

DRAFT

E-GVAP update of file naming and COST-format

A new naming scheme for the COST files with ZTD data which the analysis centres (ACs) upload to E-GVAP is introduced. The change of naming serves three purposes:

- **Enabling sub-hourly uploads of COST files**
- **Enabling selection of GTS distribution channel via the file naming**
- **Enabling continued use of the current naming scheme, for ACs not doing any changes to their processing.**

In connection with the new naming scheme updates have been made also to the COST-Format definition, the new one being version 2.2.

ACs need not change to the new format for their current deliveries.

ACs making changes to their processing, setting up test processing, or changing to the new naming scheme, must change also to the new COST-Format.

The main purpose of the changes in the COST files header information is to enable cross checking that the file naming used is intentional, and to enable inclusion of the DOMES id. The latter is expected to become very useful when GNSS meteorology gradually becomes truly global, in which case we will see more appearances of different GNSS sites with the same four letter identifier.

Background

Since the days of COST 716, MAGIC, TOUGH, etc., we have in European ground-based GNSS meteorology been doing the ZTD estimation in hourly batches, and correspondingly uploaded the ZTDs hourly.

However, Swisstopo has for a long time been producing also a real-time ZTD product (LPTR) for the AGNES sites, but uploading the ZTDs hourly. At KNMI sub-hourly processing of mainly Dutch sites has been running for a while (15 min batches), and the ZTDs are assimilated at KNMI into an NWP model running an hourly assimilation cycle. Also at UK Met Office sub-hourly processing has started. In NWP, the high resolution models are in general moving toward more frequent data assimilation and shorter cutoff times, resulting in a demand for faster access to observations.

It is clear that the large variation in access speed and access rights to the raw GNSS data used by E-GVAP ACs to estimate ZTDs prevents a system-wide move to sub-hourly processing at this stage.

On the other hand it will be very beneficial to some of the current NWP to distribute the ZTDs which are derived sub-hourly, without the artificial aging due to the old upload scheme.

The COST format itself enables inclusion of delay estimates for periods of any length. It is only the file naming convention that currently prevents multiple uploads per hour. For that reason it has been decided to introduce a **new file naming convention** to enable sub-hourly uploads to E-GVAP.

We distinguish in E-GVAP between two products "operational" and "test". Operational is uploaded to directory "nrt" on the upload server, while test goes into "test/nrt". The users can obtain the data in two ways. Via password restricted ftp, and via the meteo specific GTS network. On the download ftp-server the above directory structure is mirrored. Only the operational data, from "nrt" goes onto the GTS. Prior to being put on the GTS the COST files are converted to the meteo specific format BUFR.

What is not well known is that actually two GTS distributions are available to us (recognizable by their headers), this has enabled us to provide users that can only handle BUFR data with E-GVAP data prior to them being distributed as fully "operational", as an in between "demonstration" product. So far it is mainly Meteo France that has been taking advantage of that option. With the expansion in numbers of ACs and users, we foresee more widespread usage of this option, and therefore want to move from the current completely manual setup of the GTS channel selection, to an automated setup, in which the filename determines the handling and routing of the data.

Naming conventions

Current file naming, for hourly uploads of data from within an hourly period

`cost_yyyymmddhh_xxxx_pppp.gz`

where "cost" identifies it as a cost file, "yyymmddhh" are year, month, day and hour, "xxxx" means 'from multiple sites' (as opposed to a site name, e.g. "onsa"), "pppp" is AC&solution type identifier, e.g. "gop1" for the processing type 1 from GOP. ".gz" is due to the gzip compression of the file.

As mentioned, the COST-format description itself has been updated. New products for upload to E-GVAP should be encoded according to the new COST-format description. See the document "COST-Format File Specification for Ground-Based GNSS Delay and Water Vapour Data" version 2.2, which is available at the download ftp-server, as well as at the E-GVAP homepage: http://egvap.dmi.dk/support/formats/cost_format_v22.pdf

New file naming

`cost_b_s_yyyymmddhhmm_yyyymmddhhmm_cccc_pppp.gz`

where the first "yyymmddhhmm" is the valid time of the oldest ZTD in the file to the precision of one minute, and the next "yyymmddhhmm" is the valid time of the youngest ZTD in the file.

"cccc" is substituted with either "mult" (alternative "xxxx"), meaning ZTDs from multiple GNSS sites; or a four letter GNSS stations identifier, in case the ZTDs in the file are all from the same site.

"pppp" is substituted with the AC and processing method identifier, e.g., "gopg" for GOPs global processing. Note: New ACs must agree with the E-GVAP team upon a new AC identifier, in order not to conflict with current AC identifiers.

There are 6 alternatives for "b", indicating the time period and/or special processing:

1. "r" for real real-time processing
2. "s" for sub-hourly
3. "h" for hourly
4. "c" for combined solution (based on ZTDs from different ACs).
5. "p" for post processing
6. "u" for anything else or combinations of multiple types of the above for ZTDs in one file, or unknown.

There are 4 alternatives for "s", indicating the file status:

1. "o" for operational. COST-Format files uploaded to "nrt" in files named with an "o" will be sent in BUFR format to users via the "operational" GTS channel.
2. "d" for demo. COST-Format files uploaded to "nrt" and named with a "d" will be sent in BUFR format to users via the "non-operational" GTS channel.
3. "t" for test. COST-format files with status flag "t" will not be sent on the GTS

4. "u" for anything else than the above, or unknown.

Notice that all files named according to the new naming scheme, shall be uploaded to directory "nrt" on the upload ftp-server, also those containing "test" data.

Time tacking of "overlapping" ZTDs at end and beginning of consecutive batch intervals

For the hourly batch processing and uploads of data from it, we have the rule that *if* an AC is producing ZTDs valid at both the very ends of the hourly batch interval, then in order to update the youngest, least precise, ZTD with the updated value from the next hourly processing, the time tack of the first (youngest) ZTD is adjusted from hh:00 to hh-1:59 (e.g. 15:00 becomes 14:59) when written to the cost file.

This procedure overcomes the problem that many meteorological databases cannot accept multiple examples of "the same" observation. By having the latest, and best, ZTD estimate valid at the full hour, which is a preferred time by current data assimilation selection algorithms, the system will choose the latest of the two ZTDs when both are present in the database.

For ACs doing sub-hourly processing producing and uploading ZTDs valid at both ends of the sub-hourly batch interval, a similar time offsetting scheme should be used, with an artificial offset by 1 minute of the ZTD at the far end of the time interval. For example changing the first ZTD estimate valid at hh:15 to hh:14, if doing 15 minutes batches.

(Still to be done on the minute level, since in the WMO BUFR format time can only be handled to the precision of one minute, and we are not able to change the BUFR format regarding parameters that are shared with other observations, such as valid times and locations.)

Validation of cost file header content

The header part of the COST file entry for each site provides some info about the site itself and about the processing done by the AC. We have noticed some discrepancies between the ACs, and would be glad if you can find time to validate the content for your AC.

There is particularly one aspect which is important, **the height of the antenna reference point (ARP) with respect to the geoid**. All ACs provide ARP height with respect to the ellipsoid, but not all do with respect to geoid. In connection with the BUFR encoding at UK Met Office, the geoid height (EGM96) is calculated if the geoid height is missing in the COST file, and inserted in the BUFR file. (The ellipsoid height does not go into the BUFR file). However, we do for principal reasons not want to change the original COST files, leaving it to end-users of COST files, to convert themselves. To some this is a big obstacle. Some may not even be aware of the need for it. Please do the conversion when encoding the COST file. To the precision of NWP, it is not important whether it is one flavor of a current geoid model or another which is used.

There is another set of descriptors in the headers which are of less importance to NWP usage.

ARP height above benchmarch: Varies between AC for same site. Please use missing value identifier if not known (-999.999).

Antenna and receiver: Varies between ACs for same site.

When changing processing strategy, in particular if moving to sub-hourly processing, notice that line 7 in the cost file header "Time Increment, Update Interval, Batch Length" should be changed accordingly.

Whereas these entries are less important to NWP than geoid height, some of them are of importance when doing inter comparison of GNSS processing methods and equipment and trying to identify the

effects of those on ZTD quality. Therefore it would be good if an effort is made to streamline header content.