

A hybrid algorithm for ship clutter identification in pulse compression polarimetric radar observations

Shuai Zhang¹, Haoran Li^{2,3}, Dmitri Moisseev^{3,4}

¹Radar Meteorological Centre, China Meteorological Administration, Beijing, China

² State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China

³Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, Helsinki, Finland 4 Finnish Meteorological Institute, Helsinki, Finland

Correspondence to: Haoran Li (lihr@cma.gov.cn) and Dmitri Moisseev (dmitri.moisseev@helsinki.fi)

Abstract. With the rapid development of active-phased arrays and solid-state transmitters, pulse compression technology has

- 10 become increasingly important. Currently, pulse compression waveforms with peak sidelobe levels better than -50 dB have been developed, enabling the broader application of pulse compression technology in weather radar systems. However, existing sidelobe suppression levels are still insufficient to ensure that radar data quality is unaffected by range sidelobes for ship clutter, which have a high echo intensity and cannot be removed by conventional quality control methods. In this study, we introduce a Hybrid Ship Clutter Identification (HSCI) algorithm to address this issue in pulse compression polarimetric
- 15 radar observations. The HSCI algorithm comprises two parts: mainlobe and sidelobe identification (including the range and antenna sidelobes). Mainlobe identification uses a random forest model that integrates multiple features to identify the mainlobe of ship clutter. Sidelobe identification uses a series of heuristic criteria derived from the statistical characteristics of ship clutter to distinguish them from precipitation echoes. The analysis results of two typical cases indicate that after implementing the HSCI algorithm, the impact of ship clutter on radar data is visually imperceptible. The statistical results
- 20 show that the HSCI algorithm achieves a ship clutter mainlobe identification rate of 97.25% with a misidentification rate of only 0.08% in the precipitation data. Application of this algorithm to the University of Helsinki C-band dual-polarization Doppler weather radar data successfully reproduced ship tracks in the Gulf of Finland.

1 Introduction

As a sophisticated observation instrument, weather radar has significantly advanced research in disaster weather warning 25 (Sandmæl et al., 2023; Chen et al., 2024), precipitation microphysics (Ho et al., 2023; Li et al., 2024), and quantitative precipitation estimation (Li et al., 2023; Hanft et al., 2023). Evolving demands for meteorological applications continue to drive improvements in radar performance and data quality, thereby stimulating the development of innovative radar technologies. Pulse compression, which modulates radar signals to increase bandwidth and thus achieve a better range resolution, is a typical example (Cook and Bernfeld, 1967; Rihaczek, 1969). However, the pulse compression technique was

30 initially not widely adopted in the meteorological field because its associated range of sidelobes could obscure weak targets near strong ones.

The recent growing popularity of solid-state transmitters and phased array radars (Weber et al., 2021; Palmer et al., 2022; Kollias et al., 2022), has seen rapid advances in pulse compression technology. Bharadwaj and Chandrasekar (2012) proposed a combination of a continuous nonlinear frequency modulation (NLFM) waveform with a minimum integrated

- 35 sidelobe level filter, and its performance was validated for reflectivity steps up to 40 dB through simulation experiments. Kurdzo et al. (2014) introduced an NLFM waveform designed using a genetic algorithm. Specifically, this approach optimized the frequency function represented by a Bezier curve to generate an NLFM waveform with low sidelobes and high range resolution. Compared with traditional windowed methods, this technique offered a sensitivity gain of approximately 3 dB. Torres et al. (2017) also used a genetic algorithm to design an NLFM waveform tailored to operational requirements,
- 40 focusing on minimizing the transmission bandwidth as the primary optimization goal. Other optimization techniques used in this field include simulated annealing (Pang et al., 2015) and quadratic optimization (Argenti and Facheris, 2020). Owing to these advanced technologies, range sidelobes have been effectively suppressed, enabling the broader adoption of pulse compression technology in weather radar systems.

Like conventional short-pulse radar, pulse compression radar data can also be contaminated by non-meteorological echoes.

- 45 Therefore, quality control is crucial for the effective use of radar data. Ground clutter, a common type of non-meteorological echo, is caused by scattering from stationary targets, such as buildings or mountains (Billingsley, 2002). Typically, ground clutter exhibits a near-zero Doppler velocity and narrow Doppler spectrum width (Hubbert et al., 2009a). Numerous ground clutter identification and filtering algorithms have been developed based on this characteristic (Hubbert et al., 2009b; Torres and Warde, 2014; Golbon-Haghighi et al., 2018; Hubbert et al., 2021), and have achieved substantial success (Fig. 1).
- 50 Biological echoes, caused by scattering from airborne biological entities, such as insects and birds, represent another frequent source of non-meteorological echoes (Stepanian et al., 2016). Given that the shape, orientation, and other attributes of biological targets differ significantly from precipitation particles, simple metrics such as correlation coefficients or depolarization ratio thresholds can effectively distinguish between these two types of targets (as demonstrated in Fig. 2; Kilambi et al., 2018; Pérez Hortal and Michelson, 2023).

55

Figure 1: 0.5° elevation of Kumpula radar using linear frequency-modulation (LFM) waveform at 1800 UTC 3 May 2020. The Gaussian model adaptive processing (GMAP) algorithm built into the RVP900 signal processor was adopted to achieve this performance. (a) Raw reflectivity; (b) Reflectivity after ground clutter filtering.

60

Figure 2: 0.5° elevation of Kumpula radar using LFM waveform at 0845 UTC 4 June 2020. Precipitation echoes are concentrated within azimuthal intervals of approximately 60–180° with high correlation coefficient, while other sectors are mainly affected by biological echoes with low correlation coefficient. (a) Filtered reflectivity; (b) Correlation coefficient.

65

In addition to the commonly-observed ground clutter and biological echoes, weather radars deployed along coastlines usually detect echoes scattered from ships, referred to as ship clutter, in meteorological contexts (Overeem et al., 2020). Unlike other forms of clutter, ship clutter exhibits non-zero Doppler velocities and high correlation coefficients, making it challenging for existing quality control methods to suppress it effectively (as depicted in Fig. 3). Typically, a ship spans one

70 or more range gates. However, in pulse compression radar systems, the impact of ship clutter is not confined to these gates but also extends radially and tangentially because of the range and antenna sidelobes. As illustrated in Fig. 3, this results in numerous cross-shaped patterns that can extend over ten kilometers on the plan position indicator (PPI), significantly compromising the quality of the radar data.

75

Figure 3: 0.5° elevation of Kumpula radar using LFM waveform at 1210 UTC 5 May 2020. Several ships present cross-shaped radar variable fields, as well as non-zero Doppler velocity and high correlation coefficients. (a) Filtered reflectivity; (b) Doppler velocity; (c) Correlation coefficient.

80 In this study, we propose a Hybrid Ship Clutter Identification (HSCI) algorithm to enhance the data quality and meteorological application performance of pulse compression radars. The instruments and related datasets used in this study are described in Sect. 2. Section 3 provides an in-depth description of the HSCI algorithm, while Sect. 4 presents the algorithm performance evaluation results. Discussion and summary are presented in Sects. 5 and 6, respectively.

2 Instrument and data

- 85 At the Kumpula campus of the University of Helsinki, a C-band dual-polarization Doppler weather radar (hereinafter referred to as the Kumpula radar) was installed on the rooftop of the Department of Physics building (60.204°N, 24.269°E, 60 m above mean sea level). In 2019, the klystron transmitter of the Kumpula radar was upgraded to two solid-state transmitters supported by Vaisala Oyj. The radar currently serves as a prototype for evaluating the performance of solid-state transmitters and pulse compression technology. Observational data collected between May and June 2020 were used in this
- 90 study.

The archived data from the Kumpula radar include the reflectivity factor at horizontal polarization (Z_H) , Doppler velocity (v_r) , Doppler spectrum width (σ_v) , differential reflectivity (Z_{DR}) , differential phase (ϕ_{DP}) , and co-polar correlation coefficient (ρ_{HV}) . The scanning strategy used by the Kumpula radar diverges from that used in operational radars, such as the volume coverage pattern 21 used by the Weather Surveillance Radar-1988 Doppler (Crum and Alberty, 1993). Specifically,

- 95 the Kumpula radar conducts three PPI scans at an elevation of 0.5°, using diverse transmitting waveforms: 1) unmodulated short pulse (SP), LFM, and NLFM. For the LFM and NLFM waveforms, a frequency diversity technique was applied to address the blind-zone issue (Bharadwaj and Chandrasekar, 2012). This involves transmitting an additional unmodulated short pulse (ASP) at a slightly shifted frequency to cover the blind zones created by the modulated pulses. The detailed system characteristics of the Kumpula radar and settings for different waveforms are listed in Table 1.
- 100

Table 1: System characteristics and waveform settings of the Kumpula radar.

The Kumpula radar, situated on the north coast of the Gulf of Finland, which is a crucial waterway in Northern Europe, frequently detects ship clutter at low elevation angles. We compiled a radar dataset for ship clutter by manually identifying

- 105 the distinct strong point echoes and cross-shaped signatures. When these static signatures were insufficient to confirm the presence of ship clutter, the movement of echoes across consecutive scans was used as supplementary evidence. Given that ships typically take several hours to traverse the effective field of view of the radar, we extracted only one scan per hour to maintain a high level of diversity and independence among the different instances of ship clutter. Ultimately, this method yielded a dataset comprising nearly 1600 ship clutter events across 110 scans.
- 110 The precipitation dataset was manually extracted from four precipitation events that occurred on May 10, May 15, June 4, and June 5, 2020. Despite the Kumpula radar using pulse compression technology, which inherently produces range sidelobes, its peak sidelobe level was maintained below -50 dB (as shown in Sect. 3.2.1 below). Furthermore, the range sidelobes from strong precipitation echoes are typically overshadowed by the surrounding medium-intensity precipitation echoes. Consequently, for most precipitation echoes, the impact of range sidelobes was not significant. It is important to note
- 115 that both datasets—ship clutter and precipitation—were derived from observational results obtained using the LFM waveform.

3 Method

Figure 4 shows a flowchart outlining the HSCI algorithm, which uses a straightforward sequential structure. The process begins with the identification of the mainlobe of the ship clutter in the radar data. If the identification results indicate the

120 presence of ship clutter, the procedure continues with the identification of ship clutter sidelobes, followed by the removal of the entire ship clutter (i.e., both the mainlobe and sidelobes). The range gate exhibiting the highest reflectivity within the ship clutter is regarded as the mainlobe, whereas the remaining range gates are considered sidelobes.

125 **Figure 4: Flowchart of the HSCI algorithm.**

3.1 Mainlobe identification

3.1.1 Region limitation

A single scan from weather radar typically yields hundreds of thousands of range gate observations. If the algorithm 130 processes each gate individually, its efficiency is exceedingly low. Moreover, because precipitation echoes occur more frequently and cover larger areas than ship clutter, there is a heightened risk of misidentifying these echoes as ship clutter. Therefore, narrowing the identification area is crucial.

The HSCI algorithm incorporates three specific constraints to enhance efficiency and accuracy:

1) Identification is conducted only over sea areas, as this is naturally the most likely location for ship clutter.

135 2) Identification of range gates with local maximum reflectivity is based on the definition used in this study, in which the mainlobe is the range gate with the highest reflectivity within the ship clutter. 3) A reflectivity threshold is set (defaulting to 20 dBZ) because ship clutter with lower reflectivity typically does not compromise data quality significantly.

3.1.2 Feature calculation

140 Extracting features from radar variable fields that can differentiate between ship clutter and precipitation echoes is a crucial step in the implementation of machine learning algorithms. Upon analyzing these fields, six distinct features were identified: Reflectivity Difference (RD), Reflectivity Gradient Flag (RGF), σ_v , Spectrum Width Ratio (SWR), Z_{DR} , and Correlation Coefficient Difference (CCD).

(1) RD

145 The ship is a typical point target. As depicted in Fig. 3a, there was a sharp decrease in the reflectivity when the ship moved from being directly in the beam's mainlobe to being slightly off-center. This pattern was also observed in the pulse compression radar along the radial direction. Within the HSCI algorithm, RD was developed to quantify this phenomenon, which is defined as:

$$
RD = \max \left[Z_H(x, y) - Z_H(x - 1, y), Z_H(x, y) - Z_H(x + 1, y), Z_H(x, y) - Z_H(x, y - 1), Z_H(x, y) - Z_H(x, y + 1) \right], \quad (1)
$$

150 where x and y represent the tangential and radial indices of the mainlobe in the radar variable field, respectively, and max denotes the maximum-value function.

(2) RGF

The principle of the RGF is like that of the RD, indicating that the reflectivity decreases as the distance from the mainlobe increases. RGF is defined as follows:

155 $RGF = [Z_H(x-1, y) > Z_H(x-2, y)] \& [Z_H(x+1, y) > Z_H(x+2, y)] \& [Z_H(x, y-1) > Z_H(x, y-2)] \& [Z_H(x, y+1) > Z_H(x, y-2)]$ $Z_H(x, y + 2)$], (2)

where & represents the AND operation. It is important to note that, unlike RD and the other features, RGF yields a Boolean value.

 (3) σ_v

160 As a rigid target, a ship exhibits extremely high velocity consistency. In contrast to the precipitation echo, which is formed by a multitude of precipitation particles within the sampling volume, the mainlobe of ship clutter displays a significantly lower σ_v , comparable even to ground clutter. Consequently, σ_v was chosen as one of the features in the HSCI algorithm. (4) SWR

When analyzing the σ_{ν} of ship clutter, a sudden change in the σ_{ν} values at the position of the mainlobe and its adjacent 165 antenna sidelobes was observed such that $\sigma_v(x \pm 1, y) \gg \sigma_v(x, y)$. This phenomenon was observed by Feng and Fabry

(2016). It was explained by the sharp change in the antenna phase pattern near the mainlobe and highlights a distinct characteristic of ship clutter.

Although $\sigma_{v}(x \pm 1, y) > \sigma_{v}(x, y)$, the difference is relatively minor when compared to the σ_{v} of precipitation echoes. To better illustrate the relative relationship between σ_{ν} ($x \pm 1$, y) and σ_{ν} (x , y), we propose using the SWR, defined as follows:

170
$$
SWR = \max \left[\frac{\sigma_v(x+1,y)}{\sigma_v(x,y)}, \frac{\sigma_v(x-1,y)}{\sigma_v(x,y)} \right].
$$
 (3)

 $(5) Z_{DR}$

 Z_{DR} , the ratio of reflectivity from horizontal to vertical polarization, can to some extent indicate the shape of precipitation particles (Seliga and Bringi, 1976). During descent, raindrops encounter air resistance that causes large raindrops to split into smaller droplets. On the other hand, hail tumbles as it falls, leading to Z_{DR} values close to 0 dB. Consequently, Z_{DR} values

175 for most precipitation echoes typically fall within the specific range of (-1 to 6 dB; Kumjian, 2013). However, in our analysis of ship clutter, we observed that Z_{DR} values varied almost randomly across the entire range (-8 to 8 dB in Kumpula radar). Thus, Z_{DR} was incorporated as a feature into the HSCI algorithm. (6) CCD

As depicted in Figs. 2b and 3c, the mainlobe of ship clutter typically exhibits a high ρ_{HV} , like that of precipitation echoes, 180 while that in the antenna sidelobes of ship clutter sharply decreases. This phenomenon can be attributed to the antenna

pattern, where the horizontal and vertical polarization channels align well in the mainlobe but mismatch in the sidelobes. Consequently, this study introduces the CCD to quantify the disparity between the antenna mainlobe and the sidelobes of ship clutter. The CCD is defined as follows:

$$
CCD = \max [\rho_{HV}(x, y) - \rho_{HV}(x - 2, y), \rho_{HV}(x, y) - \rho_{HV}(x + 2, y)]. \tag{4}
$$

185

Figure 5 shows the normalized histograms for the six features of ship clutter and precipitation echoes using the datasets specified in Sect. 2. The overlapping areas of the probability distribution densities of the two echo types are listed in Table 2. This overlap quantitatively reflects the discriminatory ability of each feature, with smaller values indicating better differentiation capability. Although the statistical analysis revealed that RD offers the most effective discrimination among

190 all the features, neither ship clutter nor precipitation echoes can be accurately distinguished by relying solely on a single feature. Thus, it is essential to integrate multiple features to further enhance the identification accuracy.

Figure 5: Normalized histograms of six features across selected datasets. (a) RD; (b) RGF; (c) σ_n **; (d) SWR; (e)** Z_{DR} **; and (f) CCD.** 195 **Blue and orange denote ship clutter and precipitation echoes, respectively.**

| Feature | Overlapping area |
|--------------------|------------------|
| RD | 14.34% |
| RGF | 25.93% |
| σ_v | 26.56% |
| SWR | 34.13% |
| \mathcal{Z}_{DR} | 33.35% |
| CCD | 20.19% |
| | |

Table 2: Overlapping area between normalized histograms for ship clutter and precipitation echoes of six features.

3.1.3 Identification model

200 In this study, a random forest model was used to integrate multiple features to identify ship clutter. Random forest is a classic machine learning algorithm that primarily constructs multiple decision trees and combines their prediction results to enhance overall prediction accuracy and stability. Owing to the advantages of random forest, such as high execution efficiency, no need to scale input features, and the ability to handle missing data, it has been widely used in the field of weather radar, including tornado identification (Sandmæl et al., 2023), precipitation forecasting (Mao and Sorteberg, 2020),

205 and raindrop size distribution retrievals (Conrick et al., 2020). This study does not provide an in-depth introduction to the principle of the random forest algorithm; details can be found in Ho (1998) and Breiman (2001).

Like other supervised learning methods, the development of the random forest model involves two steps: training and testing. The ship clutter and precipitation datasets mentioned in Sect. 2 were split into training and test sets at a ratio of 3:1. Although the selected precipitation dataset is extensive, approximately 10,000 range gates only remain after applying the

- 210 region limitation described in Sect. 3.1.1 (7,500 for training and 2,500 for testing). In this study, the Python Scikit-learn machine learning library was used for training, testing, and subsequent prediction tasks (Pedregosa et al., 2011). The input for the random forest model comprises the six features of the target range gate, and the output is a Boolean identification result, where 1 and 0 represent ship clutter and precipitation echoes, respectively. The hyperparameter configurations of the random forest model in the HSCI algorithm are listed in Table 3. The "GridSearchCV" method from Scikit-learn was used to
- 215 determine the optimal hyperparameters (listed in the third column of Table 3) by tuning the model through iterations over the hyperparameter value ranges (shown in the second column of Table 3). A detailed description of these hyperparameters can be found in Pedregosa et al., (2011).

Table 3: List of hyperparameter values used in random forest model for HSCI algorithm.

3.2 Sidelobe identification

3.2.1 Adaptively determine the potential sidelobe distribution (PSD)

Although the scattered energy of the ship clutter is predominantly concentrated in the mainlobe, the extensive distribution of weaker sidelobes can significantly interfere with radar data applications. Therefore, once the mainlobe is identified, the next 225 step involves identifying the sidelobes. Unlike mainlobe identification, which is performed gate by gate, sidelobe identification leverages mainlobe identification results to determine the PSD.

Sidelobes in ship clutter typically appear in a cross-shape but vary in size. The signal-to-noise ratio (SNR) of a sidelobe at different positions ($SNR_{side}(\Delta x, \Delta y)$) is influenced by the SNR of the mainlobe (SNR_{main}), the ambiguity function, and the antenna pattern, where Δx and Δy represent the distance from the mainlobe in tangential and radial directions, respectively.

- 230 If $SNR_{side}(\Delta x, \Delta y)$ at a range gate falls below a set SNR threshold, that gate will be masked, and no radar variables will be output. Therefore, a static PSD setting may be insufficient when SNR_{main} is high, and excessive when SNR_{main} is low. To determine the sidelobe distribution settings more effectively, it is essential to make adaptive decisions for different ship clutter events. Three main factors influence the sidelobe distribution: SNR_{main} , the relative power between the mainlobe and sidelobe, and the SNR threshold used in the radar variable estimation. Although the Kumpula radar does not directly output
- 235 SNR values, this study proposes a method to obtain the SNR indirectly from reflectivity (details in the Appendix). Moreover, the SNR threshold is set by the user and is a known value (1.5 dB for the Kumpula radar). The relative power between the mainlobe and sidelobe can be calculated theoretically using the ambiguity function of the specified pulse compression waveform and antenna pattern. However, discrepancies between theoretical analysis and actual observations may arise because of unforeseen factors. Consequently, this study derived the relative power between the mainlobe and sidelobe
- 240 through a statistical analysis of actual data.

To capture as many sidelobe distribution characteristics as possible, only ship clutter with $SNR_{main} > 50$ dB was selected from the dataset. Because statistical results can be skewed by echoes from other sources that overlap with the mainlobe and/or sidelobe of the ship clutter, only eight relatively isolated ship clutter events were ultimately selected for the analysis of PSD statistics. For these clutter events, range gates where sidelobes were located were manually selected within 13.5 km

- 245 (90 gates on Kumpula radar) in the radial direction and 15 degrees (15 rays) in the tangential direction, centered on the mainlobe. To facilitate statistical analysis across different ship clutter events, the SNR values of the ship clutter were normalized (i.e., the SNR_{main} and $SNR_{side}(\Delta x, \Delta y)$ were subtracted from the SNR_{main} in dB units). Owing to several factors in actual observations, the relative relationship between the mainlobe and sidelobes of the eight ship clutter events was inconsistent. As previously discussed, to capture as many sidelobe distribution characteristics as possible (thereby
- ensuring the results are applicable across a wide range of scenarios), the maximum value of $SNR_{side}(\Delta x, \Delta y)$ from the eight ship clutter events was selected. Additionally, the maximum values of the antenna sidelobes were obtained on both sides, using the antenna mainlobe as the reference axis.

The statistical results of the relative power between the mainlobe and sidelobe are shown in Fig. 6. When a range gate was identified as the mainlobe of the ship clutter, the SNR differences between it and the surrounding range gates were calculated. 255 Range gates with SNR differences exceeding the statistical results shown in Fig. 6 were identified as PSD.

Figure 6: Statistical result of the relative power between the mainlobe and sidelobe for eight ship clutter events with high SNR.

260 **3.2.2 Velocity and SNR filter**

As shown in Fig. 7, the adaptively determined PSD is effective in identifying all affected range gates for isolated ship clutter. However, when ship clutter overlaps with other types of echoes, such as the precipitation echoes shown in Fig. 8, eliminating the PSD can lead to loss of important information. In other words, the PSD is a sufficient but unnecessary condition for the sidelobe distribution of the ship clutter. Thus, additional constraints are required to refine the PSD screening.

265

Figure 7: 0.5° elevation of Kumpula radar using LFM waveform at 1155 UTC 5 May 2020. (a) Reflectivity before ship-clutter filtering; (b) Reflectivity after filtering all range gates in the PSD; (c) Doppler velocity; (d) Reflectivity after filtering using velocity and SNR thresholds.

270

Figure 8:0.5° elevation of Kumpula radar using LFM waveform at 1955 UTC 4 June 2020. (a) Reflectivity before ship-clutter filtering; (b) Reflectivity after filtering all range gates in the PSD; (c) Doppler velocity; (d) Reflectivity after filtering using velocity and SNR thresholds.

275

When analyzing the signatures of ship clutter across different radar variables, it was found that the v_r of the sidelobes exhibits consistent patterns. This consistency makes v_r a highly effective indicator for PSD screening. As shown in Fig. 7a, among the five identified ship clutter events, Nos. 1 and 4 are particularly notable for their distinct cross-shaped patterns. Correspondingly, their v_r (Fig. 7c) also display cross-shaped distributions with small differences within each group (standard

280 deviations of 0.48 and 0.37 m/s, respectively).

To quantitatively assess the v_r distribution of ship clutter, we selected datasets based on criteria like those for PSD statistics described in Sect. 3.2.1, albeit with less stringent SNR requirements. Consequently, statistics were gathered from 65 ship clutter samples. After normalizing the SNR and v_r of the ship clutter sidelobes, the data was categorized into bins based on the power difference between the sidelobes and the mainlobe, ranging from -80 to 0 dB in 20 dB increments. The statistical

285 outcomes, illustrated in the violin plot in Fig. 9, indicate that while there are outliers, a v_r threshold of 1 m/s is adequate to encompass most ship clutter sidelobes.

Figure 9: Violin plots showing the normalizing v_r for normalizing SNR of ship clutter sidelobes from -80 to 0 dB in steps of 20 dB.

A notable exception, however, is ship clutter No. 1 in Fig. 8, where precipitation and ship clutter overlap and exhibit similar v_r values. To address this, we introduced an additional SNR threshold, that is, the SNR of the ship clutter sidelobe must exceed the lower limit used for PSD determination in Sect. 3.2.1, but not exceed the higher SNR threshold (default to 10 dB). Typically, the SNR of ship clutter sidelobes is lower than that of precipitation echoes.

295 The identification results, after applying v_r and SNR thresholds, are presented in Figs. 7d and 8d. Compared with Fig. 8a, Fig. 8d effectively isolates and removes only the regions affected by ship clutter, with minimal loss of precipitation echo information. Meanwhile, there is no observable degradation in the identification performance for the isolated ship clutter between Figs 7a and 7d.

4 Performance evaluation

300 **4.1 Case analysis**

Figure 10 presents a typical clear air scenario observed by the Kumpula radar at 1310 UTC on May 5, 2020, where several ship clutter events were observed in isolation from the precipitation and sea clutter. The results of the mainlobe identification are indicated by red circles. In Fig. 10a, which displays the Z_H before the application of ship clutter filtering, both strong cross-shaped and weaker point-shaped ship clutter are evident. Following the implementation of ship clutter filtering, as

305 shown in Fig. 10b, both strong and weak ship clutter were effectively removed. The influence of ship clutter on Z_H was substantially mitigated, rendering it visually undetectable.

Figure 10: Reflectivity on 0.5° elevation of Kumpula radar using LFM waveform at 1310 UTC 5 May 2020. (a) Before ship clutter 310 **filtering; (b) After ship clutter filtering. The mainlobe identification results are highlighted by red circles.**

A precipitation event observed by the Kumpula radar at 0040 UTC on June 5, 2020 is shown in Fig. 11, in which several ship clutter events were completely mixed with precipitation echoes. The results of the mainlobe identification are indicated by the red circles in Figs. 11a and 11b and white circles in Figs. 11c and 11d. The ship clutter mainlobe exhibits distinct

- 315 characteristics from the surrounding precipitation echoes in the v_r field, which supports the accuracy of the mainlobe identification results to some extent. In cases where ship clutter sidelobes are involved, the echo intensity of most ship clutter sidelobes is generally lower than that of the adjacent precipitation echoes. As a result, the intrinsic characteristics of the ship clutter sidelobes, such as the v_r values approaching those of the ship clutter mainlobe, are not prominently displayed in these overlapping range gates. Instead, the characteristics of the dominant precipitation echoes prevail. Consequently, it is prudent
- 320 to eliminate only the mainlobe and the sidelobes close to the mainlobe, which possess sufficiently strong echo intensities. This selective filtering approach is demonstrated in Figs. 11b and 12d following the application of ship clutter filtering.

Figure 11: 0.5° elevation of Kumpula radar using LFM waveform at 0040 UTC 5 June 2020. (a) Reflectivity before ship clutter 325 **filtering; (b) Reflectivity after ship clutter filtering; (c) Doppler velocity before ship clutter filtering; (d) Doppler velocity after ship clutter filtering. The mainlobe identification results are highlighted by red or white circles.**

4.2 Statistical evaluation

A quarter of the manually curated dataset, consisting of 400 gates for ship clutter and 2500 gates for precipitation echoes, 330 was used to objectively assess the mainlobe identification results. The remainder of the dataset (75%) served as the training set for the random forest identification model. The model achieved identification accuracies of 97.25% and 99.92% for ship clutter and precipitation echoes, respectively. The performance of the identification process was further quantified using a probability density plot like that shown in Fig. 5, where the overlapping area between the distributions of ship clutter and precipitation echoes was only 2.83%. This represents a significant improvement over the results obtained using a single

335 feature, as detailed in Table 2, and underscores the benefits of integrating multiple features to enhance the identification accuracy.

In addition to using labeled datasets to evaluate the performance of mainlobe identification, this study also incorporated observed scanning data from the Kumpula radar. Unlike other studies that typically present identification results from one or a few radar scans (Tang et al., 2014; Kurdzo et al., 2020), our analysis encompasses a 24-hour precipitation event on June 4,

340 2020. The Z_H both before and after the removal of ship clutter, was converted into precipitation rates (R) using the $Z-R$ relationship $Z_H = 300R^{1.6}$ (Marshall and Palmer, 1948), and the total precipitation rate for the entire event was accumulated. The conversion of Z_H to the precipitation rate has two important purposes. First, the precipitation rate is a critical parameter

in meteorological applications, offering a more direct reflection of identification performance. Secondly, it facilitates the accumulation of data, allowing for the analysis of long-term effects.

345 The rain accumulations before and after the removal of ship clutter are shown in Figs. 12a and 12b. As illustrated in Fig. 12a, numerous areas with high precipitation accumulation appear in linear formations over the sea. Following the application of ship clutter removal, as shown in Fig. 12b, these anomalous values were effectively eliminated, whereas the precipitation echoes in other areas remained unaffected. This confirms the efficacy of the ship clutter identification algorithm and demonstrates its capability to enhance the accuracy of precipitation measurements. Figure 12c shows the difference in 350 precipitation accumulation before and after ship clutter removal, highlighting the overestimation of precipitation caused by ship clutter. A comparison with the ship traffic density map of the Gulf of Finland shown in Fig. 12d reveals a strong correlation between the two, substantiating the assertion that these echoes originated from ships.

355 **Figure 12: 24-hour rain accumulation on 4 June 2020 from Kumpula radar at 0.5° elevation. (a) Before ship clutter filtering; (b) After ship clutter filtering; (c) Difference between before and after filtering; (d) Ship density map of the Gulf of Finland (© MarineTraffic 2024; Image source: www.marinetraffic.com).**

5. Discussion

360 The evaluation results confirm that the HSCI algorithm effectively identifies both the mainlobe and sidelobes of ship clutter; the only way to mitigate their negative impact is to mask the radar variables at the range gates where they are detected, which is the method used in this study. However, when precipitation echoes overlap with ship clutter, this inevitably results

in loss of precipitation data. An effective strategy to address this challenge is to move the stage of ship clutter identification and filtering from "data processing" to "signal processing" (Keeler and Passarelli, 1990). This approach involves using 365 spectrum processing techniques to suppress the ship clutter component in the radar signal while preserving the precipitation component like ground clutter filtering methods such as GMAP (Siggia and Passarelli, 2004) and CLEAN-AP (Torres and Warde, 2014). The center position of the notch filter needs to be adjusted from zero frequency to match the v_r of the ship clutter's mainlobe. Additionally, delayed processing across several azimuth intervals is necessary owing to the cross-radial

effect of the antenna sidelobes. Although this adjustment may affect the real-time performance of radar signal processing, it

- 370 represents a cost-effective compromise to reduce the influence of ship clutter on radar data. Phased array technology is becoming increasingly prevalent in weather radar. Since the early 21st century, the United States has conducted experiments and developed advanced phased array weather radars such as ATD (Weber et al., 2021) and Horus (Palmer et al., 2023)). Similarly, China pioneered the operational application of phased array technology in several provinces (Geng and Liu, 2023). Pulse compression technology, a cornerstone of active phased array radars, suggests that
- 375 the HSCI algorithm proposed in this study may have even wider applications in the future. Moreover, phased array radars often use digital beamforming for rapid scanning. However, this can exacerbate the deterioration of the antenna sidelobes in the direction of elevation (Schvartzman et al., 2021). Consequently, it is anticipated that future HSCI algorithms will evolve from two-dimensional to three-dimensional.

6. Summary

380 In this study, a Hybrid Ship Clutter Identification (HSCI) algorithm for pulse compression weather radar was introduced. This algorithm not only identifies the mainlobe of the ship clutter but also detects the range sidelobes resulting from pulse compression technology and the antenna sidelobes inherent to all radars. Data observed using the Kumpula radar at the University of Helsinki (from May to June 2020), which frequently captures the activities of ships sailing in the Gulf of Finland, were used in this study. For algorithm development and performance evaluation, 1,600 ship clutter and 10,000 385 range gates of precipitation echoes were manually selected.

The HSCI algorithm is structured into two parts: mainlobe and sidelobe identification. In the mainlobe identification step, the identification region is initially limited to minimize false identifications and enhance efficiency. Six features—RD, RGF, σ_{v} , SWR, Z_{DR} , and CCD—are then calculated and used in a random forest model to distinguish ship clutter mainlobes from precipitation echoes. If the model confirms the presence of a ship clutter mainlobe, the process transitions to sidelobe 390 identification.

The potential sidelobe distribution (PSD) of ship clutter is dynamic, increasing or decreasing with the SNR of the mainlobe. The first step in sidelobe identification uses an adaptive method to accurately determine the PSD, thus avoiding missed identifications that leave sidelobe residues or excessive identifications that lead to precipitation data loss. Velocity and SNR

corresponded well with the ship density map.

filters were then applied within the PSD to further protect against the loss of precipitation information due to overlapping 395 ship clutter and precipitation echoes.

Two typical cases (one in clear air and the other during precipitation) were used for the algorithm performance analysis. The results demonstrate that isolated ship clutter is accurately identified and filtered out, whereas ship clutter overlapping with precipitation is effectively identified and removed while preserving the precipitation data. In addition, the algorithm achieved correct identification rates of 97.25% for ship clutter and 99.92% for precipitation echoes on a test dataset 400 comprising 400 ship clutter gates and 2,500 range gates of precipitation echoes. This study also assessed the cumulative precipitation before and after ship clutter filtering during a 24-hour precipitation event, finding that the precipitation overestimation caused by ship clutter was effectively eliminated, and the footprint of precipitation overestimation

Appendix

- 405 SNR is commonly used as a threshold parameter to mask regions with noise and weak signals that are significantly influenced by noise. Although SNR can be included as part of the archived data along with other radar parameters such as Z_H , v_r , σ_v , Z_{DR} , ϕ_{DP} , and ρ_{HV} , in some signal processors, it is not always mandatory. For instance, the Kumpula radar used in this study did not output an SNR. Given that the SNR is a crucial factor for adaptively determining the PSD in the HSCI algorithm, an estimation technique was developed to accurately obtain the SNR using reflectivity (Z) .
- 410 In Vaisala RVP signal processors, *Z* (expressed in logarithmic units) is estimated from the SNR (expressed in logarithmic units) and a series of constants (Vaisala, 2016):

$$
Z = SNR + C \tag{A1}
$$

For simplicity, a series of constants is consolidated and denoted by C. When the SNR reached the preset threshold SNR_{thr} (1.5 dB for the Kumpula radar), the radar detected the minimum reflectivity Z_{min} :

$$
415 \quad Z_{min} = SNR_{thr} + C \tag{A2}
$$

Subtracting Eq. (A2) from Eq. (A1): $Z - Z_{min} = SNR - SNR_{thr}$. (A3)

The SNR can then be determined by transposing:

$$
SNR = Z - Z_{min} + SNR_{thr} \tag{A4}
$$

420 In Eq. (A4), Z can be sourced from the archived data, and SNR_{thr} is a predefined value set by the user. Z_{min} for each range can be determined through statistical analysis of a large dataset.

Competing interests

The contact author has declared that none of the authors has any competing interests.

References

- 425 Argenti, F. and Facheris, L.: Radar pulse compression methods based on nonlinear and quadratic optimization, IEEE Trans. Geosci. Remote Sens., 59, 3904–3916. https://doi.org/10.1109/TGRS.2020.3010414, 2020. Bharadwaj, N. and Chandrasekar, V.: Wideband waveform design principles for solid-state weather radars, J. Atmos. Ocean. Technol., 29, 14–31. https://doi.org/10.1175/JTECH-D-11-00030.1, 2012. Billingsley, J. B.: Low-angle radar land clutter: Measurements and empirical models, IET, 2002.
- 430 Breiman, L.: Random forests, Mach. Learn., 45, 5–32. https://doi.org/10.1023/A:1010933404324, 2001. Chen, Y., Zheng, H., Sun, T., Meng, D., Qin, L., Yin, J.: Improving forecasts of the "21⋅ 7" Henan extreme rainfall event using a radar assimilation scheme that considers hydrometeor background error covariance, Mon. Weather Rev., 152, 1379– 1397. https://doi.org/10.1175/MWR-D-23-0190.1, 2024.

Conrick, R., Zagrodnik, J. P., and Mass, C. F.: Dual-polarization radar retrievals of coastal pacific northwest raindrop size 435 distribution parameters using random forest regression, J. Atmos. Ocean. Technol., 37, 229–242. https://doi.org/10.1175/JTECH-D-19-0107.1, 2020.

Cook, C. E. and Bernfeld, M.: Radar signals: An introduction to theory and application, Academic Press, New York, 1967. Crum, T. D. and Alberty, R. L.: The WSR-88D and the WSR-88D operational support facility, Bull. Am. Meteorol. Soc., 74, 1669–1687. https://doi.org/10.1175/1520-0477(1993)074<1669:TWATWO>2.0.CO;2, 1993.

- 440 Feng, Y. C. and Fabry, F.: The imperfect phase pattern of real parabolic radar antenna and data quality, J. Atmos. Ocean. Technol., 33, 2655–2661. https://doi.org/10.1175/JTECH-D-16-0143.1, 2016. Geng, F. and Liu, L.: Study on attenuation correction for the reflectivity of X-band dual-polarization phased-array weather radar based on a network with S-band weather radar, Remote Sens., 15, 1333. https://doi.org/10.3390/rs15051333, 2023. Golbon-Haghighi, M. H., Zhang, G., and Doviak, R. J.: Ground clutter detection for weather radar using phase fluctuation
- 445 index, IEEE Trans. Geosci. Remote Sens., 57, 2889–2895. https://doi.org/10.1109/TGRS.2018.2878378, 2018. Hanft, W., Zhang, J., and Simpson, M.: Dual-pol VPR corrections for improved operational radar QPE in MRMS, J. Hydrol. Meteorol., 24, 353–371. https://doi.org/10.1175/JHM-D-22-0010.1, 2023. Ho, J., Zhang, G., Bukovcic, P., Parsons, D. B., Xu, F., Gao, J., Carlin, J. T., Snyder, J. C.: Improving polarimetric radarbased drop size distribution retrieval and rain estimation using a deep neural network, J. Hydrol. Meteorol., 24, 2057–2073.
- 450 https://doi.org/10.1175/JHM-D-22-0166.1, 2023.

Ho, T. K.: The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell., 20, 832– 844. https://doi.org/10.1109/34.709601, 1998.

Hubbert, J. C., Dixon, M., and Ellis, S. M.: Weather radar ground clutter. Part II: Real-time identification and filtering, J. Atmos. Ocean. Technol., 26, 1181–1197. https://doi.org/10.1175/2009JTECHA1160.1, 2009.

- 455 Hubbert, J. C., Dixon, M., Ellis, S. M., Meymaris, G.: Weather radar ground clutter. Part I: Identification, modeling, and simulation, J. Atmos. Ocean. Technol., 26, 1165–1180. https://doi.org/10.1175/2009JTECHA1159.1, 2009. Hubbert, J. C., Meymaris, G., Romatschke, U., et al: Using a regression ground clutter filter to improve weather radar signal statistics: Theory and simulations, J. Atmos. Ocean. Technol., 38, 1353–1375, 2021. Keeler, R. J. and Passarelli, R. E.: Signal processing for atmospheric radars: Radar in Meteorology: Battan Memorial and
- 460 40th Anniversary Radar Meteorology Conference, American Meteorological Society, Boston, MA, 199–229, 1990. Kilambi, A., Fabry, F., and Meunier, V.: A simple and effective method for separating meteorological from nonmeteorological targets using dual-polarization data, J. Atmos. Ocean. Technol., 35, 1415–1424. https://doi.org/10.1175/JTECH-D-17-0175.1, 2018.

Kollias, P., Palmer, R., Bodine, D., Adachi, T., Bluestein, H., Cho, J. Y. N., Griffin, C., Houser, J., Kirstetter, P. E., Kumjian, 465 M. R., Kurdzo, J. M., Lee, W. C., Luke, E. P., Nesbitt, S., Oue, M., Shapiro, A., Rowe, A., Salazar, J., Tanamachi, R.,

Tuftedal, K. S., Wang, X., Zrnić, D., Treserras, B. P.: Science applications of phased array radars, Bull. Am. Meteorol. Soc., 103, E2370–E2390. https://doi.org/10.1175/BAMS-D-21-0173.1, 2022.

Kumjian, M. R.: Principles and applications of dual-polarization weather radar. Part I: Description of the polarimetric radar variables, J. Oper. Meteorol., 1, 226-242. https://doi.org/10.15191/nwajom.2013.0119, 2013.

- 470 Kurdzo, J. M., Bennett, B. J., Smalley, D. J., Donovan, M. F.: The WSR-88D inanimate hydrometeor class, J. Appl. Meteorol. Climatol., 59, 841–858. https://doi.org/10.1175/JAMC-D-19-0271.1, 2020. Kurdzo, J. M., Cheong, B. L., Palmer, R. D., Zhang, G., Meier, J. B.: A pulse compression waveform for improvedsensitivity weather radar observations, J. Atmos. Ocean. Technol., 31, 2713–2731. https://doi.org/10.1175/JTECH-D-13- 00021.1, 2014.
- 475 Li, H., Moisseev, D., Luo, Y., Liu, L., Ruan, Z., Cui, L., Bao, X.: Assessing specific differential phase (K DP)-based quantitative precipitation estimation for the record-breaking rainfall over Zhengzhou city on 20 July 2021, Hydrol. Earth Syst. Sci., 27, 1033–1046. https://doi.org/10.5194/hess-27-1033-2023, 2023.

Li, H., Yin, J., and Kumjian, M.: ZDR backwards arc: Evidence of multi‐directional size sorting in the storm producing 201.9 mm hourly rainfall, Geophys. Res. Lett., 51, GL109192, e2024. https://doi.org/10.1029/2024GL109192, 2024.

480 Mao, Y. and Sorteberg, A.: Improving radar-based precipitation nowcasts with machine learning using an approach based on random forest, Weather Forecasting, 35, 2461–2478. https://doi.org/10.1175/WAF-D-20-0080.1, 2020. Marshall, J. S. and Palmer, W. M. K.: The distribution of raindrops with size, J. Meteor., 5, 165–166. https://doi.org/10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2, 1948.

Overeem, A., Uijlenhoet, R., and Leijnse, H.: Full-year evaluation of nonmeteorological echo removal with dual-polarization

485 fuzzy logic for two C-band radars in a temperate climate, J. Atmos. Ocean. Technol., 37, 1643–1660. https://doi.org/10.1175/JTECH-D-19-0149.1, 2020.

Palmer, R., Bodine, D., Kollias, P., Schvartzman, D., Zrnić, D., Kirstetter, P., Zhang, G., Yu, T. Y., Kumjian, M., Cheong, B., Collis, S., Frasier, S., Fulton, C., Hondl, K., Kurdzo, J., Ushio, T., Rowe, A., Salazar-Cerrenˉo, J., Torres, S., Weber, M., Yeary, M.: A primer on phased array radar technology for the atmospheric sciences, Bull. Am. Meteorol. Soc., 103, E2391– 490 E2416. https://doi.org/10.1175/BAMS-D-21-0172.1, 2022.

- Palmer, R. D., Yeary, M. B., Schvartzman, D., Salazar-Cerreno, J. L., Fulton, C., McCord, M., Cheong, B., Bodine, D., Kirstetter, P., Sigmarsson, H. H., Yu, T. Y., Zrnić, D., Kelley, R., Meier, J., Herndon, M.: Horus—A fully digital polarimetric phased array radar for next-generation weather observations, Trans. Rad. Sys., 1, 96–117. https://doi.org/10.1109/TRS.2023.3280033, 2023.
- 495 Pang, C., Hoogeboom, P., Le Chevalier, F., Russchenberg, H. W. J., Jian Dong, Tao Wang, Xuesong Wang: A pulse compression waveform for weather radars with solid-state transmitters, IEEE Geosci. Remote Sens. Lett., 12, 2026–2030. https://doi.org/10.1109/LGRS.2015.2443551, 2015.

500 Pérez Hortal, A. A. and Michelson, D.: Using a continental-scale data quality monitoring framework to evaluate a new nonweather filter for radar observations, J. Appl. Meteorol. Climatol., 62, 411–425. https://doi.org/10.1175/JAMC-D-22- 0014.1, 2023.

Rihaczek, A. W.: Principles of high-resolution radar, McGraw-Hill, New York, 1969.

- Sandmæl, T. N., Smith, B. R., Reinhart, A. E., Schick, I. M., Ake, M. C., Madden, J. G., Steeves, R. B., Williams, S. S., 505 Elmore, K. L., Meyer, T. C.: The tornado probability algorithm: A probabilistic machine learning tornadic circulation detection algorithm, Weather Forecasting, 38, 445–466. https://doi.org/10.1175/WAF-D-22-0123.1, 2023.
	- Schvartzman, D., Torres, S. M., and Yu, T. Y.: Distributed beams: Concept of operations for polarimetric rotating phased array radar, IEEE Trans. Geosci. Remote Sens., 59, 9173–9191. https://doi.org/10.1109/TGRS.2020.3047090, 2021.
- Seliga, T. A. and Bringi, V. N.: Potential use of radar differential reflectivity measurements at orthogonal polarizations for 510 measuring precipitation, J. Appl. Meteor., 15, 69–76. https://doi.org/10.1175/1520- 0450(1976)015<0069:PUORDR>2.0.CO;2, 1976.

Siggia, A. D. and Passarelli, R. E.: Gaussian model adaptive processing (GMAP) for improved ground clutter cancellation and moment calculation//Proc., Eur. Rad., 2, 421–424, 2004.

Stepanian, P. M., Horton, K. G., Melnikov, V. M., Zrnić, D. S., Gauthreaux, S. A.: Dual-polarization radar products for 515 biological applications, Ecosphere, 7, e01539. https://doi.org/10.1002/ecs2.1539, 2016.

Tang, L., Zhang, J., Langston, C., Krause, J., Howard, K., 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. https://doi.org/10.1175/WAF-D-13-00072.1, 2014.

Torres, S. M., Curtis, C. D., and Schvartzman, D.: Requirement-driven design of pulse compression waveforms for weather 520 radars, J. Atmos. Ocean. Technol., 34, 1351–1369. https://doi.org/10.1175/JTECH-D-16-0231.1, 2017.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al.: Scikit-learn: Machine learning in Python, J. Mach. Learn. Res., 12, 2825– 2830, 2011.

Torres, S. M. and Warde, D. A.: Ground clutter mitigation for weather radars using the autocorrelation spectral density, J. Atmos. Ocean. Technol., 31, 2049–2066. https://doi.org/10.1175/JTECH-D-13-00117.1, 2014. Vaisala: User's manual: RVP900 digital receiver and signal processor; Vaisala oyj: Vantaa, Finland, 513, 2016. Weber, M., Hondl, K., Yussouf, N., Jung, Y., Stratman, D., Putnam, B., Wang, X., Schuur, T., Kuster, C., Wen, Y., Sun, J., 525 Keeler, J., Ying, Z., Cho, J., Kurdzo, J., Torres, S., Curtis, C., Schvartzman, D., Boettcher, J., Nai, F., Thomas, H., Zrnić, D., Ivić, I., Mirković, D., Fulton, C., Salazar, J., Zhang, G., Palmer, R., Yeary, M., Cooley, K., Istok, M., Vincent, M.: Towards

the next generation operational meteorological radar, Bull. Am. Meteorol. Soc., 102, E1357–E1383. https://doi.org/10.1175/BAMS-D-20-0067.1, 2021.