Respond to the Anonymous Referee #2

In general: Thanks a lot for your comments and suggestions and especially for your helpful major comments!

To the major comments:

1. *You need to decide how "much" you want to discuss SMART+ in this work.:* We have added a subsection with the most important information about SMART+.
2. *I cannot understand how the Rot + Exp method shows almost no difference from the background NeQuick version.:*
   We tuned all the tested models with respect to their error models (process noise, systematic error, weights between X_a and the background model) and did several runs (up to 15 not show in the paper) to figure out the best configurations. When we increased the error terms, the results became unreasonable after some days of EnFK. At the end, the presented variants of the methods delivered the best results (reasonable and providing the smallest validation residuals). However, we think that especially for the Rot+Exp method the manuscript-presented choice of the error terms and weighting parameters could be still not optimal. For this method, these errors are probably too small, and that is the reason why the differences from the background model are small. We think about an adaptive method to calculate these errors, but this will be future work which cannot be covered in this paper.
3. *In a related point, I am surprised by the amount of differences between each of the methods. I would have expected the DA to be the dominate term, with the changes caused by the propagation being minor. Since this not seem to be the case I think it needs to be carefully described why this is.*
   We think that the main reasons therefore are the following: (1) Compared with the huge topside iono & plasmasphere volume, there are not much measurements available (compared for example to the situation when you apply data assimilation using all the available ground based data). (2) Using sequential data assimilation one has to deal with changes summed up with time. At the beginning of the period you are interested in (for example in our case, on DOY 041, early morning hours), the differences between the results using different propagation methods are not so big, but then they become more, because the propagation goes in each time step complete different directions. Like e.g. using Rotation you just rotate the analysed solution, but using Rotation with exponential decay you build in each propagation step a weighted sum between rotated analysed solution and background.
4. *For example you have picked 20mins propagation time but it would be interesting to see a histogram of the errors, for each technique, overplotted with increasing time, say 5min, 15min, 30min, 60min (or something similar).*

   We investigated the EnKF as well as SMART+ regarding their ability to reproduce assimilated STEC as well as to estimate independent STEC measurements and in-situ electron density measurements in dependency on changed temporal resolution of 60 minutes. For all tested methods as well as for NeQuick model we obtained no significant differences in the statistics between the validation results for $\Delta t=60$ versus $\Delta t=20$ minutes. We think, this indicates that the presented propagation methods work in generally well. Contrary, for example for simple persistence as propagation method, the reconstruction results become implausible more quickly if $\Delta t$ is enlarged.

5. *I would find it interesting to see the difference between t(n) and t(n+1) for each method. So following the DA step (resulting in X_a) you could apply each method separately, and show a difference from SMART+ (as that is the best model). We would then be able to see an example of the differences (and what changes there are globally).*

We have added Figure 7 showing for all the EnKF variants the differences $E\left(X^f(t_{n+1})\right) - E\left(X^a(t_n)\right)$ on the left and the percentage differences on the right calculated as $100 \cdot 2 \cdot \left(E\left(X^f(t_{n+1})\right) - E\left(X^a(t_n)\right)\right) \Big/ \left[E\left(X^f(t_{n+1})\right) + E\left(X^a(t_n)\right)\right]$ for a fixed $t_n$.

6. *In all cases comparing the methods to persistence (let X_aˆt = X_bˆ{t+1}) would also help to demonstrate the models effectiveness.*
   We indeed tried also "persistence" as a propagation method. But "persistence" delivered very bad results. Already after ca. 24 hours, reconstructions based on "persistence" had shown unreasonable effects, like completely misplaced equatorial crest region. The propagation time of 20 minutes seems to be too long for the method "persistence". We have added this information to the paper.

To the technical comments:

1. We put two Ref. to the statement "Around 50% of the signal delays…" Line 28.

2. Regarding your questions on the factors which we had chosen within the equations of the forecast methods, like 8/10 in Eq. (9): We had derived them by an internal validation. Particularly, we had conducted several runs with different factors and then analyzed the results. At the end, the factors delivering the best results have been applied in the paper. We have added a corresponding note in the paper.

3. Regarding your question about the generation of the ensembles, Line 170ff: We chose 3/100 based on our analysis of the sensitivity of the NeQuick model on F10.7 (we tested also higher factors). We did this analysis indeed only around the periods important for the paper. Thus, the factor 3/100 could be a different for others periods. We think that the generation of the ensembles with respect to their variation and amount forms another important topic to apply the EnKF (in addition to the choice of the propagation method). Until now, we have not found publications in the field of ionospheric research focusing on this topic. We would be very thankful for a reference.
   An extensive validation of the reconstruction results in dependence on the number of ensembles and the method to generate them is in work but not covered in the scope of this paper.

4. Fig. 2, top subfigures, (your question at line 228) shows the reconstructed results for DOY 076, 19:00 UT. In order to derive these figures: (1) We take the analysed ensembles from UT 18:40 and (2) propagate those 20 minutes in time using "Rot and Exp". (3) We update the propagated ensembles by the assimilation of the currently available (around 19:00 UT) STEC measurements (analysis step of EnKF). (4) We calculate the ensemble mean, which is partly shown for a fixed latitude/longitude/height in Fig. 2. The described propagation-update-cycle, which has been used to obtain the results shown in this figure, had started at time 0:20 UT on DOY 041 and has been repeated until 19:00. Consequently, data has been assimilated and "Rot+Exp" has been applied.

5. Your comments to lines 234, 238: We have added figures 4, 5, showing the differences and the percentage differences between the reconstructed and the modeled electron densities; and figure 6, illustrating the corresponding measurements geometry.