Philipps Universität Marburg

Faculty of Geography

Hanna Meyer

Tel.: +49-6421-28-25954
E-Mail: hanna.meyer@geo.uni-marburg.de
Address: Deutschhausstraße 12
35032 Marburg
Web: http://www.uni-marburg.de/fb19/

Marburg, 01 February 2017

Dear editor,

First of all, we would like to thank you and the reviewers for their very thoughtful and comprehensive comments on our manuscript. Based on these valuable recommendations, we herewith submit a revised version. We have considered the reviewers recommendations for revisions and made the necessary changes. On the following pages, we would like to outline a point-by-point response of how we have addressed each concern.
We hope our revision meets your expectations and we are looking forward to hearing from you in due course.

With best regards
Hanna Meyer

Reviewer # 1
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.

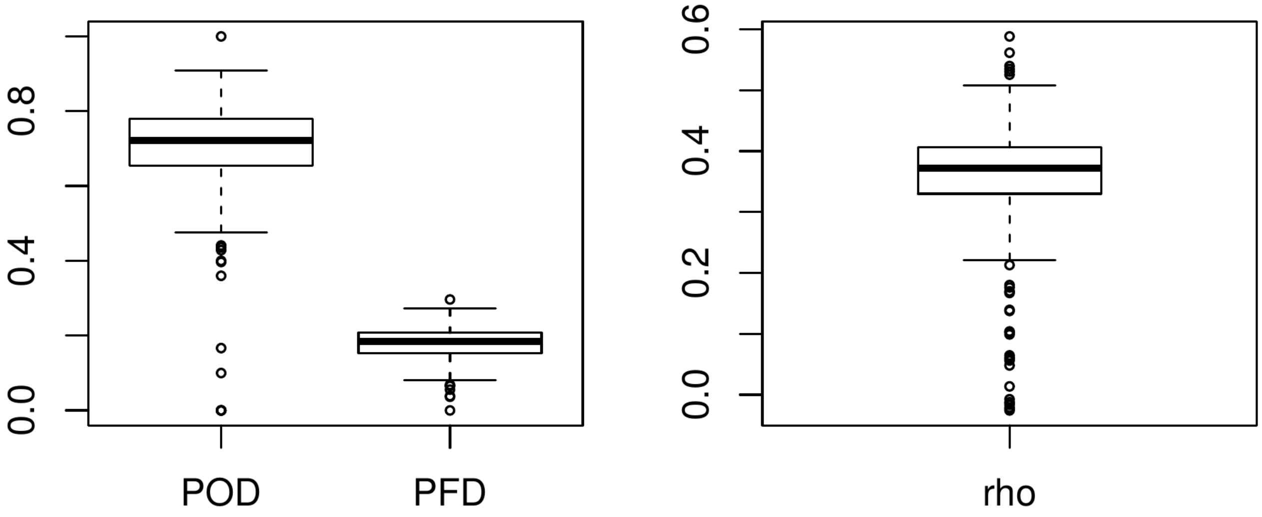
Response
We totally agree that the quality of the ground truth data is an important issue! Yes, the data distinguish between zero and "no data" otherwise it would not be possible to train a model for rainy and non rainy clouds. All gauges that were used in this study provided rainfall for the training/testing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



---

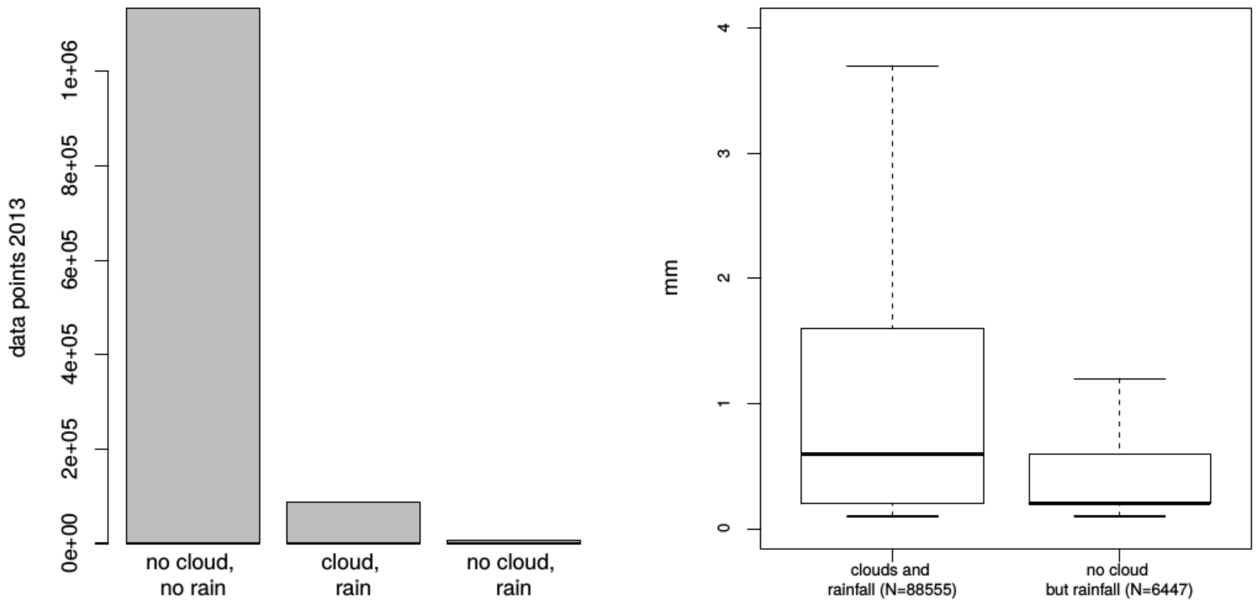
ii) The use of the cloud mask in the statistical analysis (page xxx) removes regions where the gauge might report rainfall, but the satellite does not, thus, it biases the analysis.

<u>Response</u>

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?

<u>Response</u>
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'?

<u>Response</u>
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°)"

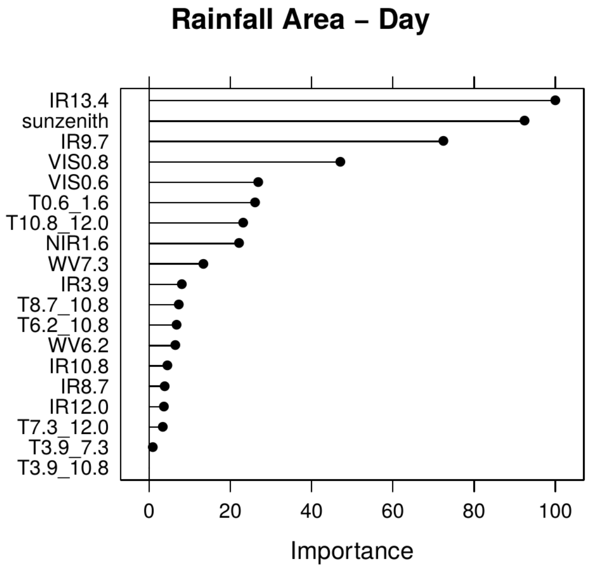
We also added a section to describe how the spatial model estimates were created. Within one MSG SEVIRI scene, the model is used consistently as the mean solar zenith angle from the entire scene was used as decisive angle:

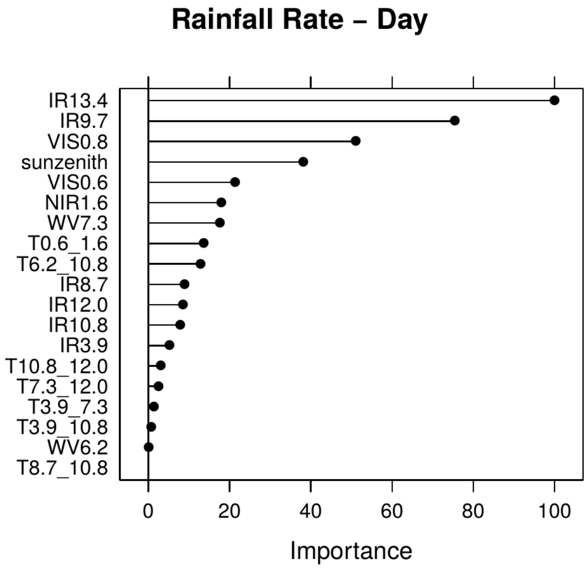
“Final models were applied to all hourly MSG SEVIRI scenes from 2010-2014 for the Southern Africa extent to obtain spatio-temporal estimates of rainfall. Therefore, the clouded areas of a scene were first classified into rainy or not rainy using the respective model. The rainfall quantities were then 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).

<u>Response</u>
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 ‘designed 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’

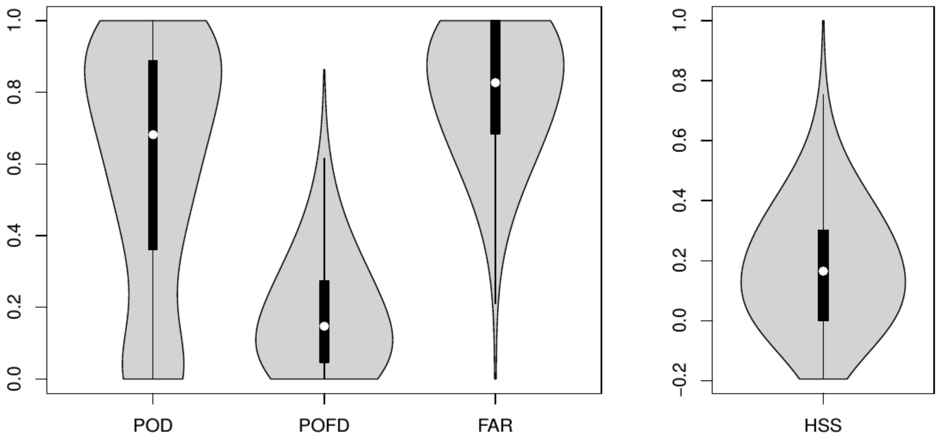
Response
We replaced the references and made all other changes as suggested

---

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.

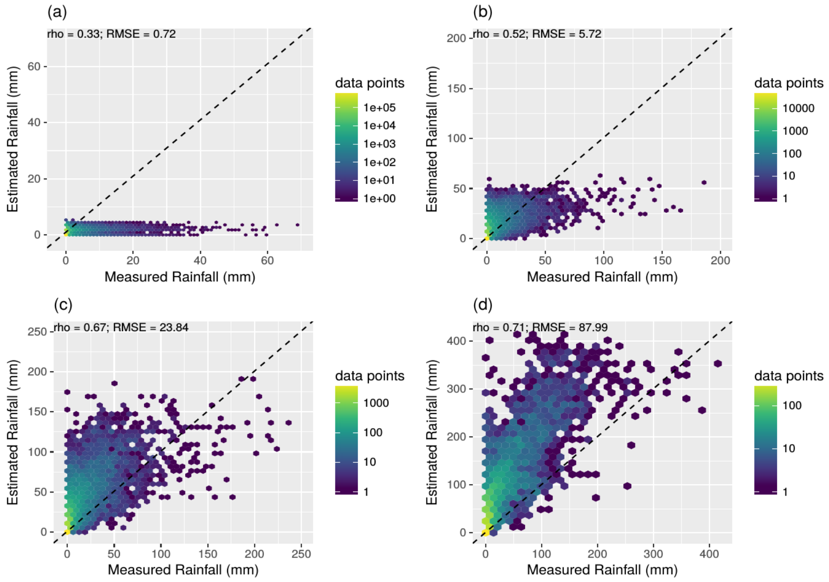
EUMETSAT: High Rate SEVIRI Level 1.5 Image Data - MSG - 0 degree, http://navigator.eumetsat.int/discovery/Start/DirectSearch/DetailResult.do?f%28r0%29=EO:EUM:DAT:MSG:HRSEVIRI, 2010.

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

<u>Response</u>
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'.

<u>Response</u>
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)

<u>Response</u>

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

Reviewer # 2

The paper describes a method for estimating hourly rainfall over Southern Africa, based on a neural network approach, using MSG SEVIRI observations for the estimation and rain gauges data as ground truth. The results are compared to those obtained from IMERG of the Global Precipitation Measurement mission. The paper is interesting because it addresses an important and complex issue, as is the estimate of the surface precipitation in a region (the African continent) with sparse rain gauge and radar networks. I would like to recommend that this paper could be published after major/minor revisions to address the following comments.

Response

Thank you very much for taking your time on our manuscript and your helpful comments! 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).

Major revisions:

1 – The description of some important aspects of the study is often done in a concise, not sufficiently complete and precise way to allow a direct and complete understanding. This fact is partly due to the use of some too general references (e.g. a conference (P2 L6 : IPWG, 2016) or books (P6, L13 : Venables and Ripley, 2002) or (P6, L17 : Kuhn and Johnson, 2013)), where more precise/accurate references (the paper in the conference or the section/pages in the books) would facilitate the understanding of the specific topics. In part it is due to the use of references that seem irrelevant/inconsistent with the text (P5, L8-9 : xxl technology .... OpenCL acceleration (see https://github.com/umr-dbs/xxl)). In part it is due to the use of specialized terms generally difficult to understand/interpret (P6, L15 : stratified 10-fold cross-validation). More attention to the aspects mentioned and a clearer description of the different topics would make it easier to read the text and would better highlight the most innovative aspects of the study.

Response

We changed/added the following information towards a comprehensive description of the applied methods:

1) References: IPWG refers to a website which is THE reference to compare different rainfall retrievals for South Africa (see also comment from Reviewer 1) and must be mentioned here. Venables and Ripley, 2002 refers to the software implementation being used. The nnet package is supposed to be cited in this way (see https://cran.r-project.org/web/packages/nnet/citation.html), though we agree it's a very general reference. We therefore added the direct reference to the software package:
Ripley, B. and Venables, W.: nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models, http://CRAN.R-project.org/package=nnet, r package version 7.3-12, 2016.

For Kuhn and Johnson we adapted the chapter and pages.

2) According to your suggestions we expanded the description of the input variables being used (see comment 2), the preprocessing of the satellite data, the architecture of the neural network (see comment 2ii) and the cross validation.

About the preprocessing of the data: *"MSG SEVIRI Level 1.5 data (EUMETSAT 2010) were preprocessed to radiance values according to EUMETSAT (2012a) and BBT values 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."*

About stratified 10-fold cross validation: *"Thus, the training samples were randomly partitioned into 10 equally sized folds with respect to the distribution of the response variable (i.e., raining cloud pixels, rainfall rate). Thus, every fold is a subset (1/10) of the training samples and has the same distribution of the response variable as the total set of training samples. Models were then fitted by repeatedly leaving out one of the folds. The performance of a model was then determined by predicting on the held-back fold. The performance metrics from the hold-out iterations were averaged to the overall model performance for the respective set of tuning values. For the rainfall areas classification models, the distance to a "perfect model", based on Receiver Operating Characteristics (ROC) analysis (see cite{Meyer2016} for its application in rainfall retrievals) was used as decisive performance metric. For the rainfall quantities regression models, the Root Mean Square Error (RMSE) was used."*

2- Since the neural network is a key point in the study, more clarification on its design and its architecture would be appropriate. The references to texts (e.g. P6, L17 : Kuhn and Johnson, 2013) or packages (P6, L13 : "nnet" package (Venables and Ripley, 2002); P6, L14 : "caret" package Wing et al (2016)) do not lead to a direct understanding of the actual network used. The following points should be clarified:

i) How the network input variables were selected (P5, L30 and P6, L1-2). The reference P6, L1 : Meyer et al. (submitted) is not available.

<u>Response</u>
Meyer et al. (submitted) was in review when this manuscript was submitted. It is now published in Remote Sensing Letters so we could include the correct reference:

Meyer, H.; Kühnlein, M.; Reudenbach, C. & Nauss, T.: Revealing the potential of spectral and textural predictor variables in a neural network-based rainfall retrieval technique. *Remote Sensing Letters,* **2017**, *8*, 647-656.

Concerning the choice of predictor variables: We agree that a paragraph describing the general idea of the relation between MSG channels and cloud properties in section 2.2.2 is missing. We therefore added the following information:

"The rainfall retrieval technique presented here works under the assumption that VIS, NIR and IR channels of MSG SEVIRI provide proxies for microphysical cloud properties, which are, in turn, related to rainfall. VIS and NIR channels have been shown to be related to cloud optical depth (Roebeling et al., 2006; Benas et al., 2017) and cloud water path (Kühnlein et al., 2014b) where the NIR channel is further related to cloud particle size (Roebeling et al., 2006). The IR channels have been shown to provide information about the cloud top temperature which was used as a proxy for cloud height (Hamann et al., 2014). The cloud droplet effective radius as well as liquid water path during night was approximated using IR channel differences (Merk et al., 2011; Kühnlein et al., 2014b)."

From a technical perspective, it is no problem to insert all available information, even though individual channels might only have minor relations with rainfall. We added a note on that issue in the method section:

"The function of the neural network is then to learn the relations between the spectral information and rainfall areas or rainfall quantities, respectively. In this context, a sophisticated pre-selection of input variables is not required, as the network is able to deal with correlated and even uninformative predictors unless their number is very high (Meyer et al., 2017), which was not the case in this study."

---

ii) What is the network architecture (number of hidden levels and perceptrons) and how it has been designed. The text P6, l6-17 : The number of hidden units were tuned for each value ...., is not clear in this regard.

<u>Response</u>
We made the architecture clear and improved the description of the hyperparameters that required tuning.

“A single-hidden-layer feed-forward neural network was applied as machine learning algorithm. The spectral channels of MSG SEVIRI as well as the channel differences served as input nodes (predictor variables). The neural network was then applied to learn the relations between these spectral information and rainfall areas or rainfall quantities, respectively. In this context, a sophisticated pre-selection of input variables is not required, as the network is able to deal with correlated and even uninformative predictors unless their number is very high (Meyer et al., 2017), which was not the case in this study. For the technical realisation, all steps of model training were performed using the R environment for statistical computing (R Core Team, 2016). The neural network implementation from the "nnet" package (Venables and Ripley, 2002; Ripley and Venables, 2016) in R was used in conjunction with the "caret" package (Kuhn, 2016) that provides enhanced functionalities for model training, estimation and validation.“

“Neural networks require two hyperparameters to be tuned to avoid under- or overfitting of the data: the number of neurons in the hidden layer, as well as the weight decay. The neurons in the hidden layer represent nonlinear combinations of the input data and their number influences the performance of the model (Panchal et al., 2011). Weight decay penalizes large weights and controls the generalisation of the outcome (Krogh and Hertz, 1992). “

The actual number of neurons in the hidden layer results from the model tuning. We added a table showing the final model settings.

“Optimal hyperparameters for the individual models revealed during the tuning study and applied in the final model fitting.”

	Number of neurons	Weight decay	Threshold
Rainfall areas at daytime	5	0.05	0.07
Rainfall areas at nighttime	5	0.07	0.01
Rainfall quantities at daytime	5	0.05	
Rainfall quantities at nighttime	5	0.05	

---

iii) What is the training procedure used in the study. Section 2.3.3 does not appear clear on this subject both for the language and the references provided (see point 1 above) and because the cited paper Meyer et al. 2016 does not provide more details about this procedure (apart from the threshold tuning methodology).

<u>Response</u>
Meyer et al. 2016 was included as a reference for the threshold tuning. We now made clear that the final step is fitting the model to all training data using the optimal set of hyperparameters:
“The optimal values for the hyperparameters that were revealed in the tuning study (Tab. 1) were adopted for the final model fitting. In this step, the model is fit to all training data using the optimal hyperparameters.”

We also included a short paragraph on the spatial estimations (see also minor comment 4):

“2.3.4 Spatial estimations of rainfall
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."

Thus, our description of the training procedure now contains the selection of predictor and response variables as well as their preprocessing, the network architecture, the model tuning and cross validation approach, the final fit of the models and the validation as well as spatial model estimations Please let us know if you still miss information.

---

3 - The use of rain gauges as ground truth requires checks on the data quality. In the paper some aspects of this issue should be developed, e.g check on no-data or no-rain, consistency between data from different networks. Is the retrieval quality depending on the rain gauges density?

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. We now 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 for example filtering of data beyond common data ranges, or situational checks for consistency with related parameters (e.g. air humidity) by SASSCAL. The data 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 of rainfall per hour. All station data that provided sub-hourly information was aggregated to a temporal resolution of 1 hour within the database."

Unfortunately, we don't see a way to ensure a consistency between data from different networks: First of all, we can't analyze weather inconsistencies exist or not because this would require having stations from the different networks at exactly the same location which we don't have. Second, even if we had, how to correct for inconsistencies? There would probably be ways in areas where a high number of stations are available so that systematic inconsistencies in the data could be studied in a robust way. But this is not possible for our study. Our study relies on (comparably) sparse data from different sensors and provider-specific ways to process data. We don't see a way to check and/or correct for inconsistencies. However, we now accounted for this point in the discussion:

"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. "

It's further difficult to test if the retrieval quality depends on the gauge density. An obvious test would be to correlate the station density with the performances. However, that wouldn't lead to meaningful results since areas with many stations allow for a robust signal while areas with low density of stations don't because there are simply too few data.

However, we assume that different densities don't affect the model performance for the following reasons: We assume that the density would certainly have an effect if we trained on few scenes only, because then the current conditions of the areas with high station densities are highlighted. However since we trained on a long period, the model saw different conditions from dry to wet to learn from. Therefore it should be able to learn the relations between cloud properties and rainfall without the risk of over-fitting towards areas with high station densities.

---

4 – Figures 3 and 4 show the box plots concerning the POD, FAR, PDF, HSS, RMSE and rho evaluated considering the whole set of data; It would be more effective to evaluate these indexes considering different ranges of precipitation values (e.g. 0-25 mm, 25-50 mm etc).

Response

Thank you for this comment! We now compared POD for different measured rainfall quantities and we compared FAR for different predicted rainfall quantities. The results give valuable information about the performance: high rainfall rates could be very well recognized as rainy clouds by the model. Though the model overestimated rainfall (high FAR), the predicted rainfall quantities for these false alarms were comparably low. We accounted for this in the results:

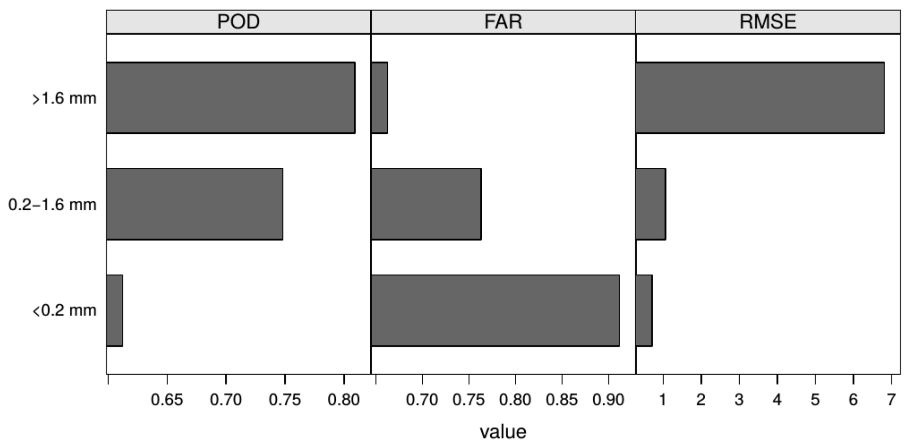
“The POD was highest for high measured rainfall quantities and decreased for lower rainfall quantities (Fig. 4). FAR was highest for low predicted rainfall quantities and decreased for higher predicted quantities. [...] Especially data points with low or medium measured rainfall could be estimated with low RMSE (Fig. 4).

And in the discussion:

“The strength of the retrieval in terms of rainfall areas classification was a high POD for heavy rainfall events. The rainfall quantities for the heavy rainfall events were, however, underestimated in most cases. The major problem of the model was the overestimation of rainfall events leading to an overestimation of rainfall quantities. However, false alarms in the retrieval were generally predicted with low rainfall quantities.“

We presented the figure as barplot since a similar boxplot representation is not possible in this context: The boxplots base on the POD/FAR/etc on a scene basis, thus the data points of the boxplot could not be assigned to a unique rainfall quantity class.

POD can't be calculated when no rainfall is measured and FAR/POFD can't be calculated when rainfall is measured/no rainfall is predicted. Therefore, we can't make sense of a HSS for different rainfall classes since it bases on POD and FAR. For that reason we only compared POD and FAR. We didn't follow your suggestion of a class-based comparison for the correlation coefficient. For the correlation we want to know the model's ability to distinguish between low and high rainfall. When we only consider small parts of the gradient we would lose too much information.



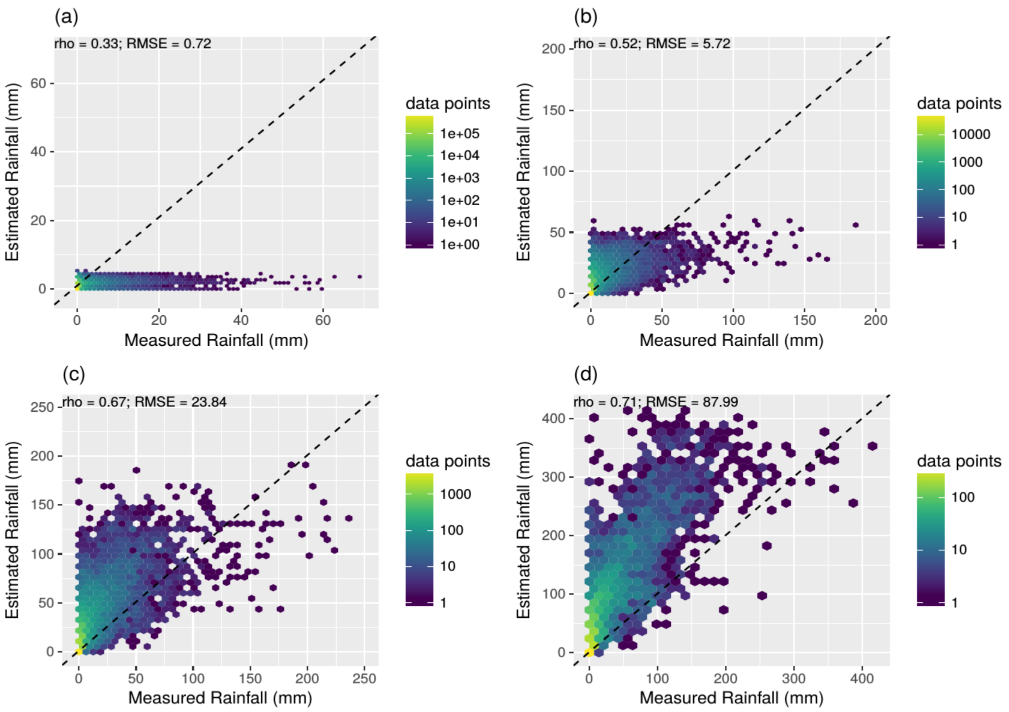
Minor revisions:
1 – The section 2.2.2 should be modified by introducing a short description of the ability of the Seviri channels to provide information on the state of the atmosphere and the ground. This is important to clarify the choices that led to the selection of the neural network inputs.

Response
See major comment 2i

---

2 – The performance of the retrieval technique (P8, L5-6) shown in fig. 5 (P11) could be presented in a more complete way by inserting in the four panels the corresponding RMSE and mean bias values. In the figure the colour bar (data point density) should be added.

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. We also added the RMSE in addition to rho.



---

3 – The reference to Smith et al. 2007 (P7, L9) can be updated with: Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The global precipitation measurement mission, B. Am. Meteorol. Soc., 95, 701-722, doi:10.1175/BAMS-D-13-00164.1, 2014.

Response
done

---

4 – P6, L3 Please explain the criteria that has allowed to split the database into day and night.

<u>Response</u>
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 (data points with a solar zenith angle < 70°) and a nighttime dataset (data points with a solar zenith angle > 70°)"

We also added a section of how the spatial estimation of rainfall were created because in this case, the mean solar zenith angle of the entire scene was decisive for the choice of the model:

"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."

---

5 – The paper contains a few typos that need to be corrected.

<u>Response</u>
We checked the manuscript again for typos.

Satellite based high resolution mapping of rainfall over Southern Africa

Hanna Meyer[1], Johannes Drönner[2], and Thomas Nauss[1]

[1]Environmental Informatics, Faculty of Geography, Philipps-University Marburg, Deutschhausstr. 10, 35037 Marburg, Germany

[2]Database Research Group, Faculty of Mathematics und Informatics, Philipps-University Marburg, Hans-Meerwein-Str. 6, 35032 Marburg, Germany

Correspondence to: Hanna Meyer (hanna.meyer@geo.uni-marburg.de)

Abstract. A spatially explicit mapping of rainfall is ~~highly required~~ necessary for Southern Africa for eco-climatological studies or nowcasting but accurate estimates are still a challenging task. This study presents a method to estimate hourly rainfall based on data from the Meteosat Second Generation (MSG) ~~spinning enhanced visible and infrared imager~~ Spinning Enhanced Visible and Infrared Imager (SEVIRI). Rainfall measurements from about 350 weather stations from ~~the years~~ 2010-2014 served as ground ~~truths~~ truth for calibration and validation. SEVIRI and weather station data were used to train neural networks that allowed ~~predicting~~ the estimation of rainfall area and rainfall quantities ~~during~~ over all times of the day. The results revealed that 60 % of recorded rainfall events were correctly classified by the model (Probability ~~of detection~~Of Detection, POD). However, the ~~false alarm ratio~~ False Alarm Ratio (FAR) was high (0.80), leading to ~~an~~ a Heidke Skill Score (HSS) of 0.18. ~~Predicted~~ Estimated hourly rainfall quantities were estimated with an average hourly correlation of rho = 0.33 and a ~~RMSE~~ Root Mean Square Error (RMSE) of 0.72. The correlation increased with temporal aggregation to 0.52 (daily), 0.67 (weekly) and 0.71 (monthly). The main weakness was the overestimation of rainfall events. The model results were compared to the ~~IMERG product~~ Integrated Multi-satellitE Retrievals for GPM (IMERG) of the Global Precipitation Measurement (GPM) mission. Despite being a comparably simple approach, the presented MSG based rainfall retrieval outperformed GPM IMERG in terms of rainfall area detection where GPM IMERG had a considerably lower POD. The HSS was not significantly different compared to the MSG based retrieval due to a lower FAR of GPM IMERG. There were no further significant differences between the MSG based retrieval and GPM IMERG in terms of correlation with the observed rainfall quantities. The MSG based retrieval, however, provides rainfall in higher spatial resolution. Though ~~it remains challenging to~~ estimate rainfall from satellite data ~~, especially on a high temporal resolution~~remains challenging especially at high temporal resolutions, this study showed promising results towards improved spatio-temporal estimates of rainfall over Southern Africa.

1 Introduction

The dynamics of rainfall play an important role in Southern Africa especially in the arid and semi-arid areas where farming is a main income and the quality of the pastures mainly depends on water availability (Fynn and O'Connor, 2000). ~~An accurate~~ Accurate nowcasting of rainfall ~~in~~ at high temporal and spatial ~~resolution~~ resolutions is therefore of interest for the farmers in Southern Africa and would help them to assess the carrying capacity of their land. It is of further importance as a baseline product for a variety of environmental research studies as rainfall is a key variable for many ecological and hydrological processes.

Rain gauges are still considered as the most accurate way to measure rainfall. Southern Africa features a network of rain gauges operated by the weather services of the individual countries as well as by a variety of research projects. However, the network does not feature a sufficient density to capture spatially highly variable rainfall dynamics. To obtain spatially explicit data, ground-based radar networks are well established to measure rainfall in other parts of the world (e.g. RADOLAN in Germany, Bartels et al. (2004)). A radar network ~~for entire~~ covering the entire region of Southern Africa, however, is currently not available and the existing radar-based rainfall estimates in South Africa are still afflicted with many uncertainties (IPWG, 2016). A satellite-based monitoring of rainfall is therefore an obvious alternative.

A number of global satellite-derived products have been developed in the last decades (e.g. TRMM, CMORPH, PERSIANN, see review in Kidd and Huffman (2011); Prigent (2010); Thies and Bendix (2011); Kidd et al. (2011); Levizzani et al. (2002)). Since 2014, the latest product from the Global Precipitation Measurement (GPM) mission, as a successor of the Tropical Rainfall Measuring Mission (TRMM), provides the most recent global ~~estimations of precipitation in~~ estimates of precipitation at high spatial and temporal ~~resolution. It can~~ resolutions. It might be expected that the GPM products would feature a high ~~accuracy as~~ degree of accuracy since the TRMM-3B42 product has been identified as the most accurate retrieval at least for east Africa (Cattani et al., 2016).

In addition to global rainfall retrievals, a number of regionally adapted retrievals were developed in the last decades (Kühnlein et al., 2014b, a; Meyer et al., 2016; Feidas and Giannakos, 2012; Giannakos and Feidas, 2013). ~~Kühnlein et al. (2014b, a); Meyer et al. (~~ Meyer et al. (2016) presented a methodology to estimate rainfall from optical Meteosat ~~second generation~~ Second Generation (MSG) ~~spinning enhanced visible and infrared imager~~ Spinning Enhanced Visible and InfraRed Imager (SEVIRI) data for Germany. In this approach, machine learning algorithms were used to relate the spectral properties of MSG to reliable radar data as a ground truth. Though the retrieval showed promising results, such spatially comprehensive ground truth data are lacking for ~~South~~ Southern Africa. An adaptation of the retrieval technique to ~~South~~ Southern Africa hence requires a model training that ~~relys~~ relies on sparse weather station data as a ground truth.

This study aims to test the suitability of a MSG and artificial neural network based rainfall retrieval which is regionally trained using rain gauge data to provide spatially explicit estimates of rainfall areas and rainfall quantities for Southern Africa. The suitability of the model is assessed by validation with independent weather station data and comparison to the ~~GPM IMERG~~ Integrated Multi-satellitE Retrievals for GPM (IMERG) product.

2 Methods

The methodology is divided into a pre-processing of satellite and rain gauge data, model tuning and training including its validation, model ~~prediction~~ estimation and comparison to GPM IMERG (Fig. 1).

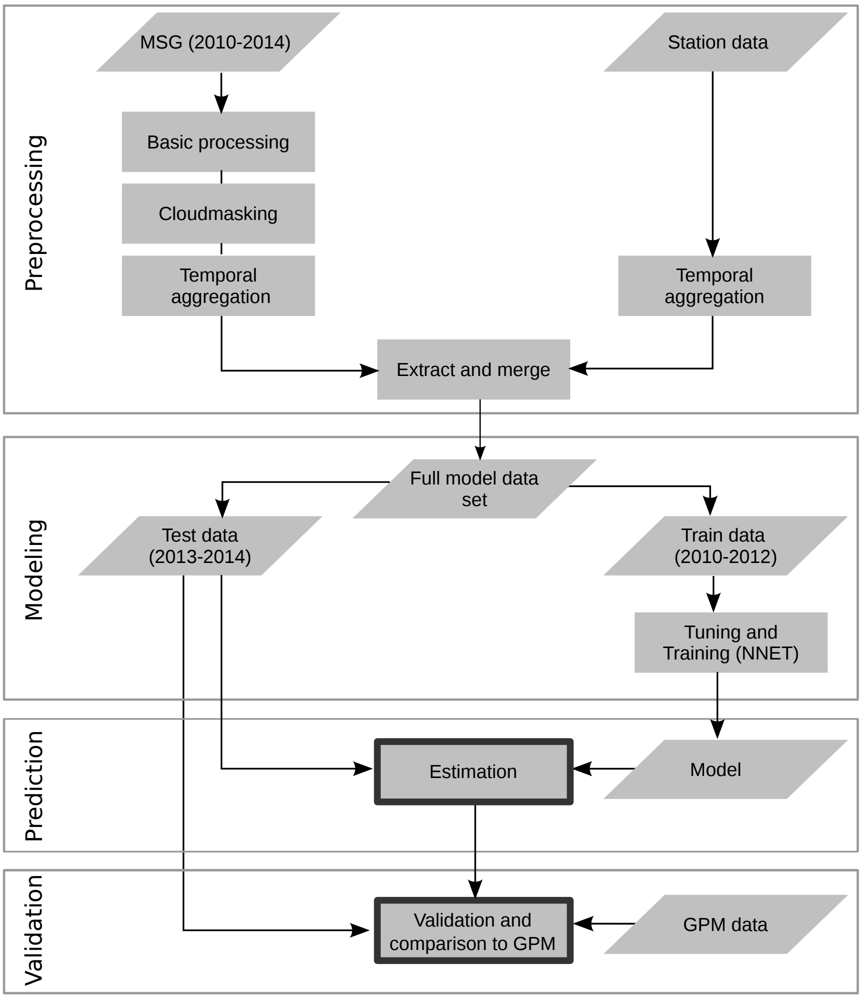


Figure 1. Flow chart of the methodology applied in this study.

2.1 Study area

The area of investigation comprises South Africa, Lesotho and Swaziland, Namibia, Botswana, Zimbabwe as well as parts of Mozambique (Fig. 2). Average ~~yearly rainfall sums~~ annual rainfall in Southern Africa roughly ~~follow~~ follows an aridity gradient from the dry west to the more humid east. With the ~~exception~~ exceptions of some coastal regions in South Africa, most rain falls during the summer months. In the coastal areas of South Africa, frontal systems cause light ~~rains~~ rain that may last over several days. The majority of interior areas are dominated by local and short-term convective heavy showers mostly with thunder in the afternoon or evening hours. Rain from synoptic systems lasting up to several days also occurs. Snow and hail only contribute a ~~neglegible~~ negligible amount to the overall precipitation ~~sums~~totals. The inter-annual variability of rainfall is high for the arid areas. For a detailed description of Southern African rainfall characteristics see ~~Kruger (2007); Kaptué et al. (2015)~~ Kruger (2007) and Kaptué et al. (2015) .

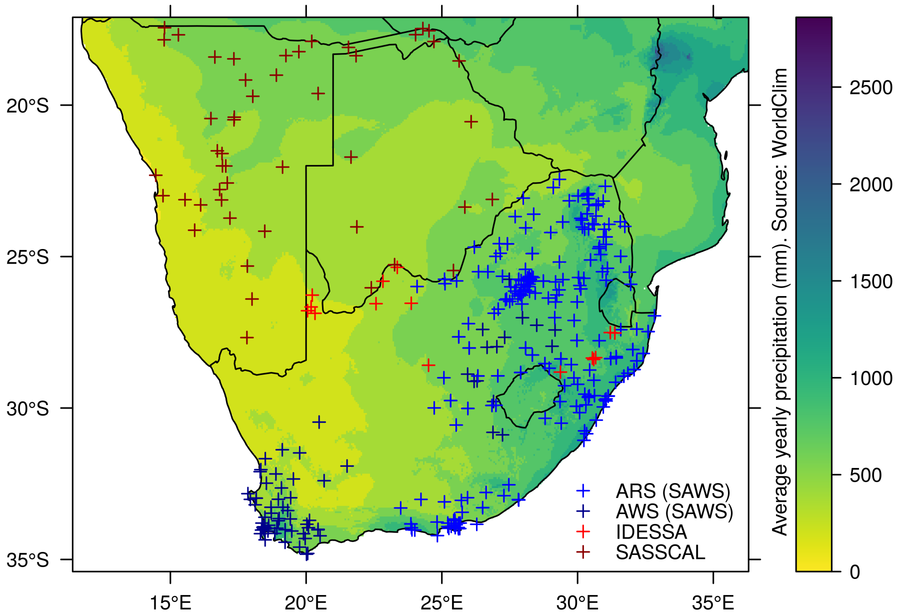


Figure 2. Map of the average ~~yearly~~ annual precipitation sums in the study area as estimated by WordClim (Hijmans et al., 2005). Points show the locations of the weather stations that were used as ground truth data in this study. Automatic ~~rainfall stations~~ Rainfall Stations (ARS) and ~~automatic weather stations~~ Automatic Weather Stations (~~ARS~~AWS) are operated by the South African Weather Service (SAWS). Further stations are operated by SASSCAL WeatherNet as well as by the IDESSA project.

2.2 Data and Preprocessing

2.2.1 Station data

Rainfall data for ~~the years~~ 2010 to 2014 were obtained from the South African Weather Service (SAWS). The data were recorded ~~from~~ at 229 automatic rainfall stations and 91 automatic weather stations (Fig. 2). They were complemented by 22 stations from SASSCAL WeatherNet (www.sasscalweathernet.org/) located in southern Namibia and Botswana. For ~~the year~~ 2014, data from an additional 15 stations in South Africa operated by the IDESSA project (An Integrative Decision Support System for Sustainable Rangeland Management in Southern African Savannas, www.idessa.org/) were available. 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 of rainfall per hour. All station data that provided sub-hourly information ~~were~~ was aggregated to a temporal resolution of 1 hour within the database. Though the station data is not randomly distributed in the model domain, it covers the entire aridity gradient, from sites with very low ($<$ ~~200mm~~200 mm) precipitation to sites in areas with highest ($\sim$ 1500 mm) yearly precipitation sums.

2.2.2 Satellite data

MSG SEVIRI (Aminou et al., 1997) scans the full disk every 15 minutes with a spatial resolution of ~~3 by 3~~ 3×3 km at sub-satellite point (3.5×3.5 km in Southern Africa). Reflected and emitted radiances are measured by 12 channels, three channels at visible (VIS) and very near infrared wavelengths (NIR, between 0.6 and 1.6 μm), eight channels ranging from near-infrared to thermal infrared wavelengths (IR, between 3.9 and 14 μm) and one high-resolution ~~visible channel . MSG SEVIRI data were preprocessed based on a Meteosat processing scheme that uses xxl technology and custom raster extensions which were designed to support OpenCL acceleration~~ VIS channel with a spatial resolution of 1×1 km which was not considered in this study.

The rainfall retrieval technique presented here works under the assumption that VIS, NIR and IR channels of MSG SEVIRI provide proxies for microphysical cloud properties, which are, in turn, related to rainfall. VIS and NIR channels have been shown to be related to cloud optical depth (Roebeling et al., 2006; Benas et al., 2017) and cloud water path (Kühnlein et al., 2014b) where the NIR channel is further related to cloud particle size (Roebeling et al., 2006) . The IR channels have been shown to provide information about the cloud top temperature which was used as a proxy for cloud height (Hamann et al., 2014) . The cloud droplet effective radius as well as liquid water path during night was approximated using IR differences (Merk et al., 2011; Kühnlein et al., 2

MSG SEVIRI Level 1.5 data (EUMETSAT, 2010) was preprocessed to radiance values according to EUMETSAT (2012b) and brightness temperatures according to EUMETSAT (2012a) 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.

2.2.3 Cloud mask

A cloudmask was used to exclude all pixels that were not cloudy in the respective SEVIRI scenes. For ~~the years~~ 2010 to 2012, the CM SAF CMa Cloudmask product (Kniffka et al., 2014) was applied. Due to the availability of the CM SAF CMa cloudmask dataset which was currently limited to the years 2004 to 2012, we used the cloud mask information of the CLAAS-2 data record (Finkensieper et al., 2016) for the years 2013 and 2014 which is the 2nd edition of the SEVIRI-based cloud property data record provided by the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF; see also Stengel et al. (2014) for further information on CLAAS). All pixels that were classified as cloud contaminated or cloud filled were interpreted as cloudy. Pixels that were classified as cloud-free were ~~masked~~ excluded from further analysis.

2.3 Model strategies for rainfall estimation

2.3.1 General model framework

The modeling methodology follows the study of ~~Kühnlein (2014); Kühnlein et al. (2014b)~~ Kühnlein et al. (2014a, b) who used the spectral channels of MSG SEVIRI to train a Random Forest model that is able to spatially ~~predict~~ estimate rainfall areas and rainfall rates over Germany. Based on this study, Meyer et al. (2016) have shown that neural networks outperform the initially used Random Forest algorithm. In these previous studies on the rainfall retrieval, the radar based RADOLAN product (Bartels et al., 2004) was used as ground truths to train the model. The high data quality and spatially explicit information allowed the model to be optimised without too ~~many confusions~~ much confusion caused by uncertainties in the training data. However, the goal of the retrieval was that it can be applied to areas where spatially explicit data for rainfall are not available, as it is the case in Southern Africa. ~~All steps of model training were performed using the R environment for statistical computing (R Core Team, 2016) .~~

2.3.2 Training and test data sets

Cloud masked MSG data from 2010 to 2014 were extracted at the locations of the weather stations. 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).

The spectral channels as well as the channel differences Δ T6.2 - 10.8, Δ T7.3 - 12.1,Δ T8.7 - 10.8, Δ T10.8 - 12.1, Δ T3.9 - 7.3, Δ T3.9 - 10.8 and the sun zenith were used as predictor variables during daytime~~. Meyer et al. (accepted) tested different texture parameters as additional predictor variables~~ , in accordance to (Kühnlein et al., 2014b) and previous studies on MSG based delineation of cloud properties (see section 2.2.2). 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 channels. As additional potential predictors, Meyer et al. (2017) tested different cloud texture parameters and have shown that ~~these spectral channels~~ the chosen spectral channels and differences are sufficient as predictors.

Since neural networks require that the predictor variables are standardized, all predictors were centered and scaled by dividing the values of the mean-centered variables by their standard deviations. Since the VIS and NIR channels of MSG are not available during the nighttime, the dataset was split into a daytime dataset (~~all scenes where VIS and NIR were available~~data points with a solar zenith angle $< 70°$) and a nighttime dataset (~~reliable VIS and NIR not available)~~ data points with a solar zenith angle $> 70°$) and were considered in separate models. 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 response variables (rainfall yes/no and rainfall quantities) were taken from the rain gauge measurements.

The years 2010 to 2012 were used for model training. The year 2013 was used for validation. The retrieval process was ~~two-folded~~ two-step and consisted of (i) the identification of precipitating cloud areas and (ii) the assignment of rainfall quantities. All 2010 to 2012 data from the rain gauges that are masked as cloudy by the cloud mask products were used for training the rainfall area model. All recorded rainfall events were used for training the rainfall quantities model. The resulting training dataset comprised 917774 (daytime) and 1409072 (nighttime) samples for the rainfall area training and 69703 (daytime) and 129325 (nighttime) samples for training of rainfall quantities from 26243 individual MSG scenes.

2.3.3 Tuning and model training

A single-hidden-layer feed-forward neural network was applied as machine learning algorithm. The spectral channels of MSG SEVIRI as well as the channel differences served as input nodes (predictor variables). The neural network was then applied to learn the relations between these spectral information and rainfall areas or rainfall quantities, respectively. In this context, a sophisticated pre-selection of input variables is not required, as the network is able to deal with correlated and even uninformative predictors unless their number is very high (Meyer et al., 2017) , which was not the case in this study. For the technical realisation, all steps of model training were performed using the R environment for statistical computing (R Core Team, 2016) . The neural network implementation from the "nnet" package ~~(Venables and Ripley, 2002)~~ (Venables and Ripley, 20 R was used in conjunction with the "caret" package ~~Kuhn (2016)~~ (Kuhn, 2016) that provides enhanced functionalities for model training, ~~prediction~~ estimation and validation. ~~Model parameters~~

Neural networks require two hyperparameters to be tuned to avoid under- or overfitting of the data: the number of neurons in the hidden layer, as well as the weight decay. The neurons in the hidden layer represent nonlinear combinations of the input data and their number influences the performance of the model (Panchal et al., 2011) . Weight decay penalizes large weights and controls the generalisation of the outcome (Krogh and Hertz, 1992) . The number of neurons as well as weight decay were tuned using a stratified 10-fold cross-validation~~where each fold~~ . Thus, the training samples were randomly partitioned into 10 equally sized folds with respect to the distribution of the response variable (i.e., raining cloud pixels, rainfall rate). Thus, every fold is a subset (1/10) of the training samples and has the same distribution of ~~rainfall areas, or rainfall quantities respectively, as the entire training dataset~~the response variable as the total set of training samples. Models were then fitted by repeatedly leaving out one of the folds. The performance of a model was then determined by predicting on the held back fold. The performance metrics from the held back iterations were averaged to the overall model performance for the respective set of tuning values. For the rainfall areas classification models, the distance to a "perfect model", based on Receiver Operating Characteristics (ROC) analysis (see Meyer et al. (2016) for its application in rainfall retrievals) was used as decisive performance metric. For the rainfall quantities regression models, the Root Mean Square Error (RMSE) was used. The number of hidden units were tuned for each value between two and the number of predictor variables~~(Kuhn and Johnson, 2013a)~~. Weight decay was tuned between 0 and 0.1 with increments of 0.02 (Kuhn and Johnson, 2013b) . For training of rainfall areas, the threshold that separates rainy from non-rainy clouds according to the ~~predicted~~ estimated probabilities was an additional tuning parameter. The optimal threshold was expected to be considerably smaller than 0.5 since the amount of non rainy samples was higher than the amount of rainy samples. Therefore, the range of tested thresholds was 0 to 0.1 with increments of 0.01, and 0.4 to 1 with increments of 0.1. See Meyer et al. (2016) for further details of the threshold tuning methodology.

The optimal values for the hyperparameters that were revealed in the tuning study (Tab. 1) were adopted for the final model fitting. In this step, the model is fit to all training data using the optimal hyperparameters.

Table 1. Optimal hyperparameters for the individual models revealed during the tuning study and applied in the final model fitting.

	Number of neurons	Weight decay	Threshold
Rainfall areas at daytime	5	0.05	0.07
Rainfall areas at nighttime	5	0.07	0.01
Rainfall quantities at daytime	5	0.05	
Rainfall quantities at nighttime	5	0.05	

2.3.4 Spatial estimations of rainfall

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.

2.4 Validation

Model ~~predictions~~ estimates and weather station records from the entire year 2013 were used as independent data for model validation. For the validation of ~~predicted~~ estimated rainfall areas, all pixels at the location of the weather stations that were classified as cloudy by the cloud mask product were considered. Therefore the information from the weather stations about whether it was raining or not was compared to the model ~~prediction~~ estimate for the respective MSG pixel. The validation data contained 403211 samples during daytime and 565415 samples during nighttime. Average hourly ~~probability of detection~~ Probability Of Detection (POD), Probability ~~of false detection~~ Of False Detection (POFD), False ~~alarm ratio~~ Alarm Ratio (FAR) and Heidke ~~skill score~~ Skill Score (HSS) were calculated as validation metrics. The POD gives the percentage of rain pixels that the model correctly identified as rain (Tab. 2, 3). POFD gives the proportion of non-rain pixels that the model incorrectly classified as rain. The ~~false alarm ratio (FAR)~~ FAR gives the proportion of ~~predicted~~ estimated rain where no rain is observed. The HSS also accounts for chance agreement and gives the proportion of correct classifications (both rain pixels and non-rain pixels) after eliminating expected chance agreement. ~~HSS is independent of the bias in the classifications.~~

Table 2. Confusion matrix as baseline for the calculation of the verification scores used for the validation of the rainfall area ~~predictions~~estimates.

		Observation	
		Rainfall	**No Rainfall**
Estimation	**Rainfall**	True positives (TP)	False positives (FP)
	No Rainfall	False negatives (FN)	True negatives (TN)

Table 3. Categorical metrics for validation of rainfall area ~~predictions~~estimates.

Metric	**Formula**	**Range**	**otimal value**
Probability Of Detection	$POD = \frac{TP}{TP+FN}$	0 - 1	1
Probability Of False Detection	$POFD = \frac{FP}{FP+TN}$	0 - 1	0
False Alarm Ratio	$FAR = \frac{FP}{TP+FP}$	0 - 1	0
Heidke Skill Score	$HSS = \frac{TP*TN-FP*FN}{[(TP+FN)*(FN+TN)+(TP+FP)*(FP+TN)]/2}$	-∞ - 1	1

To evaluate the ability of the model to ~~predict~~ estimate rainfall quantities, the correlation between the measured and the ~~predicted~~ estimated hourly rainfall was calculated using ~~Spearmans rho~~Spearman's Product Moment Correlation (rho) to account for a non-normal distribution of the data. ~~Further, the root mean square error (RMSE) was used. All clouded~~ RMSE was also calculated. All cloudy data points (including non-rainy data points) were used for the validation of rainfall quantities. The rainfall quantities were further aggregated to daily, weekly and monthly rainfall sums to assess the performance of the model on different temporal scales.

2.5 Comparison to GPM

The results of the presented rainfall retrieval were compared to the rainfall estimates of the GPM mission. GPM, as a successor of the Tropical Rainfall Measuring Mission (TRMM), consists of an international network of satellites ~~aiming at~~ designed for worldwide high resolution precipitation estimates ~~(Smith et al., 2007)~~ (Hou et al., 2014; Skofronick-Jackson et al., 2017) . GPM provides data from March 2014 onwards. The GPM IMERG product estimates rainfall by combining all available passive-microwave ~~instruments~~ estimates as well as microwave-calibrated infrared satellite estimates and data from rainfall gauges. GPM IMERG is available in 6h, 18h and 4 ~~month~~ months latency.

In this study the 4 month latency (final product) with 30 minutes temporal and 0.1° spatial resolution (~10km x 10km) was used (Huffman et al., 2014) . Due to different data availabilities of GPM IMERG, MSG as well as weather station data, the comparison was conducted for the overlapping time period late March 2014 to August 2014. GPM was aggregated from 30 minutes to 1h to match the temporal resolution of the MSG ~~predictions~~based estimates. Both products were validated using the weather station data as a reference. The performance metrics were compared between the MSG product and the GPM product on an hourly basis.

3 Results

3.1 Model performance

On average, 60 % of the rainfall observations were correctly identified as rainy by the model with a high number of scenes having much higher PODs (Fig. 3). The ~~probability of false detection~~ POFD was low (18 % in average) but the ~~predictions~~ estimates featured a high ~~false alarm ratio~~ FAR of 0.80. The average HSS per scene was 0.18. The POD was highest for high measured rainfall quantities and decreased for lower rainfall quantities (Fig. 4). FAR was highest for low predicted rainfall quantities and decreased for higher predicted quantities.

The average hourly RMSE was 0.72 ~~mm~~ $mm\ h^{-1}$ (Fig. 5). Especially data points with low or medium measured rainfall could be estimated with low RMSE (Fig. 4). The RMSE was higher for high measured rainfall. Correlation indicated by ~~Spearmans~~ Spearman's rho was 0.33 ~~in~~ on hourly average. The performance of ~~the rainfall quantities assignment~~ modeled rainfall quantities increased with the aggregation level (Fig. 6). The average correlation increased from rho = 0.33 (hourly) to 0.52 on a daily, 0.67 on a weekly and 0.71 on a monthly basis. An overestimation of rainfall ~~quantities could be~~ is observed especially when aggregated to monthly ~~rainfall sums~~totals. An example of temporally aggregated rainfall ~~predictions are shown for the year~~ estimates for 2013 are shown in Fig. 7.

3.2 Comparison to GPM

Compared to GPM IMERG, the MSG based rainfall retrieval for the period Mar-Aug 2014 showed a higher POD (0.57) than GPM IMERG (0.28) which considerably underestimated rainfall events (Fig. 8). In contrast, GPM IMERG had a lower FAR (0.70) than the MSG based model (0.81). However, the FAR was high for both retrievals. The average HSS was the same

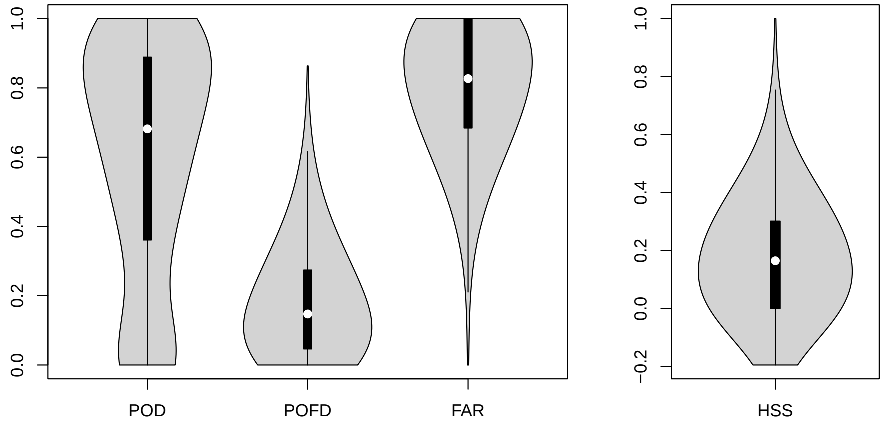


Figure 3. Validation of ~~predicted~~ estimated rainfall areas for ~~the year~~ 2013 on an hourly basis. Each of the data ~~point~~ points is the average performance of one hour. The data are visualized as "vioplot" where a boxplot is complemented by the kernel density of the data shown as grey areas at the sides of the boxplot.

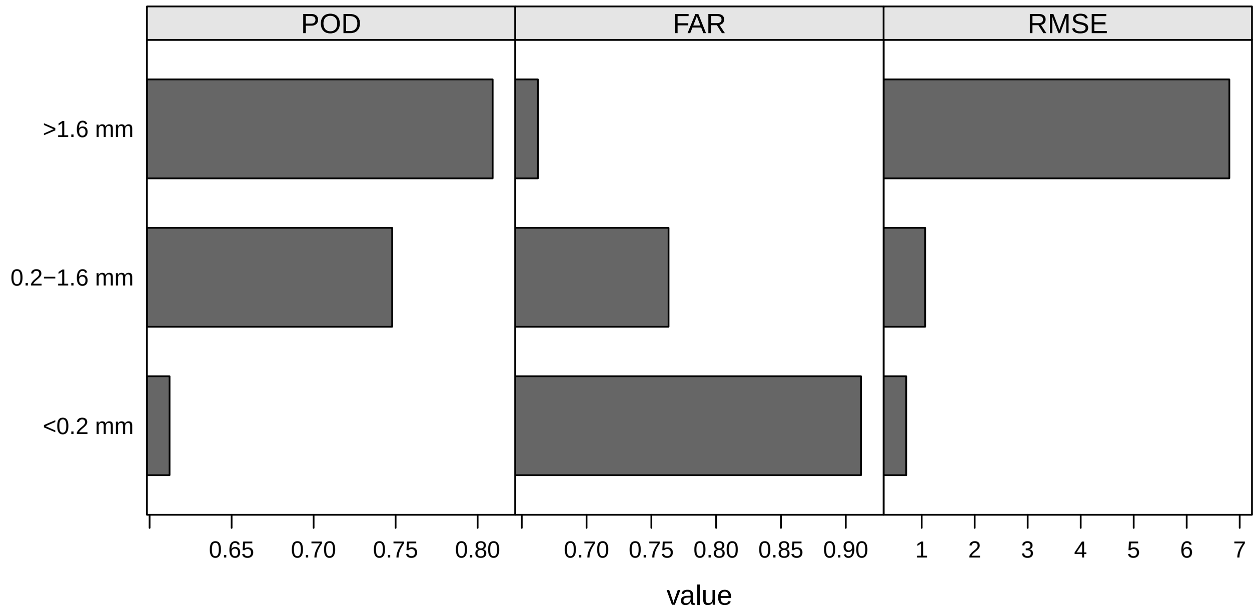


Figure 4. ~~Validation~~ Comparison of POD for different hourly measured rainfall quantities as well as FAR for different predicted rainfall quantities. RMSE was compared for ~~the year 2013 on an hourly basis~~different measured rainfall quantities. ~~Each~~ All data ~~point is~~ points from 2013 were used for the ~~average performance~~ calculation of ~~one hour~~the statistics. Thresholds for the three rainfall classes were set according to the first and third quartiles of the measured hourly rainfall quantities.

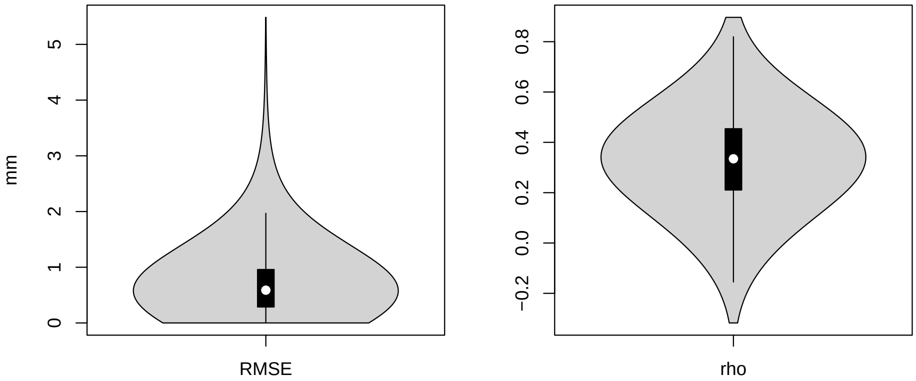


Figure 5. Validation of estimated rainfall quantities for 2013 on an hourly basis. Each of the data points is the average performance of one hour. See Fig. 3 for further information on the figure style.

for both retrievals (0.17), but the median HSS for GPM IMERG was 0 which was considerably lower than using the MSG based retrieval (0.10). Concerning the rainfall quantities, neither the correlation to measured rainfall nor the RMSE showed significant differences between both retrievals (Fig. 9). The average rho was 0.36 for the MSG based retrieval and 0.34 for GPM IMERG. The average RMSE was 0.88 for the MSG based retrieval and 0.85 for MSG IMERG.

Fig. 10 gives an example of the differences between the MSG based retrieval and GPM IMERG for 2014/04/24 12:00 UTC where severe floods occurred in the Eastern Cape province of South Africa. The colour composite of the corresponding MSG scene shows that clouds had a high optical depth in this area. The pattern is reflected in the ~~predictions~~ estimates of the MSG based retrieval that ~~predicted~~ estimated rainfall for the areas with high values of optical depth. This was partly confirmed by the weather station data. However, rainfall was also ~~predicted~~ estimated for areas where weather stations did not record any rainfall. In contrast, GPM IMERG showed an underestimation of rainfall areas, but still captured the high rainfall quantities that were recorded by the weather stations. The summary statistics for this hour are a POD of 0.75 for the MSG based retrieval and 0.19 for GPM IMERG. FAR was 0.65 and HSS 0.34 for the MSG based retrieval compared to a FAR of 0.89 and a HSS of 0.08 for GPM IMERG. The correlation between ~~predicted~~ estimated and observed rainfall was 0.39 for the MSG based retrieval and -0.06 for GPM IMERG.

4 Discussion

The presented monthly maps reflect the general spatial and temporal rainfall patterns of Southern Africa as shown in Kruger (2007). They also reflect the annual characteristics of the year 2013. For example, the heavy rainfall events over southern Mozambique and the Limpopo River basin during mid January ~~Manhique et al. (2015)~~ (Manhique et al., 2015) .

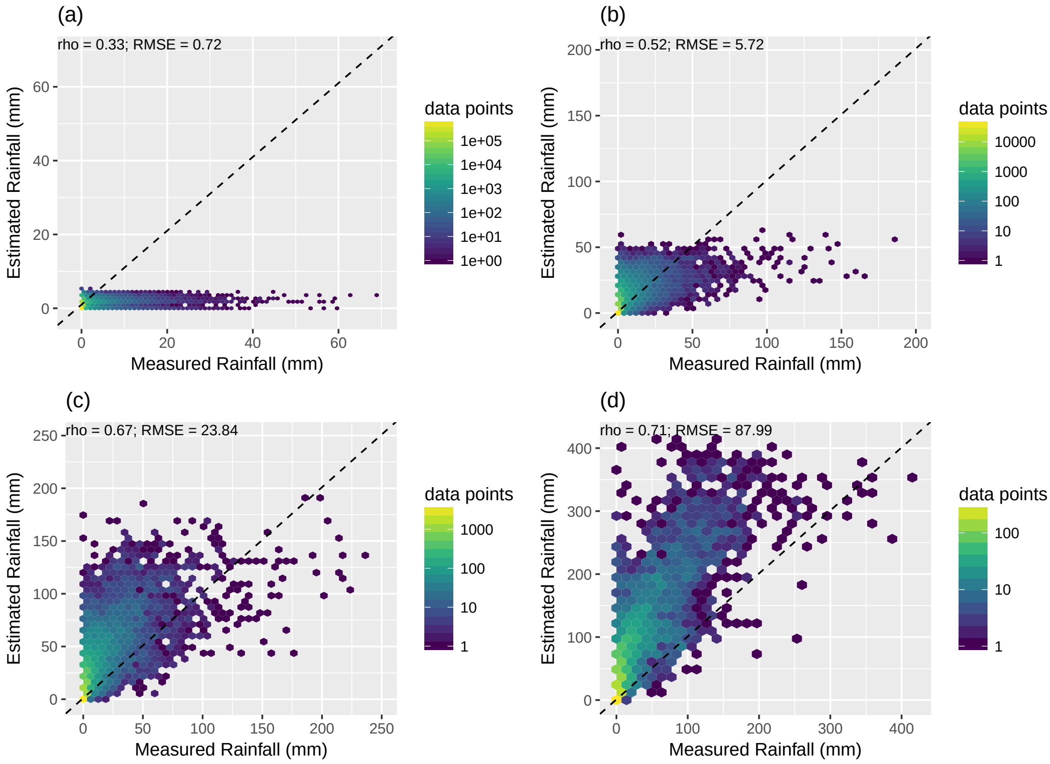


Figure 6. Validation of ~~predicted~~ estimated rainfall quantities for ~~the year~~ 2013 ~~on~~ at (a) hourly resolution ~~and~~ on the different aggregation ~~levels~~ (b) daily, (c) weekly, (d) monthly. Each of the data ~~point~~ points represents a station at the respective level of temporal aggregation. Rho represents the average correlation for each time step of the respective aggregation level. ~~The intensity of~~ For an easy visual interpretation, the ~~colour represents~~ data are presented via hexagon binning where the number of data ~~point density~~points falling in each hexagon are depicted by color.

The validation of the rainfall ~~retrieval~~ retrievals showed promising results but ~~highlights also~~ also highlights the difficulties of optical satellite-based rainfall estimates. The ~~major problem~~ strength of the retrieval in terms of rainfall areas classification was a high POD for heavy rainfall events. The rainfall quantities for the heavy rainfall events were, however, underestimated in most cases. The major problem of the model was the overestimation of rainfall events leading to an overestimation of rainfall quantities. However, false alarms in the retrieval were generally predicted with low rainfall quantities. In this context, it is of note that ~~the~~ in view to the scene-based validation strategy, FAR can easily increase ~~to elevated levels~~ in dry conditions when there are just a few false alarms in the ~~predictions~~ estimates and no rainfall was observed by any station. However, the FAR was still high for hours with a considerable number of rainfall events. This might be partly explainable by spatial displacement due to parallax shifts~~that~~. Though the shift is generally below 1 pixel in this region, even minor shifts can affect model training as well as the ~~predictions~~estimates. For future enhancement of the rainfall retrieval, a correction of the parallax shift (Vicente et al., 2002) would be appropriate. Differences in spatial and temporal scale are also an important issue especially since a

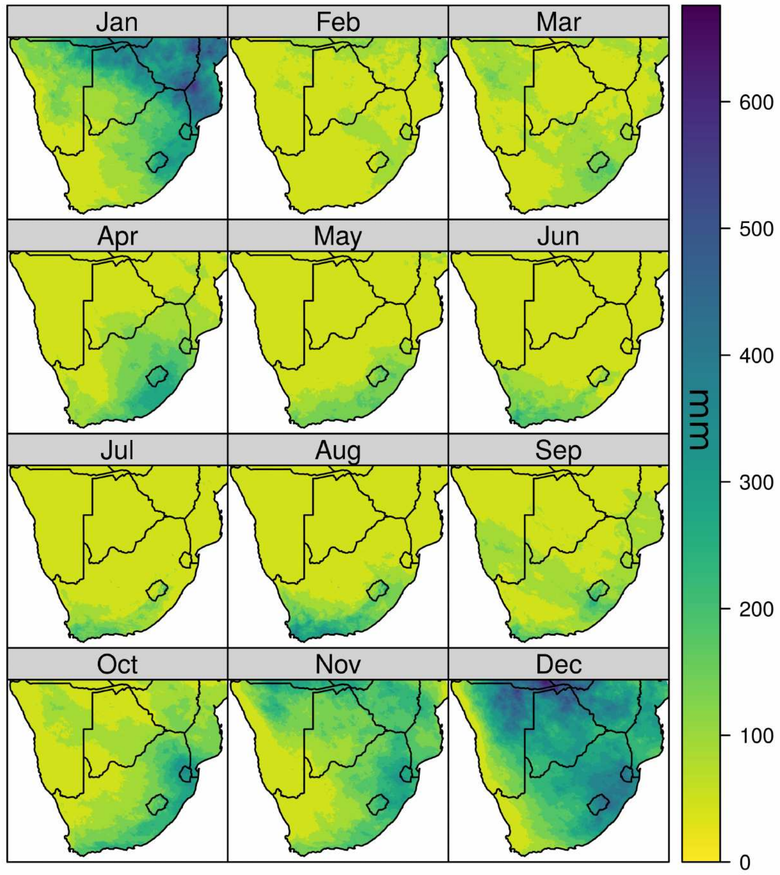


Figure 7. Monthly precipitation sums in mm of the year 2013 as estimated by this study.

majority of rainfall events in Southern Africa are of small spatial and temporal extent. The aggregation to an hour as well as the assumption that the weather station observation is representative for the entire pixel ~~,~~ are also problematic, though essential. The issue of scale especially affects the broader resolution GPM IMERG data where a several km sized pixel is validated by a single point measurement. Beside of the issue of scale and spatial displacement, the retrieval technique depends on the quality of the rain gauge observations. Although the data was quality checked, common problems associated with rain gauge measurements e.g. wind drift or evaporation leading to errors in the ground truth data and affect model training and validation remain ~~Kidd and Huffman (2011)~~ (Kidd and Huffman, 2011) . 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.

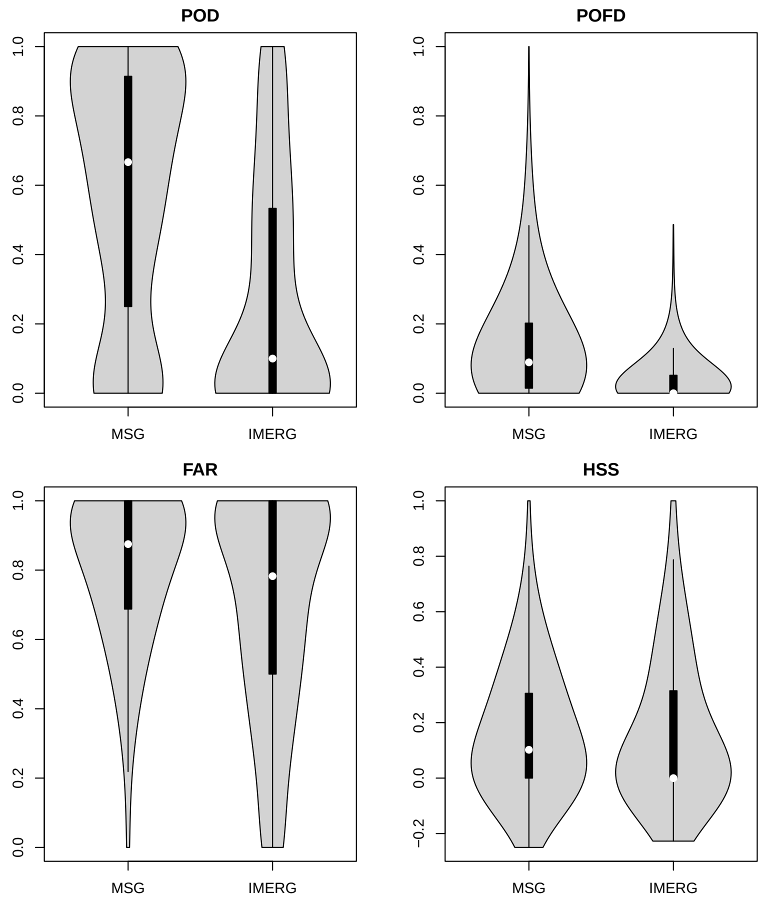


Figure 8. Comparison of the performance of the MSG based retrieval and GPM IMERG for rainfall area delineation between March and August 2014. Each of the data ~~point~~ points is the average performance of one hour. ~~The notches serve as a rough estimation of significant differences~~See Fig. 3 for further information on the figure style.

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.

Despite the errors and uncertainties associated with the presented rainfall retrieval, the combination of MSG data and neural networks are a promising approach. The model presented in this study outperformed the GPM IMERG product in terms of rainfall area detection where GPM IMERG considerably underestimated rainfall events. This behavior is partly explainable by scale because GPM IMERG has a coarser resolution of 0.1°. This makes local processes difficult to capture which is an disadvantage considering that in Southern Africa especially small scale convective showers contribute to rainfall sums Kruger

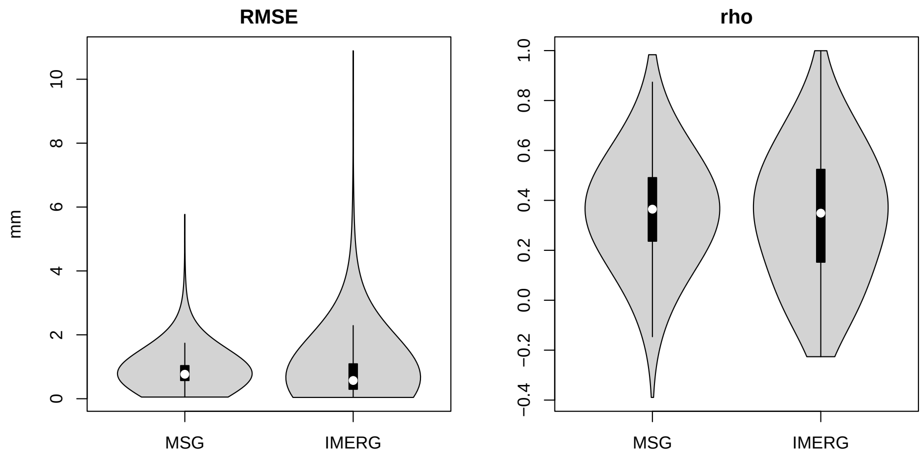


Figure 9. Comparison of the performance of the MSG based retrieval and GPM IMERG for hourly rainfall quantities between March and August 2014. Each of the data ~~point~~ points is the average performance of one hour. ~~The notches serve as a rough estimation of significant differences~~See Fig. 3 for further information on the figure style.

(2007). In terms of rainfall quantities, GPM IMERG and the presented retrieval did not show significant differences in ~~view to~~ correlation. The sample ~~predictions have shown that GMP~~ spatial comparison has shown that GPM IMERG has more differentiated rainfall estimates while the MSG based retrieval tends to ~~predict~~ estimate the mean distribution.

The presented MSG based retrieval is an easy to use method and allows for time series at a relatively high spatial resolution. Aside of the promising results compared to GPM IMERG, the daily estimates of the MSG based retrieval are at least comparable to other products incorporated in the IPWG validation study IPWG (2016). A detailed comparison could currently not be given since validation data and strategy were not identical. Incorporation of the presented retrieval scheme to the IPWG validation study is intended by the authors for future assessment.

5 Conclusions

The rainfall retrieval technique developed in this study provides hourly rainfall estimates ~~in~~ at high spatial resolution based on the spectral properties of MSG SEVIRI data and neural networks. The retrieval showed promising results in terms of rainfall area detection and estimation of rainfall quantities. However, the results also showed that the estimation of rainfall remains challenging. The main weakness of the presented retrieval was the overestimation of rainfall ~~areas~~occurrence. However, the retrieval could compete with the ~~global~~ GPM IMERG product in terms of rainfall quantity ~~assignment~~ and was even ~~advantageous~~ better for rainfall area detection.

High resolution spatial datasets of rainfall ~~are~~ is requested by a variety of research disciplines. The developed MSG based rainfall retrieval is able to deliver time series from the launch of MSG SEVIRI onward. An operationalization for near real-time rainfall estimates is intended. It can therefore serve as valuable dataset where high resolution rainfall for Southern Africa

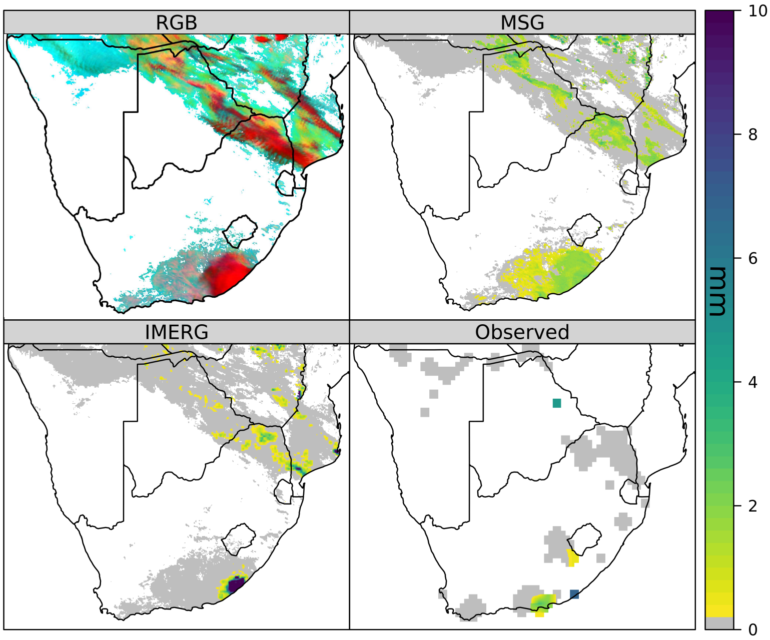


Figure 10. Sample satellite scene from 2014/04/24 ~~12~~10:00 UTC represented as a VIS0.8-IR3.9-IR10.8 false colour composite according to (Rosenfeld and Lensky, 1998) where cloud optical depth is indicated by red colouration, cloud particle sizes and phases in green and the brightness temperature modulates in blue. The rainfall ~~predictions~~ estimates for this scene (estimated using the daytime model) are shown as well as the corresponding GPM IMERG product. Observed rainfall is depicted where weather station data were available. For visualization purposes, the spatial extent of the stations was increased. White background in the colour composite as well as in the MSG based retrieval and the GPM IMERG product represent no data due to missing clouds. In addition, white background in the representation of the observed rainfall is due to the absence of weather stations.

are needed. As an example it will serve as an important parameter within the "IDESSA" (An Integrative Decision Support System for Sustainable Rangeland Management in Southern African Savannas) project that aims to implement an integrative monitoring and decision-support system for the sustainable management of different savanna types. The hourly and aggregated rainfall quantity estimations are available from the authors on request.

Author contributions. H. Meyer and T. Nauss designed the study. J. Drönner preprocessed the satellite data. H. Meyer developed the model code, performed the data analysis and prepared the manuscript with contributions from both co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was financially supported by the Federal Ministry of Education and Research (BMBF) within the IDESSA project (grant no. 01LL1301) which is part of the SPACES-program (Science Partnership for the Assessment of Complex Earth System processes). We are grateful to the South African Weather Service for providing us with their rainfall data for South Africa and to SASSCAL WeatherNet for rainfall data from Namibia and Botswana. The cloud masking was done by using Level-2 data of the CLAAS-2 data record provided by the EUMETSATs Satellite Application Facility on Climate Monitoring (CM SAF). The GPM IMERG V3 data were provided by the NASA/Goddard Space Flight Center's Mesoscale Atmospheric Processes Laboratory and Precipitation Processing System (PPS), which develop and compute the GPM IMERG V3 as a contribution to project GPM, and archived at the NASA GES DISC."

References

Aminou, D. M. A., Jacquet, B., and Pasternak, F.: Characteristics of the Meteosat Second Generation (MSG) radiometer/imager: SEVIRI, in: Proceedings of SPIE: Sensors, Systems, and Next-Generation Satellites, 3221, pp. 19–31, 3221, 1997.

Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., Gerlach, N., and GmbH, M.: Projekt RADOLAN - Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer), Deutscher Wetterdienst, Offenbach, 2004.

Benas, N., Finkensieper, S., Stengel, M., van Zadelhoff, G.-J., Hanschmann, T., Hollmann, R., and Meirink, J. F.: The MSG-SEVIRI based cloud property data record CLAAS-2, Earth System Science Data Discussions, 2017, 1–36, doi:10.5194/essd-2017-9, http://www.earth-syst-sci-data-discuss.net/essd-2017-9/, 2017.

Cattani, E., Merino, A., and Levizzani, V.: Evaluation of Monthly Satellite-Derived Precipitation Products over East Africa, J. Hydrometeorol., 17, 2555–2573, 2016.

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.

EUMETSAT: The Conversion from Effective Radiances to Equivalent Brightness Temperatures, 2012a.

EUMETSAT: Conversion from radiances to reflectances for SEVIRI warm channels, 2012b.

Feidas, H. and Giannakos, A.: Classifying convective and stratiform rain using multispectral infrared Meteosat Second Generation satellite data, Theor. Appl. Climatol., 108, 613–630, 2012.

Finkensieper, S., Meirink, J.-F., van Zadelhoff, G.-J., Hanschmann, T., Benas, N., Stengel, M., Fuchs, P., Hollmann, R., and Werscheck, M.: CLAAS-2: CM SAF CLoud property dAtAset using SEVIRI - Edition 2, Tech. rep., Satellite Application Facility on Climate Monitoring, 2016.

Fynn, R. and O'Connor, T.: Effect of stocking rate and rainfall on rangeland dynamics and cattle performance in a semi-arid savanna, South Africa, J. Appl. Ecol., 37, 491–507, 2000.

Giannakos, A. and Feidas, H.: Classification of convective and stratiform rain based on the spectral and textural features of Meteosat Second Generation infrared data, Theor. Appl. Climatol., 113, 495–510, 2013.

Hamann, U., Walther, A., Baum, B., Bennartz, R., Bugliaro, L., Derrien, M., Francis, P. N., Heidinger, A., Joro, S., Kniffka, A., Le Gléau, H., Lockhoff, M., Lutz, H.-J., Meirink, J. F., Minnis, P., Palikonda, R., Roebeling, R., Thoss, A., Platnick, S., Watts, P., and Wind, G.: Remote sensing of cloud top pressure/height from SEVIRI: analysis of ten current retrieval algorithms, Atmospheric Measurement Techniques, 7, 2839–2867, 2014.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol., 25, 1965–1978, 2005.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The Global Precipitation Measurement Mission, Bull. Amer. Meteor. Soc., 95, 701–722, 2014.

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.

IPWG: IPWG South African Validation, http://rsmc.weathersa.co.za/IPWG/ipwgsa_qlooks.html, 2016.

Kaptué, A. T., Hanan, N. P., Prihodko, L., and Ramirez, J. A.: Spatial and temporal characteristics of rainfall in Africa: Summary statistics for temporal downscaling, Water Resour. Res., 51, 2668–2679, 2015.

Kidd, C. and Huffman, G.: Global precipitation measurement, Meteorol. Appl., 18, 334–353, 2011.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., Hsu, K.-L., and Braithwaite, D.: Intercomparison of High-Resolution Precipitation Products over Northwest Europe, J. Hydrometeor, 13, 67–83, 2011.

Kniffka, A., Stengel, M., and Hollmann, R.: SEVIRI cloud mask dataset - Edition 1 - 15 minutes resolution. Satellite Application Facility on Climate Monitoring, EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF), 2014.

Krogh, A. and Hertz, J. A.: A Simple Weight Decay Can Improve Generalization, in: Advances in Neural Information Processing Systems 4, pp. 950–957, Morgan Kaufmann, 1992.

Kruger, A. C., ed.: Climate of South Africa, Precipitation,, vol. Report No. WS47, South African Weather Service, Pretoria, South Africa, 2007.

Kuhn, M.: caret: Classification and Regression Training, https://CRAN.R-project.org/package=caret, r package version 6.0-68, 2016.

Kuhn, M. and Johnson, K.: Applied Predictive Modeling, Springer, New York, first edn., 2013a.

Kuhn, M. and Johnson, K.: Applied Predictive Modeling, chap. 7.1 Neural Networks, pp. 141 – 145, Springer, New York, first edn., 2013b.

Kühnlein, M.: A machine learning based 24-h-technique for an area-wide rainfall retrieval using MSG SEVIRI data over Central Europe, Ph.D. thesis, Fachbereich Geographie der Philipps-Universität Marburg, 2014.

Kühnlein, M., Appelhans, T., Thies, B., and Nauss, T.: Precipitation Estimates from MSG SEVIRI Daytime, Nighttime, and Twilight Data with Random Forests, J. Appl. Meteor. Climatol., 53, 2457–2480, 2014a.

Kühnlein, M., Appelhans, T., Thies, B., and Nauss, T.: Improving the accuracy of rainfall rates from optical satellite sensors with machine learning - A random forests-based approach applied to MSG SEVIRI, Remote Sens. Environ., 141, 129–143, 2014b.

Levizzani, V., Amorati, R., and Meneguzzo, F.: A Review of Satellite-based Rainfall Estimation Methods, Tech. rep., European Commission Project MUSIC Report (EVK1-CT-2000-00058), 2002.

Manhique, A. J., Reason, C. J. C., Silinto, B., Zucula, J., Raiva, I., Congolo, F., and Mavume, A. F.: Extreme rainfall and floods in southern Africa in January 2013 and associated circulation patterns, Nat. Hazards, 77, 679–691, 2015.

Merk, C., Cermak, J., and Bendix, J.: Retrieval of optical and microphysical cloud properties from Meteosat SEVIRI data at night – a feasibility study based on radiative transfer calculations, Remote Sens. Lett., 2, 357–366, 2011.

Meyer, H., Kühnlein, M., Appelhans, T., and Nauss, T.: Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals, Atmos. Res., 169, Part B, 424–433, 2016.

Meyer, H., Kühnlein, M., Reudenbach, C., and Nauss, T.: Revealing the potential of spectral and textural predictor variables in a neural network-based rainfall retrieval technique, Remote Sensing Letters, 8, 647–656, 2017.

Meyer, H., Kühnlein, M., Reudenbach, C., and Nauss, T.: Revealing the potential of spectral and textural predictor variables in a neural network-based rainfall retrieval technique, Remote Sens. Lett., accepted.

Panchal, G., Ganatra, A., Kosta, Y. P., and Panchal, D.: Behaviour Analysis of Multilayer Perceptrons with Multiple Hidden Neurons and Hidden Layers, International Journal of Computer Theory and Engineering, 3, 332 – 337, 2011.

Prigent, C.: Precipitation retrieval from space: An overview, C. R. Geosci., 342, 380–389, 2010.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2016.

Ripley, B. and Venables, W.: nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models, http://CRAN.R-project.org/package=nnet, r package version 7.3-12, 2016.

Roebeling, R. A., Feijt, A. J., and Stammes, P.: Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17, J. Geophys. Res.: Atmospheres, 111, n/a–n/a, doi:10.1029/2005JD006990, http://dx.doi.org/10.1029/2005JD006990, d20210, 2006.

Rosenfeld, D. and Lensky, I. M.: Satellite-Based Insights into Precipitation Formation Processes in Continental and Maritime Convective Clouds, Bull. Amer. Meteor. Soc., 79, 2457–2476, 1998.

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., Kakar, R., Braun, S. A., Huffman, G. J., Iguchi, T., Kirstetter, P. E., Kummerow, C., Meneghini, R., Oki, R., Olson, W. S., Takayabu, Y. N., Furukawa, K., and Wilheit, T.: The Global Precipitation Measurement (GPM) Mission for Science and Society, Bull. Amer. Meteor. Soc., 2017.

Smith, E., Asrar, G., Furuhama, Y., Ginati, A., Mugnai, A., Nakamura, K., Adler, R., Chou, M.-D., Desbois, M., Durning, J., Entin, J., Einaudi, F., Ferraro, R., Guzzi, R., Houser, P., Hwang, P., Iguchi, T., Joe, P., Kakar, R., Kaye, J., Kojima, M., Kummerow, C., Kuo, K.-S., Lettenmaier, D., Levizzani, V., Lu, N., Mehta, A., Morales, C., Morel, P., Nakazawa, T., Neeck, S., Okamoto, K., Oki, R., Raju, G., Shepherd, J., Simpson, J., Sohn, B.-J., Stocker, E., Tao, W.-K., Testud, J., Tripoli, G., Wood, E., Yang, S., and Zhang, W.: International Global Precipitation Measurement (GPM) Program and Mission: An Overview, in: Measuring Precipitation From Space, edited by Levizzani, V., Bauer, P., and Turk, F., vol. 28 of *Advances In Global Change Research*, pp. 611–653, Springer Netherlands, 2007.

Stengel, M., Kniffka, A., Meirink, J. F., Lockhoff, M., Tan, J., and Hollmann, R.: CLAAS: the CM SAF cloud property data set using SEVIRI, Atmos. Chem. Phys., 14, 4297–4311, 2014.

Thies, B. and Bendix, J.: Satellite based remote sensing of weather and climate: recent achievements and future perspectives, Meteorol. Appl., 18, 262–295, 2011.

Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S, Springer, New York, fourth edn., 2002.

Vicente, G. A., Davenport, J. C., and Scofield, R. A.: The role of orographic and parallax corrections on real time high resolution satellite rainfall rate distribution, Int. J. Remote Sens., 23, 221–230, 2002.

Wöllauer, S., Forteva, S., and Nauss, T.: On demand processing of climate station sensor data, in: EGU General Assembly Conference Abstracts, vol. 17 of *EGU General Assembly Conference Abstracts*, p. 5231, 2015.