Using Markov switching models to infer dry and rainy periods from telecommunication microwave link signals

Response to reviewers

by Z. Wang, M. Schleiss, J. Jaffrain, A. Berne and J. Rieckermann

The authors would like to thank the reviewers for their valuable comments and suggestions. Enclosed, please find our point-by-point answers to the reviewers' comments. For a better visibility, all major changes have been highlighted in yellow in the attached pdf version of the manuscript.

Anonymous Referee # 1

The paper presents relevant information and procedures on how to convert a microwave/mm-wave radio link into a rain rate measurement instrument. I found it very interesting they way in which the authors have been able to increase the rate the attenuation can be read and recorded. The references provide a very good overview of related works. The application of the instrument is put in perspective as something in between point measurements and large volume measurements. The authors clearly point out the problems involved in the measurement, namely the lack of a stable baseline for identifying the dry from the wet periods. The identification of the baseline is the objective of the paper where a State switching model is proposed where different behaviors are to be expected.

Specific comments

1. *"I think that the Gaussian distribution is not a well-accepted one for rain rate; it is rather the log-normal distribution that is normally accepted, at least for the tail of the distribution."*

   The authors agree that the probability distribution of rain rates (and therefore of path-integrated attenuation during rainy periods) is not Gaussian. The effects of this model mismatch on the classification performance are, however, very limited. This can be seen in Figure 1, which shows the probability density functions of attenuation values for dry and rainy periods. The sample distributions are not exactly Gaussian but the fact that the tails of the distributions are not correctly reproduced is not critical with respect to the classification problem. In fact, the optimal classification threshold which is at the intersection between the two empirical probability density functions (i.e., about 49 dB) is very close to the threshold derived from the Gaussian model (i.e., the intersection between the two Gaussian densities). Similar results are obtained for all channels and all considered datasets and show that the Gaussian error assumption is not critical with respect to the classification problem.

   Figure 1 and some additional explanations have been added to the manuscript in Section 5 to clarify this point.

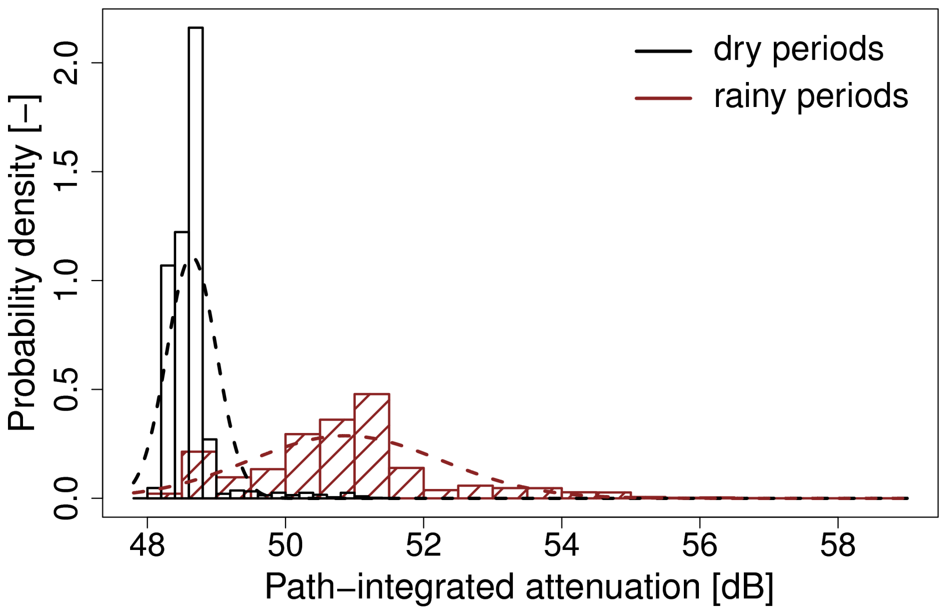

Figure 1: Empirical probability density functions of attenuation values for dry and rainy periods (for dataset 1). The dashed lines represent the fitted densities of a Gaussian distribution with same mean and variance as the samples. The dry and rainy periods are derived from the disdrometer data.

2. *"Also, in the methodology in this section, the assumption is made that the samples used are uncorrelated. This is not the case since the variations during dry periods are fairly slow, with a period of one day. As for the rainy samples, uncorrelated rain attenuation samples at 4 s are unlikely, in addition to the superposed temperature drifts."*

   Clearly, the samples are not independent. The authors performed several tests using more complex and time-dependent formulations (e.g., using an AR(1) model) but none of them significantly improved the classification performance. Moreover, in many cases, the more complex models did not converge or converged too slowly to be useful in practical applications. The AR(0) model has therefore been chosen as a good trade-off between classification performance and computational efficiency. This fact is now clearly mentioned in the manuscript, at the beginning of Section 5.

3. *"In the same section, I would take some time to present and define the elements in equation (7)."*

   These terms represent the probability densities of the attenuation measurements (for a given state) and the associated state probabilities (for a given set of model parameters Θ). An additional sentence has been added in the manuscript in order to clarify this point.

4. *"Still in Section 2.2, the authors claim that the other techniques for identify the baseline need to set an empirically derived threshold. However, in Equation (8), a threshold of $\frac{1}{2}$ is arbitrarily chosen."*

   This threshold is not arbitrary. In absence of any additional external information, the optimal way of assigning the states is by choosing the one that has the largest probability. In the case of two states (dry/wet), this leads to a threshold of $\frac{1}{2}$. This explanation has been added to the manuscript in Section 2.2.

5. *"In Section 2.3, the assumption of independence for a multivariate approach, it looks a-priory to be hardly fulfilled. I suspect that all channels are similarly affected by the same thermal drift."*

   In theory, the proposed multivariate method can account for any valid correlation structure. However, since the correlation structure is not known a-priori (and cannot be estimated before the classification), the authors prefer to treat the channels as if they were independent. This might not be optimal but the results show that the additional information from the other channels significantly improves the classification performance compared to the univariate case. This will be investigated more closely in the future.

6. *"With respect to Figures 2 and 3, what are the numbers in the ordinates? They seem to be very far from the free space loss value plus some small rain-induced loss."*

   The numbers represent the path-integrated attenuation, i.e., the difference between the received and the transmitted power (in dB). This includes the losses at the antennas, the atmospheric attenuation and possible rain-induced attenuations. In our experience, such values are common for a 1.85-km link at 38 GHz.

7. *"Could the authors quantify the wet antenna effect? I have the impression that the range of attenuation values measured in this very short link (Figures 2 and 3), i.e., not much more than 2 dB, is almost comparable with the wet antenna loss."*

   The authors are still investigating the wet antenna effect and cannot answer this question at the moment. Wet antenna effects are of interest but clearly beyond the scope of this paper. So far, preliminary analyses suggest that wet antenna attenuations are smaller than 2 dB. The rain-induced attenuations on the other hand can be much higher (e.g., more than 10 dB during intense rainfall). Further work will be conducted on this question.

8. *"In section 4 there is a discussion on errors of type I and II and their rates which seem to be fairly high, especially in the non-stationary case. I wonder, what is their impact in practical application (meteorology, flood control, etc.) for example? The statistics of rain, will they be very much off?"*

   The effects of dry/wet misclassification on rain rate estimation are very difficult to evaluate. Rain rate estimates are mostly influenced by the method used to fit the attenuation baseline, and not so much by the dry/wet classification errors. Preliminary analysis suggest that the errors are about ± 25%. In how far errors in the disdrometer and raingauge data can be taken into account, is subject of current research. The problems mentioned by the reviewer clearly require major additional investigations and are therefore far beyond the scope of this paper.

9. *"Equation (11) is missing some constants in front of the second and third terms on the right-hand side. I believe it is ρ and then $\sqrt{1-\rho^2}$."*

   This equation has been removed from the text because it was not essential for the understanding of the paper. Some additional explanations about the advantages and limitations of the AR(1) model have been added instead.

10. *"After equation (11), there is a mention of using transition probabilities p_{00}, p_{11} (and p_{01}, p_{10}). In my experience with a first order Markov model it is not possible to actually reproduce all possible durations of events, wet and dry, short and long."*

The proposed first order Markov model uses transition probabilities to relate the current state to the immediate previous state. The transition probabilities therefore only represent the likelihood of a state change, and not the event durations.

Technical comments

The suggested text corrections have been taken into account.

Anonymous Referee # 2

This paper presents a good primer on the use of microwave radio links as a method for estimating the rain rate along a given path, and clearly places the use of this method in the context of other rain measurement techniques such as rain gauges (point rainfall) and rain radar (measuring rain in volumes and over areas). It is well known that microwave links experience attenuation from sources other than rain (including clouds and atmospheric gases), especially as one gets up to frequencies such as the one used in this study (38 GHz). It is therefore important to be able to use the link itself to determine wether or not the attenuation it is experiencing is the result of rain or some other effect. Markov switching models are compared with other existing algorithms and are concluded to perform well. The experimental set-up here is of intense interest, and I sincerely hope that the measurements will continue to be made over a long period of time, as they have the potential to provide significant insights into the development of fade mitigation techniques and rainfall retrieval techniques.

Specific comments

1. *"Why was a 4 second temporal resolution chosen for logging the attenuation, rather than 1 second or 5 seconds?"*

   The sampling rate has been chosen in order to achieve a high sampling rate while limiting the amount of missing data (caused by occasional long response times of the radio equipement). A sampling resolution of 4 s was found to be a good trade-off between these two considerations.

2. *"Antenna wetting is indeed a problem, and results (in my experience) in a characteristic exponential tail in the attenuation signal at the end of the rain event. Perhaps this characteristic tail could be used to detect antenna wetting and compensate for it?"*

   The authors noticed a similar effect at the end of almost each event. This is certainly important for rain rate estimation and quantification but is beyond the scope of this paper (which only investigates the classification into dry and wet periods). It will be investigated in the future.

3. *"I am very pleased to see the authors intentions to make the experimental data publicly available via a web platform once the experiment is completed. I would strongly encourage them to avoid re-inventing the wheel, and instead submit their data to an appropriate data archive who can then issue the dataset with a formal citation and permanent id (DOI), giving them credit for their efforts in building, maintaining and processing the dataset."*

The authors thank the reviewer for his advice. This possibility will be considered at the end of the experiment. We are currently thinking about an approach using Dapper/DChart, which has been used for ocean, climate and weather data. If the reviewer has experience with public data archives, any suggestions are welcome.

4. *"The selected datasets used in the analysis are quite short, and are limited to spring and summer months. Annual and seasonal variability of rain is very high, so it would have been good to have seen tests of the algorithms done across an entire year of data. This would also allow a more accurate estimation of the amount of time the non-stationary cases represent."*

   Unfortunately, data availability for this winter is not optimal. There were very few rain events and the sky was very cloudy. The disdrometers (which run on solar power) frequently experienced power issues and had to be shut down for longer periods in order to preserve the batteries. The number of periods for which all disdrometers were simultaneously operating is rather small. The winter months were also discarded from the analysis in order to avoid problems related to mixed-phase precipitation and snow. Nevertheless, the authors think that the selected datasets are long enough to be representative of the average behavior of the system, even if they do not cover an entire year. The exact amount of non-stationary cases is not known precisely but is estimated between 10-20 % based on a visual inspection of the data for one year.

5. *"In figures 2 and 3, the path attenuation seems excessively high given the low rain rates and the short link length."*

   The path-integrated attenuation is in the order of 50 dB during dry periods and increases up to 56 dB during rainfall (>10 mmh^{-1}). It is important to remind that the path-integrated attenuation measurement includes the antenna losses, atmospheric attenuation and possible rain-induced attenuation (including wet antenna). See point 6 of our answer to Reviewer 1.

6. *"I would also like to see a further investigation into the unstationary cases, to determine what the variable baseline is correlated with, and whether there is a method for processing the unstationary case data to make it stationary, as this would reduce type I and type II errors. Future work, perhaps."*

   This is a very good comment. We think that the reviewer's suggestions are interesting but unfortunately also beyond the scope of this paper. This will be investigated in future work, where we will use the meteorological data from the airtraffic control system at the nearby airport to analyse the influence of local climatic conditions on the system performance.