



Article

Estimating Scattering Coefficient in a Large and Complex Terrain through Multifactor Association

Peng Zhao ^{1,2} , Yushi Zhang ², Dong Zhu ², Qingliang Li ², Zhensen Wu ^{1,*} , Jinpeng Zhang ², Zhiying Yin ², Huaiyun Peng ² and Longxiong Linghu ¹

¹ School of Physics, Xidian University, Xi'an 717071, China; zhaop1@crirp.ac.cn (P.Z.); lxlinghu@xidian.edu.cn (L.L.)

² National Key Laboratory of Electromagnetic Environment, China Research Institute of Radiowave Propagation, Qingdao 266107, China; zhangyus@crirp.ac.cn (Y.Z.); p21101003@stu.ahu.edu.cn (D.Z.); liql@crirp.ac.cn (Q.L.); zhangjp@crirp.ac.cn (J.Z.); yinzy@crirp.ac.cn (Z.Y.); penghy@crirp.ac.cn (H.P.)

* Correspondence: wuzhs@mail.xidian.edu.cn

Abstract: Complex-terrain clutter presents serious nonuniformity, which has a significant impact on radar target detection, communication, and navigation. The accurate acquisition of the clutter characteristics by measurement or calculation for large and complex terrains has always posed a challenge due to the high costs of the measurement, as well as the intricate and diverse environmental factors. To address this challenge, we proposed a research methodology that leverages the similarity of multidimensional terrain features to infer the clutter characteristics of unmeasured regions, particularly those that are difficult or impossible to measure directly. In order to realize this study object, we constructed a dataset consisting of multidimensional environmental and clutter data to quantitatively characterize the complex environmental information in a vast territory. Within the dataset, we selected two regions with similar terrain characteristics: one region served as the source data for mining and analyzing features, while the other was designated as the target data region for method validation. Through the application of prior-knowledge-based classification and multifactor weight analysis on the dataset, two novel estimation techniques were devised. The first method, designated as PCKRF, blended prior-knowledge classification, weighted K-means clustering, and the random forest (RF) algorithm; and the second method, labeled PCKMW, integrated prior-knowledge classification, weighted K-means clustering, and the minimum weighted distance (MW) approach. In estimating and validating the clutter data from the source region to the target region, the performances of both the PCKRF and PCKMW methods were notably superior to those of the RF, MW, and K-means minimum weighting (KMW). Specifically, the root-mean-squared error (RMSE) was enhanced from a range of 7 dB–10 dB to a range of 4 dB–6 dB, while the determination coefficient (R²) was increased from a range of −1.15–0.09 to a range of 0.25–0.66. The above demonstration illustrates that the current achievements in the clutter estimation methods offer a viable option for accurately recognizing clutter characteristics in complex-terrain environments where comprehensive data collection may be difficult or impossible, with lower human and economic costs.

Keywords: complex terrain; multifactor characterization; prior-knowledge-based classification; multifactor-weighting analysis; weighted similarity; scattering-coefficient estimation



Citation: Zhao, P.; Zhang, Y.; Zhu, D.; Li, Q.; Wu, Z.; Zhang, J.; Yin, Z.; Peng, H.; Linghu, L. Estimating Scattering Coefficient in a Large and Complex Terrain through Multifactor Association. *Remote Sens.* **2024**, *16*, 650. <https://doi.org/10.3390/rs16040650>

Academic Editor: Amin Beiranvand Pour

Received: 30 December 2023

Revised: 6 February 2024

Accepted: 7 February 2024

Published: 9 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to the complexity and diversity of ground objects, complex-terrain clutter presents serious inhomogeneities, which have a significant impact on radar target detection, communication, and navigation [1,2]. In order to grasp the variation patterns between clutter and various parameters, numerous studies on the characteristics of complex-terrain clutter, including several special test programs [3–5], theoretical algorithm-based research [6–8], and empirical modeling [9–11], have been conducted, which are of great significance to the

knowledge of complex-terrain clutter characteristics. However, different from the scattering of single-type ground objects in a uniform terrain, which can be effectively modeled as random rough surfaces [7,12] or multilayer dielectric scattering [13,14], the influence of complex-terrain parameters on scattering is much more intricate and challenging to correlate and represent. The complexity arises from two reasons: the complexity and diversity of the ground objects cannot be characterized well, and the influence of the environmental factors on the scattering coefficient presents complex nonlinear relationships that are difficult to quantify. As a result, the clutter characteristics of complex terrains cannot be accurately obtained when encountering new ground object environments and test conditions.

In recent years, in order to break through the limitation of previous models and test results of only categorizing and qualitatively describing terrain clutter [4,10,15], the generation of mixed-terrain clutter using digital elevation maps (DEMs), land-cover classification data, image data, and other remote sensing data to assist traditional clutter models has garnered serious attention. Hellard et al. [16] simulated ground-based radar clutter using digital topographic-relief and land-cover information. Feng et al. [17] studied a low-grazing-angle-scattering model using the inductive reasoning method by combining the frequency, grazing angle, and feature type. Darrah et al. [18] investigated a model that combined digital terrain elevation data, digital feature analysis data, and the empirical model of the terrain clutter amplitude to generate an accurate site-specific clutter reflectivity map. Kurekin et al. [19] proposed a clutter map generation method based on multichannel images and digital terrain data. Capraro et al. [20] developed techniques for the registration of radar echo signals and the correction of Doppler and spatial misalignment using digital terrain classification and elevation data and applied them to improve the performance of space-time adaptive processing. Based on their research results and methods, more research has been conducted on complex-terrain clutter simulations using digital elevation and traditional models in the last decade. Xie et al. [21] proposed an improved knowledge-assisted algorithm based on the Morchin clutter model. Li et al. [22] proposed a method based on DEMs and digital land feature classification data termed the nonuniform ground clutter simulation method, which selects different scattering models according to distinctive features to improve accuracy. Oreshkina et al. [23] proposed a method to optimize the azimuthal sampling interval of digital topographic maps for radar ground clutter simulations. Luo et al. [24] proposed an optimal searching strategy based on DEMs to suppress ground clutter and achieve a fast and high-probability search for low-altitude targets. Kim et al. [25] proposed a method for synthesizing clutter based on an empirical model by creating and accumulating clutter block by block to predict clutter in unexplored environments and operational conditions.

Despite the introduction of more accurate terrain and feature classification data, previous studies still relied on the traditional empirical model, which is based on terrain and ground object information that is projected onto the corresponding scattering unit to synthesize the clutter map. The traditional model is only a brief representation of the statistical results and variation rules and does not adequately characterize a complex terrain. In addition, owing to the lack of a perfect association between the measured clutter data and geomorphological and environmental elements, these studies did not perform quantitative analyses of the influencing factors in complex terrains, leading to low accuracy.

Given the nonlinear relationship between complex-terrain clutter and environmental factors, it is crucial to adopt appropriate data-mining and analysis methods. The random forest (RF) method [26,27] is a popular machine learning (ML) algorithm. Its internal estimates can be used to measure the importance of the variables for the analysis of their multivariate nonlinear weighted influence. In recent years, the RF method has been widely used in remote sensing image classification, agricultural prediction, regression analysis, and other research [28–32]. In addition, clustering and classification prediction methods based on similar parameters, such as k-means clustering and the k-nearest-neighbors (KNN) algorithm, have also been applied [33,34]. In recent years, several scholars have utilized deep learning networks to predict the characteristics and parameters of sea clutter [35–37],

achieving promising results. However, there has been limited research on the application of similar techniques to ground clutter. In theory, with a sufficiently large dataset covering a comprehensive range of parameters, both deep learning and machine learning can be effective in predictive modeling. Nevertheless, this ideal scenario contrasts with the practical limitations of real-world data acquisition, where it is impossible to achieve full coverage of all possible scenarios through measured data. The intention of this article is to develop a method for estimating similar terrain clutter based on actual measured data. This approach aims to extrapolate the clutter characteristics to unmeasured areas with comparable terrain features, especially in situations in which costs and data availability are constraints.

In response to the above research difficulties, we constructed a multifactor-associated clutter dataset by reconstructing the geometric relationship of the measurement scene, triangulating the terrain, and matching the illumination areas. The dataset contains S-band-airborne-radar-measured clutter data, Advanced Land Observing Satellite (ALOS) DEM data with a resolution of 12.5 m downloaded from the Distributed Dynamic Archive Center (DAAC) of the Alaska Satellite Facility (ASF) [38], land-cover classification data from Global Land 30 [39], soil composition data from the Harmonized World Soil Database (HWSD) [40], normalized difference vegetation index (NDVI) data attributed to the Moderate Resolution Imaging Spectroradiometer (MODIS) [41], and Google image data. This dataset solves the problem of quantitative parameter representation in the study of large-scene clutter characteristics and enables a better analysis and understanding. To overcome the challenges posed by the large dynamic ranges of the scattering coefficients and numerous parameters in complex terrains across extensive areas, prior-knowledge-based classification and multifactor weight coefficient estimation were explored on this dataset. Based on the aforementioned research, two innovative methods were developed by combining the techniques of K-means clustering, RF prediction, and the minimum Euclidean distance. The first method, named PCKRF, integrated pre-classification, weighted K-means clustering, and RF prediction. The second method, known as PCKMW, incorporated pre-classification, weighted K-means clustering, and the minimum weighted Euclidean distance. Both methods aimed to enhance the accuracy and reliability of terrain clutter estimation by leveraging the strengths of different algorithms. The pre-classification step helped to identify distinct terrain types, which were then used as inputs for the subsequent clustering and prediction steps. The weighted K-means clustering took into account the importance of different parameters, allowing for a more precise grouping of similar terrain characteristics. Compared to the measured data and evaluated across multiple metrics, such as the mean absolute error (MAE), root-mean-squared error (RMSE), coefficient of determination (R^2), and mean value error in the pulse dimension (PMVE), the PCKRF and PCKMW methods demonstrate superior accuracy compared to RF prediction, the minimum weighted distance (MW), and the K-means-clustering minimum weighted distance (KMW). Despite requiring further enhancements in scatter plot regression and prediction accuracy, the current methodologies already possess significant advantages and application value in accurately recognizing the clutter characteristics of untested areas using limited data. These approaches also provide considerable support for evaluating the radar performance in complex-terrain environments where comprehensive data collection may be difficult or impossible.

The remainder of this paper is organized as follows. Section 2 provides details on the data used and the overall method. Section 3 shows the associative multiparameter dataset and the results of the weight analysis. Section 4 details similar terrain estimation methods and discusses the results. Section 5 presents the conclusions.

2. Materials and Methods

2.1. Description of Clutter Measurement and Environmental Data

The results of this study were based on complex-terrain clutter data acquired by the airborne S-band radar and remote sensing data of multidimensional environmental elements. The experiments and data are described in detail as follows.

2.1.1. Airborne-Radar Clutter Measurement

The research team conducted airborne ground clutter measurement tests in July 2016 and collected clutter data on various terrain types. Airborne-radar clutter data were acquired using an S-band radar located under an aircraft. The radar was operated with HH polarization, and the radar resolution was 50 m, with an azimuth beam width of $<1.2^\circ$ and a pitch beam width of $<10^\circ$. The flying altitude of the aircraft was approximately 5000 m. Table 1 lists the main parameters of the radar used during the measurement.

Table 1. Radar parameters.

Parameter	Value
Polarization	HH
Frequency band	S
Depression angle	-3°
Resolution	50 m
Airborne height	5000 m
Pitch beam width	$<10^\circ$
Azimuth beam width	$<1.2^\circ$

During the clutter measurement period, an active calibration experiment was performed simultaneously to obtain the system loss of the radar and further derive the scattering coefficients of the measured terrains. Figure 1 shows a geometric diagram of the calibration test and an image map of the clutter measurement area. The aircraft flew along a predetermined route, and the radar beam was set to a fixed illumination direction with a depression angle of 3° and an azimuth angle of 45° to the heading. According to the 3 dB beam width range of the main beam, an illumination strip was formed on the ground parallel to the air route, and active radar calibrators (ARCs) were arranged within the illumination strip. The radar measurement parameters and flight states of the aircrafts were recorded. In addition, the aircraft has an automatic attitude correction device that maintains a stable measurement state.

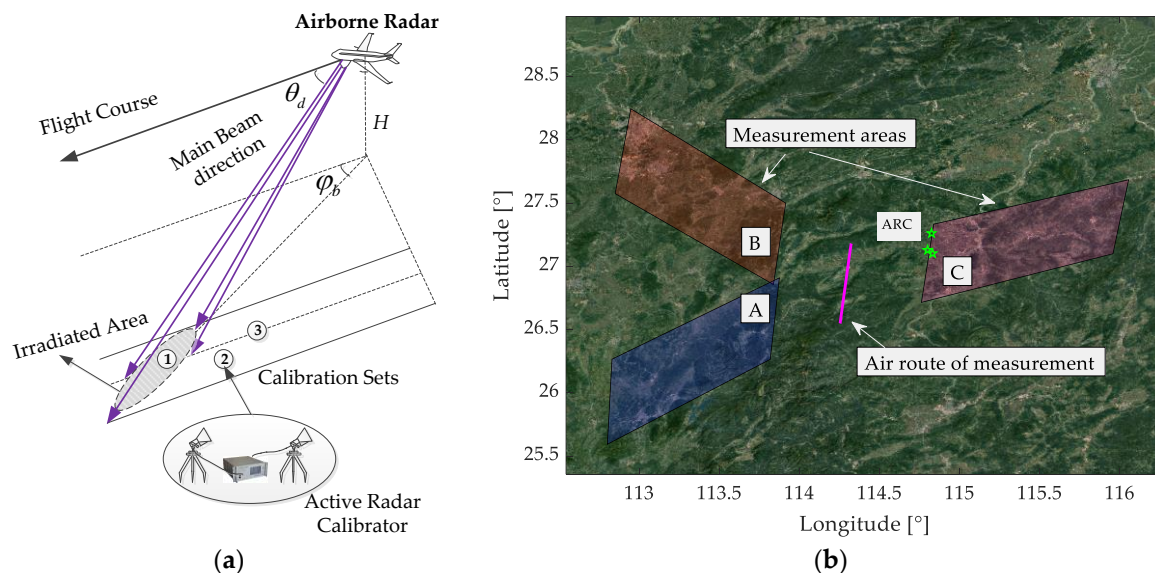


Figure 1. Geometrical schematic diagram of airborne-radar clutter measurement test and image map of test area. (a) Geometrical schematic diagram of active radar calibration and clutter measurement. (b) Air route and image map of measurement areas. The numbers ① to ③ in (a) and the three green stars in (b) represent the locations of ARCs.

The marked areas in Figure 1b are the radar test areas, where areas A and B are mainly the clutter acquisition areas, and area C is the calibration and clutter acquisition area. In

the test, six valid sorption datasets were obtained, where the total echo and calibrator data were approximately 10 TB and 1 GB in size, respectively.

2.1.2. Terrain Characteristics of the Measured Area

Figure 2 presents the geographic feature maps of the test area. Figure 2a shows the distribution image of the land cover sourced from Global Land 30. On this map, red represents anthropogenic built-up areas, green indicates forested areas, yellow denotes bare ground surfaces, and blue represents waterbodies. Figure 2b displays the DEM of the test area, depicting elevation values ranging from 0 to 2000 m above sea level. These maps provide essential information for understanding the terrain, natural features, and human-made structures within the designated test region.

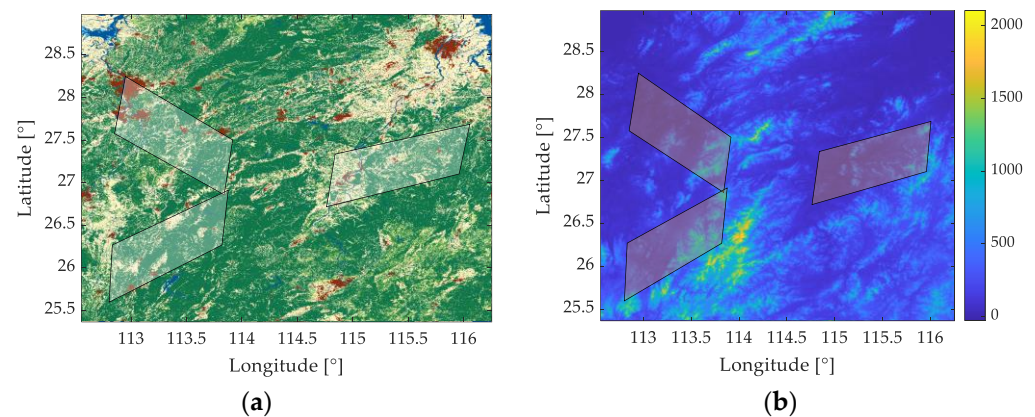


Figure 2. Topographic images of the test area. (a) Distribution image of the land cover, where red represents anthropogenic built-up areas, green indicates forested areas, yellow denotes bare ground surfaces, and blue represents waterbodies. (b) DEM of the test area. The marked areas in (a,b) are the radar test areas.

According to the information in Figure 2, the overall landform is dominated by mountains and hills, of which mountains with elevations of more than 500 m account for 36% of the total area, and the highest elevation is approximately 2000 m. Hills with altitudes of 100–500 m account for 41%, low-relief terrains below 100 m account for 20%, and waters account for 3%. In addition, the test area is dotted with scattered villages, towns, and several medium-sized cities. Overall, the terrain in the test area is very complex and diverse.

2.1.3. Multisource Environmental Data

The ground-cover scattering coefficient is a composite function of multiple factors, and complex dependencies between them and various environmental factors that need consideration exist. In particular, for complex-terrain scattering, it is meaningless to study the clutter without environmental information. Therefore, various environmental factors were introduced to characterize the terrain parameters and establish a correlation with the clutter.

To better analyze the influence of the environmental parameters on the clutter characteristics, this study used multisource remote sensing data, including ALOS DEM data with a 12.5 m resolution downloaded from the DAAC of the ASF, land-cover classification data from Global Land 30, soil composition data from HWSO, NDVI data from the MODIS, and Google image data. In this study, ALOS PALSAR public DEM data were used to represent the surface relief, the Global Land 30 dataset was used to represent the surface-cover type and distribution, and the soil composition type was used to represent the difference in the surface soil composition. Finally, remote sensing image data were used for verification. The relevant data parameters and resolutions are listed in Table 2.

Table 2. Environmental data and resolutions.

Data Source	Type of Representation	Data Resolution (m)
ALOS DEM	Topographic relief	12.5
Global Land 30	Land-cover type	30
Soil composition	Distribution of soil	1000
NDVI of MODIS	Coefficient of vegetation cover	900
Google image	Distribution of ground objects	17

Because the resolution between the land-cover, soil component, and elevation data are inconsistent, to facilitate the data association, the land-cover and soil component data were interpolated, and their resolution was increased to 12.5 m, which was consistent with that of the elevation data. In this article, we used the geographic information processing software Global Mapper version 17 to unify the data precision. Global Mapper primarily uses the method of bilinear interpolation to interpolate images and enhance their resolution. This approach estimates the value of the point to be interpolated by linearly combining the values of the four known data points surrounding it. Bilinear interpolation is a commonly used technique in image processing that can effectively improve the resolution and detail representation of images. The principle of bilinear interpolation is as follows:

Suppose there is a 2×2 grid with four points at coordinates (x_0, y_0) , (x_1, y_0) , (x_0, y_1) , and (x_1, y_1) , corresponding to function values $f(x_0, y_0)$, $f(x_1, y_0)$, $f(x_0, y_1)$, and $f(x_1, y_1)$, respectively. We now want to interpolate the function value $f(x, y)$ at point (x, y) . First, we perform two one-dimensional linear interpolations in the x-direction to obtain the following:

$$f(x, y_0) \approx \frac{(x_1 - x)f(x_0, y_0)}{(x_1 - x_0)} + \frac{(x - x_0)f(x_1, y_0)}{(x_1 - x_0)} \quad (1)$$

$$f(x, y_1) \approx \frac{(x_1 - x)f(x_0, y_1)}{(x_1 - x_0)} + \frac{(x - x_0)f(x_1, y_1)}{(x_1 - x_0)} \quad (2)$$

Then, we perform a one-dimensional linear interpolation in the y-direction to obtain the following:

$$f(x, y) \approx \frac{(y_1 - y)f(x, y_0)}{(y_1 - y_0)} + \frac{(y - y_0)f(x, y_1)}{(y_1 - y_0)} \quad (3)$$

By substituting the expressions (1) and (2) into (3), the complete expression for bilinear interpolation based on the four neighboring points can be derived.

2.2. Methods

The proposed method comprises three main parts. The first part constructs a clutter dataset with multifeature associations, including data-processing and feature representation association methods. The second part analyzes the weight coefficients of multiple features, and the third part estimates clutter similar to the terrain. Each method is explained separately. Figure 3 shows the schematic diagram of the method flow.

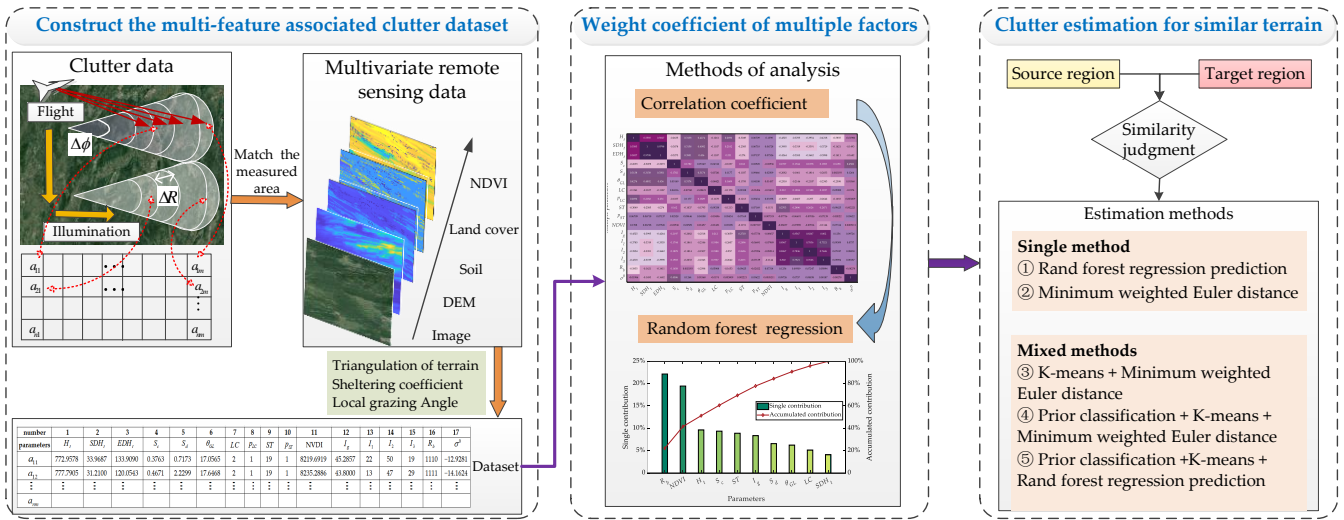


Figure 3. Schematic visualization of the method.

2.2.1. Processing Methods for Clutter Data

The acquired data were processed and classified into calibration and clutter data. Figure 4 shows the process flow of classifying calibration and clutter data.

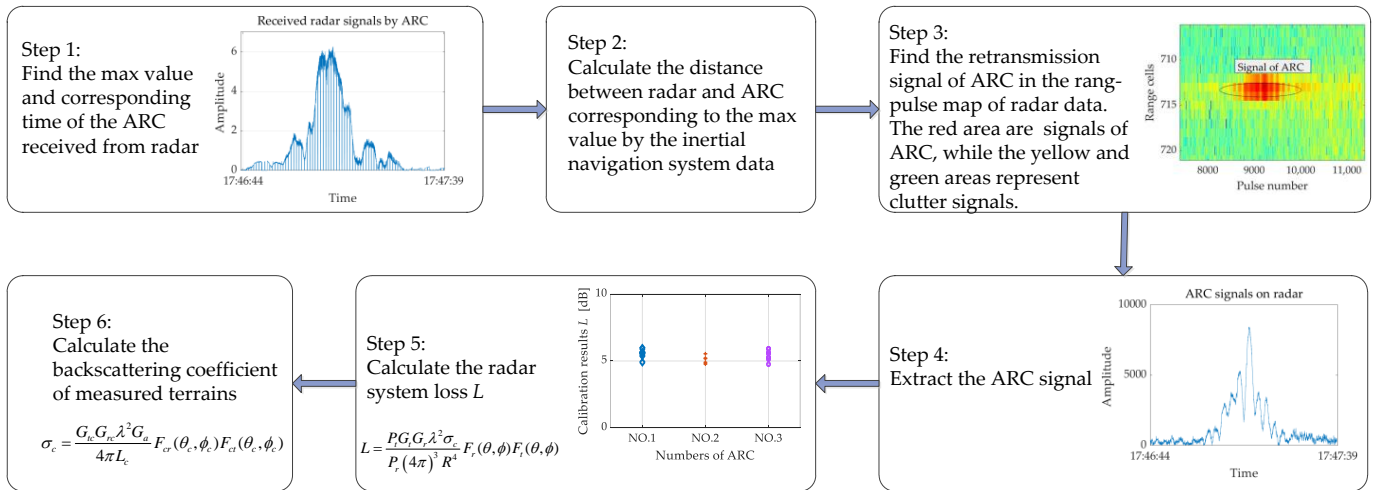


Figure 4. Schematic visualization of the clutter data processing.

The ultimate goal of radar calibration data processing is to obtain a system constant such that the radar echo data can be calibrated. According to the radar equation, the formula for the system constant can be derived as follows:

$$L = \frac{P_t G_t G_r \lambda^2 \sigma_c}{P_r (4\pi)^3 R^4} F_r(\theta_d, \Delta\phi) F_t(\theta_d, \Delta\phi) \quad (4)$$

where P_t is the transmitted power; G_t is the gain of the transmitting antenna of the radar; $F_t(\theta_d, \Delta\phi)$ is the directivity factor of the transmitting antenna; G_r is the gain of the radar-receiving antenna; $F_r(\theta_d, \Delta\phi)$ is the directivity factor of the receiving antenna; λ is the wavelength of the electromagnetic wave transmitted by the radar; θ_d is the depression angle; $\Delta\phi$ is the offset azimuth angle; R is the slant distance from the radar to the calibrator, and the atmospheric refraction effect should be considered when calculating the slant range; σ_c is the radar cross section of the active radar calibrator (ARC).

It is necessary to synchronize the times of the radar and ARC during radar calibration. When the main beam of the airborne radar is swept across the active calibrator along the illumination strip shown in Figure 1, the maximum signal received by the calibrator and the maximum signal transmitted by the radar-receiving calibrator appear. According to the maximum position and time information combined with the antenna pattern factor, aircraft flight inertial data, and other information, the system constant of the airborne radar can be calculated using Equation (4).

When the radar system loss (L) is calculated from the external calibration data according to the radar equation, the scattering coefficients of the radar-receiving ground and sea clutter are as follows:

$$\sigma^0 = \frac{P_r (4\pi)^3 R^4 L}{P_t G_t G_r \lambda^2 F_r(\theta_d, \Delta\phi) F_t(\theta_d, \Delta\phi) A} \quad (5)$$

where $A = \frac{R\phi_{az}\Delta r}{(2\cos\theta_g)}$ is the irradiated area, ϕ_{az} is the azimuth beam width of the radar main beam, Δr is the radar resolution, and θ_g is the grazing angle, which is calculated as the angle between the Earth's tangent plane and the radar beam, as shown in Figure 5.

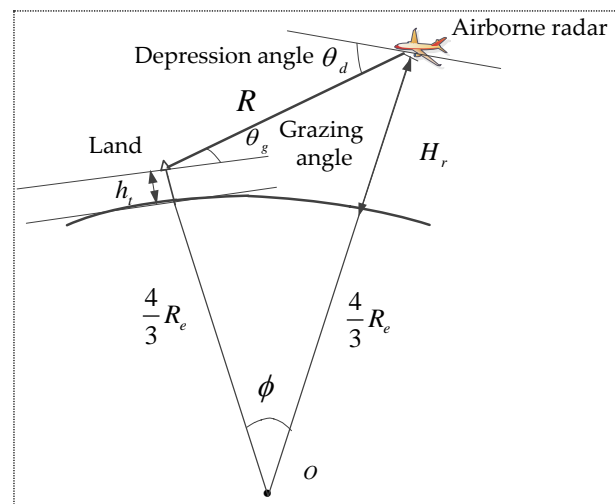


Figure 5. Geometric relationship between airborne radar and ground.

Figure 6 shows the calculated results for the scattering coefficients measured using the S-band environmental test radar. Figure 6a shows a 2D plot of the scattering coefficient, where the abscissa represents the flight distance of the aircraft, and the ordinate represents the test distance of the radar. The test area contains landforms, such as mountains, cities, waterbodies, arable land, and forests. The scattering coefficient fluctuates significantly due to the topography and ground feature distribution. Figure 6b shows the variation in the corresponding mean backscattering coefficient with the grazing angle. Within a grazing angle from 1° to 3° , the mean value of the scattering coefficient under this terrain ranges from -28 to -18 dB, with a relatively strong fluctuation, reflecting the complexity of the ground-cover scattering.

To prove the validity of the test data and calibration results, data on the low-grazing-angle scattering about the HH polarization of the S-band from [4] were introduced for verification. In [4], the depression angle was used to represent the angle size. Because of the fixed-point test adopted in [4], the height of the radar was only tens of meters, and the influence of the Earth's curvature could be ignored; thus, the depression angle in [4] can be approximated as the grazing angle. Figure 7 presents the results of the comparison. Our data are presented as means \pm standard deviations, and the data from [4] are represented by mean values. The mean values of the measured results of different ground objects and those obtained from [4] were of the same order of magnitude. The scattering coefficient of the cities was the strongest, followed by mountains, forests, arable

land, and waterbodies. Results of the comparison show that the data and processing methods tested in the experiment were effective.

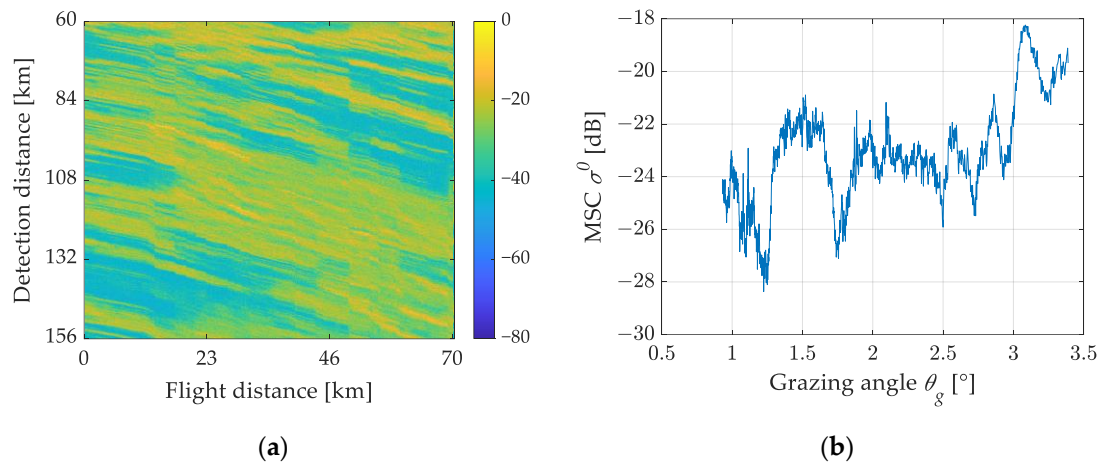


Figure 6. Calculation results of the radar backscattering coefficient. (a) Distance 2D image of backscattering coefficient. (b) Mean backscattering coefficient with grazing angle.

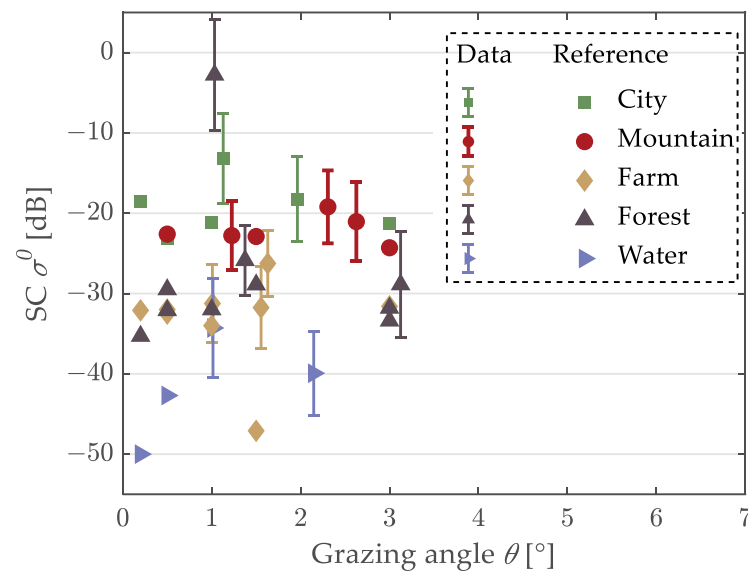


Figure 7. Comparison of scattering coefficients between measured and literature data.

2.2.2. Methods for Processing Environmental Data

The result of the clutter data is the average effect of multiple scatterers in a scattering cell. Therefore, the first step in the association between the environmental elements and clutter data is to determine the ground object information in the processing unit with each pulse and resolution cell as the scattering coefficient, especially for the shadowing information, which is related to the depth of the terrain relief in the radar view direction and illumination area. Therefore, the occlusion and local-grazing-angle information should be calculated in a pulse unit. According to the longitude and latitude information corresponding to the effective range gate in the pulse, the ground feature information contained in each resolution unit in the pulse should be determined from the ground feature information representation dataset. In summary, this study adopted the following steps to extract environmental information:

1. According to the global positioning system (GPS), (Lon_m, Lat_m, H_{rm}) , and radar-beam-pointing (φ_i, θ_{di}) information in the inertial data of the aircraft at the time T_m , and

combined with the width of the radar main beam (ϕ_{az}), the radar beam irradiation area range from R_{n1} to R_{n2} was calculated;

2. The GPS information of four vertices in the illuminated areas, $(Lon_{n1, -\frac{\phi_{az}}{2}}, Lat_{n1, -\frac{\phi_{az}}{2}})$, $(Lon_{n1, \frac{\phi_{az}}{2}}, Lat_{n1, \frac{\phi_{az}}{2}})$, $(Lon_{n2, -\frac{\phi_{az}}{2}}, Lat_{n2, -\frac{\phi_{az}}{2}})$, and $(Lon_{n2, \frac{\phi_{az}}{2}}, Lat_{n2, \frac{\phi_{az}}{2}})$, were marked;
3. The inpolygon function was used to extract the scattering points in the four vertices containing areas of multiple ground object data, such as DEM, NDVI, and image data;
4. The distance between the extracted points and the radar was calculated, and the extracted points were assigned to each scattering cell according to the length of the distance cell.

The inpolygon function uses the ray method to determine whether a point is within a polygon. Through the above steps, the extraction of the object parameter information in the radar beam irradiation range was completed.

The terrain relief is an important factor affecting ground clutter, and the shading and local grazing angle caused by the interaction between the radar viewing direction and terrain are key parameters. To calculate these parameters, it was necessary to triangulate the terrain, calculate the normal vector of the surface element, and calculate the shading and local-grazing-angle information. The steps of the terrain surface element subdivision were as follows:

1. Coordinate conversion was performed on the elevation data extracted by the radar beam irradiation, and the latitude and longitude heights were converted to Earth's XYZ coordinate system; detailed coordinate conversion methods are available in the literature [20] and will not be repeated here;
2. The transformed XYZ coordinates were decomposed into triangular surface elements to obtain the vertex and inner coordinates of each triangle. We assumed that the three vertices of the number i triangles were $A_i = (x_{i1}, y_{i1}, z_{i1})$, $B_i = (x_{i2}, y_{i2}, z_{i2})$, and $C_i = (x_{i3}, y_{i3}, z_{i3})$, and the coordinates of the reference point were the center of gravity or the inner heart, which can be calculated according to the following formula:

$$D_i = \frac{A_i + B_i + C_i}{3} = \begin{cases} \frac{x_{i1} + x_{i2} + x_{i3}}{3} \\ \frac{y_{i1} + y_{i2} + y_{i3}}{3} \\ \frac{z_{i1} + z_{i2} + z_{i3}}{3} \end{cases} = (x_{iD}, y_{iD}, z_{iD}) \quad (6)$$

3. To calculate the tangent plane normal vector, the following formula was used:

$$\vec{u} = (x_i, y_i, z_i) = \frac{\vec{AB} \times \vec{AC}}{|\vec{AB} \times \vec{AC}|} \quad (7)$$

Through this process, the GPS position and normal vector of the interior point of the subdivision surface element in the illuminated area were obtained. The local-grazing-angle and shadowing information can be further calculated by combining the radar inertial navigation information. A schematic diagram of the radar beam and subdivision terrain in the azimuth and pitch directions is shown in Figure 8. Because each scattered cell contains multiple DEM data points, to improve the accuracy of the local erasing angle and occlusion judgment, we divided the beam into 20 parts to save calculation time and data.

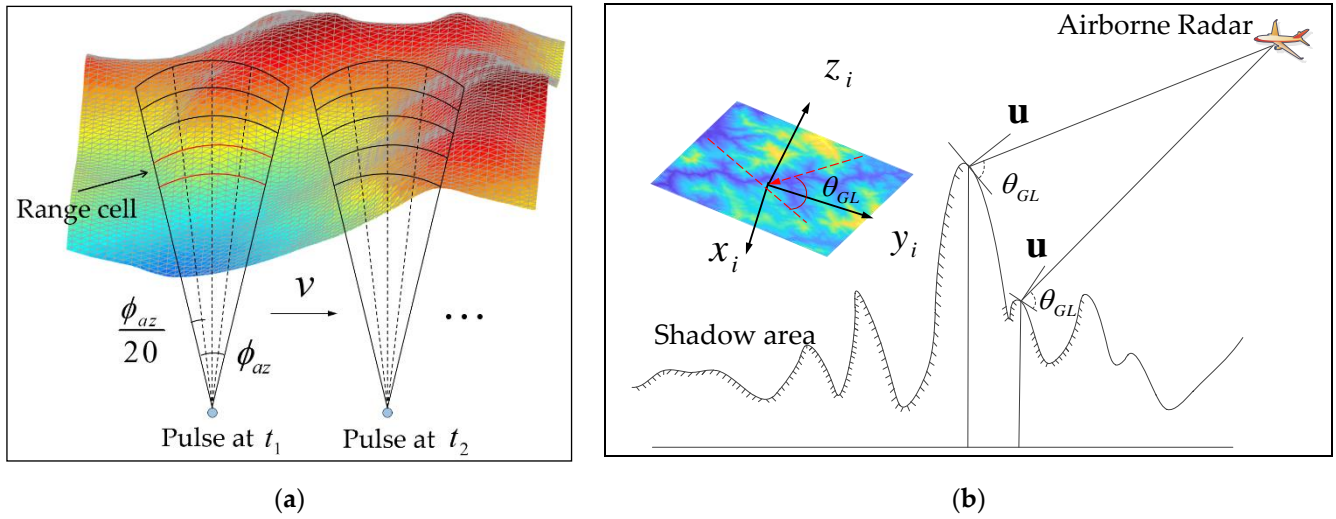


Figure 8. Geometric diagram of radar beam and terrain. (a) Azimuth view. (b) Pitch to the view.

The local grazing angle (θ_{GL}) is expressed as the angle between the radar illumination and normal direction of the tangent plane of a point on the ground; the radar GPS information (Lon_m, Lat_m, H_{rm}) is converted into XYZ coordinate information (x_m, y_m, z_m), and the elevation angle from the scattering points in the i subdivision plane element of the radar is as follows:

$$\theta_{d,i} = \arctan \frac{z_{iD} - z_m}{\sqrt{(x_{iD} - x_m)^2 + (y_{iD} - y_m)^2}} \quad (8)$$

The line-of-sight vector from the radar to the center of the subdivision surface element is $\vec{v} = (x_{iD} - x_m, y_{iD} - y_m, z_{iD} - z_m)$, and the local grazing angle is calculated as follows:

$$\theta_{GL,i} = \frac{\pi}{2} - \arccos \frac{\vec{u} \cdot \vec{v}}{|\vec{u}| |\vec{v}|} \quad (9)$$

Shadowing judgment: The scattering points in the sub azimuth beam are found and sorted from small to large according to the slant distance. Whether a scattering point is shaded by a point in front of the azimuth is determined based on the following principles:

1. The elevation angle between this point and the radar is higher than the minimum elevation angle of all the preceding points (that is, $\theta_{d,i} > \max\{\theta_{d,(i-1)}\}$). This was judged to be obscured by the line of sight, which corresponds to an area with high fluctuations;
2. If the local grazing angle of the point is ≤ 0 (that is, $\theta_{GL,i} \leq 0$), it was judged as self-shading, corresponding to an area with low fluctuations.

If one of the above two conditions is satisfied, the point is judged as an occlusion. Once we obtain the occlusion, we calculate the depth of the occlusion to represent how severely the point is occluded:

$$S_{d,i} = \begin{cases} \theta_{d,i} - \min\{\theta_{d,(i-1)}\} & \text{Sheltered line of sight} \\ -\theta_{GL,i} & \text{Self sheltered} \end{cases} \quad (10)$$

2.2.3. Construction Method for Clutter Dataset Based on Multifeature Representation

After the area matching and shadowing judgment, the corresponding results for multiple-ground-feature information and one scattering coefficient in each scattering unit were obtained. For example, if there are 3×36 elevation points in a single scattering unit,

only one scattering-coefficient value is present. Therefore, the parameters of the ground objects in a single scattering unit can be simplified as follows:

1. The mean elevation (H_t), standard deviation of elevation (SDH_t), and extreme deviation of elevation (EDH_t) were calculated as the variation in the elevation of the scattering unit to characterize the topographic relief;
2. The shadowing proportion of all scattering points in a scattering unit is considered to be the shadowing coefficient (S_c) of the scattering unit, and the shadowing depth (S_d) is characterized by the mean value. These two parameters characterize the shading intensity of the scattering unit;
3. For the unsheltered scattering points, the local grazing angle (θ_{GL}) was divided into ten intervals according to the size, and the number of local grazing angles in each interval was counted. The local grazing angle with the highest number of occurrences was used to represent the local grazing angle of the scattering unit;
4. Subsequent data were processed in the same manner. For example, for land-cover information, the types with the highest proportion of land cover (LC) of the unsheltered scatterers were counted, as well as their percentages (p_{LC}), and they were used to characterize the land cover in the scattering unit. The same method was used to obtain the soil type (ST), soil-type percentage (p_{ST}), normalized vegetation index (NDVI), image grayscale (I_g), and RGB channels of the image data (I_1 , I_2 , and I_3). All the above information was used to characterize the distribution of the ground objects;
5. The range bin number (R_b) was used to represent the relative distance between the radar and scattering unit;
6. This information is relevant for the backscattering coefficient (σ^0) of the scattering unit.

2.2.4. Multifactor Influence Weight Analysis Method

We comprehensively characterized the relationships among the scattering coefficients, environmental parameters, and radar parameters in multiple dimensions. Evidently, the effects of the above parameters on the scattering coefficient differ, and a strong coupling relationship may exist between the parameters. Therefore, we first used the Pearson correlation coefficient matrix to select the parameters before the weight analysis. A random forest regression analysis was further used to perform a nonlinear regression analysis on the above dataset, and the contribution weight of each parameter to the scattering coefficient was obtained.

The Pearson correlation coefficient formula is as follows:

$$\rho = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (11)$$

where the value of the correlation coefficient (ρ) is within the range of $[-1, 1]$, and the higher the absolute value, the stronger the correlation. A value of 0 indicates no correlation, and 1 or -1 indicates that a linear representation can be used. A value greater than 0 indicates a positive correlation and a value < 0 indicates a negative correlation.

Figure 9 shows the correlation coefficient heatmap matrix. A certain degree of parameter optimization was performed according to the values of the correlation coefficients. Parameters with strong correlations among themselves were removed, and parameters with very low correlations with the scattering coefficients were also removed.

The p_{LC} and p_{ST} contributed almost zero to the scattering coefficient; therefore, they were removed. The SDH_t was strongly correlated with the EDH_t , which, in the I_g , was strongly correlated with I_1 , I_2 , and I_3 . Therefore, only the SDH_t and I_g were retained. Based on this dimension reduction process, the dimensions of the parameters were reduced from the original 17 columns to 11 columns.

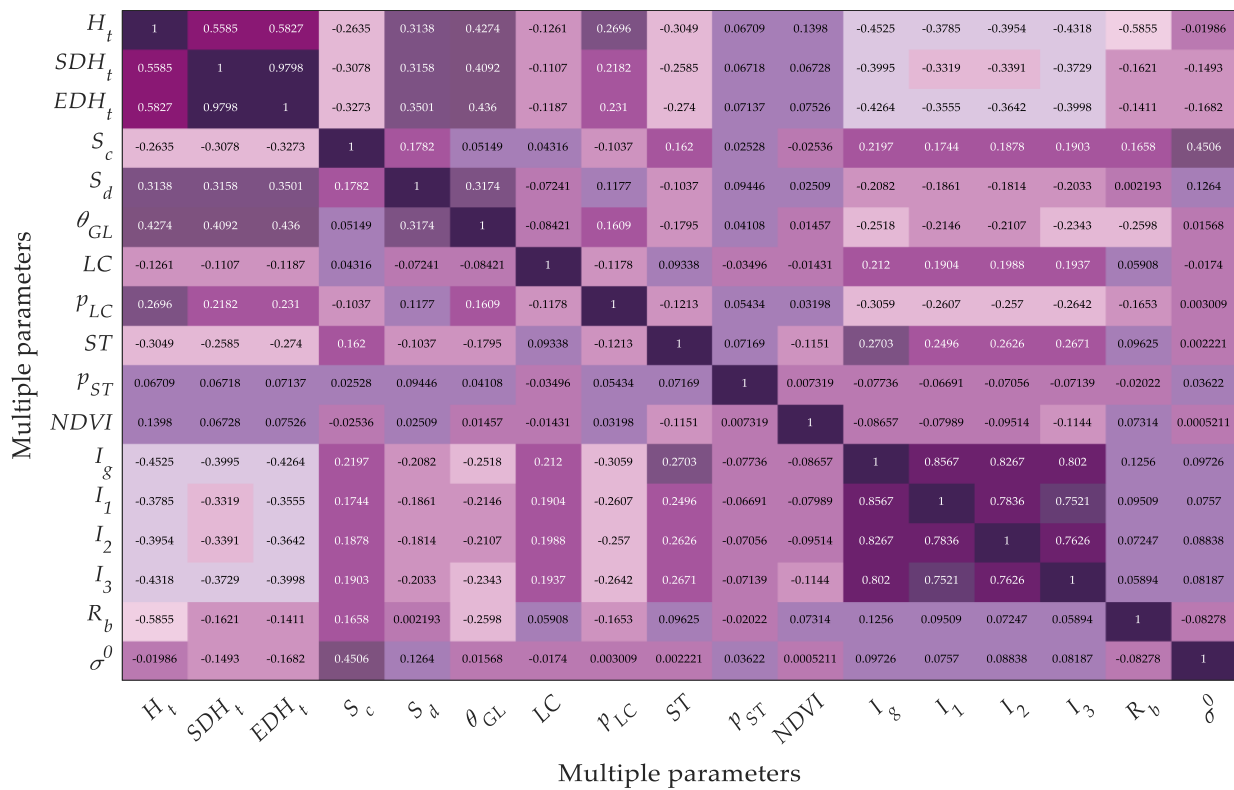


Figure 9. Correlation matrix heatmap.

The influence of multidimensional parameters on the scattering coefficient is complex and nonlinear. The above correlation coefficient analysis is a linear analysis, which does not meet the requirements of weight analysis in a complex terrain. Therefore, based on the above dataset, we used a variety of ML methods to perform the nonlinear regression analysis, such as the support vector machine (SVM), k-nearest-neighbor (K-N), RF, and multilayer perceptron (MLP) methods. The model error was evaluated using the root-mean-squared error (RMSE). These results are presented in Table 3.

Table 3. Error analysis of machine learning models.

Machine Learning Models	SVM	K-N	RF	MLP
RMSE	5.4752	5.5203	4.9302	5.7431

The RF model showed the best prediction, and the out-of-bag error parameters in the RF method could be used to measure the contribution of the parameter. Therefore, this study used the RF method for the weight analysis.

Various parameter choices are available for the random forest method, such as the numbers of trees and leaves. Considering the calculation time and prediction accuracy, we performed hyperparameter calculations to find the optimal combination of parameters. The number of trees was preset to 1–1000, and the leaf parameters were 5, 10, 20, 50, 100, 200, and 500. The above combinations were cycled, and a parameter combination with both calculation efficiency and accuracy was obtained according to the RMSE analysis. Figure 10 shows the prediction error analysis diagram for this combination. The RMSE is the lowest when the number of leaves is five, unchanged when the number of trees reaches 400, and maintained at approximately 4.9. Therefore, 400 trees and five leaves were chosen as the parameters of the random forest.

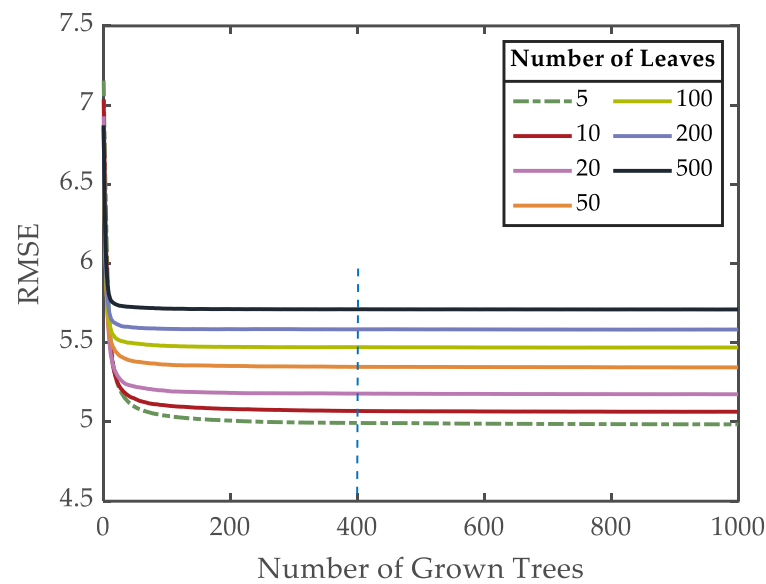


Figure 10. Hyperparameter error analysis for the random forest (RF) model.

During the prediction process by the RF method, we employed 10-fold cross-validation, taking the average accuracy from the 10 iterations as an estimate of the algorithm's accuracy. Table 4 presents the cross-validation results obtained during one of these prediction processes. Notably, the results are quite consistent, indicating that the model exhibits good stability.

Table 4. Error analysis of cross-validation for RF.

Verification Fold Number	1	2	3	4	5	6	7	8	9	10
RMSE	4.6729	4.6907	4.6465	4.6142	4.7826	4.6860	4.6670	4.5026	4.6835	4.7173

Figure 11 shows a comparison between the RF-predicted results of the above parameters and measured data in the training and test sets. The RMSE of the training set was 3, with a coefficient of determination (R^2) of 0.8782, and the RMSE of the test set was 4.6663, with a determination coefficient (R^2) of 0.7056. The comparison shows that the RF model has better convergence and accuracy for the dataset used in this study.

Using the RF model, the out-of-bag parameter error (ε_i) was obtained as the contribution of each parameter to the scattering coefficient. In view of the complexity of the ground-cover scattering, the out-of-bag parameter error (ε_i) and correlation coefficient (ρ_i) may be missing to some extent. Moreover, there are significant differences in the results between the ρ_i and ε_i . Therefore, by combining relevant methods of comprehensive weight calculation [42,43], the comprehensive weight coefficients of the ρ_i and ε_i were fused to characterize the multifactor influence in this study:

$$Q_i = \frac{\sqrt{\rho_i \varepsilon_i}}{\sum \sqrt{\rho_i \varepsilon_i}} \quad (12)$$

From the examples encountered in this study, it was evident that the results obtained by utilizing the Q_i exhibited higher accuracies compared to utilizing the ρ_i or ε_i individually. Certainly, the formulation of this approach involves a significant degree of subjective judgment, and it could be continuously refined based on practical considerations. Apparently, this is also an area worthy of further in-depth study.

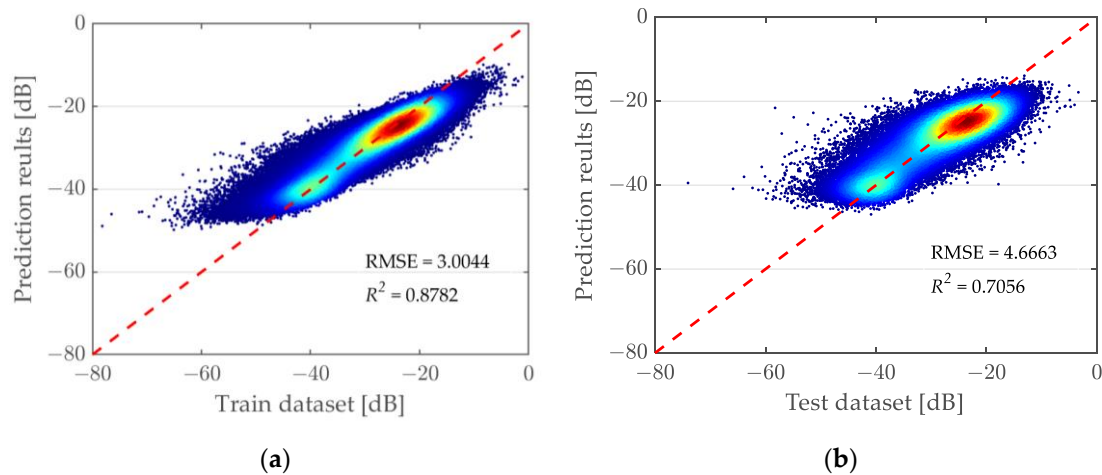


Figure 11. Comparison between RF prediction results and measured data. (a) Training set. (b) Test set. The red dashed line represents the regression line, while the data is presented in the form of scatter plot density, with red indicating areas of high scatter density and blue indicating areas of low scatter density.

2.2.5. Clutter Estimation Methods

The clutter estimation method used in this study was based on measured data and regional similarity. Measured data in region B, shown in Figure 1b, were used as source data, and region A was used as the estimated target region. In other words, the scattering unit with the closest parameters to each scattering unit in region A was found in region B, and the scattering-coefficient value of the scattering unit was assigned to the scattering unit in region A. Finally, the accuracy of the method was verified by comparing it with the measured results in region A. Figure 11 shows a flowchart of the estimation method.

Because the clutter estimation method presented in this article is based on similar terrain features, which should be exhibited in both the category and data ranges, the overall similarity of the terrain features was first judged using the Bhattacharyya coefficient. During the development of the estimation method, we explored several approaches. One approach involved training an RF model on the source data and using it to predict clutter for the target area by inputting environmental parameters. This is referred to as the RF prediction method in Figure 12. Another approach involved calculating the weight coefficients of the influencing factors and using a multiparameter weighted nearest-neighbor principle to estimate the clutter data of the target area from the source data. This is known as the MW prediction method in Figure 12. Upon comparing the results of these initial methods with the measured data, it was found that their performance was not satisfactory. As a result, the K-means-clustering method was introduced to form the KMW prediction method with the MW method, as shown in Figure 12. Although the result showed some improvement, it still did not meet the requirements. Building upon these methods, we implemented a priori-knowledge-based pre-classification. Weighted clustering was then applied to the classified sub-datasets, and the weight coefficient calculation method was improved. This ultimately led to the development of the PCKRF and PCKMW prediction methods, which showed significant improvements compared to the previous three methods. Overall, the figure illustrates the process of the iterative improvement in and increasing prediction accuracy of the clutter estimation method presented in this article. The following sections provide a detailed description of the specific methods employed.

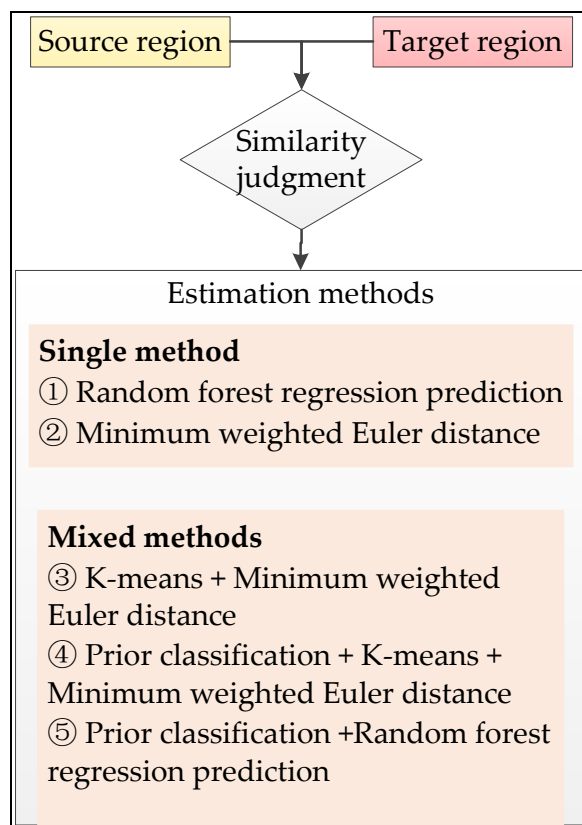


Figure 12. Schematic visualization of the estimation method.

- Regional similarity judgment method

Before performing a data-based estimation, it is necessary to verify the closeness of the source dataset to the target dataset using multiple parameters. The Bhattacharyya coefficient (BC) method was used to characterize the overall approximation:

$$BC_i(p_i, q_i) = \sum \sqrt{p_i(x)q_i(x)} \quad (13)$$

where p and q are the probability distributions of the variables at the same position on the histogram. The BC range was 0–1, where 0 was completely different and 1 was completely identical. The scattering of land cover is significantly affected by the terrain relief and ground-cover distribution, and the BC method can better measure the overall similarity of different regions based on the proximity of the probability distribution value of each parameter;

- Prior-knowledge-based classification

Terrain scattering has a significantly wide dynamic range and nonuniformity, and the weight of the influence on the ground-cover scattering changes significantly under different distribution states of the land cover. To improve the estimation accuracy, ground objects are pre-classified based on prior knowledge, and ground objects with similar scattering mechanisms are gathered together to form multiple datasets. For example, based on the prior knowledge of the shading coefficient, elevation fluctuation degree, image map, and other prior knowledge, source data are classified into six categories, as follows: the ordinary shading area, mountainous shading area, mountainous illumination area, woodland, town, and other types. On this basis, RF classification training was performed, the target area was classified, and the calculated data labels corresponding to the target area and source data were obtained;

- The weighted-k-means-clustering method

The k-means-clustering method is commonly used for data processing and analysis. Based on the traditional k-means method, this paper introduces the weight coefficient obtained previously and then clusters the dataset. The value of the clustering parameter (k) is determined according to the elbow method, and the value of the k is 12;

- Estimation methods

In the study of the scattering-coefficient estimation of similar features in different regions, we tried the following methods:

1. The RF estimation method. RF regression prediction was performed on the source data area, and the environmental parameter values of the target area were introduced into the trained RF model to obtain the scattering coefficients;
2. The minimum weighted distance (MW) method. This method utilizes the comprehensive weighted coefficient (Q) to calculate the multiparameter weighted Euclidean distance between the scattering units in the target region and source data region. It identifies the scattering unit in the source data region with the multiparameter weighted minimum distance and assigns the scattering coefficient of this unit to the target region. The overall scattering coefficient was obtained by traversing all the scattering elements in the target area;
3. The K-means-clustering minimum weighted distance (KMW) method. The weighted-k-means-clustering method was used to semi-supervise all the data in the source data region, and the comprehensive weighted coefficient (Q) was used as the weighting coefficient. The scattering coefficient was added to the comprehensive weighted coefficient, and the weight of the scattering coefficient was set to 30%. The number of clusters was determined to be 12 according to the elbow method. The weighted minimum distance method was then used to determine the clusters to which the scattering units in the target region belonged. Subsequently, the minimum weighted distance between the scattering units in the target area and source data in each cluster was calculated, and the scattering-coefficient value of the scattering units in the source data region was assigned to the target region;
4. The prior-knowledge classification + weighted k-means clustering + RF prediction (PCKRF) method. According to the above classification method and the k-means-clustering method, data of the source and target areas were divided into 6×12 clustering datasets. Subsequently, the RF regression prediction was trained on the source data region, and the weight coefficients of each cluster were obtained. The environmental data of the target area were input into the trained RF model to obtain the scattering coefficient of the target area;
5. The prior-knowledge classification + weighted k-means clustering + minimum weighted distance (PCKMW) method. According to the classification of the influence weight coefficient and the 6×12 clustering results obtained by the PCKRF method, all scattering units in the target area were cyclically calculated, and the multiparameter weighted distances between each scattering unit and the scattering units in the source area of the corresponding class were calculated. A scattering unit with the minimum weighted distance was determined, the scattering coefficient was replaced, and the overall scattering coefficient of the target area was obtained.

3. Results

3.1. Associative Multiparameter Clutter Datasets

Through the data processing and construction methods of the clutter dataset with multiparameter associations described in the previous section, the clutter dataset associated with multiple parameters based on each pulse and each resolution unit can be obtained by processing the data of the test area. Figure 13 shows the constructed clutter dataset associated with multiple elements, where columns 1–6 represent the terrain relief and shadowing effects, columns 7–15 represent the surface-cover type and distribution, column

16 represents the radar test range, and column 17 represents the corresponding scattering-coefficient value of the scattering unit. a_{nm} is a single scattering unit, where n corresponds to the number of range gates and m corresponds to the number of pulses.

Characterizing terrain relief and shading effects						Characterizing the type and distribution of land cover											
number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
parameters	H_t	SDH_t	EDH_t	S_c	S_d	θ_{gz}	LC	p_{lc}	ST	p_{sr}	NVDI	I_g	I_1	I_2	I_3	R_b	σ^0
a_{11}	772.9578	33.9687	133.9090	0.3763	0.7173	17.0565	2	1	19	1	8219.6919	45.2857	22	50	19	1110	-12.9281
a_{12}	777.7905	31.2100	120.0543	0.4671	2.2299	17.6468	2	1	19	1	8235.2886	43.8000	13	47	29	1111	-14.1624
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a_{nm}																	

Scattering units, where n is range bin numbers and m is the pulse numbers

Figure 13. Illustration of the clutter dataset represented by multifeatured association.

Figure 14 shows range–pulse 2D images of the typical parameters in test area B. The cell range was 1200–3120, and the number of pulses was 360. These three 2D images correspond to the 2D image of the clutter, as shown in Figure 6a. If this corresponds to data in the list form in Figure 13, the amount of data in region B reaches $691,200 \times 17$. The amount of data in test area A is consistent.

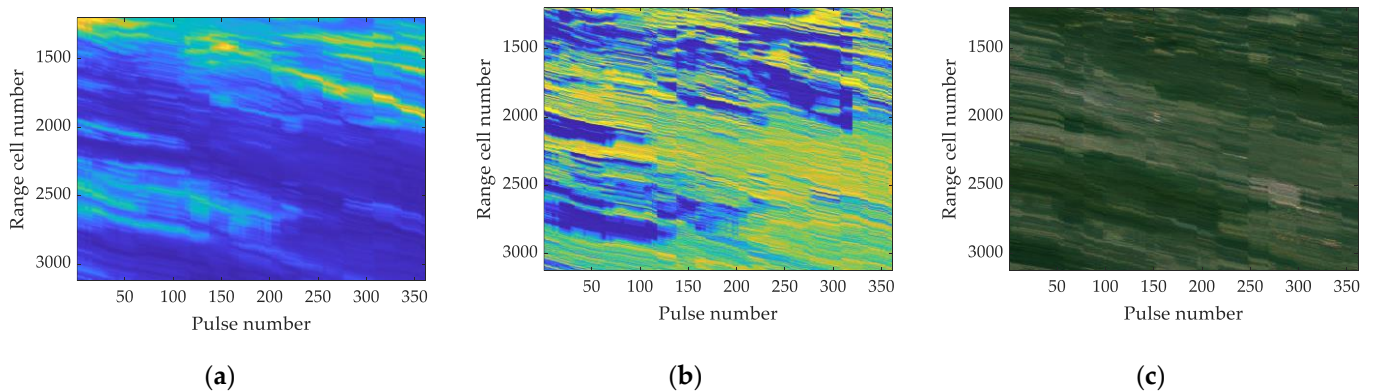


Figure 14. Range–pulse 2D images of typical parameters in measured region B. (a) Mean value of DEM (H_t). From blue to yellow represents increasing height, with blue indicating lower heights and yellow indicating higher heights. (b) Shadowing coefficient (S_c). From yellow to blue, it indicates a progressive increase in the degree of shading. (c) Google image data.

In this study, region B was used as the source data region, which was primarily used for data mining, analyzing the influence weights under complex terrain, and training the RF model. Region A was used as the target region, and the scattering coefficient of region B was extrapolated to region A by the correlation between the parameters, which was verified using measured data for region A.

3.2. Results of Influence Weight Coefficient

As shown in Figure 8, after the correlation coefficient analysis based on Equation (8), the dimensions of the dataset were reduced to 11 columns, and the environmental parameters were only 10. RF regression analysis was performed on the source dataset after dimensionality reduction to determine the contribution of each parameter to the scattering coefficient. Figure 15 shows the contribution weights of multiple elements, in which the range cell number and NVDI contributed the most, followed by the H_t and S_c . The result is quite different from the result of the correlation coefficient matrix shown in Figure 8.

Considering the complexity of the ground-cover scattering, these two weight measurement methods may be missing to some extent. Therefore, the comprehensive weight coefficients of the ρ_i and ε_i were fused to characterize the multifactor influence using Equation (9).

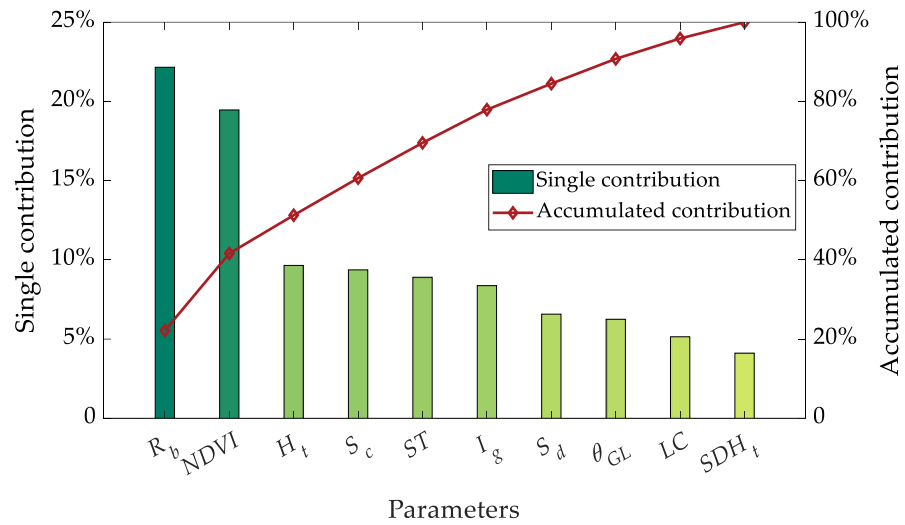


Figure 15. Results of multiple-feature weighting.

Table 5 shows the results of the multiparameter weight calculation, where ρ_i represents the correlation coefficient weight and ε_i represents the RF contribution weight. Q_i is the fusion weight obtained from Equation (9). It can be observed that there is a significant difference in the weight representation between the ρ_i and ε_i , and the Q_i reduced this discrepancy. The last column of Table 4 presents a comparison of a certain example’s results in terms of the RMSE, demonstrating that the Q_i performs relatively better. Thus, the Q_i was used to measure the influence of multiple factors on the scattering coefficients in this study.

Table 5. Results of multiple-feature weighting.

Parameters	H_t	SDH_t	S_c	S_d	θ_{GL}	LC	ST	NDVI	I_g	R_b	RMSE
ρ_i (%)	2.11	15.88	47.90	13.44	1.67	1.85	0.24	0.06	8.05	8.80	11.7357
ε_i (%)	9.65	4.11	9.38	6.57	6.25	5.15	8.90	19.46	8.37	22.16	11.1538
Q_i (%)	6.08	10.88	28.57	12.66	4.35	4.16	1.97	1.46	11.06	18.81	10.8299

Terrain scattering has a significantly wide dynamic range and nonuniformity, and the weight of the influence on the ground-cover scattering changes significantly under different land-cover distribution states. It is difficult to apply a set of weight coefficients to all ground-cover types. Therefore, this is a more reliable and accurate method for classifying ground objects into different types and calculating the influence weights of the parameters on the scattering coefficients for each type.

As stated earlier, according to prior knowledge of the shading coefficient, the source data were classified into six types. The influences of the parameter weights on the scattering coefficients for each type were calculated. Figure 16 shows the label image of the classification results and the multiparameter weight coefficient results for each type. In Figure 16a, M2, M2, M3, M4, M5, and M1 represent the ordinary shaded, mountain shelter, mountain illumination, woodland, town, and other areas, respectively. Because the label results of the classification were approximate to the land-cover parameter, they were dropped from the weight analysis of the classified regions. Therefore, there are only nine parameters in Figure 16b. A comparison of the weight coefficients in different regions shows that the influence of each parameter on the scattering coefficient changed significantly with changes

in the types of regional ground features. Table 6 presents the influence weight coefficients for each type, which were used in the scattering-coefficient estimation that is described later.

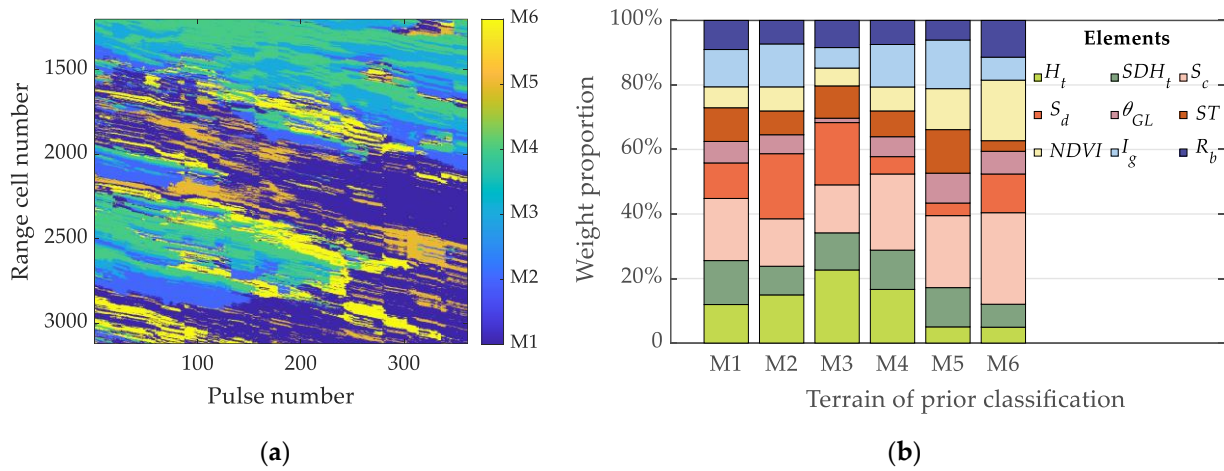


Figure 16. Classification results and weighting coefficients of each type for source data area. (a) Classification label image. (b) Weighting coefficients of each type.

Table 6. Results of multiple-feature weighting for different classifications.

Parameters	H_t	SDH_t	S_c	S_d	θ_{GL}	ST	$NDVI$	I_g	R_b	
Q_i (%)	M1	11.94	13.63	19.27	11.01	6.66	10.44	6.49	11.56	9.02
	M2	14.93	8.89	14.70	20.18	5.87	7.38	7.46	13.31	7.28
	M3	22.66	11.50	14.88	19.30	1.36	10.01	5.53	6.35	8.41
	M4	16.63	12.18	23.60	5.37	6.18	7.99	7.43	13.17	7.45
	M5	5.01	12.17	22.31	3.91	9.26	13.49	12.70	15.07	6.07
	M6	4.92	7.12	28.36	12.01	7.04	3.28	18.77	7.13	11.38

RF classification training was performed on the source region data in Figure 16a. The environmental parameters of the target region were introduced into the RF classification model of the source region to obtain the label classification map of the target region, as shown in Figure 17. In Figure 17a, the accuracy of each label classification in the target region is given. Compared to the actual classification result, the total accuracy of the label classification reached 98%. The results shown in Figure 17b were used in the a priori-knowledge classification extrapolation of the scattering coefficients.

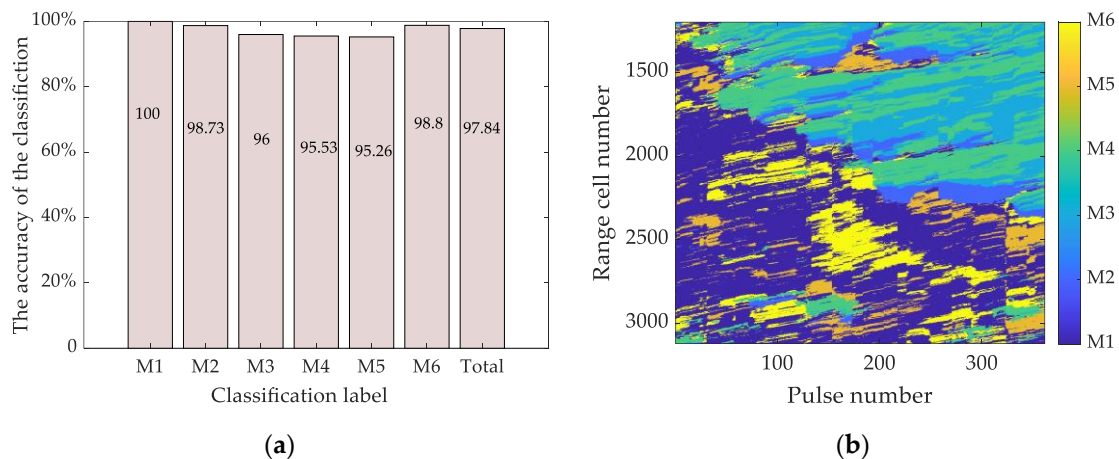


Figure 17. The result of the target region classification. (a) The accuracy of the classification. (b) Label classification image.

3.3. Results of Scattering-Coefficient Estimation for Target Area

The purpose of the scattering-coefficient estimation method is to mine the clutter data of the tested area, determine the influencing mechanisms of the environmental parameters on the clutter characteristics, and estimate the clutter characteristics of the untested area according to the association of the multidimensional environmental parameters between regions. This method can realize the clutter cognitive requirements of complex areas with accuracy and fully use existing data to reduce the test cost, and hence, it has significant engineering application. Figure 18 presents a flowchart of the scattering-coefficient estimation method.

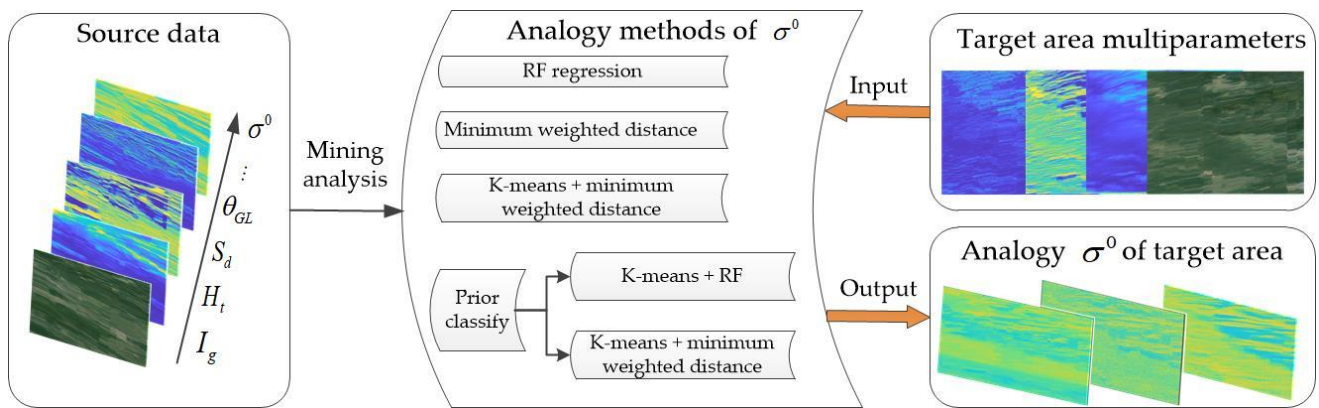


Figure 18. Flow diagram of the method for the estimation of the scattering coefficient of the target area.

3.3.1. Analysis of Similarity between Source and Target Regions

In this study, the similarity between the source and target regions was measured using the BC method, as shown in Equation (10). The amplitude distribution statistics of the corresponding parameters of the source and target regions were obtained and used in Equation (10) to obtain the BCs of the two regions under each parameter. The average of the BCs of all the parameters is the similarity between the two regions.

Figure 19 shows the results of the BC calculations for the target and source regions for multiple parameters. Figure 19a shows the probability distribution of the H_t , and Figure 19b shows the multiparameter results. By averaging the BCs of all the parameters, the overall similarity between the two regions was estimated at 0.87, which indicated that the source and target regions had similar characteristics in terms of multiple parameters. Data from the source region can be used to evaluate the target region.

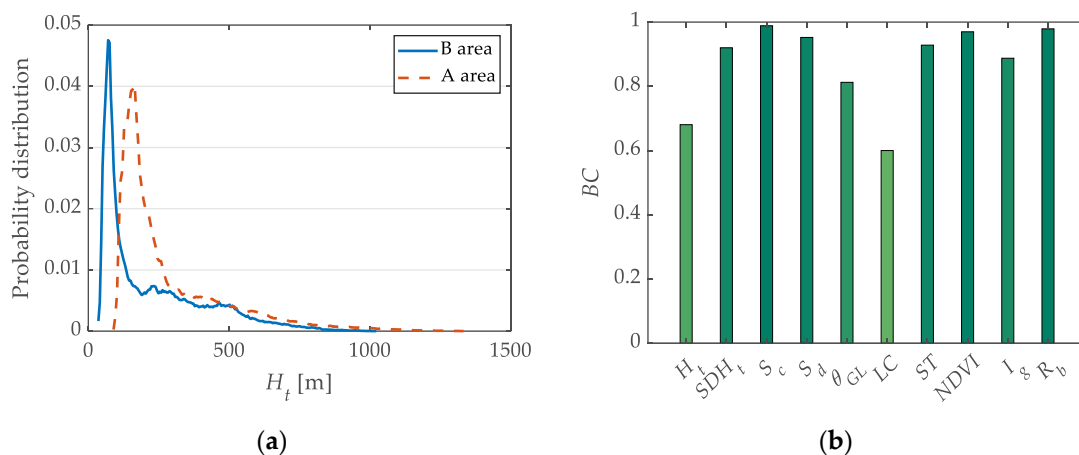


Figure 19. Results of BCs between target and source regions. (a) Comparison of probability distributions of H_t between two regions. (b) BC results for multiple parameters.

3.3.2. Evaluation Index of Prediction Results

Several metrics were used to evaluate the prediction results, including the mean absolute error (MAE), root-mean-squared error (RMSE), coefficient of determination (R^2), and pulse dimension mean value error (PMVE).

The MAE is the average of the absolute errors and is denoted by the following equation:

$$MAE = \frac{1}{k} \sum_{i=1}^k |\sigma_i - \hat{\sigma}_i| \quad (14)$$

where σ_i is the measured data, $\hat{\sigma}_i$ is the predicted result, and the dimensions of both are the same. k is the total amount of data.

The RMSE represents the difference between the predicted value of the model and measured data, and it is used to measure the prediction accuracy of the model. This is expressed by the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^k (\sigma_i - \hat{\sigma}_i)^2}{k}} \quad (15)$$

The R^2 indicates the model's ability to explain the dependent variable and ranges from $-\infty$ to 1; when it is closer to 1, it indicates a better fit of the model to the data:

$$R^2 = 1 - \frac{\sum_{i=1}^k (\sigma_i - \hat{\sigma}_i)^2}{\sum_{i=1}^k (\sigma_i - \bar{\sigma})^2} \quad (16)$$

The measured and predicted data were transformed into a range cell–pulse matrix, as shown in Figure 6a, and the mean value of the data in the pulse dimension was calculated using the mean value curve shown in Figure 6b. The mean value error is calculated as follows:

$$PMVE = \frac{1}{n} \sum_{i=1}^n |\bar{\sigma}_i - \hat{\bar{\sigma}}_i| \quad (17)$$

In the study of scattering characteristics, the PMVE is a commonly used evaluation parameter.

3.3.3. Estimation Results

Several estimation methods were used in this study, such as the RF, MW, KMW, PCKRF, and PCKMW methods. These methods have been described previously; however, the choice of the weight coefficients is explained here. Weighting was applied to the MW method based on the weights shown in Table 4. The results listed in Table 5 were used for the weighted clustering and weighted distance calculations in the PCKMW and PCKRF.

Figure 20 shows a 2D image of the scattering coefficient of the target area and the change curve of the mean scattering coefficient with the grazing angle. The results of the extrapolation methods were compared with the measured data.

Figure 21 shows the calculation results for each method, including the 2D images and scatter plots. The prediction results of the RF are close to the measured data in the two-dimensional image; however, most of the RF prediction results are concentrated between -20 and -40 dB (Figure 21b), and the estimation of the high and low values is not very ideal. The results predicted by the MW method can be spread over the dynamic range; however, a significant difference between the predicted and measured data is observed in the two-dimensional image. The regression of the scatter plot for the MW method was poor and discrete. The KMW method was superior to the MW method, and the predicted and measured data were relatively close; however, the scatter plot was still very discrete. The above three methods are relatively simple and not ideal in terms of the overall 2D image approximation and regressivity of the scatter plot. From the differences between these and

the measured data, the ground-cover scattering is very complex, and it is difficult to obtain good results for clutter prediction in unknown areas using traditional methods.

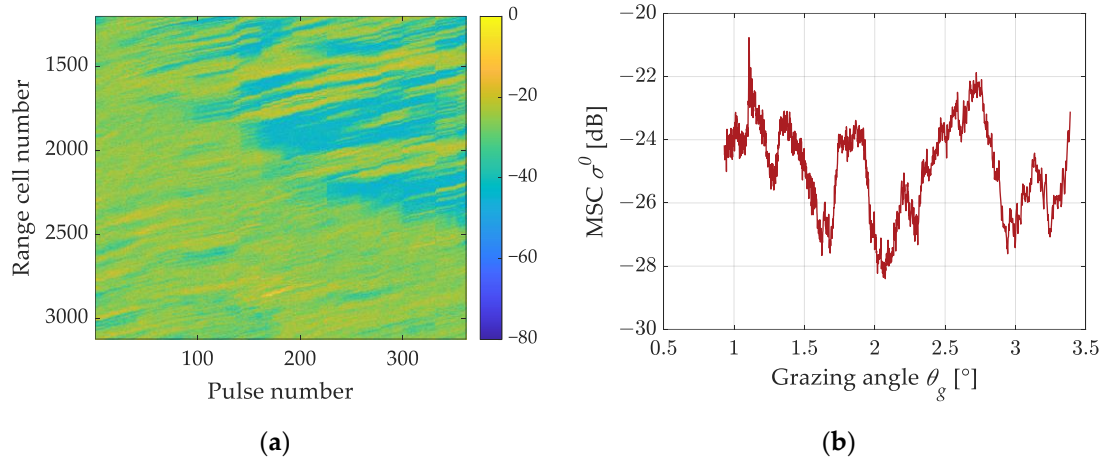


Figure 20. Range–pulse 2D image of the scattering coefficient in the target area and the variation in the mean scattering coefficient with the grazing angle. (a) Range–pulse 2D image. (b) Mean scattering coefficient.

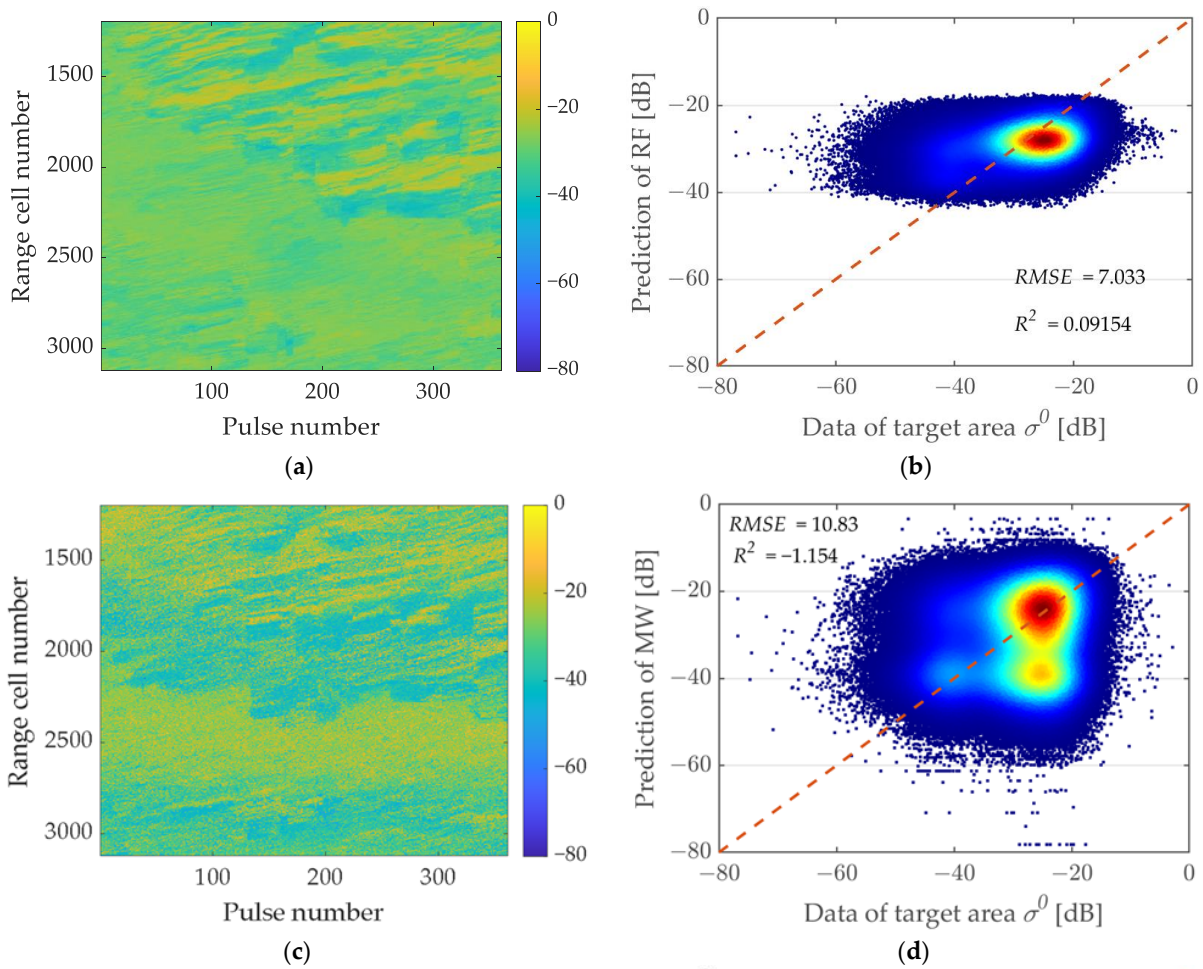


Figure 21. Cont.

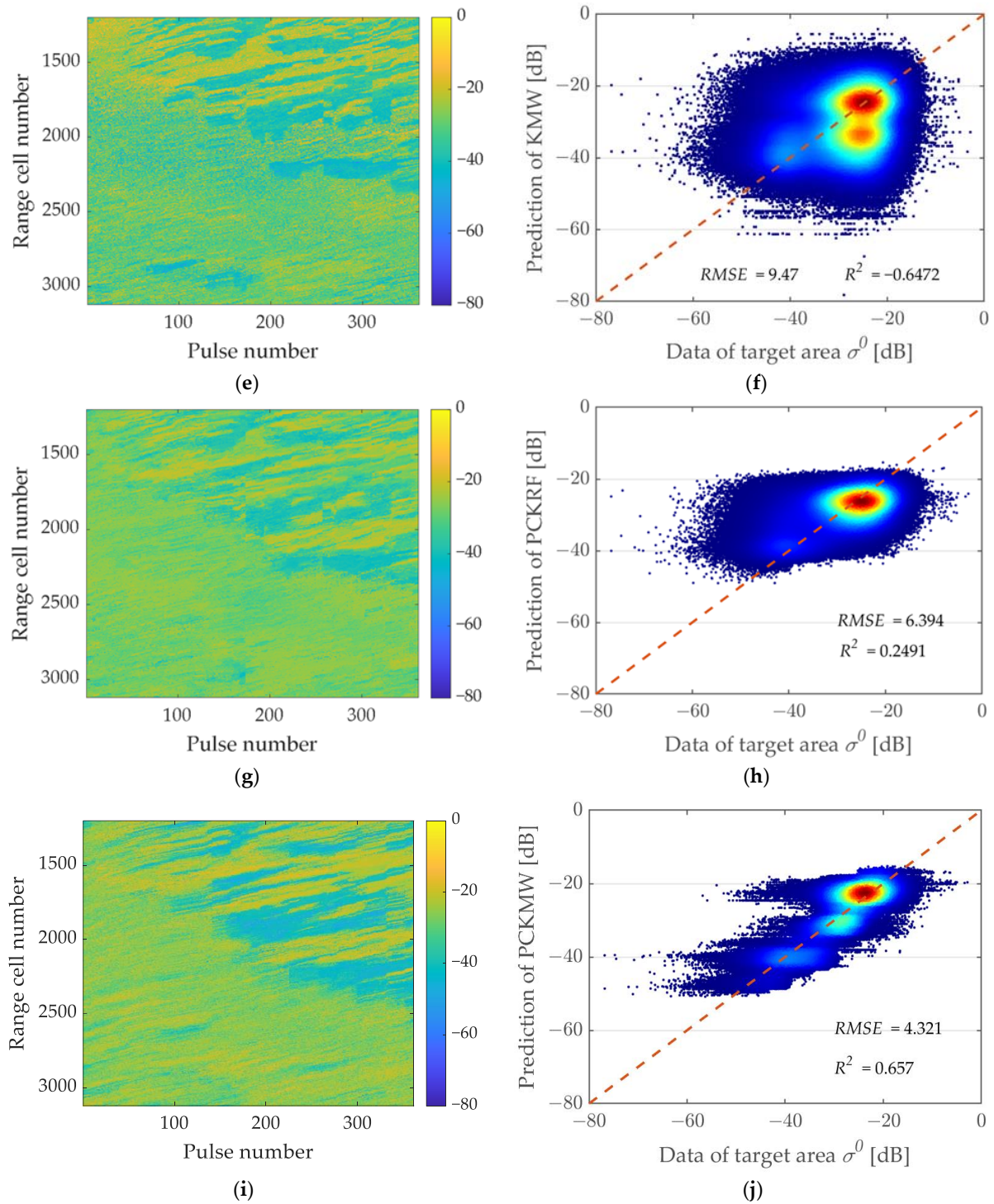


Figure 21. Comparison of prediction results of multiple estimation methods. (a) Range–pulse 2D image of RF results. (b) Scatter plots between RF results and measured data. (c) Range–pulse 2D image of MW results. (d) Scatter plots between MW results and measured data. (e) Range–pulse 2D image of KMW results. (f) Scatter plots between KMW results and measured data. (g) Range–pulse 2D image of PCKRF results. (h) Scatter plots between PCKRF results and measured data. (i) Range–pulse 2D image of PCKMW results. (j) Scatter plots between PCKMW results and measured data. In subfigure (b,d,f,h,j), the red dashed line represents the regression line, while the data is presented in the form of scatter plot density, with red indicating areas of high scatter density and blue indicating areas of low scatter density.

As shown in Figure 21g,h, the PCKRF method is significantly better than the above three methods. First, the two-dimensional scattering coefficient is very close to that of the measured data. Second, the predicted result of the PCKRF method has a wider dynamic range (from -45 to -15 dB) compared to that of the RF method. Third, the overall scatter plot is closer to the regression line. This shows that the pre-classification method combined with clustering and ML regression prediction has a good effect on ground-cover-scattering prediction in unknown areas.

Finally, the prediction results of the PCKMW method are shown in Figure 21i,j. The two-dimensional image of the predicted results is very close to that of the measured data, and the dynamic range is from -50 to -15 dB. The scatter plot is much closer to the regression line than those of the above four methods, indicating that the prediction is comparatively better. However, depending on the clustering method used, some of the predicted data points had the same or similar values.

Figure 22 compares the mean scattering coefficients of the considered estimation methods. The overall comparison shows that the maximum difference between the mean predicted scattering coefficients of the five methods and measured data was no more than 10 dB, which is relatively close, and the trends of variation with the fluctuation in the grazing angle were in agreement. The predicted results of the RF and PCKRF were very close to the mean scattering coefficient, but their values were lower than the measured data. In addition, the prediction accuracy was better for undulating terrain corresponding to grazing angles from 1.5° to 3.4° , and worse for flat terrain with grazing angles from 1° to 1.5° . The prediction results of the MW and KMW are relatively close, and they are in good agreement with the measured data in the grazing-angle range from 1.5° to 2.8° ; however, the results are poor at other angles. The mean predicted result of the PCKMW had the highest degree of agreement with the measured data, particularly when the grazing angle was $<2^\circ$.

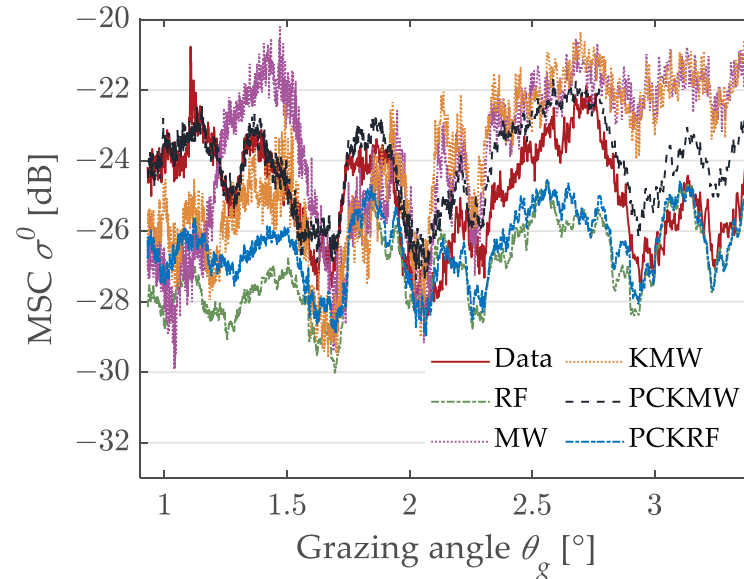


Figure 22. Comparison of mean scattering coefficients of multiple estimation methods.

Table 7 presents a comparison of the evaluation indices for the multiple estimation methods. The two estimation methods that applied pre-classification based on prior knowledge were superior to the RF, MW, and KMW methods for all indicators, and the PCKMW method in particular was much better. The results showed that the accuracy of the estimation method for an unknown region was significantly improved after pre-classification and clustering.

Table 7. Comparison of evaluation indexes of multiple estimation methods.

Estimation Method	RF	MW	KMW	PCKRF	PCKMW
MAE (dB)	5.5231	8.6095	7.4796	4.9319	3.3610
RMSE (dB)	7.0330	10.8299	9.4703	6.3942	4.3214
R^2	0.0915	−1.1541	−0.6472	0.2491	0.6570
PMVE (dB)	2.5735	2.2444	1.9740	1.8290	0.7263

4. Discussion

This study is based on airborne-radar data of a large area with complex-terrain clutter combined with ALOS DEM, Global Land 30, HWSD soil distribution, MODIS NDVI, Google image, and other environmental remote sensing data. A dataset of the association and matching between the clutter and environmental parameters was constructed by reconstructing the geometric relationship of the test scene, dividing the terrain, and matching the illuminated areas. Compared with previous clutter data [22,25], the dataset constructed in this study has the characteristics of wide coverage and many ground object parameters. This dataset overcomes the problem of quantitative parameter representation in the study of large-scene clutter characteristics and enables a better analysis and understanding.

Based on this dataset, aiming to solve the practical problems of the high cost of traditional clutter testing in large areas, difficult testing in remote and complex terrain, and lack of understanding of clutter characteristics, this study attempted to use measured data and the underlying logic of similar associations of multidimensional parameters between different regions to implement clutter estimation methods for similar terrains. This included a weight analysis of the influences of the complex terrain and multiple environmental parameters on the clutter, RF regression prediction, prior-knowledge classification, clustering, and hybrid estimation using multiple-method superposition.

Figures 11 and 21b show the significant differences between the predictions of the RF model in the training set, test set, and unknown regions. The RMSE of the training set was only 3, the R^2 was 0.8782, and the dynamic range in the scattering coefficient was from −50 to −10 dB, which are relatively ideal indicators in cases of such high amounts of data and numbers of multidimensional parameters. For the corresponding test set, the RMSE reached 4.67, the R^2 was 0.7, and the dynamic range was from −44 to −18 dB, which is in line with the expectations. However, in the prediction of the target area, all the evaluation indexes deteriorated: the RMSE was 7, the R^2 was only 0.09, and the scatter plot of the scattering coefficients was concentrated in the range from −40 to −20 dB. This comparison shows that the RF model is inefficient in clutter prediction in unknown regions.

The logic of scattering-coefficient extrapolation for untested areas based on data is somewhat different from traditional ML or DL prediction [39]. First, ML or DL was used to shuffle the entire dataset, most of which was used as the training set to train the model, while a small part was used as the test set to verify the model [30,31]. The data sources for the test and training sets were the same. In this study, the data for the source and target areas were different. Although they were in similar environments, they were not located in the same test area. Second, the high complexity of ground-cover scattering increases the data cost of ML prediction models. For example, the data of the target area were not involved in the training process for the ML prediction. The RF scattering-coefficient prediction (Figure 21b) was not ideal because of the lack of scattering coefficients for the target area. It is reasonable to train the ML model by mixing the target region with data of the source region to improve its prediction accuracy in the target region; however, this is contrary to the precondition that the scattering coefficient of the target region is unknown. Therefore, if we want to rely on complete ML modeling to predict unknown regions, we need a large number of datasets from different regions for training and learning, which will lead to high costs and difficulty in measurements in remote areas. Therefore, the purpose of this study was to use existing data or a low amount of test data to achieve the rapid and accurate cognition of unknown regions.

Based on data calculations, it is believed that the closer the environmental parameters, the more similarity is required in the scattering characteristics. Therefore, after calculating the influence weight of the source area, the MW method was used to match and assign the scattering units of the target and source areas (Figure 21c,d). The 2D plot is close to the measured data, but the scatter plot has poor regression, with an RMSE of 10.83 and an R^2 of -1.154 , indicating that the estimation method is not appropriate. The KMW first classified the source data into several clusters and used the MW method to estimate the corresponding clusters (Figure 21e,f). Compared with the MW, the KMW shows a certain improvement, but the regression is not sufficiently good; the RMSE is 9.47 and the R^2 is -0.6472 .

The prediction results of the RF, MW, and KMW methods are not ideal, mainly because ground-cover scattering is very complex; in particular, the weight coefficient of each parameter that affects the scattering changes significantly with the change in ground cover. For example, in mountainous areas, the shading coefficient is a relatively large parameter, whereas, in flat terrains, the distribution of ground objects such as towns, woodlands, and arable land is the main influencing factor. In addition, the dynamic range of the scattering coefficient in a complex terrain is large, and in many cases, the parameters of the ground objects are close to the measured results; however, the scattering coefficient varies significantly, leading to the insufficient regression of the prediction method.

To overcome these problems, based on a previous method, this study used prior knowledge to classify the source data into six classes, calculate the weight coefficient of each class of data, and perform weighted clustering. In the process of the weighted clustering, the scattering coefficient was also used as one of the clustering parameters, and the weight of the scattering coefficient was set to 30% to increase the influence of the scattering coefficient in the clustering algorithm. Therefore, a dataset of 6×12 groups was obtained, and RF and MW predictions were performed for each dataset, namely, the PCKRF and PCKMW datasets. The PCKMW method uses a weighted minimum distance judgment when judging the category of the target data, and the weights include only those of the environmental parameters. From the comparison results in Figure 21 and the evaluation indicators in Table 6, these two methods were significantly improved compared to the other three methods. Compared with the RF, the two-dimensional plot of the PCKRF is closer to that of the measured results, and the dynamic range reaches from -50 to -15 dB. The regression of the scatter plot also improved to a certain extent, with an RMSE of 6.394 and an R^2 of 0.2491, which are better. The RMSE of the PCKMW method was 4.321 with an R^2 of 0.657, the mean value is very close to that of the measured results, and the PMVE is 0.726 dB; therefore, all the indicators are better than those of the other methods. The results of the comparison of the methods show that the PCKRF and PCKMW provide better clutter estimation in unknown regions after pre-classification and superimposed weighted clustering. However, owing to the adoption of the data-clustering method with an increasing proportion of scattering coefficients, a considerable part of the prediction results of the PCKMW have many concentrated values, which has a negative impact on the regression.

In summary, the terrain clutter estimation method presented in this paper is in the early stages of research, with several areas warranting improvement. For example, the dataset construction relies on the statistical results of the scattering-unit parameters. However, in fact, a scattering unit should be made into a dataset associated with multiple sub-units for research rather than statistical results. In addition, DL neural networks and hyperparameter optimization algorithms [35] should be applied to improve the estimation methods in the future. Furthermore, by converting multidimensional parameter data into a two-dimensional range-pulse map, image recognition and prediction algorithms could be applied to estimate the range-pulse map of scattering coefficients. The verification dataset, derived from the target area, closely aligns with the test lot and parameters of the source area, demonstrating typicality. However, to extend applicability to a broader range of areas and datasets, further research is essential.

5. Conclusions

For the complex-terrain clutter in a vast territory, the complexity and diversity of the ground objects cannot be characterized well, and the influence of environmental factors on the scattering coefficient presents complex nonlinear relationships that are difficult to quantify. As a result, the clutter characteristics of complex terrain cannot be accurately obtained when encountering new ground object environments and test conditions. To address the aforementioned challenges, this study undertook the following research efforts. Firstly, a clutter dataset association with multisource environmental data was established. Secondly, through data mining employing various methods, two novel clutter estimation techniques for similar terrains were developed.

In this study, combining the measured clutter data, DEM data, land-cover classification data, soil composition data, NDVI data, and Google image data, we constructed a multifactor-associated clutter dataset by reconstructing the geometric relationship of the measurement scene, triangulating the terrain, and matching the illumination areas. There are as many as 16 types of parameters involved in this dataset, which quantitatively characterize the complex environmental information from various aspects, such as the terrain relief, shadowing effects, and surface-cover types and their distribution. Taking the dataset of area B as an example, it covers an extensive region of 90 km × 70 km and contains 691,200 rows and 17 columns. In terms of both the richness of the parameter information and the size of the covered area, the association dataset established in this study is unparalleled. This dataset not only solves the problem of the associated quantitative parameter representation of large-scene clutter, but it also provides a foundation for the development of clutter estimation methods for similar terrains.

Based on the multifactor-associated clutter dataset, to overcome the challenges posed by the large dynamic ranges of scattering coefficients and compound effects of multiple parameters on clutter, two innovative clutter estimation methods, designated as PCKRF and PCKMW, were developed by combining multiple techniques, such as prior-knowledge-based classification and multifactor weight coefficients, K-means clustering, RF prediction, and the minimum Euclidean distance. We compared the prediction results of the PCKRF and PCKMW with the measured data from multiple aspects, including the range-pulse, scattering-coefficient, two-dimensional image, scatter plot of scattering coefficients, and mean-scattering-coefficient curve. And multiple metrics, such as the MAE, RMSE, R2, and PMVE, were used to validate the accuracies of the algorithms. The results showed that the performances of both the PCKRF and PCKMW methods in predicting the scattering coefficient were notably superior to those of the RF, MW, and KMW. Specifically, compared with the RF, the two-dimensional plot of the PCKRF was closer to that of the measured results, and the dynamic range reached from −50 to −15 dB. The regression of the scatter plot also improved to a certain extent, with an RMSE of 6.394 dB and an R2 of 0.2491, which were better than those of the RF. Even more impressive was that all the indicators of the PCKMW method surpassed those of the other methods, with an RMSE of 4.321 dB, a PMVE of 0.726 dB, and an R2 of 0.657.

The above demonstration illustrates that prior-knowledge-based classification and weighted coefficient calculation are particularly beneficial for overcoming the challenges posed by the large dynamic ranges of scattering coefficients and numerous parameters in complex terrains across extensive areas, ultimately enhancing the prediction accuracy. The pre-classification classifies source data into six classes with similar features, which helps to calculate more accurate weight coefficients for each data class. And, in the process of weighted clustering for source data, the scattering coefficient is also used as one of the clustering parameters, and the weight of the scattering coefficient is set to 30% to increase the influence of the scattering coefficient in the clustering algorithm. In addition, by comprehensively considering the correlation coefficient and the RF out-of-bag parameter error, a comprehensive weight coefficient was formed to characterize the multifactor influence, which was found to be superior to the fitting effects of using either of these two factors individually. Through the aforementioned operations, a large and complex dataset was

divided into 6×12 clusters. Each cluster contained elements that were similar in terms of their influence on the scattering coefficient. Subsequently, by applying the RF and MW, a relatively good prediction performance was achieved. The current achievements in clutter estimation methods offer a viable option for accurately recognizing clutter characteristics in complex-terrain environments where data collection may be difficult or impossible, with lower human and economic costs.

Despite the certain effectiveness of the current clutter estimation methods, they still require considerable time and effort to refine and optimize the algorithms to further improve their accuracy, robustness, and applicability. In the future, we plan to make improvements in the following aspects. Firstly, we aim to enrich and optimize the clutter dataset associated with multiple elements, making it closer to real-world scenarios and providing more comprehensive representations of the element information. Secondly, we intend to apply deep learning networks to the data training and prediction, which will leverage their ability to learn complex patterns and relationships from large datasets, thereby enhancing the accuracy of the algorithms. Thirdly, we plan to enrich the knowledge information of the pre-classification, making the classification method of terrain features more detailed and accurate. This will enable us to develop more fine-grained classification schemes that can distinguish between different types of terrain features more precisely, leading to more accurate clutter estimations. Lastly, we intend to consider broader principles for weight coefficient computation and to optimize the methodology for determining these weights. By addressing these areas, we can expect to see significant advancements in clutter estimation techniques in the coming years.

Author Contributions: Conceptualization, Y.Z., Z.W. and Q.L.; methodology, P.Z.; formal analysis, P.Z. and D.Z.; investigation, H.P.; resources, Z.Y.; data curation, D.Z.; writing—original draft preparation, P.Z.; writing—review and editing, P.Z.; visualization, J.Z.; supervision, L.L.; project administration, Z.Y.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant Nos. U2006207 and 62271381, and was also supported by the Foundation of the National Key Laboratory of Electromagnetic Environment under Grant No. 6142403180204.

Data Availability Statement: Data are not publicly available due to privacy restrictions.

Acknowledgments: We thank our colleagues for their contributions during the experiments, and data pretreatment supported the manuscript. The ALOS data products used in this study were retrieved from an online Data Pool courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), the USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, https://lpdaac.usgs.gov/data_access/data_pool (accessed on 20 September 2023). We are truly grateful for the provision of these data, which were the focal point of our study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Skolnik, M.I. *Radar Handbook*, 3rd ed.; McGraw-Hill: New York, NY, USA, 2008.
2. Long, M.W. *Radar Reflectivity of Land and Sea*, 3rd ed.; Artech House: Boston, MA, USA, 2002.
3. Sloper, D.; Fenner, D.; Arntz, J.; Fogle, E. Multi-Channel Airborne Radar Measurement (MCARM) Flight Test. In *Westinghouse Electronic Systems Technical Report Contract F30602-92-C*; Westinghouse Electronic Systems location: Monroeville, PA, USA, 1996.
4. Billingsley, J.B. *Low-angle Radar Land Clutter Measurements and Empirical Models*; William Andrew Publishing: New York, NY, USA, 2002.
5. Craven, T. *Mountaintop Surveillance Sensor Test Integration Center Facility, Kauai, Hawaii*; Army Space and Missile Defense Command: Huntsville, AL, USA, 2000.
6. Pinel, N.; Boulier, C. *Electromagnetic Wave Scattering from Random Rough Surfaces: Asymptotic Models*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
7. Tsang, L.; Kong, J.A.; Ding, K.-H. *Scattering of Electromagnetic Waves: Theory and Applications*; John Wiley & Sons: New York, NY, USA, 2000.
8. Fung, A.K.; Chen, K.S. *Microwave Scattering and Emission Models for Users*; Artech House: Boston, MA, USA, 2010.
9. Ulaby, F.T.; Craig, D.M.; Álvarez-Pérez, J.L. *Handbook of Radar Scattering Statistics for Terrain*; Artech House: Boston, MA, USA, 1989.

10. Currie, N. Clutter characteristics and effects. In *Principles of Modern Radar*; Eaves, J.L., Ready, E.K., Eds.; Springer: Boston, MA, USA, 1987; pp. 281–342.
11. Morchin, W.C. *Airborne Early Warning Radar*; Artech House: Norwood, MA, USA, 1990.
12. Ulaby, F.T.; Moore, R.K.; Fung, A.K. *Microwave Remote Sensing Active and Passive: Volume II: Radar Remote Sensing and Surface Scattering and Emission Theory*; Artech House: Norwood, MA, USA, 1982.
13. Fung, A.K.; Fung, H. Application of first-order renormalization theory for cross-polarized backscatter from a half-space. *IEEE Trans. Geosci. Electron.* **1977**, *15*, 189–195. [[CrossRef](#)]
14. Ulaby, F.T.; Moore, R.K.; Fung, A.K. *Microwave Remote Sensing Active and Passive: Volume III: From Theory to Applications*; Artech House: Norwood, MA, USA, 1986.
15. Peng, S.; Tang, Z. Reflectivity Model of Ground/Sea Clutter. *J. Airf. Radar Acad.* **2000**, *14*, 1–4.
16. Hellard, D.L.; Henry, J.P.; Agnesina, E.; Moruzzis, M. Ground clutter simulation for surface-based radars. In Proceedings of the International Radar Conference, Alexandria, VA, USA, 8–11 May 1995.
17. Feng, S.; Chen, J. Low-angle reflectivity modeling of land clutter. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 254–258. [[CrossRef](#)]
18. Darrach, C.A.; Luke, D.W. Site-specific clutter modeling using DMA digital terrain elevation data (DTED), digital feature analysis data (DFAD), and Lincoln Laboratory five frequency clutter amplitude data. In Proceedings of the IEEE National Radar Conference, Ann Arbor, MI, USA, 13–16 May 1996.
19. Kurekin, A.; Radford, D.; Lever, K.; Marshall, D.; Shark, L.-K. New method for generating site-specific clutter map for land-based radar by using multimodal remote-sensing images and digital terrain data. *IET Radar Sonar Nav.* **2011**, *5*, 374–388. [[CrossRef](#)]
20. Capraro, C.; Capraro, G.; Bradaric, I.; Weiner, D.; Wicks, M.; Baldygo, W. Implementing digital terrain data in knowledge-aided space-time adaptive processing. *IEEE Trans. Aerosp. Electron. Syst.* **2006**, *42*, 1080–1099. [[CrossRef](#)]
21. Xie, M.; Yi, W.; Kong, L. Knowledge-aided space-time adaptive processing based on Morchin model. In Proceedings of the IET International Radar Conference, Hangzhou, China, 14–16 October 2015.
22. Li, H.; Wang, J.; Fan, Y.; Han, J. High-fidelity inhomogeneous ground clutter simulation of airborne phased array PD radar aided by digital elevation model and digital land classification data. *Sensors* **2018**, *18*, 2925. [[CrossRef](#)]
23. Oreshkina, M.; Stepanov, M.; Kiselev, A. Digital earth surface maps for radar ground clutter simulation. *J. Syst. Eng. Electron.* **2022**, *33*, 340–344. [[CrossRef](#)]
24. Luo, J.; Huang, Y.; Zhang, Y.; Mao, D.; Yang, W.; Zhang, Y.; Yang, J. Optimal search strategy of low-altitude target for airborne phased array radar using digital elevation model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8025–8037. [[CrossRef](#)]
25. Kim, D.; Park, A.J.; Suh, U.S.; Goo, D.; Kim, D.; Yoon, B.; Ra, W.S.; Kim, S. Accurate clutter synthesis for heterogeneous textures and dynamic radar environments. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 3427–3445. [[CrossRef](#)]
26. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
27. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
28. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [[CrossRef](#)]
29. Sales, M.H.R.; Bruin, S.D.; Souza, C.; Herold, M. Land use and land cover area estimates from class membership probability of a random forest classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4402711. [[CrossRef](#)]
30. Wang, Y.; Zhan, Y.; Yan, G.; Xie, D. Generalized fine-resolution FPAR estimation using Google Earth Engine: Random forest or multiple linear regression. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 918–929. [[CrossRef](#)]
31. Devyatkin, D.A. Estimation of vegetation indices with random kernel forests. *IEEE Access* **2023**, *11*, 29500–29509. [[CrossRef](#)]
32. Dong, L.; Du, H.; Mao, F.; Han, N.; Li, X.; Zhou, G.; Zhu, D.; Zheng, J.; Zhang, M.; Xing, L.; et al. Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique—Subtropical area for example. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 113–128. [[CrossRef](#)]
33. Fernández-González, P.; Bielza, C.; Larrañaga, P. Random forests for regression as a weighted sum of k-potential nearest neighbors. *IEEE Access* **2019**, *7*, 25660–25672. [[CrossRef](#)]
34. Ahmed, O.S.; Franklin, S.E.; Wulder, M.A.; White, J.C. Extending airborne LIDAR-derived estimates of forest canopy cover and height over large areas using kNN with Landsat time series data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3489–3496. [[CrossRef](#)]
35. Ma, L.; Wu, J.; Zhang, J.; Wu, Z.; Jeon, G.; Zhang, Y.; Wu, T. Research on Sea Clutter Reflectivity Using Deep Learning Model in Industry 4.0. *IEEE Trans. Ind. Inform.* **2020**, *16*, 5929–5937. [[CrossRef](#)]
36. Shui, P.L.; Shi, X.F.; Li, X.; Feng, T.; Xia, X.Y.; Han, Y. GRNN-Based Predictors of UHF-Band Sea Clutter Reflectivity at Low Grazing Angle. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1502205. [[CrossRef](#)]
37. Linghu, L.; Wu, J.; Jeon, G.; Wu, Z.S.; Shi, M. Sea Clutter Feature Prediction and Parameters Inversion Using Deep Learning Model. *IEEE Trans. Ind. Inform.* **2023**, *19*, 8374–8383. [[CrossRef](#)]
38. ASF DAAC. *ALOS PALSAR—Radiometric Terrain Correction*. ASF’s Radiometric Terrain Correction Project; ASF DAAC: Fairbanks, AK, USA, 2015. [[CrossRef](#)]
39. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [[CrossRef](#)]

40. Nachtergaele, F.; Velthuizen, H.; Verelst, L.; Wiberg, D. *Harmonized World Soil Database (HWSD)*; Food Agriculture Organization of the United Nations: Rome, Italy, 2009. Available online: <https://library.wur.nl/WebQuery/wurpubs/500737> (accessed on 20 September 2023).
41. Justice, C.O.; Vermote, E.; Townshend, J.R.G.; Defries, R.; Roy, D.P.; Hall, D.K.; Salomonson, V.V.; Privette, J.L.; Riggs, G.; Strahler, A.; et al. The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1228–1249. [[CrossRef](#)]
42. He, J.; Zhao, H.; Jiang, Y. Method for determining comprehensive weight vector based on multiple linear fitting. In Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, China, 24–26 November 2017; pp. 1–6.
43. Huang, Z.; Chai, J.; Li, B.; Feng, X. Index weighting methods. In Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 15–17 December 2017; pp. 1521–1525.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.