Rain detection using commercial microwave link data and k-means clustering

Raul Victor de O. Paiva.*, Tarcisio F. Maciel.*, Rodrigo Z. Prado.†,Modeste Kacou‡ and Marielle Gosset§
*Federal University of Ceará (UFC), Post-Graduation Program in Teleinformatics Engineering (PPGETI),
Fortaleza, Brazil, E-mail: {raul.paiva@alu.ufc.br, maciel@ufc.br}
†The Weather Force, Toulouse, France, E-mail: rodrigo.zambrana@weatherforce.org
‡University Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire, E-mail: modeste.kacou@ird.fr
§Institut de Recherche pour le D'eveloppement (IRD), Toulouse, France, E-mail: marielle.gosset@ird.fr

***Abstract*— Rain monitoring is crucial for preventing natural disaster damage and for agriculture. Commercial Microwave Link (CML) data has been used to predict rain events, especially when Rain Gauges (RGs) or radars are scarce. In this work, we modified a classic variance approach with k-means clustering for rainfall classification. For this, CML data was applied and validated using RG data. The results showed that the Unsupervised Learning (UL) approach is sufficient for classification. The proposed approach achieved a precision of 82% for 1 hour and 15 minutes window. The classical is limited because it needs *a priori* RG data.**

***Keywords*— CMLs, attenuation, rainfall classification, clustering.**

I. Introduction

Commercial Microwave Links (CMLs) offer a cost-effective solution for analyzing rainfall by measuring the attenuation of microwave signals caused by rain. With their extensive network coverage and high temporal resolution, CMLs provide valuable insights into precipitation patterns. By utilizing CMLs, researchers can improve flood warning, hydrological modeling, and water resource management systems. However, challenges such as signal quality and non-rainfall-related factors need to be carefully addressed for accurate rainfall estimation [1]. Withal, leveraging CML measurements for weather monitoring can enhance our understanding of rainfall and globally contribute to a more sustainable and efficient water resource planning.

In general, radio signals carried by electromagnetic waves suffer attenuation as they travel between transmitter and receiver in a wireless communication network. This attenuation depends on the propagation medium, adopted carrier frequency, and on the distance between transceivers [2]. Indeed, depending on the frequency, when the radio signals traverse a rainy path, they can suffer a significant additional amount of attenuation, primarily due to the collective influence of individual rain droplets that absorb and scatter the waves' energy in various directions [3]. Thus, the main idea of exploiting CMLs for estimating rainfall relies on relating the radio signal attenuation with the amount of rain perceived on the link. Fig. 1 illustrates this effect.

As specified in [4], this attenuation relates to rainfall according to a power law that connects the specific attenuation

This work was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES) - Finance Code 001, by FUNCAP/Universal under grant no. UNI-0210-00043.01.00/23 and INCT-Signals/CNPq under grant no. 406517/2022-3, and by the Institut de Recherche pour le Développement (IRD). Tarcisio. F. Maciel was partially supported by CNPq under grant 312471/2021-1.

Fig. 1: Illustration of the basic operating principle of CML rainfall estimation.

k [dB · km^{-1}] along a rainy path with a rain rate R [mm · h^{-1}] as

$$k = a \cdot R^b, \tag{1}$$

where a and b are the coefficients of power law. Frequency and polarization have a significant influence on both a and b, while other factors such as Drop Size Distribution (DSD) and temperature exert a comparatively milder effect [5]. The specific attenuation k quantifies the decrease in signal strength that occurs over a distance of one kilometer. This fundamental mechanism establishes a direct correlation between the extent of attenuation along the path and the intensity of rainfall, i.e., a $k \leftrightarrow R$ relation.

Recent studies have demonstrated the utilization of operational CMLs in telecommunication networks for classification of wet and dry periods, as shown in the following works. The main objective in [6] is to classify the various physical phenomena that induce the Received Signal Level (RSL) measured on the CMLs. The authors propose a classification using the decision tree algorithm based on physical characteristics to distinguish between different precipitation phenomena. In this work, 3 links are used to obtain the attenuation data and a meteorological weather sensor called OTT Parsivel disdrometer to access the precipitation measurements.

In [7], the main objective of the work is to propose a method for detecting rainfall using microwave links and classifying dry and rainy periods based on the Support Vector Machine (SVM) algorithm in order to improve the accuracy of rainfall estimation. In this study, 7 links and 8 rain gauges are used to train and evaluate the models. The modelling is carried out on each link, directly with the attenuation data, without time dependence, and the evaluation is done in terms of accuracy, true positive rate and false positive rate.

In [8], the aim is to explore the application of machine learning techniques, including both supervised and unsupervised methods, for classifying dry and rainy periods in precipitation estimation using microwave links from mobile telecommunication networks. The study utilizes data from four links, incorporating ground-based C-band weather radars and

rain gauges, and evaluates the performance of these machine learning models in comparison to traditional model-based approaches from the literature. The primary objective is to enhance the accuracy of rainfall estimation, particularly in areas with limited rainfall monitoring infrastructure. The Supervised Learning (SL) models include logistic regression, K-nearest Neighbors (KNN), decision trees, and artificial neural networks, while Unsupervised Learning (UL) models involve k-means, fuzzy C-means, and Self-organization Map (SOM). The study also assesses ensemble models like Random Forest (RF), Histogram-Based Gradient Boosting (HBGB), Stacked Machine Learning Ensemble (SMLE), and Voting Classifier (VC), employing various metrics such as accuracy, precision, sensitivity, F1 score, Area Under the ROC Curve (AUC), and binary cross-entropy for model evaluation.

In this paper, the rainfall classification of wet and dry periods is treated using firstly a classical approach, applied by [9] and, secondly one different methodology for rainfall classification are created using k-means clustering [10], an unsupervised machine learning method. The main objective of this work is to combine k-means clustering with the classical variance-based approach proposed by [9] in order to improve the classical methodology by comparing this combination with the classical variance-based. These methods are evaluated on different (no) rain conditions.

The remainder of this paper is organized as follows. A very short review on the composition of radio attenuation related to rain is presented in Section II. The methods used in this paper are explained in Section III. We present the study area and a brief data analysis, the methodology and the results in Section IV. Finally, the concluding remarks of this work are drawn in Section V.

II. CML-based rainfall classification background

In this section we describe some fundamental discrete-time models required as background for the rainfall classification considered in this work. The measurement of raw attenuation between the transmitting and receiving ends of a link involves determining the disparity between the Transmitted Signal Level (TSL) and the RSL provided by the operator, i.e.,

$$A_{raw}[n] = TSL[n] - RSL[n] \tag{2}$$

where n represents the index of the discrete-time signal sample, which are taken at each Δ_t time units [9]. Thus, the index n is used hereafter as to indicate the time instant $t = n\Delta_t$ with $n \in \{0, 1, \ldots\}$. For instance, in this work the available CML data is provided at a coarse time sampling with $\Delta_t = 15$ minutes.

In the absence of rain, $A_{raw}[n]$ varies for different reasons: dew on the antennas; variation in the refractive index of the air; attenuation by atmospherics gases; changes in the transmit power levels; noise in the electronics, and/or; quantization of the signal.

In general, the total attenuation of a CML is given by

$$A_{raw}[n] = A_{clear}[n] + A_{rain}[n] + A_{wet}[n] + A_{vap}[n] + e[n], \tag{3}$$

where $A_{clear}[n]$, $A_{rain}[n]$, $A_{wet}[n]$, and $A_{vap}[n]$ mean the attenuation due to distance, rain, wet antennas, and humidity in the atmosphere, respectively, and $e[n]$ means the errors due to quantization noise $q[n]$ plus thermal noise $z[n]$, i.e., $e[n] = q[n] + z[n]$.

III. Rain detection

A fundamental step in rainfall prediction studies is to discriminate between wet (raining) and dry (no raining) periods [9], [11] and, indeed, this is the main problem investigated in this paper. Two different classification schemes are studied in this paper, a classical and a based on UL.

A. Dry/wet discrimination based on variance

One way to classify dry and wet periods is to consider the variance of the radio signal raw attenuation of a CML within a moving window compared to a reference σ_0 variance value [9]. In this method a wet or dry classification is performed first using a statistical test considering a moving window characterized by $\tilde{n} \in \mathcal{W}$, where $\mathcal{W} = \left[n - \lfloor \frac{W}{2} \rfloor, n + \lfloor \frac{W}{2} \rfloor\right]$ and $W > 0$ is the (local) window size[1]. For this classification, the discriminant value γ_{var} is calculated as

$$\gamma_{var}(n, W) = \frac{\sum_{\tilde{n} \in \mathcal{W}} (A_{raw}[\tilde{n}] - \bar{A}_{raw})^2}{W}, \text{ with} \tag{4a}$$

$$\bar{A}^{w}_{raw}(n, W) = \frac{1}{W} \sum_{\tilde{n} \in \mathcal{W}} A_{raw}[\tilde{n}]. \tag{4b}$$

According to [9], this method for separating dry from rainy periods is based on the assumption that γ_{var} values are small during dry periods and large during wet periods. Then, the binary decision rule $\hat{y}^{wd}[n]$ for classifying $A_{raw}[n]$ samples as wet or dry in this case is given by

$$\hat{y}^{wd}[n] = \begin{cases} 1 \text{ (wet)}, & \gamma_{var}(n, W) > \sigma_0, \\ 0 \text{ (dry)}, & \gamma_{var}(n, W) \leq \sigma_0, \end{cases} \tag{5}$$

and σ_0 is the variance threshold.

In this method, initially the wet and dry distributions get segregated by utilizing the Rain Gauge (RG) data and an arbitrary rainfall threshold ρ. Subsequently, upon employing a centralized moving window of size W, variance is computed on the unprocessed attenuation subsets. To estimate σ_0, we separated the dataset into dry, for each rainfall value measured by the RG that is smaller than or equal to a threshold ρ, and wet, for each rainfall value measured by the RG that is larger than ρ. We then removed the zero values from the "dry" class and used the 3$^{\text{rd}}$ quartile of the remaining data to define the variance threshold σ_0 based on the variance computed considering the window W. This is motivated by the facts that most of the data is usually associated with dry periods for which RG measurements are zero and that both dry and wet RG measurements distributions are not Gaussian, which would lead to undesired excessive biases if the zeros were not removed in advance.

B. Dry/wet discrimination based on variance and k-means clustering

Considering the previous section, we can see that the rule in Eq. (5) can easily be replaced by another rule instead of calibrating a σ_0 value to split the data into wet/dry periods.

[1]For instance, in Section IV, odd values are used for W to ensure a window symmetrically centered around n.

The k-means clustering, for example, is an UL technique that can be used to assign the raw attenuation data to wet and dry periods without any previous information about rainfall. In the method of this section, the input data is the variance calculated from the raw attenuation given a window size W, cf. Eq. (4a). In effect, we are clustering the variances that represent the periods. As described in [9], the variance values are small during dry periods and large during wet periods, which is part of the rainfall dynamics that supports the application of clustering to the variance values calculated from the centralized moving window. However, in cases where rainfall is constant over a period of time, depending on the window size, two peaks in the variance might be formed, one at the beginning and one at the end of the period. The truth is that for a constant period of rainfall, depending on the window size the variance within the window might become close to zero, which could lead to inaccurate clustering, because the algorithm groups the data according to the variance levels.

In the following, we revisit the k-means method before employing it to discriminate between dry and wet periods based on the variance. Unsupervised clustering techniques assign data to different classes (clusters) considering no *a priori* information about the classes which the data belongs to [12]. For this, a clustering algorithm normally uses only information extracted from the data itself, building clusters based on the inherent data similarity [10], [12]. Partitional clustering algorithms divide the attribute space into cells, regions, or simply non-overlapping partitions, generally with the aid of prototype vectors. Each attribute vector is then associated with one of the existing prototypes based on a similarity criteria, for example, the smallest distance [10].

K-means is a very popular one due to its simplicity. For a number N of D-dimensional attribute vectors $\mathbf{x}_n \in \mathbb{R}^D$, k-means aims to find prototypes vectors (also termed centroids)[2]

$$\mathbf{w}_k \in \mathbb{R}^D, k = \{1, \ldots, K\}, K \ll N, \tag{6}$$

around which the attribute vectors are clustered.

The partition associated with the prototype $\mathbf{w}_i$ is defined as

$$\mathcal{V}_m = \left\{ \mathbf{x}_n \in \mathbb{R}^D \mid \|\mathbf{x}_n - \mathbf{w}_i\|^2 < \left\|\mathbf{x}_n - \mathbf{w}_j\right\|^2 \right\}, \quad \forall i, j \in \{1, \ldots, K\}, i \neq j, \tag{7}$$

where $\|\cdot\|^2$ means the Euclidean distance.

The sequential version of the k-means is then given by: 1) Choose a value for K (often determined using the *elbow* heuristic). 2) Define the K initial prototypes $\mathbf{w}_k$ (often determined selecting one attribute vector randomly). 3) Find the subscript of the prototype nearest to each attributes vector $\mathbf{x}_n$:

$$k_n^\star = \arg\min_k \|\mathbf{x}_n - \mathbf{w}_k\|_2, \forall n. \tag{8}$$

4) Update the prototype $\mathbf{w}_i$ assuming that it is equivalent to the average of all attribute vectors $\mathbf{x}_{k_n}$ currently assigned to the cluster k. 5) Repeat steps 3) and 4) until the convergence of $\mathbf{w}_k$ [10], [12].

The main objective of the classification approach in this section is to categorize the variance values. Herein the clusters represent different levels of variance with small values indicating dry periods and large values indicating wet periods. First, the variances are calculated from the raw attenuation data given a window size. Then, k-means clustering takes this 1-D variance data and categorizes it into $K = 5$ clusters (selected using the elbow criterion as shown later in Section IV-C). Finally, the first two groups are marked as non-rain, since they have most of the values equal to zero (first cluster) or smaller than the RG threshold ρ (second cluster), while the others three clusters are associated to "light rain", "moderate rain" and "heavy rain".

It is worth noticing that data used in this paper lacks long periods of continuous rainfall and that for this method the RG data is only used to evaluate the performance of the model *a posteriori*.

IV. Performance evaluation

In this section we evaluate the performance of the rain detection/classification methods described in Section III. First we provide in Section IV-A a brief presentation of the dataset considered in this paper. Afterwards, the metrics used to compare the methods are introduced in Section IV-B. Then, some details about the methodology used to evaluate each method are presented in Section IV-C while Section IV-D presents and discusses the obtained results.

A. Adopted dataset

The CML data used in this work was achieved through a collaboration with Orange Cameroun, the mobile telecommunication operator. Rainfall data is provided by the research network operated by the University of Douala (UIT). The link and pluviometric data used in this article are from Douala, Cameroon, and are provided with a time resolution of 15 minutes for each sample. Fig. 2 shows the location of the considered CML network and highlights the particular CML and RG used in this work, which details are given in Table I.

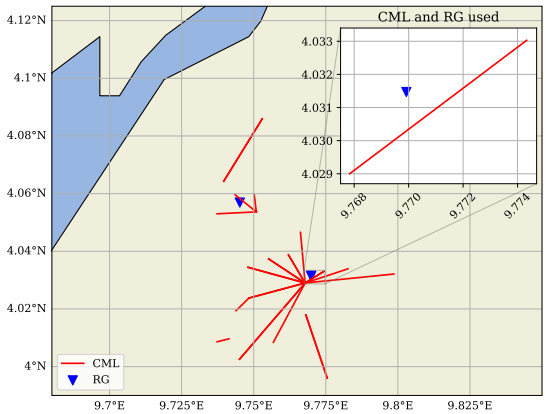


Fig. 2: Scenario of interest. Focused on one specific link.

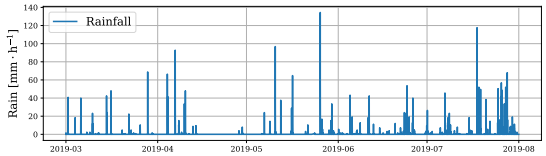
TABLE I: CML and RG details.

Parameter	Value
Frequency	14.5 GHz
Length	0.85 km
CML midpoint distance to RG	0.14 km

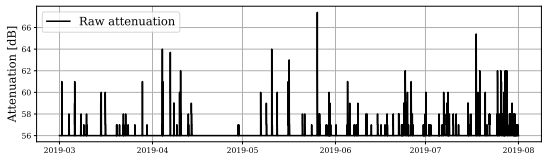
The period chosen for this study goes from March to August 2019. Fig. 3 shows the rainfall measured by the RG and the

[2]Please, observe that herein the index n does not refer to the discrete-time index used elsewhere in this paper.

attenuation measured for the CML, which have a positive correlation of around 55%. The distribution of the rainfall and



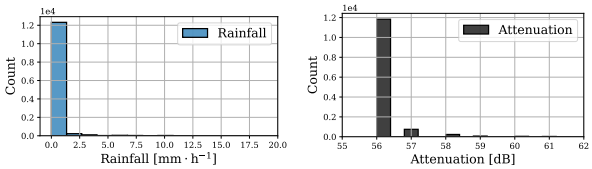
(a) Rainfall time series measured by the RG.



(b) Raw attenuation time series calculated from the CML.

Fig. 3: Rainfall and raw attenuation time series from the RG and CML, from March to August, 2019. Note the variability of both time series.

attenuation data are shown in Fig. 4. Note in Fig. 4a the most



(a) Histogram of rainfall measurements.

(b) Histogram of attenuation measurements.

Fig. 4: Distribution of the rainfall and attenuation measurements.

frequent values of rainfall around $0\,\mathrm{mm}\cdot\mathrm{h}^{-1}$, and in Fig. 4b the values of attenuation around $56\,dB$.

B. Evaluation Metrics

The metrics used to evaluate the models are accuracy, precision and recall. Accuracy Acc is used to measure how many predictions are correct out of all the predictions and is given by

$$Acc = (TP + TN) \cdot (TP + TN + FP + FN)^{-1}, \tag{9}$$

where TP, TN, FP, FN mean true positive, true negative, false positive, and false negative, respectively.

Precision $Prec$ quantifies how many cases are really positive classifications among all positive predictions, and it is given by

$$Prec = TP \cdot (TP + FP)^{-1}. \tag{10}$$

Recall Rec measures how many cases are predicted as positive among all actual positive cases. A recall of 1.0 means that there were no false negatives. It is given by

$$Rec = TP \cdot (TP + FN)^{-1}. \tag{11}$$

The F_1-score comes in handy when aiming for a balance between $Prec$ and Rec [10]. It is given by

$$F_1 = 2 \cdot (Prec \cdot Rec) \cdot (Prec + Rec)^{-1}. \tag{12}$$

C. Methodology

Several frameworks offer an off-the-shelf implementation of k-means and some of its variants. In this work, the `scikit-learn` [13] python library was used to run the k-means algorithm. To classify rainy and non-rainy events, we used the two methods explained in Section III, here called variance-based-only and variance-based-k-means, respectively. Data from March to June was used for training and from July to August for performance evaluation. The use of the rainfall threshold ρ on RG data is part of the validation process.

In the variance-based [9] methods we used the centralized moving window approach. In this case, we varied the Window Size (WS) in 3, 5, 7, 9 and 11, i.e. 45 min, 1h15 min, 1h45 min, 2h15 min and 2h45 min respectively.

In the variance-based-k-means method we have to specify the number of clusters. For this, we used the *elbow* criterion to find a suitable number K of clusters, which was found to be $K = 5$. It is worth noting that variances are considered as input data to be clustered depending on the window sizes. To come back to two clusters, i.e., the wet and dry clusters, 2 clusters were merged for the dry set and 3 for the wet set, cf. the reasons provided in Section III-B.

D. Results

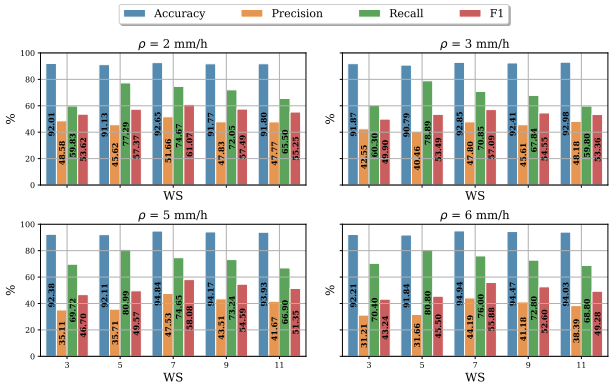
We varied the rainfall threshold ρ as 2, 3, 5 and 6 $\mathrm{mm}\cdot\mathrm{h}^{-1}$ to distinguish between dry and wet periods based on the RG data. The results are presented in Fig. 5, where the blue bars represent the accuracy, the orange bars the precision, the green bars the recall, and the red bars the F_1-score.

In terms of accuracy, the variance-based methods (Figs. 5a and 5b) performed above 90%, with the variance-based-k-means method outperforming the first one for every ρ and WS value. With an accuracy of around 95%, a WS of 7 and a $\rho \geq 5$ $\mathrm{mm}\cdot\mathrm{h}^{-1}$ give the best results for both variance-based methods. Due to the rain and attenuated signal distributions, which are unbalanced and contain much more samples of dry than wet periods, the accuracy metric does not give the best idea about false alarm rates. To address these concerns, precision and recall metrics must be evaluated.

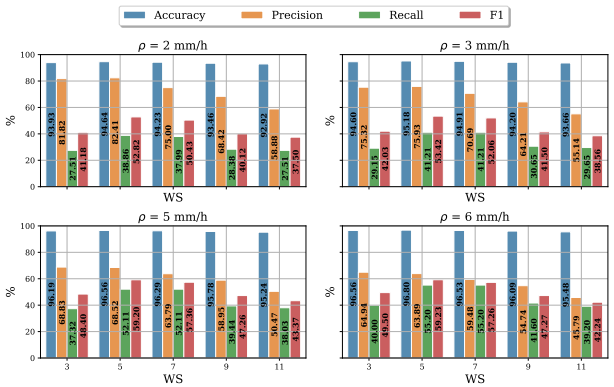
The best results in terms of precision come from the variance-based-k-means method: above 70% for WS ≤ 7 and $\rho \leq 3$ $\mathrm{mm}\cdot\mathrm{h}^{-1}$. In the variance-based-k-means method, we can observe that the precision decreases with the WS while ρ leads to a shift effect in this decrease. This happens because more samples are taken when the WS is increased to calculate γ_{var}, and in some cases there may be fewer values related to zero, thus increasing the assertiveness of the precipitation class. In this case, the best precision comes out when the WS is 5 for every ρ. On the variance-based-only method, the best precision result (around 52%) arises for a WS of 7 and $\rho = 2\,\mathrm{mm}\cdot\mathrm{h}^{-1}$, because the threshold σ_0 only works well.

The best recall results are concentrated on the variance-based-only, more specifically for WS values of 5 and 7, which lead to recall values above 70%. This can be attributed to the predominance of zero values in the data, indicating dry periods, and therefore fewer false alarms in this case.

Finally, in terms of F_1-score, both the variance-based-only method achieve results often exceeding 50% for $\rho \leq 3$ mm ·



(a) Performance evaluation for the variance-only method.



(b) Performance evaluation for the variance-based-k-means method.

Fig. 5: Performance evaluation of the two methods in terms of accuracy, precision, recall and F_1-score.

h^{-1}. On the other hand, the variance-based-k-means method exhibits the highest F_1-score results for $5 \leq \text{WS} \leq 7$, regardless of the value of ρ.

In general, the performance results from the variance-based-only method are satisfactory compared with the other method, mainly for $\rho \leq 3$ mm $\cdot$ h^{-1}. However, this method needs previous RG information to split the data into wet/dry distributions and calculate the variance on the moving window subsets. Thus, when there is a RG close enough of the link, it could be interesting to apply this method.

In case the use of RG data is somehow inconvenient, we presented a k-means based solution that cope with the classification of dry/wet periods relying only on CML information. The variance-based k-means method, has presented good results for $\rho \geq 5$ $mm \cdot h^{-1}$ and WSs equal to 5 and 7.

V. Conclusions

In this work the basic concept of using CML data for rainfall classification has been investigated. Two methods for wet and dry classification have been presented and evaluated. The classical approach based on variance had shown good results, but it is still a supervised method that depends on a previous analysis of rainfall data information associated to RGs. On the other hand, the unsupervised method based on clustering have shown good results without requiring *a priori* RG data, only for training. Thus, UL methods may have a great potential for rainfall classification problems, especially when rainfall data is only coarsely available. As perspectives for future studies, rainfall forecasting techniques based on raw attenuation after the dry/wet classification shall be investigated, as well as tensor decomposition models to work with multiple CML signals.

References

[1] C. Kidd, A. Becker, G. J. Huffman, *et al.*, "So, how much of the earth's surface is covered by rain gauges?" *Bulletin of the American Meteorological Society*, vol. 98, no. 1, pp. 69–78, Jan. 2017. DOI: 10.1175/bams-d-14-00283.1.

[2] T. S. Rappaport, *Wireless communications: principles and practice*, 2nd ed. Prentice Hall, Jan. 2002.

[3] G. Mie, "Beiträge zur optik trüber medien, speziell kolloidaler metallösungen," *Annalen der Physik*, vol. 330, no. 3, pp. 377–445, 1908. DOI: 10.1002/andp.19083300302.

[4] R. Olsen, D. Rogers, and D. Hodge, "The aR b relation in the calculation of rain attenuation," *IEEE Transactions on Antennas and Propagation*, vol. 26, no. 2, pp. 318–329, Mar. 1978. DOI: 10.1109/tap.1978.1141845.

[5] C. Chwala and H. Kunstmann, "Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges," *WIREs Water*, vol. 6, no. 2, Feb. 2019. DOI: 10.1002/wat2.1337.

[6] D. Cherkassky, J. Ostrometzky, and H. Messer, "Precipitation classification using measurements from commercial microwave links," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2350–2356, 2014. DOI: 10.1109/TGRS.2013.2259832.

[7] K. Song, X. Liu, M. Zou, D. Zhou, H. Wu, and F. Ji, "Experimental study of detecting rainfall using microwave links: Classification of wet and dry periods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5264–5271, 2020. DOI: 10.1109/jstars.2020.3021555.

[8] E. V. Kamtchoum, A. C. N. Takougang, and C. T. Djamegni, "A machine learning approach for the classification of wet and dry periods using commercial microwave link data," *SN Computer Science*, vol. 3, no. 3, Apr. 2022. DOI: 10.1007/s42979-022-01143-8.

[9] M. Schleiss and A. Berne, "Identification of dry and rainy periods using telecommunication microwave links," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 3, pp. 611–615, Jul. 2010. DOI: 10.1109/lgrs.2010.2043052.

[10] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4.

[11] H. Leijnse, R. Uijlenhoet, and J. N. M. Stricker, "Rainfall measurement using radio links from cellular communication networks," *Water Resources Research*, vol. 43, no. 3, Mar. 2007. DOI: 10.1029/2006wr005631.

[12] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Oakland, CA, USA, vol. 1, 1967, pp. 281–297.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in python," 2012. DOI: 10.48550/ARXIV.1201.0490.