Geoscientific Instrumentation Methods and Data Systems Discussions

Geosci. Instrum. Method. Data Syst. Discuss., referee comment RC1
https://doi.org/10.5194/gi-2021-17-RC1, 2021

Comment on gi-2021-17

Anonymous Referee #1

Referee comment on "Swarm Langmuir probes' data quality validation and future improvements" by Filomena Catapano et al., Geosci. Instrum. Method. Data Syst. Discuss., https://doi.org/10.5194/gi-2021-17-RC1, 2021

Review of gi-2021-1

General comments

This paper presents an interesting discussion about the Swarm Langmuir Probe data quality monitoring efforts. The paper offers a few strong points that make it worthwhile to report to the community

- It proposes two modifications from the 04 to 05 baseline. These modifications, in itself, do not represent much, yet they improve data quality and availability. As I explain further down in my review, I think the major modification would be much stronger if it would be corroborated by a physical understanding of the issue rather than only be an empirical improvement.
- A comparison with the TII front plate data suggests that the night side data products are actually not so good. While I personally think this is the most relevant finding reported in this paper, I have the impression that the authors do not sufficiently emphasize its importance.

The scientific basis of the paper is solid, although I want to raise a few questions (see my specific comments below).

The overall structure of the manuscript is logical, with a few exceptions as indicated below. I recommend the authors to introduce clear definitions of "data quality", "data availability", etc.. Some of these concepts are mixed up in the text and that does not help the clarity of the presentation. The paper suffers from many language and typographical errors; I list some of them at the end of this review without any attempt at being complete.

In conclusion, I do have a substantial number of comments and questions. Addressing these issues would likely make the manuscript suited for publication.

Specific comments

In the introduction the authors list some of the most noteworthy of Swarm's accomplishments up to now. This does not really prepare the reader for the contribution that is presented in the manuscript, so I would consider this as largely superfluous material; the reference to the Swarm publication list would be sufficient – unless the authors would indicate, for the cited contributions, what role the Langmuir probes have played in these accomplishments. That would then raise the reader's awareness that these data have been proven to be of considerable scientific importance and thus that quality control is an absolute necessity.

I was puzzled by the explanation (around line 80) of the difference between low and high gain, even after consultation of Knudsen et al. (2017) – in any case I suggest to add that reference here explicitly. Later on, at the beginning of section 3.1.1, the authors give a more precise explanation. I strongly recommend to move that explanation here; this will considerably improve the readability of the text.

The end of the paragraph on line 81 is a surprising one. The authors say that the use of TiN was questioned and that therefore one of the probes received an Au coating. But then they end the paragraph with a statement about the uncertainty regarding the preservation of the Au coating. I would expect the final statement to address whether the TiN coating is still preserved, because that was the initial concern.

I am curious to know how geomagnetic indices are used to process the LP data (line 94). Could the authors expand on this?

On line 100, the authors state that interpolating the 2 Hz data product at exact UTC leads to the 1 Hz sampled data. Is linear interpolation used? To what extent is linear

interpolation justified as the signal likely contains faster time variations? I would expect that one would first perform a smoothing or filtering in order to remove time variability faster than 1 s before doing such an interpolation. Does one interpolate the densities and temperatures or their logarithms – which is more appropriate for quantities that vary over orders of magnitudes and that are always strictly positive? One could also interpolate the measured currents, and then do the processing with those interpolated values ...

Section 3.1.1 gives a clear explanation of the old algorithm and of the new one. It presents an evaluation of the changes and indicates that this is an improvement because of the better correspondence with ground-based measurements. However, I miss a fundamental point: Why do the low and high gain measurements differ? As the authors have indicated, the difference is a resistor in the measurement circuit. Is there a physical underpinning of the measured difference? I believe significant effort should be invested into this. After all, this change in electron temperature calculation is the main point of improvement in data quality upon which the whole paper is built, so it has to be well documented and justified.

I would welcome a clear definition of "data quality" as it plays such a central role in this paper. My intuitive understanding seems to be at odds with that of the authors. For instance, I read on line 213: "A larger percentage of invalid measurements obviously indicates a poorer data quality." That is not evident to me. One could argue otherwise: "if one obtains a higher percentage of invalid measurements, one is apparently able to catch very well those situations where the measurement process fails, so that one can have more confidence in the remaining data." Indeed, if there is an ADC overflow, that measurement clearly is not reliable, but that does not immediately say anything about the quality of the measurements performed before or after.

The authors discuss the evolution of quality with the solar cycle. How certain are they that an apparent systematic trend with the cycle does not mask detector aging (such as cumulative damage to LP coating, consequences of nanodust impacts, etc.)? Wouldn't one need at least a complete solar cycle to evaluate this? The topic of detector aging is only briefly touched upon. I think it deserves more attention, as this would be one of the major aspects of instrument quality monitoring. Such a discussion could be part of a more extended discussion section (which is rather short at present).

A very interesting point is the comparison of LP and FP densities presented in Figs. 6 and 8. On line 235ff the authors say that the FP processor does not need any assumptions regarding ion composition, while the LP processor does. But then line 236 states that measurement differences between both are due to different assumptions – that contradicts the statement that FP uses no assumptions at all. This is an important point, because it suggests that the night side FP density measurements are considerably more reliable than the LP densities there. And consequently, in view of the observed relatively poor correlation, I do not understand the assessment on p. 245 "The Swarm LP measurements are of very good data quality" – shouldn't this be qualified somewhat, e.g. restricted to the day side? The conclusions section does list this problem as one that will drive future attempts for improvement.

Detailed issues

Abstract: The abstract reads well, explaining that the paper discusses the quality control approach. I suggest to add a sentence that states how good the data quality is, before then saying that there is an anomaly. Please also do not use the LP abbreviation in the abstract as it is not explained before.

Throughout the paper there are numerous issues with punctuation – lots of unnecessary or misplaced commas. I have indicated only some of them below.

Line 18: divers -> diverse

Line 21: satellite’ s -> satellite’s (remove blank space)

Line 22: plasma and electric field -> electric field and plasma (so that the order of both corresponds to the names of the sensors mentioned before)

Line 33: reveled -> revealed

Line 41 mission, -> mission (drop comma)

Line 45 fluctuations, -> fluctuation (drop comma)

Line 45: described -> describes

Line 50: largely -> extensively (?)

Line 54: remind -> refer

Line 70: instruments -> instrument’s or instruments’

Line 91: to simple flow-chart -> to the simple flow-chart

Line 95: essentials -> essential

Line 98: remove comma (or have it after “Finally”)

Line 106: expectantly: not sure what is meant by this; I propose to drop this word

Line 108: “This data is” or “These data are”

Line 112: In the next two sessions are described -> The next two sessions describe

Line 116: I do not think that explaining the file naming convention that is used by the team is of any value to the reader; I propose to drop this sentence.

Line 121: baseline lower than 04 are -> “baseline lower than 04 is” or “baselines lower than 04 are”

Line 124: ESA (2018, 2020b) -> (ESA, 2018, 2020b)

Line 126: consisting on -> consisting of

Line 126: the first item you mention is not a “difference”. Replace by “consisting of an updated electron temperature computation” or something of that style.

Line 132: differ which –> differ, which

Line 137: appropriated -> appropriate

Line 145: drawback, -> drawback

Line 147: region -> regions

Line 176: where -> was

Line 178: PLASMA -> the PLASMA

Line 179: occurred -> that occurred

Line 184: impact in -> impact on

Line 186, 193: remove comma

Line 195: sun spot -> sunspot

Line 197: in-orbit -> in orbit

Line 224, 226: remove comma

Line 268: In the next baseline will be introduced -> The next baseline will introduce

Line 282: the calibrating the LP measurements yield -> the calibration of the LP measurements yields

Caption of Fig 3: each one-degree bins in latitude -> each one-degree bin in latitude

Figure 5: I see little value of showing the entire sunspot record. The figure would be much

more clear if all panels would focus on the Swarm mission period.