https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

Technical Note: A simple feedforward artificial neural network for high temporal resolution classification of wet and dry periods using signal attenuation from commercial microwave links

Erlend Øydvin[1], Maximilian Graf[2, 3], Christian Chwala[2, 3], Mareile Astrid Wolff[1, 5], Nils-Otto Kitterød[4], and Vegard Nilsen[1]

[1]Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway
[2]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Campus Alpin, Garmisch-Partenkirchen, Germany
[3]Institute of Geography, University of Augsburg, Augsburg, Germany
[4]Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway
[5]Norwegian Meteorological Institute, Oslo, Norway

Correspondence: Erlend Øydvin (erlend.oydvin@nmbu.no)

Abstract. Two simple feedforward neural networks (MLPs) are trained to classify wet and dry periods using signal attenuation from commercial microwave links (CMLs) as predictors and high temporal resolution reference data as target. MLP_{GA} is trained against nearby rain gauges and MLP_{RA} is trained against gauge-adjusted weather radar. Both MLPs perform better than existing methods, showcasing their effectiveness in capturing the intermittent behaviour of rainfall. This study is the first using both radar and rain gauges for training and testing for CML wet-dry classification. Where previous studies has mainly focused on hourly reference data, our findings show that it is possible to predict wet and dry periods with a higher temporal precision.

1 Introduction

By exploiting the relation of rainfall intensity to signal attenuation, commercial microwave links (CMLs) can be used to estimate path-average rainfall between telecommunication towers (Messer et al., 2006; Leijnse et al., 2007). As the signal is also attenuated by factors other than rain, such as air humidity, these non-rainy factors must be taken into account in what is often called the baseline attenuation. Rain-induced attenuation can then be estimated by subtracting the estimated baseline from the total loss, where the baseline is typically estimated from the mean signal attenuation shortly before the rainfall event (Chwala and Kunstmann, 2019). This makes rain event detection a crucial step in deriving rainfall rates from CMLs. CML signal loss is recorded differently depending on the network operator and can for instance be available as instantaneous measurements every minute. Another popular format is to record the minimum and maximum signal loss over a period, typically 15 minutes. In this work, we focus on instantaneously sampled CML data as this data is becoming more and more available, see for instance Andersson et al. (2022).

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

During wet periods, the CML signal loss tends to fluctuate more than during dry periods. Based on this, a simple method for rain event detection was developed by Schleiss and Berne (2010). They suggested using these fluctuations to predict wet periods by taking the standard deviation of a 60-minute rolling window and setting time steps with values above a certain threshold to wet. This threshold is different between CMLs, but can be derived from local climate characteristics. Graf et al. (2020) expanded this method by recognizing that climate characteristics is not necessarily valid for different locations, individual years and in particular specific rainy periods that might be of interest. They proposed to estimate the threshold by computing the 80 % quantile of the 60-minute rolling standard deviation for each CML and multiplying this number by a constant that was found to be similar for all CMLs in the study. A more data-driven approach was explored by Polz et al. (2020). They trained a convolutional neural network (CNN) to detect wet periods using 800 CMLs in Germany. As a reference, they used the gauge-adjusted radar product RADOLAN-RW from Germany's National Meteorological Service (DWD) which has an hourly resolution. Another approach is to include the signal loss from nearby CMLs (Overeem et al., 2011). This method was shown to work for dense CML networks. The literature describes several other approaches (Habi and Messer, 2018; Reller et al., 2011; Rayitsfeld et al., 2012; Wang et al., 2012).

Most of the mentioned methods have been evaluated using hourly reference data. This might be a reasonable approach as rainfall detection is mostly used for estimating the baseline, which is typically set as a constant throughout a rainfall event (Chwala and Kunstmann, 2019; Uijlenhoet et al., 2018; Messer and Sendik, 2015). However, existing methods are not optimized for predicting rainfall on a higher temporal resolution, and thus, the predictions might not reflect the true intermittency of rainfall. Predicting too long wet periods could result in the CML baseline not being adapted to new time steps, possibly introducing a bias in the rainfall retrieval. Further, a drawback of predicting too long rainy periods is that some of the predicted rainy time steps could contain non-liquid precipitation. As dry snow induces a very low signal attenuation, these time steps appears as dry in the CML time series. In an event where the precipitation type changes between rain and snow, classifying dry snow events in the CML signal as dry is important as the presence of precipitation in a nearby rain gauge could then indicate that it is in fact snowing.

In this paper, we present two methods to better detect wet periods in highly intermittent rainfall. One method is trained on radar reference data and the other method is trained on rain gauge reference data. Both methods are tested against rain gauge and radar data, highlighting their differences. We also examine the performance of the developed methods in comparison to existing approaches, aiming to gain a clearer understanding of the distinctions between these various methodologies.

2 Methods

2.1 Data

A large dataset with 3901 CMLs from Germany was used, providing transmitted and received signal levels with a temporal resolution of one minute from 01-07-2021 to 31-07-2021. The total signal loss (TL) was computed by subtracting the transmitted signal level from the received signal level. Each CML consists of two time series called sublinks, reflecting the signal loss in the beams going from location 0 to 1 and vice versa. More information on this dataset can be found in Graf et al. (2020). As

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

ground truth, two different sources were explored. The first used rain gauges near the CMLs provided by DWD. The rain gauge data was provided with a temporal resolution of one minute and volume resolution of 0.01 mm. We consider a minute to be wet if the rain gauge records any rainfall. The other source was the radar product *RADKLIM-YW* (Winterrath et al., 2018). This product from DWD is a gauge-adjusted, climatologically corrected product with a temporal resolution of 5 minutes. For the comparison with CML data, the radar product was averaged over the CML path intersections, with each grid value weighted by the length of the CML path in each grid cell. For comparison of the path-averaged *RADKLIM-YW* reference and the CML rainfall estimates, *RADKLIM-YW* was resampled from a 5-minute resolution to a 1-minute resolution by linear interpolation and then dividing the rainfall sums by 5. To make it comparable to the rain gauges, minutes with rainfall above 0.01 mm were set to wet.

2.2 Neural network

In our approach, we have used a simple feed-forward neural network provided by the python library *sklearn* (Pedregosa et al., 2011). This network consists of an input layer, fully connected hidden layers, and an output layer. Networks with simple architecture of this type are often referred to as a Multilayer perceptron (MLP). The input layer takes the total signal loss from a 40 time steps long centered moving window over both sublinks. The CML predictor data is organized in a so-called design matrix (Equation 1) where $tl_{s_1,t}$ and $tl_{s_2,t}$ represents the total signal loss at time step t for sublink 1 and sublink 2 respectively.

$$\begin{bmatrix} tl_{s_1,t_0-20} & \dots & tl_{s_1,t_0+20} & tl_{s_2,t_0-20} & \dots & tl_{s_2,t_0+20} \\ \vdots & & \vdots & \vdots & & \vdots \\ tl_{s_1,t_i-20} & \dots & tl_{s_1,t_i+20} & tl_{s_2,t_i-20} & \dots & tl_{s_2,t_i+20} \\ \vdots & & \vdots & \vdots & & \vdots \\ tl_{s_1,t_n-20} & \dots & tl_{s_1,t_n+20} & tl_{s_2,t_n-20} & \dots & tl_{s_2,t_n+20} \end{bmatrix} \tag{1}$$

We experimented with longer windows, but could not find any improvements by increasing the window size beyond 40 time steps. There was also an improvement from using both sublinks rather than one. We do not show these findings in detail in this note.

As pre-processing, we subtracted the 12 hours centered rolling median from the signal level for each CML. This removes longer trends from the signal level making the time series stationary. We experimented with other detrending methods such as differencing, but got poorer results.

Next, two approaches were explored, one where we trained the neural network against radar data (MLP$_{RA}$) and one where we trained the MLP against rain gauge data (MLP$_{GA}$). For testing, the optimal MLP$_{RA}$ and MLP$_{GA}$ were integrated in to *pycomlink*, a python library for CML processing (Chwala et al., 2023). Since the current *pycomlink* environment does not support *sklearn*, the weights and network architecture were exported to *tensorflow* using the *Keras* API (Abadi et al., 2015). The final testing was performed by loading the exported MLPs from the pycomlink environment.

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

2.3 Reference methods

Two reference methods were used for comparing the MLP results. The σ_{80} method from Graf et al. (2020) and the CNN method from Polz et al. (2020). Both methods are described in the introduction and can be run from pycomlink.

2.4 Performance metrics

The performance of the methods was evaluated by recording the predicted CML wet and dry periods against the reference data (rain gauge or radar) in a confusion matrix. In our case, the confusion matrix is a 2x2 matrix listing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Although no perfect performance metric exists, a balanced way of describing the confusion matrix as a single number can be done by the Matthews correlation coefficient (MCC) (Chicco and Jurman, 2020). The MCC is a diagnostic that gives a number between -1 and 1, where 1 represents a perfect prediction, 0 is no better than a random prediction, and -1 is a perfect disagreement with the reference.

2.5 CMLs close to rain gauges

Pairs of CMLs and rain gauges that are closer to each other than 5 km were selected for training and testing the MLPs. This resulted in 395 pairs of CMLs and rain gauges spread out across Germany. All pairs are covered by the *RADKLIM-YW* radar product.

2.6 Train-test split

In order to assess how well the models performed, the CML data was split into a training set and a test set. Due to, for instance, noisy CMLs, malfunctioning rain gauges, or spatio-temporal uncertainties, some CMLs showed a poor correlation with the rain gauges or the radar. As these pairs could result in poor training data, we opted to exclusively include pairs with high MCC in our training set. We selected training pairs for MLP$_{RA}$ and MLP$_{GA}$ by predicting the CML wet periods using the σ_{80} method. The top 26 CML-radar pairs with the highest MCC, estimated using radar data, were chosen for MLP$_{RA}$. MLP$_{GA}$ used CML-rain gauge pairs with the highest MCC, estimated using rain gauge. The remaining 369 pairs were used for testing. A possible drawback of this approach is that the MLPs are not trained on noisy CMLs, hindering their effectiveness in dealing with erratic signal fluctuations. However, erratic CMLs are usually removed before the rain event detection step for instance by removing CMLs where the rolling standard deviation of the total loss exceeds 2 dB at least 10% of the time or where the 1 hour rolling standard deviation of the of the total loss exceeds 0.8 dB at least 33% of the time (Graf et al., 2020; Blettner et al., 2023).

2.7 Hyperparameter estimation and cross-validation

During training, the MLP classifier can be tuned using several hyperparameters such as activation function, hidden layers, initial learning rate, and L2 regularization. The optimal hyperparameters were found by using k-fold cross-validation over a grid search over the hyperparameter values listed in Table 1. We performed k-fold cross-validation by splitting the CMLs in

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

Table 1. MLP hyperparameters used in grid search

Hyperparameter	**Values**
Hidden Layer Sizes	[[1], [10], [20], [70], [5, 5], [10, 10], [50, 50], [100, 100]]
Activation Function	['relu', 'logistic']
Regularization	[0., 0.175, 0.35, 0.525, 0.7]
Initial Learning Rate	[0.0000001, 0.00000147, 0.00002154, 0.00031623, 0.00464159, 0.06812921, 1]

the training data into 5 folds and iteratively trained the MLP on 4 folds of data and validated on the 5th fold using the MCC. The final score is the mean of all 5 validation MCC scores.

The rainfall time series is characterized by extended periods of no rain, leading to an imbalance that can impede the effectiveness of neural network training. A common method to address this issue is random undersampling, where samples from the majority class are discarded to create a balanced dataset (Hoens and Chawla, 2013). However, rainfall time series often include short intermittent dry periods within longer events, which are of particular interest in our approach. If we were to use random undersampling, these events might be underrepresented in the training dataset. Recognizing that the total signal loss moving window can include rainy time steps during dry periods close to wet ones, we have adopted a modified undersampling strategy. Specifically, we only discard dry steps more than 30 minutes away from any rainfall events as detected by the reference methods.

3 Results and discussion

3.1 Training the MLP

The MCC, given optimal given optimal initial learning rate and regularization, as a function of an increasing number of neurons and hidden layers for the MLP$_{RA}$ and the MLP$_{GA}$ for both activation functions is presented in Fig. 1. For each hidden layer configuration, the optimal regularization and initial learning rate that yielded the highest mean MCC were selected and plotted together with the minimum and maximum of all 5 folds obtained from k-fold cross-validation.

We can observe that the MLP$_{GA}$ generally has a lower score than the MLP$_{RA}$ method. This could be because of the spatial differences between the CMLs and rain gauges, causing errors related to spatial uncertainty. For the radar data, this spatial representation is most likely mitigated by the comparison based on CML path-weighted intersects. Another reason could be that the spatial averaging performed by the radar and CMLs produces less intermittent rainfall time series than what is the case for the rain gauges, resulting in better agreement between the CML and radar.

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

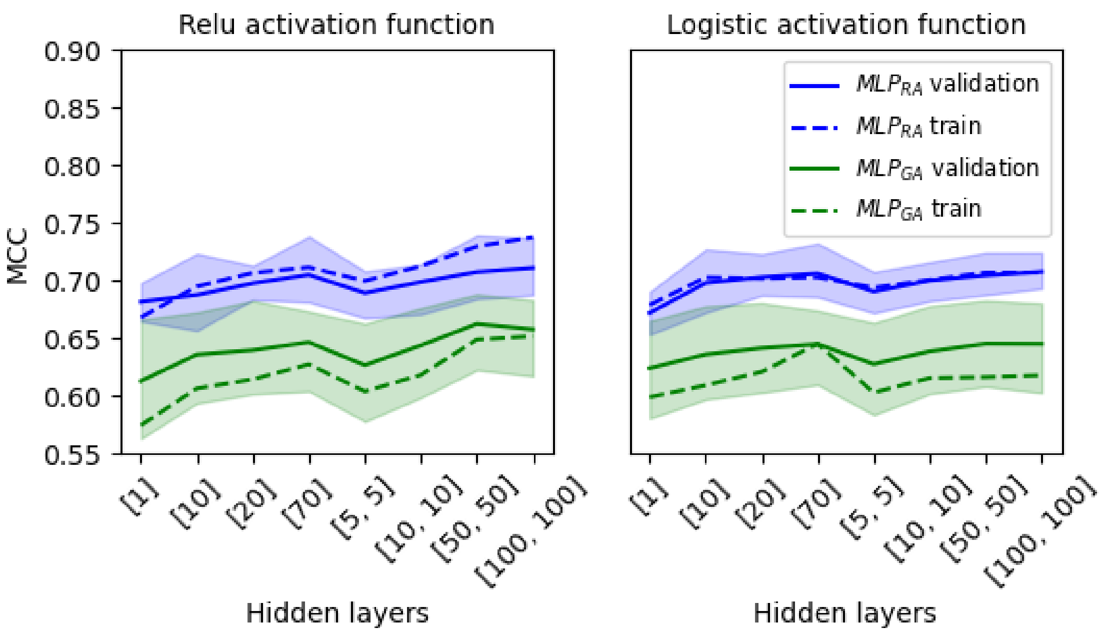

Figure 1. MCC as a function of network architecture for the relu and logistic activation function. [5, 5] means two layers with 5 neurons in each layer. The MLP was trained using k-fold cross-validation with 5 folds over 26 CML-rain gauge pairs using radar (MLP$_{RA}$) and rain gauge (MLP$_{GA}$) as reference. The solid line is the mean value of the 5 folds while the shaded area shows the minimum and maximum score of the 5 folds.

We can also observe that models using the logistic activation function generally seem to perform more consistently for all network architectures than the relu activation function. The relu activation function has a lower score for simple network architectures (for instance [1]), but produces larger scores with increased network architecture compared to the logistic activation function. Further, for the relu activation function with larger networks ([70] and [100, 100]), MLP$_{RA}$ shows a larger deviation between the train set and validation set, indicating that the model is not generalizing very well. MLP$_{RA}$ has a smaller deviation between train and validation when the logistic activation function is used, indicating more general fits. Thus MLP$_{RA}$ seems to have a good compromise between model complexity and score when using a single layer with 20 neurons and the logistic activation function. MLP$_{GA}$ on the other hand has a smaller deviation between the train and validation set and provides a good compromise between model complexity and score when using two layers with 50 neurons in each and the relu activation function. The optimal hyperparameters for MLP$_{RA}$ and MLP$_{GA}$ are shown in Table 2.

3.2 Testing the MLP

The MCC scatter plot density for the MLP$_{RG}$ and MLP$_{RA}$ method compared with the benchmark methods σ_{80} and CNN using the radar and rain gauge test data as reference is presented in Fig. 2. For both radar and rain gauge reference we can observe

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

Table 2. Optimal hyperparameters for the MLP trained with radar reference (MLP$_{RA}$) and the MLP trained with rain gauge reference (MLP$_{GA}$)

Hyperparameter	**MLP$_{RA}$**	**MLP$_{GA}$**
Network architecture	[20]	[50, 50]
Activation function	logistic	relu
Regularization	0.175	0.175
Initial learning rate	0.00031623	0.00031623

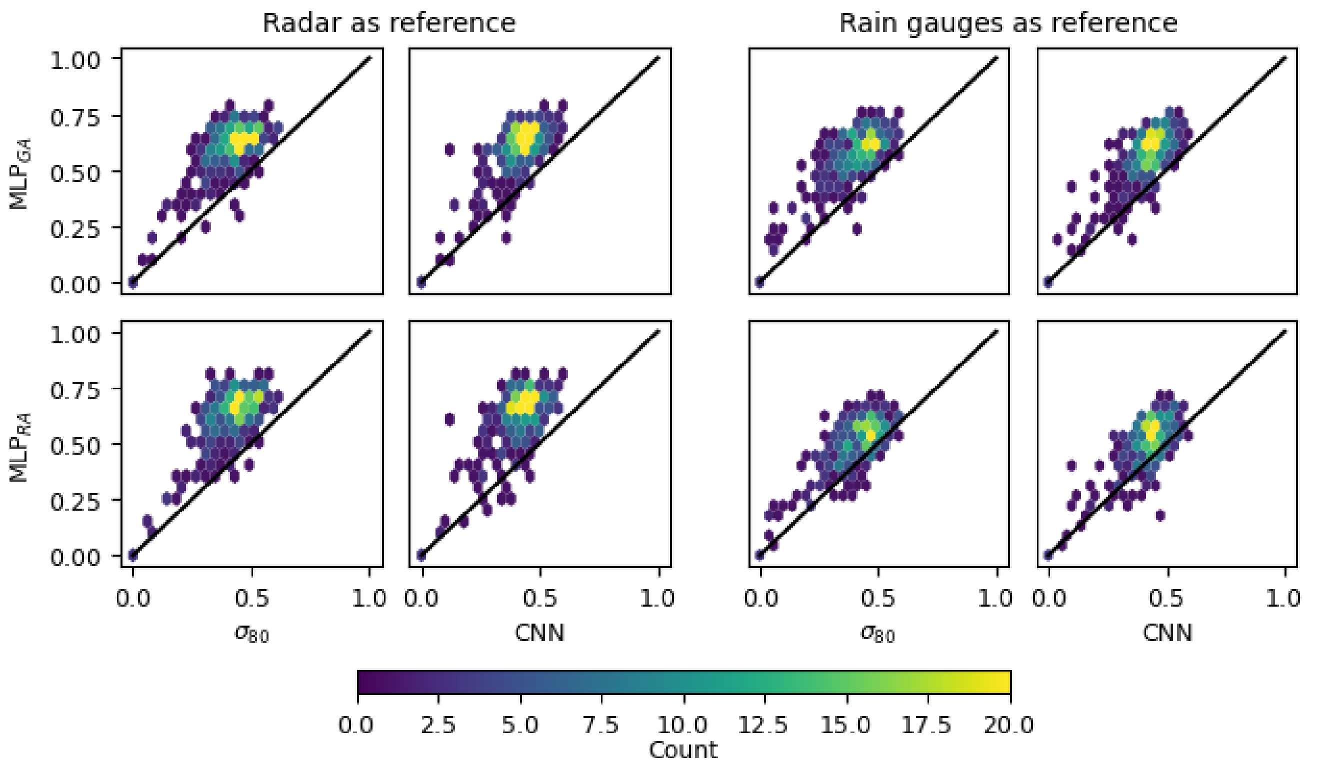

Figure 2. Scatter density plot of the MCC score for the MLP trained on the rain gauge reference (MLP$_{GA}$) and the MLP trained on the radar reference (MLP$_{RA}$) compared with the benchmark methods σ_{80} and CNN. The left plot used radar as reference and the right plot used rain gauges as reference. CML, radar and rain gauge uses a one minute resulution. Scores were computed based on 369 CML-radar data pairs over one month.

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

that for most data pairs, the MCC score is higher when using one of the MLP methods than when using one of the reference methods. Another observation is that MLP_{GA} performed slightly better (median MCC of 0.59) than MLP_{RA} (median MCC of 0.52) when the rain gauge was used as a reference. When the radar was used as a reference MLP_{RA} scored slightly better (median MCC of 0.62) than MLP_{GA} (median MCC of 0.66). This difference could be explained by the inherent differences in the measurement methods, where the rain gauge captures the rainfall differently than the weather radar due to for instance wind.

In Fig. 3, Fig. 4, and Fig. 5 we plot the CML signal loss as a function of time as well as the predicted wet periods for all methods and the ground truth. We also plot the confusion matrix and the corresponding MCC score for each method using the rain gauge as a reference.

Fig. 3 shows the results from a 10-hour long period for a CML where the MLP_{RA} method (MCC: 0.71) and MLP_{GA} method (MCC: 0.74) outperformed the CNN method (MCC: 0.08) and the σ_{80} (MCC: 0.47). Looking at the CML total loss (TL) we can observe that the CML behaves nicely with a relatively constant baseline outside of wet periods. Around time 06:00 the radar reference (RA) shows a short wet period, while the rain gauge shows a longer highly intermittent wet period. The intermittent behavior of the rain gauge might be due to low-intensity rainfall or smaller droplets falling into the bucket from the collector. Both MLPs were able to detect a short wet period at this time. For the full 10 hours, the CNN in general predicts a very long wet period, missing several dry events and leading to a poorer MCC. This is not surprising as it was trained to detect wet events on an hourly basis. The σ_{80} method was better in classifying the dry events but still predicted longer wet periods than the MLPs. Further, MLP_{RA} tended to predict wet periods that started shortly before the CML TL starts to rise, while the MLP_{GA} tended to predict wet periods shortly after the TL has started to rise. This is an interesting feature and could be due to the rain gauges showing short breaks at the beginning of rainfall events due to low rainfall intensity. If the beginning of a wet event has more dry minutes than wet minutes, as seen by the rain gauge, this could lead MLP_{GA} to just predict no rain on these occasions. It could also be due to that the radar observes rainfall before it is measured on the ground, making the MLP_{RA} estimate rainfall shortly before MLP_{GA}

Fig. 4 shows a 6-hour case for a different CML where the difference between the MLP_{RA} and MLP_{GA} method is easier to spot. Like in Figure 3, MLP_{RA} predicts the wet starting point before the MLP_{GA} does. As in the previous case, the CNN predicts a very long wet period, while the σ_{80} predict rain before and after the rain gauge and radar reference rainfall prediction. In this instance, none of the methods can accurately predict the reference wet periods. Looking at the TL we can see that it increases gradually over an extended period, suggesting a longer wet period. In contrast, the reference data only indicates one or two short wet events. This discrepancy may be attributed to very low rainfall rates, causing an elevated TL due to CML wet antenna attenuation. However, these rates might be too small to register on the rain gauge or radar.

Fig. 3 and Fig. 4 also raise some interesting questions. The final rainfall amounts is often derived from a baseline that is typically estimated based on the values of the dry periods before the wet event. Since these baseline values are estimated differently for the different methods we have explored in this study, the resulting rainfall rates are expected to vary. For instance, if the MLP_{GA} is used, the baseline would be placed at a higher level than if the MLP_{RA} method was used, resulting in a lower rainfall rate estimate. Looking at Figure 3 and the first and last rainfall event detected by MLP_{GA} (time steps 01.00 and 08.00),

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

it is clear that MLP$_{GA}$ predicts rainfall shortly *after* the TL has started to rise. If we assume that the TL in these two cases is only affected by raindrops, then MLP$_{GA}$ would produce a too-high baseline estimate. MLP$_{RA}$, on the other hand, seems to better capture the entire rainfall event and thus is might be more suitable for baseline estimation. A more detailed analysis of these effects is beyond the scope of this paper.

In Figure 5 we have depicted the TL as well as the predicted wet periods and reference wet periods for a CML with more erratic signal fluctuations. For σ_{80}, multiple wet periods are estimated. While these estimated wet periods may seem plausible when observing the TL, the reference data reveals that there is no actual rainfall during this time. Therefore, the wet predictions likely stem from a noisy CML signal.

Overall it must be noted that while the MCC is a useful and balanced metric, its score must be seen in relation to the reference chosen for evaluation. As weather radar provides average rainfall intensities for the entire radar grid cell, we expect that the radar rainfall estimates are less intermittent than what is observed by a rain gauge. This is supported by the findings in Figure 3, Figure 4, and Figure 5 where the weather radar rainfall events are less intermittent than what is the case for the rain gauges. The CML, like the weather radar, also measures spatially averaged rainfall. However, the CML measures rainfall closer to the ground and might thus be able to better capture the intermittency as seen by the rain gauge. In this study MLP$_{GA}$ was able to better detect rainfall events as seen by the rain gauge than MLP$_{RA}$. This suggests that there is no single best reference or method for evaluating CML rainy periods. Rather, the CML rain event detection method must be seen in relation to its application.

4 Conclusions

In this technical note, we introduced a simple feedforward neural network (MLP) designed to detect intermittent rainfall from CML signals at a higher temporal resolution compared to existing methods. Our approach involved training the MLPs on reference data from rain gauges (MLP$_{GA}$) with a temporal resolution of 1 minute and gauge-adjusted radar (MLP$_{RA}$) with a temporal resolution of 5 minutes. Both MLPs outperformed the two reference methods. MLP$_{GA}$ typically predicts rainfall shortly after MLP$_{RA}$ and often after the CML total loss has started to increase. Thus, if the MLP$_{GA}$ method is used, the user should consider setting for instance 5 minutes before and after a wet event to wet, similar to Pastorek et al. (2022). Moreover, MLP$_{GA}$ better predicts wet periods as recorded at the nearby rain gauges than what is the case for MLP$_{RA}$, while both methods perform equally well when radar data is used as reference. Thus, the different methods capture different nuances of the rainfall patterns.

Future work may involve further refining the model architecture and testing its robustness in generalization to other datasets. Another interesting topic could be to better understand how different wet and dry classifications affect the resulting baselines and the effect this has on rainfall rate estimation from CML data. Overall, both MLPs showed successful skill for the challenge of rainfall event detection in CML attenuation time series.

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

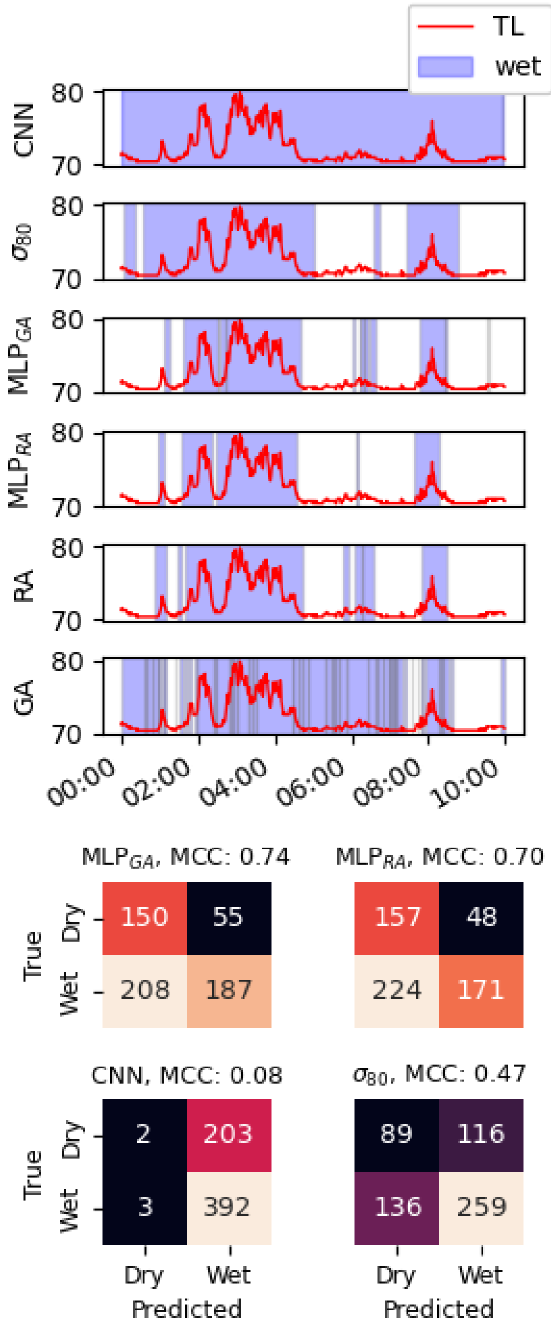

Figure 3. CML signal loss (TL) for a 10-hour long interval and its corresponding confusion matrix (compared to rain gauge reference) and MCC score for the CNN, σ_{80}, MLP$_{RA}$, MLP$_{GA}$ methods. The reference wet periods for the rain gauge (RG) and gauge-adjusted radar (RA) were also plotted. The blue shaded area mark the wet periods and its borders were colored grey to highlight the intermittent behavior.

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

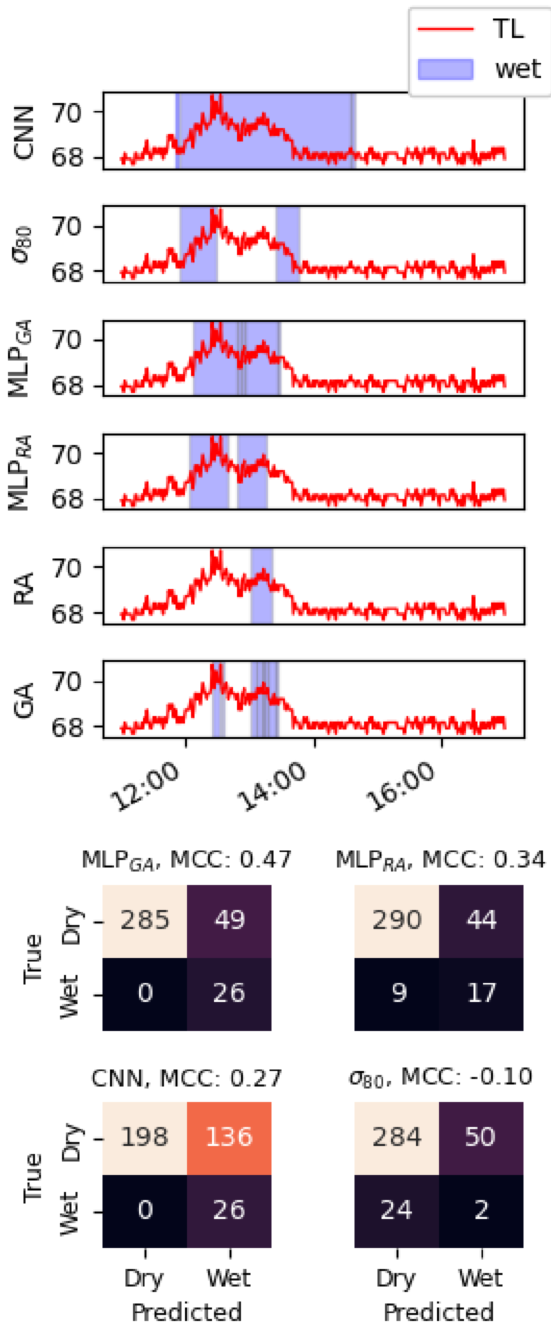

Figure 4. CML signal loss (TL) for a 6-hour long interval and its corresponding confusion matrix (compared to rain gauge reference) and MCC score for the CNN, σ_{80}, MLP$_{RA}$, MLP$_{GA}$ methods. The reference wet periods for the rain gauge (RG) and gauge-adjusted radar (RA) was also plotted. The blue shaded area mark the wet periods and its borders was colored grey to better show the intermittent behavior.

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

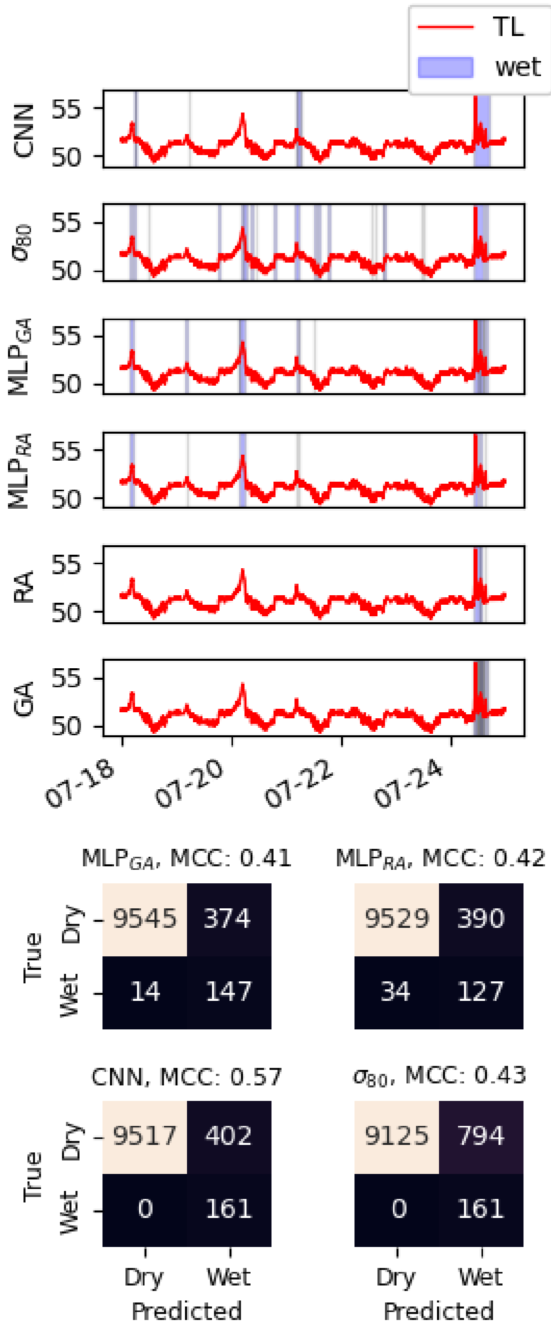

Figure 5. CML signal loss (TL) for a 6 day long interval and its corresponding confusion matrix (compared to rain gauge reference) and MCC score for the CNN, σ_{80}, MLP_{RA}, MLP_{GA} methods. The reference wet periods for the rain gauge (RG) and gauge adjusted radar (RA) was also plotted. The blue shaded area mark the wet periods and its borders was colored grey to better show the intermittent behavior. Here the CNN outperformed the other methods as it was able to better classify the noisy CML signal as dry, which was more in line with the reference.

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

Code availability. The MLP_{RA} method and the MLP_{GA} method are available from pycomlink under https://github.com/pycomlink/pycomlink/tree/master/pycomlink/processing/wet_dry. An example notebook running the different wet dry classification methods is available under https://github.com/pycomlink/pycomlink/tree/master/notebooks

Data availability. The rain gauge data was derived from the open data server of the German Meteorological Service and can be found here: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/1_minute/precipitation/.

Author contributions. Conceptualization: EØ, CC. Data curation: CC. Methodology: EØ, CC, MG, VN, MW, NOK. Software: EØ, MG, CC. Supervision: VN, MW, NOK. Writing – original draft preparation: EØ. Writing – review and editing: EØ, MG, CC, VN, MW, NOK.

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Acknowledgements. The authors thank co-supervisor Etienne Leblois for nice discussions. The authors would also like to express their gratefulness to the OpenSense COST action (CA20136) for facilitating a short term scientific mission to Garmisch that contributed in making this work possible. We would also like to acknowledge the access to RADOLAN-YW data from the German Weather Service and Ericsson for providing CML data. This work is funded by the Norwegian University of Life Sciences and the German Research Foundation via the SpraiLINK project (Grant CH-1785/2-1).

https://doi.org/10.5194/egusphere-2024-647
Preprint. Discussion started: 4 April 2024

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/, software available from tensorflow.org, 2015.

Andersson, J. C. M., Olsson, J., van de Beek, R. C. Z., and Hansryd, J.: OpenMRG: Open data from Microwave links, Radar, and Gauges for rainfall quantification in Gothenburg, Sweden, Earth System Science Data, 14, 5411–5426, https://doi.org/10.5194/essd-14-5411-2022, 2022.

Blettner, N., Fencl, M., Bareš, V., Kunstmann, H., and Chwala, C.: Transboundary Rainfall Estimation Using Commercial Microwave Links, Earth and Space Science, 10, https://doi.org/10.1029/2023EA002869, 2023.

Chicco, D. and Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genomics, 21, 6, https://doi.org/10.1186/s12864-019-6413-7, 2020.

Chwala, C. and Kunstmann, H.: Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges, WIREs Water, 6, https://doi.org/10.1002/wat2.1337, 2019.

Chwala, C., Polz, J., Graf, M., Sereb, D., Blettner, N., Keis, F., and Boose, Y.: pycomlink/pycomlink: v0.3.2, https://doi.org/https://doi.org/10.5281/ zenodo.4810169, 2023.

Graf, M., Chwala, C., Polz, J., and Kunstmann, H.: Rainfall estimation from a German-wide commercial microwave link network: optimized processing and validation for 1 year of data, Hydrology and Earth System Sciences, 24, 2931–2950, https://doi.org/10.5194/hess-24-2931-2020, 2020.

Habi, H. V. and Messer, H.: Wet-Dry Classification Using LSTM and Commercial Microwave Links, pp. 149–153, IEEE, ISBN 978-1-5386-4752-3, https://doi.org/10.1109/SAM.2018.8448679, 2018.

Hoens, T. R. and Chawla, N. V.: Imbalanced Datasets: From Sampling to Classifiers, pp. 43–59, Wiley, https://doi.org/10.1002/9781118646106.ch3, 2013.

Leijnse, H., Uijlenhoet, R., and Stricker, J. N. M.: Rainfall measurement using radio links from cellular communication networks, Water Resources Research, 43, 1–6, https://doi.org/10.1029/2006WR005631, 2007.

Messer, H. and Sendik, O.: A New Approach to Precipitation Monitoring: A critical survey of existing technologies and challenges, IEEE Signal Processing Magazine, 32, 110–122, https://doi.org/10.1109/MSP.2014.2309705, 2015.

Messer, H., Zinevich, A., and Pinhas, A.: Environmental Monitoring by Wireless Communication Networks, Science, 312, 17–18, https://www.jstor.org/stable/3846088, 2006.

Overeem, A., Leijnse, H., and Uijlenhoet, R.: Measuring urban rainfall using microwave links from commercial cellular communication networks, Water Resources Research, 47, https://doi.org/10.1029/2010WR010350, 2011.

Pastorek, J., Fencl, M., Rieckermann, J., and Bares, V.: Precipitation Estimates From Commercial Microwave Links: Practical Approaches to Wet-Antenna Correction, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–9, https://doi.org/10.1109/TGRS.2021.3110004, 2022.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Polz, J., Chwala, C., Graf, M., and Kunstmann, H.: Rain event detection in commercial microwave link attenuation data using convolutional neural networks, Atmospheric Measurement Techniques, 13, 3835–3853, https://doi.org/10.5194/amt-13-3835-2020, 2020.

Rayitsfeld, A., Samuels, R., Zinevich, A., Hadar, U., and Alpert, P.: Comparison of two methodologies for long term rainfall monitoring using a commercial microwave communication system, Atmospheric Research, 104-105, 119–127, https://doi.org/10.1016/j.atmosres.2011.08.011, 2012.

Reller, C., Loeliger, H.-A., and Díaz, J.: A model for quasi-periodic signals with application to rain estimation from microwave link gain, European Signal Processing Conference, 2011.

Schleiss, M. and Berne, A.: Identification of Dry and Rainy Periods Using Telecommunication Microwave Links, IEEE Geoscience and Remote Sensing Letters, 7, 611–615, https://doi.org/10.1109/LGRS.2010.2043052, 2010.

Uijlenhoet, R., Overeem, A., and Leijnse, H.: Opportunistic remote sensing of rainfall using microwave links from cellular communication networks, WIREs Water, 5, https://doi.org/10.1002/wat2.1289, 2018.

Wang, Z., Schleiss, M., Jaffrain, J., Berne, A., and Rieckermann, J.: Using Markov switching models to infer dry and rainy periods from telecommunication microwave link signals, Atmospheric Measurement Techniques, 5, 1847–1859, https://doi.org/10.5194/amt-5-1847-2012, 2012.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Lengfeld, K., Walawender, E., Weigl, E., and Becker, A.: RADKLIM Version 2017.002: Reprocessed quasi gauge-adjusted radar data, 5-minute precipitation sums (YW), https://doi.org/10.5676/DWD/RADKLIM_YW_V2017.002, 2018.