SPEKTRA: Jurnal Fisika dan Aplikasinya
Volume 8, Issue 3, December 2023
Homepage: doi.org/10.21009/SPEKTRA

Check for updates

DOI: doi.org/10.21009/SPEKTRA.083.05

IMPLEMENTATION OF GABELLA METHOD AND RANDOM FOREST FOR GROUND CLUTTER DETECTION IN PADANG WEATHER RADAR DATA

Wildan Nurrahman[1,*], Abdullah Ali[2], Riser Fahdiran[1]

[1]*Department of Physics, State University of Jakarta Jl. Rawamangun Muka, Jakarta 13220, Indonesia*
[2]*Remote Sensing Data Management Division, Indonesia Agency for Meteorology and Geophysics Jl. Angkasa 1, Jakarta 10720, Indonesia*

*Corresponding Author Email: wldannurrahman2412@gmail.com

Received: 19 September 2023
Revised: 14 December 2023
Accepted: 17 December 2023
Online: 31 December 2023
Published: 31 December 2023

SPEKTRA: Jurnal Fisika dan Aplikasinya
p-ISSN: 2541-3384
e-ISSN: 2541-3392

ABSTRACT

Weather radar is an active remote sensing instrument for various hydrological and meteorological applications. One advantage of weather radar is its ability to detect rainfall in space and time with high spatial resolution. However, one of the issues that contaminate radar observations is ground clutter. Ground clutter is a signal or echo from non-meteorological objects on the earth's surface that are stationary in the time domain. Detecting and mitigating clutter effects is crucial to achieve precise weather measurements. This research aims to implement the Gabella and random forest methods to detect ground clutter in Padang weather radar data and determine the optimal method between the two. The implementation of the Gabella method for detecting ground clutter in Padang weather radar data was suboptimal. This was due to the most duplicated data at the same point being only 15.97% of the total data. Meanwhile the random forest method obtained a kappa value of 92.03%. This indicates that the random forest model created using 2000 trees as the parameter performs well. Based on these results, the random forest method identified as the most optimal approach for detecting ground clutter in Padang weather radar data.

Keywords: gabella method, ground clutter detection, random forest, weather radar

INTRODUCTION

Weather radar is an active remote sensing instrument for various hydrological and meteorological implementations [1]. It serves as a crucial tool for monitoring and analyzing weather patterns, providing valuable data for hydrological and meteorological applications. Weather radar has the advantage of detecting rainfall space and times due to its high spatial resolution (less than 1 km) and temporal resolution (5 - 10 minutes) [2-3]. These capabilities enable meteorologists and hydrologists to track the evolution of precipitation events with remarkable detail. A fundamental problem before radar-derived precipitation amounts can be utilized for hydrological purpose is to ensure that they possess enough accuracy and robustness [4]. Rainfall estimation on weather radar is extremely complicated due to signal attenuation effects, ground clutter contamination, and radar calibration errors [5-6]. Ground clutter contamination presents another challenge, as non-meteorological objects such as terrain features and man-made structures can generate false echoes on radar screens, obscuring genuine precipitation signals and affecting the accuracy of rainfall estimates.

Ground clutter is a signal or echo from non-meteorological objects on the earth's surface that is stationary in the time domain [7]. When the sensor operates in forward-looking mode, the signal it transmits travels through the surrounding medium and encounters the resolution volume, primarily occupied by scatterers on the ground plane. This interaction results in the creation of ground clutter, which comprises unexpected echo data that interferes with weather observation tasks. Consequently, the near-ground meteorological signal may be submerged in strong ground clutter totally [8]. Objects commonly referred to as ground clutter include trees, mountains, rocks, snow, and other tall building structures [9]. The presence of ground clutter in weather radar observations results in a decline in the quality of data and can result in misinterpretation of radar echoes. Detecting and mitigating clutter effects is crucial to achieve precise weather measurements [10]. Ground clutter mitigation is usually performed at the lowest elevation angles of the weather radar because ground clutter is usually detected at the lowest elevation angles [11-12]. By applying filtering techniques and algorithms specifically designed to identify and remove clutter signals, meteorologists can improve the reliability of radar observations and enhance the accuracy of weather forecasts. Effective clutter mitigation contributes to more precise measurements of meteorological phenomena, ultimately leading to better-informed decision-making in various sectors reliant on weather information, such as agriculture, aviation, and disaster management.

Many studies have been conducted to detect and remove ground clutter on weather radar data. Heistermann and Jacobi researched to detect and remove clutter using a texture-based filter developed [15]. The Gabella method is based on the fact that ground clutter has higher reflectivity values compared to meteorological objects [15]. In that study, Gabella's method was implemented on the image of weather radar observations using adjusted parameters. To detect ground clutter, the study used the accumulation of several weather radar data for the past year [13].

Another method that can be used to detect ground clutter is random forest. A major benefit of using random forest for prediction modeling is the ability to handle datasets with a large number of predictor variables; however, often in practice, the number of predictors required for obtaining outcome predictions should be minimized to improve efficiency [14]. The use of random forest has been done by Ali et al. to classify echo interference, precipitation echo, and clear/no echo. The study was conducted on the Tangerang weather radar at the lowest elevation angle. The testing results of the model in that study yielded a kappa value of 0.9, indicating that the model can be used for classification [16].

This research aims to implement the Gabella method and random forest to detect ground clutter in Padang weather radar data. Both methods will be implemented using several parameter variations. Padang was chosen as a case study because the geographical conditions in West Sumatra mostly consist of mountains and highlands which are categorized as ground clutter.

METHOD

The study applied Padang Weather radar data on 21 March 2019. The Gabella method uses one day of data from 00.05 to 23.55 Western Indonesia Time (WIB) with a time interval of 10 minutes between data. The data is in volumetric format, with reflectivity values as the main input. The random forest method uses TIFF-formatted data with reflectivity values as input. The time range for the random forest method was from 00:05 to 00:55 WIB.

Volumetric data is processed using Python to extract radar coordinates, azimuth, elevation, range, and radar data which are then stored in a variable. This variable is used as input for the Gabella method. Raster data processing is done using ArcMap 10.8 for data labeling. The labeling is divided into three classes: ground clutter, precipitation echo, and clear/no echo. After labeling, the data is imported into Python as input for the random forest method.

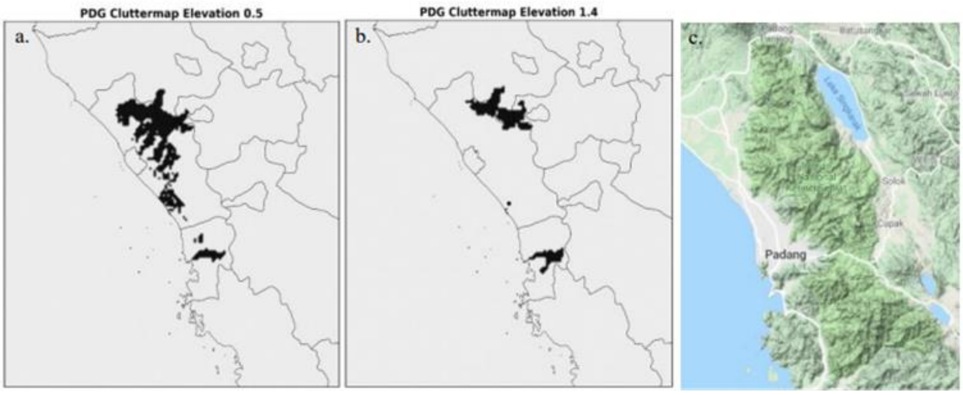
Gabella method will use the threshold parameter tr_{var} which is in the interval (3 dBZ ≤ tr_{var} ≤ 9 dBZ). The number of pixels n_p surrounding the tested pixel will be within the interval (6 ≤ n_p ≤ 10) and the window size will be 5x5 pixels. The selection of tr_{var} and n_p affects the process of clutter detection, but if both parameters are set at the interval, the results will not show drastic differences. The random forest parameter used is the number of trees. The number of trees used is 2000, 2500, and 3000. Varying the number of trees is done to observe the performance of the created random forest model.

The output of the Gabella method produces Boolean values, namely True and False. The output will be true if a specific pixel is indicated to be contaminated by ground clutter, and False if it is not indicated. Then, the True values are extracted using Python and saved in an Excel file. This step is repeated for one day of data with a distance of 10 minutes between data. Then calculate how much ground clutter is duplicated at the same point from the total amount of data used. This is because one of the properties of ground clutter is stationary in the time domain, meaning that ground clutter will remain in a fixed position at a certain point. The performance of the Gabella method can be evaluated using parameters that have been studied by Scovell et al. In that study, a point can be indicated as ground clutter if the count of occurrences at a particular point is 60% of the total accumulated data used [17]. In this study for the Gabella method, the data used totaled 144 data. So that a point can be categorized as ground clutter if it is duplicated 86 times.

The random forest method uses raster data and predefined parameters as input for classification. The classification is based on the three class labels: ground clutter, precipitation echo, and clear/no echo. The performance of the random forest method is evaluated using the kappa value. If the kappa value is less than 0.75 then the number of trees will be increased to increase the kappa value. The kappa value is obtained as follows.

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{1}$$

$\Pr(a)$ is percentage of total consistent measurements among raters, $\Pr(e)$ is percentage of total measurement changes among raters. Both methods will be compared in detecting ground clutter. The output from the Gabella method will focus on the top 10 duplicated data at the same point to see if those 10 data points account for more than 60% of the overall data. The random forest method will evaluate the model that has been made, if the kappa value produced is above 0.75 then the model can be made for classification.



(a) (b) (c)

FIGURE 1. Clutter maps (a) and (b), Topographic Map (c)

The technique used is spatial analysis technique to analyze the most optimal method. The spatial analysis involves the calculation and logical evaluation to discover geographic relationships within digital data [18]. Spatial analysis in this research is done by overlaying. The overlay method is an information system in graphical form formed by combining various individual maps [19]. The overlay method is implemented to analyze specific areas categorized as ground clutter by overlaying the output results of the random forest and Gabella methods onto the actual map. Both methods will be focused on points classified as ground clutter which are adjusted to the research conducted.

RESULT AND DISCUSSION

The parameter variations of Gabella method were implemented on 144 data from 00.05 to 23.55 WIB to find which parameter combination could identify ground clutter duplicated at the same point as much as 60% of the total data. Based on the results of Gabella parameter variations that have been carried out, it is found that there is no significant difference in the amount of duplicated data at the same point as shown in the table below.

TABLE 1. Duplicated data at the same point for tr_{var} of 4 dBZ with Gabella method

	tr_{var} = 4 dBZ								
	$n_p = 6$			$n_p = 8$			$n_p = 10$		
	Lat	**Long**	**Dup**	**Lat**	**Long**	**Dup**	**Lat**	**Long**	**Dup**
1	0.769422	100.3057	15	0.770564	100.3057	19	0.766281	100.3057	23
2	0.718167	100.3175	15	0.766281	100.3057	18	0.767621	100.3057	21
3	0.786584	100.305	14	0.791757	100.3054	18	0.766607	100.3057	21
4	0.766281	100.3057	14	0.786584	100.305	17	0.780478	100.3047	20
5	0.813225	100.305	14	0.849573	100.3128	17	0.885576	100.3053	20
6	0.783753	100.3048	14	0.749082	100.3014	17	0.770564	100.3057	20
7	0.791757	100.3054	14	0.761938	100.3038	17	0.802044	100.3043	20
8	0.783594	100.3048	14	0.718167	100.3175	17	0.811602	100.3045	20
9	0.765961	100.3057	14	0.811602	100.3045	16	0.749982	100.3014	20
10	0.701846	100.2998	14	0.793769	100.3041	16	0.849573	100.3128	20

TABLE 2. Duplicated data at the same point for tr_{var} of 6 dBZ with Gabella method

	tr_{var} = 6 dBZ								
	$n_p = 6$			$n_p = 8$			$n_p = 10$		
	Lat	**Long**	**Dup**	**Lat**	**Long**	**Dup**	**Lat**	**Long**	**Dup**
1	0.8028	100.3043	13	0.770564	100.3057	16	0.849573	100.3128	20
2	0.766281	100.3057	13	0.850688	100.3128	16	0.770564	100.3057	20
3	0.761549	100.3068	13	0.766281	100.3057	16	0.791757	100.3054	19
4	0.769422	100.3057	13	0.793769	100.3041	16	0.766281	100.3057	19
5	0.765961	100.3057	13	0.849573	100.3128	16	0766607	100.3057	19
6	0.762887	100.3033	12	0.800251	100.3033	15	0.813238	100.305	18
7	0.793874	100.3064	12	0.775522	100.3045	15	0.811602	100.3045	18
8	0.813225	100.305	12	0.813238	100.305	15	0.765961	100.3057	18
9	0.764253	100.3033	12	0.770178	100.3057	15	0.850688	100.3128	18
10	0.781406	100.3048	12	0.791757	100.3054	15	0.757631	100.3062	17

TABLE 3. Duplicated data at the same point for tr_{var} of 8 dBZ with Gabella method

	tr_{var} = 8 dBZ								
	$n_p = 6$			$n_p = 8$			$n_p = 10$		
	Lat	**Long**	**Dup**	**Lat**	**Long**	**Dup**	**Lat**	**Long**	**Dup**
1	0.769422	100.3057	12	0.766281	100.3057	15	0.849573	100.3128	18
2	0.813242	100.305	11	0.769422	100.3057	14	0.765961	100.3057	17
3	0.813225	100.305	11	0.749082	100.3014	14	0.791757	100.3054	17
4	0.761549	100.3068	11	0.850688	100.3128	14	0.766281	100.3057	17
5	0.764253	100.3033	11	0.781406	100.3048	14	0.813238	100.305	17
6	0.765961	100.3057	10	0.793769	100.3041	14	0.818327	100.306	16
7	0.799179	100.3034	10	0.76984	100.3031	14	0.770564	100.3057	16
8	0.227728	100.0683	10	0.8028	100.3043	13	0.767278	100.3057	16
9	0.786584	100.305	10	0.770564	100.3057	13	0.800251	100.3033	16
10	0.541589	100.4557	10	0.803923	100.3051	13	0.793769	100.3041	16

Lat is latitude, long is longitude, and dup is the number of duplicates. Based on all the tables above, it can be observed that the value of n_p increases, the number of data points duplicated at the same location also increases, while decreasing the threshold value tr_{var} leads to a higher number of duplicated data points at the same location. Based on the results of Gabella parameter variations that have been carried out, it is found that there is no significant difference in the amount of duplicated data at the same point. This is consistent with Gabella's statement that the selection of tr_{var} and n_p values affects the clutter detection process, but if the values are still in the interval (3 dBZ ≤ tr_{var} ≤ 9 dBZ) and the interval (6 ≤ n_p ≤ 10) then the results will not show drastic differences.

Based on the results of this study, the parameter configuration of $tr_{var} = 4$ dBZ and $n_p = 10$ that can produce the most number of duplicated data at the same point. which is 23 times at the point latitude 0.766281 longitude 100.3057 or only 15.97%. This result is significantly different from what was stated [14], where a point can be categorized as ground clutter if its occurrence count at a specific point accounts for 60% of the total accumulated data used. So the Gabella method with parameter variations used in this study cannot properly detect ground clutter at a certain point.

The random forest model that has been made is evaluated using kappa values calculated from EQUATION 1. For each variation of the number of trees parameter 2000, 2500, and 3000, a kappa value of 92.03% was obtained from the model that has been made. Based on these results, the random forest model can be used to classify ground clutter.

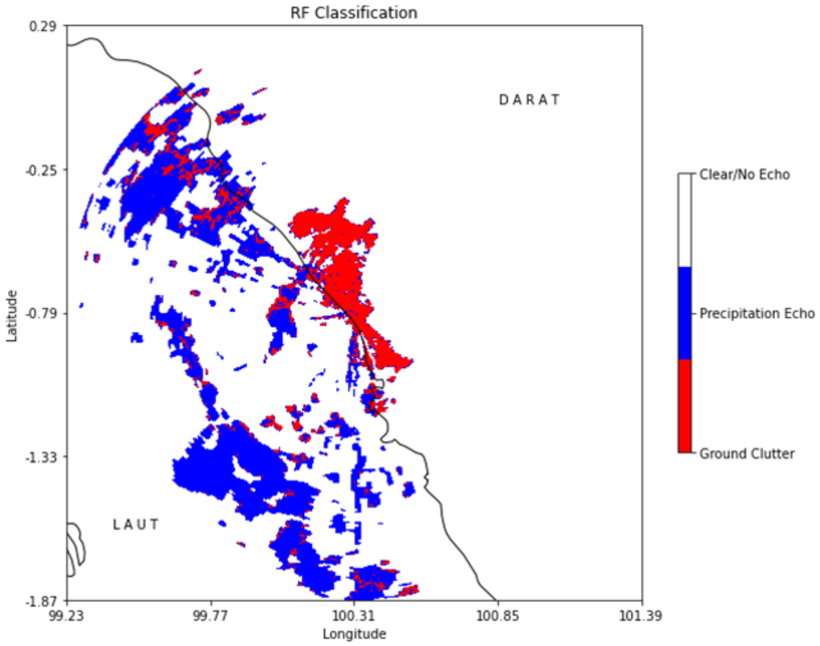


FIGURE 2. Ground Clutter Classification Result from Random Forest Method

The results of the classification showed that ground clutter marked with red color spread along the coast while rain clouds or precipitation echo spread across the sea. Based on FIGURE 1c, the ground clutter is present in the coastal area, as the region's topography is predominantly mountainous and highland. The ground clutter pattern obtained from the random forest model created exhibits a similar pattern to FIGURE 1a, where the ground clutter tends to dominate the mountainous regions with hilly topography in the northern, northeast, and southeastern parts. The results of the implementation of the random forest method obtained are still not perfect because some points are still wrongly classified as ground clutter. These points are indicated by red dots outside the coastline, meaning that the presence of red dots to the left of the coastline is a classification error. This occurs because the predicted classification results for some pixels do not match the actual pixels, resulting in misclassifications.

Based on the results of both methods used in this research, the random forest method is the most optimal method for detecting ground clutter in Padang weather radar data. This is due to the result of the random forest model that has been made to get a kappa value of 92.03% meaning the model already has a good performance, while the Gabella method for the variation of parameters that have been made only gets the amount of data duplicated at the same point of 15.97%. The spatial analysis of the random forest method's results, as shown in FIGURE 2, was performed using overlay techniques, referencing the study conducted. It can be observed that the distribution pattern of ground clutter obtained using the random forest method resembles the pattern in FIGURE 1a. Hence, the random forest method is considered the most optimal approach for detecting ground clutter in Padang weather radar data.

CONCLUSION

The implementation results of the Gabella method for detecting ground clutter in Padang weather radar data are suboptimal. This is because, after exploring different parameter settings, the maximum data duplication at the same point is only 15.97%. Meanwhile the random forest method obtained a kappa score of 92.03%. This means that the random forest model that has been made using the parameters of a number of trees of 2000 has a good performance. This is evidenced by the results showing a ground clutter distribution pattern consistent with previous research findings which is ground clutter tends to dominate the mountainous regions with hilly topography in the northern, northeast, and southeastern parts. Therefore, the random forest method is the most optimal approach for detecting ground clutter in Padang weather radar data.

REFERENCES

[1] N. Nanding and M. A. Rico-Ramirez, “Precipitation measurement with weather radars,” *ICT for Smart Water Systems: Measurements and Data Science*, pp. 235-258, 2021.

[2] A. A. Arbain, F. Sunarto and E. Mulyana, “Deteksi Es dan Hail di Atmosfer dengan Radar Polrimetrik X-Band Furuno WR-2100,” *Jurnal Sains & Teknologi Modifikasi Cuaca*, vol. 19, no. 1, pp. 21-31, 2018.

[3] R. Y. Mardyansyah *et al*., “Optimalisasi Saluran Komunikasi Berbasis Gelombang Mikro Sebagai Alternatif Sistem Pemantauan Curah Hujan,” *Elektron Jurnal Ilmiah*, pp. 21-29, 2022.

[4] J. Yan and A. Bárdossy, “Short time precipitation estimation using weather radar and surface observations: With rainfall displacement information integrated in a stochastic manner,” *Journal of Hydrology*, vol. 574, pp. 672-682, 2019.

[5] D. Putri, E. Y. Nugroho and J. R. Pratama, “Kajian Impelementasi Quality Control Faktor Bright Band dan Atenuasi Radar Cuaca C-Band,” *Jurnal Teori dan Aplikasi Fisika*, vol. 9, no. 1, pp. 111-120, 2021.

[6] J. C. Hubbert, M. Dixon and S. M. Ellis, "Weather radar ground clutter. Part II: Real-time identification and filtering," *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 7, pp. 1181-1197, 2009.

[7] M. H. Golbon-Haghighi and G. Zhang, "Detection of ground clutter for dual-polarization weather radar using a novel 3D discriminant function," *Journal of Atmospheric and Oceanic Technology*, vol. 36, no. 7, pp. 1285-1296, 2019.

[8] Y. Wang *et al*., "A weather signal detection algorithm based on EVD in elevation for airborne weather radar," *Digital Signal Processing*, vol. 116, 2021.

[9] K. Osródka and J. Szturc, "Improvement in algorithms for quality control of weather radar data (RADVOL-QC system)," *Atmospheric Measurement Techniques (AMT)*, vol. 15, no. 2, pp. 261-277, 2022.

[10] Golbon-Haghighi *et al*., "Detection of ground clutter from weather radar using a dual-polarization and dual-scan method," *Atmosphere*, vol. 7, no. 6, pp. 2-11, 2016.

[11] D. A. Warde and S. M. Torres, "Staggered-PRT sequences for Doppler weather radars. Part II: Ground clutter mitigation on the NEXRAD network using the CLEAN-AP filter," *Journal of Atmospheric and Oceanic Technology*, vol. 34, no. 3, pp. 703-716, 2017.

[12] T. Y. Yu *et al*., "Variational analysis of oversampled dual-Doppler radial velocity data and application to the analysis of tornado circulations," *Journal of Atmospheric and Oceanic Technology* (*JAOT)*, vol. 24, no. 4, pp. 616-626, 2007.

[13] J. L. Speiser *et al*., "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, vol. 134, no. 6, pp. 2-11, 2019.

[14] S. Jacobi and M. Heistermann, "Benchmarking attenuation correction procedures for six years of single-polarized C-band weather radar observations in South-West Germany," *Geomatics, Natural Hazards, and Risk*, vol. 7, pp. 93-101, 2016.

[15] M. Gabella and R. Notarpietro, "Ground clutter characterization and elimination in mountainous terrain," *Proceedings of ERAD Use of radar observations in hydrological and NWP models*, pp. 305-311, 2002.

[16] A. Ali *et al*., "Preliminary Study Of A Radio Frequency Interference Filter For Non-Polarimetric C-Band Weather Radar In Indonesia (Case Study: Tangerang Weather Radar)," *International Journal of Remote Sensing and Earth Sciences*, vol. 18, no. 2, pp. 189-202, 2022.

[17] R. Scovell, N. Gaussiat and M. Mittermaier, "Recent improvements to the quality control of radar data for the OPERA data centre," *In Proc. 36th Conference on Radar Meteorology*, 2013.

[18] R. H. Anasiru, "Analisis Spasial dalam Klasifikasi Lahan Kritis di Kawasan Sub-DAS Langge Gorontalo," 2017.

[19] Z. Rachmah, M. M. Rengkung and V. Lahamendu, "Kesesuaian lahan permukiman di kawasan kaki Gunung Dua Sudara," *Spasial*, vol. 5, no. 1, pp. 118-129, 2018.