Reviewer # 2
General: The paper describes the formulation of a satellite-based precipitation estimation scheme based upon the MSG SEVIRI observations over southern Africa, and provides a comparison of this technique, together with that of the GPM IMERG product against gauge data. As such it is an interesting and useful paper since it covers a region that is often neglected.
My overall recommendation is that the paper is acceptable for publication following (minor/) major revision. The technical issues need to be addressed, in particular the ones relating to the masking of the data in the comparison (masking to just the MSG-identified cloud regions could bias the statistics).

Response
Thank you very much for the comprehensive and detailed revision of our manuscript! In the following we would like to outline our response (green color) to your concerns (red color) as well as the subsequent changes that we made for the final version of the manuscript (green, italic).

I would point the authors to the work of the International Precipitation Working Group team working on the South Africa data, also using gauge data to inter-compare daily precipitation products.

Response
Yes, the IPWG does a lot of work on comparing different rainfall products for South Africa. We intend an incorporation of the presented retrieval to the IPWG validation study for future assessment and we are sure this would bring further insights into the strengths and weaknesses of the retrieval technique.

Key issues: i) Need to check the gauge data. First, ensure that the quality control is optimal, e.g. do some gauges never report rainfall? Do gauges distinguish between 'zero' and no-data? It is possible, once you have the satellite estimates, to check the performance of individual gauges – are there individual gauges that always are 'incorrect' compared to the satellite data? It would be unlikely that the satellite product would be consistently wrong over a particular gauge if it is correct over a neighbouring gauge.

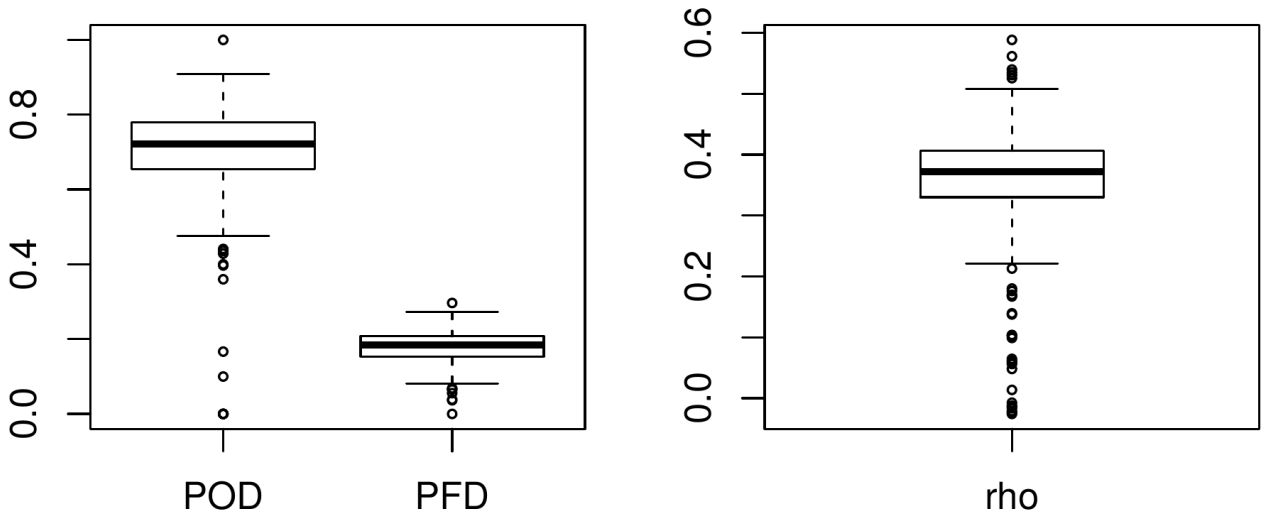
Response
We totally agree that the quality of the ground truth data is an important issue! Yes, the data distinguish between zero and "no data" otherwise it would not be possible to train a model for rainy and non rainy clouds. All gauges that were used in this study provided rainfall for the training/testing period, however, with gaps. We added the following point to the discussion:

"Also, due to different installation dates of the individual weather stations as well as the natural challenge of maintaining weather stations in remote areas, no gapless dataset could be compiled. Therefore, different sensor and data provider dependent calibration techniques, gaps in the time series of the data as well as the general problems associated with rain gauge measurements might lead to inconsistencies and uncertainties. However, no reliable alternatives are available and rain gauge measurements are still considered as most reliable source of rainfall data. "

We further included information about the pre-processing of the data:

"The data passed general provider-dependent quality checks before it was used in this study. This includes filtering of data beyond common data ranges, or situational checks for consistency with related parameters (e.g. air humidity) by SASSCAL. SAWS payed attention to rainfall values > 10 mm within 5 minutes and deleted those values if unreliable. Data from all providers was then included in an on-demand processing database system (Wöllauer et al., 2015) where it was automatically cross-checked for reliability by filtering values <0 and >500 mm rainfall per hour. All station data that provided sub-hourly information was aggregated to a temporal resolution of 1 hour within the database."

We checked the results for station-dependent errors (see Figure below) and there were only few stations where the model constantly showed a low performance. We could identify 9 outliers that had a POD < 0.3 and a rho <0.05. The reason for this were large data gaps for the validation period so that the resulting statistics on a station basis are not meaningful for these stations. Only two stations showed abnormalities in the data that are probably associated with errors as they constantly featured very low rainfall values that were not at all in accordance with neighboring stations. These problems were not captured by the quality check methods. Therefore, it would be reasonable to remove such stations from the analysis. However, the problem is not critical for the results of this study. Since only very few stations (and very few data points in total) were affected, these data must be regarded as extreme outliers that have negligible effects.



---

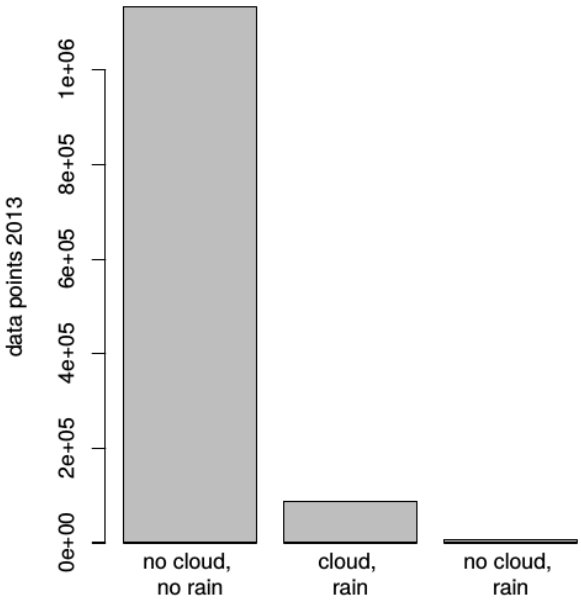
ii) The use of the cloud mask in the statistical analysis (page xxx) removes regions where the gauge might report rainfall, but the satellite does not, thus, it biases the analysis.

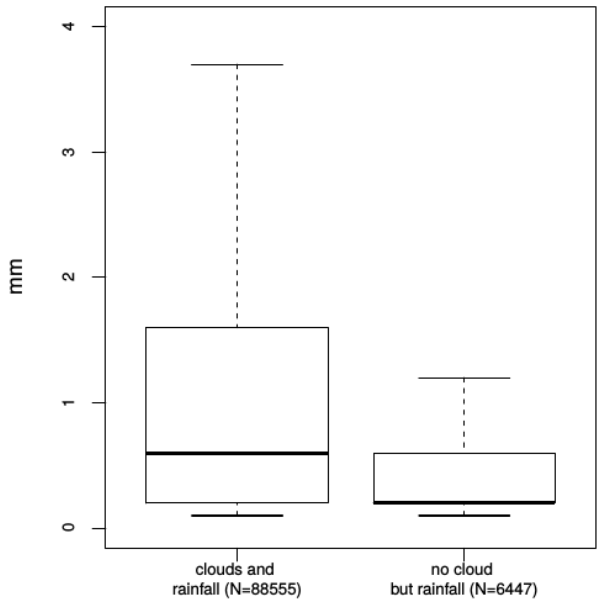
Response

That is true! Our model relies on the cloud mask product that makes an initial selection of areas that come into question for rainfall. Honestly, we didn't think about this point as a potential source of error because we assume that rain clouds are easy to be captured as clouds by a cloud masking algorithm. To get an idea about the bias we visualized the fraction of data points for the year 2013 where this problem occurred: From 2108958 data points in total, roughly half of the data had no clouds and were not raining (1133885). 880071 data points were cloudy but it was not raining and 88555 data points were cloudy and it rained. It total, 6447 data points were not cloudy but it rained thus have to be regarded as problematic due to the cloud mask as initial selection (Figure A). This fraction is comparably small and if we compare the measured rainfall of those data points that were masked as cloudy with those that were not masked as cloudy (Figure B), the problematic data points had significantly lower measured rainfall, thus the problem luckily only slightly contributed to the rainfall totals.

Without going into detail with this analysis in the manuscript, we now accounted for this issue in the discussion section:

"The retrieval techniques relied on the cloud mask for an initial selection of relevant data points used for model training, validation and the final spatio-temporal estimates. Therefore, it can't be excluded that some data points were falsely excluded from the analysis as they were falsely masked as being not cloudy but rainfall was measured on the ground. However, we assume that rainy clouds are easy to capture by common cloud masking algorithms and that the resulting bias is therefore comparably small."





---

iii) Although it is mentioned that the IMERG product is aggregated from the 30 minute product resolution to a 1-hour resolution, I could not find how the 15-minute MSG observations are aggregated into hourly estimates. Also, the authors should be careful with the time stamp of the products – do these relate to the start or end time (UTC) of the product? Also, is the gauge data in UTC or local time?

Response
We agree that this information was missing in the manuscript. We now added the following information: "*To match the temporal resolution of all available rain gauge data, the extracted data were aggregated to hourly values. This was done by taking the median value of the four scenes available every hour. However, only if all four scenes were masked as cloudy, the corresponding hourly values for a respective station were used for further analysis. The extracted and aggregated MSG data were then matched with the corresponding rain gauge information under consideration of the time shift between MSG data (UTC) and rain gauge data (UTC + 2)*".

---

iv) (P6, first paragraph) Since there is a daytime and a nighttime 'algorithm', how do the two compare? In particular, since (presumably) the nighttime algorithm can be used both night and day, it could be used to assess the differences in performance. This is somewhat critical since a smooth transition in rainfall estimates between day and night is clearly desirable. Also, how do you define 'day' and 'night'?

Response
We now added the information about how the data were split into day and night:
"Since the VIS and NIR channels of MSG are not available during the nighttime, the dataset was split into a daytime dataset (scenes with a solar zenith angle < 70°) and a nighttime dataset (scenes with a solar zenith angle > 70°)"

We also added a section to describe how the spatial model estimates were created. Within one MSG SEVIRI scene, the model is used consistently as the mean solar zenith angle from the entire scene was used as

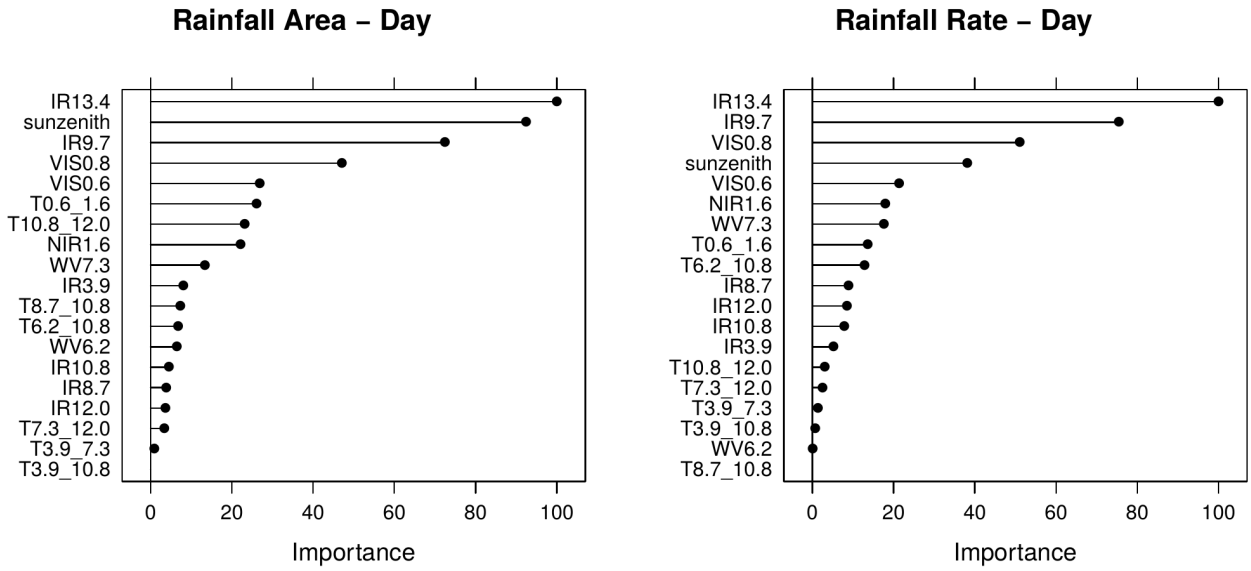
decisive angle:

"Final models were applied to all hourly MSG SEVIRI scenes from 2010-2014 for the Southern Africa extent to obtain spatio-temporal estimates of rainfall. Therefore, the clouded areas of a scene were first classified into rainy or not rainy using the respective model. The rainfall quantities were then estimated for the estimated rainfall areas. To ensure consistency within one scene, the choice of the model being applied (either the daytime or nighttime model) was made according to the mean solar zenith angle of the respective scene. If the mean solar zenith angle was <70°, rainfall for the entire scene was estimated using the daytime model. For scenes with a mean solar zenith angle > 70°, the nighttime model was applied."

We agree, the nighttime algorithm can be used for both, however, the VIS channels can increase the performance of the daytime models. In this context, we rated high performance for the daytime data being more important than smooth transitions. We added a short comment on that issue in the manuscript:

"Though two different models might lead to rough transitions between daytime and nighttime estimates, accurate estimates were in the foreground of this study, leading to the decision of separate models according to data availability."

The importance of the VIS channels is confirmed in by importance of the variables within the models. Though the importance of the variables within a neural network can only roughly be estimated, we certainly would yield a lower performance if the VIS information were not included.



General Technical issues: Check use of capitals for acronyms, e.g. P1, L3: 'Spinning Enhanced Visible and InfraRed Imager (SEVIRI)' Check the consistency of capitals, e.g. P1, L6/7: '. . .(Probability of Detection, POD). However the False Alarm Ratio (FAR). . .'. Check use of acronyms: The general rule is, define all acronyms on first usage, after this only use the acronym (usually following on after the abstract). Only use an acronym if used more than once – and only if it is a commonly-used acronym (i.e. don't make up acronyms).

Response

We now defined all acronyms in the abstract (if they are used there) and then on the first appearance in the main text. We also made sure that the capitals are used consistently.

Specific Technical issues: P1, L1: consider ‘necessary’ instead of ‘highly required’ P1, L3 (and elsewhere): use of capitals for acronyms – ‘Spinning Enhanced Visible and InfraRed Imager (SEVIRI)’ P1, L4: remove ‘for years’ and replace ‘truths’ with ‘truth’ P1, L5: replace ‘predicting’ with ‘the estimation of’, and replace ‘during’ with ‘over’ P1, L6/7: ‘. . .(Probability of Detection, POD). However the False Alarm Ratio (FAR) . .’.P1, L10: Define ‘IMERG’ P1, L16: replace ‘on a’ with ‘at’ and replace ‘resolution’ with ‘resolutions’ P1, L20: replace ‘An accurate’ with just ‘Accurate’ P1, L21: replace ‘in’ with ‘at’ and ‘resolution’ with ‘resolutions’

Response
We made all the suggested changes.

P2, L5: replace ‘for entire’ with ‘covering the entire region of’ P2, L11: replace ‘resolution’ with ‘resolutions’ and ‘in’ with ‘at’ P2, L12: replace ‘can’ with ‘might’; insert ‘would’ after ‘products’; insert ‘degree of’ before ‘accuracy’ and replace ‘as’ with ‘since’ P2, L16: replace ‘;’ with ‘and’; capitals for ‘Meteosat Second Generation’ and ‘Spinning Enhanced Visible and InfraRed Imager’ P2, L19: should ‘South Africa’ be ‘southern Africa’ (middle and end of line)? P2, L27: replace ‘prediction’ with ‘estimation’ P2, L30: replace ‘yearly’ with ‘annual’, remove ‘sums’ and replace ‘follow’ with ‘follows’. P2, L32: replace ‘rains’ with ‘rain’ P4, L1: replace ‘sums’ with ‘totals’ P4, L2: replace ‘;’ with ‘and’. P4, L5: remove ‘the years’ P4, L6: replace ‘from’ with ‘at’ P4, L7: remove ‘the year’

Response
Thank you, all changed.

P5, L4: The 3 x 3 km resolution is the IR resolution; i) the visible channels are about 1 x 1 km, but ii) the resolution over southern Africa for both the IR and visible channels is of course, poorer. P5, L9/10: the last sentence is gobbledygook: ‘xx1 technology’ if you google it, is to do with cycling, and the link to the web-page provided does not exist.

Response
Only the high resolution visible channel has a spatial resolution of 1 x 1 km. The “normal” visible channels still have a resolution of 3 x 3 km. We now accounted for the approx. resolution in southern Africa in our revised manuscript and made sure that we didn't use the high resolution channel:

“MSG SEVIRI (Aminou et al. 1997) scans the full disk every 15 minutes with a spatial resolution of 3 x 3 km at sub-satellite point (~ 3.5 x 3.5 km in Southern Africa). Reflected and emitted radiances are measured by 12 channels, three channels at visible and very near infrared wavelengths (between 0.6 and 1.6 µm), eight channels ranging from near-infrared to thermal infrared wavelengths (between 3.9 and 14 µm) and one high-resolution visible channel with a spatial resolution of 1 x 1 km which was not considered in this study.“

We added more explanation about the processing scheme:
“MSG SEVIRI Level 1.5 data (EUMETSAT 2010) were preprocessed to radiance values according to EUMETSAT (2012a) and brightness temperatures according to EUMETSAT (2012b) using a processing scheme based on a custom raster processing extension of the eXtensible and fleXible Java library (see https://github.com/umr-dbs/xxl) which enables parallel raster processing on CPUs and GPUs using OpenCL.”

However, we cannot understand your concern about the link to this web-page as we could reach this page via the link provided.

Reword/revise. P5, L11: remove ‘the years’ P5, L17: consider ‘excluded’ rather than ‘masked’ P5, L21:

replace ‘predict’ with ‘retrieve’ P5, L25: replace ‘many confusions’ with ‘much confusion’

Response
changed

P5, L30: If all the channels are included in the NN, surely any channel differences should also considered within the NN without having to include them as separate entities?

Response
We agree since this is a point that we also discussed quite a bit. The predictor variables contains duplicated information in some way. However, we highly assume that these combinations are able to highlight patterns that are not obvious when only the individual channels are used. We found a significant increase of performance when the channel combinations were included which supports our assumption. Also, the neural networks are robust to duplicated information (at least considering comparably small numbers of predictor variables as we did in this study.), so that including this information is not of disadvantage for the model but allows taking advantage of highlighted patterns.

We now justified our decision in the manuscript:

*“Thus, the predictor variables contain the SEVIRI channels as well as channel combinations. Although this partially duplicates information, the channel combinations allow highlighting patterns that might not be apparent in the individual chan*nels.”

P6, L7: replace ‘two-folded’ with ‘two-step’ (?) P6, last paragraph: see above regarding use of cloud mask, acronyms, use of capitals. P6, L34: the HSS can be bias-dependent since if all retrievals are zero and surface data non-zero, it will be dependent.

Response
We changed everything accordingly.

P7, L2: By ‘Spearmans’ I presume you mean the ‘Spearman’s Product Moment Correlation’; suugest rewording ‘Spearmans rho’ to ‘Spearman’s Product Moment Correlation (rho)’ (or use the greek letter ‘rho’) P7, L2/3: replace ‘Further the root mean square error (RMSE) was used’ with ‘The root mean square error (RMSE) was also calculated’. P7, L3: replace ‘clouded’ with ‘cloudy’ P7, L8: replace ‘aiming at’ with ‘de-signed for’ P7, L9: The reference to ‘Smith et al., 2007’ is somewhat antiquated: use ‘Hou et al., 2014 and Skofronick-Jackson et al., 2017.’ (Full references below) P7, L10: replace ‘instruments’ with ‘estimates’

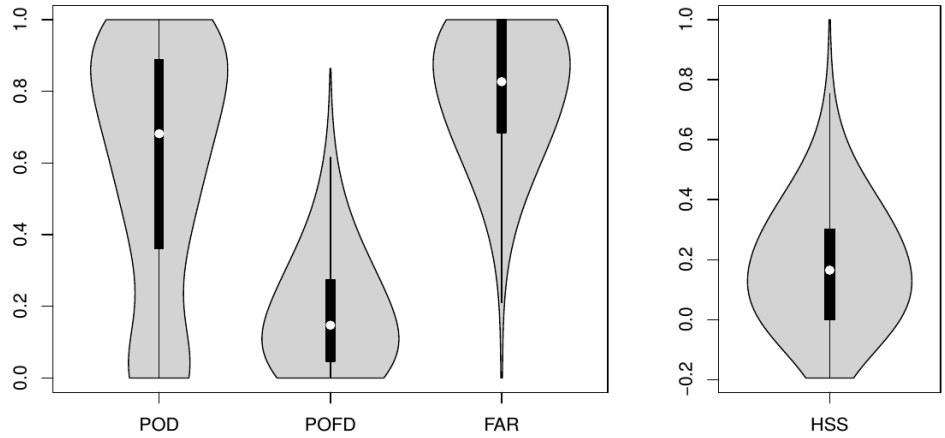
Response
We replaced the references and made all other changes as suggested

---

P8, L3: The initial sentence here is not evident from Figure 3. (see comments below about the box-plots). P8, L4: replace ‘predictions’ with ‘estimates’ P8, L5: presumably the ‘0.72 mm 4’ should be ‘0.72 mmh-1’ (use journal style for mm/hr) P8, L5: replace ‘in’ with ‘on’ P8, L6: reword ‘rainfall quantities assignment’ (I don’t know what is meant by this). P8, L7/8: replace ‘quantities could be’ with ‘is’ P8, L8: replace ‘rainfall sums’ with ‘totals’ and ‘predictions’ with ‘estimates’ P8, L9: replace ‘are show for the year 2013’ with ‘for 2013 are shown’ P8, L30: replace ‘Manhique et al. (2015).’ with ‘(Manhique et al., 2015).’

Response

We made all required changes and the boxplots are changed to violinplots.



P9, L1: replace ‘retrieval’ with ‘retrievals’ and replace ‘highlights also’ with ‘also highlights’ P9, L3: remove ‘to elevated levels’ P9, L5: parallax shifts would generally be < 1 pixel at this region.

Response
Changed. We added the information that the parallax shift is rather small.

P10, L3: move comma from after ‘pixel’ to after ‘problematic’ P10, L8: replace ‘Kidd and Huffman (2001)’ with ‘(Kidd and Huffman, 2011)’ P10, L8/9: see comment above about checking gauge data. P10, L15: remove ‘view to’ P10, L16: replace ‘GMP’ with ‘GPM’
Response
Changed

P11, L2: Insert ‘scheme’ after ‘retrieval’ P11, L5: Insert ‘technique’ after ‘retrieval’ and replace ‘in’ with ‘at’

Response
Changed

P12, L1: ‘overestimation of rainfall areas’ – care is needed here – is there an over-estimation of ‘rain area’ or ‘rain occurrence’ (these are different, but linked). P12, L2: remove ‘global’; remove ‘assignment; replace ‘even advantageous’ with ‘better’ P12, L6: replace ‘are’ with’is’

Response
We changed the wording to rain occurrence and made the other suggested changes

---

References: Include data set references (most data sets now have doi’s – and the GPM ones certainly do so).

Response
SASSCAL and SAWS weather station data as well as the cloud mask products were all cited and

acknowledged in personal consultation with the providers and have no doi's. However, we now included more appropriate citation of GPM and MSG SEVIRI.

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., and Xie, P.: GPM L3 IMERG Late Half Hourly 0.1 degree x 0.1 degree Precipitation V03, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC). Accessed 15 June, 2015, doi:10.5067/GPM/IMERG/HH/3B, ftp://gpm1.gesdisc.eosdis.nasa.gov/data/s4pa/GPM_L3/GPM_3IMERGHH.03/, 2014.

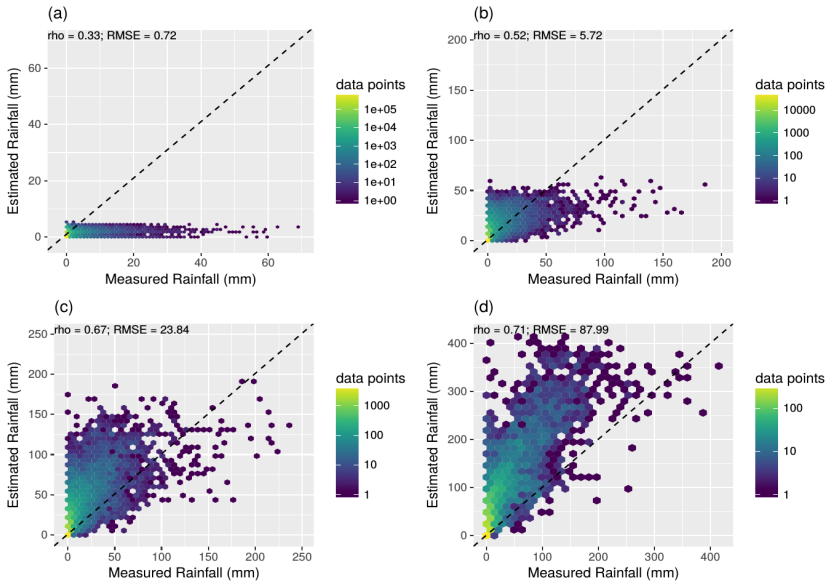
EUMETSAT: High Rate SEVIRI Level 1.5 Image Data - MSG - 0 degree, http://navigator.eumetsat.int/discovery/Start/DirectSearch/DetailResult.do?f%28r0%29=EO:EUM:DAT:MSG:HRSEVIRI, 2010.

Captions/Figures Figure 2: replace ‘yearly’ with ‘annual’; Figure 3: replace ‘predicted’ with ‘estimated’.

Response
We changed both figure captions

Figure 5: replace ‘predicted’ with ‘estimated’; remove ‘the year’; replace ‘on’ with ‘at’; remove ‘and on. . .levels’. Also, the colours seem to be smeared – particularly in (d) where each green point appears to be surrounded by a yellow ‘ring’.

Response
We improved the figure by providing a clearer binning of the values and a comprehensive color scheme with a legend showing the amount of data points.



Figures 3,4,7 & 8: The box plots are not terribly good at conveying the necessary information. It would be much more valuable to display these a ‘violin’ plots (see Figure 5 of *http://dx.doi.org/10.1175/JHM-D-16-0079.1*)

Response

Agreed. To take additional advantage of the data density, we changed the figure style to violin plots.

Figure 9: would be good to include the gauge locations. Also, note that the MSG-estimate is a daytime retrieval scheme.

Response
We included the note that the daytime scheme was used.
Including the location of the station turned out to be not really helpful: The relevant information is covered by the location of the stations, which is in our opinion, however, not very important since the location of all stations is already shown in Figure 1 and the location of the stations that recorded rainfall is shown in section d of this figure.

References: Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurements Mission. Bull. Amer. Meteor. Soc., 95, 701-722, doi:10.1175/BAMS-D-13-00164.1. Skofronick-Jackson, G., and Coauthors, 2017: The Global Precipitation Measurement (GPM) Mission for Science and Society. Bull. Amer. Meteor. Soc., doi:10.1175/BAMS-D-15-00306.1, in press.

Response
Thank you, both references are now included
