This interesting study is well developed and executed. The topic of estimating/comparing precipitation/rainfall by different remote sensing-based data sources is quite challenging. My comments mostly are textual where further detailing and description of used terms is advised to increase readability. As an example, *sampling* is widely used but it is not always clear how to read the meaning of the term. E.g. sampling method, sampling pattern, sampling sensitivity, sampling area are used but remain somewhat undefined. Also the term pixels is used but also undefined in terms of size shape. It would be good to add a figure that shows how spatial units of sampling overlays the pixel structure of the SRE products. Also *uncertainty* and *error* are used throughout. It would help to define what these terms actually imply; right now I read a somewhat overlapping interpretations of the terms. What is considered an error, is that a delta? E.G. a mismatch between a reference and a counterpart (or sample estimate? Or from a SRE? Or sensor observation?), and what is considered uncertainty. In case (different) errors counter act that may reduce the overall estimation error (i.e. overall mismatch), so is uncertainty of the estimate reduced then, or increased? I, myself, associate uncertainty with a random character of an error. A suggested, I think there is scope to improve respective text sentences. Findings of the study are relevant and interesting.

I read that the authors use *precipitation* but it is not very clear If precipitation other than rainfall (in liquid phase) is represented (e.g. Line 161 suggests that only rain is targeted). The title suggests precipitation, the opening sentence in the Introduction is on floods and suggests only rainfall, but it should be made more clear what precipitation actually implies as used throughout the manuscript. It is relevant to know if findings also apply to snow and hail, precipitation by frost and rhyme can be discarded.

Abstract

The (long) abstract reads a bit complicated by the large amount of information provided.

Ln 23: I struggled a bit with the conclusion. Uncertainty now becomes an error? should it not be other way around; effects of accumulated/combined errors result in ((un)predictable) uncertainty.

Introduction:

Ln. 26: Opening statement on “uniform” distributions is questionable as, especially, for extreme events we often seek information at highest time and space intervals to best represent real world rainfall forcing. Does the statement not reflect more on use of rainfall data sources, that are obtained at fixed discrete time and space steps/units (I read here “uniformly”).

Lns 32 and 38: Please limit to only 2 (or max 3) most relevant references (unless truly meaningful) (please revisit the manuscript)

Lns 47-54: The GPROF algorithm is introduced and briefly described. It would be meaningful to comment on weaknesses of the algorithm in the discussion of results. The way the algorithm operates has quite some limitations that to some extend effects findings of the study.

Ln 106: Section 2.1.2 . It is not clear why in-situ data was not used directly so to perform a direct *point to (SRE) pixel* comparison. The radar estimates also is a precipitation (or rainfall) product that has errors as described but that effect on the research findings remains now undefined. Actually, a lot of findings revert back to aspects of performance of the KNMKI radar (e.g. line 285), although the radar serves as a

refence to replace in-situ gauge observations. Throughout the manuscript there are descriptions that address performance aspects of the ground radar, and at the same time GPROF, and thus aspects of performance of both sensor products are intercompared. I think this aspect of the methodological approach at the base of the study can be improved. It only takes few sentences.

Ln 106: I indicate line 106 but the comment applies to the manuscript. What also could be further clarified is on how many spatial units the analysis are based. The KNMI ground radar operates at $1km^2$ resolution. The footprints are of different number, size and shape. I found it difficult to read how the comparisons are actually done, e.g. it is not very clear what the sample sizes are at the back of findings in figures. Often descriptions halt by simply stating, *sampling*.

Ln 115: Simply add few words what precipitation height actually implies. Please clarify on these types of terms throughout the paper to increase readability.

Ln 155: Eq. 2 is similar to Eq 1 except errors are taken as absolute. So Eq. 2 also is a relative error but now as absolute RE. A mean value (division by N is required) is not performed, NMAE simply is/becomes ARE (absolute relative error). Also check the units for consistency.

Ln 171: Why to mention explicitly “independent”, is that not trivial?

Ln 182: To relate foot print size to number of observations (this reads like sampling frequency) to threshold is quite speculative I would say. To write it differently “Is there any prove that threshold has relation to the foot print size to affect the number of observations” (Do you mean larger size of the foot print results in less spatial units for analysis)? To work with mean values is not a problem.

Ln 196: Here the “ash-tray” term *sampling* is used *bit I guess what sampling* actually complies in this study needs a better description. We have scale of sampling (the foot print), the frequency of sampling, the way a sample is taken (the sampling method), and the sample size (nr. Of observations). I advise to go through the paper again to read and to further clarify on these issues so to create better understanding. In my opinion there is scope to improve readability. Also, the term observation(s) is frequently used (e.g. annotation Fig 5) but does an observation now also have the meaning of a single sample?

Fig 3: It is not clear how “volume” is defined? Make clear what the meaning of volume is. Also in the writing of the paragraph (Section 3.1).

Fig 3: The legend with respective CDFs is undefined in the text. Explain what the terms *CDF GPROF; CDF 19 GHz; CDF GPRO-own and CDF 19 GHz-own* actually represent to clarify on information by the CDFs.

Fig 4: Not clear why the number of “4” methods is mentioned? The figure shows results of 3 methods applied at two heights.

Line 256: Figure 2 does not show an (comparative) numbers (or reference) to the statement. The word “likely” is used three times re argue for findings but the term actually has a dimension of “speculation”. I’m not in favor of such wording as research aims to create clarify.

Ln 285: This reads difficult. How can ETH influence performance of GPROF? Ground radar is used as a reference.

Kind regards,

Tom Rientjes