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 for Southern Africa for eco-climatological studies or now-casting 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 (SEVIRI). Rainfall measurements from about 350 weather stations from the years 2010-2014 served as ground truths for calibration and validation. SEVIRI and

5 weather station data were used to train neural networks that allowed predicting rainfall area and rainfall quantities during all times of the day. The results revealed that 60 % of recorded rainfall events were correctly classified by the model (Probability of detection, POD). However, the false alarm ratio (FAR) was high (0.80), leading to an Heidke Skill Score (HSS) of 0.18. Predicted hourly rainfall quantities were estimated with an average hourly correlation of rho = 0.33 and a 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

10 the overestimation of rainfall events. The model results were compared to the IMERG product of the Global Precipitation Measurement (GPM) mission. Despite being a comparably simple approach, the presented MSG based rainfall retrieval out-performed 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

15 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, 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

20 a main income and the quality of the pastures mainly depends on water availability (Fynn and O'Connor, 2000). An accurate nowcasting of rainfall in high temporal and spatial resolution 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

5   in Germany, Bartels et al. (2004)). A radar network for entire 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)).

10 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 high spatial and temporal resolution. It can be expected that the GPM products feature a high accuracy as 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 (Kühn-

15 lein et al., 2014b, a; Meyer et al., 2016; Feidas and Giannakos, 2012; Giannakos and Feidas, 2013). Kühnlein et al. (2014b, a); Meyer et al. (2016) presented a methodology to estimate rainfall from optical Meteosat second generation (MSG) 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 Africa. An adaptation of the retrieval technique to South

20 Africa hence requires a model training that relys 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 product.

2   Methods

25

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

2.1   Study area

The area of investigation comprises South Africa, Lesotho and Swaziland, Namibia, Botswana, Zimbabwe as well as parts of

30 Mozambique (Fig. 2). Average yearly rainfall sums in Southern Africa roughly follow an aridity gradient from the dry west to the more humid east. With the exception 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 that may last over several days. The majority of interior areas
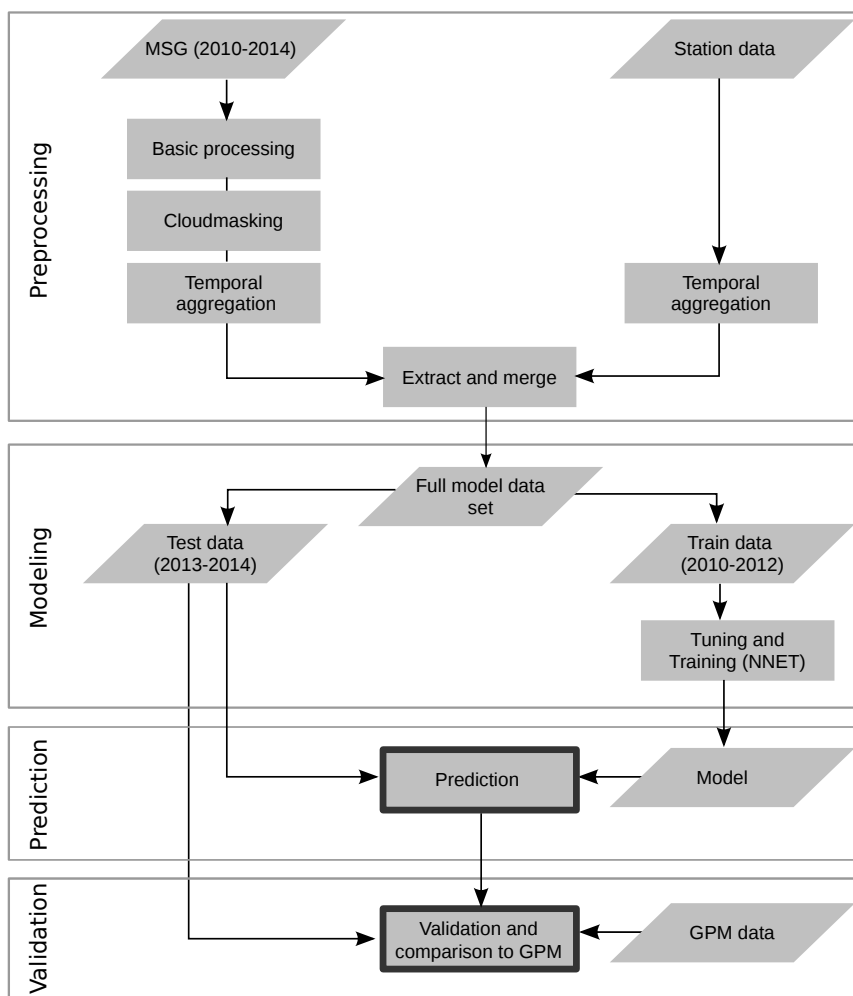
Figure 1. Flow chart of the methodology applied in this study.

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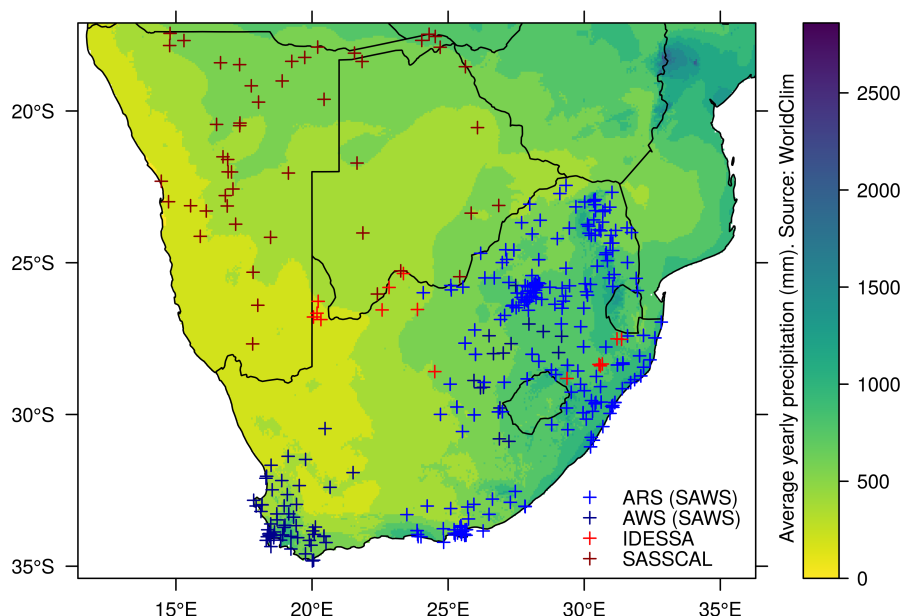
Figure 2. Map of the average yearly precipitation sums in the study area as estimated by WordClim (Hijmans et al., 2005). Points show the locations of the weather stations that were used as ground truth data in this study. Automatic rainfall stations (ARS) and automatic weather stations (ARS) are operated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

precipitation sums. 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).

2.2 Data and Preprocessing

2.2.1 Station data

5   Rainfall data for the years 2010 to 2014 were obtained from the South African Weather Service (SAWS). The data were recorded from 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. All station 10   data that provided sub-hourly information were aggregated to a temporal resolution of 1 hour. Though the station data is

not randomly distributed in the model domain, it covers the entire aridity gradient, from sites with very low (< 200mm) precipitation to sites in areas with highest (~ 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 km at sub-satellite point. 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. 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 (see https://github.com/umr-dbs/xxl).

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 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) who used the spectral channels of MSG SEVIRI to train a Random Forest model that is able to spatially predict 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 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. 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. (Submitted) tested different texture parameters as additional predictor variables and have shown that these spectral channels are sufficient as predictors. 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) and a nighttime dataset (reliable VIS and NIR not available). The response variables (rainfall yes/no and

5   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 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

10   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

The neural network implementation from the "nnet" package (Venables and Ripley, 2002) in R was used in conjunction with the "caret" package Wing et al. (2016) that provides enhanced functionalities for model training, prediction and validation.

15   Model parameters were tuned using a stratified 10-fold cross-validation where each fold has the same distribution of rainfall areas, or rainfall quantities respectively, as the entire training dataset. The number of hidden units were tuned for each value between two and the number of predictor variables (Kuhn and Johnson, 2013). Weight decay was tuned between 0 and 0.1 with increments of 0.02. For training of rainfall areas, the threshold that separates rainy from non-rainy clouds according to the predicted probabilities was an additional tuning parameter. The optimal threshold was expected to be considerably smaller

20   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.

2.4   Validation

Model predictions and weather station records from the entire year 2013 were used as independent data for model validation.

25   For the validation of predicted 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 for the respective MSG pixel. The validation data contained 403211 samples during daytime and 565415 samples during nighttime. Average hourly probability of detection (POD), Probability of false detection (POFD), False alarm ratio (FAR) and Heidke skill score (HSS) were calculated as validation metrics. The POD gives

30   the percentage of rain pixels that the model correctly identified as rain (Tab. 1, 2). POFD gives the proportion of non-rain pixels that the model incorrectly classified as rain. The false alarm ratio (FAR) gives the proportion of predicted 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 1. Confusion matrix as baseline for the calculation of the verification scores used for the validation of the rainfall area predictions.

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

Table 2. Categorical metrics for validation of rainfall area predictions.

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	$HSS = \frac{TP*TN-FP*FN}{[(TP+FN)*(FN+TN)+(TP+FP)*(FP+TN)]/2}$	$-\infty$ - 1	1

To evaluate the ability of the model to predict rainfall quantities, the correlation between the measured and the predicted hourly rainfall was calculated using Spearmans rho to account for a non-normal distribution of the data. Further, the root mean square error (RMSE) was used. All clouded 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
5 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 worldwide high resolution precipitation estimates (Smith et al., 2007). GPM provides data from March 2014 onwards. The GPM IMERG
10 product estimates rainfall by combining all available passive-microwave instruments as well as microwave-calibrated infrared satellite estimates and data from rainfall gauges. GPM IMERG is available in 6h, 18h and 4 month latency.

In this study the 4 month latency (final product) with 30 minutes temporal and $0.1°$ spatial resolution ($\sim$10km x 10km) was used. 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
15 temporal resolution of the MSG predictions. 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 (Fig. 3). The probability of false detection was low (18 %) but the predictions featured a high false alarm ratio of 0.80. The average HSS per scene was 0.18.

The average hourly RMSE was 0.72 mm 4). Correlation indicated by Spearmans rho was 0.33 in average. The performance of the rainfall quantities assignment increased with the aggregation level (Fig. 5). 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 observed especially when aggregated to monthly rainfall sums. An example of temporally aggregated rainfall predictions are shown for the year 2013 in Fig. 6.

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. 7). 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 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. 8). 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. 9 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 of the MSG based retrieval that predicted rainfall for the areas with high values of optical depth. This was partly confirmed by the weather station data. However, rainfall was also predicted 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 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).
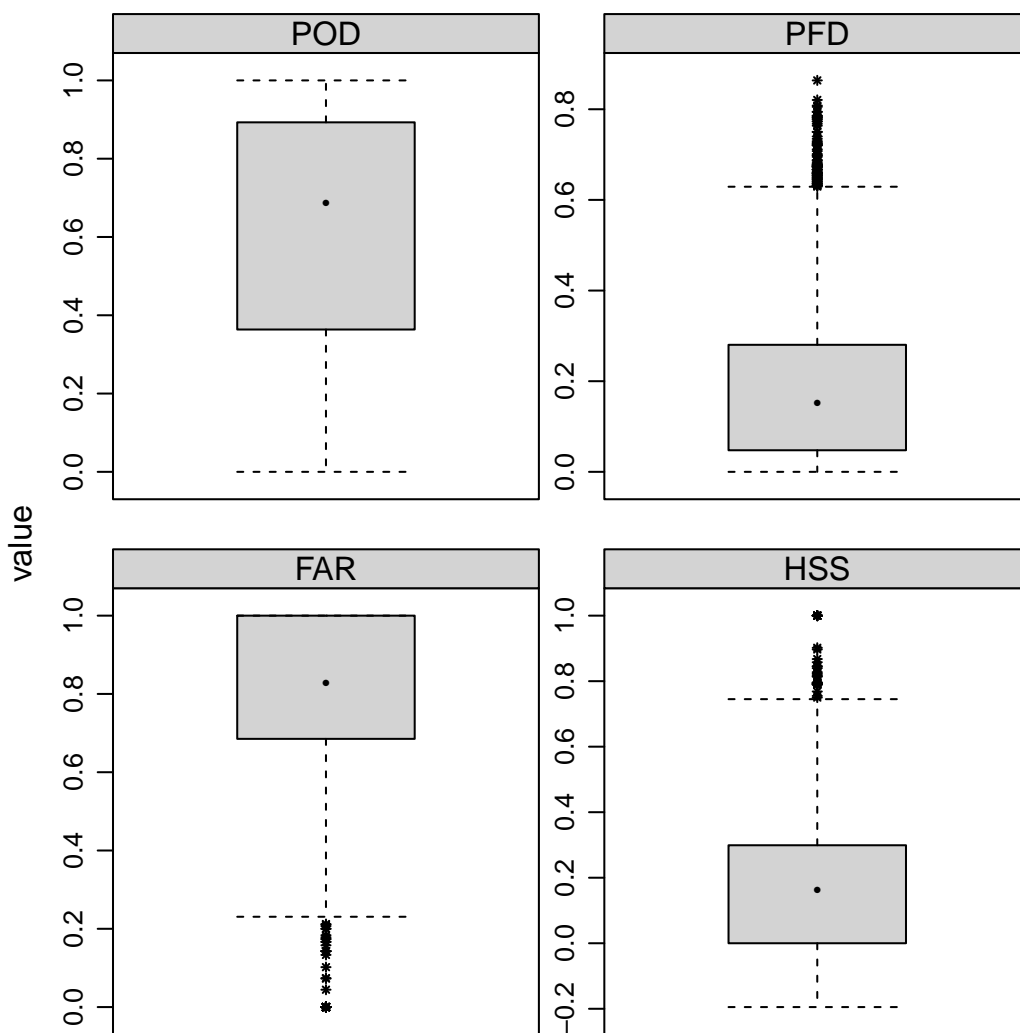
Figure 3. Validation of predicted rainfall areas for the year 2013 on an hourly basis. Each data point is the average performance of one hour.

The validation of the rainfall retrieval showed promising results but highlights also the difficulties of optical satellite-based rainfall estimates. The major problem was the overestimation of rainfall events leading to an overestimation of rainfall quantities. In this context, it is of note that the FAR can easily increase to elevated levels in dry conditions when there are just a few false alarms in the predictions and no rainfall was observed by any station. However, the FAR was still high for hours with

5   a considerable number of rainfall events. This might be partly explainable by spatial displacement due to parallax shifts that affect model training as well as the predictions. For future enhancement of the rainfall retrieval, a correction of the parallax
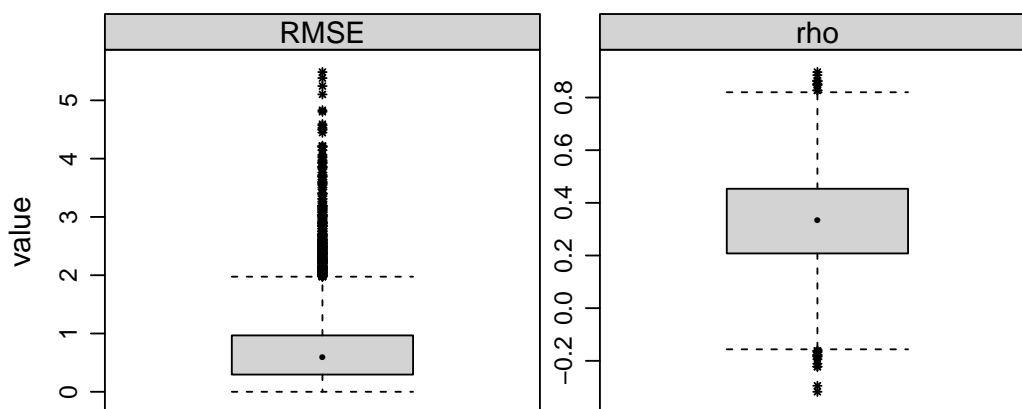
Figure 4. Validation of predicted rainfall quantities for the year 2013 on an hourly basis. Each data point is the average performance of one hour

shift (Vicente et al., 2002) would be appropriate. Differences in spatial and temporal scale are also an important issue especially since a 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). However, no reliable alternatives are available and rain gauge measurements are still considered as most reliable source of rainfall data.

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 (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 IMERG has more differentiated rainfall estimates while the MSG based retrieval tends to predict 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 compa-
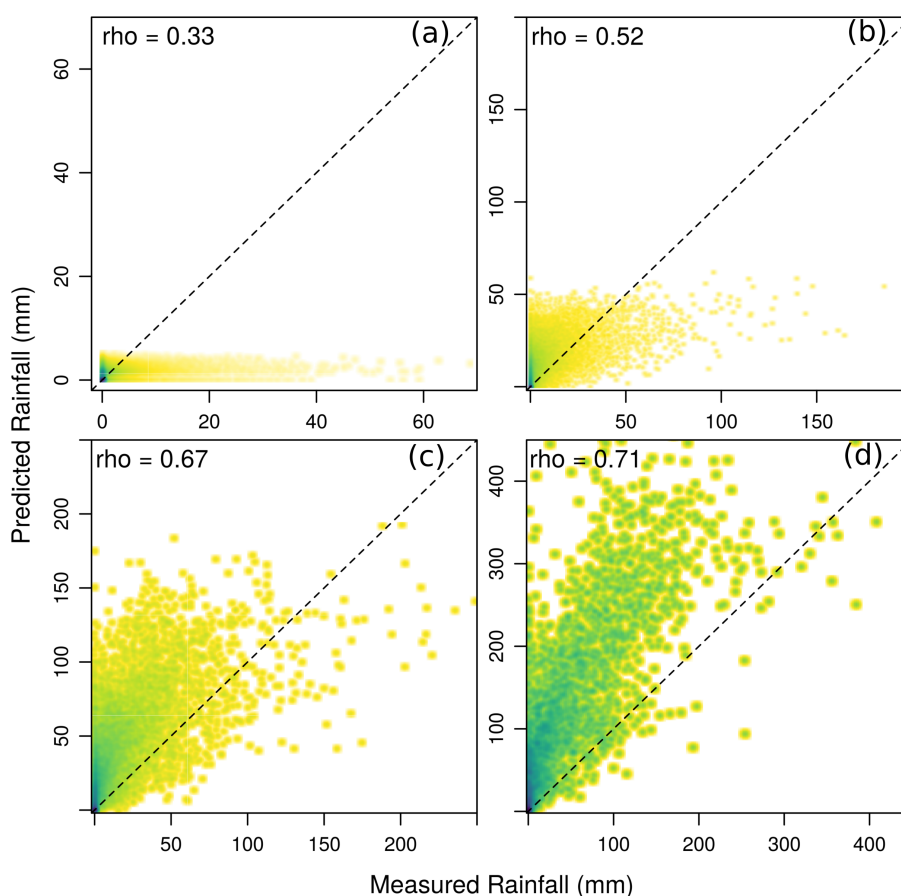
Figure 5. Validation of predicted rainfall quantities for the year 2013 on (a) hourly resolution and on the different aggregation levels (b) daily, (c) weekly, (d) monthly. Each data point represents a station at the respective level of temporal aggregation. Rho represents the average correlation for each time step of the respective aggregation level. The intensity of the colour represents the data point density.

rable 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 to the IPWG validation study is intended by the authors for future assessment.

5 Conclusions

5   The rainfall retrieval developed in this study provides hourly rainfall estimates in 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
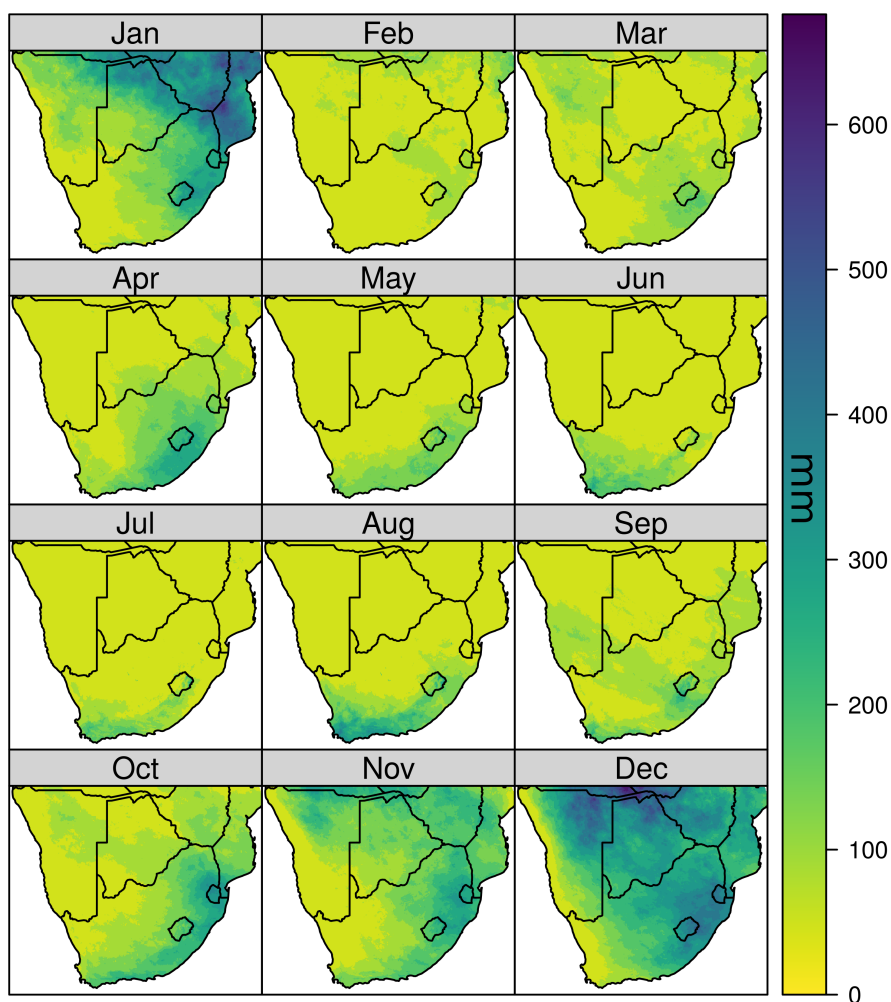
Figure 6. Monthly precipitation sums in mm of the year 2013 as estimated by this study.

main weakness of the presented retrieval was the overestimation of rainfall areas. However, the retrieval could compete with the global GPM IMERG product in terms of rainfall quantity assignment and was even advantageous for rainfall area detection.

High resolution spatial datasets of rainfall are 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 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
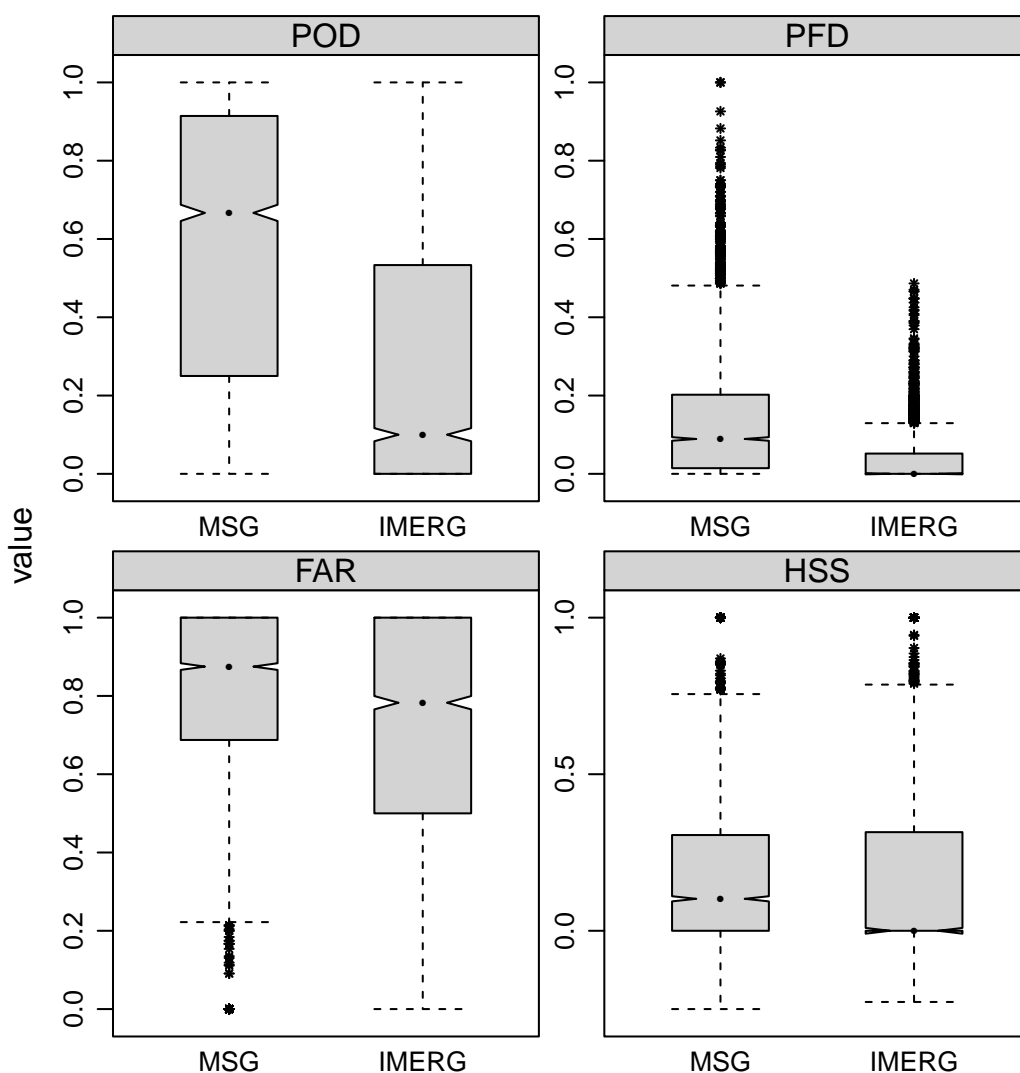
Figure 7. Comparison of the performance of the MSG based retrieval and GPM IMERG for rainfall area delineation between March and August 2014. Each data point is the average performance of one hour. The notches serve as a rough estimation of significant differences.

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 and performed the data analysis. H. Meyer prepared the manuscript with contributions from both co-authors.
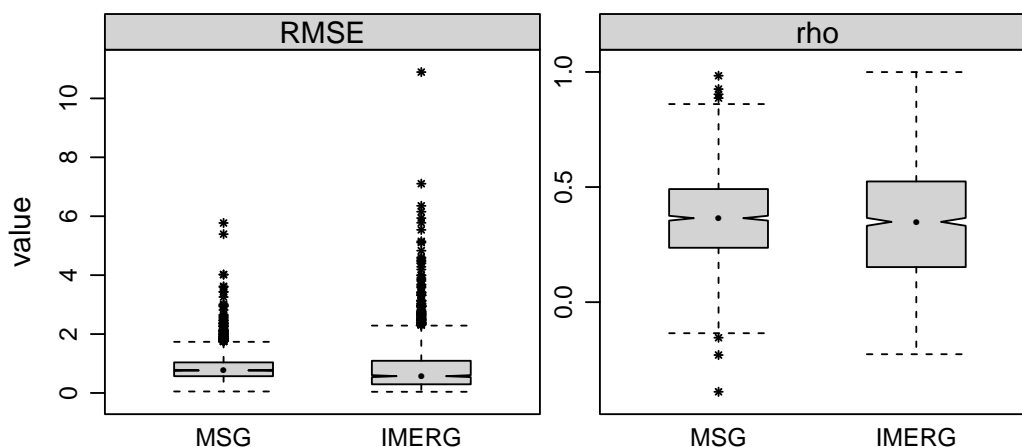
Figure 8. Comparison of the performance of the MSG based retrieval and GPM IMERG for hourly rainfall quantities between March and August 2014. Each data point is the average performance of one hour. The notches serve as a rough estimation of significant differences.

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
5 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).
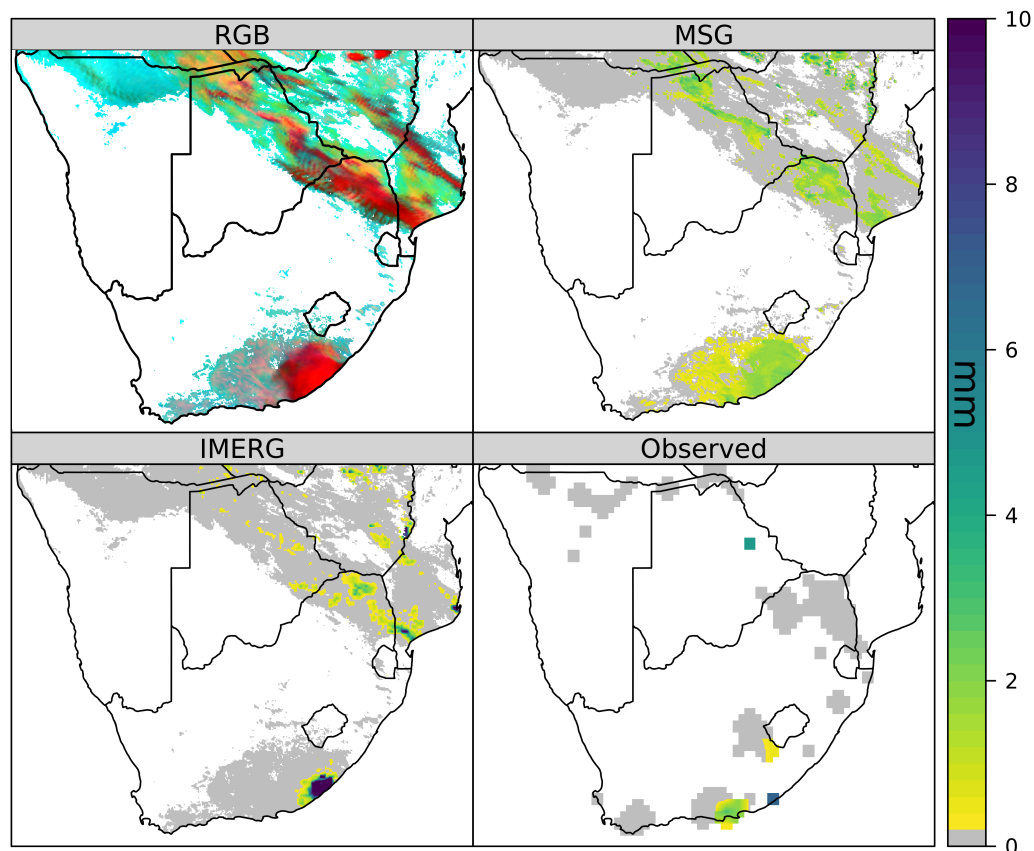
Figure 9. Sample satellite scene from 2014/04/24 12:00 UTC represented as a VIS0.8-IR3.9-IR10.8 false colour composite according to (Rosenfeld and Lensky, 1998) where cloud optical depth is indicated by red colouration, cloud particle sizes and phases in green and the brightness temperature modulates in blue. The rainfall predictions for this scene 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.

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.

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

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.

5 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.

10 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.

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.

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

15 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.

20 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.

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

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

25 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
30 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.

35 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 texture predictor variables in a neural network based rainfall retrieval technique, Remote Sens. Lett., Submitted.

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.

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.

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.

Wing, M. K. a. J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., and Candan., C.: caret: Classification and Regression Training, https://CRAN.R-project.org/package=caret, r package version 6.0-68, 2016.