Answer to comments of H. Leijnse (Referee #2)

We thank Hidde Leijnse (Referee #2) for his constructive comments on our manuscript and work which enables us to further improve the quality of our manuscript. In this document we repeat the referee's comments (black font) and add our reply to each point (blue font).

General comments

This paper describes a study of rainfall estimation from recorded RSL levels from five commercial microwave links. The emphasis of this paper is on the classification of wet and dry periods, for which the authors propose a new algorithm. The paper is well-written, well-referenced, and shows interesting results. This is the first paper (that I know of) where data loggers were installed by the researchers at the antennas of commercial links, so that the sampling time and the quantization of the logged signal could be controlled. I have some comments, some of which require some additional analyses and explanations. I think that the paper can be published after revisions. Specific comments are given below.

We thank the referee for acknowledging our work. Answers to the specific comments are given below.

Specific comments

1. When discussing the uncertainties in radar rainfall estimation in general at the bottom of p. 742, you should probably include the effects of the vertical profile of reflectivity (VPR), see e.g. Smith (1986, J. Atmos. Oceanic Technol., 3, 129-141), Joss and Pittini (1991, Meteorol. Atmos. Phys., 47, 61-72), and Hazenberg et al. (2011, Water Resour. Res., 47, W02507).

We agree that the effect of VPR has to be mentioned when uncertainties of radar rainfall estimation are discussed. It will be added to the revised manuscript.

2. On lines 15-16 of p. 747, it is stated that the RADOLAN Z - R relation is used. Could you give this relation here?

We use

$Z = 125 * R^{1.4}$

which is the first part of the three part Z-R relation from the RADOLAN project for stratiform rain events. We will include this in the revised manuscript.

And what is the height of the radar measurements above the terrain for the locations of the different links?

The elevation angle of the radar is 0.5°. The two links for which comparison with the radar is possible have one end on a communication tower 1 km east of the radar. The radar antenna is at 1000 m a.s.l., the link antennas at approximately 1020 m a.s.l. with the antennas pointing downwards ~1-2°. That is, close to the radar, link and radar are at the same height. Towards the end of the link the radar beam center is approximately 450 m above the link. The delay between the observation of falling rain by the radar and by the links is thus negligible, e.g. 100 seconds for a terminal velocity of 5 m/s. (We will not include these details in the revised manuscript)

3. In Eq. (5) the spectrum is normalized by the mean spectrum occurring in dry weather. It is not clear how this mean dry spectrum ($P_{meandry}(f)$) is determined. The problem is that you're trying to distinguish between dry and wet periods, and that you somehow use information about which periods are dry in the process. Please elaborate on this, and explain if additional information about wet/dry periods is needed for this method.

The mean dry spectrum ($P_{meandry}(f)$) was determined by visual inspection of the spectrogram before the normalization. We agree that this at first requires other wet/dry information, e.g. from gauges, to gain the experience needed to be able to judge a spectrum by visual inspection. However, the selection of the dry period has to be done only once. Thus, a period of stable weather could be chosen, either by gauge records or by using a reanalysis data set, to calculate the mean dry spectra. This procedure would also be feasible for a large number of links.

If the spectra in dry weather show a 1/f behavior, simply multiplying the spectra by f would also solve this problem.

This would be an elegant solution to overcome the problem of selecting a dry period to calculate the mean dry spectrum ($P_{meandry}(f)$). However, the dry spectra differ for the different links. None of them shows a real 1/f behavior. In the manuscript we only mention that all spectra exhibit a 1/f-like shape (page 751, line 13). That is, a decrease from high amplitudes at low frequencies to low amplitudes at high frequencies, forming a more or less straight line in a loglog-plot (not shown in the manuscript).

4. An important parameter of the wet/dry classification method is the frequency at which the slow and fast signal variations are separated. This parameter (which corresponds to f_{low2} and f_{high1}, which are close together) is not fitted but chosen based on visual inspection of frequency spectra. I think that because this is such an important parameter, it would be a good idea to optimize it in a manner similar to σ.

We agree that the optimization of this parameter would be interesting, in particular regarding comment 6 where the objection is raised, that the link length and orientation could influence the optimal frequency thresholds. First results for short RSL time series however show that the performance of our wet/dry classification is mainly governed by the choice of σ.

5. The same holds for the parameter L. This parameter is currently also chosen based on visual inspection of results, but would also be a good candidate for optimization. If it should be a power of 2, then the parameter log(L)/ log(2) could be optimized.

The problem is, that for the many different parameters (σ, L, f_{low1}, f_{high1}, f_{low2}, f_{high2}), an optimization, which then must be done for all different combinations, is complicated to overlook, not to mention visualizing it.
For the analysis shown in the discussion paper we already tested our wet/dry classification algorithm with different window length and found that a length of 256 points performed best. The sentences at page 755, line 5-8, are just the explanation of this behavior, not the cause for choosing 256 points as window length.
We will again look into this analysis and think of a way to present its results together with the optimization of the separation frequency of the spectrum mentioned in comment 4. But we think that we should only present the dependence on the most crucial parameters.

6. I had expected a discussion on the space-time structure of rainfall (which may very well be influenced by topography) and its relation to link length and orientation and the employed frequency thresholds. The longer the links, the more averaging occurs, and the longer the typical timescales. This could influence the choice of $f_{low1;2}$ and $f_{high1;2}$, which now correspond to timescales of approximately 4 hours (256 minutes), 4 min., 4 min., and 2 min., respectively. I think the authors should devote some attention to this. For example, how well does the assumption hold that all rainfall events have time scales greater than approximately 4 minutes?

As mentioned in the answer to comment 5, we will analyze the dependence of the algorithm's performance for different separation frequencies of the spectrum. We agree that an accompanying discussion of the space-time structure of rain events will fit in well in the light of different link length, i.e. different spatial and temporal scales of averaging.
We however have to note that we already performed an analysis of the dependence of the optimal threshold value σ on link length, orientation, altitude difference, frequency and sensitivity (a combination of frequency and length) without a clear outcome.

7. The order in which things are presented could be improved. For example the example presented in Section 6.3 should probably be presented after Section 6.4, so that the reader does not have to guess on what the value of σ = 2:5 is based. The choice of the window length could also be moved to the bottom of p. 750, where is was introduced.

We think that the example illustrating how the algorithm works should be presented before the optimization of the parameters. But we also see that the choice of the parameters used in the examples causes confusion because it is not explained well or not at all. In the revised manuscript we will explain in more detail why the values have been chosen and we will also refer to the later section where the optimization of the values is explained. (see also answer to comment 9 of referee #1)

8. If I consider Eq. (9), then I would conclude that for any given $P_{sumdiff}$ (t), the higher σ, the higher the number of dry hours and the lower the number of wet hours. In other words, the number of wet hours is a monotonically decreasing function of _ and the number of dry hours is a monotonically increasing function of σ. This also means that $N_{link\&gauge=wet}$ (Eq. (12)) is a monotonically decreasing function of σ and $N_{link\&gauge=dry}$ (Eq. (13)) is a monotonically increasing function of σ. This in turn means that "$_{wet}$ (Eq. (12); wet detection error rate) should be a monotonically increasing function of σ and that "$_{dry}$ (Eq. (13); dry detection error rate) should be a monotonically decreasing function of σ. However, looking at Fig. 5, it can be seen that this is not always the case ("$_{wet}$ sometimes decreases with σ and "$_{dry}$ sometimes increases with σ). The reason for this should be explained clearly in the paper.

The cause for this behavior is that the number of dry or wet hours for the link data is not derived from the wet/dry flag each minute value gets from the classification algorithm. We rather use the mean hourly rain rate for the link data here. If it is zero, the hour is classified as dry. If it is larger than zero, the hour is classified as wet.
For low threshold values, where most of the hours are classified as wet, the baseline stays constant over long periods, because it is only changed during a dry period. That is, it is often too low. Hence negative attenuation and rain rates emerge. These are then set to zero during processing (page 753, line 2). An hour which is classified as wet, because it contains "wet" minutes, can thus yield a rain rate of zero if all minute rain rates within it were negative. In our analysis such an hour is then counted as dry. In particular for threshold values far below the optimal value, the number of dry and wet hours is hence not monotonically changing.

It would be possible to use only the initial wet/dry classification for the error analysis, but we think that, as the negative rain rates have to be excluded anyway, it is better to use the resulting hourly rain rate.
The sentence on page 754, line 12 only mentions the hours which are classified as wet and dry, but does not mention, that this is not based on the initial minutely wet/dry classification. We will clarify this in the revised manuscript and explain the non-monotonically increasing and decreasing course of the error rates.

9. It is shown in Fig. 9 that the comparison is between link and radar is better than that between link and gauge. It would be interesting to see how the values of σ would change if radar data would be used to compute wet and dry errors for those links where radar data are available (hop2-murn1 and hop2-wh0). This may also shed some light on the different behavior of the wet detection error rate of the hop2-murn1 link discussed on lines 15-24 of p. 755.

We will add an analysis of the wet/dry classification performance using radar data as ground truth.

10. Because the focus of the paper is on wet/dry classification, I think more attention should be devoted to the discussion of the behavior of the wet and dry detection error rates, as these reflect the quality if the algorithm. This discussion should also include radar-based error rates (see also my previous comment).

As mentioned in the answer to comment 9, we will analyze the use of radar data as ground truth for the performance analysis. This will extend the discussion of the error rates, in particular regarding their dependence on the spatial representativeness of the used ground truth.