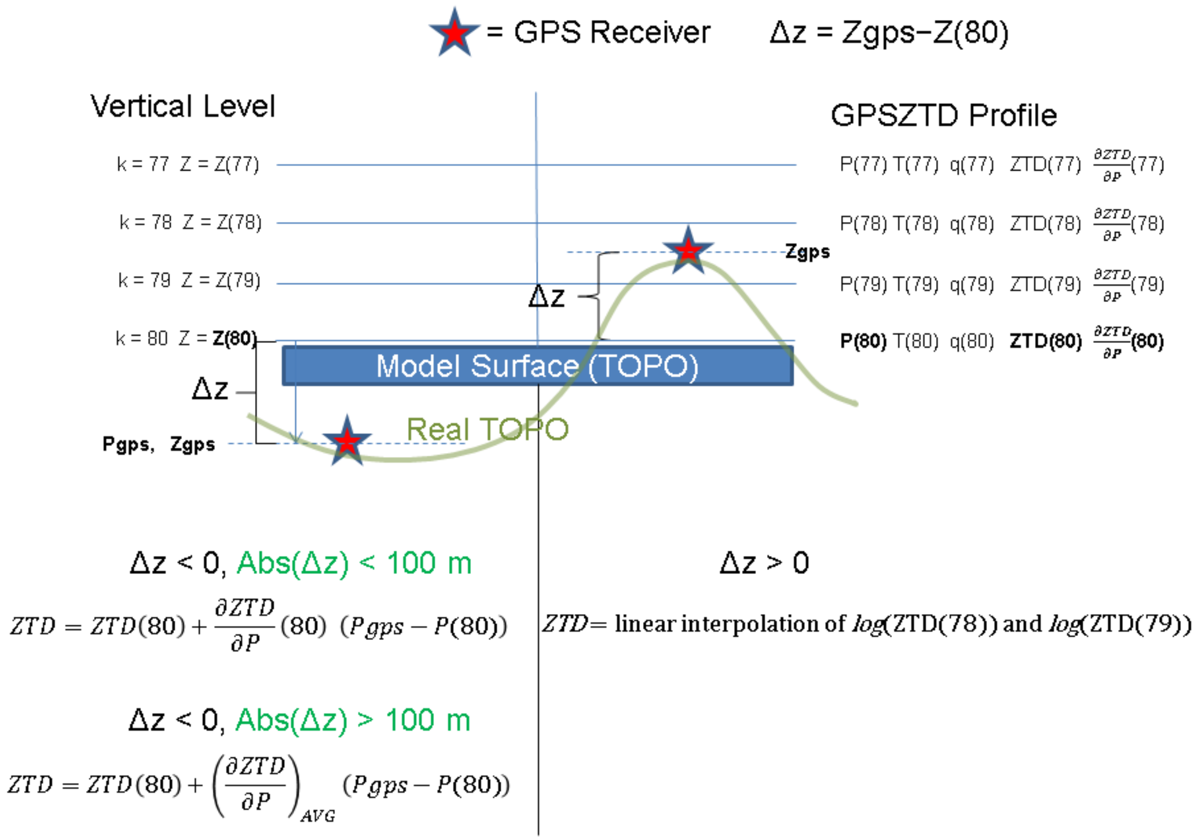


Figure A2: Computation of model ZTD at GPS observation height for 80-level GEM model (*nlev* = 80). *Pgps* is *P*(*Zgps*) in the text (pressure at GPS antenna height) and is computed hydrostatically from model lowest level pressure *P(80)*.

6. APPENDIX B: The dynamic ZTD bias corrections scheme

This appendix describes the files and scripts used to activate, configure and run the ZTD bias correction scheme that is integrated into the assimilation cycle MAESTRO suite (the **genbiascoeff** part). For details on the MIDAS Fortran code and namelist files used to apply the corrections as part of the background check (**bgckalt/BgckConv**) part of the suite, see Appendix C.

Activating the scheme

MAESTRO suite file:
config/main/assimcycle/bgckalt.cfg

```
BGCKALT_gp_bcor_mode=dynamic
```

File:
config/main/assimcycle/genbiascoeff.cfg

```
GENBIASCOEFF_conv=yes
```

Configuring the scheme (user options)

File:
config/main/assimcycle/genbiascoeff.cfg

```
GENBIASCOEFF_conv_obstyp="ai,gp"
GENBIASCOEFF_conv_ndays="365,45"
GENBIASCOEFF_conv_vars="TT,ZTD"
GENBIASCOEFF_conv_trim="yes,no"
GENBIASCOEFF_conv_cleandb="yes,no"
GENBIASCOEFF_conv_newdb="no,no"
GENBIASCOEFF_conv_builddb="no,no"
GENBIASCOEFF_conv_omacol="OMP"
GENBIASCOEFF_conv_namelist=${SEQ_EXP_HOME}/config/main/assimcy
cle/genbiascoeff/gen_bcor.json
GENBIASCOEFF_conv_thindt="0.0,2.0"
```

Values for GB-GPS are the second values in the strings (the first values are for AI data bias correction).

GENBIASCOEFF_conv_ndays = Max number of days of O−A data to keep in database (45)
GENBIASCOEFF_conv_vars = Observation to bias correct (ZTD)
GENBIASCOEFF_conv_trim = Remove O−A data from database using SEM [see NOTE below] (no)
GENBIASCOEFF_conv_cleandb = Vacuum the database file to minimize file size (no)

GENBIASCOEFF_conv_newdb = (R&D only) Create a new master O−A database (no)
GENBIASCOEFF_conv_builddb = (R&D only) Build the O−A database only. Don't generate a bias correction file (no)
GENBIASCOEFF_conv_omacol = Block in GP BURP file where MIDAS oMinusF program puts the "O−A" values and hence the column in the SQLite dbase table produced by converting the BURP file to SQLite dbase with *burp2rdb* (OMP = O−P block, BFAM=14, OMP column in dbase DATA table)
GENBIASCOEFF_conv_namelist = JSON file containing additional configuration parameters read by ***gen_bcor*** Python script that generates the bias correction text file
GENBIASCOEFF_conv_thindt = Temporal thinning applied to O−A data prior to appending to O−A database (2.0 hours). 0.0 = No thinning.

NOTE: The bias correction scheme can be configured to consider the Standard Error of the Mean O−A (*SEM*) when writing bias corrections to the bias correction file and for trimming the master O−A database. The *SEM* is defined as

$$SEM = \frac{Std(O - A)}{\sqrt{Nobs}}$$

If the *SEM* is too high (over a prescribed maximum value), the mean is considered unreliable and so the bias correction will be output as "missing". The *SEM* can also be used to "trim" the database by removing unneeded O−A data for a particular station. A minimum desired "optimal" *SEM* can be prescribed. If the number of observations (*Nobs*) exceeds the number needed to obtain the minimum *SEM* value, then the excess oldest data can be removed. This is useful to minimize the size of very large O−A database files, such as the one for AI data, but is not needed for GB-GPS data so
GENBIASCOEFF_conv_trim = no.

File:
config/main/assimcycle/genbiascoeff/gen_bcor.json

Note that this file contains config parameters for both AI TT and GP ZTD data bias correction. Here we show only the GP config section "config_gp".

```
"config_gp": {
   "min_num_obs": 10,
   "ndmax": 9999,
   "sem_max": 9000.0,
   "sem_opt": 0.001,
   "stdoma_max": 10.0,
   "str_codetypes": "(189)",
   "AddCondNdays_opt": true,
   "numdays_min": 7,
   "percent_obs_min": 42.0,
   "nite_only": false,
   "missing": -999.0,
   "table_vars": ["ZTD", "P0"],
```

```
  "oma_col": "OMP"
}
```

min_num_obs = minimum number of observations needed to compute mean/std O−A stats. If Nobs < min_num_obs, bias correction will be missing.

ndmax = print warning when number of days in O−A database exceeds this value

sem_max (mm) = output Mean O−A as "missing" in bias correction file if SEM > sem_max; here 9000 means don't consider the SEM (output correction regardless of SEM)

sem_opt (mm) = if GENBIASCOEFF_conv_trim = "yes" in config/main/assimcycle/genbiascoeff.cfg file, then, if SEM < sem_opt, identify "excess" observations that can be removed from database file to maintain SEM at sem_opt value. Excess records exist if the actual SEM is less than the optimal minimum SEM value sem_opt. Here we set the value to very low (0.001 mm) so no excess observations will ever be flagged for removal, although GENBIASCOEFF_conv_trim = "no" also ensures that this will not happen.

stdoma_max (mm) = (if GENBIASCOEFF_conv_trim = "yes" in config/main/assimcycle/genbiascoeff.cfg file) If stdoma_max > 0, then use this value of Std(O−A) to determine Nobs needed to maintain SEM = sem_opt using SEM equation above. If stdoma_max = 0, then use actual Std(O−A). The Nobs needed to maintain SEM = sem_opt is

$$Nobs = \frac{(Std(O-A))^2}{sem\_opt^2}$$

str_codetypes = code types for O−A reports (not used for GP data)

AddCondNdays_opt = If "true", then the following conditions must also be met to output a non-missing bias correction for a station to the bias correction file:

- The observations must span a period of at least {**numdays_min**} days in the O−A database
- There must be at least {**percent_obs_min**} % of expected observations in the observation time span

The number of expected observations per day is 12 (when 2-hour temporal thinning is applied). In this case, the condition is:

"Observations must span at least 7 days with 42% of expected observations in that period"

If the condition is not met, then the correction for the GPS station will be "missing value".

nite_only = If "true", only consider nighttime O−A data when computing Mean O−A stats (using lat/lon and date to get solar elevation angle and time-of-day). This option exists because the supposedly unbiased analysis A is obtained from assimilation of radiosonde observations and other conventional data, and radiosonde humidity observations for some sonde types (such as LMS type used in USA) have a pronounced daytime bias that could be reflected in the analysis A, which in turn will negatively affect the ZTD bias estimation.

missing = The value output to bias correction file to indicate a "missing" value.

table_vars = The variables found in the O−A database file. Surface pressure (P0) is included in the GP database in case bias correction is ever applied to Psfc as well as ZTD.

oma_col = Read the O−A values from this column (OMA or OMP) of the O−A database DATA table. Note that the OMA and OMP columns should both contain the “O−A” values in the operational database files so the value of oma_col should not matter.

Files output by the bias correction scheme

The following two files are output by the ***gen_bcor*** Python script. They are found in the assimilation cycle suite’s **`.../hub/…/dynbcor`** directory. The **{date}_gp_mean.txt** file is used by MIDAS to apply the bias corrections. The **{date}_gp_stats.txt** file is for information only and is optional. The master O−A SQLite database also resides in the suite’s **`.../hub/…/dynbcor`** directory.

{date}_gp_stats.txt: Contains the ZTD O−A statistics for each GPS station.

Header (columns):

STN_ID	= GPS station name
Nobs	= Total number of O-A observations in database for this station
MeanOMA	= Mean O-A (mm)
StdOMA	= StdDev O-A (mm)
SEM	= Standard Error of Mean O-A (mm)
NDays	= Number of days with reports (range); can be data gaps within this range
PrcntD	= Percentage of expected observations within the NDays range
Cond	= 0 or 1: 1 means that, if AddCondNdays_opt = “true”, the optional added condition was met so a non-missing correction should have been output to the bias correction file for this station. If 0, the condition was not met and the correction will be missing in the bias correction file.

{date}_gp_mean.txt: The bias correction file containing the ZTD Mean O−A (mm) for each GPS station.

This is the file read by the MIDAS ObsSelection program for application of the ZTD bias corrections. The Mean O-A values are subtracted from the raw ZTD observations to get the corrected values that are assimilated. However, the negative of the Mean O-A value is written to output BURP file element 15234 as the “bias correction” because the convention in this case is to add the correction to the raw value.

Bias correction scripts (in assimilation MAESTRO suite)

(a) …/modules/genbiascoeff/Conv.tsk

This is the main shell script (task) that handles the appending of the O−A database with the latest O−A data, calls ***gen_bcor*** Python script to generate the bias correction files, removes old data (e.g., older than 45 days) from the database and, optionally, removes records from the database based on SEM by calling ***trim_dbase*** Python script, and cleans (vacuums) the database. It is a generalized script used for both AI and GP data bias correction.

(b) …/modules/genbiascoeff/scripts/genbc.gp_temp_thin.py

This Python script is used to apply (optional) temporal thinning to GB-GPS O−A data before appending the data to the O−A database.

(c) …/modules/genbiascoeff/scripts/genbc.gen_bcor.py

This Python script generates the O−A statistics and bias correction files for AI or GP data (section 2 above).

(d) …/modules/genbiascoeff/scripts/genbc.trim_dbase.py

This Python script trims the AI or GP O−A database by removing the oldest excess records as identified by the ***gen_bcor*** script (excess records exist if the actual SEM is less than the optimal minimum SEM value). This is currently not done for GP observations (GENBIASCOEFF_conv_trim = no)

7. APPENDIX C: GB-GPS related MIDAS files and routines

The MIDAS namelist file: `VAR.nml` (`&NAMGPSGB` Section)

Various options can be set here:

```
!*    Namelist variables for Ground-based GPS (ZTD)
!
!     DZMIN:       Minimum DZ = Zobs-Zmod (m) for which DZ adjustment to ZTD
!                  will be made.
!     YSFERRWGT:   Weighting factor multiplier for GPS surface met errors (to
!                  account for time series observations with error correlations)
!     DZMAX:       Maximum DZ (m) over which the ZTD data are rejected
!                  due to topography (used in SOBSSFC when LTOPOFILT = .TRUE.)
!     YZTDERR:     If < 0 then read ZTD errors from data blocks in input
!                  files (i.e. the formal errors).
!                  If > 0 then use value as a constant error (m) for all ZTD
!                  observations.
```

```
!                   If = 0 then compute error as a function of ZWD.
!      LASSMET:     Flag to assimilate GPS Met surface P, T, T-Td
!      LLBLMET:     Flag to indicate that surface met data have been blacklisted
!                   for GPS sites close to surface weather stations.
!      YZDERRWGT:   Weighting factor multiplier for GPS ZTD errors (to account
!                   for time series observations with error correlations)
!      LBEVIS:      .true.  = use Bevis(1994)  refractivity (k1,k2,k3) constants
!                   .false. = use Rueger(2002) refractivity (k1,k2,k3) constants
!      IREFOPT:     1 = conventional expression for refractivity N using k1,k2,k3
!                   2 = Aparicio & Laroche refractivity N (incl. compressibility)
!      L1OBS        Flag to select a single ZTD observation using criteria in
!                   subroutine oonl_gpsgb
!      LTESTOP      Flag to test ZTD observation operator (O-P, Bgck modes only)
!      IZTDOP       1 = normal mode: use stored ZTD profiles to get ZTDmod
!                   2 = Vedel & Huang ZTD formulation: ZTDmod = ZHD(Pobs) + ZWD
```

Current values of `&NAMGPSGB variables:`

```
&NAMGPSGB
   DZMIN        = 2.0D0,
   DZMAX        = 1000.0D0,
   YZTDERR      = 0.0D0,
   LASSMET      = .TRUE.,
   LLBLMET      = .TRUE.,
   YSFERRWGT    = 1.00D0,
   YZDERRWGT    = 1.00D0,
   LBEVIS       = .TRUE.,
   IREFOPT      = 1,
   L1OBS        = .FALSE.,
   LTESTOP      = .FALSE.,
   IZTDOP       = 1,
```

NOTES:

- Selecting IZTDOP=2 (Vedel & Huang ZTD formulation) causes the EnVar GPS ZTD observation operator to compute ZTD at observation height in a different manner. The dry delay (ZHD) component of ZTD is computed separately using model lowest level pressure (P(nlev)) adjusted hydrostatically to the GPS observation height Zgps with Equation 7. The wet delay part (ZWD) is computed by integration of model q/T from model top to model lowest level Z(nlev). The model ZWD is then adjusted to GPS height Zgps by computing a mean Δz layer "wet refractivity" N_w from extrapolated lowest level values of P, T and q and multiplying N_w by Δz=Zgps-Z(nlev) (from Equation 1). Then the ZTD = ZHD + ZWD.
- Currently GB-GPS serial observation error correlations are not taken into account and each GPS observation in the 6h assimilation window is given full (equal) weight (YZDERRWGT=1, YSFERRWGT=1). The *ERRWGT values can be set > 1 to reduce the weight given to the observations to compensate for correlated-in-time observation errors.

Namelist used by ObsSelection program (file `nml.conv_bgck`):

```
&namObsSelection
   doThinning = .true.
/
&thin_gbgps
   deltemps = 8       thinning Δt = 8 15-minute time steps (2 hours)
   deldist  = 50      thinning Δx = 50 km
/
&NAMGPSGB
{…as above in VAR.nml file}
/
&NAMBIASCONV
  gpBiasActive=.true.,
  gpRevOnly=.false.
```

`gpBiasActive` = Apply (or reverse only) GP (ZTD) bias corrections
`gpRevOnly` = Reverse existing bias corrections only (do not apply new corrections)

GB-GPS related routines and functions in MIDAS Fortran source code

The MIDAS data assimilation source code and scripts can be found here:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas
Specifically, the latest master branch MIDAS Fortran source code (modules) is found here:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/tree/master/src/modules

NOTE: This documentation refers to the latest version of MIDAS as found in the master branch as of May 2021 which has not been used operationally. The operational version of MIDAS for Innovation Cycle 3 is taken from branch "v_3.6".

(a) Apply bias corrections for GB-GPS observations:

Module:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/blob/master/src/modules/biascorrectionConv_mod.f90
Subroutine:
This module contains various subroutines for bias correction AI and GP observations including routines to:

- read the &NAMBIASCONV part of the namelist file
- read the bias correction text file
- apply the bias corrections to the observations in obsSpaceData

Note that GB-GPS derialt BURP files do not have the data block BURP element to store the value of the bias correction applied to the ZTD observation (elem 15234). This is added to the input derialt BURP file by routine **`brpr_addElementsToBurp`** found in module

https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/blob/master/src/modules/burpread_mod.f90
The bias correction applied to ZTD will be stored in this element in the output evalalt/bgckalt files.

(b) Set the observation errors for GB-GPS observations:

Module:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/blob/master/src/modules/obserrors_mod.f90
Subroutine:
oer_SETERRGPSGB

- This routine sets the observation errors for GB-GPS ZTD observations using the selected error model (see section 4b OBSERVATION ERRORS). Also set are the StdOMP(ZWD) for all GPS-ZTD observations needed for the O−P background check (see section 4b BACKGROUND CHECK section).

(c) Read and initialize all GB-GPS namelist variables:

Module:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/blob/master/src/modules/gps_mod.f90
Subroutine:
GPS_SETUPGB

(d) The GB-GPS ZTD observation operator and related routines:

Module:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/blob/master/src/modules/gps_mod.f90

Subroutines:

gps_structztd_v2

- fills GPS profiles of type "gpsprofilezd" (for the ZTD and PW operators)
- GPS profiles (structures) include P(k), z(k), Psfc (P0), T(k), q(k) plus ZTD(k) and $\frac{\partial ZTD}{\partial P}$(k) at each observation location where z(k) is the altitude (AL) profile.

gps_ztdopv

- **ZTD non-linear observation operator** which returns model ZTD at observation location/height.
- uses GPS profiles at observation locations created with *gps_structztd_v2*
- uses operator overloading and automatic differentiation to provide both the value of ZTD and the derivative of ZTD with respect to the analysis control vector (the Jacobian)

gps_pw

- returns model PW at observation location/height
- uses GPS profiles at observation locations created with *gps_structztd_v2*
- used to compute background PW for the filtering applied in subroutine *oop_gpsgb_nl* (see below).

Module:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/blob/master/src/modules/obsoperators_mod.f90

Subroutines:

oop_gpsgb_nl

- Applies the non-linear ZTD observation operator (*gps_ztdopv*) to get the trial ZTD (P) for the O−P calculations.
- The subroutine reads all GB-GPS (GP family) ZTD observations in memory and computes the initial value of Jo(GPS-ZTD) from the O−P's
- Also filters out ZTD data for assimilation when the following conditions are met:
  i. background (trial) PW is very low (dry) (trial PW < 2 mm)
  ii. surface pressure observation is missing or out-of-range (bad), if namelist option `LASSMET`=true[1]
  iii. if Abs(P(Zgps)−Pobs) > 2 hPa where Pobs = GPS Psfc observation, P(Zgps) = model pressure at observation height[1].

[1] Conditions (ii,iii) are applied only to ZTD observations from the NOAA (FSL) GPS network (by checking for the string "FSL_" or "NOAA" in the GPS site names). At the time the routine was written, the only GP family observations with Psfc reports were from the NOAA/FSL network. When the replacement UCAR network observations (which also contain Psfc) became operational, the filter code in the operator was not modified such that conditions ii and iii are no longer applied to filter ZTD observations. This will be fixed in the future such that the filtering using these conditions will be applied to the UCAR network data.

The filters (i), (ii) and (iii) listed above are used to reject ZTD assimilation locally for certain cases that could be problematic. For example, background check programs

> **Filter (i)**: When the background is very dry, ZTD innovations (O−P) can lead to significant and undesirable increments in analysis surface pressure P0. Examples are over the arctic regions in winter.
>
> **Filter (ii)**: Psfc observations (Pobs) help constrain the analysis increments from ZTD assimilation. Also, cannot apply Filter (iii) if Pobs is missing.
>
> **Filter (iii):** When surface pressure O−P is too large, it could indicate a large error in the forecast pressure, and the corresponding large ZTD innovations (O−P) will produce undesired increments in analysis humidity HU rather than P0 (unless the background is very dry, but in Filter (i) will reject ZTD in this case). An example that could produce such

conditions would be a land-falling hurricane or a poorly forecast intense low pressure system.

oop_calcGPSGBJacobian

- computes the ZTD Jacobian (dZTD/dx) at each observation location for *oop_Hgp*, *oop_HTgp* , where x is the analysis control vector of temperature (TT), humidity (HU), height (GZ) and surface pressure (P0) (using the trial profile values).
- calls routine *gps_ztdopv* (ZTD observation operator) from module *gps_mod.f90* to get the Jacobians
- Jacobians are only computed at the first iteration of the minimization and stored for use in the remaining iterations.

oop_Hgp

- applies the tangent linear (TL) of the ZTD observation operator at all ZTD observation locations using the ZTD Jacobian and analysis increments in temperature (TT), humidity (HU), height (GZ) and surface pressure (P0).

oop_HTgp

- applies the adjoint (AD) of the ZTD observation operator.

(e) Other MIDAS modules containing important GB-GPS related code:

Module:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/blob/master/src/modules/varqc_mod.f90

Subroutine:
vqc_setup

- sets parameters that determine QCVar rejection for GPS ZTD observations

Module:
https://gitlab.science.gc.ca/atmospheric-data-assimilation/midas/blob/master/src/modules/backgroundCheck_mod.f90

Function ***isetflag***:

- contains the criteria zzdcrit() for background check O−P based rejection of GB-GPS ZTD observations

Subroutine ***bgck_data***:

- performs the background (O−P) check for conventional observations including GP family (surface met and ZTD)