Soil pedogeochemical attributes prediction by interpolators in ice-free areas of Antarctica

Predição de atributos pedogeoquímicos do solo por interpoladores em áreas livres de gelo da Antártica

Predicción de atributos pedogeoquímicos del suelo por interpoladores en áreas libres de hielo de la Antártida

Adriano Luis Schünemann
ORCID: https://orcid.org/0000-0001-7227-7074
Universidade Federal do Pampa, Brazil
E-mail: adrianoschunemann@unipampa.edu.br
André Thomazini
ORCID: https://orcid.org/0000-0001-5613-0494
Universidade Federal de São João Del-Rei, Brazil
E-mail: andre.thz@gmail.com
Pedro Henrique Araújo Almeida
ORCID: https://orcid.org/0000-0002-7166-0693
Universidade Federal de Viçosa, Brazil
E-mail:phalmeida06@gmail.com
Márcio Rocha Francelino
ORCID: https://orcid.org/0000-0001-8837-1372
Universidade Federal de Viçosa, Brazil
E-mail: márcio.francelino@gamail.com
Elpídio Inácio Fernandes Filho
ORCID: https://orcid.org/0000-0002-9484-1411
Universidade Federal de Viçosa, Brazil
E-mail: elpidio@ufv.br
Gérson Rodrigues dos Santos
ORCID: https://orcid.org/0000-0002-4306-8334
Universidade Federal de Viçosa, Brazil
E-mail: gerson.santos@ufv.br
Mayara Daher de Paula
ORCID: https://orcid.org/0000-0003-4937-2289
Universidade Federal de Viçosa, Brazil
E-mail: Mayara.dahher@gmail.com
Carlos Ernesto Gonçalves Reynaud Schaefer
ORCID: https://orcid.org/0000-0001-7060-1598
Universidade Federal de Viçosa, Brazil
E-mail: carlos.schaefer@ufv.br
Antônio Batista Pereira
ORCID: https://orcid.org/0000-0003-0368-4594
Universidade Federal do Pampa, Brazil
E-mail:ambatistape@gmail.com

Abstract
The main objective of this paper is to predict soil attributes in unsampled areas using geostatistical models, By improving the prediction parameters of selected data, using environmental covariates characteristic of Antarctic ice free areas. In this study, 58 soil samples from a grid were collected at 0-10 cm depth in Keller Peninsula, King George Island, Antarctica. The soil chemical analysis was performed, and the values of potassium, calcium and magnesium were determined for each soil sampled. Simple kriging (SK) and Random Forest interpolator were used in this work to predict the values of the studied soil attributes in non-sampled areas. We used a Terrestrial Laser Scanner (TLS) to generate a cloud of points, to obtain digital elevation models (DEMs) of 1, 5, 10, 20 and 30 meters cell size. The use of covariates did not improve the parameters of soil bases prediction in the studied area. The final maps did not show great differences based on RMSEs, mainly related to the great spatial variability of soil attributes in this region, despite soil thematic maps evidencing visual difference.
Keywords: Predictive covariates; Interpolation; Digital mapping.

Resumo
Este trabalho tem como objetivo principal predizer atributos do solo em áreas não amostradas utilizando modelos geoestatísticos, através da melhora nos parâmetros de predição dos dados selecionados, utilizando covariáveis ambientais características de áreas de gelo da Antártica. Neste estudo, 58 amostras de solo de uma grade foram coletadas a 0-10 cm de profundidade na Península Keller, Ilha Rei George, Antártica. Foi realizada a análise química do solo e determinados os valores de potássio, cálcio e magnésio para cada solo amostrado. A krigagem simples (SK) e o interpolador *Random Forest* foram utilizados neste trabalho para predizer os valores dos atributos do solo estudados em áreas não amostradas. Usamos um *Terrestrial Laser Scanner* (TLS) para gerar uma nuvem de pontos, para obter modelos digitais de elevação (DEMs) de tamanho de célula de 1, 5, 10, 20 e 30 metros. O uso de covariáveis não melhorou os parâmetros de predição de bases do solo na área estudada. Os mapas finais não apresentaram grandes diferenças com base em RMSEs, principalmente relacionadas à grande variabilidade espacial dos atributos do solo nesta região, apesar dos mapas temáticos de solo evidenciarem diferença visual. O uso de mapas com melhor resolução obtidos de TLS não melhora a previsão usando *Random Forest* para Ca2+. Os interpoladores de previsão de solo podem ser aplicados para determinar os atributos do solo em áreas não amostradas, mas áreas com alta complexidade precisam de uma grade de amostragem mais densa para melhorar o desempenho da previsão.
Palavras-chave: Covariáveis preditivas; Interpolação; Mapeamento digital.

Resumen
El objetivo principal de este trabajo es predecir los atributos del suelo en áreas no muestreadas utilizando modelos geoestadísticos, mejorando los parámetros de predicción de los datos seleccionados, utilizando covariables ambientales características de las áreas de hielo antárticas. En este estudio, se recolectaron 58 muestras de suelo de una cuadrícula a una profundidad de 0 a 10 cm en la Península Keller, Isla Rey Jorge, Antártida. Se realizó el análisis químico del suelo y se determinaron los valores de potasio, calcio y magnesio para cada suelo muestreado. En este trabajo se utilizaron interpoladores de *kriging simple* (SK) y *Random Forest* para predecir los valores de los atributos del suelo estudiados en áreas no muestreadas. Utilizamos un escáner láser terrestre (TLS) para generar una nube de puntos, para obtener modelos digitales de elevación (DEM) de tamaño de celda de 1, 5, 10, 20 y 30 metros. El uso de covariables no mejoró los parámetros de predicción de bases de suelo en el área de estudio. Los mapas finales no mostraron grandes diferencias basadas en RMSE, principalmente relacionadas con la gran variabilidad espacial de los atributos del suelo en esta región, a pesar de que los mapas temáticos de suelos evidencian diferencia visual. El uso de mapas con mejor resolución obtenidos de TLS no mejora la predicción usando *Random Forest* para Ca2+. Los interpoladores de predicción del suelo se pueden aplicar para determinar los atributos del suelo en áreas no muestreadas, pero las áreas con alta complejidad necesitan una cuadrícula de muestreo más densa para mejorar el rendimiento de la predicción.
Palabras clave: Covariables predictivas; Interpolación; Mapeo digital.

1. Introduction

Terrestrial ecosystems of Antarctica are characterized as a heterogeneous soil environment (Thomazini et al., 2014). This characteristic is mainly related to the active periglacial processes action (i.e. glacier melting, rising temperatures, colonization by animals and vegetation, guano deposition) (Mendonça et al., 2010; Thomazini et al., 2014). The heterogeneity in vegetation distribution and soil patches may increase as periglacial processes are enhanced in the current warming scenario. However, the characterization of the spatial variability of soil attributes and its controlling factors is little studied in Maritime Antarctica.

Some geostatistical models have been applied to characterize the spatial variability of soil attributes. The most used geostatistical methods are kriging. It has many advantages over other interpolation methods, such as providing on average estimates without bias and minimum variance, in addition to being able to know the variance of the estimate (Ferreira, et al. 2013). Kriging method can be highly supportive in the prediction of soil attributes, especially when point based field measurements are available (Wasklewicz et al., 2013).

There are several types of kriging, such as simple kriging, ordinary kriging, universal kriging, co-kriging, among others. The most used types are simple and ordinary kriging (Webster and Oliver, 1990), the first being used when the average is assumed to be known throughout the study area and the second considers the floating or moving average (Ferreira, et al. 2013).

Kriging methods do not use any selection mechanism of variables, such as decision tree. Random Forest (RF) is an

alternative, as it uses decision trees tools. Breiman and Breiman (2001), it is a conceptualize the method known as Random Forest, method with as successive decision trees that are developed by the introduction of random elements in their construction. This statistical method can be used to predict both categorical and continuous data.

Random Forest shows some theoretical advantages over kriging techniques, as it can reduce the sample size and predictive covariates as well as can be collected by remote sensors. This method is important to perform surveys in remote areas, that is not possible to access or when logistical costs are expensive to obtain detailed field data, such as Antarctica. Soil sampling in ice-free areas in Antarctica is greatly hard to perform, mainly due to heavy weather conditions, such as significant increase on wind speed, low temperatures as well as soil freezing.

Kirkwood et al. (2016) showed that auxiliary variables (covariates) can be introduced to predict geochemical attributes, such as gamma spectrometry and data from other surveys. In this case, data obtained by LIDAR (Light Detection and Ranging) surveys, airborne and/or Terrestrial Laser Scanner (TLS), which collect points with three-dimensional coordinates can be a powerful tool to generate environmental covariates and improve prediction parameters (Schünemann et al., 2018), especially in areas where field work is difficult to perform.
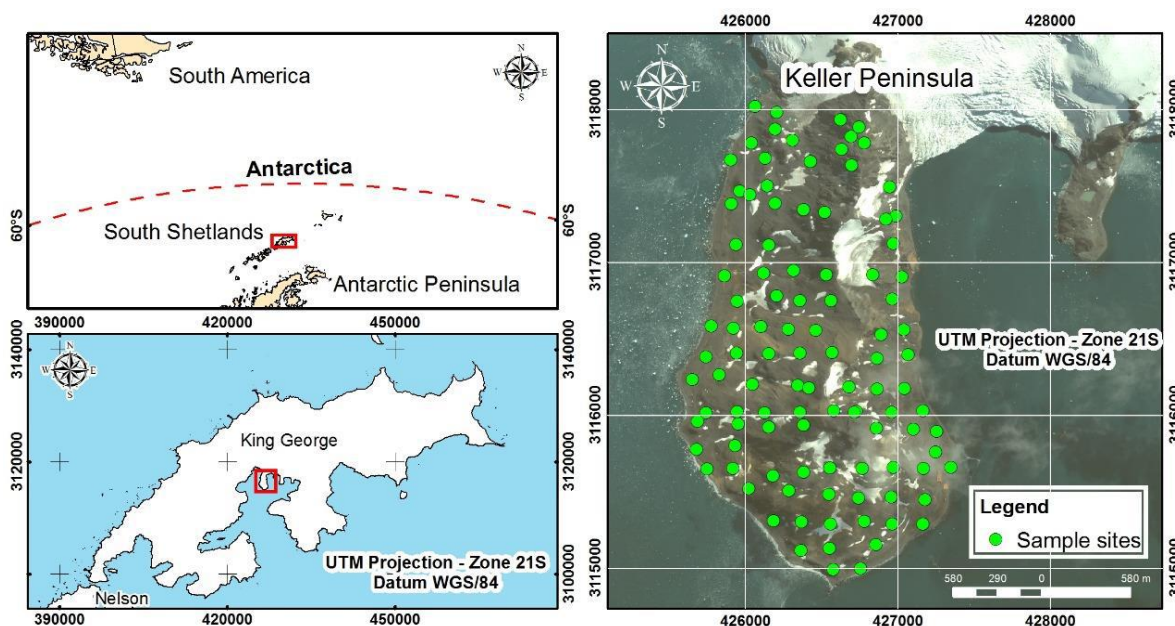
In this context, the objective of this study was to predict soil attributes in non-sampled areas using geostatistical models, as well as to verify the improvement on prediction parameters of the selected attributes by using environmental covariate in ice-free areas of Antarctica.

2. Material and methods

2.1 Study area

The study was carried out in Keller Peninsula, King George Island, located in the South Shetlands Archipelago, maritime Antarctica (Figure 1). Keller Peninsula com- prises approximately 500 ha, being 4×2 km of length and width respectively (Francelino et al., 2011). According to the meteorological station located at Brazilian Commandant Ferraz ($62°4′33″$ S, $58°23′46″$ W), mean air temperatures from 1986 to 2013 are reported to be +1.6 during the summer (December–March) and −5.3 °C during the winter (June–September) (INPE, 2015). Mean annual precipitation is about 400mm. The altitude ranges from 0 to 340 m above the sea level (Francelino et al., 2011). Outcrops of piritized andesites occur and generates acid sulphated sediments (Simas et al., 2006). The main soil classes on the peninsula are Leptosol, Regosol, Cryosol and Cambiso. About 17% of the peninsula is covered by some soil typel, the other areas are occupied by glaciers, snow bank, rock outcrop and unconsolidated sediments (Francelino et al. 2011).

Figure 1. Illustrative map of South Shetlands Archipelago (a), where King George Island is localized (b) and sampling points in Keller Peninsula (c).

Source: Authors.

2.2 Soil Sampling and soil attributes determinations

Soil samples were collected during the summer of 2012 and 2015 at 0-10 and 10-30 cm depth, to determine pedogeochemical attributes. Soil samples were collected spatially dispersed, in the form of a semi-regular grid (Fig. 1c). Soil samples were air-dried, grounded, and sieved through a 2 mm mesh. Exchangeable Ca^{2+} and Mg^{2+} were extracted with 1 mol L^{-1} KCl and element contents in the extracts were determined by atomic absorption. K^+ were extracted with Melich-1 and element contents quantified by flame emission photometry (Embrapa, 2011).

2.3 Predictive interpolator

Random Forest (RF) with cell sizes of 1, 5, 10, 20 and 30 m and simple kriging (SK) with a cell size of 30 m were used in this study as predictive interpolators of pedogeochemical attributes. All analyses were performed in R program (R Core Team, 2019).

Only sample values of Ca^{2+}, Mg^{2+} and K^+ obtained from the chemical analysis were used in the SK modeling. It was necessary to obtain a local grid, where the values predicted by the interpolator were marked to obtain the predictive map. This grid was made from digital elevation models (DEMs) obtained from a cloud of points, generated by a RIEGL VZ-1000 Terrestrial Laser Scanner (TLS), during the summer of 2015. Grids were produced from the cloud containing approximately 270 million points, generating DEMs with 1, 5, 10, 20 and 30 m of geometric resolution.

The RF uses sample values of Ca^{2+}, Mg^{2+} and K^+ and covariates. In this work, the covariates used were lithology (Francelino et al., 2011), geomorphology maps (Francelino et al., 2011), Euclidian distance of bird nests present in the Keller Peninsula, generated in the program ArcGIS 10.1 program from the data obtained by Carneiro et al. (2010), DEM generated by the TLS as well as the morphometric covariates derived from the DEM.

A script was elaborated in the R program to obtain the covariates from each DEM with five different spatial resolutions and different morphometric themes, adding a total of 50 covariables (Tables 1, 2 and 3). The "RSAGA" (Brenning, 2008) and "raster" (Hijmans, 2015) packages were used to automatically obtain the covariates. The other tables and figures are

based on the treatment of data obtained in the field and in the laboratory by the author and the support team.

Table 1. Morphometric covariates of the curvatures group.

Nº	Morphometric variables	Brief description
	Curvature derivatives	
1	Section Curvature.	
2	Flow line curvature.	
3	General curvature.	
4	Longitudinal curvature.	
5	Maximum curvature	in the normal local section.
6	Minimal curvature	referring to the normal local section.
7	Flat curvature.	
8	Profile curvature	describes the second accumulation mechanism.
9.	Tangential curvature.	describes the first accumulation mechanism.
10.	Curve classification.	classifies curves into eight classes TOTAC = MEANC2 – DIFFC2
11.	Total curvature.	TOTAC = MEANC2 – DIFFC (average curvature less differences in curvatures)

Source: Authors.

Table 2. Morphometric covariates of the local solar radiation group.

Nº	Morphometric variables	Brief description
	Derived from solar radiation	
12	Solrad diffuse 1	Diffuse solar radiation incident in January
13	Solrad diffuse 2	Diffuse solar radiation incident in June
14	Solrad diffuse 1	Diffuse solar radiation incident in January
15.	Solrad diffuse 2	Diffuse solar radiation incident in June
16	Solrad diffuse 1	Diffuse solar radiation incident in January
17	Solrad diffuse 2	Diffuse solar radiation incident in June
18	Solrad diffuse 1	Total radiation (direct +diffuse) incident in January
19.	Solrad diffuse 2	Total radiation (direct +diffuse) incident in June

Source: Authors.

Table 3. Morphometric covariates of the others groups.

	Other covariates	
Nº Morphometric variables	Brief description
20 Relief face exposition	Orientation in relation to cardinal points of each face of the relief;
21 Convergence index	Calculates a convergence / divergence ratio in relation to runoff;
22 Difference	Difference in the gradient slope;
23 Daytime anisotropic heating	Represents the asymmetry in heating, comparing east-west face based on hours of radiation;
24 Gradient	Corresponds to the hydraulic gradient;
25 Mass balance index	Index representing the mass balance of each pixel;
26 Digital elevation model	Represents the elevation in each cell of the model;
27 Half slope position	Represents the top valley distance, ranging from 0 to 1;
28 Multiresolution index flatness - top	Uses flatness and valley features, varies 0 -1;
29 Normalized elevation	It is a measure of relative height. A value of 1 is the highest point and 0 to the lowest.
30 Real surface area	Considers the real surface and not the projected;
31 Declivity	Considers the actual surface and not the projected;
32 Slope rise	
33 Standardized elevation	It is the product of the normalized height multiplied by the absolute height
34 Specific points surface	Indicates the difference in specific points of surface change;
35 Land roughness index	Calculates the difference in elevation values from a center cell and eight Neighboring cells;
36 convexity of the Earth's surface	Is calculated as the ratio of the number of cells with the positive curvature to the number of all valid cells;
37 Earth surface texture	Divides the terrestrial texture into 8, 12 or 16 classes;
38 Depth of the valleys.	Reverses the elevation, drifts the drainage network
39. Roughness measurement value	Measures the variation in the roughness of the terrain.
40 Topographic unit index	Describes the tendency of each cell to accumulate water in function of the relief.

Source: Authors.

SK processing was performed from a script for R program with the following packages: "geoR", "moments", "scatterplot3d", "tcltk2", "sp", "rgdal" and "RSAGA". The descriptive statistics and normality test were performed using the Shapiro and Wilk test

The first processing was the preparation of data set, determining the maximum distance that was considered to adjust data range and experimental semivariogram. At this step, it was performed interactively in an omnidirectional way by changing the maximum distance values and verifying the semivariogram generated visually. Then, anisotropy analysis was performed, plotting the semivariograms in four directions: 0º, 45º, 90º, and 135º. The similarity between the four variograms allows the observation of anisotropy behavior of the studied variables. The adjustment of theoretical models was performed visually and interactively, where the best values of step (C), range (a), nugget effect (C_0) and the theoretical mathematical model were determined. Data were adjusted using the Ordinary Least Squares (OLS).

Data validation was performed to identify the quality of the method used (OLS), with the following indicators: standardized error means (SEM) and standardized error deviation (SED). The SEM is better the closer to the zero value and the SED is better the closer to the value 1 (Vieira, 2000). Another indicator of data quality is the variance of errors, which can be interpreted as the kriging variance (Webster and Oliver, 1990). Vieira (2000) suggests that a linear regression fitted between the values observed and predicted by the model, ideally shows values of β_o = zero and β_1 =1.

The predicted map was exported in the ASCII format. For both SK and RF, root mean square error (RMSE) was

performed through the cross-validation known as "leave-one-out," based on the following Eq. 1 (Willmott and Matsuura, 2005):

(1) $$AE = [n^{-1}\sum_{i=1}^{n}|e_i|^2]^{1/2}$$

Where MAE =mean absolute error; where is the root mean square error and is the error of each point from the "leave-one-out" difference.

SK processing was also performed from a script for R program with the following packages: "gstat", "randomForest", "raster", "sp", "rgdal" and "SDMTools".

During the predictor training tests, values of each covariate were extracted at the sites corresponding to the sampling points and saved in a training matrix. This new matrix associated with sampled values, was carried out the first round of the interpolator. After processing, the interpolator provides an order importance value of each covariate. The higher values the most important is. One of the indicators of this process is the percentage of the variable that is explained with all the covariates being used. The most important variables were selected by ordering. The predicted maps were exported in the ASCII format.

Predicted maps were compared by the degree of conformity. The grid of 30 m was used as spatial reference to generate the points (each cell). The cell values of all generated rasters were extracted. With the vectors produced from the extraction and excluding the lines that had missing values, the degree of agreement between each pair of maps was calculated.

The quantitative comparison between the predicted Ca^{2+}, Mg^{2+} and K^+ (predicted maps) values for RF and KS was performed through the Mean Absolute Error (MAE) (Willmott and Matsuura, 2005).

2) $$MAE = [n^{-1}\sum_{i=1}^{n}|e_i|]$$

This error is considered by the difference between all values of the predicted maps (4670 cells) with 30 m size.

3. Results

It is important to emphasize initially that this is an unprecedented work for the studied areas and for the equipment used. In addition, due to extreme and unfavorable environmental conditions, where meteorological conditions reduce the periods of data collection in the field, associated with the reduced permanence of researchers in Antarctica, modeling work is very useful for future research, establishing predictable parameters to be validated.
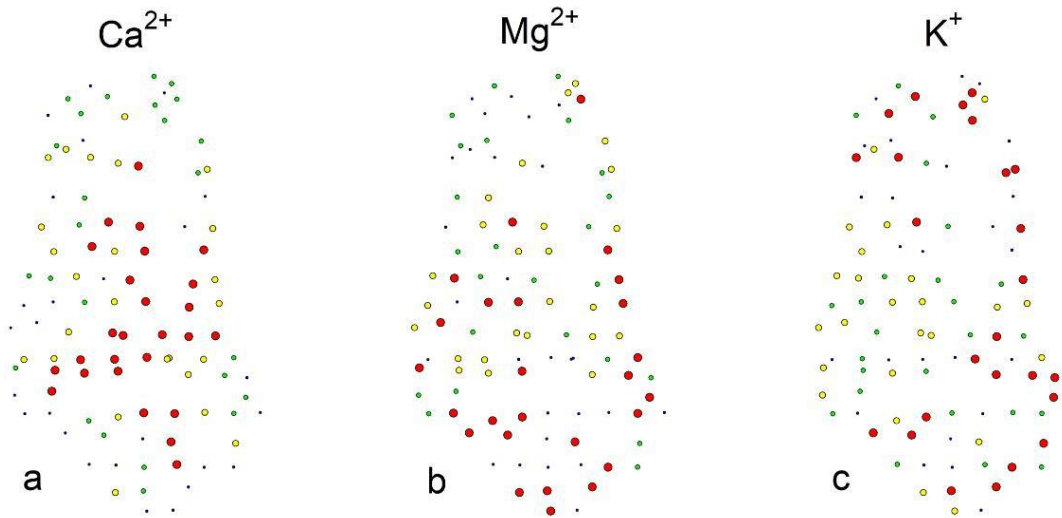
3.1 Sampling description

Soil attributes values are presented in table 4. The quartile distribution represents how the values of the determined variables were distributed accordingly to the data set in the study area (Figure 2). No treatment was performed to eliminate outliers' values predicted by the interpolators during data processing.

Table 4. Ca^{2+}, Mg^{2+} and K^+ contents obtained from the 102 soil samples in Keller Peninsula.

Nº	Long. E (m)	Lat. N (m)	Ca^{2+} ($cmol_c\ dm^{-3}$)	Mg^{2+} ($cmol_c\ dm^{-3}$)	K^+ ($mg\ dm^{-3}$)	Nº	Long. E (m)	Lat. N (m)	Ca^{2+} ($cmol_c\ dm^{-3}$)	Mg^{2+} ($cmol_c\ dm^{-3}$)	K^+ ($mg\ dm^{-3}$)
			Dados Moraes (2013)						Dados Schünemann (2016)		
1	426181,2	3115307,1	7,05	1,75	42,17	48	426360,1	3115115,9	14,46	5,58	91,18
2	426363,0	3115300,7	2,20	1,65	24,48	49	426548,9	3115127,1	10,46	8,85	133,43
3	426554,9	3115286,7	12,65	1,80	38,38	50	426854,0	3115152,4	7,19	4,95	107,74
4	426776,4	3115304,0	19,85	0,85	39,64	51	426019,4	3115518,4	8,09	7,64	113,53
5	426961,4	3115286,7	6,15	5,90	126,83	52	426280,8	3115504,0	12,82	6,94	113,80
6	427161,6	3115288,2	5,15	2,85	53,54	53	426544,3	3115479,5	0,73	0,35	38,40
7	425744,3	3115648,5	3,70	2,05	52,28	54	426737,9	3115458,4	29,68	5,19	91,08
8	425913,6	3115652,5	1,35	5,95	32,06	55	427175,8	3115447,8	17,32	2,26	94,25
9	426178,8	3115601,6	11,60	10,70	77,55	56	425675,4	3115774,5	3,04	2,68	82,30
10	426379,4	3115625,8	17,55	7,90	99,03	57	425930,0	3115800,4	30,12	3,07	45,11
11	426552,6	3115654,7	28,05	1,45	43,43	58	427243,2	3115759,3	9,27	9,77	141,31
12	426761,8	3115650,0	19,75	0,85	35,85	59	427250,9	3115893,1	7,19	2,88	99,78
13	426966,6	3115657,1	14,15	1,95	51,01	60	427100,0	3115908,2	11,95	5,23	133,42
14	427162,8	3115651,0	8,30	5,35	54,80	61	426854,7	3115913,3	18,56	4,76	142,74
15	427342,9	3115657,4	0,80	1,95	90,18	62	426711,5	3116018,6	16,34	1,82	96,43
16	425739,6	3116014,7	13,20	1,95	32,06	63	426378,2	3115935,9	42,09	5,43	58,25
17	425943,7	3116021,2	13,30	4,60	45,96	64	426150,0	3115922,4	23,16	3,29	74,39
18	426123,8	3116013,8	23,90	3,55	23,21	65	425950,2	3115942,3	27,30	3,74	46,85
19	426354,2	3116017,1	25,10	1,00	35,85	66	425682,2	3115959,4	11,01	6,45	88,77
20	426574,5	3116029,4	23,35	0,40	28,27	67	425648,3	3116231,6	6,63	3,96	82,00
21	426724,0	3116024,5	17,25	0,95	30,79	68	425824,8	3116265,1	7,93	4,78	88,54
22	426957,4	3116017,1	17,40	2,35	38,38	69	426043,2	3116202,3	17,14	2,83	48,53
23	427161,7	3116027,9	9,90	5,75	62,38	70	426341,3	3116195,1	35,18	4,50	75,04
24	425738,7	3116381,4	4,15	4,25	55,35	71	426412,4	3116178,3	35,85	3,41	75,11
25	425937,5	3116405,3	0,45	1,00	38,59	72	426677,2	3116184,2	25,47	2,06	58,60
26	426149,4	3116402,3	12,75	6,05	82,66	73	426858,2	3116169,8	32,39	3,44	103,65
27	426353,1	3116405,2	13,20	8,25	62,83	74	427039,5	3116175,1	30,57	3,67	59,79
28	426566,1	3116407,9	30,30	3,70	56,22	75	427036,0	3116556,6	14,83	9,47	149,69
29	426860,9	3116370,3	21,05	3,75	67,24	76	426887,5	3116526,1	19,08	2,61	60,08
30	427063,4	3116395,1	13,10	8,30	67,24	77	426458,9	3116553,6	19,78	2,74	50,59
31	425942,8	3116745,8	15,65	3,00	69,44	78	426278,1	3116562,6	1,83	0,03	85,82
32	426201,8	3116780,4	22,15	3,00	6,90	79	426096,5	3116576,7	16,90	2,68	56,93
33	426355,3	3116746,5	18,95	3,40	1,62	80	425918,8	3116566,4	9,14	5,23	77,90
34	426558,2	3116749,3	19,15	3,90	-0,14	81	425773,4	3116582,0	8,75	4,10	70,82
35	426959,5	3116757,0	19,35	5,90	3,98	82	425858,0	3116911,6	16,87	1,13	69,11
36	425936,9	3117115,0	6,90	2,55	20,94	83	426114,6	3116927,1	11,12	3,22	74,39
37	426149,8	3117110,1	9,65	3,60	16,61	84	426313,3	3116944,9	19,38	5,09	142,58
38	426967,2	3117120,5	0,75	2,30	0,00	85	426526,0	3116917,6	29,49	3,62	58,50
39	425960,1	3117463,6	12,50	2,20	27,34	86	426830,3	3116916,3	6,74	2,24	20,75
40	426138,4	3117498,6	8,10	2,15	33,74	87	427020,4	3116902,8	16,45	4,38	104,92
41	426378,7	3117344,1	17,20	3,55	40,13	88	426983,2	3117303,1	13,62	4,60	150,58
42	426940,9	3117493,9	9,65	3,60	24,14	89	426919,7	3117279,9	9,76	2,61	123,58
43	426190,5	3117868,8	4,80	2,65	49,73	90	426516,4	3117325,1	25,19	1,55	37,22
44	426621,8	3117932,8	8,70	2,25	38,53	91	426191,4	3117383,2	15,86	1,60	106,40
45	426740,3	3117883,1	11,80	4,50	27,18	92	426025,2	3117438,9	14,34	1,67	76,31
46	426571,8	3114989,7	4,10	8,80	86,50	93	425904,9	3117382,1	13,17	1,42	93,17
47	426750,9	3114996,8	2,05	1,70	33,74	94	425902,2	3117669,6	6,61	2,27	60,06
						95	426125,7	3117680,6	10,72	1,67	101,05
						96	426423,2	3117658,7	15,98	0,81	59,68
						97	426693,4	3117634,9	8,65	2,80	107,06
						98	426776,8	3117778,0	9,96	5,16	68,91
						99	426628,3	3117738,4	12,70	1,29	93,79
						100	426304,4	3117796,4	13,08	1,72	186,96
						101	426037,3	3117780,7	9,30	1,48	32,86
						102	426690,3	3117820,3	7,02	3,38	102,24

Source: Authors.

Figure 2. Distribution of the quartiles for Ca^{2+}, Mg^{2+} and K^+ in Keller Peninsula. Red = 25% of the highest values; Blue = 25% of the lowest values.
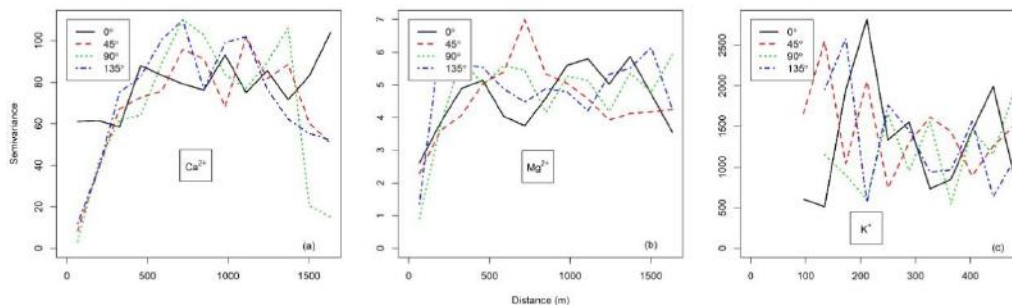

Source: Authors.

Results suggest that highest Ca^{2+} values are found in the central part of the peninsula, evidenced by the quartile represented by the red points (4th quartile). However, the lowest Ca^{2+} values are observed in the coastal regions, represented by the blue points (1st quartile). The intermediate values between the second and third quartiles are distributed randomly throughout the entire peninsula. This spatial characteristic was not observed for Mg^{2+} and K^+, which directly influenced them as covariates in the RF model.

3.2 Interpolators

The analysis of semivariograms in the four directions (0º, 45º, 90º and 135º) allow the visual pattern of spatial distribution of soil variables. The close relation between the directions for Ca^{2+} and Mg^{2+} indicates lower anisotropy. However, this behavior was not observed for K^+, indicating high anisotropy, which indicates that such variable changes or assumes more spatial patterns in different directions than Ca^{2+} and Mg^{2+}.
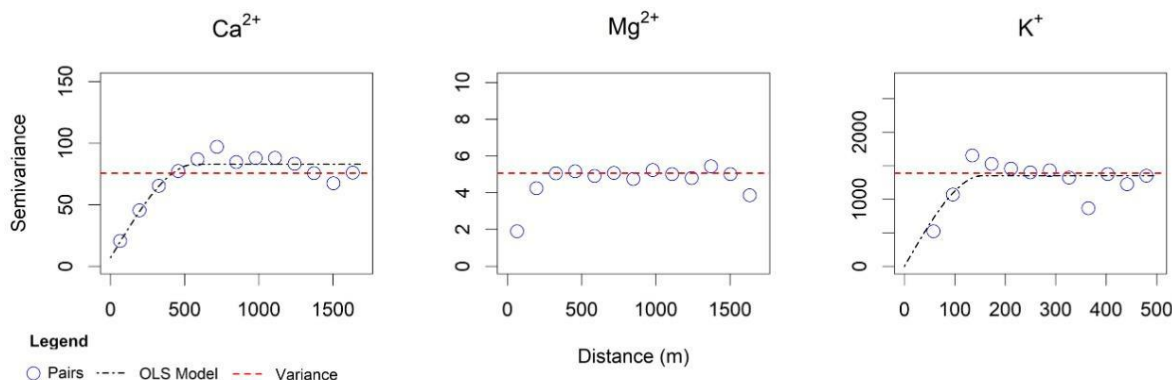
Figure 3. Omnidirectional semivariogram Ca^{2+}, Mg^{2+} and K^+.

Source: Authors.

The SK with the OLS adjustment method is presented in Figure 4. Kriging parameters are presented in Table 5. Results of RMSE indicates that SK and RF have similar values (Table 5).

Figure 4. Semivariograms models obtained using the Ordinary Least Squares (OLS).

Source: Authors.

Table 5. RMSE obtained from cross-validations for all interpolators and SK indicators.

Base	Interp.	RMSE	Unidade	P1 (nº)	P2 (nº)	P3 (nº)	MEP	DEP	Beta 0	Beta 1	AP	Pep
Ca^{2+}	KS	6,686	($cmol_c$ dm^{-3})	17	196	246	-0,004	1,08	1,26	0,91	569,457	7,00
Ca^{2+}	RF	7,569	($cmol_c$ dm^{-3})	-	-	-	-	-	-	-	-	-
Mg^{2+}	KS	2,134	($cmol_c$ dm^{-3})	17	196	246	-0,004	1,06	0,80	0,77	303,096	0,57
Mg^{2+}	RF	2,234	($cmol_c$ dm^{-3})	-	-	-	-	-	-	-	-	-
K^+	KS	37,893	(mg dm^{-3})	3	9	25	0,000	1,12	83,26	-0,22	154,979	0,00
K^+	RF	37,269	(mg dm^{-3})	-	-	-	-	-	-	-	-	-

Source: Authors.

P1: number of pairs in 1st point; P2: number of pairs in the 2nd point; P3: number of pairs in the 3rd point; MEP: mean of standardized errors; SDP: standardized error deviation; R: practical range; C_0: nugget effect.

When RF interpolation was performed for Ca^{2+}, with different cell sizes, the RMSE values between the different sizes remained numerically similar (Table 6). The lowest value of RMSE can be observed for the SK interpolator, with a resolution of 30 m. (Table 6).

Table 6. RMSE values in different covariant resolutions for the Ca^{2+} variable.

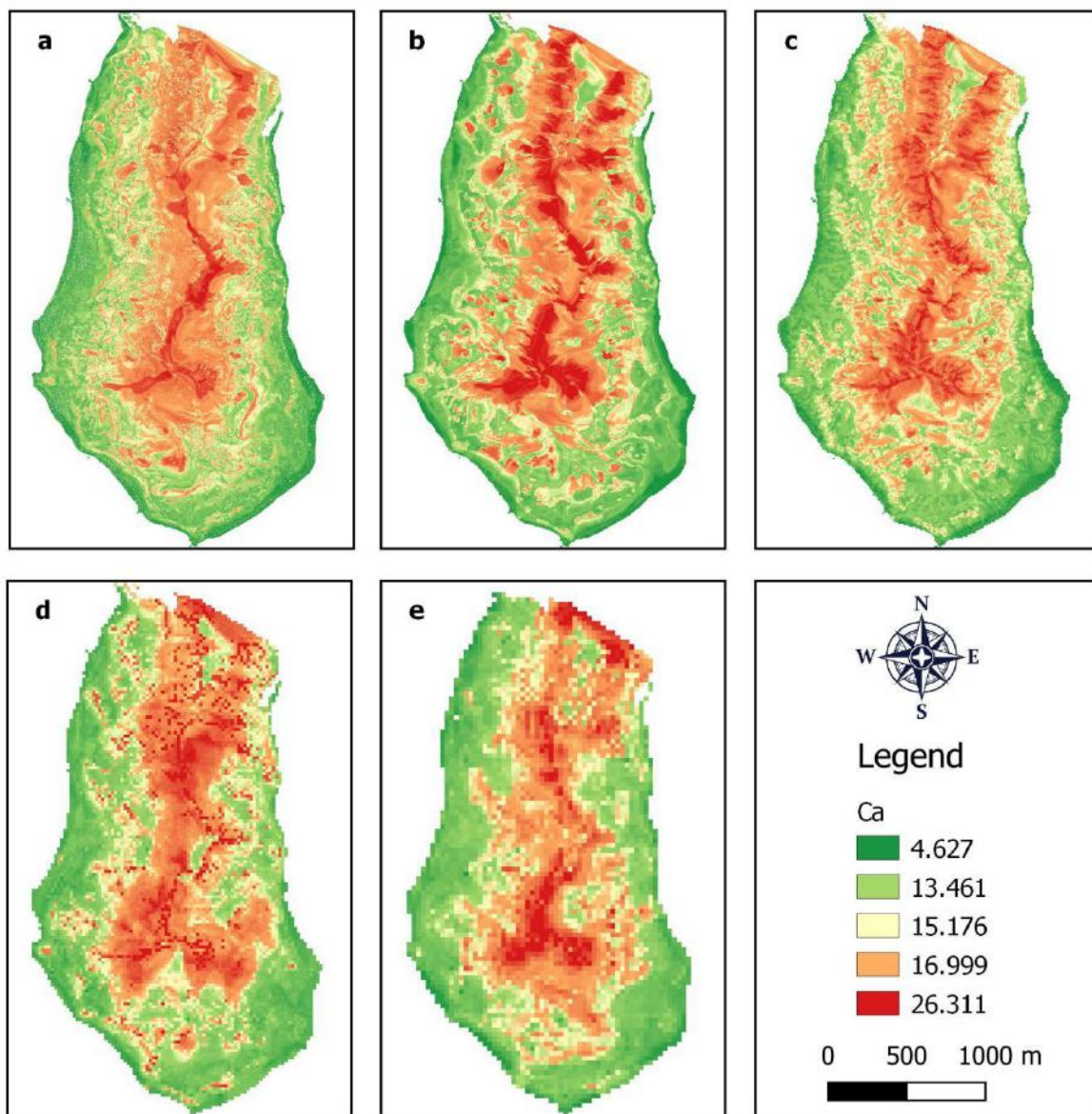| Resolução (m)	Interpolador	RMSE ($cmol_c$ dm^{-3})
1	RF	7,781
5	RF	7,863
10	RF	7,792
20	RF	7,905
30	RF	7,569
30	KS	6,686

Source: Authors.

The DEM, topographic humidity index and standardized elevation covariates were the most important covariates in explaining the variation of Ca^{2+} in all cell sizes. For Mg^{2+}, the topographic position index, curvature profile, convergence

index, surface convexity, half-slope position, multiresolution index of valley bottom flatness and face of exposure covariates, were the covariates that best explained the phenomenon. For K^+, the covariates selected by RF were diurnal anisotropic heating, longitudinal curvature, curvature profile and surface convexity.
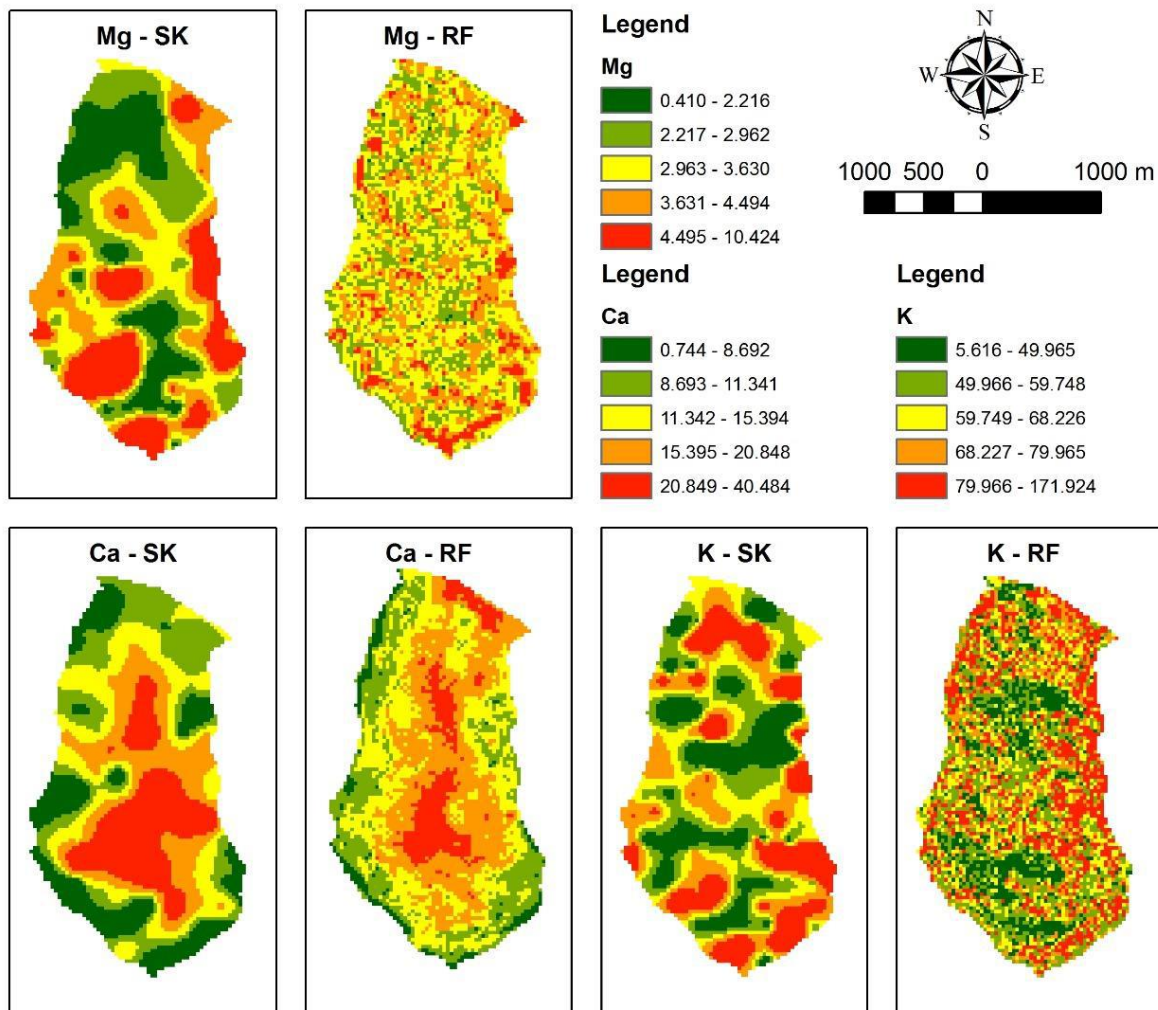
3.3 Comparison between predictions

Thematic maps produced for Ca^{2+} with different cell sizes (1, 5, 10, 20 and 30 m) by RF modeling are presented in Figure 5.

Figure 5. Predicted maps of Ca^{2+} ($cmol_c$ dm^{-3}) by the RF interpolator. (a) 1 m cell size; (b) 5 m cell size; (c) 10 m cell size; (d) 20 m cell size; (e) 30 m cell size.

Source: Authors.

Figure 6. Predicted maps of Ca^{2+} ($cmol_c$ dm^{-3}), Mg^{2+} ($cmol_c$ dm^{-3}) and K^+ ($cmol_c$ dm^{-3}) by SK and RF interpolators with 30 m cell size.

Source: Authors.

The map of errors comparing the two methods (RF and SK) cell to cell is shown in Figure 7.

Figure 7 – Error maps between methods. Ca and Mg values in $cmol_c$ dm^{-3}. K values in mg dm^{-3}.

Source: Authors.

Table 7. Error Summary between maps generated from RF and SK.

	Error Summary between RF and SK		
	Element		
Summary	**Ca** ($cmol_c$ dm^{-3})	**K** (mg dm^{-3})	**Mg** ($cmol_c$ dm^{-3})
Mínimum	0.0005	0.0034	0.0001
1st Quartile	1.7506	7.0061	0.4590
Median	3.4369	15.1256	0.9232
Mean (MAE)	4.1503	18.8267	1.1702
3rd Quartile	5.7167	27.2098	1.5566
Maximum	20.4817	101.7942	7.2639

Source: Authors.

4. Final Considerations

Simple kriging and randon forest interpolators did not present significant differences when considered their predicted surfaces for the three studied variables. The use of covariates increases computational processing, as well as the use of grids and smaller cell sizes, without enhancement in the quality of the prediction.

The covariates: Digital Elevation Model, Topographic Moisture Index, and Standardized Elevation are important in predicting Ca^{2+}. For the RF interpolator these covariates were the most important regardless cell, but with low explanation of the variability (~ 24%).

Considering the indicators used in this work and a relatively small sampling scheme, in ice-free areas with high complexity in their chemical attributes due to spatial heterogeneity, SK can be easily applied.

A denser sample grid is required to better model the spatial dependence of K^+. This presupposes a decrease in the distance between the samples, which may enhance spatial patterns detection for this soil chemical attribute. Further studies are needed to evaluate the best covariates for K+ in the RF modeling process in this area.

Ca^{2+} mapping in periglacial areas with lithological and geomorphological complexity is possible using kriging techniques. In this case, less computational processing is required associated with the fact that no environmental covariates are need (such as in random forest modeling). The RF did not improve the prediction with the use of covariates in this study. Further studies with denser grid sampling are needed to better elucidate the spatial dependence of soil attributes in Antarctica.

Acknowledgments

References

Breiman, L., Breiman, L. (2001). Statistical Modeling: The Two Cultures. Statistical Science, 16(3): 199–215.

Brenning, A. (2008). Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In: SAGA--Seconds Out (Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, 19: 23–32.

Carneiro, A. P. B., Polito, M. J., Sander, M. W Z. (2010). Trivelpiece. Abundance and spatial distribution of sympatrically breeding Catharacta spp. (skuas) in Admiralty Bay, King George Island, Antarctica. Polar Biology, 33(5): 673–682.

Embrapa. (2011). Manual de métodos de análise de solo, 3rd ed. EMBRAPA/CNPS, Rio de Janeiro (212 pp).

Ferreira, Í. O., Santos, G. R. DOS, Rodrigues, D. D. (2013). Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas. Revista Brasileira de Cartografia, 65(5): 831–842.

Francelino, M. R., Schaefer, C. E. G. R., Simas, F. N. B., Filho, E. I. F., Souza, J. J. L. L., Costa, L. M., (2011). Geomorphology and soils distribution under paraglacial conditions in an ice-free area ofAdmiralty Bay, King George Island, Antarctica. Catena 85: 194–204.

Hijmans, R. J. (2015). Geographic Data Analysis and Modeling. R package version 2.4-20. http://CRAN.R-project.org/package=raster.

INPE, 2015. Instituto Nacional de Pesquisas Espaciais-CPTEC. http://antartica.cptec.inpe. br/~antar/weatherdata.shtml.

Kirkwood, C., Cave, M., Beamish, D., Grebby, S., & Ferreira, A. (2016). A machine learning approach to geochemical mapping. Journal of Geochemical Exploration, 167.

Thomazini, A., Teixeira, D. B., Turbay, C. V. G., La Scala Jr, N., Schaefer, C. E. G. R., Mendonça, E. S., (2014). Spatial variability of CO2 emissions from newly exposed paraglacial soils at a glacier retreat zone on King George Island, Maritime Antarctica. Permafr. Periglac. Process. 25: 233–242.

R Core Team. R. (2015). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL https://www.R-project.org/, consulted novenber 2015.

Mendonça, E. S., La Scala, N., Panosso, A. R., Simas, F. N. B., Schaefer, C. E. G. R., (2010). Spatial variability models of CO2 emissions from soils colonized by grass (Deschampsia antarctica) and moss (Sanionia uncinata) in Admiralty Bay, King George Island. Antarctic Science. 23: 27–33.

Simas, F. N. B., Schaefer, C. E. G. R., Albuquerque Filho, M. R., Francelino, M. R., Fernandes Filho, E. I., & Costa, L. M. (2006). Genesis, properties and classification of Cryosols from Admiralty Bay, Maritime Antarctica. Geoderma 144: 116–122.

Schünemann, A. L., Almeida, P. H. A., Thomazini, A, Fernandes Filho, E I., Francelino, M, R., Schaefer, C. E. G. R., & Pereira, A B. (2018). High-resolution topography for Digital Terrain Model (DTM) in Keller Peninsula, Maritime Antarctica. Anais da Academia Brasileira de Ciências, 90(2, Suppl. 1), 2001-2010.

Vaysse, K., Lagacherie, P., (2015). Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Reg. 4: 20–30.

Vieira, S. R. (2000) Geoestatística em estudos de variabilidade espacial do solo. In: Novais, R.F., Alvarez V., V.H., Schaefer, G.R., eds. Tópicos em ciência do solo. Viçosa, Sociedade Brasileira de Ciência do Solo, 1, 1-54.

Wasklewicz,. T., STaley D M., Reavis K., & Oguchi T. (2013). 3.6 Digital Terrain Modeling. In: Treatise on Geomorphology, p. 130-216.Webster, R., Oliver, M.A., (1990). Statistical Methods in Soil and Land Resource Survey. Oxford University Press, Oxford, p. 316.

Willmott, C. J, & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30:79–82.