

Article

Not peer-reviewed version

Machine Learning and Linear Regression Models for Mapping Soil Properties and Albedo in Periglacial Areas Using Sentinel Imagery (Byers Peninsula, Marine Antarctica)

[Susana del Carmen Fernandez](#)*, [Ruben Muñiz](#), [Juanjo Peón](#), [Ricardo Rodriguez Cielos](#), [Alfonso Pisabarro](#), [Javier Fernandez Calleja](#)

Posted Date: 28 February 2024

doi: 10.20944/preprints202402.1632.v1

Keywords: Antarctic periglacial areas; albedo bare soils; Machine Learning; neural networks; Soil properties modeling; Sentinel images



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Machine Learning and Linear Regression Models for Mapping Soil Properties and Albedo in Periglacial Areas Using Sentinel Imagery (Byers Peninsula, Marine Antarctica)

Susana del Carmen Fernández ^{1,*}, Rubén Muñiz ^{2,†}, Juanjo Peón ^{3,†},
Ricardo Rodríguez-Cielos ⁴, Alfonso Pisabarro ⁵ and Javier F. Calleja ⁶

¹ ICTEA (Instituto de Ciencias y Tecnologías de Ciencias Aeroespaciales de Asturias), Oviedo, Spain

² Department of Computer Science. University of Oviedo, Spain; rubenms@uniovi.es

³ juanjopeon@gmail.com

⁴ Department of Signals, Systems and Radiocommunications (SSR), Polytechnic University of Madrid, Spain; ricardo.rodriguez@upm.es

⁵ Department of Geography and Geology, University of León, Spain; apisp@unileon.es

⁶ Department of Physics, University of Oviedo, Spain; jfcalleja@uniovi.es

* Correspondence: fernandezmsusana@uniovi.es

† Current address: Affiliation 1.

‡ These authors contributed equally to this work.

Abstract: Byers Peninsula (62°34'–62°40'S–60°54'61°13'W), 60 km² in size, is considered one of the largest ice-free areas in Antarctica. Since 2006, the Spanish Polar Program has taken part in a large number of environmental studies involving effects of climate changes, limnology and microbiology, live cycles, but not albedo. Surface albedo is one of the key physical parameters in the surface energy budget of polar regions. Most of Antarctica is covered by ice sheets; only about 0.44% of the area is permanently ice-free. However, in maritime Antarctica, the ice-free areas, corresponding to small islands, peninsulas, and coastal beaches, account for about 3% of the territory. To incorporate the contribution of these areas into global albedo models, it is necessary to relate the soil properties of the ice-free areas to the surface albedo response. Also, mapping soil properties and albedo have and special interest in these ice-free areas. Image classification using machine learning methods trained with georeferenced soil data could be useful for mapping soil properties and albedo in multispectral optical satellite images. A shallow neural network implemented using the Keras Python module was used to define and train models of soil properties using 15 explanatory variables corresponding to bands and spectral indices of a Sentinel image and a population of 49 soil samples taken from the top 5 centimeters of the soil profile. The soil samples were analyzed in the laboratory and a spectral library in the Vis-Nir range (350–2500 nm) was created. The albedo of the samples was integrated from the ADS spectra. At the same time, a linear regression model of albedo using the soil properties as explanatory variables was performed. The R² fit of this new model was about 0.82 and the error of estimation was 4.1. The model was extended to the entire Byers Peninsula using the ML soil property models as explanatory variables. The RMSE of the extended models increased up to 8.2. We think that the main cause of this increase (2.6 points) is related to the error propagation, phenomenon due to the use of models as explanatory variables. However, another important part of the error increase could be due, among other reasons, to the use of an image that, although corrected, is not completely free of clouds.

Keywords: Antarctic periglacial areas; albedo bare soils; Machine Learning; neural networks; Soil properties modeling; Sentinel images.

1. Introduction

The Polar Regions are crucial to the Earth's climate system and are highly susceptible to the effects of climate change. Surface albedo is a critical physical parameter in studies of the terrestrial surface energy budget [1,2] and its temporal and spatial variations are closely linked to global climate change and regional weather patterns due to the albedo feedback mechanism. The calculation of surface albedo holds significant importance in the global energy balance [3]. In polar regions, the albedo of snow and ice primarily governs surface albedo [4,5]. However, properties such as color, texture, mineralogy, and organic carbon content of bare soils are strongly related to the albedo response of ice-free surfaces [6]. Although most of Antarctica is covered by ice sheets, a small portion of the continent remains permanently free of ice. These ice-free regions account for approximately 0.44% of the total area of Antarctica and consist mainly of small islands, beaches along the coast, mountain ranges, and nunataks. However, in maritime Antarctica, the ice-free areas are significantly greater, reaching about 3%. Along the borders of the ice caps, there are areas that are free of ice and expose bare soil for several months per year [7]. The contribution of these areas to the global albedo in Antarctica is significant. To establish the relationship between soil properties and the albedo response of the surface, it is necessary to consider relevant soil properties. Some of these properties have high temporal variability [6,8] while others remain stable for long periods of time. These properties significantly contribute to albedo, and it is possible to track them over long periods. This paper aims to elaborate a structural albedo index for bare soil areas based on soil properties such as soil organic compounds, texture, and concentration of elements such as iron or calcium. This structural index does not attempt to explain the total variability of albedo in bare soils, but only aims to explain the variability of albedo linked to the main soil components. The experiment took place in a 60 km² ice-free area on the Byers Peninsula in Maritime Antarctica.

Conducting field surveys on Byers Peninsula presents significant challenges due to the extreme environmental conditions and restricted access. Byers Peninsula is highly sensitive to trampling impact, as noted by [9,10]. These restrictions also affected the mapping and soil sampling processes. For extreme situations, some researchers, such as [11], argue that digital soil mapping techniques are necessary to interpolate soil properties between sparse observations in order to map the soil properties of restricted Antarctic areas.

The free availability of satellite images of different wavelengths nowadays facilitates the creation of environmental variable maps in areas that are difficult to access. Several studies have been conducted in several Antarctic regions using remote sensing technologies to analyze environmental variables such as vegetation [12,13] and soils [14–16].

However, there are not many works on soil mapping and assessing the spatial variability of soil properties [11]. In 2020 [13] presented a soil-geomorphological map of the Fildes Peninsula and Ardley Island (King-George Island). The map was created using the conventional method and was based on 130 soil samples. For Byers Peninsula, although access is limited, there is a relatively abundant amount of information available on soil formation dating back several decades [15,17–20]. However, these works are insufficient for the elaboration of extensive maps of soil properties for the entire Byers Peninsula [21].

This study utilized modeling techniques to establish correlations between the values of radiometric bands and spectral indices derived from Sentinel imagery and the properties of soil measured in collected samples. By employing algorithms, these methods enable us to interpolate soil property values in areas that have not been previously explored [11]. Several authors have attempted to model the spatial distribution of soil properties using explanatory variables [22,23]. In this study, we utilized a multilayer perceptron implemented in Python with Keras (<https://keras.io/>), a high-level API for building and neural network models. To calculate the structural albedo index based on soil properties, we used a linear regression model (LRM) with soil properties as explanatory variables to predict the albedo values and VIS-NIR spectra of the soil samples. The main objective of linear regression is to determine the most accurate straight line that describes the relationship between the

dependent variable and the independent variables. Linear regression is a powerful and straightforward algorithm due to its simplicity and interpretability. In addition, incorporating predictions as variables introduces uncertainty, and analyzing the error propagation of the model is fundamental to ensuring reliability and accuracy. Calculating error propagation in linear regression models is feasible following the recommendations of the International Error Metrology Guide.

The main aims of the present study are:

1. To predict soil and sediment properties, Machine Learning techniques are used with radiometric bands and indexes from Sentinel Image as explanatory variables.
2. To extend the soil properties models to all Byers peninsula using GIS techniques.
3. To conduct a structural index of albedo for all of Byers Peninsula based on soil and sediment properties, as well as VIS-NIR spectra of the soil samples.
4. To apply the albedo structural index in Byers peninsula using Sentinel.

2. Materials and Methods

2.1. Byers Peninsula

Byers Peninsula (Figure 1) is considered one of the largest ice-free areas in maritime Antarctica and is an Antarctic Special Protected Area (ASPA No. 126) designated as Site of Special Scientific Interest (SSSI). The main purpose of designating Byers Peninsula as an ASPA is to safeguard the terrestrial and lacustrine habitats within the area. Since 2006, the Spanish Polar Programme has participated in several studies on the effects of climate change, limnology, microbiology, and life cycles, among others. Byers Peninsula was established as an 'International Antarctic Reference Site for Terrestrial, Freshwater and Coastal Ecosystems' during the four International Polar Year. A review of all science conducted in the area between 1957 and 2012 is provided by [24]. The automatic mapping of soil properties in this area has special value due to the SSSI consideration of the peninsula and restrictions on access to most of the territory.

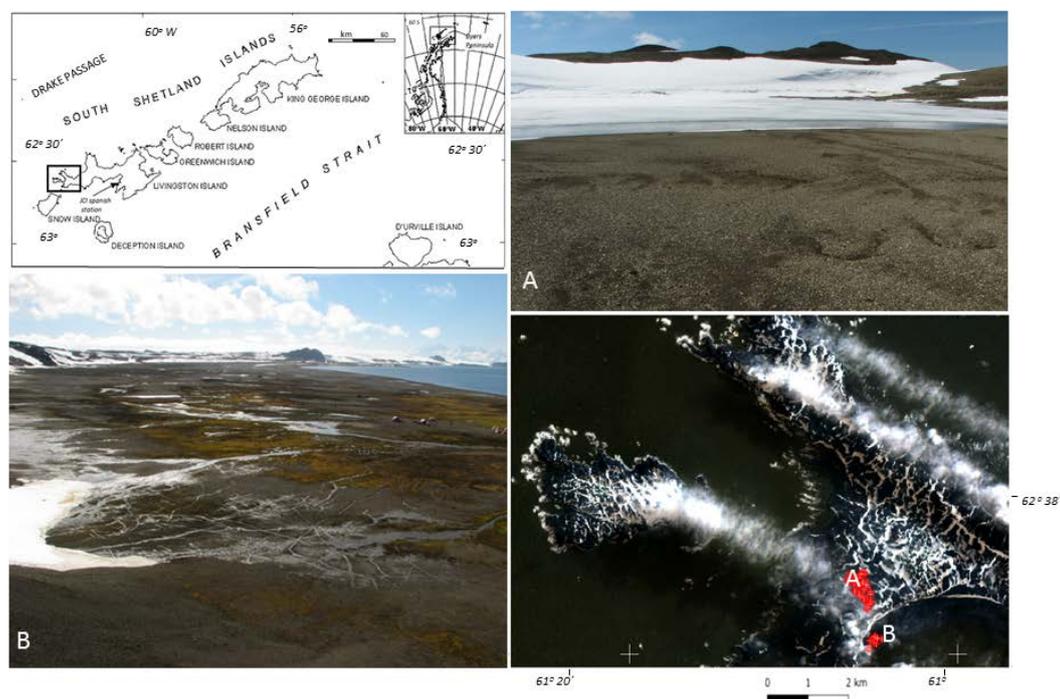


Figure 1. Location of the Byers peninsula. Image Sentinel used in this study with the location of soil samples and photos showing main types of terrains sampled. Photo A corresponds to plot A in the sentinel image, Photo is from the beaches (Plot B).

2.2. Geological and Geomorphological Setting

Byers Peninsula is situated in a Maritime Antarctic environment, which includes the Antarctic Peninsula and surrounding islands. The peninsula is located at the western end of Livingston Island, South Shetland Islands (approximately $62^{\circ}13'70\text{S}$, $61^{\circ}1'60\text{W}$). The environmental conditions in this region are more conducive to soil development than in other parts of Antarctica. The region experiences an annual precipitation of over 500 mm and an average temperature of -2°C . During the summer, precipitation levels rise above 100 mm and the average daily temperature at sea level is above 0.1°C [25].

Livingstone Island is mostly covered by glaciers, except for Byers Peninsula, which is the largest ice-free area in the South Shetland Islands. The chronology of the deglaciation process was obtained by dating the deeper layers of sediments in the lakes [26]. The last major deglaciation of the peninsula occurred between 5,000 and 4,000 years before present [27]. The rocky outcrops of the peninsula consist mainly of sandstones, slates, micro-conglomerates, volcanic and volcanoclastic rocks (Upper Jurassic to Lower Cretaceous) with igneous bodies [26,28,29]. With the exception of the northwestern part of the Peninsula, which reaches 268 m.a.s.l. above sea level, few areas exceed altitudes of 100 m.a.s.l., except for some residual hills (usually volcanic plugs). The highest of these, Chester Cone, is a prominent feature in the central part of the peninsula, rising 193 meters high. The inland part of the peninsula is a regular, undivided platform, covering about 40 km² and ranging between 85 and 100 meters in height, known as the 'Byers Plateau' [30,31]. The majority of the peninsula displays erosive and depositional features of glacial origin, as well as an incipient drainage system. During the austral summer, over 65 shallow lakes and numerous streams become active. The majority of the peninsula is encompassed by vast beaches, with a series of raised beaches located above them. The Rotch Ice Dome glacier, which reaches a height of approximately 360 m, defines the eastern edge of the peninsula. The

soils in Byers are a result of periglacial and nival processes, as indicated by the patterned ground [17,26]. The soil parent materials vary from marine sedimentary to volcanic and volcanoclastic rocks, intruded by igneous bodies. According to the World Reference Base for Soil Resources (WRB) system, 23 soil profiles were classified as Fluvisols, Regosols, Leptosols, or Cryosols [21]. According to [21], the soils in northern Byers Peninsula are generally shallow and coarse-textured, with low organic matter content. The northern coastal region's rocky platforms have ornithogenic characteristics, with lower pH, higher P, Al³⁺, and organic C values. The soils' clay fraction indicates physical weathering occurs with limited chemical alteration of primary minerals.

2.3. Sampling and Analysis

In [19], a map of the substrata of the Byers Plateau and Southern Beach can be found, which includes lithological units differentiated based on diverse criteria, such as rock composition, presence of organic matter, and texture. Samples of the materials that form these units were taken at different locations in two plots (see Figure 1, plots and photos A and B): one on the Byers Plateau and the other on the Southern Beach. In the Limnopolar Lake watershed, located in the Byers Plateau at an average altitude of 80-100 m.a.s.l., we collected 42 samples. The remaining 7 samples were taken from the Southern Beaches, which are situated at sea level altitude. These beaches extend along the southern side of Byers Peninsula, between Devils Point to the west and Rish Point to the east. To ensure comparable data from all sites and for the satellite images, soils were sampled at the surface (0-5 cm). A map displaying the coordinates of the collected samples has been created. The following properties were analyzed in the fraction <2mm: pH, electrical conductivity (EC), density, texture (percentage of sand, silt, and clay), dissolved organic carbon (mg/L), Fe³⁺ (mg/kg), Fe associated with organic matter extracted with pyrophosphate, manganese (g/kg), and Ca²⁺ (g/kg). Some of these properties are directly related to soil color. The albedo of the 49 samples was calculated using a Vis-Nir spectral library obtained with an ADS spectroradiometer.

2.4. Satellite Imagery

The search for Sentinel-2 data for the Byers Peninsula has been conducted, and various options have been analyzed. These options include the use of the Sentinel-2 Global Mosaic (S2GM), a Sentinel-2 image mosaic obtained in Google Earth Engine, and daily images. The use of mosaics has been discarded due to the high presence of clouds and snow in the region, which does not provide an acceptable result. Out of the more than 200 daily images, only one image is suitable for use - the Sentinel 2-A image from March 28, 2016, at 13:29. The other images contain too many clouds and snow. The image was downloaded at L1B level (TOA reflectivity) and corrected for atmosphere using Sen2Cor in SNAP to obtain the L2A level product (BOA reflectivity). This work will use 9 bands of Sentinel-2, as indicated in blue in Table 1. Additionally, six radiometric indexes were applied to the BOA reflectivity (L2A) image, as shown in Table 2.

Table 1. Nine bands of Sentinel-2 used in this study.

Spatial resolution	Band	Spectral region	Central wavelength (nm)	Bandwidth (nm)
10	2	VIS-Blue	496.6	98
	3	VIS-Green	560.0	45
	4	VIS-Red	664.5	38
20	5	NIR	703.9	19
	6	NIR	740.2	18
	7	NIR	782.5	28
	8a	NIR	864.8	33
	11	SWIR	1613.7	143
	12	SWIR	2202.4	242

Table 2. Radiometric indexes used in this study; the expression and Sentinel-2 bands involved in their formulation.

Indexes	Expression	Sentinel 2 Bands	Authors
Ferric iron (Fe3)	$\frac{\rho_{RED}}{\rho_{GREEN}}$	B4-VIS - ρ_{RED} B3-VIS - ρ_{GREEN}	[32]
Hue	$\frac{\rho_{RED}}{\rho_{GREEN}}$	B2-VIS - ρ_{BLUE} B3-VIS - ρ_{GREEN} B4-VIS - ρ_{BLUE}	[33]
IR550	$\frac{\rho_{RED}}{\rho_{GREEN}}$	B3-VIS - ρ_{GREEN}	[34]
IR700	$\frac{\rho_{RED}}{\rho_{GREEN}}$	B5 - ρ_{NIR}	[34]
Missa Soil Brightness Index (MSBI) v2	$MSBI = 0.406 \times \rho_{GREEN} + 0.600 \times \rho_{RED} + 0.645 \times \rho_{NIR1} + 0.243 \times \rho_{NIR2}$	B3-VIS - ρ_{GREEN} B4-VIS - ρ_{RED} B6-NIR - ρ_{NIR1} B8a-NIR - ρ_{NIR2}	[35]
I/O (Oxides)	$IO = \frac{\rho_{RED}}{\rho_{BLUE}}$	B2-VIS - ρ_{BLUE} B4-VIS - ρ_{RED}	[36]

3. Modelling Soil Properties and Albedo

3.1. Multilayer Perceptron

To generate a non-linear regression model we defined and trained a multilayer perceptron [37] (Figure 2). This is a common resource to solving problems where a linear approach is not good enough [38,39]. In summary, the topology of the neural network used in this work is described as follows:

1. Input layer (size = 15)
2. First hidden layer (500 neurons)
3. Second hidden layer (100 neurons)
4. Third hidden layer (50 neurons)
5. Output layer (1 neuron)

The model was trained using the Mean Squared Error (MSE) loss function [40], while the chosen optimizer for this purpose was Adam [41].

Table 3. Descriptive parameters of the soil properties compared to ML_soil properties.

n 49	H ₂ O PH	Density (g/cm ³)	Fe ³⁺ (mg/kg)	DOC (mg/L)	Organic matter Fe (mg/kg)	Mn (g/Kg)	Ca (g/Kg)	Clay (%)	Silt (%)	Salt (%)
Mean	7.32	1.15	96.81	190.30	289.54	6.87	16.93	15.14	20.15	64.71
ML_Mean	7.40	1.10	91.09	117.00	305.34	6.39	16.86	12.66	20.42	60.91
Min	5.07	1.16	65.12	0.00	53.60	1.28	2.20	4.13	4.37	43.85
ML_Min	5.72	0.65	74.90	5.04	94.01	1.76	7.14	8.26	9.05	51.15
Max	8.26	1.50	129.18	3671.09	1768.00	16.49	29.20	32.42	38.59	87.38
ML_Max	8.33	1.40	114.23	1114.42	1278.48	10.91	23.99	21.10	30.13	80.98
Std	0.72	0.27	13.30	656.77	208.54	4.28	7.62	5.33	8.39	11.26
ML_Std	0.71	0.20	7.96	262.31	374.26	2.04	5.27	21.10	5.48	7.07
MAE	0.51	0.17	9.93	55.60	116.70	2.61	4.68	4.01	5.17	7.99
RMSE	0.69	0.22	13.04	156.20	218.39	3.69	5.89	5.43	6.43	9.74

Predictor variables included spectral bands and indices (Tables 1 and 2). We included 15 explanatory variables in the model, but the importance of each variable varied from pixel to pixel. As a result, it is difficult to determine which variables were most important in predicting each soil property using this approach. The study of soil properties was extended to the entire Byers Peninsula using the Sentinel image. Figure 3 shows the spatial distribution of six soil properties: pH (in H₂O), clay content (%), Fe³⁺ concentration (mg/kg), Fe associated with organic matter (mg/kg), dissolved organic carbon (mg/L), and calcium concentration (g/kg). These soil properties were used to predict the albedo distribution in the linear regression model shown in Figure 4. To simplify the figure, we have omitted the coordinates and scale on each map. This information has already been presented in Figure 1. A cloud filter that we created in the SNAP program is included in the figures. The images have not been cropped by coastline, but the area of the Sentinel image for which the soil property models would be valid has been highlighted in the figures. It is worth noting that some properties, such as dissolved organic carbon (DOC) or iron associated with organic matter, could be identified over the sea surface because the methods used in the laboratory were wet extractions.

Although there have been many studies of the soils of the Byers Peninsula [42,43], few have produced valid soil maps for the entire peninsula. Of note is the work of [21], which presents a map of Byers soils based on the study of 23 soil profiles. This article examines the soil profiles by horizon and presents their geolocation on the peninsula. To classify the soils described in Byers, the authors used the work of [15], which describes the soils in terms of the parent material. The main reason for this is that soil properties in maritime Antarctica are closely related to the properties of the parent material due to the incipient degree of chemical weathering. The main soil classes for [21] are soils derived from volcanic tuffs; soils derived from basalts; soils derived from mixed contributions of volcanoclasts (tuffs, breccias) and andesitic basalts in areas dominated by volcanoclasts and andesitic basalts; soils derived from marine mudstones, sandstones, and conglomerates; ornithogenic soils; and cryosols. This classification is very useful for us to interpret our results. However, the map presented by these authors has a legend based on World Reference Base for Soil Resources [44] and does not agree with the distribution of geological units used for the description of the profiles. In this way, the classic works [26,28,29] which include maps of geological units, were very useful to interpret the cartographic patterns of our models of soil properties. In addition, the values of pH, Fe and calcium presented in [45] for the Byers Plateau and [21] for the northern sector of the Byers Peninsula were useful to validate our models.

4.2. Spatial Distribution of Soil Properties

On the one hand, the spatial distribution of our pH values shows the influence of the geological substrate. The highest pH values, in dark blue (Figure 3), are associated with the distribution of basaltic and volcanic rocks [29] and agree with the surface pH values of the profiles of [21]. The pH values below 5 are also consistent with the location of the ornithogenic soils described by these authors.

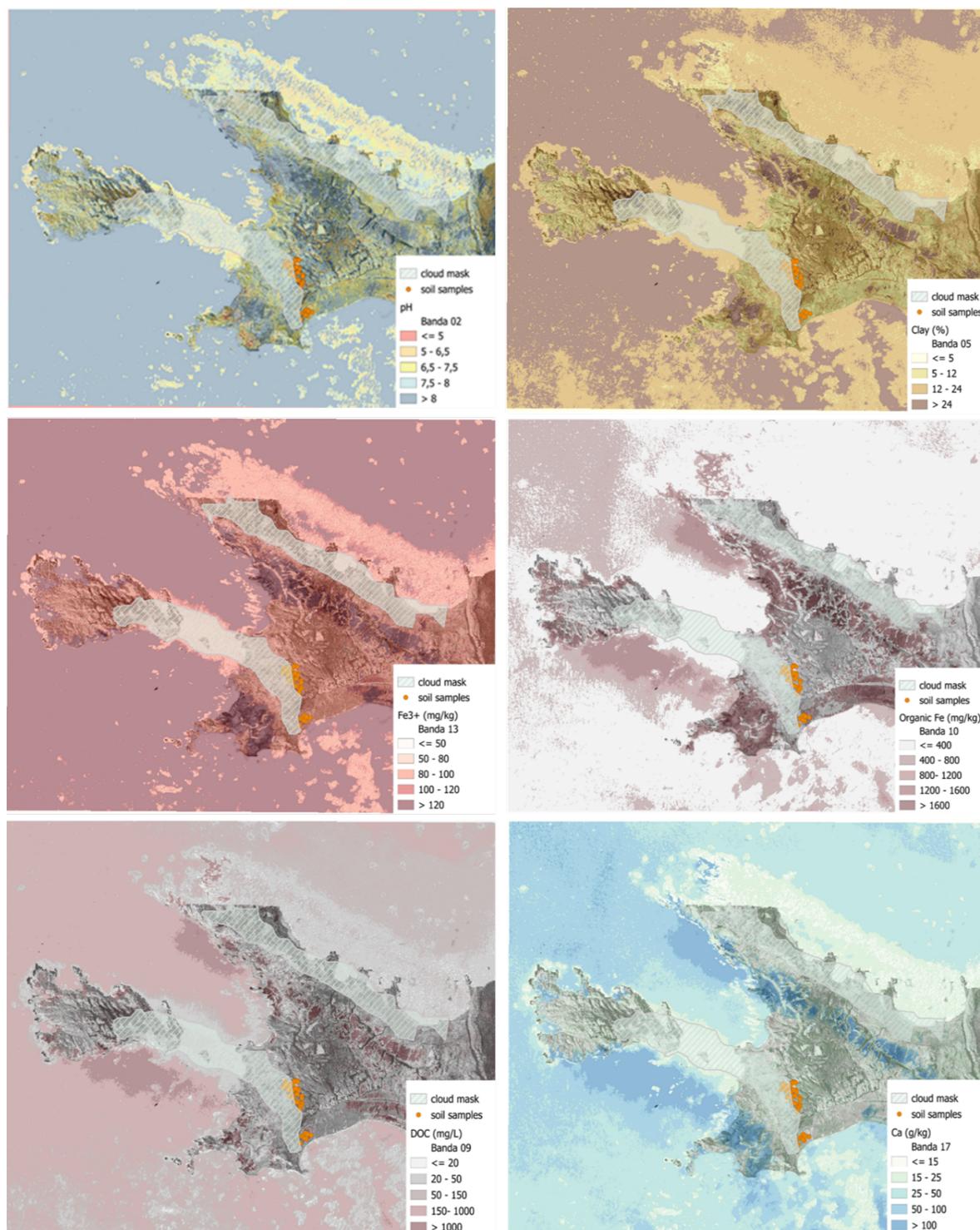


Figure 3. Patterns of soils properties obtained to extended ML soil properties models to entire Byers Peninsula over Sentinel image.

Similarly, the spatial distribution of calcium, iron and clays follows the geological patterns described by [21], with higher values in the areas occupied by basalts and pyroclastic rocks. Calcium has been described in association with basaltic substrates, but also with marine sedimentary rocks. Dissolved organic carbon (DOC) and iron associated with organic matter (Fe) have different distributions. In the case of Fe, the highest values are located in the beaches and in the limit of the coastline. DOC has a similar distribution and is consistent with the greater presence of fauna [12].

The hotspots related to organic carbon and iron associated with organic matter can be explained by the preservation of volcanic ash deposits, which are characterized by their ability to absorb organic carbon [46]. In the geological and geomorphological maps of the Byer Peninsula [26,29], these areas are covered mainly by ash, which exists there in the form of cryoconite with a granular structure, which contributes to the absence of erosion processes [47] and keeps the pH at elevated values above 7.5. The distribution of sand and silt contents is interrelated. The processes of weathering of basaltic and andesitic rocks caused the predominantly sandy structure of the soils. Mapping the soils based on the classical method of soil sampling has strong limitations in the Byers Peninsula. However, there are enough soil samples that have been collected in different projects over time and that can be put together in a geodatabase to train models such as the one presented in this paper.

4.3. Generated Albedo RLM Models

The main objective of this research is to obtain a structural index based on soil properties to calculate the albedo of bare soils in periglacial areas of Antarctica. Table 3 presents the results of two linear regression models (LRM) of albedo formation on Byers Peninsula. One of them uses soil properties as explanatory variables and the other one uses spectral index and sentinel bands (15 variables, Tables 1 and 2). We want to find a structural index of albedo for bare areas based on soil properties and with RLM it is possible to know the weight of the explicative variables in the model. The variables in the RLM were also determined using a backward stepwise selection procedure and Akaike Information Criteria (AIC) to find the model that best explains the data with the least number of parameters. We also built an albedo model with our neural network, which is also the one that best fits the data with an MAE of 6.3 and RMSE of 7.5 (Figure 4). However, the main problem with this approach is the impossibility of knowing the importance of the variables in the model. Using RLM, we can know the weight of the variables in the models. We believe that although RLM is a simple algorithm, it is a powerful tool due to its simplicity and interpretability.

In Table 4 we can see that the linear model that best fits the albedo data is the one with edaphic properties (Adjusted R² 0.80; Std 4.1). In addition, all the explanatory variables of this model are statistically significant. The backward stepwise method selected as significant edaphic variables those closely related to the color of the soils. Calcium or iron associated with organic matter have an important weight in the albedo model. Calcium is the most important property in the albedo model, which is consistent with the white color of this mineral. The RLM albedo model, which uses bands and indices taken directly in each of the samples, has a much poorer fit (Adjusted R² 0.28; Std 7.9). However, all variables in the model are statistically significant except for the B5 band (S_B5; p-level 0.164), which corresponds to the red band, 703.9 nm (Table 1). The other four variables in the model correspond to near infrared bands (S_B6 and S_B7, NIR) and an IR550 spectral index. The method used to select the variables was also backward stepwise.

Table 4. Summary of Linear Regression Models.

RML albedo	Beta	Std.Err. -of Beta	B	Std.Err. - of B	t(44)	p-level
Intercept			-43.255	15.499	-2.791	0.008
H ₂ O pH	0.358	0.158	4.643	2.048	2.267	0.028
DOC(mg/L)	-0.163	0.092	-0.002	0.001	-1.767	0.084
Organic matter Fe(mg/kg)	0.536	0.162	0.013	0.004	3.310	0.002
Ca(g/kg)	0.938	0.089	1.129	0.108	10.491	0.000
R=0.90614835 R ² =0.082110483 Adjusted R ² =0.80484164 F(4.44)=50.489 p<0.0000 Std. Estimation error: 4.1076						
Intercept			24.165	5.021	4.813	0.000
S_IR550	-0.605	0.235	-1.177	0.457	-2.575	0.013
S_B6	-5.694	2.824	-426.339	211.479	-2.016	0.050
S_B7	3.298	1.622	266.015	130.858	2.033	0.048
S_B5	2.231	1.575	150.049	105.936	1.416	0.164
R=0.58198905 R ² =0.33871125 Adjusted R ² =0.27859410 F(4.44)=5.6342 p<0.000095 Std. Estimation error: 7.8955						

These results led us to choose the RLM model built with edaphic properties as the possible structural albedo index we were looking for. Nevertheless, we extended the three models (model built with neural network or ML Sentinel; models built with linear regression, RML soil properties, and RML Sentinel) to the Byers Peninsula to observe the albedo distribution patterns (Figure 4).

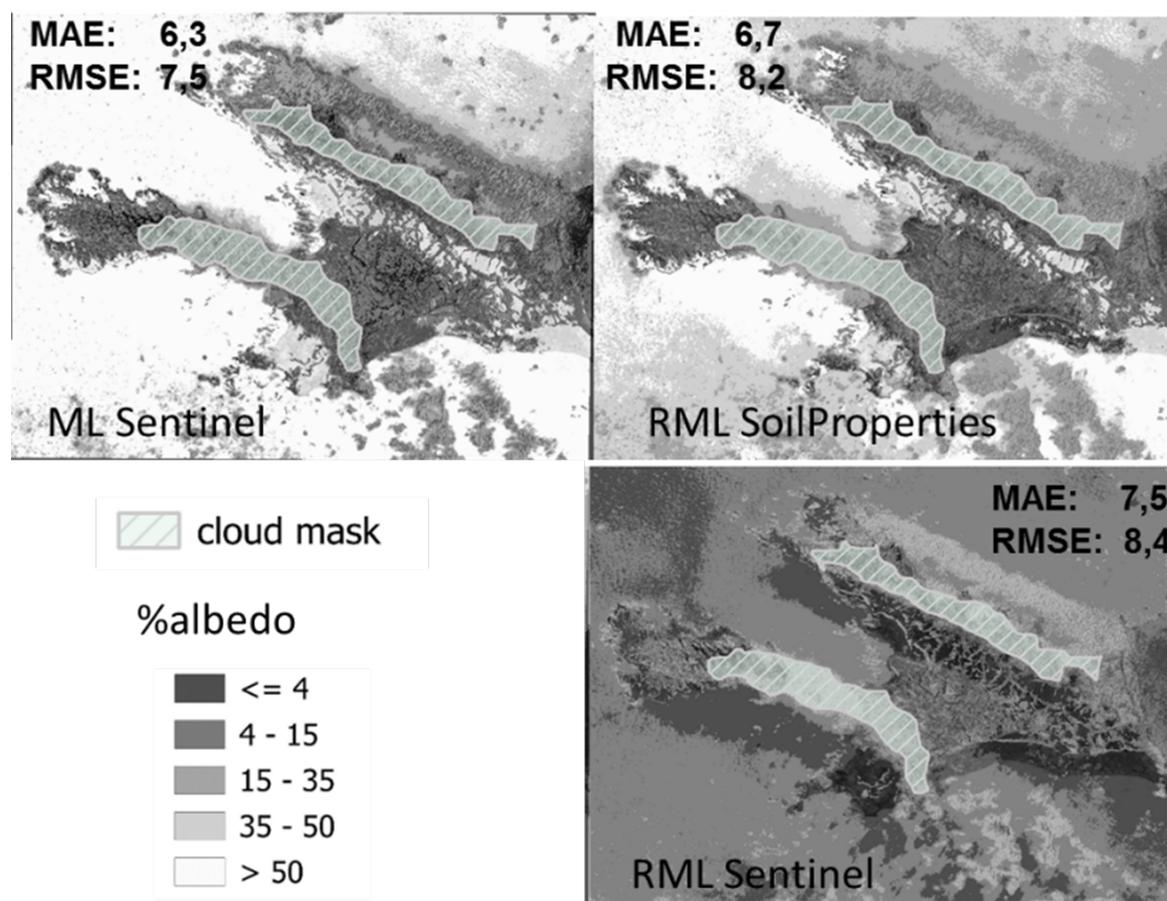


Figure 4. Albedo models extended to entire Byers Peninsula. MAE and RMSE error indexes are represented for each extended model.

The RML model with sentinel bands and indices extended over the entire Byers Peninsula (Figure 4) appears to have an opposite distribution to the other two models. In this model, the highest albedo values (>50) are found along the coast, while the lowest values are associated with areas of higher carbonate concentrations (see calcium model, Figure 3). However, the error measures (MAE 7.5; RMSE 8.4) are similar to those of the other two models. This RMSE is higher than the standard deviation of the linear model (Table 4: standard error of estimate 7.9). Nevertheless, the other two models (ML Sentinel and RML Soil Properties) have a spatial distribution corresponding to the calcium distribution. In particular, the model built with RML soil properties that seems to capture the influence of dissolved organic carbon (DOC) in the areas with low concentration of calcium as beaches. However, despite the fact that the mapping of albedo seems consistent with the distribution of soil properties and with the general geology of the peninsula, the regionalized model has significantly increased the error with respect to the linear model. The RMSE increases from 4.1 in the linear model (Table 4) to 8.2 in the extended model (Figure 4). This may be due to error propagation caused by the use of explanatory variables that are themselves models with an associated standard error (JCGM, 2020). We calculated the percentage of variance of the soil properties represented by the RMSE (Root Mean Square Error) of the estimate for each property included in the RML model (Table 5). The average, 63.3%, represents the increase in the standard error of the albedo estimate (Table 4) caused by the propagation of the

errors due to the use of models as explanatory variables. In the LRM soil properties, the standard error of the estimates is 4.1 (Table 4). The 63.3% of 4.1 is 2.6. The total error could be the sum of 4.1 plus 2.6, which implies 6.7. Nevertheless, the RMSE of the extended model increased up to 8, 2. We interpreted this additional error (1, 5) related to other sources of error.

Table 5. RMSE of the estimations soil properties involved in the RML_soil properties, albedo model, standard deviation of the measured soil properties and percentage of standard deviation of the soil properties represented by the RMSE (Root Mean Square Error).

	pH	DOC (mg/L)	Organic matter Fe (mg/kg)	Ca(g/kg)	Mean
ML_RMSE	0.69	156.2	218.39	5.43	
Std. Dev.	0.72	656.77	374.26	7.62	
% of Std. Dev.	95.8	23.5	58	76	63.3

The fit of the extended models using Sentinel imagery is not very good, but could be significantly improved if higher quality imagery were available. We also know that there are other soil properties that strongly influence albedo, such as moisture [48], but this variable has a high temporal variability and was not the aim of this study. The primary objective of our study was to select those edaphic soil physical and chemical properties of bare soils that most significantly influence albedo in the periglacial environments of maritime Antarctica.

5. Conclusions

It is possible to build models of edaphic properties in periglacial areas of Maritime Antarctica using multispectral optical Sentinel imagery. Only the spectral bands do not allow modeling of edaphic properties, so it is necessary to calculate spectral indices related to chemical composition and color. Machine learning techniques, especially deep learning neural networks, are an appropriate tool to train soil property models. However, these methods do not allow to know the weight of the explanatory variables in the models. The chemical composition of the bare soils in these areas, which have been exposed to edaphogenic processes for a short time, are very similar to the parent material properties. This condition has allowed us to interpret the models of edaphic properties based on the geology of the substrate and the biological activity of the local fauna. In this case, it was very useful to interpret the results of the models on the basis of the geological substrata, given the difficulty of taking soil samples in the field. Regarding the albedo models, it seems that the albedo model made with neural network (ML), is the one that best fits the data. This model provides the lowest values of MAE and RMSE error metric parameters. However, if one of the objectives of modeling is to know which variables best explain the variability of albedo, linear regression models (RML), whenever they can be run, are a very good option. In addition, linear regression makes it easy to analyze the propagation of errors when other models are used as explanatory variables. Finally, the extended albedo models do not provide a good fit to the data, but this can be greatly improved by using more detailed, cloud-free imagery. We used Sentinels because they are freely distributed and have high temporal and radiometric resolution. However, this structural index could be greatly improved in other optical images such as Worldview or Pleiades.

Author Contributions: Conceptualization, S.C., A.P and R.R.; methodology, S.C. and A.P, R.R.; software, R.M.; validation, S.C. and R.M.; formal analysis, S.C, R.M. and J.P.; investigation, S.C., A.P, R.R and J.P.; resources, R.M. and J.F.C.; data curation, S.C. and R.M.; writing—original draft preparation, S.C and R.M.; writing—review and editing, S.C. and R.M.; visualization, S.C.; supervision, J.P and J.F.C.; project administration, S.C and J.F.C.; funding acquisition, S.C. and J.F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by Ministerio de Ciencia e Innovación: PID2021-127060OB-I00, PID2020-113051RB-C31, CTM2017-84441-R, and CTM2014-52021-R.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study, as stated in the paper.

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

References

1. Helsen, M.M.; van de Wal, R.S.W.; Reerink, T.J.; Bintanja, R.; Madsen, M.S.; Yang, S.; Li, Q.; Zhang, Q. On the importance of the albedo parameterization for the mass balance of the Greenland ice sheet in EC-Earth. *The Cryosphere* **2017**, *11*, 1949–1965. doi:10.5194/tc-11-1949-2017.
2. Liang, S.; Wang, J. Chapter 6 - Broadband albedo. In *Advanced Remote Sensing (Second Edition)*, Second Edition ed.; Academic Press, 2020; pp. 193–250. doi:https://doi.org/10.1016/B978-0-12-815826-5.00006-4.
3. He, T.; Wang, D.; Qu, Y. 5.06 - Land Surface Albedo. In *Comprehensive Remote Sensing*; Liang, S., Ed.; Elsevier: Oxford, 2018; pp. 140–162. doi:https://doi.org/10.1016/B978-0-12-409548-9.10370-7.
4. Jakobs, C.L.; Reijmer, C.H.; van den Broeke, M.R.; Van de Berg, W.; Van Wessem, J. Spatial Variability of the Snowmelt-Albedo Feedback in Antarctica. *Journal of Geophysical Research: Earth Surface* **2021**, *126*, e2020JF005696.
5. Calleja, J.F.; Muñoz, R.; Fernández, S.; Corbea-Pérez, A.; Peón, J.; Otero, J.; Navarro, F. Snow albedo seasonal decay and its relation with shortwave radiation, surface temperature and topography over an Antarctic ice cap. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, *14*, 2162–2172.
6. Post, D.F.; Fimbres, A.; Matthias, A.D.; Sano, E.E.; Accioly, L.; Batchily, A.K.; Ferreira, L. Predicting Soil Albedo from Soil Color and Spectral Reflectance Data. *Soil Science Society of America Journal* **2000**, *64*, 1027–1034.
7. Brooks, S.T.; Jabour, J.; Van Den Hoff, J.; Bergstrom, D.M. Our footprint on Antarctica competes with nature for rare ice-free land. *Nature Sustainability* **2019**, *2*, 185–190.
8. Braghieri, R.; Wang, Y.; Gagné-Landmann, A.; Brodrick, P.; Bloom, A.; Norton, A.; Ma, S.; Levine, P.; Longo, M.; Deck, K.; others. The importance of hyperspectral soil albedo information for improving Earth system model projections. *AGU Advances* **2023**, *4*, e2023AV000910.
9. Tejedo, P.; Justel, A.; Benayas, J.; Rico, E.; Convey, P.; Quesada, A. Soil trampling in an Antarctic Specially Protected Area: tools to assess levels of human impact. *Antarctic Science* **2009**, *21*, 229–236.
10. Pertierra, L.; Santos-Martin, F.; Hughes, K.; Avila, C.; Caceres, J.; De Filippo, D.; Gonzalez, S.; Grant, S.; Lynch, H.; Marina-Montes, C.; Quesada, A.; Tejedo, P.; Tin, T.; Benayas, J. Ecosystem services in Antarctica: Global assessment of the current state, future challenges and managing opportunities. *Ecosystem Services* **2021**, *49*, 101299. doi:https://doi.org/10.1016/j.ecoser.2021.101299.
11. Roudier, P.; O'Neill, T.A.; Almond, P.; Poirot, C. Soil priorities for Antarctica. *Geoderma Regional* **2022**, *29*, e00499.
12. Abakumov, E.; Gagarina, E.; Sapega, V.; Vlasov, D.Y. Micromorphological features of the fine earth and skeletal fractions of soils of West Antarctica in the areas of Russian Antarctic stations. *Eurasian Soil Science* **2013**, *46*, 1219–1229.
13. Alekseev, I.; Abakumov, E. Soil organic matter and biogenic-abiogenic interactions in soils of Larsemann Hills and Bunger Hills, East Antarctica. *Polar Science* **2023**, p. 101040. doi:https://doi.org/10.1016/j.polar.2023.101040.
14. Lacerda, A.L.d.F.; Rodrigues, L.d.S.; Van Sebille, E.; Rodrigues, F.L.; Ribeiro, L.; Secchi, E.R.; Kessler, F.; Proietti, M.C. Plastics in sea surface waters around the Antarctic Peninsula. *Scientific reports* **2019**, *9*, 3977.
15. Simas, F.N.B.; Schaefer, C.E.G.R.; Michel, R.F.; Francelino, M.R.; Bockheim, J.G., Soils of the South Orkney and South Shetland Islands, Antarctica. In *The Soils of Antarctica*; Springer International Publishing: Cham, 2015; pp. 227–273. doi:10.1007/978-3-319-05497-1_13.
16. Siqueira, R.G.; Moquedace, C.M.; Francelino, M.R.; Schaefer, C.E.; Fernandes-Filho, E.I. Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma* **2023**, *432*, 116405. doi:https://doi.org/10.1016/j.geoderma.2023.116405.
17. Campbell, I.; Claridge, G. *Antarctica: Soils, Weathering Processes and Environment*; ISSN, Elsevier Science, 1987.
18. Ugolini, F.; Bockheim, J. Antarctic soils and soil formation in a changing environment: A review. *Geoderma* **2008**, *144*, 1–8. Antarctic Soils and Soil Forming Processes in a Changing Environment, doi:https://doi.org/10.1016/j.geoderma.2007.10.005.

19. Otero, X.; Fernández, S.; de Pablo Hernandez, M.; Nizoli, E.; Quesada, A. Plant communities as a key factor in biogeochemical processes involving micronutrients (Fe, Mn, Co, and Cu) in Antarctic soils (Byers Peninsula, maritime Antarctica). *Geoderma* **2013**, *195*, 145–154.
20. Beyer, L. Properties, formation, and geo-ecological significance of organic soils in the coastal region of East Antarctica (Wilkes Land). *CATENA* **2000**, *39*, 79–93. doi:[https://doi.org/10.1016/S0341-8162\(99\)00090-9](https://doi.org/10.1016/S0341-8162(99)00090-9).
21. Moura, P.A.; Francelino, M.R.; Schaefer, C.E.G.; Simas, F.N.; de Mendonça, B.A. Distribution and characterization of soils and landform relationships in Byers Peninsula, Livingston Island, Maritime Antarctica. *Geomorphology* **2012**, *155-156*, 45–54. Advances in Permafrost and Periglacial research in Antarctica, doi:<https://doi.org/10.1016/j.geomorph.2011.12.011>.
22. Wang, Y.; Shao, M. Spatial variability of soil physical properties in a region of the Loess Plateau of PR China subject to wind and water erosion. *Land Degradation & Development* **2013**, *24*, 296–304.
23. Shit, P.K.; Bhunia, G.S.; Maiti, R. Spatial analysis of soil properties using GIS based geostatistics models. *Modeling Earth Systems and Environment* **2016**, *2*, 1–6.
24. Benayas, J.; Pertierra, L.R.; Tejado, P.; Lara, F.; Bermudez, O.; Hughes, K.A.; Quesada, A. A review of scientific research trends within ASPA No. 126 Byers Peninsula, South Shetland Islands, Antarctica. *Antarctic Science* **2013**, *25*, 128 – 145.
25. Rakusa-Suszczewski, S., King George Island — South Shetland Islands, Maritime Antarctic. In *Geoecology of Antarctic Ice-Free Coastal Landscapes*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2002; pp. 23–39. doi:10.1007/978-3-642-56318-8_3.
26. Lopez-Martinez, J.; Thomson, M.; Arche, A.; Björck, S.; Ellis-Evans, J.C.; Hatway, B.; Hernández-Cifuentes, F.; Hjort, C.; Ingólfsson, O.; Ising, J.; others. Geomorphological map of Byers peninsula, livingston island. *British Antarctic Survey* **1996**.
27. Björck, S.; Hjort, C.; Ingólfsson, O.; Zale, R.; Ising, J. Holocene deglaciation chronology from lake sediments. *Geomorphological map of Byers Peninsula, Livingston Island. BAS GEOMAP Series, Sheet* **1996**, pp. 49–51.
28. Hobbs, G. The geology of the South Shetland Islands: IV. The geology of Livingston Island. *British Antarctic Survey* **1968**.
29. Hathway, B. Nonmarine sedimentation in an Early Cretaceous extensional continental-margin arc, Byers Peninsula, Livingston Island, South Shetland Islands. *Journal of Sedimentary Research* **1997**, *67*, 686–697.
30. Simms, A.R.; Milliken, K.T.; Anderson, J.B.; Wellner, J.S. The marine record of deglaciation of the South Shetland Islands, Antarctica since the Last Glacial Maximum. *Quaternary Science Reviews* **2011**, *30*, 1583–1601. doi:<https://doi.org/10.1016/j.quascirev.2011.03.018>.
31. van Zinderen Bakker, E.M. *Palaeoecology of Africa & of the Surrounding Islands & Antarctica*; Vol. 16, AA Balkema, 1966.
32. Mars, J.C.; Rowan, L.C. ASTER spectral analysis and lithologic mapping of the Khanneshin carbonatite volcano, Afghanistan. *Geosphere* **2011**, *7*, 276–289, [<https://pubs.geoscienceworld.org/gsa/geosphere/article-pdf/7/1/276/3341201/276.pdf>]. doi:10.1130/GES00630.1.
33. Escadafal, R.; Girard, M.C.; Courault, D. Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data. *Remote Sensing of Environment* **1989**, *27*, 37–46. doi:[https://doi.org/10.1016/0034-4257\(89\)90035-7](https://doi.org/10.1016/0034-4257(89)90035-7).
34. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* **1996**, *58*, 289–298. doi:[https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7).
35. Misra, A.; Prasad, R.C.; Chauhan, V.S.; Srilakshmi, B. A theoretical model for the electromagnetic radiation emission during plastic deformation and crack propagation in metallic materials. *International journal of fracture* **2007**, *145*, 99–121.
36. Hewson, R.; Robson, D.; Mauger, A.; Cudahy, T.; Thomas, M.; Jones, S. Using the Geoscience Australia-CSIRO ASTER maps and airborne geophysics to explore Australian geoscience. *Journal of Spatial Science* **2015**, *60*, 207–231.
37. Murtagh, F. Multilayer perceptrons for classification and regression. *Neurocomputing* **1991**, *2*, 183–197.
38. Muñoz, R.; Cuevas-Valdés, M.; de la Roza-Delgado, B. Milk quality control requirement evaluation using a handheld near infrared reflectance spectrophotometer and a bespoke mobile application. *Journal of Food Composition and Analysis* **2020**, *86*, 103388.

39. Gaudart, J.; Giusiano, B.; Huiart, L. Comparison of the performance of multi-layer perceptron and linear regression for epidemiological data. *Computational statistics & data analysis* **2004**, *44*, 547–570.
40. Hodson, T.O.; Over, T.M.; Foks, S.S. Mean squared error, deconstructed. *Journal of Advances in Modeling Earth Systems* **2021**, *13*, e2021MS002681.
41. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.
42. Navas, A.; López-Martínez, J.; Casas, J.; Machín, J.; Durán, J.J.; Serrano, E.; Cuchi, J.A. Soil characteristics along a transect on raised marine surfaces on Byers Peninsula, Livingston Island, South Shetland Islands. *Antarctica: contributions to global earth sciences* **2006**, pp. 467–473.
43. Navas, A.; López-Martínez, J.; Casas, J.; Machín, J.; Durán, J.J.; Serrano, E.; Cuchi, J.A.; Mink, S. Soil characteristics on varying lithological substrates in the South Shetland Islands, maritime Antarctica. *Geoderma* **2008**, *144*, 123–139.
44. World Reference Base. <https://www.fao.org/soils-portal/data-hub/soil-classification/world-reference-base/en/>. [Accessed 21-02-2024].
45. Otero, X.L.; Fernandez, S.; Hernández, M.A.; Nizoli, É.C.; Quesada, A. Plant communities as a key factor in biogeochemical processes involving micronutrients (Fe, Mn, Co, and Cu) in Antarctic soils (Byers Peninsula, maritime Antarctica). *Geoderma* **2013**, *195*, 145–154.
46. Ciriminna, R.; Scurria, A.; Tizza, G.; Pagliaro, M. Volcanic ash as multi-nutrient mineral fertilizer: Science and early applications. *JSEA reports* **2022**, *2*, 528–534, [<https://onlinelibrary.wiley.com/doi/pdf/10.1002/jsf2.87>]. doi:<https://doi.org/10.1002/jsf2.87>.
47. Suleymanov, A.; Abakumov, E.; Alekseev, I.; Nizamutdinov, T. Digital mapping of soil properties in the high latitudes of Russia using sparse data. *Geoderma Regional* **2024**, *36*, e00776. doi:<https://doi.org/10.1016/j.geodrs.2024.e00776>.
48. Li, Z.; Yang, J.; Gao, X.; Yu, Y.; Zheng, Z.; Liu, R.; Wang, C.; Hou, X.; Wei, Z. The relationship between surface spectral albedo and soil moisture in an arid Gobi area. *Theoretical and Applied Climatology* **2019**, *136*, 1475–1482.

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.