Geophysics and soil data

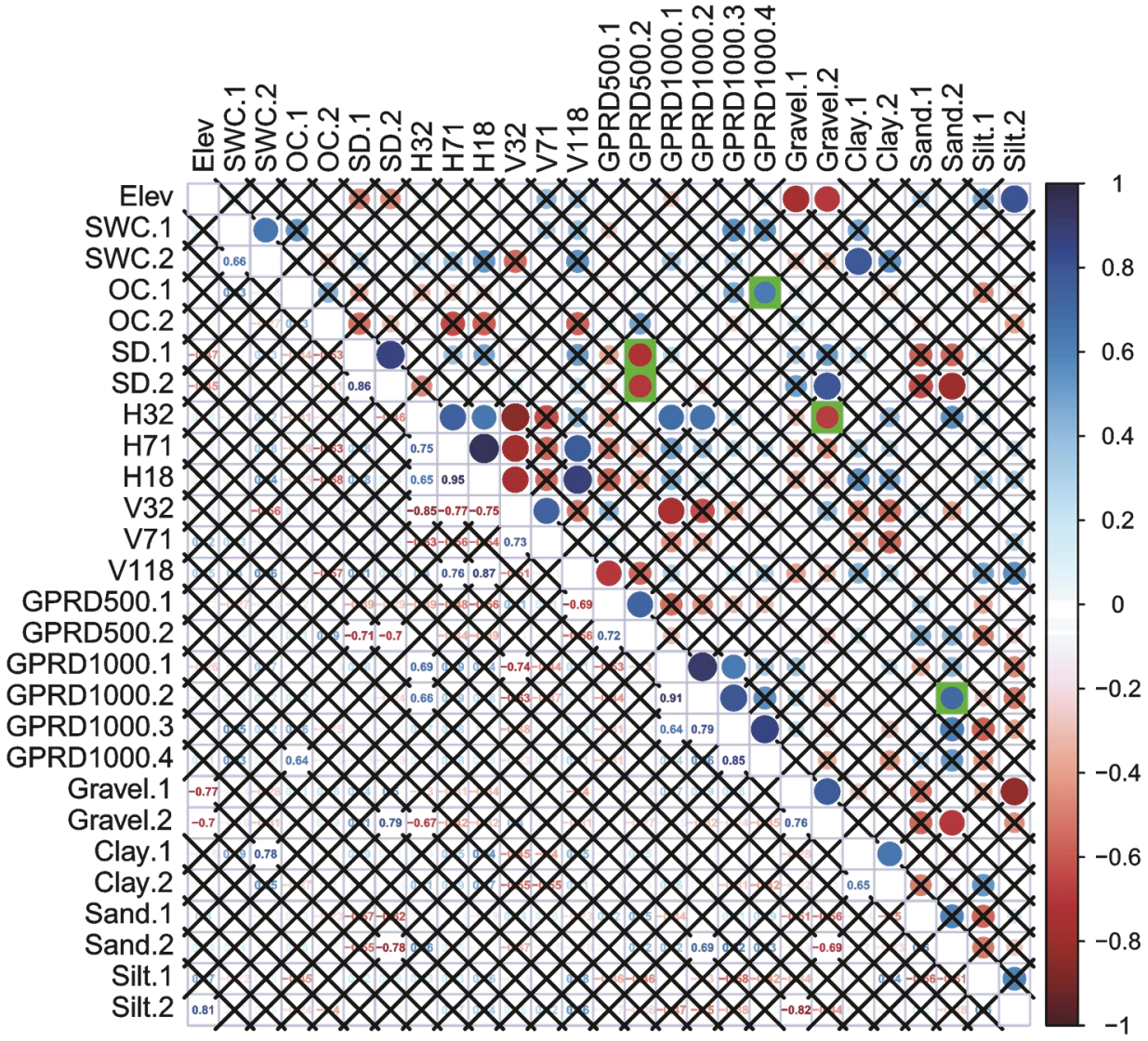
Basic statistical analysis

Data overview

Soil sample	Lat	Elev	Clay 1	Clay 2	Sand 1	Sand 2	Silt 1	Silt 2	SWC 1	SWC 2	Corg 1	Corg 2	SD 1	SD 2	HCP32	HCP71	HCP118	VCP32	VCP71	VCP118	GW 1000	R 1000 1	R 1000 2	GW 500	R 500 1	R 500 2
Soil 01	-32.9562	741.286	0.08	0.09	0.81	0.85	0.11	0.06	1.66	2.03	2.26	1.37	0.17	0.27	3.9336	3.4584	1.8058	1.2108	0.8167	1.0288	0.1706	0.1406	0.1351	0.1284	0.1026	0.1094
Soil 02	-32.956	743.867	0.12	0.13	0.77	0.77	0.12	0.10	2.72	3.68	1.71	1.55	0.30	0.46	2.8992	4.0484	3.3684	4.8068	2.2076	2.4308	0.1698	0.1259	NaN	0.1588	0.1098	NaN
Soil 03	-32.9561	735.219	0.10	0.12	0.76	0.77	0.14	0.11	2.03	1.84	2.53	1.39	0.30	0.60	1.6146	1.9312	1.3026	4.9284	1.696	1.3416	NaN	NaN	NaN	NaN	NaN	NaN
Soil 04	-32.9558	750.763	0.11	0.12	0.75	0.72	0.14	0.16	3.28	3.27	2.50	1.65	0.18	0.32	2.6036	2.4596	1.9888	2.8748	1.9224	1.752	NaN	NaN	NaN	NaN	NaN	NaN
Soil 05	-32.9558	757.308	0.07	0.09	0.75	0.70	0.18	0.21	1.61	1.62	1.52	0.89	0.20	0.40	3.0136	3.0436	2.0164	4.7975	2.4761	1.972	0.1687	0.1339	0.1249	0.1706	0.1073	NaN
Soil 06	-32.956	752.337	0.08	0.13	0.74	0.77	0.18	0.10	2.19	NaN	2.12	1.40	0.20	0.36	1.69	2.9504	1.7184	5.5308	2.1404	1.752	0.1457	0.1288	0.0979	0.1449	0.1068	0.1035
Soil 07	-32.9571	749.578	0.09	0.06	0.81	0.77	0.10	0.17	3.44	2.32	4.06	0.49	0.22	0.40	2.5524	3.7274	2.656	3.8738	2.2426	2.2656	0.158	NaN	NaN	0.1605	0.1153	0.0671
Soil 08	-32.9571	745.645	0.11	0.13	0.74	0.72	0.15	0.15	2.70	3.35	2.04	0.33	0.28	0.47	3.855	5.2593	4.9286	0.5218	0.6786	2.8016	0.1482	0.1389	0.1229	0.166	0.1157	0.0688
Soil 09	-32.9571	746.044	0.10	0.11	0.73	0.72	0.17	0.17	2.21	2.84	1.26	0.67	0.30	0.44	4.058	5.841	4.9574	-0.849	0.7296	2.8618	0.1487	0.1106	0.1237	0.1542	0.1284	NaN
Soil 10	-32.9559	745.076	0.10	0.10	0.76	0.78	0.14	0.12	3.42	3.29	2.28	0.50	0.30	0.51	3.0292	3.6888	2.856	1.7852	2.499	2.3966	NaN	NaN	NaN	NaN	NaN	NaN
Soil 11	-32.9559	747.333	0.09	0.10	0.76	0.76	0.15	0.14	3.11	3.07	1.81	1.22	0.25	0.40	3.3591	2.8187	2.204	2.255	1.8433	2.033	NaN	NaN	NaN	NaN	NaN	NaN
Soil 12	-32.956	748.444	0.10	0.11	0.74	0.77	0.16	0.12	3.25	2.36	4.77	2.14	0.20	0.40	3.2768	3.1368	2.5728	1.2716	1.3628	2.1832	NaN	NaN	NaN	NaN	NaN	NaN
Soil 13	-32.956	748.127	0.11	0.12	0.77	0.73	0.12	0.15	2.59	1.91	3.45	1.26	0.25	0.47	3.9712	3.1856	2.1748	2.9235	1.8736	2.0812	NaN	NaN	NaN	NaN	NaN	NaN

The table above is an example of the data being analyzed and displayed in the next two pages. Soil samples were collected, from which several parameters were obtained after lab analysis (e.g. Clay, Sand, Silt, SWC, Corg, SD). These where then correlated with geophysical data (e.g. HCP32, GW 1000, R500 1) to understand if a multiple linear regression model could be useful in predicting soil parameters away from the samples points using non invasive methods. Data cleaning was carried out in Matalb, where the NaN were discarded.

Correlation plot



As it can be seen from the correlation plot, only a small percentage of variables (both responses and predictors) has a significant correlation. Among this, the five green squares represent the case when the significance is between a geophysical variable and a soil response. This suggests that there might not be any possibilities of predicting ground truth data using geophysics for this specific case study.

Multiple linear regression

Response	Predictors	R^2	RMSE
Clay layer 1	-	-	-
Clay layer 2	-	-	-
Sand layer 1	-	-	-
Sand layer 2	GPRD1000.1, GPRD1000.2	0.609	0.0338
Silt layer 1	GPRD1000.3	0.258	0.0231
Silt layer 2	Elev	0.417	0.0315
Gravel layer 1	Elev	0.539	1.6574
Gravel layer 2	Elev, HCP32	0.703	3.1008
OC layer 1	-	-	-
OC layer 2	HCP71	0.280	0.4518
SWC layer 1	-	-	-
SWC layer 2	-	-	-
Depth layer 1	Elev, VCP71, VCP118, GPRD1000.4	0.941	0.0127
Depth layer 2	-	-	-

Less than half of the responses show a connection with the geophysical predictors using a multiple linear regression model. RMSE are generally low, except for the gravel, which indicates that the result might not be reliable.

Having a look at the correlation table in the previous slide, it can be noticed that significant correlations exist only for the following models:

1) Sand layer 2 with GPRD1000.2
2) Gravel layer 2 with elevation (Elev) and HCP32

Geophysics and soil data

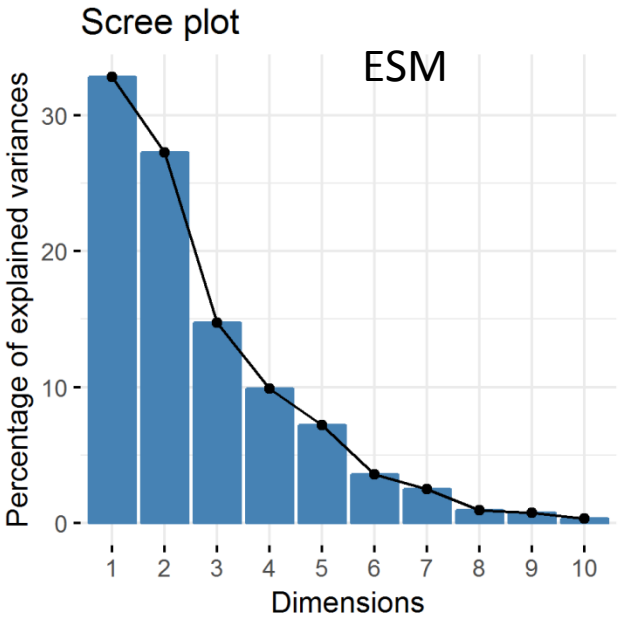
Dimensionality reduction and variance analysis using principal component analysis (PCA)

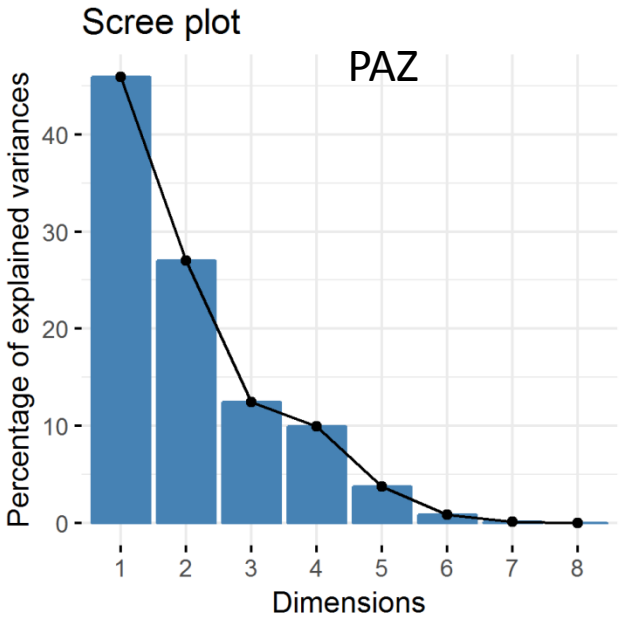
Data overview

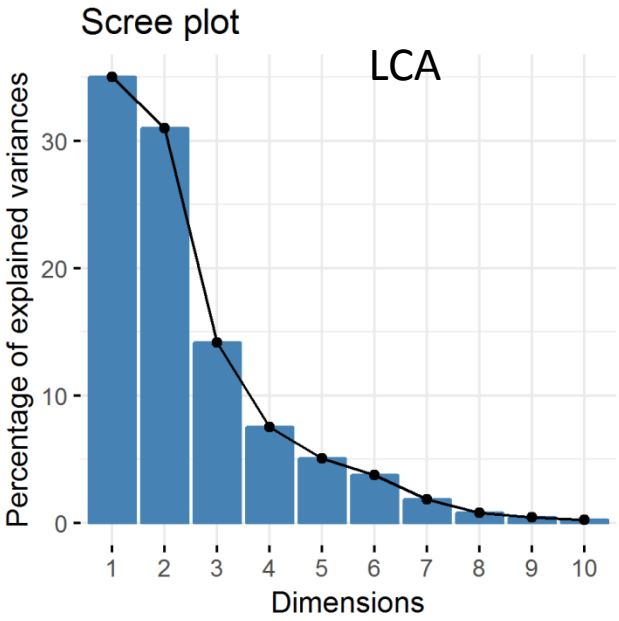
Depth	Density	Sand	Silt	Clay	LOI	CIA	Ca	Na	K	Mg	Zr	Vol strair	10 Be	pH	CEC	Envelope 500	Envelope 1000
Pan de Azucar																	
2.5	1.35	52.8	33.0	14.2	5.08	49.48	-0.44	-0.56	-0.52	1.26	1.08	-0.23	NaN	8.08	NaN	3.93E+11	1.83E+12
7.5	1.19	56.3	28.2	15.5	4.84	46.15	0.11	-0.44	-0.23	0.75	0.51	0.00	NaN	8.19	NaN	2.75E+11	1.51E+12
15	NaN	43.8	32.9	23.2	6.32	17.93	5.47	-0.45	-0.37	2.63	0.21	NaN	NaN	8.17	NaN	6.02E+11	2.69E+13
30	NaN	41.2	45.7	13.0	6.39	24.36	3.03	-0.09	-0.33	1.84	0.15	NaN	NaN	8.04	NaN	1.52E+13	3.01E+13
50	NaN	NaN	NaN	NaN	4.39	45.31	-0.10	0.00	-0.30	0.06	0.33	NaN	NaN	NaN	NaN	2.21E+13	1.93E+13
70	NaN	NaN	NaN	NaN	4.18	49.14	-0.13	-0.12	-0.27	0.01	0.22	NaN	NaN	NaN	NaN	1.06E+13	1.10E+13
90	NaN	NaN	NaN	NaN	3.73	48.70	-0.02	0.01	0.06	0.02	-0.13	NaN	NaN	NaN	NaN	6.22E+12	6.95E+12
110	NaN	NaN	NaN	NaN	3.88	47.88	0.00	0.00	0.00	0.00	0.00	NaN	NaN	NaN	NaN	4.82E+12	4.11E+12
2.5	1.19	NaN	NaN	NaN	4.77	52.51	-0.40	-0.08	-0.07	0.74	0.24	0.10	5.57	8.01	NaN	3.37E+11	1.96E+12
7.5	1.59	81.3	11.2	7.5	4.49	36.98	1.24	0.21	0.18	0.13	-0.13	0.00	NaN	8.10	NaN	2.16E+11	1.91E+12
15	NaN	NaN	NaN	NaN	3.22	52.45	-0.44	0.21	0.32	-0.14	-0.07	NaN	4.78	NaN	NaN	7.01E+11	3.65E+13
30	NaN	NaN	NaN	NaN	3.22	50.77	-0.38	0.19	0.13	-0.23	0.02	NaN	3.86	NaN	NaN	1.60E+13	3.06E+13
50	NaN	NaN	NaN	NaN	3.96	49.82	-0.16	0.17	0.10	-0.06	-0.08	NaN	NaN	NaN	NaN	2.32E+13	1.36E+13
70	NaN	NaN	NaN	NaN	3.78	51.78	-0.35	-0.01	-0.03	-0.16	0.03	NaN	3.33	NaN	NaN	1.47E+13	7.76E+12
90	NaN	NaN	NaN	NaN	4.88	42.91	0.56	-0.09	0.08	0.25	-0.09	NaN	NaN	NaN	NaN	1.38E+13	6.06E+12
110	NaN	NaN	NaN	NaN	4.03	50.20	-0.19	0.08	0.14	-0.07	-0.01	NaN	NaN	NaN	NaN	7.58E+12	5.85E+12
130	NaN	NaN	NaN	NaN	4.76	48.55	0.00	0.00	0.00	0.00	0.00	NaN	1.52	NaN	NaN	5.32E+12	7.41E+12
2.5	1.14	69.4	21.5	9.1	6.31	42.31	-0.09	-0.16	-0.11	0.23	0.09	-0.18	NaN	8.11	NaN	2.92E+11	1.69E+12
7.5	1.39	63.5	22.7	13.7	7.00	36.63	0.55	-0.34	0.08	0.38	0.02	0.00	NaN	8.11	NaN	2.12E+11	1.63E+12
15	NaN	44.4	43.8	11.8	7.48	25.64	2.06	-0.63	0.12	0.37	-0.13	NaN	NaN	8.24	NaN	7.47E+11	3.00E+13
30	NaN	NaN	NaN	NaN	4.49	52.63	-0.36	-0.70	0.15	0.10	0.05	NaN	NaN	NaN	NaN	1.81E+13	3.25E+13
50	NaN	NaN	NaN	NaN	5.20	50.52	-0.41	-0.27	-0.06	-0.14	0.15	NaN	NaN	NaN	NaN	2.33E+13	1.30E+13
70	NaN	NaN	NaN	NaN	6.00	43.77	-0.11	-0.10	-0.03	-0.18	0.07	NaN	NaN	NaN	NaN	7.12E+12	6.31E+12
90	NaN	NaN	NaN	NaN	7.88	41.11	0.25	-0.39	0.07	0.16	0.03	NaN	NaN	NaN	NaN	4.37E+12	5.91E+12
110	NaN	NaN	NaN	NaN	6.48	42.28	0.00	0.00	0.00	0.00	0.00	NaN	NaN	NaN	NaN	2.45E+12	5.75E+12
2.5	1.20	68.4	21.6	10.0	4.49	46.47	6.83	-0.10	-0.19	2.30	-0.02	-0.36	5.99	8.10	NaN	2.73E+11	8.47E+11
7.5	1.37	78.7	12.2	9.1	4.14	47.11	7.36	0.02	-0.10	1.73	-0.15	-0.23	NaN	8.12	NaN	1.73E+11	6.77E+11
15	1.32	70.7	10.9	18.4	3.05	49.19	6.01	-0.19	-0.18	1.34	-0.05	0.00	5.99	8.09	NaN	4.60E+11	1.80E+13
30	NaN	53.6	27.9	18.5	4.88	40.57	11.24	-0.13	-0.27	1.12	0.10	NaN	3.37	7.96	NaN	1.01E+13	2.16E+13
50	NaN	NaN	NaN	NaN	3.06	57.30	1.53	-0.15	-0.20	0.51	-0.02	NaN	NaN	NaN	NaN	1.67E+13	2.18E+13
70	NaN	NaN	NaN	NaN	2.44	53.53	2.58	0.11	0.17	0.16	-0.09	NaN	2.89	NaN	NaN	9.33E+12	1.85E+13
90	NaN	NaN	NaN	NaN	2.03	58.79	0.00	0.00	0.00	0.00	0.00	NaN	2.89	NaN	NaN	9.64E+12	7.90E+12
Santa Gracia																	
2.5	1.57	78.1	14.0	7.8	2.94	48.50	-0.03	0.01	1.43	-0.17	-0.16	0.33	NaN	7.01	89.26	2.40E+11	8.29E+11
7.5	1.50	75.7	14.2	10.1	2.94	47.18	-0.10	-0.22	0.80	-0.09	0.03	0.14	NaN	6.43	73.12	1.47E+11	5.66E+11
15	1.49	71.8	15.2	13.0	2.86	46.90	-0.14	-0.18	0.48	-0.17	0.11	0.05	NaN	6.38	116.94	5.09E+11	7.15E+12
30	1.56	72.0	15.1	12.9	1.82	47.70	-0.16	-0.15	0.48	-0.23	0.11	0.00	NaN	6.61	119.59	1.15E+13	1.93E+13
50	NaN	NaN	NaN	NaN	1.50	46.34	-0.11	-0.16	0.00	-0.26	0.19	NaN	NaN	NaN	NaN	1.48E+13	1.56E+13
70	NaN	NaN	NaN	NaN	1.55	45.91	-0.13	-0.21	0.49	-0.13	0.16	NaN	NaN	NaN	NaN	7.81E+12	6.78E+12
90	NaN	NaN	NaN	NaN	1.57	46.31	-0.12	-0.10	1.02	-0.20	0.03	NaN	NaN	NaN	NaN	5.73E+12	3.88E+12
110	NaN	NaN	NaN	NaN	1.52	45.71	0.00	0.00	0.00	0.00	0.00	NaN	NaN	NaN	NaN	4.69E+12	3.07E+12
2.5	1.48	76.8	17.5	5.6	1.89	47.70	-0.42	-0.48	-0.19	-0.40	0.69	-0.41	2.84	6.28	44.75	2.55E+11	7.84E+11
7.5	1.55	78.9	15.3	5.8	1.96	48.40	-0.49	-0.54	-0.26	-0.43	0.84	-0.46	NaN	6.29	51.05	2.26E+11	7.94E+11
15	1.55	78.6	13.9	7.5	1.93	48.84	-0.47	-0.50	-0.20	-0.42	0.72	-0.41	2.69	6.18	52.49	7.59E+11	1.45E+13
30	1.44	78.1	13.7	8.2	1.85	49.21	-0.44	-0.46	-0.23	-0.40	0.63	-0.37	2.79	6.11	59.95	1.87E+13	2.08E+13
50	1.43	76.9	13.7	9.5	1.97	49.17	-0.07	-0.13	-0.07	0.06	0.05	0.00	NaN	5.97	70.38	2.33E+13	1.76E+13
70	NaN	77.6	14.2	8.2	2.09	49.61	-0.14	-0.06	-0.17	0.06	0.05	NaN	2.02	5.48	108.79	1.31E+13	1.71E+13
90	NaN	NaN	NaN	NaN	2.20	51.30	-0.35	-0.17	-0.28	-0.04	0.17	NaN	NaN	NaN	NaN	1.65E+13	9.04E+12
110	NaN	NaN	NaN	NaN	1.81	48.23	-0.43	-0.38	-0.52	-0.20	0.62	NaN	NaN	NaN	NaN	8.37E+12	4.29E+12
130	NaN	NaN	NaN	NaN	1.28	47.06	-0.10	-0.07	-0.27	0.01	0.10	NaN	1.16	NaN	NaN	4.60E+12	3.31E+12

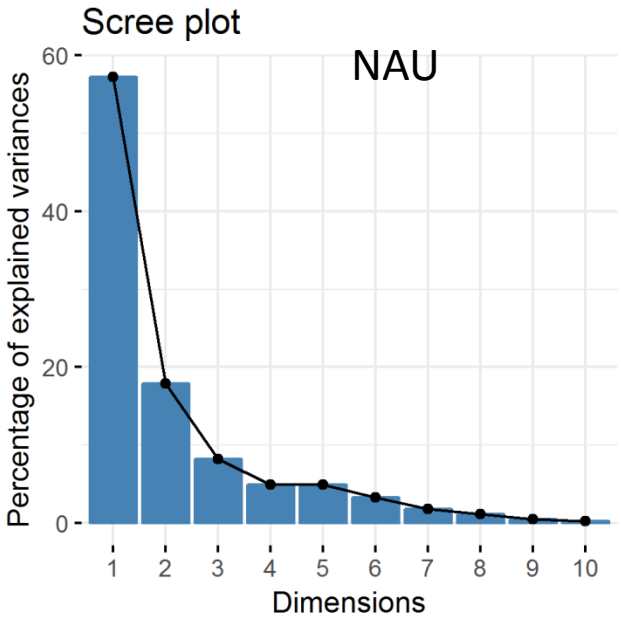
The table shows part of the data used in the analysis displayed in the next two slides. A set of physical and chemical properties where obtained for different depths, where also geophysical measurements are available (envelope 500 and 1000) after the reanalysis of ground penetrating data (GPR). The main goals are understanding the effect of each variable in the total variance and within each location (e.g. Pan de Azucar, Santa Gracia). NaN data were discarded/interpolated.

PCA – scree plots



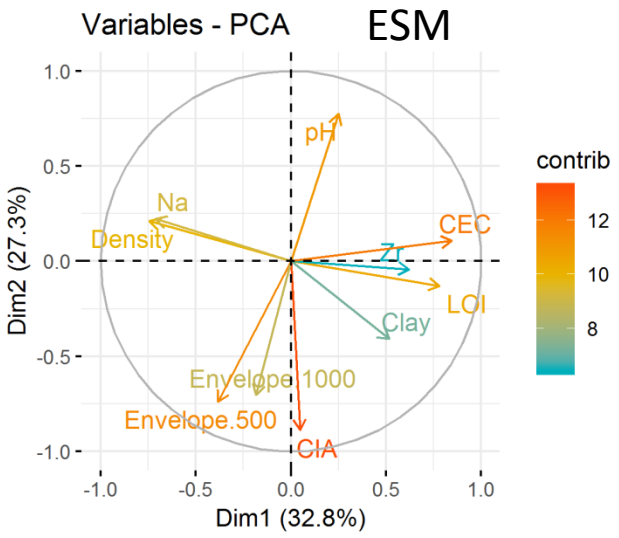


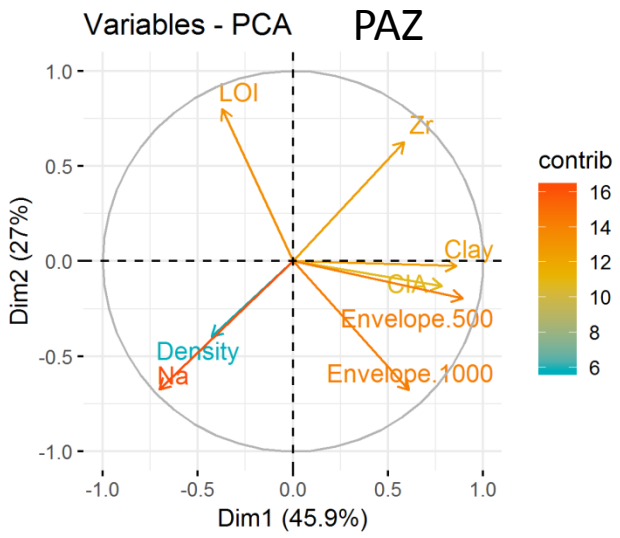


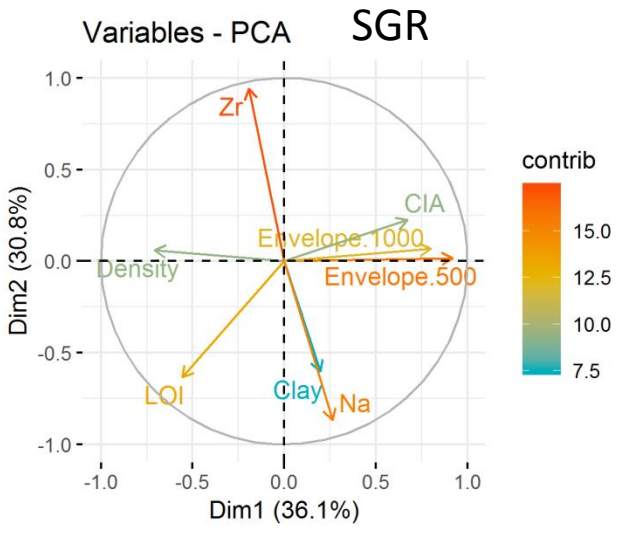


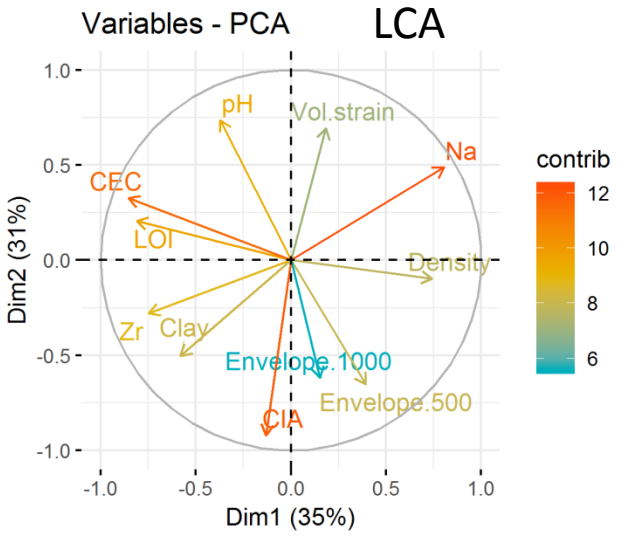
The five scree plots for the five studied datasets summarize the variance explained from each principal components (PC). Usually, most of the variance is contained in the first two PCs. The number of PCs to consider is a user dependent. Here, we only look at PC1 and PC2.

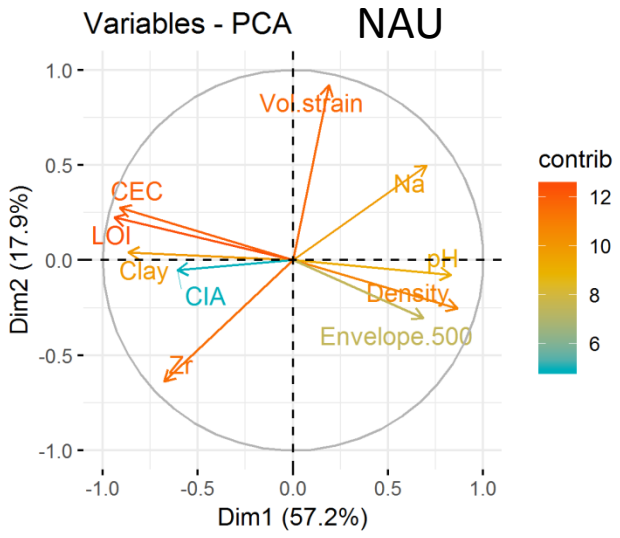
PCA – contribution plots











Each plot shows the contribution of each variable for PC1 and PC2. It can be seen that considering the entire dataset (plot upper right) CEC contributes the most to PC1 and CIA contributes the most to PC2. No real common trend is visible in the other four plots.

Geophysics and soil data

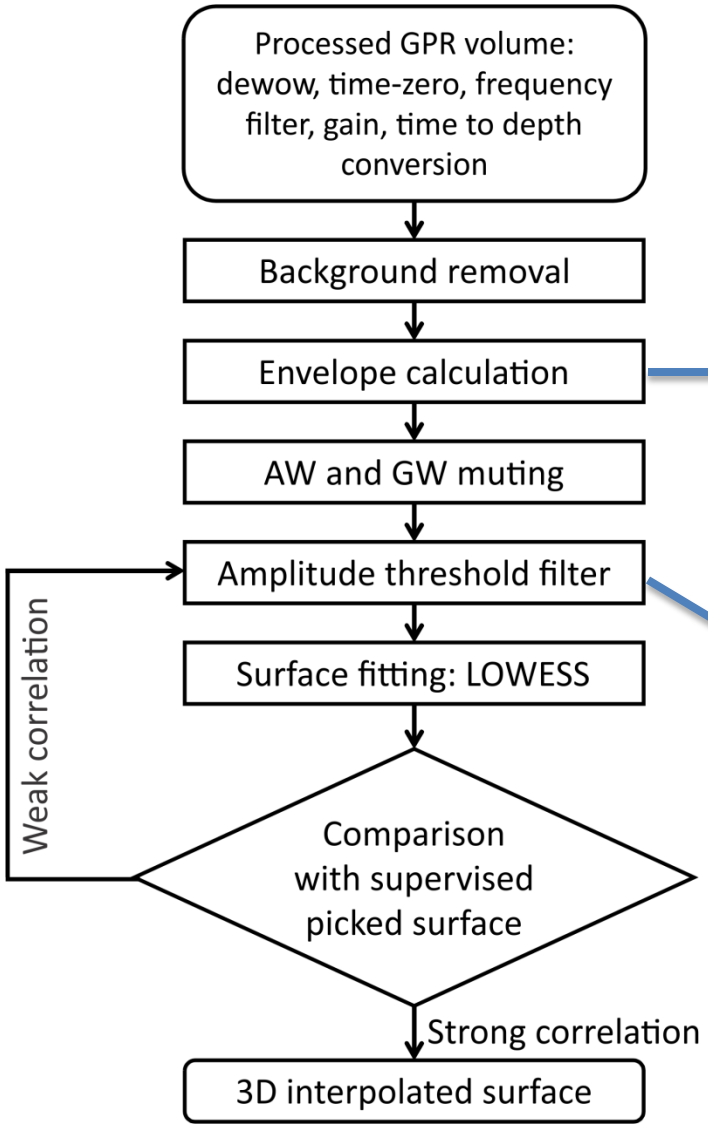
Statistical inference of a soil interface from geophysical data using non-parametric methods (lowess)

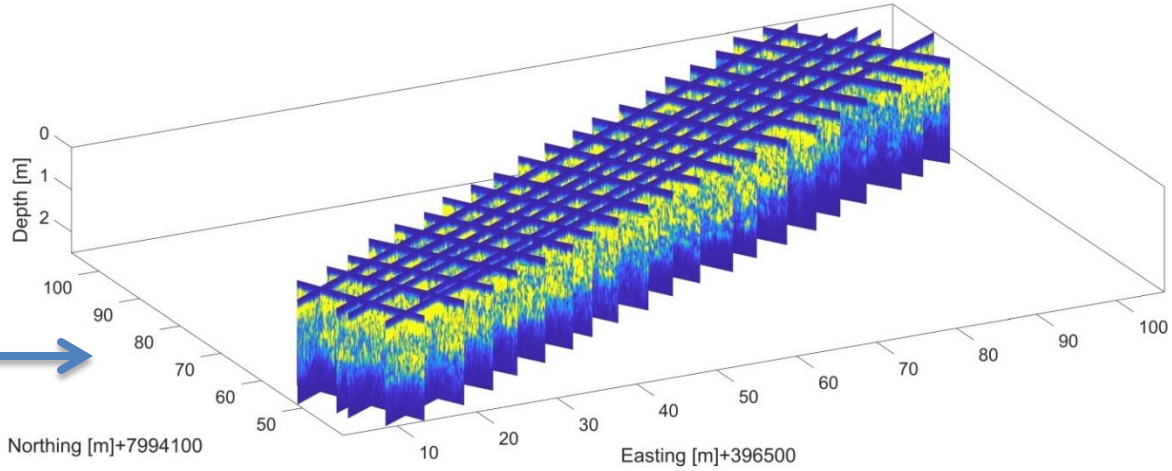
Data overview

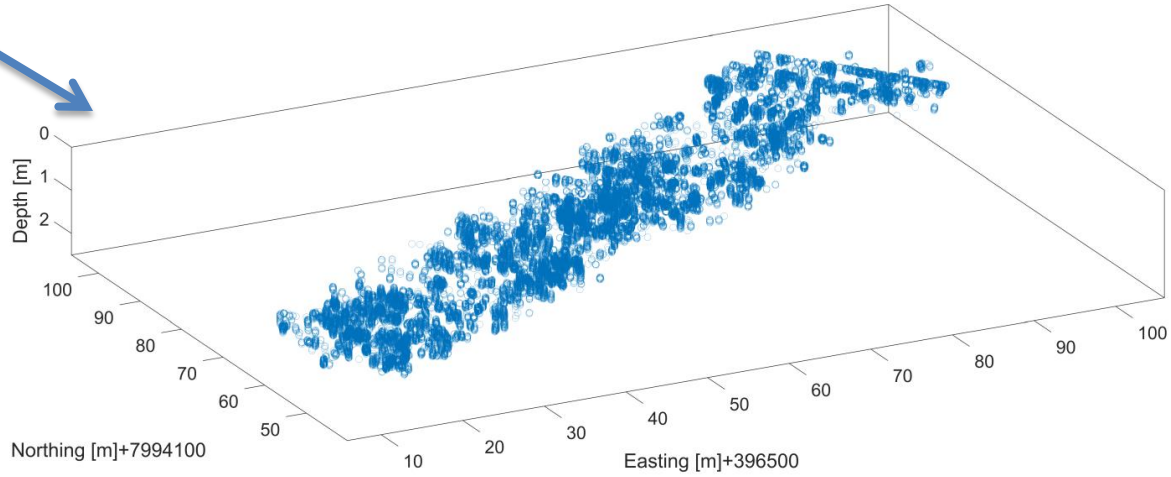
Longitude	Latitude	Elevation	Depth	Thickness
396503.2732	7334151.233	1053.415	1053.429189	1052.351975
396503.5768	7334151.412	1053.42	1053.428944	1052.35158
396503.8308	7334151.425	1053.471	1053.42929	1052.351036
396504.1548	7334151.605	1053.48	1053.429085	1052.350607
396504.377	7334151.805	1053.471	1053.428533	1052.350422
396504.6404	7334151.918	1053.444	1053.428462	1052.350025
396504.965	7334152.02	1053.469	1053.42859	1052.349486
396505.2988	7334152.233	1053.43	1053.428127	1052.349088
396505.5209	7334152.445	1053.453	1053.427403	1052.348908
396505.8046	7334152.569	1053.428	1053.427317	1052.348492
396506.1799	7334152.682	1053.441	1053.427524	1052.347885
396506.4127	7334152.817	1053.431	1053.427359	1052.347591
396506.6549	7334153.062	1053.428	1053.426561	1052.347395
396506.8877	7334153.197	1053.425	1053.426421	1052.347094
396507.2429	7334153.277	1053.388	1053.426797	1052.346506
396507.4964	7334153.356	1053.386	1053.426976	1052.346117
396507.7387	7334153.59	1053.362	1053.426438	1052.34589
396507.9711	7334153.78	1053.407	1053.426075	1052.34563
396508.2854	7334153.893	1053.384	1053.426208	1052.345151
396508.5798	7334153.94	1053.419	1053.426684	1052.344664
396508.7922	7334154.085	1053.391	1053.426389	1052.344391
396509.0446	7334154.331	1053.391	1053.425628	1052.344093
396509.267	7334154.499	1053.385	1053.425248	1052.343772
396509.5915	7334154.601	1053.392	1053.425565	1052.34324
396509.9259	7334154.747	1053.377	1053.425691	1052.342681
396510.1682	7334154.97	1053.37	1053.425173	1052.342279

A set of points with east (wrongly called latitude here) and north (wrongly called longitude here) coordinates have been used to infer the thickness of a soil interface using a lowess non-parametric algorithm. This was calculated by subtracting the depth from the elevation. As can be seen in the next slides, the points were obtained from a 2.5D envelope GPR volume and were filtered using an amplitude threshold.

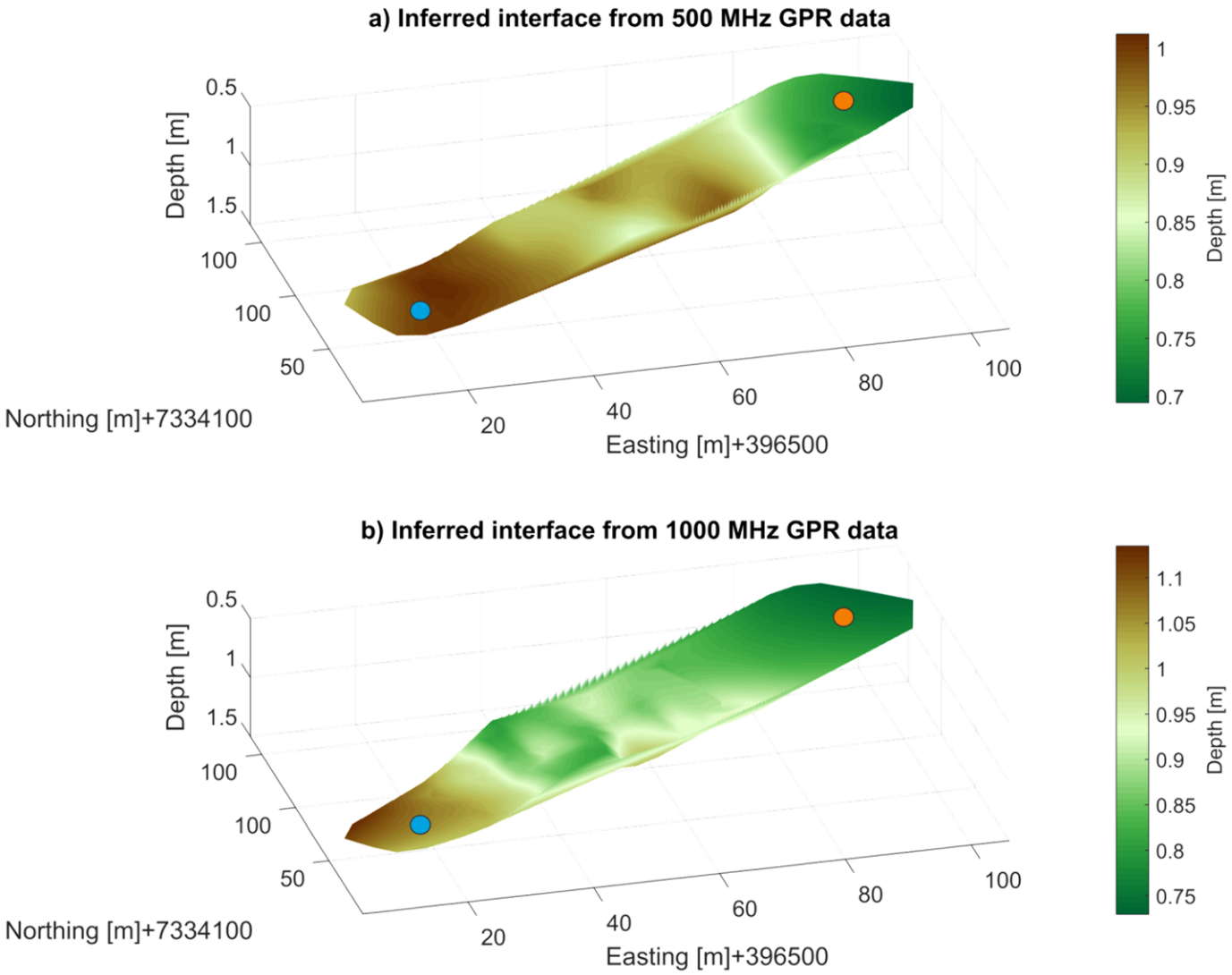
Workflow of the analysis







Interface calculation with lowess



The two inferred interfaces comes from different frequencies (500 and 1000 MHz) and show similar depths and lateral trends. The blue and orange dots represent ground truth data, where the interface was compared with the real soil interfaces. The two results agreed indicating that the surfaces are a good approximation of the reality.