# Responses to Anonymous Referee #2

## Comments 1

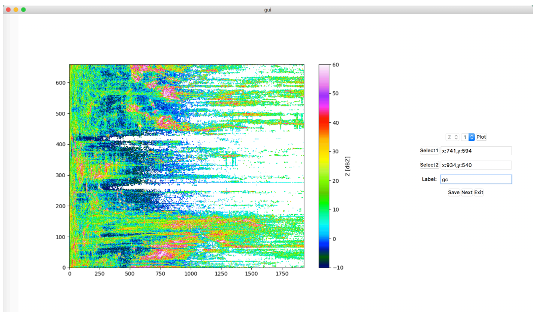
The main weakness of the paper is section 4 "Evaluation". While it is acceptable the that the authors classify, by them self, events to develop and test the algorithm improvements (as done in section 3 based on table 3) the same is not so straight forward for the evaluation step. We assume that in the training MET and NMET echoes are well separated, so it is relatively simple verify that the algorithm outputs meet the expected "expert" classification. For the evaluation phase, we hope that the authors use the improved algorithm in a more complex and mixed situation (like as probably happen in most events where MET and NMET are both present). If this is the case expert classification could be not so easy to do on a pixel basis. Further there is no evidence at all how is the "true" in these events compared and what the outputs of both algorithms. The evaluation results are present in table 5 where performances of both algorithms are shown. Here is not clear if the percentage are respect the "expert true" or what? So this section need to be reformulated and a more deep analysis is needed.

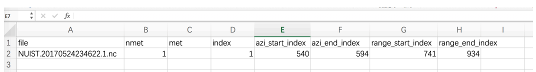
## Response

The doubt proposed by reviewer about the credibility of training/test set is very insightful, which is a very difficult (unresolved) problem in the field of radar meteorology. The details of this problem are as follows.

Generally speaking, the evaluation methods of echo classification are generally divided into two categories: qualitative and quantitative. The qualitative evaluation method generally adopts the way of showing typical cases (Krause 2016), whose disadvantage is that it cannot show the universality of the algorithm, and the cases shown may have the suspicion of luck. The quantitative evaluation method is generally evaluated by the performance of the algorithm on the selected test set in advance (i.e., the expert classification as the reviewer said; Lakshmanan et al. 2014, Tang et al. 2014). In addition, it can also be evaluated by comparing the performance of meteorological application (e.g., quantitative precipitation estimation (QPE)) before and after echo classification (Cho et al. 2006), whose disadvantage is that other error sources will be introduced and also affect the performance of QPE, such as the instability of Z-R relationship and radar miscalibration. Therefore, this paper uses the method of expert classification for evaluation, which is widely used at present.

There are two methods to select training set and test set. The first method is to select cases that only have single echo type in the whole sweep; once there are multiple echo types mixed in a case, it will be discarded (Grecu and Krajewski 2000). This method is relatively simple and convenient, and more suitable for sufficient data. The disadvantage of this method is that it cannot extract the cases with multiple echo types mixed together, which has a high frequency of occurrence as the reviewer said. The second method is to extract data through graphical user interface (GUI; Lakshmanan et al. 2007, Rennie et al. 2015), which is the method adopted in this paper. The GUI program is developed by the author, and its interface is shown in the figure below (left). Mark the region where the data is to be extracted with the mouse, and store the four vertices of the region in the CSV file (as shown in the figure below (right)). Although the software looks “rough”, it's enough to complete my research.



| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | file | nmet | met | index | azi_start_index | azi_end_index | range_start_index | range_end_index | |
| 2 | NUIST.20170524234622.1.nc | 1 | | 1 | 540 | 594 | 741 | 934 | |
| 3 | | | | | | | | | |

At present, it is very difficult to verify the truth of the training/test set. This is because there are not enough observations as evidence, so it can only rely on subjective examine by manual. However, this is not only the problem that this paper faces, but also a problem that all the researches related to radar echo classification face, especially to the hydrometeor classification (Park et al. 2009, Zrnic et al. 2001). The classification between MET and NMET is relatively easier than that of the hydrometeor classification, because it has obvious criterions in some cases, such as data from satellite and rain gauge can prove when is clear air. As proposed by reviewers, it is necessary to evaluate the algorithm in a more complex and mixed situation (e.g., use the GUI program mentioned above to extract the core of the convective storm embedded in the clear-air echoes). However, it is very difficult and unrealistic to clearly distinguish between MET and NMET at the boundary between them by manual classification (other means is even more impossible), which will inevitably lead to subjective bias. Therefore, the author tries to avoid selecting too “complex” data (e.g., the boundary between MET and NMET) to ensure that no subjective bias is introduced, which is also the default view of all related researches (Lakshmanan et al. 2014, Tang et al. 2014). Therefore, the percentages shown in Table 5 are indeed "expert truth". But this "expert truth" is obtained from the evaluation results of the “not very complex” (without dispute and subjective bias) test set.

The author complements and improves the paper according to the reviewer's suggestion, especially in the Section 4, including how to select training/test set (same as the second and fourth comments; corresponding to Lines 102-106 and 235-236 in the revised manuscript), adding an overall skill score method (HSS; corresponding to Lines 234-239 in the revised manuscript), and analyzing the sensitivity of the improvement steps mentioned in Section 3.2 (corresponding to Lines 256-274 in the revised manuscript).

## Comments 2

Section 3 line 104. Please indicate how events in table 4 are been classified.

## Response

Reviewer's comment is adopted and how to select training set has been supplemented in the revised manuscript (corresponding to Lines 102-106).

## Comments 3

Section 3.2.4 line 187 Please discuss how sensible is your method respect to time-schedule used. Since you use the "previous volume scan" at the "same location" to prevent misclassification of ML what happen if you have a quite fast-moving system?

## Response

The part questioned by the reviewer (i.e., use the "previous volume scan" at the "same location" to prevent misclassification of ML) appears in Section 3.1 for the first time (corresponding to Lines 122-124). The role of Section 3.1 in this paper is to describe the overall framework of MetSignal algorithm, and the post-processing rule (use the "previous volume scan" at the "same location" to prevent misclassification of ML is one of the three post-processing rules) mentioned in Section 3.1 was not proposed by authors but an integral part of the raw MetSignal algorithm (Krause 2016). Just like the question raised by the reviewer, I'm also confused about it. Fortunately, however, this part has been removed in the improved algorithm (i.e., MetSignal_noise) due to the potential risk of misclassifying AP into MET (corresponding to Lines 204–212 for details).

## Comments 4

section 4 line 215 Please explain how are classified the events in table 4.

## Response

Reviewer's comment is adopted and how to select test set has been supplemented in the revised manuscript (corresponding to Lines 235-236).

## Comments 5

Section "conclusion" line 232 Correct the section number.

## Response

Reviewer's comment is adopted, and the section number has been corrected in the revised manuscript (corresponding to Line 275).

## Comments 6

references lines 265-266 Please complete the reference.

## Response

Reviewer's comment is adopted, and the reference has been completed in the revised manuscript (corresponding to Lines 312-313).

## Comments 7

references lines 282-283 Please add the journal.

## Response

Reviewer's comment is adopted, and the missing journal name has been added in the revised manuscript (corresponding to Lines 329-330).

## Comments 8

figure 5 In Panel (a) what means "aggregation value"? Please explain the color scale?

## Response

The "aggregation value" questioned by the reviewer appears in Section 3.1 for the first time (corresponding to Lines 92-98), which is a terminology in fuzzy logic technology (Zrnic et al. 2001). The formula for the calculation of aggregate value for the MET ($A_{MET}$) is shown in Eq.1 (the aggregation value for the NMET can be computed by subtracting $A_{MET}$ from 1). The $A_{MET}$ of a specific pixel indicates the possibility that this pixel is MET, in other words, 1 in the color scale represents 100% (very likely to be MET), 0 represents 0% (next to impossible to be MET). After getting the $A_{MET}$ by the calculation of Eq.1, compare it with a preset threshold. The target pixel will be classified as MET if $A_{MET}$ exceeds the threshold; otherwise, it will be classified as NMET (corresponding to Lines 113–118 for details).

## Comments 9

figure 8 As for point 7 related to panels (a) and (b).

## Response

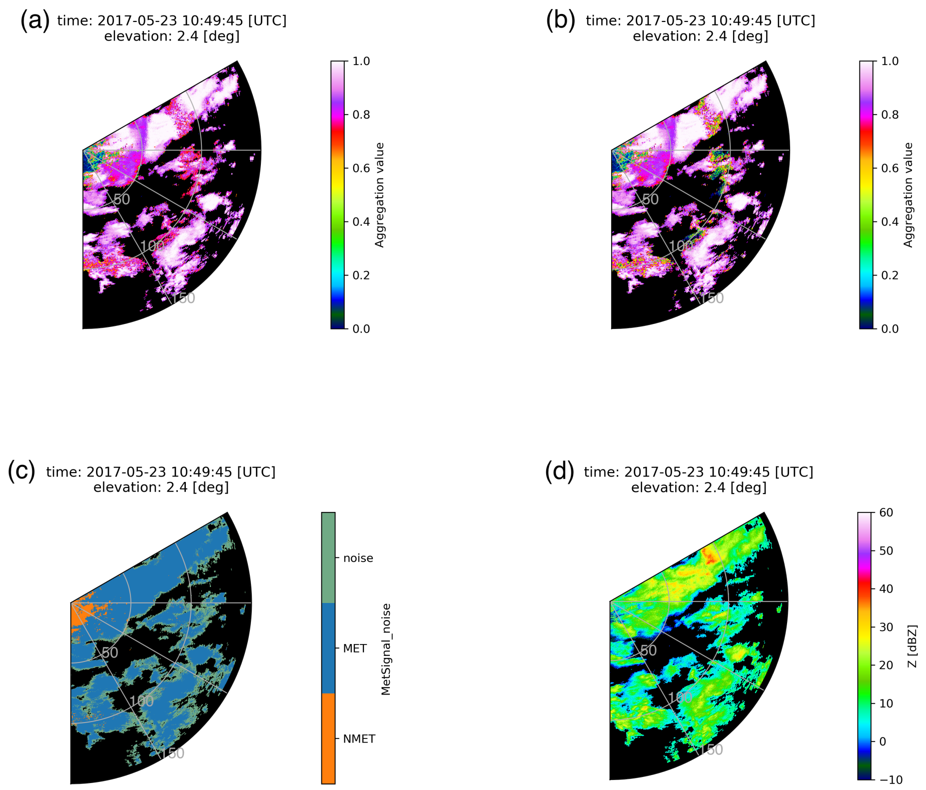
As for response to comment 8.

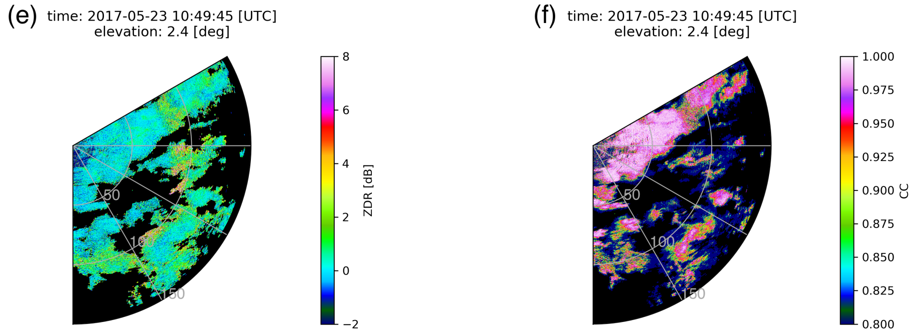
# Comments 10

figure 11 As above Further, in order to increase the readability of this figure could you evaluate to display a zoomed area around the ML at 100 km range from radar in the SE quadrant?

# Response

As for response to comment 8.
As shown in the figure below, Fig. 11 has been modified according to the reviewer's comments. However, revised Fig. 11 may raise some readers' doubts about what happened in other quadrants (whether the authors want to hide something in the rest of this case). Therefore, the author suggests that the reviewer reconsider whether to make this modification. Absolutely, if the reviewer still insists on making this modification, I'll make some adjustments.






## Reference

Cho, Y. H., Lee, G., Kim, K. E., and Zawadzki, I.: Identification and Removal of Ground Echoes and Anomalous Propagation Using the Characteristics of Radar Echoes, Journal of Atmospheric & Oceanic Technology, 23, 1206-1222, 2006.

Grecu, M. and Krajewski, W.: An Efficient Methodology for Detection of Anomalous Propagation Echoes in Radar Reflectivity Data Using Neural Networks, Journal of Atmospheric & Oceanic Technology, 17, 121-129, 2000.

Lakshmanan, V., Fritz, A., Smith, T., Hondl, K., and Stumpf, G.: An Automated Technique to Quality Control Radar Reflectivity Data, Journal of Applied Meteorology & Climatology, 46, 288-305, 2007.

Lakshmanan, V., Karstens, C., Krause, J., and Tang, L.: Quality Control of Weather Radar Data Using Polarimetric Variables, Journal of Atmospheric & Oceanic Technology, 31, 1234-1249, 2014.

Park, H., Ryzhkov, A. V., Zrnic, D. S., and Kim, K. E.: The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS, Weather & Forecasting, 24, 730-748, 2009.

Rennie, S. J., Curtis, M., Peter J., Seed A. W., Steinle P. J., and WEN G.: Bayesian Echo Classification for Australian Single-Polarization Weather Radar with Application to Assimilation of Radial Velocity Observations, Journal of Atmospheric & Oceanic Technology, 32, 1341-1355, 2015.

Tang, L., Zhang, J., Langston, C., Krause, J., Howard, K., and Lakshmanan, V.: A Physically Based Precipitation-Nonprecipitation Radar Echo Classifier Using Polarimetric and Environmental Data in a Real-Time National System, Weather & Forecasting, 29, 1106-1119, 2014.
Krause, J. M.: A Simple Algorithm to Discriminate between Meteorological and Nonmeteorological Radar Echoes, Journal of Atmospheric & Oceanic Technology, 33, 1875-1885, 2016.
Zrnic, D. S., Ryzhkov, A., Straka, J., Liu, Y., and Vivekanandan, J.: Testing a Procedure for Automatic Classification of Hydrometeor Types, Journal of Atmospheric & Oceanic Technology, 18, 892-913, 2001.

Thanks again for the helpful comments.