Response to Reviewer 1

**Question 1:** Lack of Rigorous Validation Metrics

The model evaluation focuses on overall accuracy and a small overlap percentage in histograms. However, these metrics are insufficient for a classification task with imbalanced classes (e.g., 400 vs. 2,500 gates in the test set). The manuscript should report precision, recall, and F1-score, especially for ship clutter, as false negatives can lead to significant data quality issues, and false positives can unnecessarily degrade precipitation data.

*Agree. In the revised manuscript, the additional evaluation metrics suggested by the reviewer have been included in Section 4.2* (Line 360; highlighted content)*. We have also incorporated these metrics into the abstract and summary have also been revised* (Line 19 and 450; highlighted content).

**Question 2:** Limited Generalization and Dataset Diversity

The random forest model is trained and tested on data derived from the same radar (Kumpula), location (Gulf of Finland), waveform (LFM), and limited events. There is no evidence that the algorithm generalizes to other waveform types (e.g., NLFM), elevation angles, or environmental conditions (e.g., high sea clutter, near-shore echoes, different clutter types). The authors must either test the model on independent cases or clearly state the generalization limitations.

Agree. This is a pilot study on identifying ship signatures in polarimetric observations, and we may not overstate its potential general applications. Although we believe that the features and the algorithmic logic proposed for identifying ship clutter are generally applicable, the manuscript lacks an evaluation of the algorithm’s robustness across multiple scenarios (e.g., different radars, sea areas, waveforms, etc.). To address this issue, we have added the discussion of its limitations in Section 5 (Line 425; highlighted content) and more in-depth analysis should be made in future.

**Question 3:** Overreliance on Manual Labeling

Both the ship clutter and precipitation echo datasets are manually labeled, and the methodology for doing so is not sufficiently described. This introduces potential biases. What criteria were used to define clutter? Were multiple annotators involved? Was any inter-annotator agreement measured? These questions should be addressed or acknowledged as limitations.

We feel sorry for the unclear description in the previous version. The datasets were labelled by the first author, and the validity of ship clutter identification is assured by checking its movement continuity in consecutive radar scans. We acknowledge that an objective approach could be more comparable to future studies, but no such methods are available. In the revised manuscript, we have supplemented the relevant explanations and figure in Section 2 (Line 112 and 134; highlighted content).

**Question 4:** Sidelobe Suppression Logic May Be Overly Aggressive

The PSD definition and filtering logic—especially the combination of velocity and SNR thresholds—may lead to over-removal of precipitation echoes, particularly in overlap regions. While some case studies suggest selective filtering is achieved, the possibility of precipitation loss is real and must be quantified. For example, is there a statistical estimate of how many precipitation gates were removed in mixed scenes?

According to the current ship clutter identification and removal method presented in this paper, when precipitation overlaps with ship clutter, the precipitation information within the same range gate will be lost if the ship clutter is successfully identified (as mentioned in the Discussion, this issue needs to be addressed through signal-level processing). The HSCI algorithm employs velocity and SNR filters with the aim of minimizing the loss of precipitation information. However, the authors acknowledge that there is still a risk of over-suppression even with these measures.

Since the algorithm proposed in this paper is based on data-level processing, a truly quantitative evaluation of precipitation loss would require signal-level techniques—specifically, the construction of ground truth fields through either full signal simulation (Curtis 2025) or semi-simulation (Li et al., 2013). Although we have already acquired the relevant technical capabilities, discussing these in the current paper would risk either deviating from the main topic (due to excessive description of signal processing techniques) or lacking sufficient detail. Therefore, we are preparing a manuscript, focusing on ship clutter suppression from the signal processing perspective to address this evaluation.

While a quantitative assessment of precipitation loss is not provided in this manuscript, Figure 13b offers a qualitative perspective: after suppressing the impact of ship clutter, the accumulated precipitation field exhibits a smooth and natural distribution. Moreover, no abnormally low values are observed in the main navigation channel area (Figure 13d), indirectly suggesting that the HSCI algorithm does not cause detectable damage to precipitation echoes. Related analysis on precipitation loss has been added to Sections 4.2 of the revised manuscript (Line 394; highlighted content).

Li Y, Zhang G, Doviak R J, et al. A new approach to detect ground clutter mixed with weather signals[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 51(4): 2373-2387.

Curtis C D. Weather Radar Time Series Simulation: Rapidly Looping through Signal Parameters[J]. Journal of Atmospheric and Oceanic Technology, 2025, 42(3): 269-279.

**Question 5:** No Independent Test Dataset or Cross-validation

The authors should demonstrate model robustness through k-fold cross-validation or by holding out an entire day or event as an independent test case. Without this, it is difficult to assess whether the model is overfitting or merely capturing spatiotemporal autocorrelation patterns.

The author sincerely apologizes for not explaining this issue clearly in the original manuscript. In Section 3.1.3, we mention the use of the GridSearchCV method from Scikit-learn for optimizing model parameters. This method combines exhaustive search and cross-validation to determine the optimal parameters. The cross-validation referred to here is exactly the k-fold cross-validation suggested by the reviewer (with the default value of k being 5). Additional details have been added to Section 3.1.3 in the revised manuscript (Line 237; highlighted content).

**Question 6:** Lack of Code or Reproducibility Path

For a method combining machine learning and empirical filtering, reproducibility is essential. At a minimum, a flowchart covering the entire algorithmic sequence, and pseudocode or a link to a repository, should be provided. Currently, implementation details are scattered and would be difficult for others to reproduce.

Good suggestion. In our resubmission, we have included the relevant code and data. By combining the code with the flowchart shown in Fig. 5, we believe that readers will find it easier to understand and master the algorithm.

**Question 7:** Feature Descriptions (Sect. 3.1.2): Clarify if all features are used as-is or normalized/scaled. Include distributions (box plots or ranges) for ship vs. precip echoes for added transparency.

Agree. We have supplemented the relevant explanations in Section 3.1.2 (Line 207; highlighted content).

**Question 8:** Figures 6 and 9: Use larger labels and more distinct color schemes. Consider adding contour lines for better interpretability.

Agree. The revised manuscript has optimized the figures (Fig. 7 and 10 now) as requested by the reviewer.

**Question 9:** Discussion Section: Expand on the feasibility of signal-level suppression in commercial processors (e.g., how would this be integrated in RVP900 or similar systems?). A brief note on real-time considerations would also help.

Agree. We have supplemented the relevant explanations in Section 5 (Line 405; highlighted content).

**Question 10:** Line 134: “efficiency is exceedingly low” → consider “computational efficiency is low”

Agree.

**Question 11:** Line 373: Remove extra parenthesis in “Palmer et al., 2023))”.

Agree.