

## 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.

## Response to Reviewer 2

**Question 1:** Precipitation Dataset Clarification (Line 110): The manuscript mentions that the precipitation dataset was manually extracted. However, it is unclear why this was necessary, especially since the algorithm is intended to work under both clear-sky and rainy conditions. If the algorithm is designed to perform differently depending on weather conditions, this distinction should be clearly stated and justified in the paper.

Weather radars are primarily used to observe precipitation echoes. However, in addition to precipitation, they also detect non-meteorological echoes such as ground clutter, biological echoes, and ship clutter, which is the focus of this study. These non-meteorological echoes can interfere with the observation and application of precipitation data, and therefore often need to be removed. As mentioned in Introduction, ground clutter and biological echoes are common types of non-meteorological echoes, and mature techniques have already been developed to suppress them effectively. In contrast, ship clutter is a less studied type of non-meteorological echo, and suppression methods for it have not yet been established—this is the primary research focus of our study.

Given that ground clutter and biological echoes can be effectively suppressed, we assume they have minimal or negligible impact on our research. Therefore, the classification task in this study is simplified to a binary classification problem between ship clutter and precipitation echoes, which is a common practice in the weather radar field (Kilambi et al., 2018; Hubbert et al., 2009).

When applying machine learning to solve a binary classification problem, it is necessary to identify features that can distinguish between the two classes and train a model to learn the feature distribution of each. This process requires a large amount of labeled data (i.e., with known ground truth), which explains why manual extraction of precipitation data is needed. In short, the reason for manually extracting precipitation data is not because the algorithm depends on specific weather conditions, but because the problem is formulated as a binary classification task, which requires a dataset for model training. Further clarification on this point has been added in Introduction of the manuscript (Line 83; highlighted content).

Hubbert J C, Dixon M, Ellis S M. Weather radar ground clutter. Part II: Real-time identification and filtering[J]. *Journal of Atmospheric and Oceanic Technology*, 2009, 26(7): 1181-1197.

Kilambi A, Fabry F, Meunier V. A simple and effective method for separating meteorological from nonmeteorological targets using dual-polarization data[J]. *Journal of Atmospheric and Oceanic Technology*, 2018, 35(7): 1415-1424.

**Question 2:** Threshold Justification (Line 136): The choice of a 20 dBZ threshold is not explained. Where does this value come from? Additionally, does ship clutter with lower reflectivity (e.g., 15 dBZ) not affect data quality? This requires clarification.

For pulse compression radars, in addition to the range gate of the ship target itself, ship clutter may also be observed in surrounding areas due to range sidelobes and antenna sidelobes. However, due to the inherent sensitivity limitations of radars, weaker sidelobe returns from ship clutter are often undetectable. Taking the high-sensitivity Kumpula radar as an example, the minimum detectable reflectivity at a range of 50 km is approximately -20 dBZ. As shown in the statistical results in Fig. 7, when a ship clutter with a mainlobe reflectivity of 20 dBZ is observed at 50 km, detectable sidelobe echoes are limited to a few range gates near the mainlobe (i.e., those exceeding -40 dB in Fig. 7). Namely, no significant sidelobes are observed by the radar if a lower threshold is used. Therefore, the selection of 20 dBZ is a compromise between radar sensitivity and the sidelobe levels. If better sensitivity can be achieved, a lower threshold can be made.

A more intuitive example is the ship clutter labeled as "2" in Fig. 9a. In such cases, when the ship clutter appears in isolation (i.e., not overlapping with precipitation echoes), it can be effectively suppressed using methods commonly applied to mitigate point clutter (Vaisala 2016). When the ship clutter overlaps with precipitation, the impact on the observation data is usually minimal. Therefore, by applying this threshold, the efficiency of the algorithm can be improved without affecting its performance. Additional explanation regarding this has been included in Section 3.1.1 of the manuscript (Line 154; highlighted content).

Vaisala: User's manual: RVP900 digital receiver and signal processor; Vaisala oyj: Vantaa, Finland, 513, 2016.

**Question 3:** Manual Selection of Sidelobe Region (Line 244): The sidelobe region is manually defined within a 13.5 km radial range and 15° in the tangential direction. The origin and justification for these values are not provided. Were these same values used later in operational scenarios? Furthermore, what automated method replaces the manual process in operational applications?

The radial range of 13.5 km (90 range gates) was determined based on the pulse width (90  $\mu$ s) of the pulse compression waveform. In other words, the maximum range affected by the range sidelobes of ship clutter is 13.5 km. The tangential direction was limited to 15 degrees, based on the antenna pattern. As the angle deviates from the mainlobe, the power of sidelobe echoes significantly decreases. As shown in Fig. 7, when the deviation from the mainlobe reaches 15 degrees, the sidelobe level is 80 dB lower than the mainlobe. Therefore, data beyond this 15-degree range can be considered below the radar's sensitivity and thus undetectable. Additional explanation regarding this has been included in Section 3.2.1 of the manuscript (Line 270; highlighted content).

Regarding the reviewer's question about the manual data selection, this was performed during the algorithm design phase. The purpose was to analyze the characteristics of ship clutter under a specific radar (the Kumpula radar in this case) and a specific transmitted waveform (the 90  $\mu$ s LFM waveform used by the Kumpula radar). In the subsequent application phase, manual data selection is no longer involved; instead, the statistical results obtained during the design phase are used. It is worth noting that the characteristics obtained in this study are limited to the Kumpula radar under the LFM waveform. When applying the algorithm to other radars, a new statistical analysis must be conducted.

**Question 4:** Case Study Placement (Section 3): Two case studies are introduced in the 'method' section. However, they seem more appropriate for Section 4, where other case analyses are presented. Additionally, the purpose of second case (Fig. 8) is unclear—what characteristics is it meant to highlight, and why was it chosen?

This concern is valid. We did give thinking on this point in previous submission. However, considering that the readers of this paper may not all be experts in the field, relying solely on textual descriptions might not be reader-friendly. Therefore, following the writing style of Tang et al. (2014, 2020), we adopted a visual approach to present the algorithm's principles and design logic, which we believe will better aid readers in understanding the proposed method.

Secondly, the reason of presenting the second case is to highlight the effectiveness of the velocity and SNR filters in preventing excessive suppression of precipitation information. Specifically, under clear-sky conditions (i.e., the first case), since there is no precipitation echo, there is no significant difference whether the velocity and SNR filters are applied (Figs. 8b and 8d); in contrast, when ship clutter and precipitation echoes are mixed (i.e., the second case), using the velocity and SNR filters can greatly reduce the algorithm's loss of precipitation echoes (Figs. 9b and 9d).

Tang, L., J. Zhang, C. Langston, J. Krause, K. Howard, and V. Lakshmanan, 2014: A Physically Based Precipitation–Nonprecipitation Radar Echo Classifier Using Polarimetric and Environmental Data in a Real-Time National System. *Wea. Forecasting*, 29, 1106–1119

Tang, L., J. Zhang, M. Simpson, A. Arthur, H. Grams, Y. Wang, and C. Langston, 2020: Updates on the Radar Data Quality Control in the MRMS Quantitative Precipitation Estimation System. *J. Atmos. Oceanic Technol.*, 37, 1521–1537

**Question 5:** Discussion Section Issues: The final paragraph of the discussion section appears to be more suitable for the introduction or summary. Once this paragraph is removed, the discussion becomes quite short. It would benefit from more in-depth coverage of how the algorithm operates in operational case—especially regarding the parts that were handled manually (e.g., sidelobe removal). Does the algorithm only remove sidelobes near the mainlobe, or is there a more general solution? These points should be explicitly addressed.

Agree. In the revised manuscript, the final paragraph of the Discussion section has been moved to the Summary, and the Discussion has been further improved (Line 456).

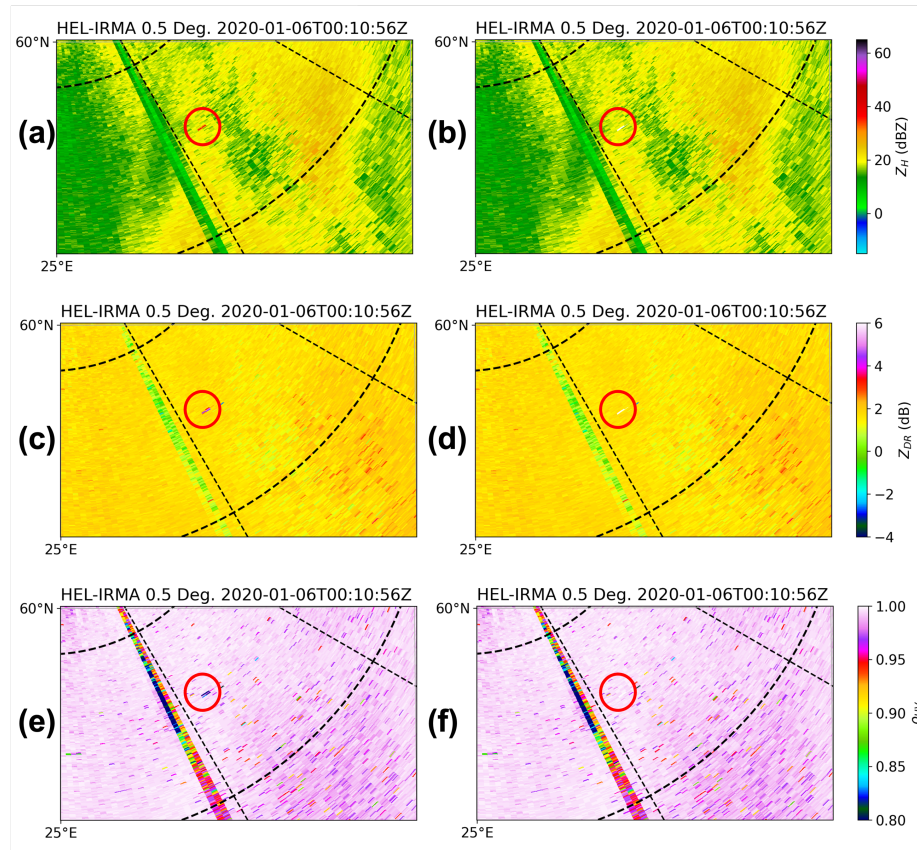
However, we would like to further clarify to the reviewer that, as described in Section 3.2.1 of the manuscript, the determination of the PSD is an adaptive process. Specifically, for strong ship clutter, the PSD closely matches the statistical results shown in Fig. 7. For weak ship clutter, however, the PSD is constrained by the radar's sensitivity — most sidelobes cannot be detected by the radar, resulting in a significantly smaller range limited to a few range gates adjacent to the mainlobe.

**Question 6:** Operational Performance for Winter Events: Since the algorithm is intended for operational use, it is strongly recommended to include at least one winter precipitation case to demonstrate its robustness in varying weather conditions.

This is a good suggestion. The occurrence of super-refraction which is conducive to ship clutter signatures has seasonal dependence (Lopez, 2009). Usually, ship clutter signatures are more common in summer than in winter. We have looked through our observations in the winter of 2020, but identified very limited ship clutter cases in precipitation. We found a single case where ship clutter was mixed with

precipitation, with just one ship clutter present across the entire PPI. Fortunately, the HSCI algorithm successfully detected and filtered this ship clutter without excessively suppressing the surrounding precipitation echoes. Further clarification on this point has been added in DATA section of the manuscript (Line 112; highlighted content).

Lopez, P. (2009). A 5-yr 40-km-resolution global climatology of superrefraction for ground-based weather radars. *Journal of applied meteorology and climatology*, 48(1), 89-110.



**Figure 1: 0.5° elevation of Kumpula radar using LFM waveform at 0010 UTC 6 Jan 2020. (a) Reflectivity before ship clutter filtering; (b) Reflectivity after ship clutter filtering; (c) Differential reflectivity before ship clutter filtering; (d) Differential reflectivity after ship clutter filtering; (e) Correlation coefficient before ship clutter filtering; (f) Correlation coefficient after ship clutter filtering.**

**Question 7:** Figure 3 is explained in the introduction but appears on page 4. Consider repositioning it closer to the relevant text.

Agree. The revised manuscript has been updated according to the reviewer's suggestion to move Figure 3 forward.

**Question 8:** In several instances, figures are referenced before they are introduced or explained (e.g., Line 321: "This selective filtering approach is demonstrated in Figs. 11b and 12d..."). However, Figure 12 is not described until Line 345. This disrupts the flow and may confuse readers. Please revise for consistency.

We mistakenly wrote 11d as 12d, which would indeed mislead readers. We are very grateful to the reviewer for their careful review and for pointing out this error on our behalf.

**Question 9:** Fig 12c: it is not clear if the plot shows the difference 'before-after' filtering or vice versa

Agree. In the revised manuscript, we have supplemented the relevant explanations in Section 4.2 (Line 389; highlighted content).