

Authors' Response to Reviews of

Electron radiation belt safety indices based on the SafeSpace modelling pipeline and dedicated to the internal charging risk

N. Dahmen et al.

Response to referee comments from Yihua Zheng

We would like to thank the referee Yihua Zheng for the constructive feedback regarding this article. We report below, our detailed response to all the major and minor comments.

The reviewer would like to see more discussion regarding possible ways to overcome such shortcomings.

While the paper is relatively clear and easy to understand, the paper falls a bit short in an adequate demonstration of the warning system and possible ways for its further improvement.

-The new version of the manuscript will contain more clarified statements on the warning system performances and more insights about the possible ways to mitigate its limitations (in section 4 and in the conclusion), that might as well serve for the community.

Minor comments

Line 6: 'radiation belt dynamic' change to 'radiation belt dynamics'

Line 19: 'the space radiative environment' -> 'the space radiation environment'

Line 61: 'More particularly' -> 'More specifically?'

Line 150: 'measures' -> 'measurements'

Line 160: 'embarked on the GOES-15' -> 'onboard the GOES-15'

Line 187: 'As regards to' should be 'As regards'

Line 190: 'beforehand' -> 'before'?

Line 193: 'As regards to' -> 'As regards'

- The Typos referred above will be amended in the new version of the manuscript.

Technical details

The SafeSpace safety service website is not accessible. It needs to be fixed.

- We are experiencing technical difficulties with the service at the moment, and are working to put it back online as soon as possible.

Response to referee comments from referee 2

We would like to thank the referee for the constructive feedback regarding this article. During the revision of the manuscript, we have strived to address all the remarks, comments and suspended questions, for which we report our detailed response below.

Major comments

Line 17: cite Mann et al. 2018

-The new version of the manuscript will include this reference.

Please explain why B_y is used and not other components.

- B_z is a better first indicator of geoeffectiveness. However, the Helio1D model (and, to the best of our knowledge, most, if not all, other models in the literature) used to forecast the solar wind and IMF parameters at L1, cannot at the moment provide reliable B_z estimates (see e.g. [1]). That is why we use B_y instead. This is probably one of the main performance-limiting features of the SafeSpace pipeline, but one for which we have no other choice. How the absence of B_z limits the performance of the forecasting pipeline is discussed in more details in Brunet et al. 2023 (currently in review at AGU Space Weather). We have added a couples of sentences at the end of paragraph 1 in Section 2 of the manuscript to highlight this discussion.

Line 100: “For each one of the 21 members of the ensemble of parameters” Please list them. A Table may work well, with a comment on each, like a definition, the utility of that parameters, may an index of criticality you could define. Please do the same for the “ensemble of magnetospheric parameters consisting of 10 members”. This is a major correction as it is very important readers know what parameters are chosen.

-The construction of the ensemble members is already reported in the manuscript: The heliospheric segment of the pipeline provides a 21 member ensemble of solar wind parameters : velocity, temperature, density and the y component of the magnetic field in the GSE coordinates system (listed in section 2 line 74). These parameters will serve as inputs in the OGNN code to produce an estimation of the Kp index (as reported in section 2 line 78) and to produce an estimation of radial diffusion coefficients DLL (as reported in section 2 line 94) and the boundary condition (BC) for Salammbô (as reported in section 2 line 99). Then, each member of the solar wind ensemble is assigned to an internal ensemble of 10 members (section 2 line 101) of Kp values that will be fed to the VLF model, SPM model and FARWEST to produce pitch angle and energy diffusion rates (Dyy, DEE) (section 2 line 96). Finally the total ensemble that will be fed to Salammbô-EnKF contains 21*10 members of input data in the form of the previously cited BC, DLL, Dyy,DEE (section 2 line 92). As the objective of this paper is mainly focused on the definition and the testing of the SafeSpace risk assessment system, we decided to remain superficial in the description of the codes in the pipeline for the sake of clarity. Nevertheless, we provide the reader with several references dedicated to the codes to go further into their theoretical principles and performances. These

details are also gathered in the Brunet et al. 2023 paper (currently under review), dedicated to the inner-magnetosphere section of the SafeSpace pipeline.

P6, major: I would like more information about the historical data in the inner belt, in a dedicated small paragraph. I think there is a caveat of the method that should be brought in. In the inner belt, electrons above 1 MeV have been found to be below background instrument (Fennel et al. GRL, 2015). It is a main result from Van Allen Probes that has caused a revisit of the AE electron radiation belt model. It was confirmed by cubesat measurements (Li, X. et al., 2015). An important corollary, is that past measurements which show electron flux above 1 MeV, have been corrupted from proton background contamination. Even if the Van Allen Area was less active than previous solar cycles. Are the historical data cleaned up in the inner belt above ~ 1 MeV or not? From Figure 4 and the rest of the article, I understand it is not. Please state explicitly in the text whether it is done or not done. How does that affect the criteria defined by the authors of making a warning if “>1.2 MeV” electron flux above 2 or 20% are detected? About higher than 1 MeV intrusion in the inner belt, which do still exist, please see/cite, Pierrard et al. 2021 (figure 2, top panel) and Claudepierre et al. 2017. Also please mention that only 6 penetrations at >1.2 MeV have been measured at LEO between May 21, 2013 and December 31, 2019, from Proba V (Figure 1 in Pierrard et al. 2020). These articles can help to back-up the choice of the authors.

The reviewer thinks that because of the above, the 1.2 MeV channel of past measurements could be proton contaminated and is thus not the best target to choose for a criterion. I am also worried by the ‘>’ sign: selecting above 1.2 MeV channels can even select higher proton energies. Lowering the energy to 0.8 or 0.9 MeV would fall around the upper bound of the inner belt (Fennell et al., 2015; Li X et al., 2015) and select for sure electrons, which would be consistent with Salamambo’s physics. (Proton dynamics is not consistent with Salamambo’s physics at all). There would be no ‘>’ or just a 0.8--1MeV range, whatever energy bin of the instrument is available in that energy range. If done, that would make the 3 criteria defined by the authors p5 to select the same energy, ~ 0.8 MeV. I ask as a major correction, this caveat to be written. The author can cite the reviewer’s comment as differing of their choice. I note that the choice of the St Patrick 2015 storm for validation is a good choice because for that one storm is one among the 6 storms which had above 1 MeV intrusion in the inner belt... Please cite again Pierrard et al. 2019 (figure 2, top panel) and Claudepierre et al. 2017 showing it. Note and write that in their figure you can see how faint is the flux signal at 1.2 MeV so that defining an index based on that flux could be an issue. I do not ask the authors to agree with my claim but to report some questioning.

-Indeed, the raised point of proton contamination in the inner belt measurements (LEO) is very relevant and we acknowledge its absence in our paper. Nevertheless, before the construction of the historical distribution on the LEO orbit, we were fully aware of the proton contamination of the POES data and the latter was cleaned accordingly and is supposed to be free from the mentioned bias and in accordance with (Pierrard et al. 2019) and (Claudepierre et al. 2017). Therefore, the new version of the manuscript will include more insights on this aspect in the historical distributions section (section 3) with the appropriate references mentioned above.

Fig2: what are the energy/ies of the flux shown in each panel? Figure 2 is not discussed enough in the text. Please complement

-The figure reports the integral fluxes at the energies stated in section 3.1 after the cross calibration and cleaning step. These flux data served for the construction of the historical distribution. We added in the figure's caption the energies of the fluxes shown in each panel.

Major: From Fig 4: please make a Table that reports the exact flux values chosen for the alarm flux threshold for all three orbits and the 2 alarm levels (2 and 20 %).

- Noted, the new version of the manuscript will contain a new table that reports the exact flux values chosen for each alarm threshold.

P9: about the results. At LEO, there are 11 alarms from the observations and none from the simulation. The simulation has 11 missed alarms. Please write it explicitly. I am asking this feature to be related to the above discussion on the >1.2 MeV flux criteria.

-As the data was already cleaned from proton contamination, we can assume that the poor performance of the LEO index is inherent to the Salammbô estimation rather than to the definition of the thresholds.

Regarding to Figure 5: the authors acknowledge "very poor" results (which turns into Table 2—please write it). They write "This behavior is explained by the low grid resolution used in Salammbô near the loss cone (modelling interactions with the atmosphere). One can consider refining the grid to improve the LEO index estimation, but this operation will impose a too intensive computational cost on the daily pipeline." I do not think it is the only reason. The authors just don't have the physics embedded to solve for inner belt physics (for instance wave particle interactions and radial diffusion, both not accurate enough today—because both parts of open subjects). So please write a caveat here and refer also to the argument given above and to the fact that inner belt physics does not bring yet a full package of models (see below).

-We acknowledge that our discussion on the origin of the poor results was superficial. To be precise, we think that the problem with the LEO results has different origins. First there is the accuracy of the physical description of the inner belt processes as explained above. Second, the LEO region showcases high gradient distributions and steep physical processes that inevitably put constraints on Salammbô's numerical solver (which also prevents the assimilation of data at LEO). To mitigate this issue, the most accessible improvement within the perimeter of the SafeSpace project would be to refine the grids at LEO. But, this operation will degrade the operability of the pipeline or even make it obsolete for the presented application. However, outside the scope of SafeSpace, one should consider the improvement of the physical description for sure, but also the adoption of a new numerical method, dedicated to the radiation belt numerical constraints. This aspect is under study by the Salammbô team and led to the prototyping of a new finite volume implicit solver [1], that will

be presented in future studies. As this aspect is related to the perspectives of the Salammbô code development we decided to add this discussion in the conclusion and leave the results section (section 4) with the perspectives of the SafeSpace project.

Table 2 legend: please write explicitly that median and ensemble are from the simulation. Inside Table 2, consider adding the word “simulation” for greater clarity. Table 2: why not giving the median as well for the ensemble, for direct comparison with simulation? You can write something like: “3 (2--5)” with median (min, max)

- Indeed, gathering the median and the ensemble results should make the table clearer. We will amend the table and its legend as suggested above.

Please write explicitly (legend and/or text) that missed and false alarm are not available from observation. Consider adding a ‘/’ in the Table

- By definition, the observation-based indices are considered as a reference that the simulation seek to replicate. For the sake of clarity, the manuscript text and tables will be amended with the abbreviation N/A (not applicable) and with a mention of that observations are taken as reference and thus supposed to be perfect.

Line 194: “worrying” is a bit naïve. Maybe « critical » for the highest level of alarm associated to active times.

- The new version of the manuscript will integrate this correction.

Line 194: “adequate”. Please add “for MEO and GEO”. Results are not adequate for LEO\@.

- The new version of the manuscript will integrate this correction.

Line 195: results are mentioned as “good” in the previous sentence and though we have “However, 7 out of 8 moderate time alarms were missed”

- It is true that missing 7 out of 8 moderate time alarms on the GEO orbit seem as a mediocre performance. However, if we look at the time evolution of the index and to the active time alarm counting on the GEO orbit, one can see that most of the missed moderate alarms were replaced with active alarms (referred as false alarms) as the SafeSpace GEO index overestimated the observed index. From a risk assessment point of view, this is adequate as it shows that the system is globally conservative. Considering the complexity of the task, we think that being able to differentiate between a non-risky and risk activity can be considered as a good result. Nevertheless, we will consider in the new version of the manuscript, the use of more suitable words to portray the performances of the alarm system.

Still about LEO results, please write that LEO flux values may differ by orders of magnitude between flux at LEO orbits (a 500--700 km) and flux measured at low L-shell more in the magnetic equatorial plane ($L=1.1$ to 3), which is a region uncovered by the MEO and GEO indices defined by the authors. Please state in the article if you think (or not) that the LEO index should be used for this region. As written, for that region of space, the authors implicitly suggest their LEO index would be the one to consider. Maybe, but there is this caveat that flux are very different in the inner belt due to strong pitch angle anisotropy. This anisotropy is well visible and discussed in the MadEIS measurements shown in Fig 2 of Ripoll et al., 2019. Also, the flux difference between high latitude flux at LEO and flux in the equatorial plane are well visible in Fig 10 of Pierrard et al., 2021 where both are compared with each other. Please refer to them.

- The reviewer is right that there is a large gradient in the LEO fluxes regarding the altitude of the considered orbits, and that the predicted index value, even if the model was performing accurately, could not be used as the prediction of the flux value for any other spacecraft than POES satellites. However, we note first that the fluxes at different LEO altitudes are well correlated, so that if one percentile of the distribution is observed at one altitude, a similar percentile of the distribution would be observed at a different one. Hence the idea to present the index value as a percentile of the climatological distribution (as in plot 3 of the initial version of the article), and the choice of thresholds at 2% and 20% percentiles. Moreover, the choice of the orbits from the three indices was made considering the user needs for such indices. These orbits are by far the most populated ones. We could introduce similar indices covering other regions of space (such as, as suggested, lower equatorial regions), but few satellites populate such regions. For these specific regions, it is best to look directly at the PSD distributions without introducing daily averaged indices.

Table 3 legend: you don't have moderate alarms for observations as written in the legend. Please correct.

- We don't understand this remark. All three observations raised moderate alarms during the four-day active period.

Line 216, in the conclusions, please add "Reason of poor prediction at LEO were discussed for possible future improvement." (I am assuming that the reasons I give above will be inserted so that it can be briefly mentioned in the conclusions).

-The conclusion was amended with the discussion presented above on the LEO orbit poor results.

In the conclusions, about my comments on the so-called good results at GEO. Please develop a bit and moderate accordingly in the conclusions.

-The conclusion was amended with the appropriate choice of words to describe the GEO index results.

At the end of the conclusions, I suggest you had a few information and references, all linked to the fact that radiation belt science has to improve. At GEO, Van Allen Probes have revealed the major role of dropouts as a critical process responsible for removing the radiation belts (Turner and Ukhorskiy 2020 and references in there). This process is not yet fully accounted for in Salamambo (unless I am wrong and it needs to be mentioned) and requires more elaborated magnetic field line models. On the other hand, the other processes responsible for flux decrease or enhancement in the radiation belt involve wave particle interactions (WPI). In particular, in the outer belts, WPI are also complexified by the knowledge of the plasmaspheric density (see/cite recent review of Ripoll et al. Frontiers 2023) and of the waves. Many open physical subjects remain (Denton et al., 2016, Ripoll et al. 2020 (already cited) and Li and Hudson (2020)). Please make sure you have one or a few sentences opening on these facts and discussing that better predictions will come as physical modeling and understanding progress.

-The conclusion was amended with more discussions on the proposed improvements to the SafeSpace warning system in connection with the pipeline. For the GEO index, we suggest the integration of magnetopause shadowing dropouts modelling. This improvement should be straightforward as Salamambo already accounts for dropouts [2] (though not retained in the SafeSpace physical description). For the LEO index, we advocate for an improvement of the inner belt physics description aside the adoption of robust numerical schemes that withstand limitations such as steep gradients. The improvement of the other processes is mentioned also along with the Ripoll et al. 2023 reference as an open research field and a long term perspective.

References

[1] Dahmen, Nour, François Rogier, and Vincent Maget. "On the modelling of highly anisotropic diffusion for electron radiation belt dynamic codes." *Computer Physics Communications* 254 (2020): 107342.

[2] Herrera, D., V. F. Maget, and A. Sicard-Piet. "Characterizing magnetopause shadowing effects in the outer electron radiation belt during geomagnetic storms." *Journal of Geophysical Research: Space Physics* 121.10 (2016): 9517-9530.

Response to community comments from Yann Pfau-Kempf

We would like to warmly thank Yann Pfau-Kempf and the members of the Journal club of the Space physics group for their very constructive feedback and relevant remarks regarding this paper and we report below our detailed response to all their major and minor comments.

Major comments

l. 75 and following: We are surprised by the use of B_y , isn't B_z a better first indicator of geoeffectiveness?

- Yes, indeed B_z is a better first indicator of geoeffectiveness. However, the Helio1D model (and, to the best of our knowledge, most, if not all, other models in the literature) used to forecast the solar wind and IMF parameters at L1, cannot at the moment provide reliable B_z estimates (see e.g. [1]). That is why we use B_y instead. This is probably one of the main performance-limiting features of the SafeSpace pipeline, but one for which we have no other choice. How the absence of B_z limits the performance of the forecasting pipeline is discussed in more details in *Brunet et al. 2023* (currently in review at AGU Space Weather). We have added a couple of sentences at the end of paragraph 1 in Section 2 of the manuscript to highlight this discussion.

l. 111: Can the authors explain why they chose to use daily averages? Are these not underestimating strong events and wouldn't e.g. 12 or 24 hour fluences be more adapted to the purpose? Large flux events can be localized on timescales much less than a day and certainly do not obey Gaussian statistics, so the average will most certainly underestimate spacecraft exposure.

- The philosophy of the indices construction is based on evaluating the average risk over an orbit. Hence, for the GEO orbit, the satellite needs at least one day to complete its revolution around earth. This is all the more important for the GEO orbit in order to distinguish between dynamical effects or day/night asymmetry effects. Therefore, and for the sake of coherence of the warning system, the daily average was adopted for all the orbits. In our opinion, considering finer time averaging would be relevant to assess the surface charging risk (outside the scope of this paper). In fact, the latter involves low particle energies (outside the range of Salammbô-3D) with dynamic time scales lower than the day.

Section 4: We would like to suggest an additional validation step: could the authors compare the ensemble output of the solar wind propagation to observations at L1? A good match would indeed narrow down the sources of mismatch of the alert indices to observation to the latter part of the pipeline, but possibly the heliospheric part of the pipeline already introduces discrepancies with respect to observations?

- To eliminate the heliospheric source of uncertainties in the pipeline, in this study we have used a synthetic, non-biased solar wind forecast ensemble, generated from OMNI2 hourly data. As this was not discussed in the article, we have added some details in section 4 regarding this ensemble.

l. 185–186: We would suggest a discussion of the computational cost of the pipeline in general and the Salammbô step in particular.

In its operational phase, the pipeline follows an organized timeline, that orchestrates the different computations and data exchange required by its components. Precisely, at 7am (UTC), HELIO-1D output data is ingested in the OGNN neural network to construct solar wind ensembles. Almost 3 hours are required to build the inner-magnetosphere ensemble components (of size 210). This segment involves not less than 4 codes (VLF code, SPM, EMERALD, FARWEST) and ends with the provision of hourly-updated wave-particle interaction and radial diffusion rates. By 10am (UTC), the Salammbô-EnKF simulation has already begun with a hindcast of the previous day, a nowcast of the current day and the forecast of the two following days, hence a 4-day data assimilation simulation that lasts around 2 hours. Finally, the post processing segment with the computation of the indices that are available every day by noon. We decided to add in the new version of the manuscript, a scheme that represents the timeline described above with the principal timecodes associated to the major steps/computations in the inner-magnetosphere pipeline. The scheme should give enough indication on the computation cost of the pipeline and its components.

Firstly, we feel that the present study would be a lot more convincing if the indices matched better, we are reluctant to call the results "fine" as presented. If indeed the critical factor is the low resolution of the Salammbô grid, we would suggest to show the same analysis run with a good grid and accordingly better matches between forecast/nowcast and observed indices. Secondly, how quickly can the code be run on what kind of architecture? What elements can be e.g. parallelised to obtain a better performance? What would it cost in terms of required computing to run the Salammbô step with a better grid operationally? As this manuscript presents a tool with expected operational applicability, we would appreciate seeing some quantitative estimations of the requirements for operational deployment, in particular regarding the grid resolution of Salammbô pointed out by the authors.

- At first glance, the use of the word "fine" may seem unrealistic, as quantitatively, the indices estimated by the SafeSpace pipeline barely reproduce the behavior of the observed indices if not worse. However, knowing the complexity of the radiation belt modelling, the results are qualitatively encouraging in our opinion and show the importance of uncertainty propagation in the improvement of radiation belt nowcast and forecast. Precisely, for the GEO and MEO indices, the system was able to detect a non-negligible number of true alarms or at least overestimate them (raise an active time alarm during a moderate activity phase). From a risk assessment point of view, this is adequate as it shows that the system is globally conservative. To avoid any confusion, we will consider in the new version of the manuscript, the use of more suitable words to portray the performances of the alarm system. Second, the issue with the LEO orbit index is multi-factorial. This region portrays high gradient distributions that are barely captured by the simulation, which also prevents a robust assimilation of in-situ data. We propose to consider a more refined grid in this region to overcome the issue. However this operation will inevitably degrade the operability of the pipeline or even make it obsolete for the presented application. With the current grid refinement, the already parallelized FARWEST and Salammbô-EnKF segments both require at least 2 hours. Refining the grid would not only inflate the wave particle interaction segment (VLF code, SPM and FARWEST) but also impose an even more restrictive time step choice on Salammbô (CFL stability condition). Thus, a trade-off had to be made. To be rigorous, the treatment of this issue requires the integration of a robust and adequate numerical solver that can withstand rapidly varying strong gradients. This improvement has been studied lately by the Salammbô team and led to the prototyping of a new finite volume implicit solver [2], that will be presented in future studies.

Minor comments

l. 71–74: What CME model is used in the propagation modelling?

l. 73: What is the source of the magnetograms used?

- As explained above, we use a synthetic solar wind forecast derived from the OMNI2 dataset, so no CME model is used in this study. In the operational case, the Multi-VP coronal model in EUPHORIA ingests magnetogram inputs from WSO (Wilcox Space Observatory) and from GONG (Global Oscillation Network Group) as reported in [3].

l. 114–116: Can the authors develop a little this point (here or elsewhere in the text): what distinguishes the proposed indices from already-existing three-colour warning levels or other such indices?

- The SafeSpace warning system provides several new features such as:

- Its dependence to the uncertainty propagation from the sun-to-belts pipeline.
- The definition of the alarm thresholds on the distribution of events and their occurrences (20%, 2%), rather than on exposure limits of materials or components.
- Those same alarm thresholds were fixed in partnership with industry stakeholders, based on their feedback. They can always be customized to their satellite or equipment sensitivity.

Technical details

l. 85: Could the authors introduce the variables used as they may not be familiar to all readers?

- The variables will be introduced in the new version of the manuscript.

Figure 4 has poor resolution, and the third panel's horizontal axis line is missing. It would be more useful to have integer tick marks and values for the last panel with the Kp index.

- We will use a higher quality version of the figure.

l. 187: as regards the forecast results

l. 190: beforehand -> before

Table 2 and throughout: certainly the authors mean 10th-90th percentile, not decile.

Figure 5: median

l. 210: no ``of''

l. 218: sensibility -> sensitivity

- The Typos referred above will be amended in the new version of the manuscript.

References

[1] MacNeice, P., et al. "Assessing the quality of models of the ambient solar wind." *Space Weather* 16.11 (2018): 1644-1667.

[2] Dahmen, Nour, François Rogier, and Vincent Maget. "On the modelling of highly anisotropic diffusion for electron radiation belt dynamic codes." *Computer Physics Communications* 254 (2020): 107342.

[3] Samara, Evangelia, et al. "Implementing the MULTI-VP coronal model in EUHFORIA: Test case results and comparisons with the WSA coronal model." *Astronomy & Astrophysics* 648 (2021): A35.