

Interactive comment on “Towards variational retrieval of warm rain from passive microwave observations” by David Ian Duncan et al.

Anonymous Referee #2

Received and published: 1 June 2018

This is an excellent study which helps understand the discrepancy between GPM/GPROF and Cloudsat precipitation retrievals, points a way forward for better retrievals of warm rain using passive microwave observations, and finally introduces some innovative and useful techniques, such as correlated observation errors that account for the unknown droplet size distribution. In places the methods need to be explained more precisely and comprehensively, and there are a few other minor comments, but otherwise this manuscript is close to being ready for publication. Many of the comments focus on the a-priori and its error covariance matrix, details of which are essential for understanding the basic character of the retrieval algorithm, i.e. how closely it attempts to fit observations, how much of the retrieval is determined by the a-priori, and whether or not the retrieval is truly optimal.

C1

Minor comments

1) CloudSat has trouble seeing the lowest km, as acknowledged on page 16, lines 1-2. But the proposed retrieval makes use of CloudSat rain profiles to generate a relationship between RWP and surface rain rate (page 13, lines 6-10). Surely the proposed retrieval is thus also affected by these errors? In any case, it would be good to explain briefly how the CloudSat rain rates are generated and hence what assumptions are being made.

2) On page 11: "Figure 5 translates the simple model containing in situ DSD data into error covariances matrices usable by the retrieval, via Eq. 2". This is not strictly true. Equation 2 gives only the derivation of the diagonal elements of the covariance matrix. Please also define how the off-diagonal elements are computed.

3) On page 11, this sentence is not clear: "To apply these analyses of in situ data as realistically as possible, the errors and DSD assumptions derived from extratropical and tropical sites are dependent on the observed latitude." How is this done? How does this relate to what is shown in figure 5? What is the relationship between "extratropical and tropical sites" and "dependent on observed latitude"? All needs to be explained in a little more detail, or perhaps rephrased to be clearer.

4) On page 11, line 14 (and correspondingly, in Figure 6) please define the criteria for non-convergence used in the non-raining algorithm. More generally the phrases "not converged" and the notation " $\chi^2 < 1$ " crop up throughout the paper and it needs to be clear whether these are one and the same, or whether "not converged" has any further meaning. For example, depending on the solver being used, there can be many other failure modes for a variational retrieval.

5) Page 11, line 27-28 the observation error covariance matrix " S_y for the non-raining retrieval is the same as that given by Duncan and Kummerow (2016)". The reader needs a few sentences to describe how this is constructed. It should not be necessary to have to go to the cited paper.

C2

6) Further, the a-priori error covariance matrix S_a for the non-raining retrieval needs a brief description and quantification (see point 9).

7) The χ^2 statistic of equation 3 is inexact, with a number of consequences. The true Chi squared test for the consistency of an optimal variational retrieval given the specified error covariance matrices (see Rodgers, 2000, equation 12.9, as cited in the manuscript) would replace the observation error covariance matrix S_y in this manuscript's approximate Chi squared test with $S_y(KS_aK^t + S_y)^{-1}S_y$. Here K (Rodgers' notation) would correspond to the Jacobian of the observation operator (the derivatives of $f(x, b)$ in this manuscript) and S_a the error covariance of the a-priori. The issues are, first, that it is confusing to label a test χ^2 when it is not the true test - and a word of explanation could be given - why did the authors choose this less exact formulation? Second, when the a-priori error covariance in observation space is small relative to the observation errors (i.e. $KS_aK^t > S_y$) the manuscript's inexact Chi squared becomes a poor measure of convergence - broadly, the retrieval will not get much closer to the observations, but this is OK because it is consistent with the specified errors. Third, the true Chi squared test could have given a powerful statistical tool for assessing the validity of the assumed error covariance matrices, which would have been very useful for understanding the retrieval methodology.

8) In section 4.3, the source of the a-priori for the warm rain retrieval does not seem to be described. This would be helpful for its own sake, but also to support section 6.2 which describes what happens when the a-priori is changed - but with respect to what?

9) In section 4.3, the a-priori error covariance matrix is not introduced, but it is essential for understanding the character of the retrieval algorithm (essentially what are the relative sizes of KS_aK^t and S_y , i.e. how closely does it attempt to fit the observations). For example, the description, "Even with four variables, the a priori errors on LWP and PC1 of water vapour are decreased" is not precise enough, especially since there is currently no quantification of the non-raining error covariance matrix (see point 6)

C3

10) In table 2, does the first row of data correspond to S_y for the non-raining retrieval? If so it would be helpful to state this explicitly. If not, this would be a good place to illustrate it.

11) Page 14, line 16: "if neither converges" - please be more precise on how this is defined.

Technical comments

1) Figure 5 caption makes use of the notation S_y before it has been introduced.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-96, 2018.

C4