RainRunner

The role of water vapor observations in satellite-based rainfall information highlighted by a Deep Learning approach

by

Fabio Curzi

to obtain the degree of Master of Science
Track: Environmental Engineering
at the Delft University of Technology,
to be defended publicly on June 28, 2022.

Student number:	5136148	
Project duration:	September, 2021 – June, 2022	
Thesis committee:	Dr. Ir. Marie-Claire ten Veldhuis,	TU Delft
	Prof. dr. ir. Nick van de Giesen,	TU Delft
	Dr. Riccardo Taormina,	TU Delft
	Mónica Estébanez Camarena,	TU Delft

The role of water vapor observations in satellite-based rainfall information highlighted by a Deep Learning approach

FABIO CURZI[a] , MONICA ESTEBANEZ CAMARENA[a] , MARIE-CLAIRE TEN VELDHUIS[a] , NICK VAN DE GIESEN[a] , RICCARDO TAORMINA[a]

[a] *Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands*

Corresponding author: Fabio Curzi, fcurzi@tudelft.nl, Monica Estebanez Camarena M.EstebanezCamarena@tudelft.nl

ABSTRACT: Rain-fed agriculture is the main source of food in Ghana therefore improving quantitative rainfall estimates is essential for local farmers to predict crop growth using vegetation models. Rainfall dynamics in the tropics is an ongoing topic of research due to their complexity and sub-grid precipitation variability. At the same time, tropical areas such as Ghana are the most affected by a lack of proper rain gauge network coverage. Traditional methods rely on empirical assumptions and statistical theories that require continuous calibration and still struggle to accurately represent local variability. The aim of this paper is to demonstrate the potential of a Deep Learning (DL) approach using bi-spectral information of water vapor imagery (WV) and thermal infrared (TIR) as a starting point to develop an effective alternative to the Cold Cloud Duration (CCD) method which is a widely applied statistical technique by satellite rainfall products like Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) and Tropical Applications of Meteorology using SATellite data (TAMSAT) that are specifically designed for Africa.

WV inhibition of low-level features assures the correct depiction of strong convective motions usually related to heavy rainfall which is crucial in tropical areas where convective rainfall is dominant. The addition of WV 7.3μm is particularly beneficial in North Ghana as tropical systems are advecting dry air from the nearby Sahara desert creating discontinuities in precipitation events which translates into dry intrusions and dry slots seen in the images of the WV channel.

The developed Deep learning model showed strong performances in binary classification where it outperformed IMERG-Final false alarms count resulting in lower rainfall overestimation (FBias < 2.0), although further research is needed to overcome the very poor relation between GEO-IR images and actual rainfall estimates at the surface.

1. Introduction

Precipitation plays a crucial role in Ghana's agriculture which accounts for most of the total country's economy: 54% of the total Gross Domestic Product (FAO 2020). This applies to Ghana as well as other regions in the world, where rain-fed agriculture is predominant. Rainfall dynamics in tropical regions are still an ongoing topic of research due to their complexity and small-scale drastic variations of convective precipitation. At the same time, tropical areas are the most affected by a lack of proper rain gauge network coverage (Bechtold 2019) (Coz and van de Giesen 2020). Tropical rainfall is primarily influenced by seasonality. Northern Ghana has a uni-modal rainfall regime, which means there is only one maximum (peak) per seasonal cycle that usually happens in the months of July and August. Dry season in this region starts in November until late March, during this period of time there are virtually no significant precipitation events (Knippertz and Fink 2008). Rainfall in this region is a complex process governed by the seasonal northward shift of the Inter-tropical Convergence Zone (ITCZ) and the West African Monsoon (WAM), a low-level south westerly moist flow from from the Atlantic ocean.

Wind shear generated by the monsoonal flow creates a strong temperature contrast between the extremely dry hot Sahara desert and the cool moist Guinea coast, this contrast exists mostly during the summer months from June to September. This extreme temperature gradient favours the formation of the African Easterly Jet, an exceptional tropical feature associated with the West African monsoon. It is a unique zonal wind feature located in the mid-level troposphere around 600 hPa (Figure 1.3) and is most intense at the end of August. The meridional temperature contrast previously mentioned induces this jet via thermal wind balance promoting the development of the African Easterly waves (AEWs) through baroclininc and barotropic instability. Due to small temperature gradients (typically <1 K/1000 km) all the other tropical regions elsewhere are generally void of jet streams (Bechtold 2019).

Ground measurements are essential for common tropical rainfall products specifically designed for Africa such as TAMSAT and CHIRPS (Funk et al. 2015). Both products make use of thermal infrared images (TIR) from geostationary satellites and assume there is a positive linear correlation between the length of time a cold cloud top resides on a given pixel and the amount of rainfall at the surface. This is the basic working principle of the Cold Cloud Statistics. Different temperature thresholds (T_b) for the cloud tops are tested within a certain area for a dekadal or pentadal (v3.0) period, linear regression is applied between gauge measurements and the cold cloud fields for each threshold. Once the optimum T_b is found, calibration parameters (slope and intercept) are derived using the median rainfall rate from the gauge observations (Tarnavsky et al. 2014). Temporal resolution of such products is daily for TAMSAT and 6-hourly for CHIRPS. Results from a calibration of the CCD method applied in the Sahel region has shown inadequate results due to spatial averaging and temporal aggregation ass well as low gauge density leading to less reliable calibration (Dybkjær 2003).

The most accurate rainfall observations from space are the ones that make use of Passive Microwave sensors (PMW), as emission and scattering of MW radiation by rain droplets gives a more direct estimation of precipitation-sized particles. Major drawback of this retrieval method is that observations in MW spectrum are carried out by LEO satellites, which means there are only two observations per day per satellite. Complex algorithms are applied to merge and translate infrequent observations into a high-resolution gridded rainfall product, an example is the Tropical Rainfall Measuring Mission (TRMM) later evolved into Integrated Multi-satellitE Retrieval for GPM (IMERG) (Kidd and Huffman 2011).

Parametrization of rainfall is a challenging process and susceptible to a huge amount of conditions especially in West Africa where higher land temperatures and higher aerosols concentration offset conventional precipitation dynamics. Many research (Tomassini et al. 2017) (Tomassini 2020) (Berry and Thorncroft 2005) have shown that moist convection is the main support for the intensification of African Easterly Waves (AEWs), latent heat release from condensation of water vapor in the atmosphere is the key promoter together with strong solar irradiation in generating unstable atmospheric conditions that leads to sparse precipitation events in the form of thunderstorms. Two MSG water vapor channels (WV), 6.2μm and 7.3μm, are sensitive to water vapor content at different altitudes dependent on the intrinsic property of water vapor absorption spectrum.

The advent of big data has promoted a spike in machine learning applications within the hydrological domain. The objective of Deep learning is to minimize human intervention and facilitate automated feature extraction from large raw data sets (Shen et al. 2021). This new data-oriented approach could be a valid method to detect and possibly estimate rainfall when theoretical or process-based approaches fail to accurately parameterize such complex atmospheric processes. The present study aims to expand and improve the existing RainRunner model that currently makes use solely of the MSG SEVIRI 10.8μm (TIR) imagery by adding the 7.3μm (WV) channel. The use of WV channels could bring useful information into rainfall dynamics that take place in West Africa, considering that there is still a general lack of products that would allow an efficient use of these channels in a synoptic scale analysis (Georgiev and Santurette 2009).

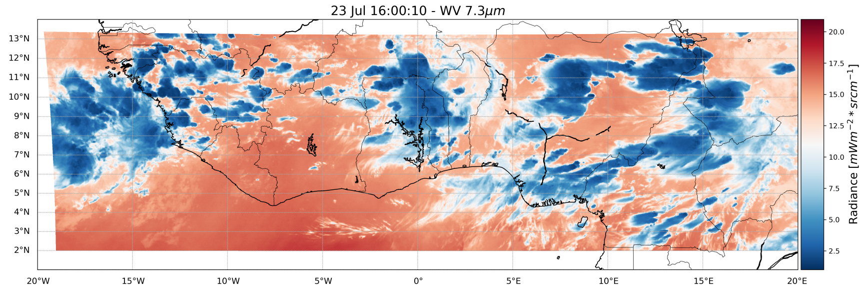
2. Data and study area

This section briefly introduces the data sources used for this study along with a short description of the region of interest.

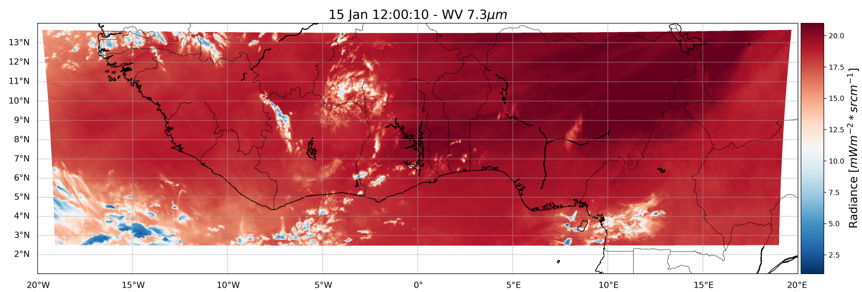
a. Input data

The Spinning Enhanced Visible and InfraRed Imager (SEVIRI) instrument that is onboard Meteosat Second Generation (MSG) located at 0 degree, captures images every 15 minutes of the Earth surface using 12 different spectral bands and has a sampling distance of 3km for the infrared channels. The images used in this study are geolocated and radiometrically pre-processed to Level 1.5 (ESA 2016).

At 6.2μm (Channel 5), the radiation is promptly absorbed by water vapor as this wavelength is located in the center of water vapor absorption band. The use of this channel is only limited to observe mid-to upper level water vapor dynamics. This research will therefore make use exclusively of Channel 6 (7.3μm) radiation. Images in 7.3μm channel are able to detect water vapor content further down in the atmosphere and are useful to interpret humidity features associated to mid-level jets in strong convective environment (Figure 1). This is valid for Ghana where the rainy season is heavily dependant on a mid-level air current that effectively transports moisture horizontally. An important characteristic of water vapor imagery is the inability in detecting low-level clouds like stratocumulus or nimbostratus clouds in moist environments as they are normally located below the effective layer. Only with a very dry troposphere the water vapor channel is able to reach such low levels (e.g eastern Sahara desert and Antarctica) (Selami et al. 2021).



(a) Eastward moisture transport during boreal summer under the influence of mid-level jets.



(b) Dry low-level Harmattan wind blowing from Sahara desert slightly visible during dry season.

FIG. 1: Difference between rain (a) and dry (b) season as seen in 7.3μm imagery.

b. TAHMO network

TAHMO stations data are used as target features for this rainfall estimation model. The main advantage of such stations is the simple installation process and the robustness while still being able to deliver reliable atmospheric measurements without continuous calibration (TAHMO 2016). The exact location of the stations in the northern region of Ghana is displayed in Figure 3. Time range of ground measurements used for this study starts in July 2018 until December 2020 and the temporal resolution of the data before preprocessing is hourly. Faulty measurements and missing data could occur if there is no signal or the rain gauge gets clogged. Figure 2 shows the amount of missing data per station during the selected time period. Given the high number of missing data in certain weather stations, only the ones that have at least two full consecutive years of observations will be considered to analyse rainfall patterns in the ground data analysis. However the model is still trained using all the available and complete sequences from the eight stations.

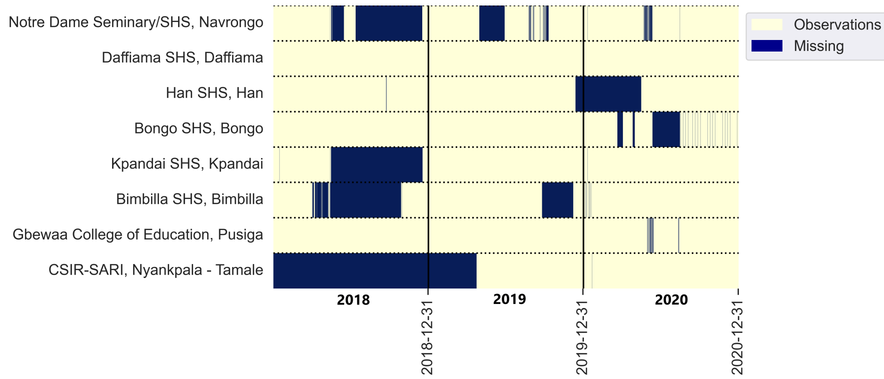


FIG. 2: Missing data per each TAHMO station in north Ghana in 2018-2019-2020.

c. Reference model

IMERG: The baseline model used in this study is IMERG. It is a rainfall product developed by NASA and its goal is to use as many Low Earth Orbiting (LEO) satellites as possible in combination with different geosynchronous earth orbit satellites IR data to fill in gaps between PMW measurements. Gauge analysis from the Global Precipitation Climatologic Centre (GPCC) are also used to correct any bias at local scale. The GPM Core Satellite is used as both calibration and evaluation tool for the PMW and IR-based products in IMERG. The multiple inputs coming from different sources are combined into a 'best' data set, which requires a space and time resolution corresponding to PMW spatial scale (0.1°) and IR temporal scale (30 minutes). The algorithm is decomposed in different stages which start with the intercalibration of microwave estimates, upsampling to finer scale using Kalman filters and finally use IR estimates to fill missing data from PMW products. Gauge data is the final stage and is intended to control local bias. There are three different products available and they have different latency starting from IMERG-Early with a latency of 4 hours, IMERG-Late has a latency of 12-hours then IMERG-Final adds gauge estimates and has a considerable higher latency of 3.5 months (Tan et al. 2019).

d. Study area: Northern Ghana

The study area of this research is the northern region of Ghana located within the latitudes 8°N and 11°N and longitudes 3°W and 0°30'E. West Africa region considered as a whole has one of the most extreme climatic gradients in the world. Rainfall is by far the most significant climatic element of West Africa, the mean annual rainfall steadily increases southward towards the equator with extremes ranging from near zero in the arid part of the Sahel up to over 2000mm/year in the coastal zones (Nicholson 2013).

On average, this part of Ghana is more under the influence of the hot and arid North Easterly trade wind which blows air that comes from the Sahara desert usually carrying a considerable amount of dust, while the southern part of the country receives more maritime influx through moist SW winds.

Vegetation in North Ghana region is mostly a Guinea Savanna with many croplands that relies on rain-fed irrigation. Nicholson (2013).

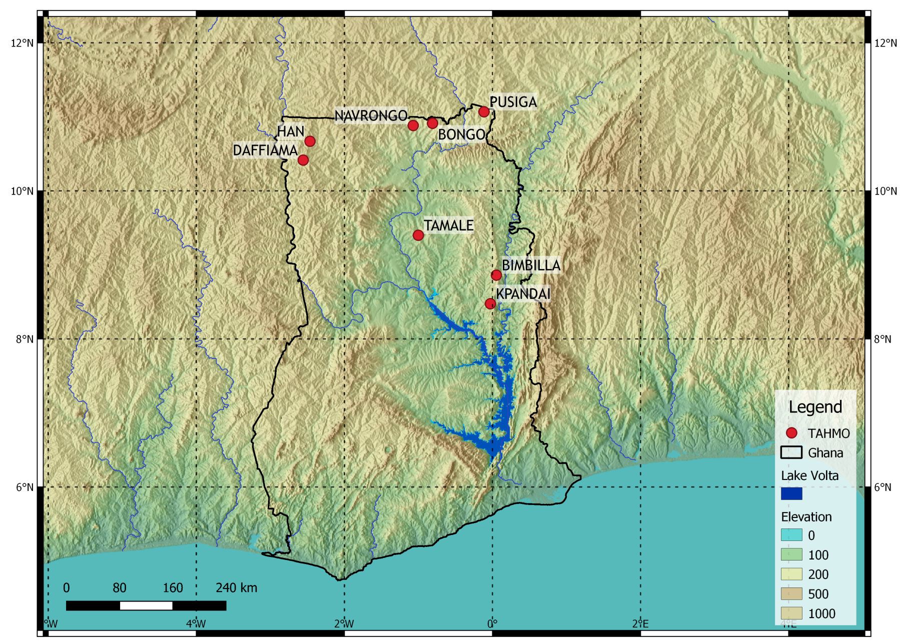


FIG. 3: Digital elevation map of the study area (GRASS QGIS). Overview of the TAHMO stations. Data retrieved from https://www.usgs.gov/

3. Methodology

Rainfall detection described in this study can be formulated as a supervised imbalanced binary classification problem where images from the MSG satellite are used as input into the model and TAHMO measurements are the target labels (ground-truth data) to distinguish rain from no-rain sequences.

a. Ground data analysis

The first step of this study is a thorough data analysis of TAHMO hourly data with at least two full consecutive years of observations (no missing data for at least 66% of the considered period 2018-2020) are used to analyse rainfall dynamics within the case study region.

Four stations were selected for this purpose out of the eight available stations: Daffiama (TA00251), Pusiga (TA00264), Bongo (TA00254), Kpandai (TA00259). Analysis of the precipitation patterns like seasonality, median rainfall duration and diurnal precipitation cycle play an important role in the selection of the temporal scale of the model and for the integration of additional inputs such as the timestamp.

b. Data preprocessing

Sparse ground data pose a challenge to any ML-based model as they perform best when there is a dense grid data network. The methodology described in this section presents a way to overcome the lack of ground data by using an image to point approach. The model is trained only using point rainfall observations as the large distance between stations prevent us from using any type of spatial interpolation without introducing big sampling errors. For this purpose, both satellite images of TIR and WV are cropped to create a matrix of 32x32 pixels (covered area is $96x96km$) with the TAHMO station located in the central square. The spatial resolution of the model corresponds to the pixel size, which is approximately 3.1km (Camarena et. al 2022). Cropped images are then grouped together to form 3-hrs sequences, each sequence is made of 12 images. The chosen temporal resolution is in-line with the rainfall duration pattern explained in the previous section. Integrity of the sequences is mandatory, if any sequence includes missing data it is discarded from the process.

Hourly TAHMO ground measurements and IMERG data with 30-minutes temporal resolution were accumulated to 3-hrs intervals to match the temporal scale of the sequences. A threshold of $1mm/3h$ following the American Meteorology Society classification has been established to distinguish between rain or no-rain sequences which corresponds to the class of moderate drizzle or very light rain.

To correctly represent the timestamp as an actual cyclical feature representative of diurnal variability we map each cyclical variable onto a circle such that the lowest value for that variable appears right next to the largest value. By doing this we are taking into account jump discontinuities (e.g from 11pm to midnight), hence the single value needs to be transformed into a two dimensional-array using sine and cosine transformation. The following relations provide the correct feature encoding:

$$X_{sin} = sin(\frac{2\pi * X}{max(X)}) \tag{1}$$

$$X_{cos} = cos(\frac{2\pi * X}{max(X)}) \tag{2}$$

c. Dataset construction and division

The dataset is highly skewed. This means that the number of no-rain sequences is much larger than rain sequences. An hybrid approach of data resampling and weighted loss function is applied. The dataset is split in such a way that the rain sequences in the training dataset are over-sampled with a ratio of 4:1 dry/rain, while both validation and test dataset have a more realistic ratio of 28.2:1 dry/rain which is representative of the 2020 full data. Training dataset contains sequences of 2018, 2019, 2020 while validation and test have only sequences of 2020. The table below shows the number of sequences available per dataset.

Dataset	**Year**	**Dry samples**	**Rain samples**	**Total n_samples**	**Ratio dry/rain**
Training	2018, 2019, 2020	4218	1055	5273	3.998
Validation	2020	6627	235	6862	28.2
Test	2020	6627	235	6862	28.2

TABLE 1: Number of samples per dataset with respective year.

d. Model development

1) ARCHITECTURE DESIGN

The inputs of the model are two different streams of twelve $32x32x1$ matrices for a total of 24 input images, one stream contains the IR channel information while the other one contains WV channel.

The model used for this study is RainRunner, the architecture of the model is similar to the one already described in Camarena et al. 2022 with a few adaptations such as an increase in the number of nodes from 8 to 16, this increase is proportional to the number of input images (from 12 to 24). Figure 4 illustrates a condensed diagram of the bi-spectral model structure. The inputs of WV and TIR are convoluted separately in order to learn information individually from each channel. The output of the convolution and pooling layers is a 2-dimensional (8x8x1) single tensor generated from each image of the sequence which means two convolutions are applied in series. The tensors are flattened and concatenated into a multilayer perceptron. The timestamp (month and time of the day) is added directly into the fully connected layer after preprocessing along with the 2-D tensors from the convolutional layers. The total numbers of learnable parameters is 11,019,197, the batch size is set to 64 and the learning rate is fixed to 0.0001. The number of passes trough the training dataset is fixed at 300 epochs with an early stopping callback set to 50 to halt the training in case the model was overfitting. The function for the dense layer(s) is a rectified linear function (ReLu) while the output layer function is a logistic function, called sigmoid, which returns a probabilistic output between 0 and 1, where 1 represents 100% rain and 0 is 100% dry. A decision boundary line at 0.5 is used for the classifier to make a distinction between the two classes. Lastly, a weighted loss function is applied to deal with the imbalanced dataset where dry sequences have 0.2 and rain sequences 0.8 coefficients which reflects the ratio of dry/rain sequences of the training dataset.

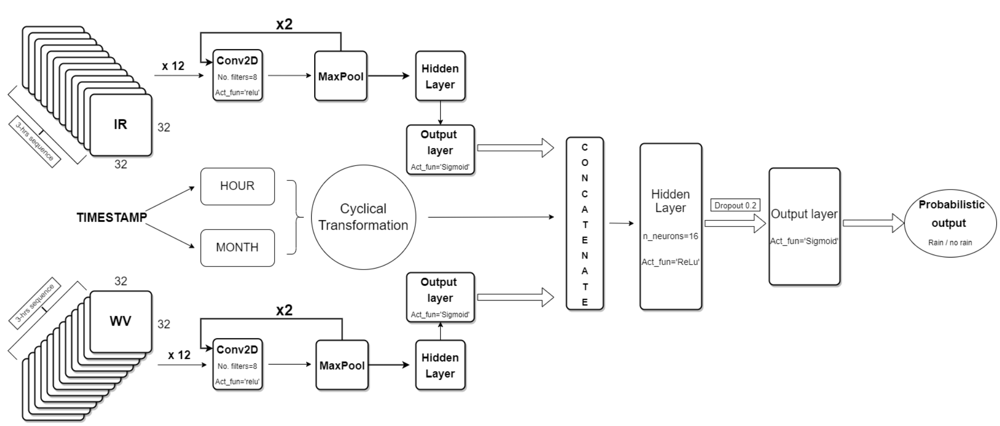


FIG. 4: Schematic overview of the proposed bi-spectral (WV+TIR) RainRunner architecture.

e. Performance evaluation

1) Assessment of data contribution

Four models are tested in this study, in order to assess the contribution of water vapor and timestamp they all have the same architecture/hyperparameters and number of samples, the only difference is the input source which is intended to highlight the contribution of water vapor imagery. The diurnal heating cycle and seasonality patterns are closely tied to rainfall in this region, this is the main reason for the inclusion of timestamp information into the model.

1. RainRunner 10.8μm (TIR)
2. RainRunner 7.3μm (WV)
3. RainRunner 10.8μm + 7.3μm (TIR + WV)
4. RainRunner 10.8μm + 7.3μm (TIR + WV) + Timestamp

Each model was trained 10 times. Ensemble average is applied to every model in order to reduce variance of the output and make the model predictions more stable. Model averaging is an approach to ensemble learning where each member contributes an equal amount to the final prediction. The stochastic nature of a machine learning model makes the probabilistic output subject to fluctuations that generate uncertainty within the same prediction, however when applying an ensemble average there is more coherence between the input sequence and the output which highlights the underlying value of the input sequence. Two different sets of metrics were used to evaluate the models performance: the first set of metrics consists of common DL metrics used to benchmark the model performance on a given binary classification problem. Accuracy, precision, recall and F1-score are all derived from the confusion matrix which shows the number of correctly (Hits and Correct negatives) and incorrectly classified sequences(False alarms and Misses). However, accuracy is not a good measure when dataset is very imbalanced, as the model might still reach a high level of accuracy by only detecting the majority class for all the sequences. F1-score becomes valuable in this type of problems, it represents the harmonic mean of precision and recall. Hence, the best averaged models are ranked according to the highest F1-score.

$$F_1 = 2 * \frac{precision * recall}{precision + recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (3)$$

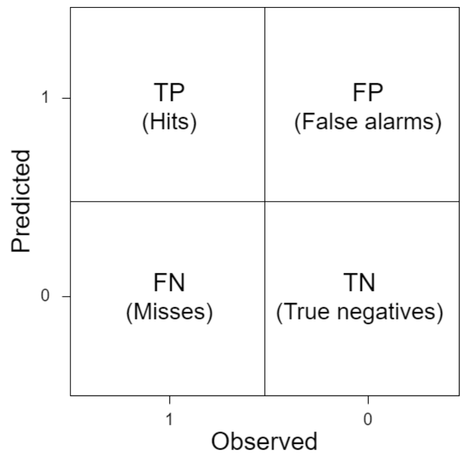


FIG. 5: Square confusion matrix used for this rain/no-rain binary classification problem.

The variance of the 10 different runs is shown by means of box plots where the quantiles of a probability distribution of a variable is displayed. The box contains the values between the upper quartile and lower quartile (50% of the distribution). The second set of metrics aims to compare different model performances using categorical evaluation metrics:

$$POD = \frac{TP}{TP + MS}, \quad \text{Range: } 0 \text{ - } 1, \quad \text{Target: } 1 \tag{4}$$

$$SR = 1 - FAR = 1 - \frac{FP}{TP + MS}, \quad \text{Range: } 0 \text{ - } 1, \quad \text{Target: } 1 \tag{5}$$

$$CSI = \frac{TP}{TP + FP + MS}, \quad \text{Range: } 0 \text{ - } 1, \quad \text{Target: } 1 \tag{6}$$

$$Bias = \frac{POD}{SR}, \quad \text{Bias} < 1\text{: Under-forecast}, \quad \text{Bias} > 1\text{: Over-forecast} \tag{7}$$

These metrics are geometrically related and can be used to construct the Roebber diagram, a visual assessment of the forecast quality of the four different models which is in general preferable for ease of interpreting the statistics (Roebber 2009). The distribution of misclassified sequences is also investigated for an in-depth understanding of the models strengths and weaknesses, also to pinpoint any improvement brought by the inclusion of the timestamp into the model. On this matter, the misclassification analysis consists of misclassifications per time of the day, per month, per season and per TAHMO station. Rain sequences were also grouped in different categories according to the rainfall accumulation per hour as recorded by the TAHMO station. Different rainfall categories are listed in Section 3f

2) WV contribution using pixel analysis

Pixel analysis has a significant role in this study and is intended to demonstrate the actual differences between the two spectral channels of the satellite and why it is preferred to use a bi-spectral channel approach over single channel. A top-down approach is applied which means firstly large scale MSG images from the two satellite channels - 10.8μm and 7.3μm - are compared against each other to have a visual aid of the water vapor exclusion of low-level non-convective features hidden by the West African monsoon during rain season. The images spans longitudinally across West Africa from 20°W to 20°E showing atmospheric conditions at synoptic scale.

The second part of the pixel analysis is focused on highlighting the differences at smaller scale (mesoscale) using cropped MSG images from relevant sequences used to validate the predictions of the model. Firstly, the pixel values are normalised between 0 and 1 then the pixel distribution of that sequence is shown using a gray-level histogram. A gray-level histogram indicates how many pixels of an image share the same pixel value. Each pixel value corresponds to a certain measured radiance that can be related to the equivalent brightness temperature of a layer. We know that temperature is not constant with height, if the atmosphere is conditionally unstable there is a negative temperature lapse rate between the Earth surface and a layer at $height = Z$ that can be simplified using the following relation:

$$\Gamma = -\frac{dT}{dz} \tag{8}$$

In raw satellite imagery, pixel radiances with values approaching the unity are bright pixels and they translate into absorption at lower levels of the atmosphere, which corresponds to higher temperatures. The effective layer is then located at low levels. Darker pixels have values closer to 0, which indicates colder temperatures of the effective layer and therefore its location will be at higher altitude. Since the gray-level of each pixel in WV imagery gives information about the layer depth, an example using 3-dimensional surface plot is aimed at better showing convective motions of a violent rain event as seen from both channels that happened in Pusiga is also provided. A number of other meaningful events are selected according to the misclassified probabilistic output value of the averaged models. A second criterion applied to find more events was to calculate the mean pixel value of TIR sequence and check whether the WV mean pixel value was at least a standard deviation away from TIR mean value.

f. Rainfall intensity estimation

1) Model setup

The estimation of rainfall intensity requires a slightly different model setup, since there are more than two possible outcomes, the problem becomes a multiclass classification. This type of classification requires a function that is able to computes a discrete categorical distribution of K possible categories. The most common approach, that is also used in this study, is to apply the softmax function in the output layer.

In order to reduce bias in the model, all the no-rain sequences were discarded from this type of classification. Following the definition of rain in the Glossary of Meteorology of the American Meteorological Society (AMS), rainfall categories were defined as:

- **Very light rain**: $1mm/3h < RR < 1mm/h$
- **Light rain**: $1mm/h < RR < 2.5mm/h$
- **Moderate rain**: $2.5mm/h < RR < 7.6mm/h$
- **Heavy rain**: $RR > 7.6mm/h$

	Very light rain	Light rain	Moderate rain	Heavy rain
Training	412	268	271	104
Validation	79	68	58	30
Test	66	64	70	35

Table 2: Distribution of precipitation events per each rainfall category in training, validation, test.

The category 'Very light rain' is firstly introduced in (Camarena et al. 2022) and is used here as additional category between dry and light rain sequences. The model is trained only using rain sequences divided per rainfall category. Results of the multiclass classification can be found in Appendix A Figure A1 as additional material.

4. Results

a. Rainfall dynamics in West Africa

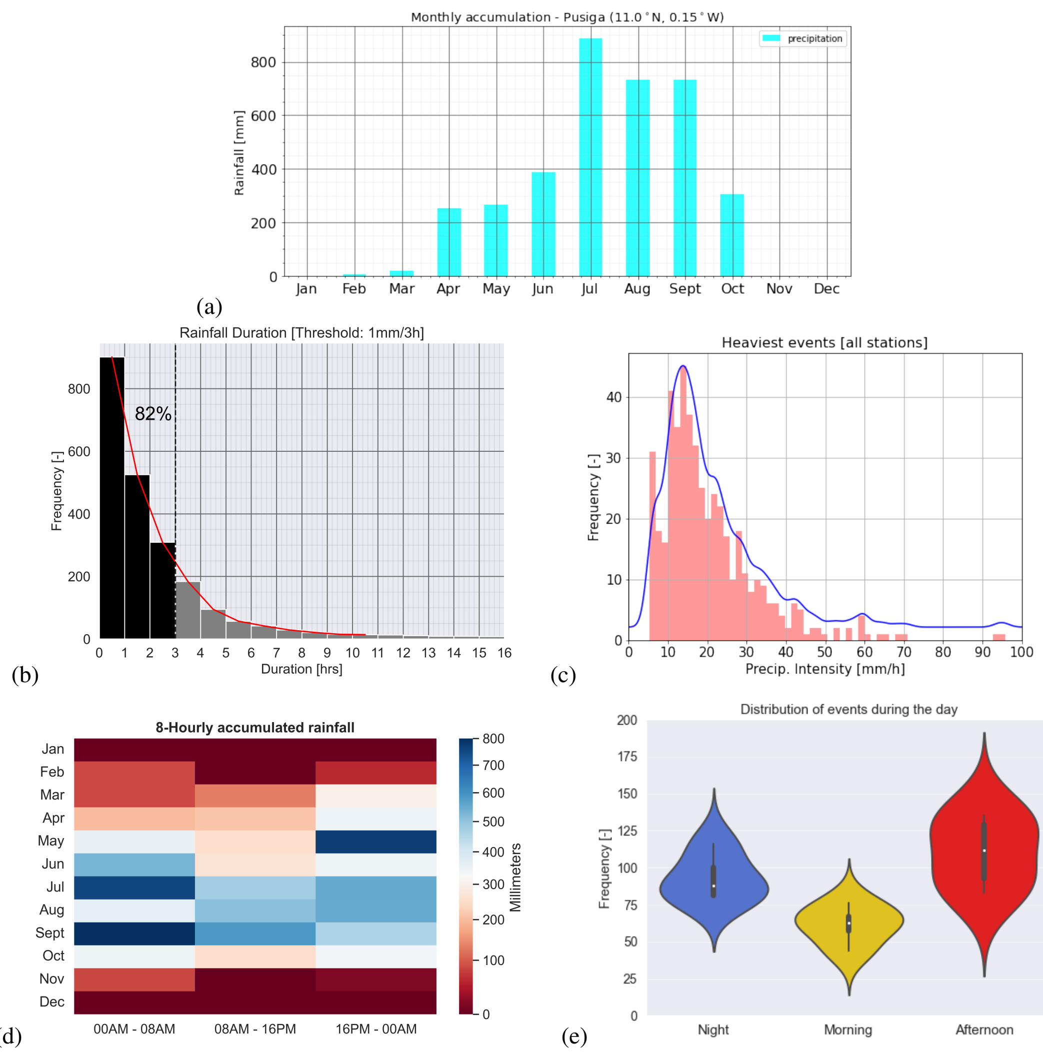


FIG. 6: Rainfall patterns in North Ghana based on hourly data from four TAHMO stations. (a) Rainfall duration. (b) Precipitation intensity. (c) Rainfall accumulation per time of the day. (d) Frequency of precipitation events ($RR > 1mm/3h$) based on time of the day.

Measurements from TAHMO stations are processed in Figure 6 to visualize the main characteristics of the rainfall regime in the region of interest. It can be seen that precipitation resembles the main characteristics of a convective rainfall regime which translates into seasonal heavy short-lived thunderstorms, Figure 6b and Figure 6c clearly show this. Most of the rainfall events (82%) do not last more than 3-hrs and the median value of the heaviest rainfall events is close to 20mm/h which indicates very heavy rainfall regime. 3-hrs is the temporal scale of each sequence and it allows to keep track of the development of the storms.

Figure 6d shows a progressively erratic diurnal cycle of convection during the rain season starting from May due to the strengthening of the AEJ consequently enhancing horizontal moisture transport (visible in 7.3μm) and formation of large mesoscale systems which are propagating overnight resulting in large accumulated rainfall values. This pattern culminates in early September where almost 1000mm falls during nighttime. Morning hours (6am - 12pm) have generally the least rainfall accumulation as well as fewer numbers of precipitation events, stable atmospheric conditions are more often found around this time of the day.

b. Model performance for independent test dataset

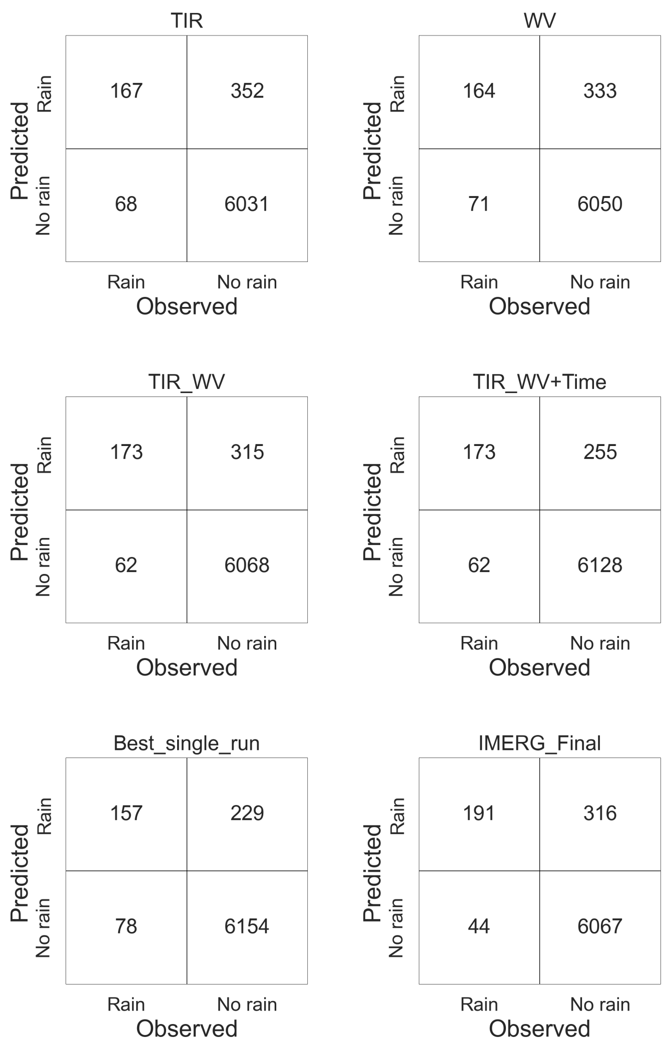


Fig. 7: Contingency tables of ensemble averaged RainRunner using different inputs (WV, TIR, WV+TIR, WV+TIR+Time, the single best run and IMERG-Final.

Confusion matrices are displayed in Figure 7. Initially WV and TIR separated performs similarly, TIR model has a slightly lower number of missed rain events whereas the WV model outcome showed less false alarms. Combining them together leads to a fewer number of misclassified rain and dry sequences. It is observed that the number of false alarms (false positives) is constantly decreasing when water vapor and timestamp are included in the model, from 352 down to 255. IMERG on the other hand has much lower number of misses (false negatives), which makes sense considering the fact that it makes use of a constellation of LEO satellites that have a more direct relationship to rainfall. The best single run has the lowest number of false alarms (229) at the expenses of a high number of misses (78), which corresponds to a third of the all rainy sequences.

c. Roebber performance diagram

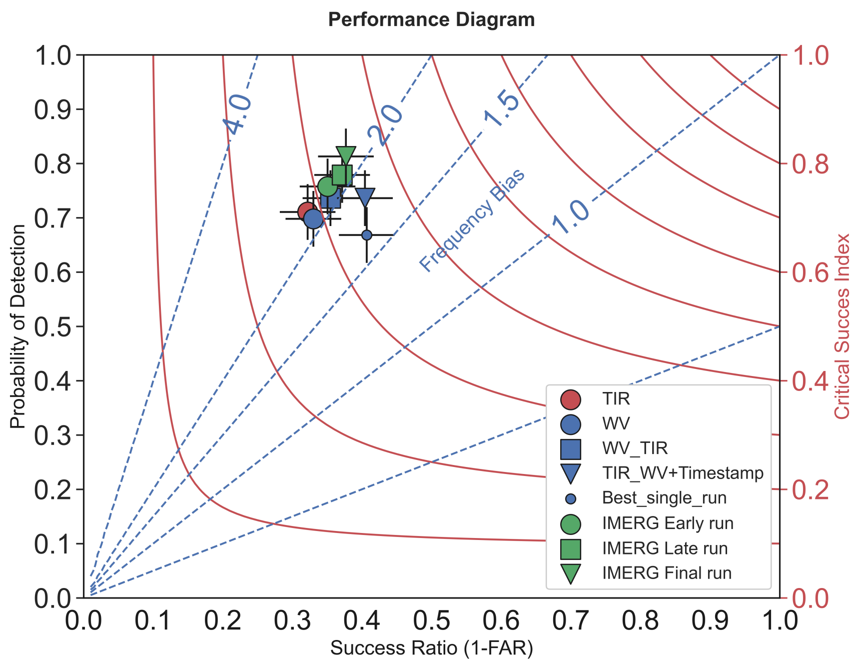


FIG. 8: Roebber performance diagram on the test dataset.

The values shown on the contingency tables are used to construct the Roebber performance diagram in Figure 8, where all IMERG products are plotted as reference models.

A perfect forecast would show values of POD, SR and CSI approaching unity value and it would be placed on the upper-right corner of the diagram. IMERG-Final has the highest number of hits (true positives) as a consequence it also has the highest POD of all the models, however with a bias well above 2.0 it is over-forecasting rainfall. The small performance increase between IMERG-Final and IMERG-Early is not enough to justify the great difference in latency time between the two products. The benefit of adding WV and timestamp is noticeable in this diagram as it progressively leads to higher success ratio (SR) as well as a lower bias compared to all other models ($1.5 < bias < 2.0$).

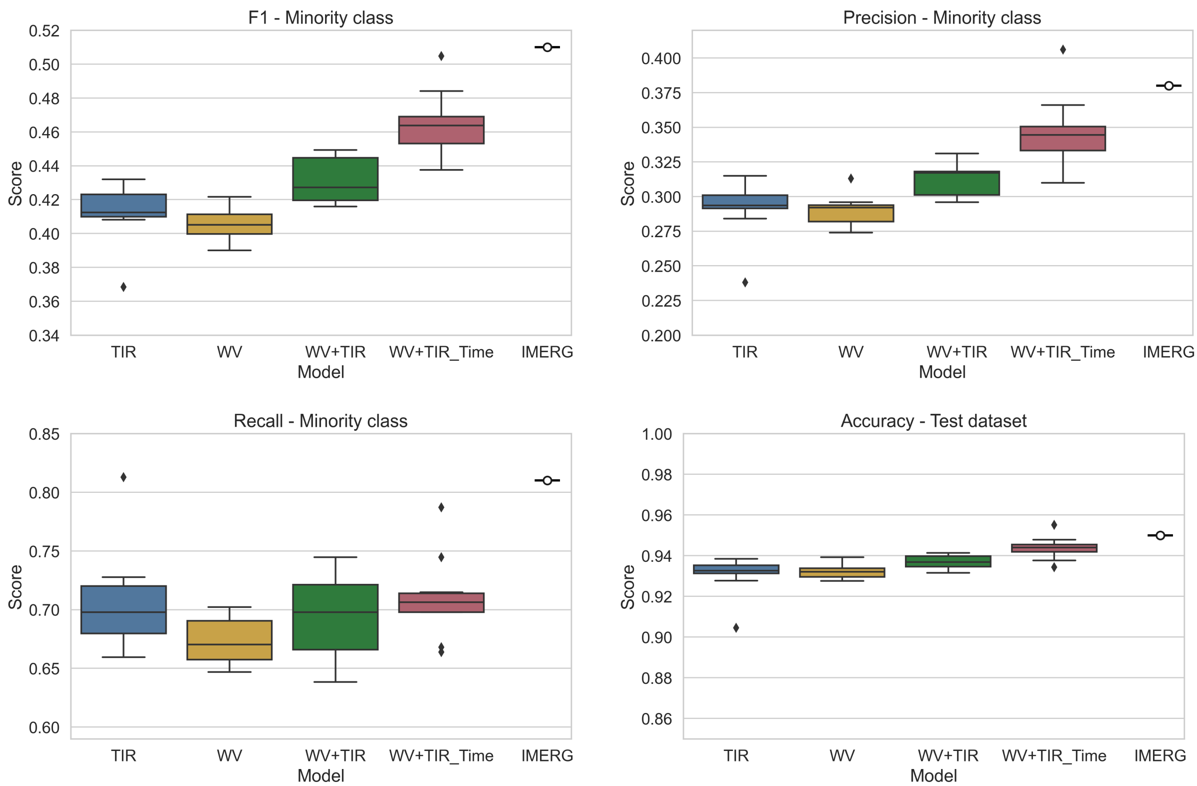


FIG. 9: Variance of the performance metrics between different runs of each model.

The variance of the 10 runs in the minority class is shown in Figure 9. Given the highly skewed nature of the dataset towards zero (dry events), all models have a high accuracy score that varies in a tight range between 93% - 95% except for one outlier in TIR. F1-score is a more indicative score for this type of classification problem. The two initial models of WV and TIR have similar performances although TIR shows slightly higher F1-scores. The combination of the two channels into a single model shows some improvements with a clear upward trend of performances that culminate at a 0.50 on the F1-score achieved on the best single run.

As precision score is inversely proportional to the number of false positives (false alarms), the gradual reduction of this number results in an increased precision which in turns improves the overall F1-score, being the harmonic mean of precision and recall. Recall is related to the number of false negatives and has a median value of approximately 0.70 for all models except for WV and IMERG which scored 0.67 and 0.81 respectively.

These runs are then averaged together to generate an ensemble probabilistic output that gives a more stable prediction in accordance to each input sequence of the test dataset.

d. Misclassification analysis

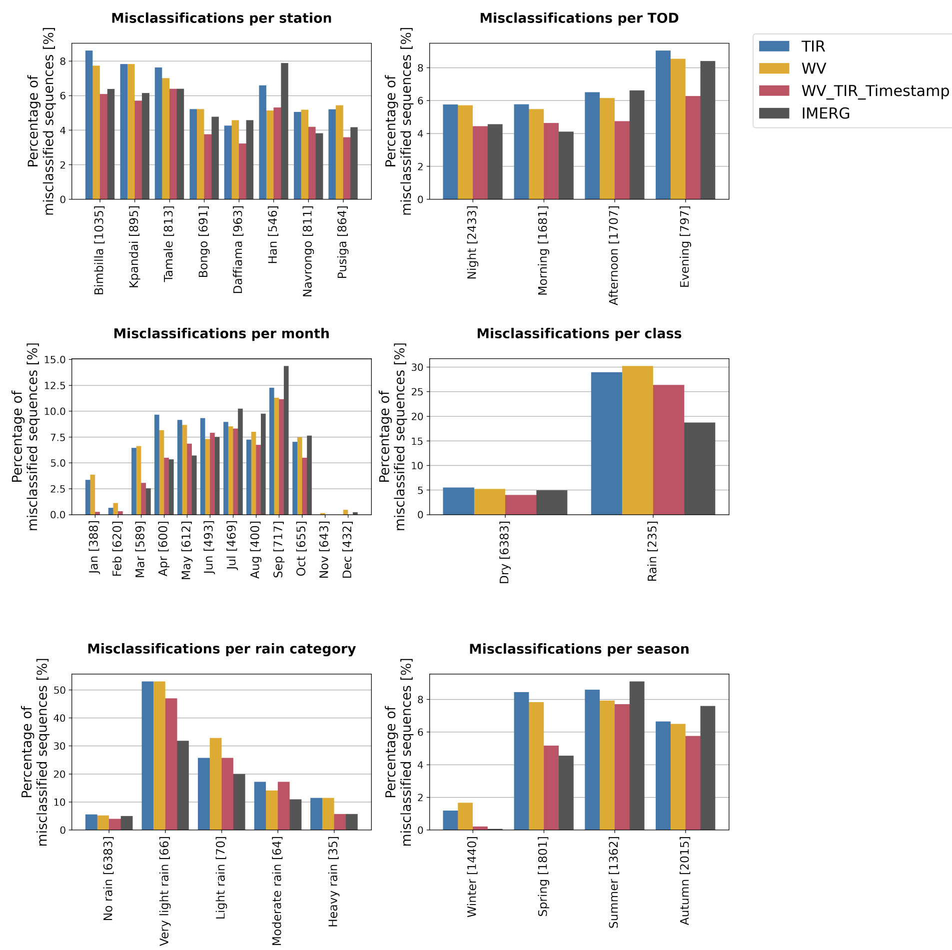


FIG. 10: Misclassification analysis based on selected parameters: Station, time of the day, month, class, rain category and season.

Figure 10 displays the distribution of misclassified sequences using defined parameters of interest. Northernmost stations have overall less misclassified sequences compared to stations located in central regions. The combination of WV and TIR with timestamp has the least number of misclassifications overall. The addition of timestamp is particularly valuable during dry season, here the combined model with timestamp has very little number of misclassifications. Rain season (boreal summer) shows the poorest performances for all models. It is worth to mention that IMERG has the highest number of incorrectly classified sequences during the rain season, highlighting the fact that the AEJ influence on rainfall patterns is a true challenge even for the most advanced models. WV model being used to track convective motions is the one struggling the most when it comes to light and very light (stratiform) rain detection.

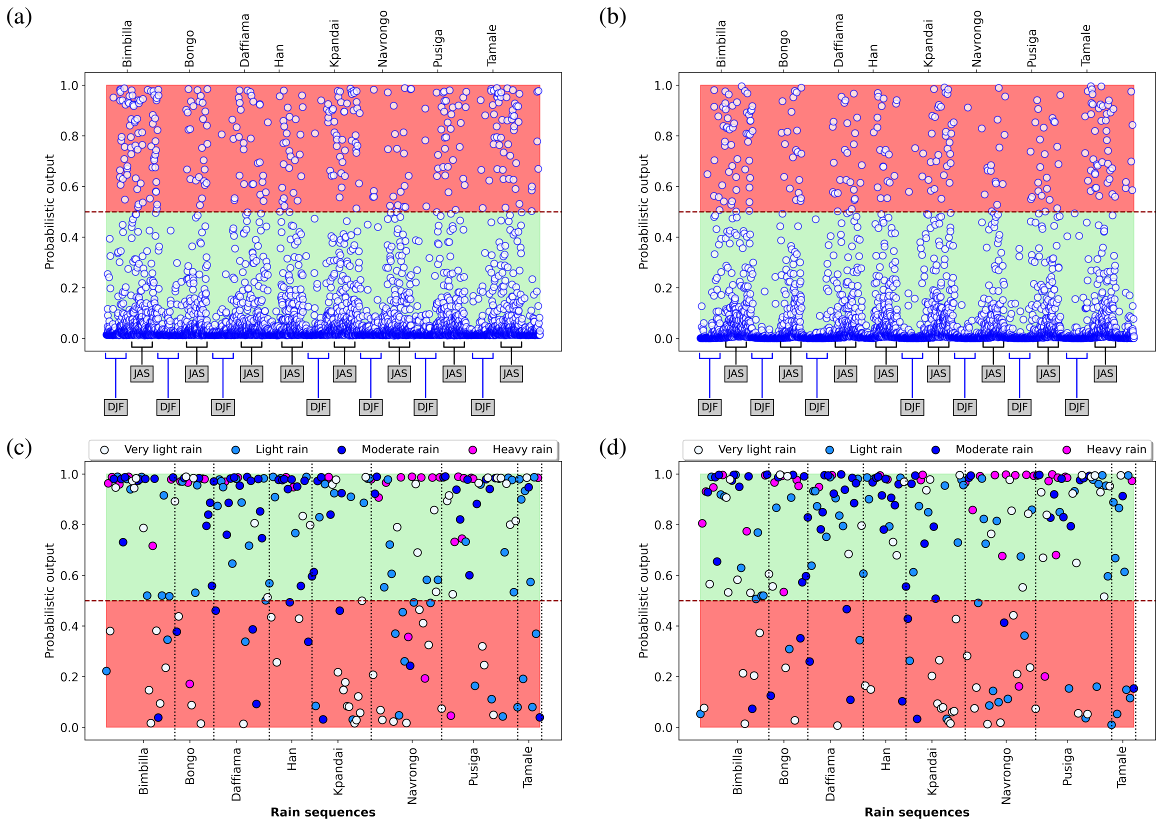


FIG. 11: Comparison of the ensemble probabilistic output of the test dataset sequences. (a) TIR ensemble probabilistic output of dry sequences; (b) TIR + WV + Timestamp ensemble probabilistic output of dry sequences; DJF is representative of the dry season while JAS is representative of rain season. (c) TIR (d) TIR + WV + Timestamp ensemble probabilistic output of rain sequences

Figure 11 illustrates the contribution of the timestamp information in the model by comparing the probabilistic output of the combined model + Timestamp with RainRunner 10.8μm. The line at 0.5 is the decision boundary for the rain/no-rain distinction. It is more visible to distinguish seasonal changes across every station by looking at the model output that uses timestamp. The addition of number of the month makes the predictions for the trimester December-January-February (DJF) much lower with values close to 0, the mean predicted value for TIR was 0.14 vs. 0.005 for the model using timestamp. On the other hand, dry events during July-August-September (JAS) are the still the most difficult to classify on both models. The addition of the time of the day is beneficial during early rain season when the AEJ is not yet offsetting the diurnal convective cycle, and rainfall is still occurring during late afternoon hours.

Figure 11c-d shows how TIR predictions of rain sequences are closer to unity than the combined model with timestamp. This is particularly true for some rain events that occurred during the shoulder season (March/April or October/November) received a lower probabilistic prediction in the model using timestamp. Four heavy rainfall events were misclassified in TIR while only two heavy events are misclassified in the combined model. This is probably due to WV capture of strong convective motions associated to heavy rainfall.

e. Pixel analysis comparison

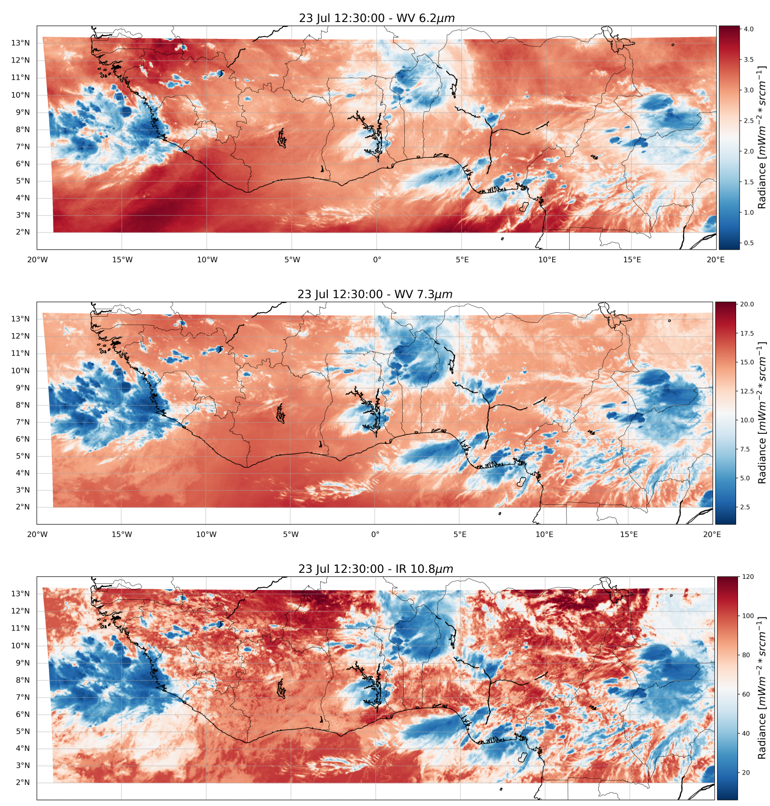


FIG. 12: Visual assessment of the three relevant SEVIRI infrared channels. From top to bottom: WV 6.2 μm (Channel 5); WV 7.3 μm (Channel 6); IR 10.8μm (Channel 9).

Satellite images over large areas are useful to understand the differences between the thermal window channel 10.8μm and the two water vapor channels. Figure 12 shows a snapshot of West Africa atmospheric dynamics on July 23 2020 at noon. Midday is the time at which solar heating cycle is at its peak and early convection is visible. The image retrieved at 10.8μm shows a lot of information not always related to rainfall, such as many low-level clouds spread across the whole region and where the sky is clear the radiance is an indicator of the land surface temperature (dark red area on the upper part of the figure is near Sahara desert). Areas of intense convection are highlighted in water vapor imagery and are located within the dark blue areas. The softer red shade shown in 7.3μm is clearly the top of the West African Monsoon layer that acts as threshold level for this channel, hiding low-level clouds. Above this level the AEJ transport moisture eastward and promotes slanted convection. The largest sensitivity range for channel 5 6.2μm is around 350 hPa, which makes this channel completely blind to the WAM as well as most of the lower level features associated with it. It is still a useful channel to locate deep convective motions that takes place in the upper-troposphere.

Figure 13 displays some of the analysed misclassified sequences used as input where the addition of water vapor proved to be functional for the model and reflected some insightful atmospheric dynamics. The images on the left side are extracted directly from the input sequence and are representative of the atmospheric event. On the right hand side the gray-level histogram shows the pixel distribution of the entire sequence. A grey level histogram indicates the frequency of occurrence of each gray-level value in the sequence.

Starting from top to bottom, Figure 13a shows a clear dry intrusion that happens when a tropical system advects air from a dry source, it is visible from the sharp gradient in water vapor imagery. Dry intrusions normally happen right after a precipitation event and they are difficult to locate in TIR imagery because warmer clouds linger for a longer period of time.

A dry slot is seen in 13b, dry slots can be a consequence of dry intrusions or they might happen along the transition zone between convective and stratiform rain in larger mesoscale convective systems. Figure 13c is a sequence from January 2020 (peak dry season) that was misclassified by WV but thanks to TIR we know there were no rainy clouds at that moment. This happens when an anomalous low level moist southerly circulation peaks up during certain days of the dry season while at higher levels dry air is present. This allows the 7.3μm channel to retrieve water vapor content from lower levels resulting in incorrect predictions. Two distinct peaks are observable in each gray-level histogram meaning that the two channels generate an asymmetric bimodal distribution of pixels at different brightness temperatures, in the case of WV imagery the peak is an indication of the most occuring height of the effective layer during the sequence.

EVENT	Groundtruth	TIR	WV	Combined	Timestamp
(a) Kpandai_2020.09.30_18	0	0.60	0.18	0.48	0.47
(b) Bimbilla_2020.05.27_09	0	0.51	0.10	0.42	0.21
(c) Tamale_2020.01.23_18	0	0.42	0.64	0.32	0.14
(d) Pusiga_2020.05.27_12	1	0.04	0.16	0.45	0.20

TABLE 3: Predicted probabilities from each ensemble model for the selected events in Fig 13.

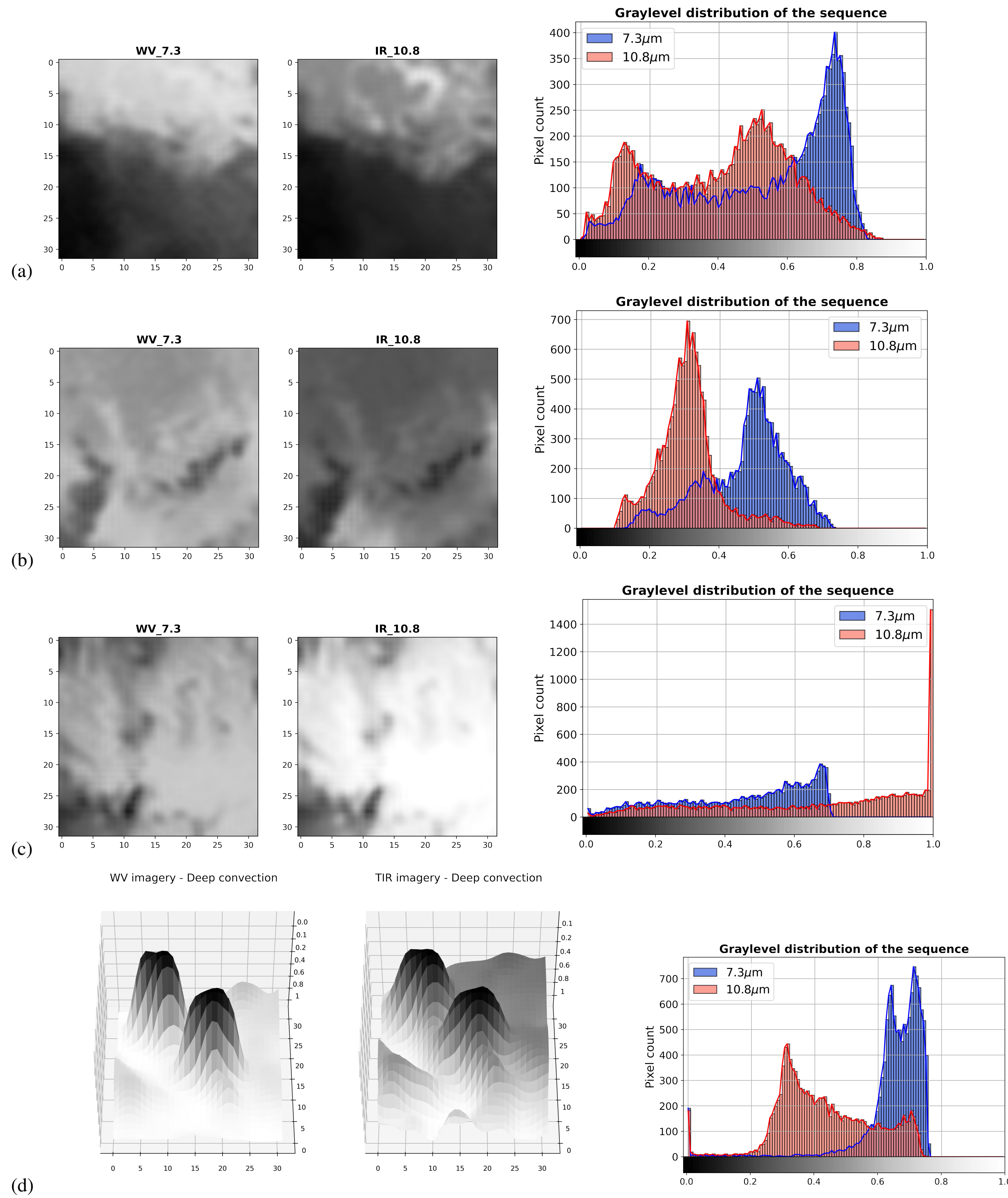


FIG. 13: Pixel analysis of relevant atmospheric events. (a) Dry intrusion from North, (b) Dry slot, (c) Low-level moisture detected in WV, (d) 3D deep convective motions of a heavy precipitation event as seen in WV and TIR imagery.

5. Discussion

This study proposed a bi-spectral approach to tackle the challenge of rainfall detection in a tropical region by means of a Deep learning model. Rainfall dynamics in West Africa are complex showing erratic behaviour, the African Easterly mid-level Jet plays a crucial role in the formation of a favourable thermodynamic environment to develop deep convection. The newly added water vapor channel 6 proves to be useful at detecting this mid-level jet and depicting where local convective motions are taking place without possible contamination from low-level clouds. The WV channel true added value relies in revealing dry intrusions and dry slots in between tropical systems (Figure 13a-b, more examples of such events are in the appendix) which translates into a reduced value of false alarms in the confusion matrices and a higher success ratio (SR) in the Roebber diagram. Certain events are more difficult to identify, that is why the gray-level histogram becomes helpful in this situations to see how the combination of the two channels usually generates a bimodal pixel distribution where WV pixels mode with respect to TIR is an indicator of either dry or wet conditions during dry intrusions and dry slots.

The addition of water vapor remarkably improved the performance of the model in a binary classification context. It is seen in Table 3 that the predicted probabilities of the model using a combination of both satellite channels resulted in values that are usually in-between the single channels model probabilities or in some cases closer to the true groundtruth binary value, this is mostly seen for low-level moisture events. Mesoscale dry intrusions are the most challenging to depict because of the sharp gradient present in the image. The probabilities of the dual-channel model for this type of event can be misleading as it sometimes struggles to capture the correct development of this dry air advection into the rainfall area. Most of the dry slots and dry intrusions analysed events (tables in the Appendix.) take place during early or late rain season that is when a more dynamical atmospheric situation is found and is reflected by larger values of the combined probabilistic output compared to the dry season. The intrusion of dry air into the tropical system could also promote the development of 'virga' which consists of precipitation evaporating before reaching the ground due to a dry patch combined with high air temperature, resulting in an actual mismatch of predictions. Contribution from the timestamp is mostly observed during the dry season where the model using time information is correctly expecting dry sequences most of the time.

Water vapor channel is not a good channel to retrieve stratiform rainfall. Stratiform or warm rain is the precipitation that falls from low-level clouds and usually is associated with light rainfall events. However considering that more than 80% of the rainfall in tropical inland areas comes from mesoscale convective systems (MCs) and the relationship between low level clouds and warm rain is still very uncertain as the presence of low clouds is sporadically linked to rainfall events (Liu and Zipser 2009), water vapor imagery could be skillful in focusing the model only towards strong convective events. Presence of such low clouds or high clouds like thin iced cirrus lead to an over-forecast of precipitation in models that only make use of TIR imagery, this overestimation can be seen in the comparison of the probabilistic output of the two models (see Figure 11) where TIR ensemble predicted probabilities shows much more uncertainty, as well as in the contingency tables, here TIR model has the highest number of false positives.
On the other hand, the results of the misclassification analysis per month and season tell us that the model using only WV images struggles more compared than the other models during dry season months. The explanation of this is the variable height of the effective layer: since during the dry season there is very dry air aloft it might happen that the satellite sensor is able to detect some anomalous low-level moist currents that are not related with any rainfall however the model without TIR information about clouds would still classify them as rain sequences, Figure 13c shows one example of such event that happened in mid-January 2020.
The misclassification per rain category also shows that water vapor alone is the worst model in detecting both very light and light rain, this type of rain is likely to be found in stratiform clouds (not seen in WV channel). The discrepancy in misclassified sequences between northern and more southern stations is in agreement with literature and it is most likely due to a progressively higher availability of moisture when approaching the coastal areas that leads to a slight increase of rain from warm clouds (Reinares Martínez et al. 2020).

The synergy between TIR and WV leads to an important consideration: the inclusion of WV channel into a deep learning model can provide the basis to develop a full alternative solution to the established CCD method. CCD method is a cloud indexing statistical approach applied to the TIR channel to distinguish convective rain clouds from non-rain low clouds. It assumes a positive linear relationship between cloud tops and rainfall to find an optimal temperature threshold for a certain area (Domenikiotis et al. 2003). However, because of the complexities of convective rainfall, both the temperature threshold and the linear regression relationship depend on local characteristics of the area under consideration. Even if the region of interest is divided in many calibration sub-areas, the results exhibit several discontinuities in the rainfall estimates. Additionally, each calibration area requires a lot of ground measurements. At the moment North Ghana gauge coverage is far from sufficient to make this method a viable option. The strengths of CCD method relies on its simple approach to get reliable results at very low temporal resolutions (POD: 0.69, SR: 0.75, BIAS: 0.9 for wet dekadals detection) (Tarnavsky et al. 2014). Deep learning models are more complex algorithms with the advantage of being fully observations-driven, and with the inclusion of water vapor imagery they do not have to rely on any a priori assumptions. The combination of the two satellite channels automatically excludes non-convective features within the whole region of interest. Temporal resolution is higher than, for instance TAMSAT (3hrs vs daily) which is very beneficial in a convective precipitation context. Looking at Figure 12 it is visible that Channel 5 is detecting only upper level WV structures, the adoption of this channel offers a way to focus the model even more on deep convective events that are strictly related to heavy rainfall events. However, no information on stratiform rain and shallow convection can be extracted from this channel and for this reason WV 7.3μm was the preferred channel in this study.

The model developed in this study is specifically designed for equatorial Africa, similar to CHIRPS and TAMSAT. Water vapor channel is expected to be less effective in detecting rainfall outside the tropics where convective rainfall is less dominant. Different factors play a role in rainfall formation in mid-latitudes, in particular frontal systems. Due to its geographic location, North Ghana has frequent intrusions of dry air from the Sahara desert. Dry intrusions are visible in water vapor imagery and for this reason the model here performs particularly well.

A promising direction for further development of the presented DL model is to trasform this binary rainfall detection into a full gridded rainfall product. Results of the rainfall categorization (multiclass classification) has shown promising results given the simplicity of the model, but still geostationary (GEO) IR images do not have any meaningful relation with precipitating hydrometeors since they are only able to sense objects like clouds and water vapor based on their thermal emissions. For this reason the model is particularly struggling in differentiating intermediate categories like light and moderate rainfall. Passive microwave sensors (PMWs) are the only satellite observations capable of retrieving the rainfall rate by receiving the backscattered signal of hydrometeors, however the latency time between consecutive measurements over the same location is just too large to give any reliable estimate. It was observed that certain rainfall events are so highly localised in space and time that the covered area of the input images (32x32) as well as the temporal resolution (3hrs) were still too coarse to make such events detectable by the model, for example the heavy rain event in Pusiga in May (Fig. B1) was incorrectly classified by all models including IMERG-Final. A well defined small dark blob in WV imagery appears only at the end of the sequence while the previous images contained mostly bright pixels, suggesting the model it was a total dry sequence. Higher temporal resolution are readily available for both SEVIRI images and TAHMO observations, it is advised as part of future development to move to an hourly time scale so that heavily localised rainfall has more chances to be detected. Moreover, by applying a higher temporal resolution less data will be discarded, because our methodology only considered full sequences with 12 images.

Precipitation estimates have an important operational value, in fact they are crucial to give a quantitative measure of plants development if used as input into any vegetation model that aims to predict crop growth. Farmers and local people in Ghana can really benefit from this model, once fully-developed, as the latency of rainfall product is near real-time, allowing this model to be used in real world applications such as flood and drought monitoring. Models like IMERG-Final have way too large time latency (3.5 months) for these kind of applications. IMERG-Late on this matter might be more useful, having a latency of 12hrs.

6. Conclusion and future improvements

This work showed that a Deep learning model is able to tackle rainfall detection in regions where sparse rain gauge networks and erratic precipitation patterns pose a challenge to traditional methods. The contribution of water vapor into the model is noticeable and resulted in a reduced number of false alarms, rainfall is then less overestimated. Water vapor inability to detect non-convective features can be seen as an equivalent of applying a temperature threshold in the CCD method. Water vapor imagery true value relies in detecting dry air intrusions into tropical easterly waves which can be very skillful since the region of interest is very close to the Sahara desert. Despite their coarse spatial and temporal resolution, the addition of passive microwave sensors from LEO satellites seems a promising way to transform this binary rainfall detection model into a continuous rainfall product estimate like TAMSAT or CHIRPS. As a starting point, rainfall estimates derived from GEO-IR imagery could be locally adjusted whenever a PMW observation is available for that region, post-processing calibration is required to account for grid mismatch (Hsu et al. 2020).

Expanding the dataset has already proved to increase the performance of the model, however the addition of more data from other stations across West Africa in a binary detection model might introduce new local climate variability that will alter the learning algorithm of the rainfall pattern distribution. The advances in the field of machine learning is bringing promising applications in both data augmentation and samples generation. Generative adversarial network (GAN) are neural network models that can replicate the data distribution of the training dataset through a generator and a discriminator model, so by using historical satellite rainfall fields it is actually possible to generate a spatial dependent probabilistic output of rainfall field for nowcasting purposes.

The combination of multiple SEVIRI channels to enhance low-level features by applying a temperature brightness difference between relevant channels might improve the detection of warm rainfall, however it is likely that precipitation will be more overestimated unless a better relation between the two variables is not defined through a defined $T_b - RR$ relationship. On the other hand, the adoption of the other WV channel 6.2μm would bring more reliable results on the detection of heavy rainfall events which they account for most of the accumulated rainfall on the ground. The use of 3-hrs temporal scale, even though it is suitable to keep track of the development of storms seems too coarse to capture certain short-lived rainfall events. Performances at hourly temporal scale are worth to be investigated since both TAHMO and EUMETSAT data are readily available, less data will be discarded since the complete sequence requires only four images instead of twelve.

Moreover, using a smaller area centered around the point observation at the expenses of a simpler model could be beneficial as precipitation in this region is attributed to localised pockets of rapid most air ascent which are sometimes not larger than few kilometers so the covered area of the input images (96x96km) is still way too large. On this matter, the launch of the new MTG-I1 scheduled for Q4 2022 is set to deliver higher pixel resolution (2km) and even faster image BRC (Baseline Repeat cycle) of 10 minutes (ESA 2021). This would potentially allow the WV channel to detect smaller scale rising air motions and keep the input shape untouched.

Acknowledgments. I want to spend few words to thank my supervisors Marie-Claire ten Veldhuis, Riccardo Taormina and Nick van de Giesen for their guidance and the precious suggestions they gave me during the past nine months. A special thanks goes to my daily supervisor Monica Estebanez Camarena, without her commitment into this project neither the model and this study would have existed. Our weekly meetings encouraged me to explore more and more the field of satellite remote sensing. Finally, my family is also part of this study through the constant moral and financial support over the past years.

Data availability statement. TAHMO stations data is available on request through the TAHMO data portal website: https://tahmo.org/climate-data/

MSG SEVIRI data is available on the Earth Observation Portal (EOP) after registration: https://eoportal.eumetsat.int/

APPENDIX A

Multiclass classification results

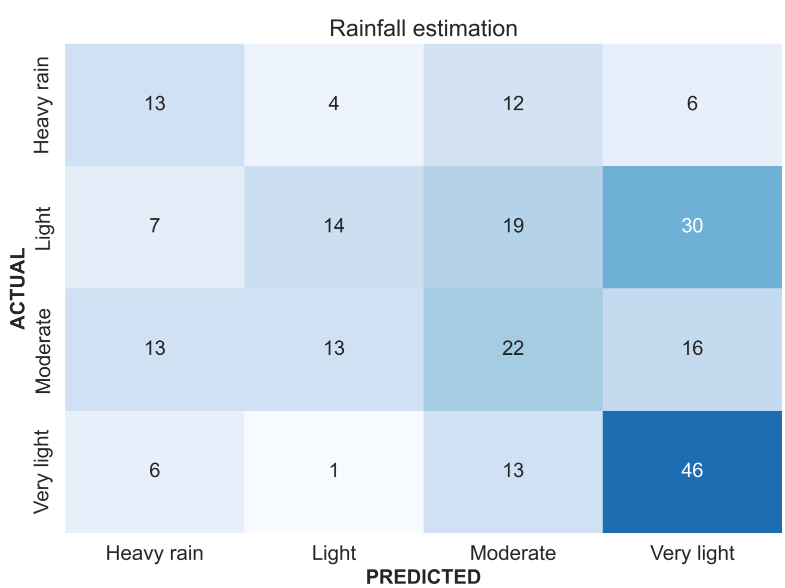


Fig. A1: Results of RainRunner WV+TIR+Timestamp multiclassification of rain sequences in the test dataset by precipitation intensity [mm/3hrs].

APPENDIX B

Pixel analysis

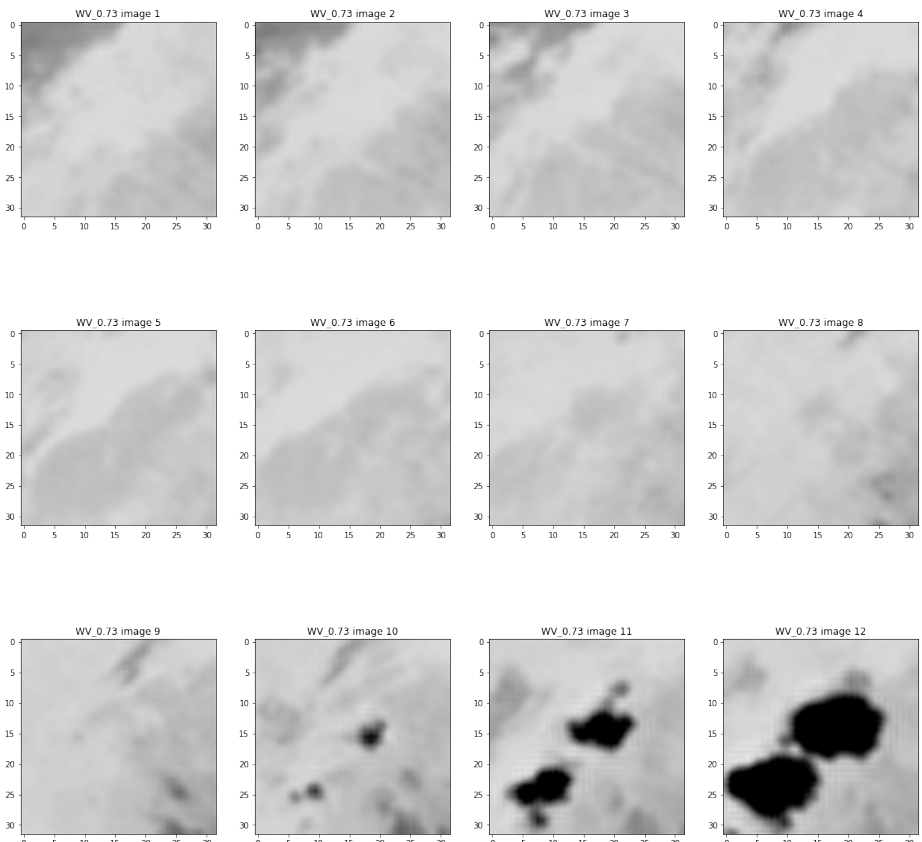


FIG. B1: Example of a misclassified rain sequence in WV imagery due to coarse temporal resolution. Pusiga - May 2020

a. Dry slots

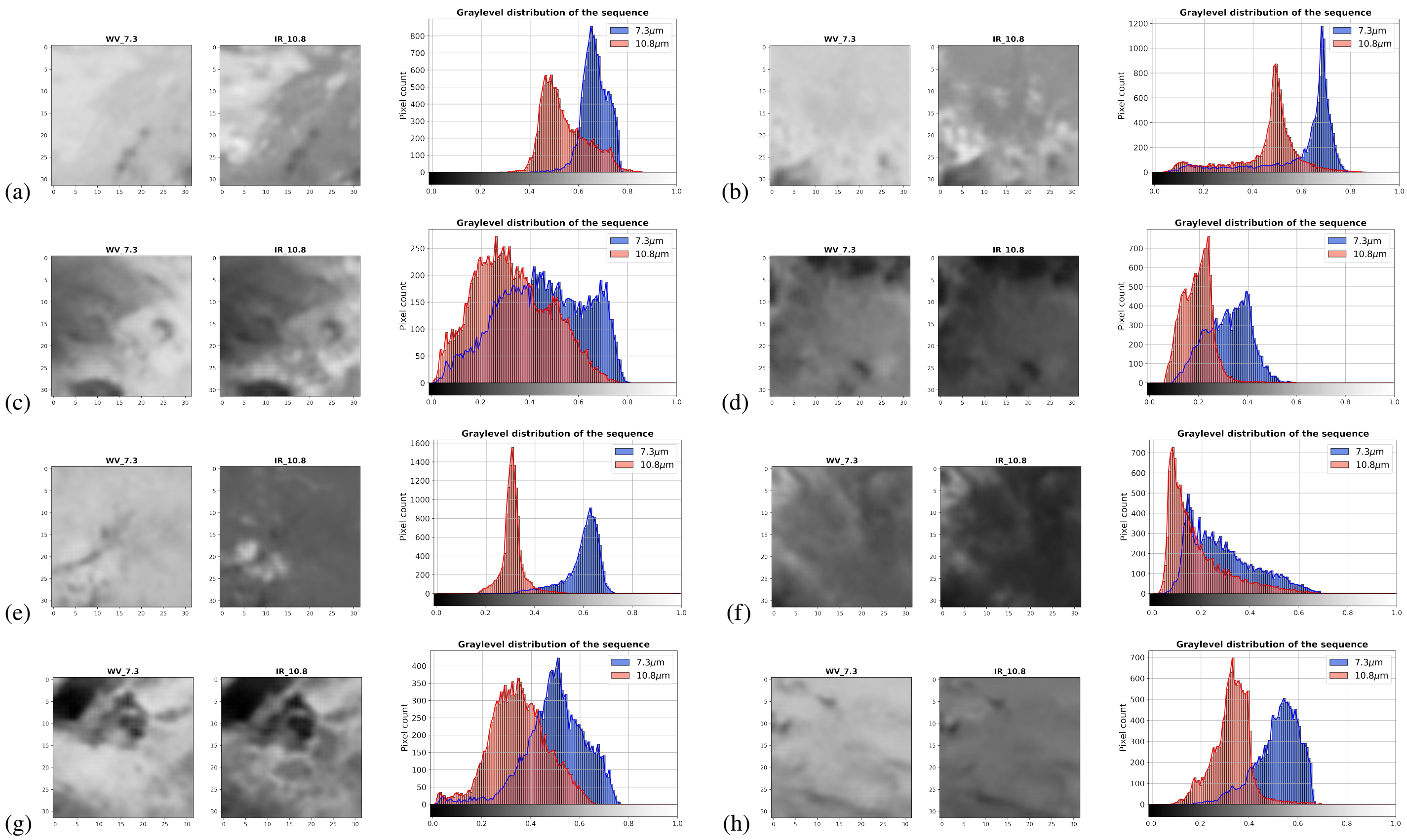


FIG. B2: Dry slots observed at TAHMO locations.

EVENT	Groundtruth	TIR	WV	Combined	Timestamp
(a) Bimbilla_2020.09.07_09	0	0.51	0.03	0.007	0.17
(b) Bimbilla_2020.09.12_06	0	0.56	0.54	0.23	0.39
(c) Han_2020.10.01_15	0	0.73	0.34	0.42	0.33
(d) Bongo_2020.04.10_15	0	0.66	0.28	0.49	0.38
(e) Bongo_2020.05.09_12	0	0.41	0.11	0.32	0.35
(f) Daffiama_2020.05.15_15	0	0.74	0.28	0.47	0.25
(g) Tamale_2020.06.14_12	0	0.52	0.09	0.29	0.23
(h) Navrongo_2020.06.20_15	0	0.45	0.04	0.35	0.50

b. Dry intrusions

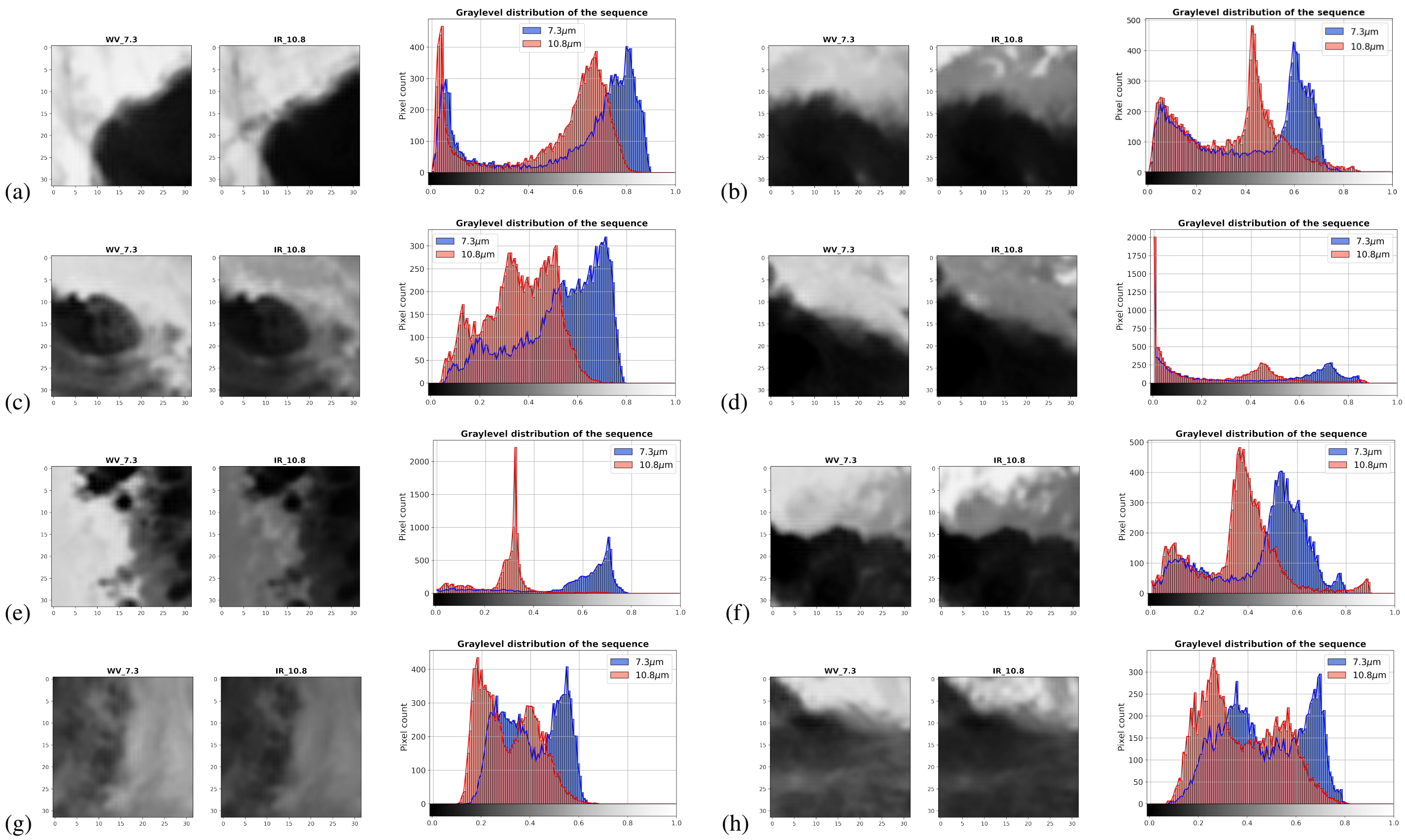


FIG. B3: Dry intrusions observed at TAHMO locations.

EVENT	Groundtruth	TIR	WV	Combined	Timestamp
(a) Bimbilla_2020.03.22_15	0	0.57	0.14	0.67	0.17
(b) Bimbilla_2020.05.06_21	0	0.92	0.45	0.64	0.23
(c) Bimbilla_2020.07.26_12	1	0.51	0.26	0.75	0.78
(d) Bimbilla_2020.09.30_15	0	0.77	0.69	0.46	0.59
(e) Navrongo_2020.05.17_12	0	0.81	0.41	0.69	0.48
(f) Pusiga_2020.05.06_00	0	0.50	0.34	0.24	0.29
(g) Pusiga_2020.07.15_03	0	0.51	0.37	0.46	0.54
(h) Bongo_2020.09.25_12	0	0.62	0.27	0.44	0.44

c. Low-level moisture

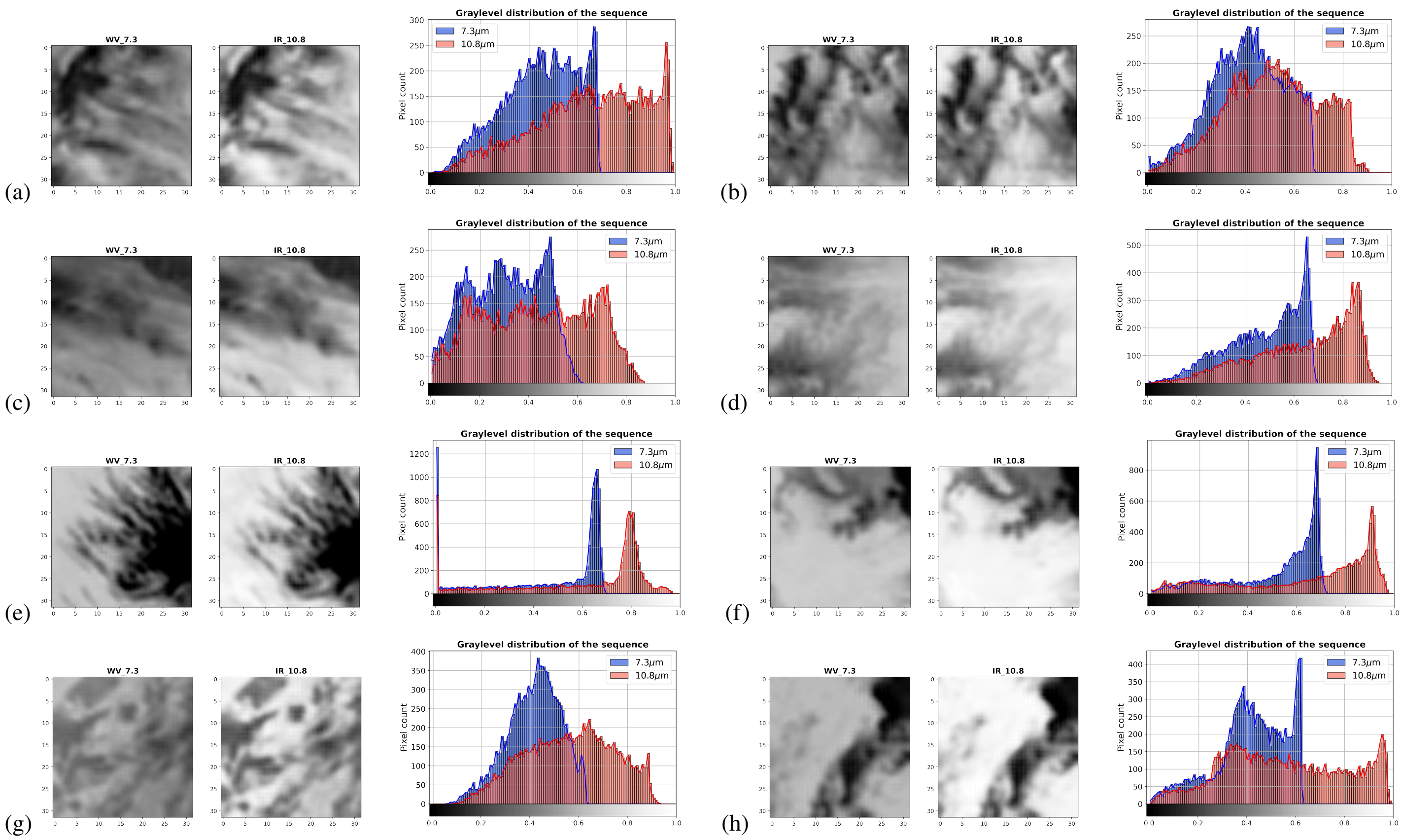


FIG. B4: Low-level moisture events observed at TAHMO locations during dry season.

EVENT	Groundtruth	TIR	WV	Combined	Timestamp
(a) Bimbilla_2020.02.11_18	0	0.42	0.83	0.28	0.30
(b) Bimbilla_2020.12.22_12	0	0.41	0.65	0.37	0.18
(c) Daffiama_2020.01.22_21	0	0.28	0.54	0.20	0.02
(d) Daffiama_2020.01.23_06	0	0.22	0.66	0.14	<0.01
(e) Kpandai_2020.01.25_00	0	0.43	0.58	0.18	0.02
(f) Kpandai_2020.10.21_00	0	0.30	0.63	0.37	0.23
(g) Navrongo_2020.02.12_00	0	0.07	0.50	0.24	<0.01
(h) Pusiga_2020.02.11_18	0	0.08	0.55	0.24	0.05

References

Bechtold, P., 2019: Convective precipitation. ECMWF, URL https://confluence.ecmwf.int/display/FUG/9.6+Convective+Precipitation#id-9.6ConvectivePrecipitation-ConvectivePrecipitation.

Berry, G. J., and C. Thorncroft, 2005: Case study of an intense african easterly wave. *Monthly Weather Review*, **133 (4)**, 752 – 766, https://doi.org/10.1175/MWR2884.1, URL https://journals.ametsoc.org/view/journals/mwre/133/4/mwr2884.1.xml.

Coz, C. L., and N. van de Giesen, 2020: Comparison of rainfall products over sub-saharan africa. *Journal of Hydrometeorology*, **21 (4)**, 553 – 596, https://doi.org/10.1175/JHM-D-18-0256.1, URL https://journals.ametsoc.org/view/journals/hydr/21/4/jhm-d-18-0256.1.xml.

Domenikiotis, C., M. Spiliotopoulos, E. Galakou, and N. Dalezios, 2003: Assessment of the cold cloud duration (ccd) methodology for rainfall estimation in central greece.

Dybkjær, G., 2003: A simple self-calibrating cold cloud duration technique applied in west africa and bangladesh. *Geografisk Tidsskrift-Danish Journal of Geography*, **103 (1)**, 83–97, https://doi.org/10.1080/00167223.2003.10649482, URL https://doi.org/10.1080/00167223.2003.10649482, https://doi.org/10.1080/00167223.2003.10649482.

ESA, 2016: Msg's seviri instrument. *ESA EUMETSAT*, 33, URL https://www.esa.int/esapub/bulletin/bullet111/chapter4_bul111.pdf.

ESA, 2021: Meteosat third generation takes major step towards its first launch. URL https://www.eumetsat.int/meteosat-third-generation.

FAO, 2020: Ghana at a glance. URL https://www.fao.org/ghana/fao-in-ghana/ghana-at-a-glance/en/#:~:text=Agriculture\%20contributes\%20to\%2054\%20\%25\%20of,food\%20needs\%20of\%20the\%20country.

Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes. *Scientific Data*, **2**, 150 066, https://doi.org/10.1038/sdata.2015.66.

Georgiev, C., and P. Santurette, 2009: Mid-level jet in intense convective environment as seen in the 7.3 μm satellite imagery. *Atmospheric Research - ATMOS RES*, **93**, 277–285, https://doi.org/10.1016/j.atmosres.2008.10.024.

Hsu, K., N. Karbalee, and D. Braithwaite, 2020: *Improving PERSIANN-CCS Using Passive Microwave Rainfall Estimation*, 375–391. https://doi.org/10.1007/978-3-030-24568-9_21.

Kidd, C., and G. Huffman, 2011: Global precipitation measurement. *Meteorological Applications*, **18**, 334 – 353, https://doi.org/10.1002/met.284.

Knippertz, P., and A. H. Fink, 2008: Dry-season precipitation in tropical west africa and its relation to forcing from the extratropics. *Monthly Weather Review*, **136 (9)**, 3579 – 3596, https://doi.org/10.1175/2008MWR2295.1, URL https://journals.ametsoc.org/view/journals/mwre/136/9/2008mwr2295.1.xml.

Liu, C., and E. J. Zipser, 2009: “warm rain” in the tropics: Seasonal and regional distributions based on 9 yr of trmm data. *Journal of Climate*, **22 (3)**, 767 – 779, https://doi.org/10.1175/2008JCLI2641.1, URL https://journals.ametsoc.org/view/journals/clim/22/3/2008jcli2641.1.xml.

Nicholson, S., 2013: The west african sahel: A review of recent studies on the rainfall regime and its interannual variability. *ISRN Meteorology*, **2013**, https://doi.org/10.1155/2013/453521.

Reinares Martínez, I., J.-P. Chaboureau, and J. Handwerker, 2020: Warm rain in southern west africa: A case study at savè. *Atmosphere*, **11**, 298, https://doi.org/10.3390/atmos11030298.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Weather and Forecasting*, **24 (2)**, 601 – 608, https://doi.org/10.1175/2008WAF2222159.1, URL https://journals.ametsoc.org/view/journals/wefo/24/2/2008waf2222159_1.xml.

Selami, N., G. Sèze, M. Gaetani, J.-Y. Grandpeix, C. Flamant, J. Cuesta, and B. Noureddine, 2021: Cloud cover over the sahara during the summer and associated circulation features. *Atmosphere*, **12**, 428, https://doi.org/10.3390/atmos12040428.

Shen, C., X. Chen, and E. Laloy, 2021: Editorial: Broadening the use of machine learning in hydrology. *Frontiers in Water*, **3**, https://doi.org/10.3389/frwa.2021.681023, URL https://www.frontiersin.org/article/10.3389/frwa.2021.681023.

TAHMO, 2016: Trans-african hydro-meteorological observatory. URL https://tahmo.org/.

Tan, J., G. J. Huffman, D. T. Bolvin, and E. J. Nelkin, 2019: Imerg v06: Changes to the morphing algorithm. *Journal of Atmospheric and Oceanic Technology*, **36 (12)**, 2471 – 2482, https://doi.org/10.1175/JTECH-D-19-0114.1, URL https://journals.ametsoc.org/view/journals/atot/36/12/jtech-d-19-0114.1.xml.

Tarnavsky, E., D. Grimes, R. Maidment, E. Black, R. P. Allan, M. Stringer, R. Chadwick, and F. Kayitakire, 2014: Extension of the tamsat satellite-based rainfall monitoring over africa and from 1983 to present. *Journal of Applied Meteorology and Climatology*, **53 (12)**, 2805 – 2822, https://doi.org/10.1175/JAMC-D-14-0016.1, URL https://journals.ametsoc.org/view/journals/apme/53/12/jamc-d-14-0016.1.xml.

Tomassini, L., 2020: The interaction between moist convection and the atmospheric circulation in the tropics. *Bulletin of the American Meteorological Society*, **101 (8)**, E1378 – E1396, https://doi.org/10.1175/BAMS-D-19-0180.1, URL https://journals.ametsoc.org/view/journals/bams/101/8/bamsD190180.xml.

Tomassini, L., D. Parker, A. Stirling, C. Bain, C. Senior, and S. Milton, 2017: The interaction between moist diabatic processes and the atmospheric circulation in african easterly wave propagation: The interaction between moist processes and circulation in aews. *Quarterly Journal of the Royal Meteorological Society*, **143**, https://doi.org/10.1002/qj.3173.