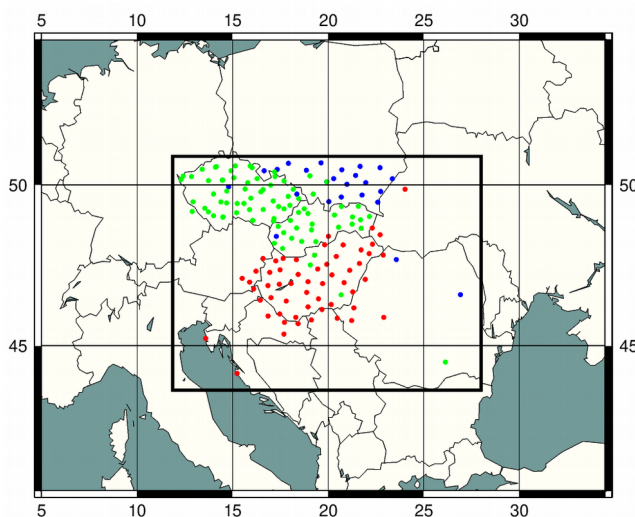
Interactive comment on "The use of GNSS zenith total delays in operational AROME/Hungary 3D-Var over a Central European domain" by Máté Mile et al.

We would like to thank referee #3 for his/her valuable suggestions. In the following, you will find our responses, separately for each comment/concern. We are confident, that we can provide a revised version of the manuscript, which is trying to address all of your points.

In black: referee observations
In blue: our response

Anonymous Referee #3

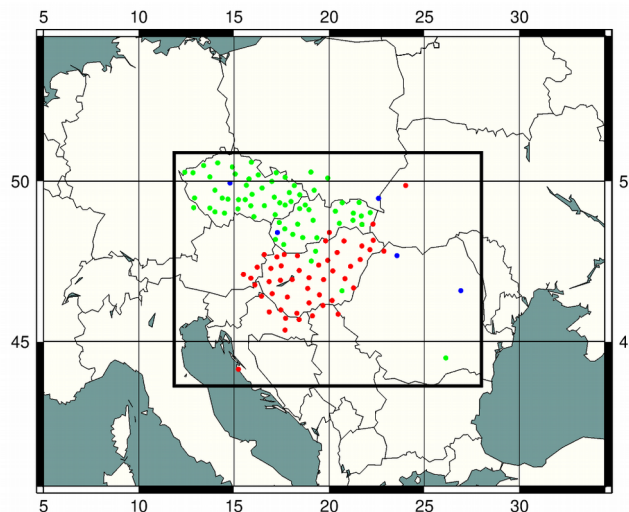
This paper describes experiments made with ground based GNSS observations as an additional observation type in the Hungarian NWP setup. There is nothing novel presented but it is an interesting contribution to the field in the sense that it confirms previous results also in a model domaine over central Europe. The paper describes the setup and assimilation procedure well and can be published. There are however a few things that need to be clarified first. One thing that is valid through the entire manuscript is that the figures are unreadable and not well explained. For example are the labels on the x-axis on figures 3, 5 and 6 impossible to read.
The quality of above mentioned figures is now improved in the manuscript.

Specific comments:
Page 2, line 33: The authors should also mention the project TOUGH, Targeting Optimal Use of Gps Humidity measurements in meteorology which was the predecessor to E-GVAP. Perhaps also a reference to the projects mentioned here, final report or a publication (e.g. http://tough.dmi.dk/deliverables/d14-final-rep.pdf)
This is now mentioned in the manuscript.
Page 3, line 23, Remove "to" in "...started in the 1990s joining to..."
Corrected.
Page 3, line 27, Missing reference right after Meso-NH?
Fixed (Lafore et al. (1997)). Lac et al. (2018) is also added.
Page 3, line 36-37: Remove "called OI_main"
Removed.
Page 3, figure 1: I would like to see a map over a slightly larger area with the model domain indicated by lines. It will make it easier to orient oneself and get a better feeling of where we are in the world.
New figure is added which shows GNSS stations as well.

Page 4, lines 7-10: Please explain figure 2 in more detail. What does absolute and relative mean?

Explained and more details are added. Figure's caption is also improved.

"The relative DFS is a normalized value by the amount of observations for each observation subset providing the diagnostic information regardless the actual amount and geographical coverage in the assimilation system."

Page 4, section 3: It would be helpful with a figure indicating the position of the GNSS stations. Perhaps include it in figure 1? It would also be nice to see which of all stations that are active in the assimilation, i.e. highlighted in some way.

We agree and new figures are added. Figure 1 shows now all available GNSS stations inside NWP domain (see above), and another figure highlights only selected GNSS stations (see below).

Page 5, lines 31-33: Very strange sentence. I do not really understand. Please re-formulate and clarify.

The sentence is reformulated.

"Although quality control procedure of the variational scheme contains the so-called background check (which is dedicated to reject observations far from model background state), one also needs to ensure that only observations with Gaussian, zero mean and uncorrelated errors are selected in the assimilation (i.e. reliable stations)."

Page 5, line 35: "...GNSS ZTDs by default are blacklisted in the system." Why is this the case and what does it mean?

This sentences was reformulated and the above mentioned (unnecessary) part was removed as well.

"For that purpose, a specific pre-selection procedure has to be performed by checking passive observation minus first-guess (OMF) departures on a training period."

This half-sentence about blacklisting is basically a technical remark which is trying to highlight that GNSS ZTD is a special observation type in the AROME's assimilation system. Due to historical reasons, the use of E-GVAP ZTD data required this special whitelisting approach, because processing upgrades, network expansion, and new analysis centers do not support to apply regular blacklisting solution (i.e. all data are active). It might explain why the pre-selection and a whitelist generation are necessary, but not really important to mention such technical detail in the manuscript if the reviewer can agree.

Page 5, line 37: It is assumed that the training period is sufficient. Can this be verified in some way.

Due to AROME/Hungary has 8 analyses in a day, the 15 days (or actually 16 days from 15th to 31st of May) together gives 128 observation minus background departures per GNSS station to compute

statistics (static bias, observation error, etc). It was not verified, but similar approach and data amount were applied in other studies (e.g. Yan et al. (2009) with AROME/France).

Page 6, line 13: Since the study includes a bias correction the predefined limits of OMF can be set more tolerant during the training period. This may include stations with a large bias that after bias correction can give a useful contribution.

We agree with this comment. There is still room for better pre-selection and bias correction in the future. We accepted the limits proposed by Yan et al. (2009), hence only 9 stations out of the total 197 (less than 5%) were excluded due to bias and stdv maximum thresholds.

Page 7, line 8: "...a space/time -average..." How does the space-part come into this? This is done fore each observation station individually is it not?

This is a mistake and the manuscript was corrected accordingly (i.e. "space" is now removed).

Page 7, equation 3: What is n here? All observations in total or all observations from one station during the training period? I hope for the latter and that BIAS really is BIAS(i).

We agree that equation 3 was not fully correct and as the reviewer remarked the latter is correct. The sentence and the equation are modified as the following.

"The observation bias of a GNSS station (station) is detected as a time-average of observation (oi) minus model-background (bi) differences considering the number of analyses (n) during the time period (3)."

(3)
$$BIAS_{station} = \frac{1}{n} \sum_{i=1}^{n} o_i - b_i$$

Page 7, line 14: OK, so the bias is calculated for each station separately. It is not really clear from the above.

The previous sentence was changed to make it more understandable. See the new sentence above.

Page 8, figures 4 and 5: Is the time period the same for both figures?

Yes, the figure's caption is now improved.

Page 8 line 24: "...default set to 60..." 60 what?

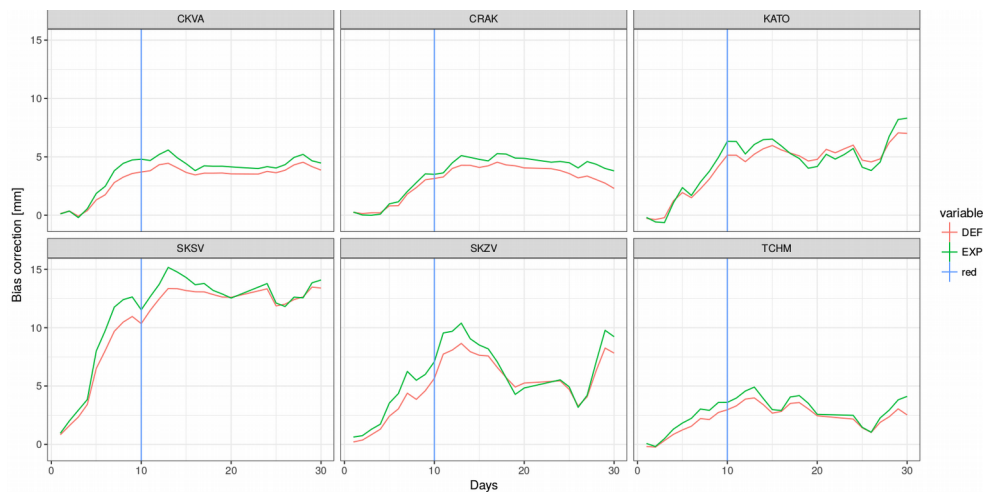
The stiffness parameter is a positive scalar parameter without physical dimension which determines the adaptivity of bias correction.

Page 8 line 25: Is 15 days enough for all stations, even those with a very large bias? Has this been checked?

According to variational bias correction scheme, the bias halving time is 5 days with default adaptivity parameter (Nbg = 60) and 8 AROME analyses per a day. There is a paper from Cameron and Bell (2016) (now cited in the manuscript) which explains it in details. The total bias elimination time is replaced by bias halving time which is more appropriate.

"The magnitude of the adaptivity is decided by the stiffness parameter which is by default set to 60 and taking into consideration that AROME/Hungary has 8 analyses in a day, the bias halving time corresponds to about 5 days (Cameron and Bell, 2016)"

"Cameron, J., and Bell, W.: The testing and planned implementation of variational bias correction (VarBC) at the Met Office.
https://cimss.ssec.wisc.edu/itwg/itsc/itsc20/papers/11_01_cameron_paper.pdf, 2016."

For 30 days of the summer period, the time evolution of bias correction was checked for various GNSS stations with the default and an experimental (Nbg = 40) setup. It can be seen on figure below that bias was eliminated with default (DEF - red curves) setup within 15 days. It is true for SKSV station which has generally larger bias than other stations.

Additionally, the first 5 days from summer impact study was excluded, but the variational bias correction was performed continuously resulting 21 days (i.e. 3 weeks between 15th of May and 5th of June, 2017) "bias spin-up" before first long AROME forecast in OSE.

Page 10, lines 5-7: Why only 30 synop stations? How are these selected? Are the ZTD observations distributed over the entire model domain?

These are the most reliable SYNOP stations from the Hungarian Met Service (called main synoptic stations) exchanged internationally in Global Telecommunication System (GTS) as well. These SYNOP stations are not covering the entire NWP domain, but Hungary.

Page 11, figure 7: Perhaps outside the scope of this study but there is a very strange diurnal (?) cycle in the verification. Just a few words to explain would be good.

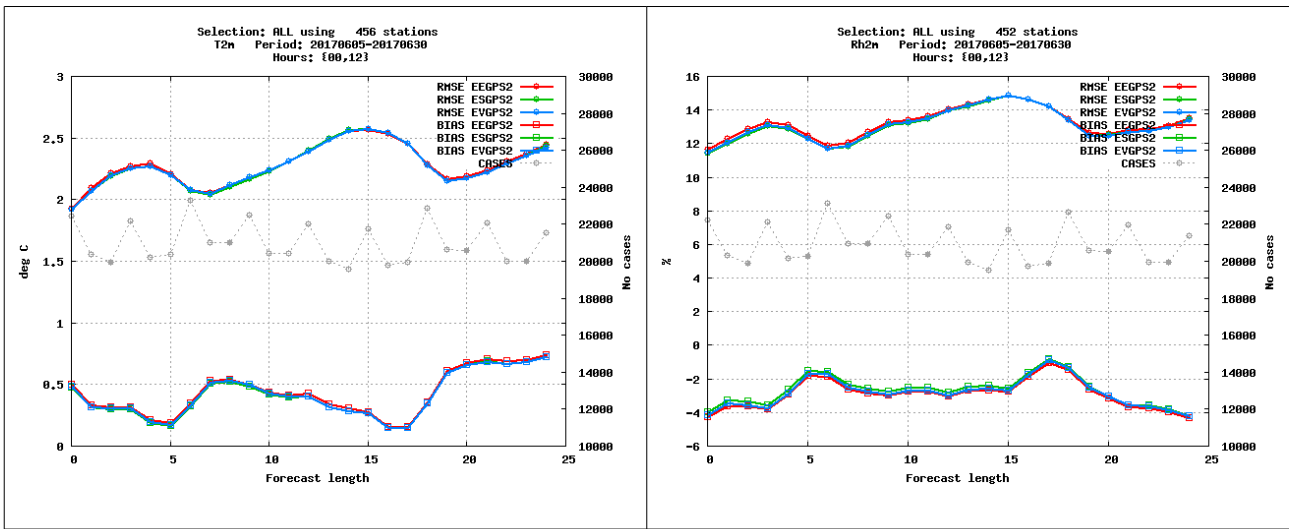
We agree and that can be explained by the characteristics of the surface model (SURFEX) and surface assimilation scheme which is mentioned in the Section 2. There are ongoing surface modelling activities at OMSZ which is aiming to improve surface fields of AROME model. However, the GNSS ZTD is assimilated in upper-air 3D-Var assimilation. A short sentence is now added to Section 5 reflecting to this behavior as the following.

"The AROME/Hungary forecasts have usually warm and dry bias during night-time, however, the assimilation of GNSS ZTD cannot mitigate this issue."

Page 11, line 20: Overfitting is not likely the explanation here. That would show up in the 1-3 hour forecast range and would be clearly visible in figure 10.

The winter OSE is proposed to be removed from the manuscript according to the remarks of other reviewer.

Page 11, lines 26-27: I do not understand the conclusion that the results are neutral since it was verified against Hungarian synop stations. Would the conclusion be different if it was verified against more stations?

Verification against all available SYNOP stations shows limited impact (closer to neutral), because we are not assimilating GNSS ZTDs from many neighboring countries like Romania, Austria, Slovenia, Croatia, and/or Serbia. Please see below RMSE and BIAS verification scores against all available SYNOP stations for T2m and RH2m.

Page 14, figure 10: Again a strange cycle in the verification that could be explained with a few words (if possible).

Yes, please see above.