Note: All Line numbers refer to the No Markup version of the revised Manuscript.

Reviewer #3

Review comments to the paper ”Assimilation of GNSS Zenith Delays and Tropospheric Gradients: A Sensitivity Study utilizing sparse and dense station networks”, manuscript ID egusphere-2025-19.

General comments:

The topic of the paper is really interesting and could potentially be very valuable for areas with sparse networks of GNSS receivers. I write “potentially” since from what I read from the paper it is not possible to determine very much from the results. More specific comments will follow but as the paper is written it is not suitable for publication. In fact it is rather poorly written with very little explanations of figures and results which makes it impossible to understand what is plotted and how it should be interpreted. The recommendation is therefore major revisions with a much better description of what is verified and why. I would as well like to see results for forecasts longer than 5 hours.

Specific comments:

1. A minor comment from the introduction: “ZTDs are the only source of moisture data used operationally” – This is not true. Moisture is obtained from radiosondes as well as from several satellite observations. Nowadays, many operational NWP models also include relative humidity at 2 meter height from synop stations.

Thank you for the comment. The sentence has been corrected. Please refer to the second comment from reviewer #1.

2. Line 123: Does entire Figure 1 show the model domain? If so (1), there is a lot of area that is not covered by GNSS observations. If so (2), it does not match the domain shown in Figure 4.

Yes the entire Figure 1 shows the model domain. We have now changed the text in the manuscript to clarify the Figures 1 and 4.

Please refer to **Lines 123-127.**
“In this sensitivity study, we are using the GNSS observations from the Benchmark data set which was collected within the European COST Action ES1206 GNSS4SWEC (Advanced GNSS tropospheric products for monitoring severe weather and climate; Douša, Jan, et al., 2016). The GNSS stations in central Europe covering Germany, the Czech Republic, and part of Poland and Austria provided the data during this campaign.”

Please refer to **Lines 298-299.**
“Figure 4 is a zoomed-in map that covers only the countries where the GNSS stations were located.”

3. Line153: The CV5 option does not mean anything unless you are familiar with WRF. Explain further or remove.

Yes we agree with the reviewer. The CV5 option has been elaborated for the readers.

Please refer to **Lines 186-192.**
"The $\boldsymbol{B}$ matrix for the regional simulations was derived from the forecast statistics by analyzing differences in 24-hour and 12-hour predictions over a month using data from May 2013. We chose the CV5 option for independent control of moisture levels, as it minimizes interference from other control variables. CV5 refers to a version of the background error covariance matrix used in the WRF model. It incorporates five control variables: stream function (Ψ), unbalanced velocity potential (χ_u), unbalanced temperature (T_u), pseudo-relative humidity (RH_s), and unbalanced surface pressure ($P_{s,u}$). Pseudo-relative humidity is the ratio of Q to $Q_{b,s}$, where $Q_{b,s}$ represents the saturated specific humidity of the background field."

4. Line 154: A 6 hour assimilation cycle is not a rapid update cycle. In fact it is the opposite. Most NWP models nowadays run with 3 hour cycling or a rapid update cycle of 1 hour.

Yes we agree with the reviewer. Rapid update cycle has been removed and replaced with "Six-hourly DA cycle".

5. Lines 155-165: This is a bit hard to follow since the authors have named the experiments after the included observations. The observations are described in the following section. I would suggest to describe the observations first, i.e. switch sections so that 3.2 becomes 3.1 and vice versa.

Thank you for the comment. The flow of the manuscript section 3 has now been restructured.

6. Line 165, Figure 2. A spin up of 12 hours is mentioned in the figure caption. Why is this spin up run and is 12 hours enough? It should be explained in the text and not in the figure caption.

Yes we agree with the reviewer. Now there is an explanation in the manuscript and not in the figure caption.
Please refer to **Lines 193 to 196.**
"Spin-up is essential for the model to stabilize with the initial and boundary conditions, enabling it to respond accurately to any desired inputs. Only after a sufficient spin-up period can the model forecasts be considered reliable for further analysis through data assimilation (DA). For our study, we adopted a 12-hour spin-up before the assimilation (Lauer et al., 2023)."

Reference included:
Lauer, A., Devaney, J., Kieu, C., Kravitz, B., O'Brien, T. A., Robeson, S. M., ... & Vu, T. A. (2023). A convection-permitting dynamically downscaled dataset over the Midwestern United States. Geoscience Data Journal, 10(4), 429-446.

7. Line 159: Why only run 5 hour forecasts? In an impact experiment you would normally want to run at the very least 12 hour but up to 24 is recommended in order to see how persistent the impact from the observations is. If there is a big impact on the analysis and very short forecasts that quickly disappears it can be an indication that your observation error is wrong and you give too much weight to the observations.

Thank you for the comment. As per the reviewers' comment we have now included a 24 hour forecast from the analysis for 12 DA cycles. We have included a new figure which depicts the observation impact forecast for 24 hours. The impact of the assimilation holds for not more than 12 hours which is quite reasonable for moisture data assimilation. We have now introduced a section 4.3 in the results as below:

"**4.3 Forecast impact**

To understand how long the effects of GNSS observations assimilation persist within the model, we conducted simulations of 24-hour forecasts based on a three-day analysis. Each day included four assimilation cycles, resulting in a total of 12 forecasts, each covering 24 hours. The forecast is better validated with independent observations that are not assimilated into the model. With the 18 excluded GNSS stations, we can directly compare the model forecast with observations from the GNSS stations. Figure 6 compares the 12 forecast average with the GNSS ZTDs and TGs, including the North and East Gradients, to compute the standard deviation. We analyzed three impact experiments: ZTD_1.0°, ZTD_0.5°, and ZTDGRA_1.0°, in addition to the control run. As anticipated, the effects of the three impact experiments gradually diminish and converge with the control run. If we define the endpoint of the impact as the moment when the standard deviation of the impact experiment aligns with that of the control run, then the duration of the impact is 12 hours. The effects of the assimilation last for no more than 12 hours, which is quite reasonable for moisture data assimilation. Additionally, it is important to note that incorporating TGs along with ZTDs enhances the forecast. Furthermore, the forecast impact of ZTDGRA_1.0° is comparable to that of ZTD_0.5°."

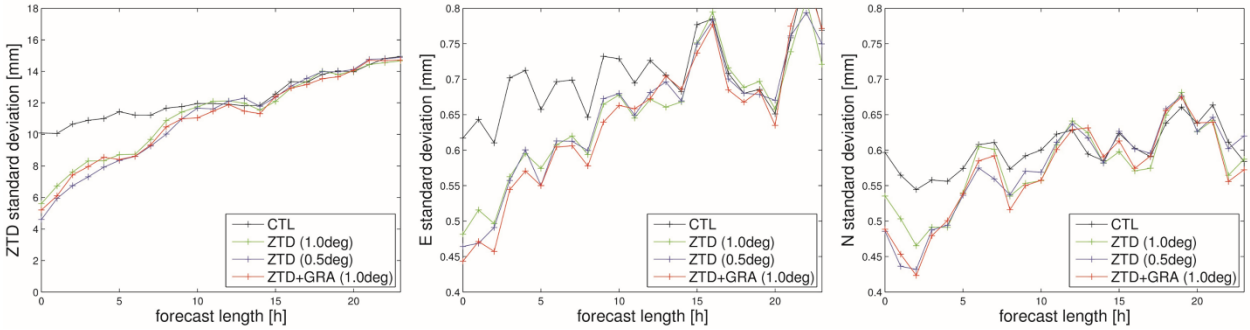

Figure 6. Average forecast impact with a 24-hour lead time initiated from 12 analyses over 3 days starting from 6 May 2013 00 UTC. The control run (black), ZTD_1.0° run (green), ZTD_0.5° run (blue), and ZTDGRA_1.0° run (red) forecasts are compared to independent GNSS stations. Subplots- **Left:** ZTD standard deviation; **middle:** East Gradient standard deviation; and **right:** North Gradient standard deviation.

8. Section 3.2: Why was the resolutions 0.5 degrees for the dense network and 1.0 degrees for the sparse network selected?

Thank you for the comment. We have now elaborated with additional sentences.
Please refer to **Lines 148 to 156.**
"To ensure a homogeneous set of observations across the domain, we excluded collocated and clustered stations and specifically chose GNSS stations with data availability exceeding 75%. In addition to complying with our WRF model domain, we carried out a simple thinning of observations ('homogenization' of station distribution). **The thinning method was conducted in two steps. First, a 0.5-degree mesh was constructed. Then, GNSS stations were selected based on their proximity to the mesh-grid point. Finally,** we obtained a station network with a resolution of about 0.5 degrees. After these steps, we were left with around 250 GNSS stations over the Benchmark domain. For the sensitivity experiment, we created another thinned station network with a resolution of about 1 degree that contained around 110 stations (see Fig. 1). **The same thinning procedure was used again.**"

9. Line 178: "a station specific bias correction" – Is this a variational bias correction or a fix one? In case of variational, what are the predictors and if fixed, how was it derived?

Thank you for the comment. We have now added a sentence to clarify the point.
Please refer to **Lines 161 to 162.**
"We utilized analyses from our control experiment to implement a "fixed" bias correction, addressing potential biases in the GNSS dataset (Thundathil et al., 2024)."

10. Lines 179-180: The observation errors, how were they selected or derived?

Thank you for the comment. We have now added an explanation as follows:
Please refer to **Lines 164-170.**
"The standard observation error for ZTDs in operational forecasting typically ranges from 5 to 15 mm. Similar to our previous study, Thundathil et al. (2024), the same observation errors were adopted: 8 mm for the ZTDs and 0.65 mm for the TGs. Given the high quality of the observations from the Benchmark campaign dataset, we have maintained these same error values in this study. The North and East gradient observation errors were calculated based on an analysis of the observation-minus-background (OB) statistics from the control run. OB statistics encompass both observation and model errors. An observation error of 0.65 mm was conservative since we did not want to force the model too much to the observations."

11. Line 180: "we set up a thorough network of surface reports" – What does this mean? That you install your own observations?

Thank you for the comment. We understand that the sentence did not convey the message due to its inaccurate construction. The sentence is now clarified as below. Please refer to **Lines 170 to 171**.
"To improve the analysis, we assimilated a set of conventional observations in addition to the GNSS observations. The conventional observations included a network of SYNOP stations across Europe."

12. Entire section 4: The authors compare the RMSE and the improvement of RMSE. The RMSE of what? Is is the comparison of the analysis and observations or forecast and observations? If the latter, what forecast lengths?

Thank you for the comment. We have now included an explanation how the RMSEs were calculated.
Please refer to **Lines 216 to 220.**
"The model simulation for two months in each experiment comprises six-hourly analyses and a five-hour forecast in between two DA cycles. Hence, the model simulations consist of hourly model outputs of analyses and forecasts. This hourly model output is compared to the corresponding GNSS station data for each experiment to calculate the root mean square error (RMSE). The model ZTDs and TGs at the locations of each specific station are computed for the RMSE. We term this the "station-specific RMSE.""

13. Lines 193-204 and Figure 3: The authors state that they "clearly demonstrate" a positive impact. I don't really see how. First, Figure 3 shows RMSE change (again of what) in percent while the text describes the same change in mm. Please be consistent. Secondly, the figure caption need to describe the figure better, e.g. what is NG and EG (one could guess but still)?

Thank you for the comment. The word "clearly" has now been removed. The results are now explained more clearly.
Please refer to **Lines 222 to 227.**
"Figure 3 is simplified in a percentage analysis of the station-specific RMSE plot (please refer to Figures A1-A6 in the appendices) for a more straightforward interpretation. Here, the control experiment was kept as the base experiment, and the ZTDs and ZTDs plus TGs assimilation experiments in dense and sparse configurations were compared. A reduction in the RMSEs indicates improvement in the assimilation experiment. The reduction in RMSE is represented as a percentage increase, which means the higher the reduction, the higher the percentage. Table 1 lists the average of the station-specific RMSEs of all the DA experiments for the two months."

Also the Table A1 in the Appendices section has been moved to the main section for reference to the mm values of the mean station-specific RMSEs.

14. Line 206: "we extend the analysis to include 18 independent GNSS stations" – Does this mean the these are in addition to the other stations or instead of the other stations? And why only 18? There are 430 available stations, you select 250 which means you should have an additional 180 stations to use for verification.

Thank you for the comment. We have now briefly answered the idea behind choosing the 18 stations.
Please refer to **Lines 157 to 161.**
"In line with the approach of Thundathil et al. (2024), we intentionally excluded 18 stations from our dataset for validation purposes. These excluded stations were chosen strategically to maintain a balanced spatial distribution, aligning with the locations of the German Weather Service (DWD) radar stations. The remaining stations included in the model are referred to as "allowed" stations. This method enabled us to analyze improvements with respect to independent observations."

15. Line 214: “assimilation of ZTD and TGs significantly enhances that accuracy” – Have you tested the significance or is it just a feeling? If you write this it needs to be validated.

We agree with the reviewer. More evaluation is needed for this statement. We would like to change the statement to the following. Please refer to **Lines 248 to 249.**
“The two-month-long statistical evaluation confirms that the combined assimilation of ZTDs and TGs improves the humidity field.”

16. Line 234: Again, significant, what does it mean?

We have now removed the word significant. Please see the previous comment.

17. Figure 4: The unit should be at the large color bar. The panels are labeled a-j but these are never referred to.

Thank you for the comment. This has now been rectified. Please refer to **Lines 300 to 301.**

“DA cycle 1 refers to Figures 4a and 4f, and DA cycles two, three, four, and five refer to Figures 4b and 4g, Figures 4c and 4h, Figures 4d and 4i, and Figures 4e and 4j, respectively.”

18. Line 246: Perhaps it could be of interest to separate the very lowest model levels and the slightly higher ones, i.e. separate the boundary layer and the free atmosphere?

Yes, we agree with the reviewer that it could be of interest. We would like to publish further analysis on the vertical distribution of the humidity and the influence of tropospheric gradients in a future manuscript.

19. Lines 251-252: SSIM index parameter – Please explain shortly what this is and how it is calculated. If the reader is not familiar with SSIM the number 0.98 does not mean much without reference.

Yes, thank you for the comment. We have now described the SSIM index and how it is computed.
Please refer to **Lines 307 to 315.**
“The Structural Similarity Index (SSIM) is a metric used to quantify the similarity between two images (Wang et al., 2004). Here is a short explanation of the computation of SSIM in our study:

$$SSIM(A,B) = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)}$$

Here A and B represent the images in the left column and the right column respectively. WVMR is the moisture variable presented here in the images with a span of 101 colors. μ_A and μ_B are the mean, σ_A^2 and σ_B^2 are the variance, and σ_{AB} is the covariance. The variables c_1 and c_2 are computed based on the color span in the images.

$$c_1 = (k_1 L)^2$$

$$c_2 = (k_2 L)^2$$

Here k_1 and k_2 are 0.01 and 0.03 by default. L here is the total number of colors in the color bar minus one. Hence the values of c_1 and c_2 comes to 1 and 3."

20. Figure 5: Again, what is shown? RMSE of what?

Thank you for the comment. The explanation in provided in the manuscript as below. Please refer to **Lines 317 to 320.**

"In order to analyze the humidity profile correction in the assimilation experiment, we computed the RMSE of specific humidity profiles from model simulations with respect to ERA5 at five locations spread equidistantly across the domain (for details, see Thundathil et al., 2024)."

Please refer to the updated Figure caption:

"**Figure 5.** The RMSE of specific humidity profiles compared to ERA5 for Control run (black), ZTD_0.5° run (blue), ZTD_1.0° run (green), ZTDGRA_1.0° run (red), and ZTDGRA_0.5° run (purple). Profiles were compared at five selected stations for 220 DA cycles, totaling 1100 profiles for the average plot."