Angeo-2019-160

Reviewer #3

Comments on the paper:

Modeling Total Electron Content derived from radio occultation measurements by COSMIC satellites over the African region

Reference file: angeo-2019-160-manuscript-version4.pdf

The authors present a regional modeling of the TEC deduced from the vertical profiles of the ionospheric density, obtained by radio occultation with the COSMIC satellites over Africa. Despite using the entire database spanning a decade (2008-2018) and limiting themselves to magnetically calm days, they do not have enough measurements to fit these measurements to different variables in their model. First, they present their interpolation algorithm to have a value at each point of a geometric grid. The result of their model is discussed on our knowledge of variations in the ionosphere.
It is difficult to comment on an article which has already been extensively modified following numerous comments from the 2 previous referees. However, I still have a list of remarks that I have positioned in minor (m) and major (M):

(m) Paragraph 'Introduction' on 3 pages (p2-4).
The first 24 lines present the ionospheric models used in single frequency for the GPS and Galileo systems. The following 7 lines relate to the IRI model. The following 16 lines talk about GIM maps obtained by processing GNSS measurements. Finally, on line 31, p.3, the word COSMIC appears to present a global model of GNSS and RO measurements and to conclude that the result over Africa is different from CODE's GIM maps.
That's a lot of text that takes a bit away from the topic being discussed. I would have preferred a more direct introduction based on the 3 keywords of the title:
- Why did you choose an area above Africa where additional measurements are and when a choice above Europe or North America would have made it possible to use the many vertical ionosondes to validate the model?
- Why use only RO / COSMIC measurements to build a model and which constitutes the originality of this work?
- Why a mathematical model with spline functions? Currently, the word "spline" occurs only once, in the sentence on line 24, p4.
[It is likely that much of the current text of the introduction would have been found as a result of this questioning].

(m) p.5, line 9. The given database contains ionization profiles. To obtain VTEC, you must integrate up to the altitude of the COSMIC satellites. This altitude (~ 800 km) is not always specified (or later in the text) but it is important for a future comparison with other VTECs. It

is certain that all the profiles do not give the same final altitude and then two questions arise:
- if a profile stops at 600 km for example, what do the authors with this measurement?
- and more generally, do the authors make a selection and if so, on what criteria?

(m) p.9, lines 6-12. Following the comment on referee # 2, the text has been modified to understand that the 36 solar variables are indeed the 3 levels (L, M, H) repeated each month (3 * 12 = 36, table 2). The rationale is '**to represent seasonal TEC**'. But then, why keep a variable "months" (to study seasonal) which will require a multiplication of the number of variables by 12 (line 11)?

(m) p.9, table 2. I have not read the information, but I assume that the monthly flow values in table 2 relate only to the dates of the measurements used in the model. However, in 2012, we are in strong solar activity of SC#24 and perhaps with values of flux higher than those present. In this case, the validation will relate to an extrapolation, which can quickly lead to important TEC values?

(M) p.10, line 7. When there is little (or no) value in a box, the authors adopt a smoothing by splines. There are a lot of spline functions and the authors don't specify their choice. For cubic splines, one can obtain oscillations with stronger extreme values since the smoothing passes through the measurement points. For other spline functions, there is a reduction in variability since the interpolated curve passes between the points. The approach followed by the authors is important for the rest of the work and I think it would have been interesting to illustrate some typical cases with figures.
When there are multiple points in a box, the authors take the mean value? What is the variability (min / max) which gives information on the uncertainty of the measurement?
When a node has a sufficient number of measurements, do the authors keep the average or opt for the interpolated value?
Only one mathematical reference (deBoor, 1978) in this article is, in my opinion, too weak (compared to 38 geophysical references in the bibliography).

(m) Pages 10-11. The authors propose a 3-step algorithm for filling the geometric grid of measurements. It's a bit of an empirical method. Have the authors analyzed other interpolation methods starting from an irregular grid?

(m) p.11, line 15. I did not understand the convergence of the procedure after 3 rotations. Need to iterate until all the boxes are filled and maybe a number of 3 is not enough? At this stage, I think that the authors could have presented TEC histograms on 3,981,312 bins against the 121,447 bins input. Is it the same distribution (mean, rms)?

(m) p.13, line 13. The authors justify the quality of their model by the existence of a secondary peak at the magnetic equator already observed elsewhere. However, if I make a vertical line around 16 LT for example on the observed or on the model (Figure 1), I will see an irregular variation of the TEC (~ 10 tecu) in latitude with many secondary peaks (southern hemisphere for example) and not a steady decrease as expected. These secondary peaks are not physical and are due to averaging (hence my question about variability in a cell) and to interpolation. What do the authors think?

(m) p.14, line 11-20. I do not see quite the same thing that the authors describe in particular for the graph c in strong solar activity. Maximum north is on the equator (the 2 bubbles red colored) when expected at 20 ° N? The south EIA maximum appears to be well positioned.

(m) p.16, line 10. A first evaluation is made only on longitude 37.5 ° E due to the existence of GPS measurements and publications around these measurements. Fortunately, there are African stations at other longitudes. My question: since the model is built between -20 and 60 °E longitude, are the conclusions of 37.5 °E longitude valid for other longitudes? I would have seen an overall statistical result but I have no idea because the difference (observed-modeled) is less than 0.1 tecu on the 2 examples. Is this same conclusion for all longitudes?

(m) p.18, line 10. The authors do not provide any positioning on the 1600 points with an absolute difference modeling of at least 10 tecu. It is certainly for the year 2012 and not 2018 but the points relate to a particular hour or month? [I already pointed out a possible divergence of the model in one of my previous remarks in the case of an extrapolation with solar activity].

(M) p.21, lines 14-16. The authors validated their model with ionosondes in South Africa, therefore located in mid-latitudes. I think it is an exaggeration to say that we would have the same result (‘predicted **fairly well** using our model.’) With a low latitude ionosonde, the study remains to be done!

(M) p.25, line 4. I do not agree with this conclusion. It’s because the TEC variations are more irregular with the spline model compared to the NN model that it is the best! Admittedly, the variations of TEC with NN are over-smoothed (but GIM / CODG also for example) but the many variations in Figure 7 are first linked to the error on the profile estimated by RO and by the procedures of interpolation to give values the nodes of the grid which is a **mathematical** filling and not a **physical** one.
I also regret that the comparison of Figures 6 and 7 is purely visual and that there are no statistical figures of differences in the proposed text.

My conclusion is that there is a **real work of exploiting the RO data** for modeling purposes. The initial difficulty is the lack of measures to cover the Africa zone. Also, the authors were forced to introduce mathematical approaches to cover all the variables retained. They justified their model on a physical result of maps reproducing the large known variability’s. The model does not allow a fine-grained approach to the ionosphere compared to a more regional modeling with GNSS measurements. Their current conclusion is that their model leads to better results than the 2 empirical models (IRI and NeQuick) widely used. If the authors want to see their results applied to future studies, they must publish the coefficients of their model. Is this an objective of the authors?