Inversion of downhole resistivity properties through infrared spectroscopy and whole-rock geochemistry using machine-learning

Mehdi Serdoun, Frédéric Sur, Gaétan Milesi, Elodie Williard, Pierre Martz, J. Mercadier

Inversion of downhole resistivity properties through infrared spectroscopy and whole-rock geochemistry using machine-learning

Mehdi Serdoun[*], Frédéric Sur[†], Gaétan Milesi[*],
Élodie Williard[‡], Pierre Martz[‡], Julien Mercadier[*]

Abstract

The electrical properties of rocks are widely used in the geophysical exploration of natural resources, such as minerals, hydrocarbons and groundwater. In mining exploration, the primary goal is to map electrically anomalous geological features associated with different mineralization styles, such as clay alteration haloes, metal oxides and sulphides, weathered crystalline rocks or fractured zones. As such, the reconciliation of geophysical data with geological information (geochemistry, mineralogy, texture and lithology) is a critical step and can be performed based on petrophysical properties collected either on core samples or as downhole measurements. Based on data from 189 diamond drill cores collected for uranium exploration in the Athabasca Basin (Saskatchewan, Canada), this paper presents a case study of reconciliation of downhole resistivity probing with core sample geochemistry and short-wave infrared spectroscopy (350–2500 nm) through three successive steps: (i) multivariate analysis of resistivity and other petrophysical properties (porosity, density) against geochemical and infrared spectroscopy information to characterize electrical properties of rocks with respect to other physical parameters, (ii) a machine-learning workflow integrating geochemistry and spectral signatures in order to infer synthetic resistivity logs along with uncertainties. The best model in the basin was Light Gradient-Boosting Machine with pairwise log-ratio, which yielded a coefficient of determination $R^2 = 0.80$ (root mean square error = 0.16), and in the basement, support vector regression with data fusion of infrared spectroscopy and pairwise log-ratios on geochemistry yielded $R^2 = 0.82$ (root mean square error = 0.35); (iii) the best model was then fitted on an area that was excluded from the original dataset (Getty Russell property) in order to infer synthetic resistivity logs for that zone. Software code is publicly available. This workflow can be re-used for the valorization of legacy datasets.

Keywords: borehole geophysics, data processing, inversion, petrophysics, resistivity.

This is the author-manuscript version of

Mehdi Serdoun, Frédéric Sur, Gaétan Milesi, Élodie Williard, Pierre Martz, Julien Mercadier, Inversion of downhole resistivity properties through infrared spectroscopy and whole-rock geochemistry using machine-learning, *Geophysical prospecting*, vol. 73, no 1, p. 355–379, 2025.

DOI: 10.1111/1365-2478.13627

[*]GeoRessources – CNRS, Labcom CREGU, Université de Lorraine, Vandoeuvre-lès-Nancy, France
[†]LORIA – CNRS, INRIA, Université de Lorraine, Vandoeuvre-lès-Nancy, France
[‡]Orano Mining – Châtillon, Courbevoie, France

1 Introduction

As easy-to-explore, shallow ore deposits are progressively depleting, exploration is now targeting more deeply-seated mineralized bodies. This involves an increasing need for extensive geophysical surveys under overburden, regolithic zones or thick sedimentary covers to characterize variations in physical properties related with metallogenic processes, such as hydrothermalism, e.g. density lows associated to alteration haloes [61].

In this context, coupling petrophysical measurements on cores and geophysical downhole logging is an essential tool for exploration geoscientists because it is the only way to connect parameters observed by geophysicists to actual rock properties observed during drilling and fieldwork. The use of petrophysical data to model deposit footprints has been applied to various types of mineral systems, such as unconformity-related uranium deposits [62], orogenic gold deposits [12], porphyry copper deposits [16] and volcanogenic massive sulphide deposits [63].

However, in mineral exploration, downhole geophysical logging is historically less widespread than in the oil-andgas sector, resulting in scarcer petrophysical data volumes, and geoscientists often need to rely on incomplete datasets to infer the properties of a whole area [19]. Being able to use proxies coming from other types of data systematically collected across prospects, such as geochemistry and infrared (IR) spectroscopy in this case study, would be an important asset in expanding entry data for 3D geophysical models. However, and despite a growing concern in the geoscience exploration community for the integration of multi-source (geochemical, geophysical, geological and mineralogical) datasets in unified 3D models [44], there are also substantial difficulties in reconciling data that were often collected at varying spatial scales without such prospects in mind [63].

Furthermore, physical properties of rocks are a complex type of data as they are controlled by several parameters occurring at the micro-scale (e.g., crystallographic assemblages, crystal defects), meso-scale (e.g., mineralogy, pore geometry, pore saturation, permeability...) and macro-scale (e.g., fault zones geometry). Influences on physical properties have therefore been observed through geochemical studies [40], spectroscopic studies [46] as well as lithological variations [9], but because of this important number of parameters involved, no single measurement method can be thought to reflect all geological features susceptible to influence physical properties.

Machine-Learning (ML) offers opportunities to integrate multi-source datasets for the inversion of petrophysical parameters using other available parameters, like geochemistry and IR spectroscopy, as demonstrated by the important increase in the number of publications associating petrophysics and ML over the last few years [43]. Various issues need to be overcome before applying data fusion, e.g., differences in sampling intervals across different datasets, problems associated to closed compositional datasets or high number of dimensions in the case of geochemical and IR spectroscopic data. Here, through a data collection from uranium exploration in the Athabasca Basin (Saskatchewan, Canada), we aim to demonstrate that consistent predictions along with associated confidence intervals can be obtained in various geological media. A workflow using data fusion of geochemistry and Vis-NIR-SWIR spectroscopy (350-2500nm) data is proposed to infer electrical properties using downhole resistivity data as a learning dataset. Following a first step of descriptive statistics and multivariate analysis, we used various ML methods (Light Gradient Boosting Machine – LGBM; Partial-Least Squares Regression – PLS-R; Random Forest; eXtreme Gradient Boosting – XGBoost; Support Vector Regression – SVR) to predict values of resistivity for two different datasets, sandstone rocks and crystalline rocks. After testing our method on an external dataset (Getty Russell property), synthetic resistivity logs are obtained and compared to actual values.

2 General context

2.1 Geological setting

The studied area (Figure 1) is located in the South-East of the Athabasca Basin (Saskatchewan, Canada). The Athabasca Basin is an intracratonic sedimentary basin made of a suite of unmetamorphosed, chemically highly mature quartzose sandstone, formed around 1.7 Ga [36] and deposited in an epicontinental or fluvial setting [74]. The maximum thickness of the sedimentary basin currently reaches 2km in its centermost part [6] decreasing towards its edges, although it has been estimated by studies on fluid inclusions that it could have originally been 5 to 6km thick [49]. The basin consists of several conglomeratic and sandstone formations but in the studied area, only the Manitou Falls formation is observed [60]. The latter breaks down into seven members: the Clampitt-Dunlop, Hodge, Warnes, Collins, Bird, Read and Smart members [10] all being considered very similar in geochemical composition [6].

The Athabasca Basin unconformably overlies a crystalline basement [46] divided between the Rae province to the West, and the Hearne province to the East of the basin [68]. To the East, the Hearne province itself is divided into three domains, namely the Wollaston, Mudjatik and Wollaston-Mudjatik transition zone which are mostly comprised of Archean gneisses and Paleoproterozoic metasedimentary rocks which underwent extensive tectonic folding at ca. 1.8 Ga [68]. Interested readers are referred to [7] for more details on the geology of the Athabasca basement.

2.2 Uranium exploration in the Athabasca basin

Unconformity-related uranium (URU) deposits are thought to have formed as a result of the reaction between oxidizing basin brines and either reducing basement fluids [35], or reducing basement lithologies themselves [39], precipitating uranium at the basin-basement unconformity as massive ore, veins or disseminated along fault zones both in the sandstone and the basement [32]. A complete review of the evolution of conceptual genetic models of URU deposits goes well beyond the scope of this study and is detailed in [39].

URU deposits were originally explored using surface methods (airborne radiometry, radioactive boulder chasing, surface geochemistry) and most of the first deposits discovered were located at or near the surface, or were discovered through surface anomalies (e.g., [64] for the Midwest Lake discovery; [24] for the Key Lake deposit). Over the past three decades, increasing reliance has been put on geophysical surveys using electrical and electromagnetic methods [59] such as resistivity for delineating prospective targets at depth. Although some recent discoveries have been made at or near the surface, the current exploration paradigm is that most favorable zones are located in the immediate vicinity of the intersection between conductive reactivated graphitic shear zones in the basement and the basin-basement unconformity [41]. Electrical properties are also thought to be associated with various other processes related to hydrothermal alteration that are interpreted as evidence of uranium mobility, such as silicification, de-silicification, chloritization and argilization [59]. Indeed, broad hydrothermal alteration haloes surrounding these deposits are often considered as exploration guides. In basin rocks, it is mainly expressed by an evolution from a background cemented sandstone with a matrix of dickite (clay), iron oxides or hydroxides [15] toward an altered sandstone where iron oxides have been leached and the clay matrix is entirely replaced by various illite polytypes [56] [38] and small crystals of aluminum-phosphate-sulphate (APS). In basement rocks, the alteration is mainly expressed by a replacement of primary magmatic and metamorphic minerals such as feldspars, biotites and muscovites by an assemblage of clay minerals including illite, sudoite plus in a lesser proportion Mg-foitite [75] [57]. In some areas of extreme alteration, silica dissolution is additionally observed in both the basin and basement surrounding major U deposits [48].

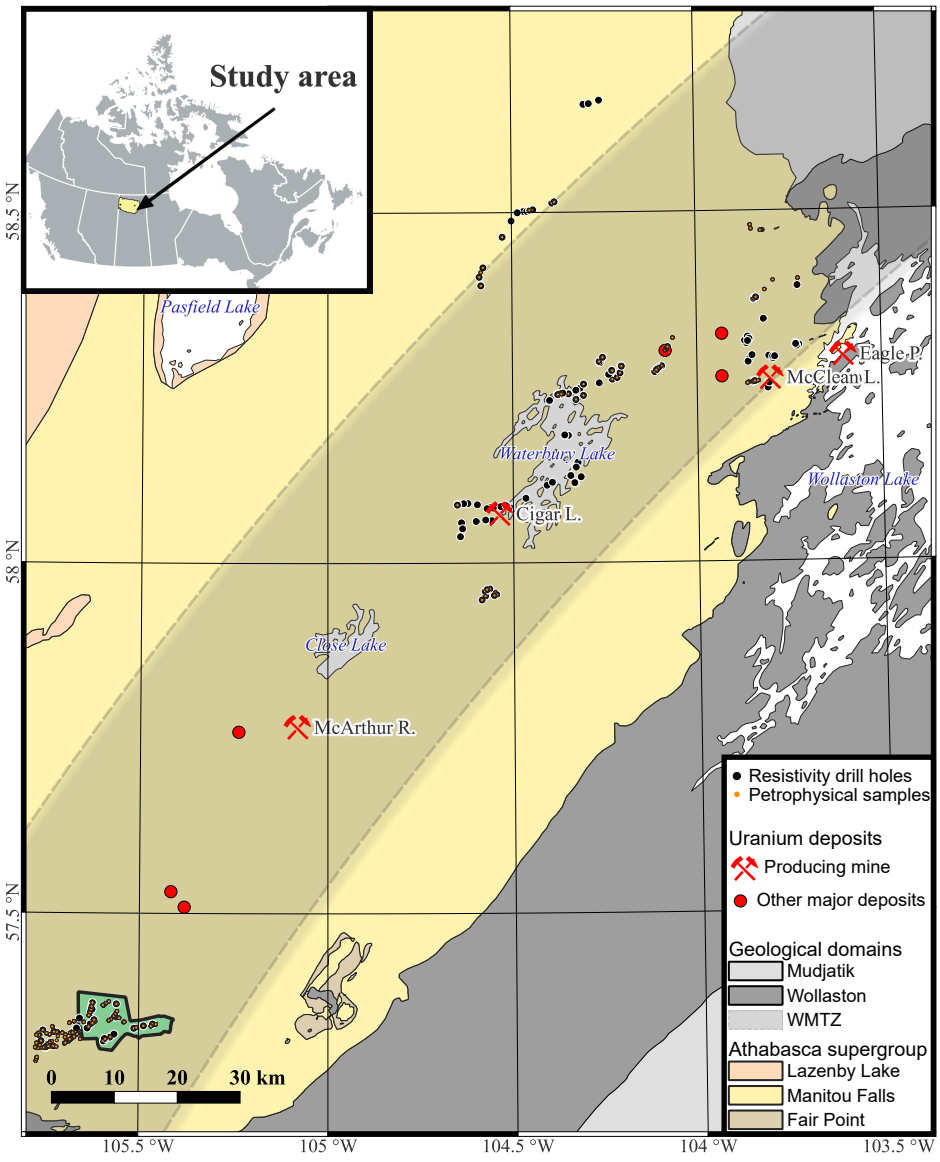

Figure 1: Location of resistivity drill holes and petrophysical samples studied. WMTZ = Wollaston-Mudjatik Transition Zone, drawn after [31]). Uranium mines are reported: Mc Arthur R.: MacArthur River Uranium Mine; Cigar L.: Cigar Lake; Eagle P.: Eagle Point; McClean L.: McClean Lake. Coordinate reference system is NAD83 UTM Zone 13. Light green area represents the Getty Russell property. Drillholes within this area were used as an external dataset for generating synthetic logs.

3 Data and methods

As the Athabasca Basin sandstone and its underlying basement rocks strongly differ in physical properties, alteration minerals and lithogeochemistry, basement samples and basin samples were divided into two separate subsets based on the depth of the unconformity. The depth of the unconformity strongly varied between different zones, ranging from ca. 100m at the edges of the basin to ca. 450m in the innermost prospects, with a median value of ca. 200m between all drill holes. In the external dataset (Getty Russell), the depth of the unconformity is approximately 180m. All subsequent steps including pre-processing, data analysis and resistivity prediction were performed separately for the two subsets. For each subset – basement and sandstone, two different datasets were studied:

- A dataset comprising resistivity logs, geochemistry and IR spectroscopy was used for the Machine-Learning workflow. A summary of the different processing steps performed before obtaining synthetic resistivity logs is given in Figure 2.
- A second dataset comprising geochemistry, IR spectroscopy and petrophysical core measurements (resistivity, porosity and density) was constituted for descriptive statistics and multivariate analysis in order to assess their relationships with resistivity relative to other petrophysical parameters (porosity and density). The purpose of this second dataset was to ensure that resistivity data measured on cores displayed similar geological responses to downhole probing measurements. It was also necessary to verify that no significant bias was introduced by the lack of information about groundwater resistivity.

3.1 Resistivity logs

A total of 189 drill holes with downhole resistivity, IR spectroscopy and whole-rock geochemistry logged between 2006 and 2018 by Orano Canada Inc. (formerly AREVA Resources Canada Inc.) for uranium exploration in the South-East of the Athabasca Basin were considered. Downhole resistivity logs were obtained using a dual focused resistivity probe LL3 manufactured by Geovista, United Kingdom [27]. The probe gives two resistivity readings, one shallow measurement with a lateral investigation depth of ca. 10cm, and a deep measurement with a lateral investigation depth of ca. 1m. Vertical resolution of the logs is 0.1m. Prior to processing, all data above 30.000 Ω.m were set to 30.000 Ω.m, since they are considered outside of the device detection limits [27].

According to the usual practice [47] [46] [72], resistivity values were taken as log values in all subsequent steps because of the wide spread of their distributions depending of the medium investigated.

3.2 Petrophysical samples

We integrated petrophysical properties measured by Orano Canada Inc. on cores in the same drillholes where geochemistry and IR spectroscopy were available, to give a comparative view of the responses of resistivity, porosity and density relative to IR spectroscopy and geochemistry. Porosity and density measurements were made by weighting samples after oven drying, water saturation and comparing measurements. Resistivity measurements were performed using a Sample Core Induced Polarization (SCIP) device manufactured by Instrumentation GDD Inc. The device features a holder equipped with two copper disks that function as electrodes. After having been saturated in a copper sulphate solution before every measurement, two cellulose sponges are then placed on the electrodes

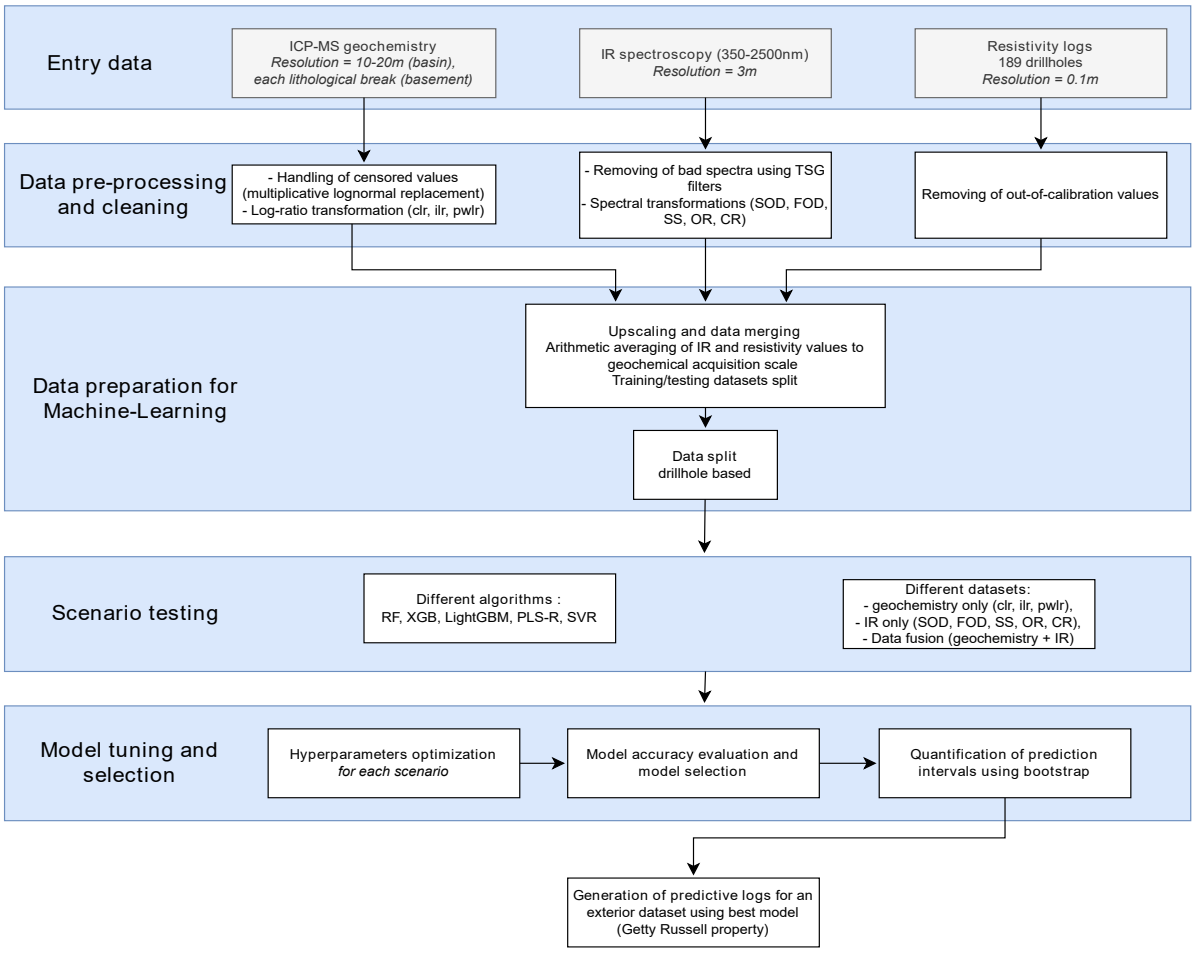

Figure 2: Steps followed for the generation of synthetic resistivity logs. Petrophysical samples were not used for Machine-Learning but only for data analysis and observing correlations.

to enhance connection between the sample and the electrodes [25]. Descriptive statistics for the different petrophysical properties are provided in Table 1.

	Count	Mean	Median	STD	Kurtosis	Skewness	IQR	Q5	Q95	Min.	Max.
Petrophysical samples											
Basin											
Resistivity (ohm.m^{-1})	1530	1941.21	1208.61	3530.72	289.85	14.42	1346.67	375.1	5608.69	89.78	85339.3
Density (g/cm^3)	1530	2.46	2.46	0.06	2.59	0.42	0.08	2.37	2.56	2.22	2.86
Porosity (%)	1530	8.92	8.91	3.05	-0.26	0.11	4.2	3.99	13.93	0.35	20.38
Basement											
Resistivity (ohm.m^{-1})	1380	28626.66	548.46	154777.03	143.73	10.92	5579.57	50.8	96035.13	3.98	2677440
Density (g/cm^3)	1380	2.62	2.61	0.16	7.39	0.94	0.12	2.36	2.85	1.7	3.58
Porosity (%)	1380	4.38	2.63	5.56	10.29	2.72	5.16	0.1	14.8	0	41.95
Downhole logs											
Basin											
Resistivity (ohm.m^{-1})	2643	5083.9	3956.22	3832.86	2	1.32	5091.18	1024.72	12728.77	12.6	29147.79
Basement											
Resistivity (ohm.m^{-1})	1485	5058.97	993.26	7452.09	1.55	1.63	7346.89	109.58	22758.34	10.97	29920.59

Table 1: Summary statistics for petrophysical properties

Core measurements were taken once every 20m in the basin and near the central part of each lithological break in the basement. All samples taken for petrophysical analysis were cut to 100mm in length with a rock saw, so that the core edges are perpendicular to the core first axis.

3.3 Infrared spectroscopy

3.3.1 Data collection

Spectral point measurements on cores in the Vis-NIR-SWIR range were made available for the same drill holes by Orano Canada Inc. These data were obtained between 2006 and 2018 in the 350-2500nm range using a TerraSpec® device, manufactured by Analytical Spectral Devices Inc. (ASD). The device consists of three distinct sensors covering the spectral ranges of 350-1000nm, 1000-1800nm and 1800-2500nm respectively. The measurements obtained from these sensors are subsequently concatenated to generate a composite spectrum. In total, there are 72,641 spectra available for analysis. Core samples were taken and either air-dried for 2 to 4 days or dried in an oven for 8 to 12 hours to avoid water contamination and ensure good-quality analyses. A white calibration standard was used for sandstone samples, and a dark grey standard for basement samples to improve the reflectance of darker rocks (e.g., gneisses and granites). As such, basin and basement samples must be studied separately if one intends to treat them with a quantitative approach using ML.

3.3.2 Pre-processing steps

Spectra were trimmed from the 350-500 and 2400-2500nm windows which are usually highly affected by noise. Following methodology proposed by [17]), convex hull was removed from the raw spectra

in order to remove background effects and better highlight the concentration in alteration minerals. Three other usual spectral transformation methods (simple Savitzky-Golay smoothing of the original reflectance and Savitzky-Golay first- and second-order derivatives,[33]) were also used as a comparison. Mean centering was then applied in order to enhance spectral features and remove baseline shift [33].

3.4 Geochemistry

3.4.1 Data collection

Geochemical data was collected based on two different sampling methodologies, one for basement samples and one for basin samples. Basin samples are composites made of crushed chip samples taken every meter. Composites were collected every 20m in the upper part of the sandstone, and every 10m from 0 to 100m from the unconformity. In the basement, chip samples were taken every meter and then crushed and grouped into composites based on lithological breaks noted by field geologists. These two sampling methodologies were developed by uranium exploration companies several decades ago for the Athabasca Basin and have been applied consistently in their different exploration prospects since then.

Samples were analyzed using the uranium exploration package in the Saskatchewan Research Council (SRC) geoanalytical laboratory in Saskatoon, Saskatchewan. Basement and basin samples were prepared and analyzed in different batches and facilities in order to avoid contamination by one another. The samples were processed using near-total digestion through 3 acids leaching (HClO4, HNO3, HF). Most trace elements were analyzed using ICP-MS, while major oxides were measured using ICP-OES. SiO2, C and loss on ignition (LOI) were not measured. Summary statistics of the geochemical data available is given in Table 2 for the basin and the basement subsets.

3.4.2 Pre-processing steps

Geochemical data is said to be compositional [3], each element given as a concentration and forming part of a whole (e.g. 100%, 106ppm...). This feature of geochemical data needs to be taken into account when dealing with non-detects (censored values) and by the use of log-ratios instead of raw geochemical variables for any statistical treatment.

Handling of censored values Censored values in a geochemical dataset correspond to values that are either "greater than" (right-censored) or "less than" (left-censored) the analytical detection limit (Sanford et al., 1993), and this problem needs to be addressed before any further step in geochemical data analysis. Only left-censored data were present in the considered datasets. Traditionally, single replacement values have been used, such as a fraction of the detection limit (DL), usually $1/3$ to $1/2$ DL. However, using arbitrary substitution values does not take into account the geometric structure of geochemical compositions [53] and can lead to biased estimates of the distributions of geochemical elements [28]. More robust methods that take into account the constraints of compositional data analysis (scale invariance, subcompositional coherence) are usually preferable to single-values replacement strategies [54] [51]. The robust multiplicative lognormal replacement method from the R package *ZCompositions* [52] was used in this study. Elements with over 50% of censored values were also discarded. Using this threshold, no element was discarded in the basement and Bi, CaO, Cd, Cs, MnO, and Na_2O were discarded in the basin.

Log-ratio transformations Due to their compositional nature, geochemical variables are subjected to the total sum constraint, making the use of log-ratio transformations necessary to correctly

Basin	Mean	Std	Min	Q25	Q50	Q75	Max	Basement	Mean	Std	Min	Q25	Q50	Q75	Max
Ag	0.12	0.34	0.02	0.07	0.09	0.12	14.6	Ag	0.26	0.19	0.02	0.15	0.23	0.33	3.4
Al_2O_3 (%)	2.56	1.55	0.54	1.57	2.26	3.21	20.5	Al_2O_3 (%)	15.39	3.45	1.04	14.1	15.5	17	28.9
B	50.03	111.98	1.48	14	22	47	3740	B	136.9	289.23	1.42	35	89	167	6000
Ba	15.68	9.14	3	11	14	18	263	Ba	311.68	314.54	4	51	179	534.5	1980
Be	0.32	0.34	0.08	0.2	0.2	0.4	9.1	Be	3.1	2.37	0.09	1.6	2.5	4	22.5
Ce	45.89	41.37	7	32	40	50	1630	Bi	0.96	4.37	0.07	0.2	0.3	0.6	155
Co	1.63	24.02	0.03	0.12	0.19	0.33	795	CaO	0.8	1.92	0.01	0.12	0.26	0.88	35.3
Cr	12.48	10.63	0.09	8	10	14	288	Cd	0.27	0.31	0.08	0.08	0.2	0.3	7.8
Cu	1.96	25.98	0.2	0.6	0.8	1.2	1310	Ce	74.47	49.77	0.88	42	71	97	661
Dy	1.59	12.42	0.27	0.84	1.06	1.4	633	Co	16.74	26.87	0.11	4.76	12	20.8	550
Er	0.86	6.08	0.2	0.48	0.59	0.78	310	Cr	64.54	54.43	0.78	24.5	59	87	556
Eu	0.34	0.27	0.11	0.24	0.29	0.37	8.3	Cs	3.55	4.79	0.08	1.4	2.6	4.3	114
Fe_2O_3 (%)	0.65	0.99	0.02	0.1	0.29	0.84	20.4	Cu	37.01	83.5	0.3	3.6	10.6	37.2	1030
Ga	3.86	3.25	0.9	2.3	3.1	4.3	63.2	Dy	3.72	15.71	0.06	1.9	3.11	4.23	633
Gd	2.49	3.58	0.7	1.6	2	2.6	119	Er	1.95	7.73	0.04	0.9	1.59	2.2	310
Hf	5.19	2.67	1	3.4	4.7	6.3	33.5	Eu	1.11	0.72	0.07	0.78	1.06	1.33	19.7
Ho	0.28	2.33	0.05	0.14	0.18	0.24	119	Fe_2O_3 (%)	4.16	3.58	0.06	1.56	3.23	5.84	40.8
K_2O (%)	0.29	0.3	0.01	0.11	0.2	0.38	4.94	Ga	21.94	6.82	1.5	18.8	21.3	24.6	97.6
La	20.98	16.93	3	14	18	23	602	Gd	4.4	3.81	0.2	2.7	4.2	5.55	119
Li	13.36	12.73	0.09	6	10	16	211	Hf	4.95	3.07	0.2	3.3	4.6	5.9	41.6
MgO (%)	0.08	0.23	0	0.02	0.03	0.07	7.71	Ho	0.66	2.95	0.02	0.31	0.54	0.73	119
Mo	0.32	1.35	0.06	0.16	0.21	0.3	49	K_2O (%)	3.46	1.76	0.02	2.28	3.38	4.53	11.6
Nb	5.64	3.59	0.7	3.4	4.9	6.9	70.3	La	39.39	25.72	0.86	23	37	51	364
Nd	18.6	23.95	3.2	12	15.4	19.7	973	Li	89.16	47.93	2	53	84	117	401
Ni	3.37	16.43	0.2	0.5	0.8	1.6	355	MgO (%)	3.46	2.35	0.02	1.82	3.12	4.81	24.9
P_2O_5 (%)	0.04	0.02	0.01	0.03	0.03	0.04	0.66	MnO (%)	0.07	0.26	0	0.01	0.02	0.04	5.46
Pb	6.08	8.06	1.04	3.76	4.74	6.06	193	Mo	2.53	7.66	0.02	0.3	0.82	2.63	248
Pr	5.27	5.68	0.9	3.4	4.4	5.7	223	Na_2O (%)	0.8	1.21	0	0.03	0.08	1.29	6.33
Rb	3.5	3.52	0.2	1.6	2.4	4.2	51	Nb	13.56	7.99	0.09	8.15	13	17.4	89
Sc	1.24	1.06	0.3	0.8	1	1.4	25.9	Nd	33	21.73	1.1	18.9	31.6	42.9	310
Sm	3.03	4.13	0.7	2	2.5	3.2	172	Ni	49.23	69.63	0.6	16.3	37.9	62.1	1810
Sn	0.82	1.77	0.11	0.31	0.49	0.88	52.5	P_2O_5 (%)	0.11	0.14	0	0.05	0.09	0.13	2.02
Sr	147.64	95.83	16	87	121	181	1580	Pb	13.78	18.46	0.5	4.89	9.8	18.3	419
Ta	0.6	0.52	0.06	0.3	0.47	0.73	12.1	Pr	9.25	6.11	0.2	5.3	8.9	12	84.1
Tb	0.29	1.32	0.07	0.17	0.21	0.28	65.6	Rb	127.49	75.83	0.6	68.6	120	178	453
Th	22.47	19.64	2.49	11.3	17.3	27.1	426	Sc	12.69	8.53	0.2	5.1	12.9	17.7	66.4
TiO_2 (%)	0.19	0.13	0.03	0.12	0.16	0.23	1.79	Sm	5.77	3.69	0.2	3.4	5.6	7.5	50.7
U	4.53	17.01	0.53	1.18	1.58	2.62	378	Sn	1.47	1.02	0.06	0.8	1.19	1.81	8.93
V	13.77	39.18	0.8	4.2	6.9	12.3	1340	Sr	102.07	106.69	6	43	74	122	1370
W	0.62	2.96	0.01	0.3	0.4	0.6	146	Ta	0.86	0.83	0.01	0.48	0.76	1.05	21.8
Y	8.11	79.1	1.3	3.9	5	6.7	4050	Tb	0.59	1.64	0.02	0.33	0.52	0.7	65.6
Yb	0.86	4.04	0.19	0.51	0.65	0.85	205	Th	19.64	15.72	0.54	10.1	17.5	24.5	184
Zr	247.51	128.34	52	160	221	303	1550	TiO_2 (%)	0.55	0.33	0.01	0.29	0.54	0.75	3.21
								U	9.09	30.88	0.28	2.84	4.42	7.28	847
								V	89.35	81.36	0.08	28.8	82	123	1340
								W	2.27	12.75	0.07	0.4	0.9	1.8	456
								Y	18.75	100.3	0.3	8.6	14.7	20	4050
								Yb	1.73	5.26	0.03	0.74	1.34	2.03	205
								Zn	43.24	51.84	0.78	17	31	56.5	1070
								Zr	159.69	99.18	2	104	148	191	1340

Note: Default unit for trace elements is ppm.

Table 2: Summary statistical table of geochemical elements for the basin and basement datasets.

compare them using multivariate methods [3]. As SiO_2, C and LOI were not measured and some elements were discarded from the sandstone dataset due to their high proportion of non-detects, the remaining elements did not add to a constant sum. In this case, it is usually advised [70] [67] to close the system to a constant sum by adding a filler variable. With N variables in ppm, the filler (F) is given by:

$$F = 10^6 - \sum_{k=1}^{N} x_k \tag{1}$$

with x_k the concentration of elements in parts per million. [71] proposed the use of the pair-wise log-ratio (pwlr) transformation when using geochemical variables along with algorithms that do not show affine equivariance, which is the case for XGBoost, Random Forest and LightGBM among the methods we tested. Pair-wise log-ratio transformation is given by [3]:

$$\text{pwlr}(x) = \begin{bmatrix} 0 & \ln\left(\frac{x_1}{x_2}\right) & \cdots & \ln\left(\frac{x_1}{x_N}\right) \\ \ln\left(\frac{x_2}{x_1}\right) & 0 & \cdots & \ln\left(\frac{x_2}{x_N}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \ln\left(\frac{x_N}{x_1}\right) & \ln\left(\frac{x_N}{x_2}\right) & \cdots & 0 \end{bmatrix} \quad \text{with} \quad x = [x_1, x_2, \ldots, x_N]. \tag{2}$$

The pwlr matrix contains all the possible log-ratios between elemental variables and is readily interpretable in terms of chemical meaning, while capturing all of the information in the dataset [45]. The vector representing the observation x in the ML datasets is the vectorized upper triangular matrix extracted from pwlr(x). It has the drawbacks of having a high number of variables (N(N-1) / 2), yielding high redundancy between variables and increased computational cost. In the present case, these drawbacks are not prohibitive as we are using essentially ensemble algorithms (LightGBM, RFR, XGBoost), which include an in-built feature selection [11] and support vector regression which suffers little from the curse of dimensionality thanks to the kernel trick.

Other common methods that were tested for comparison include the centered logratio (clr) transformation which computes the log-ratio of each component against the geometric mean [3]:

$$\text{clr}(x) = \ln\left(\frac{x_i}{G(x)}\right) \quad \text{with} \quad G(x) = \sqrt[N]{x_1 x_2 \ldots x_N} \tag{3}$$

and the isometric logratio (ilr) transformation [23]:

$$\text{ilr}(x) = V \cdot \text{clr}(x) \tag{4}$$

with V a $(N-1)xN$ matrix whose columns are orthogonal vectors, each summing to zero [67]. ILR transformation was applied using the default function of the pyrolite library [73].

3.5 Data fusion

Given the significant variance in vertical resolutions across the datasets, ranging from 0.1m for resistivity readings to 20m for geochemical composites, the challenge of creating a unified dataset suitable for machine learning (ML) analysis needed to be resolved. Averaging the finest resolution signals to the resolution of the lowest resolution dataset (geochemistry in the present case study) is used by most practitioners and is usually the preferred method for performing such a task [1] [43].

For the present study, resistivity readings were averaged on the resolution of geochemical composites using the arithmetic average. Examples of five randomly sampled drillholes from the dataset

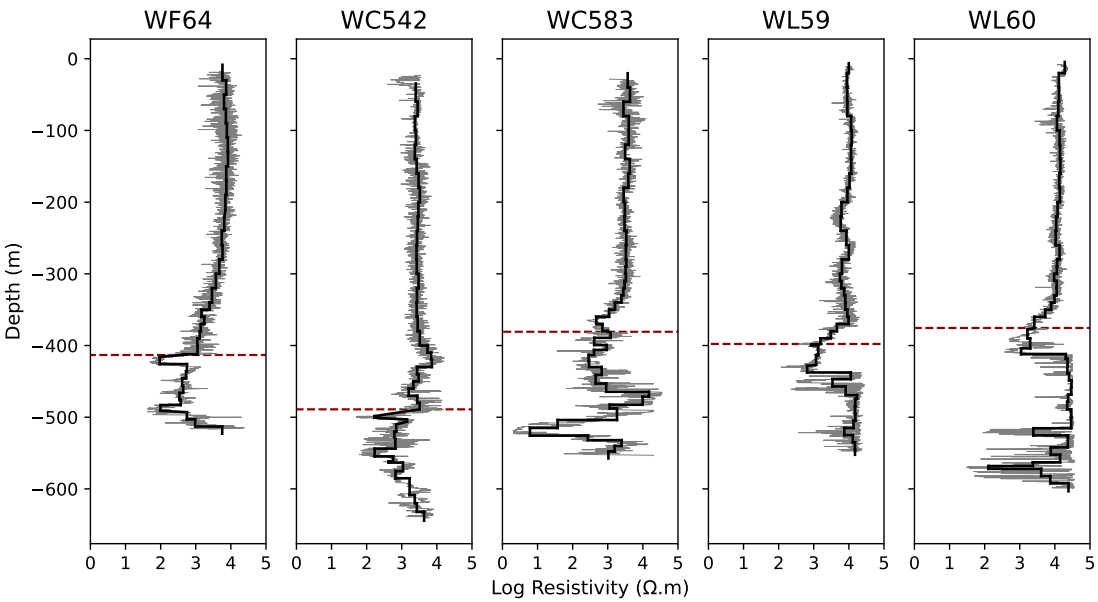

Figure 3: Examples of comparison between raw resistivity values (grey) and averaged resistivity values (black) for 5 randomly sampled drillholes. The dashed red line represents the depth of the unconformity.

prior to and after averaging are given in Figure 3. Petrophysical samples were attributed the geochemistry of the composite interval to which they belonged. All spectra belonging to the given interval were averaged as well, yielding a resulting averaged spectra as a result of the averaging of all wavelengths from the different individual spectra located within the interval.

After data fusion, the dataset was divided into two subsets using the depth of the unconformity noted by field geologists in each drill hole during campaigns. The depth of the unconformity strongly varied between different zones, ranging from ca. 100m at the edges of the basin to ca. 450m in the innermost prospects, with a median value of ca. 200m between all drill holes. In the external dataset (Getty Russell), the depth of the unconformity is approximately 180m. Samples with a depth below the unconformity were grouped into a basement subset, and those with a depth above that of the unconformity into a basin subset. All subsequent steps were performed separately on the two different subsets.

3.6 Training, testing and external datasets

Following data analysis, an external dataset (Getty Russell property, Figure 1) was excluded from the remaining data. It was neither used for training nor testing. The best model was used with this external dataset to infer synthetic resistivity logs to give an illustrative view of the results that could be produced on an entire area where no downhole electrical measurements are available. As in any machine-learning application, it is important to assess the prediction results over an external dataset independent from both the training and testing data to avoid estimation bias.

Each subset (i.e., basement and basin) outside of the Getty Russell property was then split between a training dataset and a testing dataset. 25% of the drillholes were randomly attributed to the testing dataset and 75% to the training dataset (Figure 4). This method was preferred to

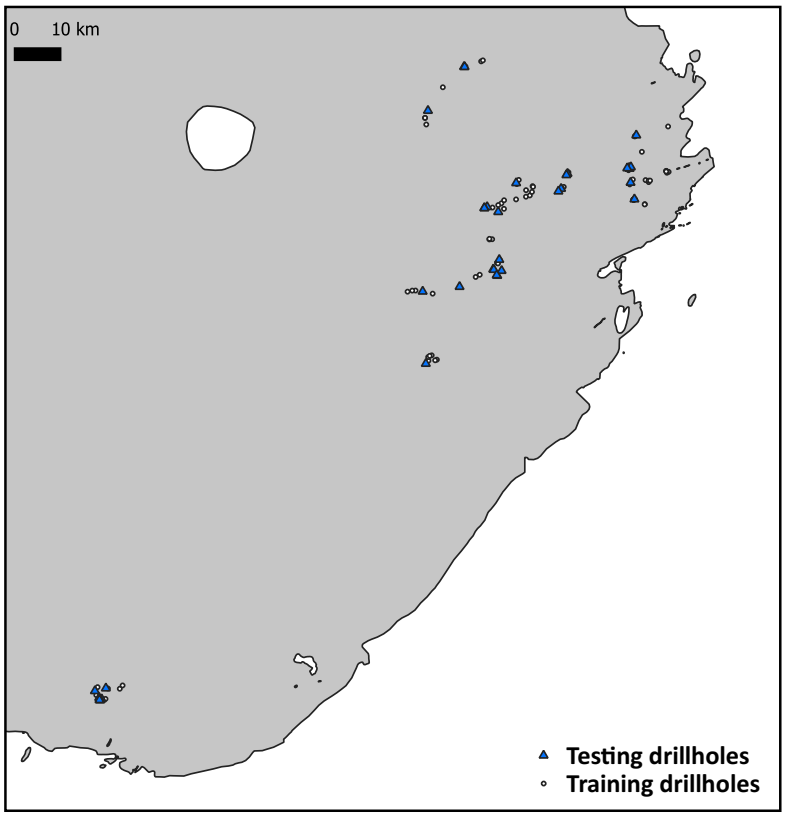

Figure 4: Location of the testing and training drillholes split using a blind well splitting strategy.

random sampling of individual samples, as otherwise neighboring samples within a same drillhole could be split between the training and testing data, yielding spurious high-accuracy results that would not be reproducible in a practical case (e.g., [18] [42]).

3.7 Data analysis methods

Statistical analysis was performed using Python. Machine-Learning methods were drawn from the *scikit-learn* [55], *lightgbm* [34] and *xgboost* [14] libraries.

3.8 Prediction methods

3.8.1 Machine-Learning methods

Five state-of-the-art regression methods representing the main families of ML models (SVR, RFR, PLS-R, LGBR and XGBR) were chosen. They represent representative families of ML methods: kernel-based machines (SVR), decision tree based methods (RFR), recent ensemble methods (LGBR and XGBR) and PLS-R, which has been efficient for calibration problems applied to IR spectra. A brief description of each of the methods is given in Table 3. Interested readers are referred to references therein for more details about the specific implementation of each model.

Method	Description	Hyperparameters	Reference
Random Forest	RF works by building a set of decision trees (Bagging technique)	n_estimators, max_depth, max_leaf_nodes, min_samples_split, etc.	[11]
Partial Least Squares Regression	Partial Least Squares (PLS) regression is a statistical method that finds the fundamental relations	number of components	[26]
XGBoost	XGBoost regression is a popular machine learning technique designed to improve gradient boosting performance	max_depth, gamma, learning_rate, subsample, colsample_bytree, etc.	[14]
LightGBM	LGBM regression, also known as LightGBM regression, is designed to handle large datasets efficiently	n_estimators, reg_alpha, reg_lambda, colsample_bytree, etc.	[34]
Support Vector Regression	Support Vector Regression (SVR) is a machine learning method that performs linear regression in a higher dimension	kernel, degree, gamma, epsilon, regularization	[20]

Table 3: Summary of the different prediction methods used in this paper.

3.8.2 Hyperparameters optimization

For a given prediction method, hyperparameters are the ensemble of parameters chosen to build the model, including, for example, the number of decision trees used in a random forest or the maximum number of variables to be randomly selected for each decision tree at each iteration. There is no rule of thumb as to what are the best hyperparameters to choose for a given study case and this task should always be performed prior to model selection. Hyperparameter optimization was performed using the Tree-structured Parzen Estimator Sampler [8] implemented in the optuna library [4]. Each model was run for 200 trials for each possible scenario, after which no significant improvement in performance metrics was observed. Hyperparameters were selected by cross-validation of the coefficient of determination on the training dataset.

3.8.3 Prediction intervals

In Machine-Learning, the overall quality of a regression method can be estimated using the Root Mean Squared Error (RMSE) computed on the test dataset:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

with $\hat{y}_i$ and y_i the predicted and actual values associated to observation i respectively. Another common performance metrics is the coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

with $\hat{y}$ the arithmetic mean of actual values.

However, neither the RMSE nor the R^2 provide any clue about what regions in the dataset were predicted with the most accuracy or stability [66]. We argue that this point is particularly important when dealing with petrophysical downhole data, as having information about uncertainty for each prediction could raise flags on anomalous zones and geological processes at stake, thus enhancing the interpretability of the model.

Confidence intervals are usually computed using bootstrapping. Bootstrapping is a statistical resampling procedure that aims at estimating the variability or uncertainty of a given statistic by repeatedly resampling the original data, with or without replacement [22]. Algorithm 1 summarizes the procedure used, termed 'non-parametric bootstrap resampling' as it works without making any assumptions about the distribution (most importantly, the normality) of the target variable [13]. The number of bootstrap samples n is typically taken as a value between 50 and 200 [69] and was set to 100 for this study.

Algorithm 1 Procedure followed for estimating prediction intervals through bootstrap resampling

1: **Input**: Original dataset divided into a testing dataset of m samples (25% of the original data) and a training dataset containing the remaining samples.
2: Run the model using the complete training data and obtain predictions for each of the m samples.
3: Generate n bootstrap samples of size 100, with replacement, from the original training data.
4: **for** $i = 1$ to n **do**
5: Fit a given predictive model using the i-th bootstrap sample as training data, and make predictions on the testing data.
6: Append the array of the m predictions as a row to a matrix.
7: **end for**
8: **for** each sample m_i **do**
9: Compute a 0.95 confidence interval: $\mathrm{CI}_i = \hat{y}_i \pm 1.96\sigma_{m_i}$, where σ_{m_i} is the standard deviation of the bootstrap predictions for m_i and $\hat{y}_i$ is the prediction for m_i using the complete dataset.
10: **end for**
11: **Output**: An array of length m with prediction intervals associated with each individual sample.

3.8.4 Model selection and model accuracy evaluation

Prediction models were fit for the 8 different scenarios considered for each of the two subsets (basement rocks and basin rocks). For each model, log values of resistivity are taken as the dependent variable. 8 different scenarios for independent variables were tested:

- Using geochemistry only. Geochemical data was considered as centered log-ratios (CLR), isometric log-ratios (ILR) and pairwise log-ratios (PWLR).
- Using IR spectroscopic data only. The different scenarios tested include IR spectroscopy taken as raw spectra after simple smoothing (SS), using first-order derivative (FOD), second-order derivative (SOD) and continuum removal (CR).
- Data fusion of both datasets, each time using the data transformation for IR and geochemistry that performed best in the two previous series of models.

Each model was run for 100 optuna trials using 5-fold cross-validation on the training dataset and optimized by maximizing the coefficient of determination (R). The best scenario for using IR only

	PLS-R		LGBR		XGBR		RFR		SVR	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
IR										
Sandstone										
SS	0.55	0.24	0.55	0.24	0.52	0.25	0.51	0.25	−0.54	0.45
FOD	0.65	0.21	**0.71**	**0.19**	0.66	0.21	0.65	0.21	0.70	0.20
SOD	0.55	0.24	0.67	0.21	0.59	0.23	0.58	0.23	0.66	0.21
CR	0.55	0.24	0.44	0.27	0.39	0.28	0.51	0.25	−0.31	0.41
Basement										
SS	0.38	0.64	0.35	0.65	0.23	0.71	0.28	0.69	0.49	0.58
FOD	0.38	0.64	0.50	0.57	0.48	0.59	0.27	0.69	**0.55**	**0.54**
SOD	0.46	0.60	0.48	0.58	0.41	0.62	0.27	0.69	0.50	0.57
CR	0.39	0.64	0.34	0.66	0.28	0.69	0.28	0.69	0.42	0.62
Geochemistry										
Sandstone										
CLR	0.77	0.17	0.77	0.17	0.65	0.21	0.65	0.21	0.78	0.17
ILR	0.77	0.17	0.78	0.17	0.70	0.20	0.62	0.22	0.79	0.17
PWLR	0.77	0.17	**0.80**	**0.16**	0.70	0.20	0.74	0.18	0.77	0.17
Basement										
CLR	0.75	0.41	0.79	0.37	0.71	0.44	0.68	0.46	0.78	0.38
ILR	0.75	0.41	0.79	0.37	0.65	0.48	0.68	0.46	0.79	0.37
PWLR	0.75	0.41	**0.80**	**0.36**	0.72	0.43	0.70	0.45	0.80	0.36
Data fusion										
Sandstone	0.78	0.17	**0.80**	**0.16**	0.75	0.18	0.72	0.19	0.31	0.30
Basement	0.75	0.40	0.79	0.37	0.66	0.47	0.68	0.46	**0.82**	**0.35**

Note: The best scenario for each subset and each type of data is highlighted in green. The overall best scenario for the basin and basement were used for the generation of synthetic logs.
Abbreviations: CLR, centred log-ratio; CR, continuum removal; FOD, first-order derivative; ILR, isometric log-ratio; PWLR, pairwise log-ratio; SOD, second-order derivative; SS, simple smoothing.

Table 4: Results of R^2 and root mean square error (RMSE) for resistivity prediction on the testing dataset.

and geochemistry only was then kept for data fusion. The best model was then selected based on the R obtained on the testing dataset. Table 4 gives a complete summary of $RMSE$ and R for all tested scenarios. R and $RMSE$ were computed for the training and testing datasets in order to ensure the models were able to aptly generalize to unseen data.

4 Results

4.1 Descriptive statistics

In the basin (Figure 5a and 6a), resistivity follows a roughly log-normal distribution, although with a heavy tail towards high values as exemplified by positive coefficients of skewness in core measurements and mean being consistently higher than the median, both for core measurements and downhole logging values. Resistivity values between core measurements and downhole probing cannot be compared as they were obtained using different devices, including different induced currents, and only a comparison of the characteristics of the distributions is possible. Resistivity probing measurements display similar patterns to that of core measurements with a slightly heavier tail towards

positive values (median = 3956.22 Ω.m; mean = 5083.90 Ω.m; skewness = 1.32) and a low number of outliers (kurtosis = 2.00).

Density is the only parameter showing a slightly negative skew in the sandstone, indicating a slightly stronger weight of lower values on the overall statistical distribution, as well as a highly peaked distribution with kurtosis values of 9.21 and 14.47 respectively, indicating a low number of outliers.

Porosity appears to follow a strongly symmetrical distribution, with similar mean and median values ($\approx 9\%$) and a spread of values similar to that of a standard normal distribution as evidenced by its kurtosis and skewness close to 0, with 90% of the porosity values lying between 4 and 14%.

In the basement (Figures 5b and 6b), resistivity values for both core measurements and downhole probing display a much wider spread than in the basin, with an IQR of 7346 against 5091 Ω.m for probing measurements and 2686 against 1346 Ω.m for core measurements as well as higher standard deviations for basement rocks compared to their sandstone counterparts. The distribution of downhole probing values appears very strongly bimodal, with two distinct peaks around ≈ 1000 and 15000 Ω.m. This is probably due to the differentiation between unaltered lithologies and rocks affected by either graphitization or clay or chloritic alteration [59].

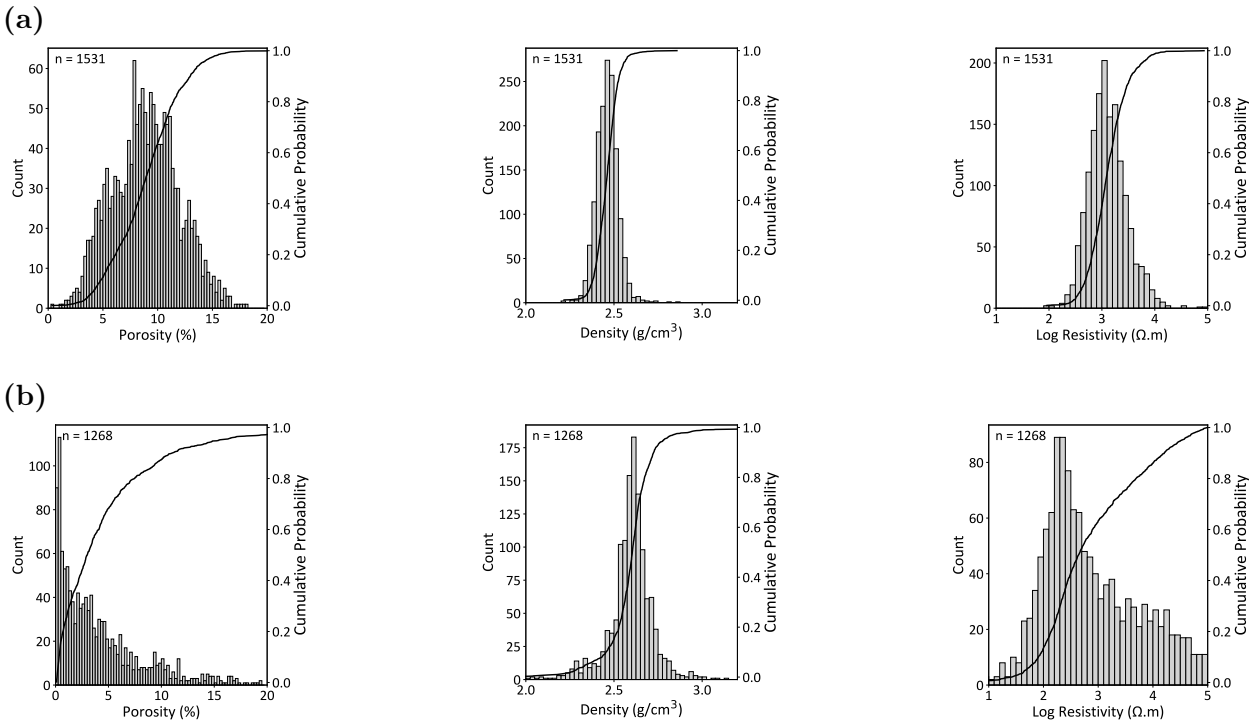

Figure 5: Histograms of physical properties (porosity, density and log-resistivity) for petrophysical samples in the Athabasca (a) basin and (b) basement, along with their associated empirical cumulative distribution functions

4.2 Multivariate analysis

4.2.1 Geochemical controls on physical properties

Geochemical data was transformed using CLR and used for exploratory data analysis with Partial-Least Squares (PLS) modelling using the different physical properties as target variables. PLS can

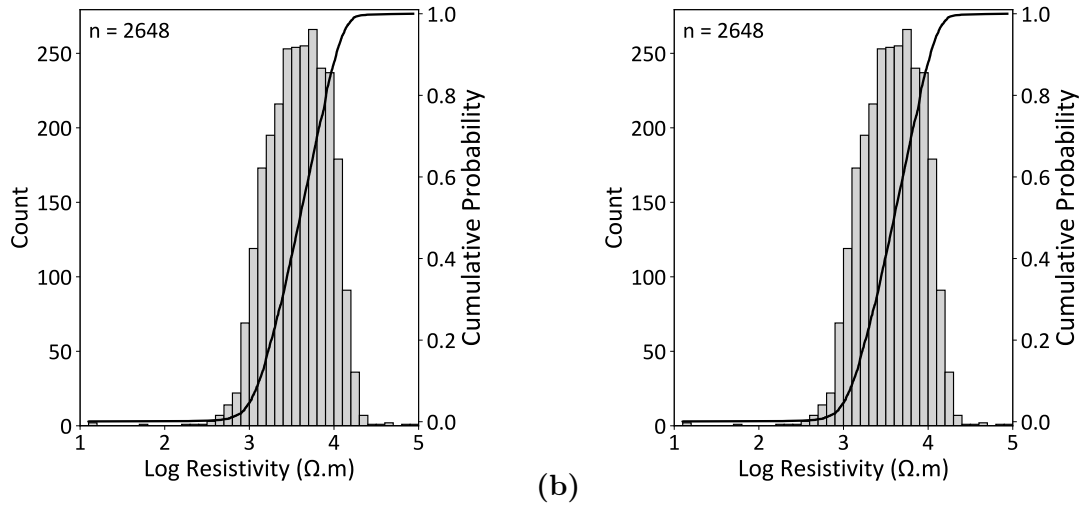

Figure 6: Distribution of resistivity values for downhole logs in the Athabasca basin (a) and basement (b).

be used in a similar fashion to traditional Principal Components Analysis (PCA) for exploratory data analysis, with the advantage that PLS latent variables will be correlated to a target variable, in this case a physical parameter, making it more suitable to the study of the relationships between a whole dataset and a specific parameter [26]. Results for the two first latent variables are shown in Figures 7 and 8. No significant visual information was observed in the following latent variables.

In the basin (Figure 7), samples with low density and resistivity values, along with higher porosity values appear to be somewhat correlated to a suite of elements including mainly K_2O, CaO, MgO, Ni and Li associated for the most part to the presence of illite (being the main primary bearer of K_2O in the sandstone), Mg-chlorites or magnesiofoitite (MgO), sudoite (Li), and sulphides (Ni, Co). Higher density and resistivity samples are mostly related to a group of elements including most of the immobile and High-Field Strength Elements (TiO_2, Th, Zr, Hf, Nb, Sn and some rare-earths like Sm, Pr, La, Ce, Yb) reflecting higher proportions of heavy minerals such as zircons and titanium oxides.

In the basement (Figure 8), lower resistivity and density values tends to be associated with a group of elements including Al_2O_3, TiO_2, Ni, V, Li, Co, MgO, B, U and some HFSE (Zr, Ta, Nb, Sn), while most REE appears to show a very light association to higher density and resistivity samples (Gd, Er, Ho, Yb, Sm, Ce, La, Pr, Eu). K_2O, MnO, Zn, Ba, and CaO on their side are associated with higher density and resistivity samples. In the case of MnO, the very high resistivity observed in garnetites and mafic gneisses from the basement [46] is most certainly the origin of its strong positive correlation with resistivity.

4.3 Mineralogical controls on physical properties

Figures 9a to 9d present spectra of the mean correlation of different wavelengths with physical parameters for the different spectral transformations used. According to the Beer-Lambert law, the intensity of the absorption feature of a chemical species is proportional to its concentration in the considered material [17]. However, this determination is considerably more complex in practice [29], and no clay content data was available for this study. Therefore, as the nature of the relationship between clay content and intensity of the absorption features was not known, Spearman's rank correlation was preferred to Pearson's linear correlation coefficient. Hull quotient removed spectra

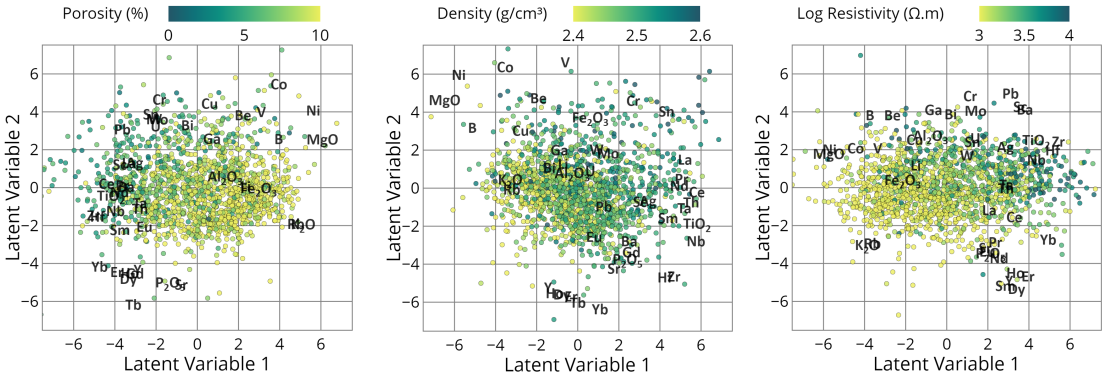

Figure 7: PLS latent variables for geochemical variables against porosity, density, log-resistivity collected on core samples in the sandstone

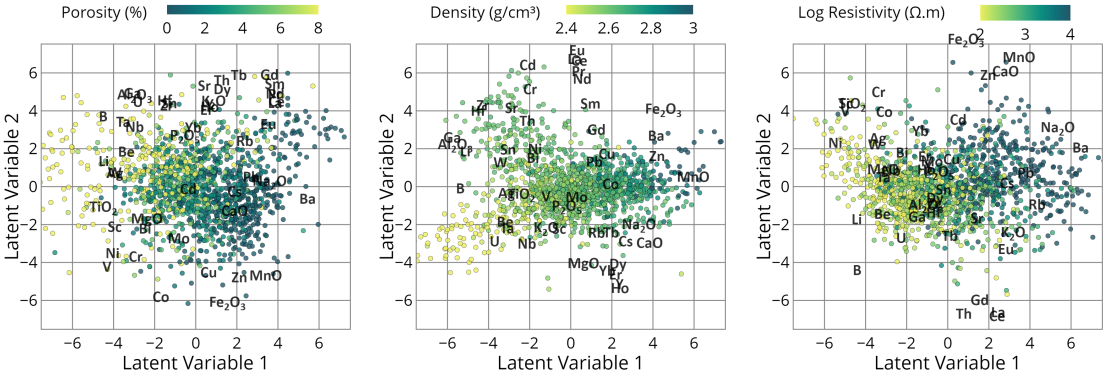

Figure 8: PLS latent variables for geochemical variables against porosity, density and log-resistivity collected on core samples in the basement

are the most interpretable in terms of mineralogical significance, as hull quotient removal preserves and enhances peaks associated to different mineral species.

In the basement, the most significant absorption features are seen around 1900nm, 2200nm, and 2250nm which corresponds to the signatures of hydroxyl bounds, Al-OH bounds and Fe-OH respectively [37]. All three petrophysical parameters appear to roughly reproduce the absorption spectrum of an ideal chlorite, and, to a lesser extent, to that of an illite (Figure 9e). The importance of the Fe-OH bound relative to the Al-OH bound is more marked for resistivity than for density and porosity, hinting at a possible increased effect of chlorite on resistivity compared to the two other properties.

Hematite displays a diagnostic absorption peak around 900nm [37] positively associated to density and negatively to porosity. Hematized sandstone are usually slightly denser and less porous than their bleached counterpart, where hematite coatings have been removed.

4.4 Prediction of resistivity signals

Results of resistivity prediction using the various methods are summarized in Table 4. Details regarding the specific implementation, computing time, definition of hyperparameters search spaces

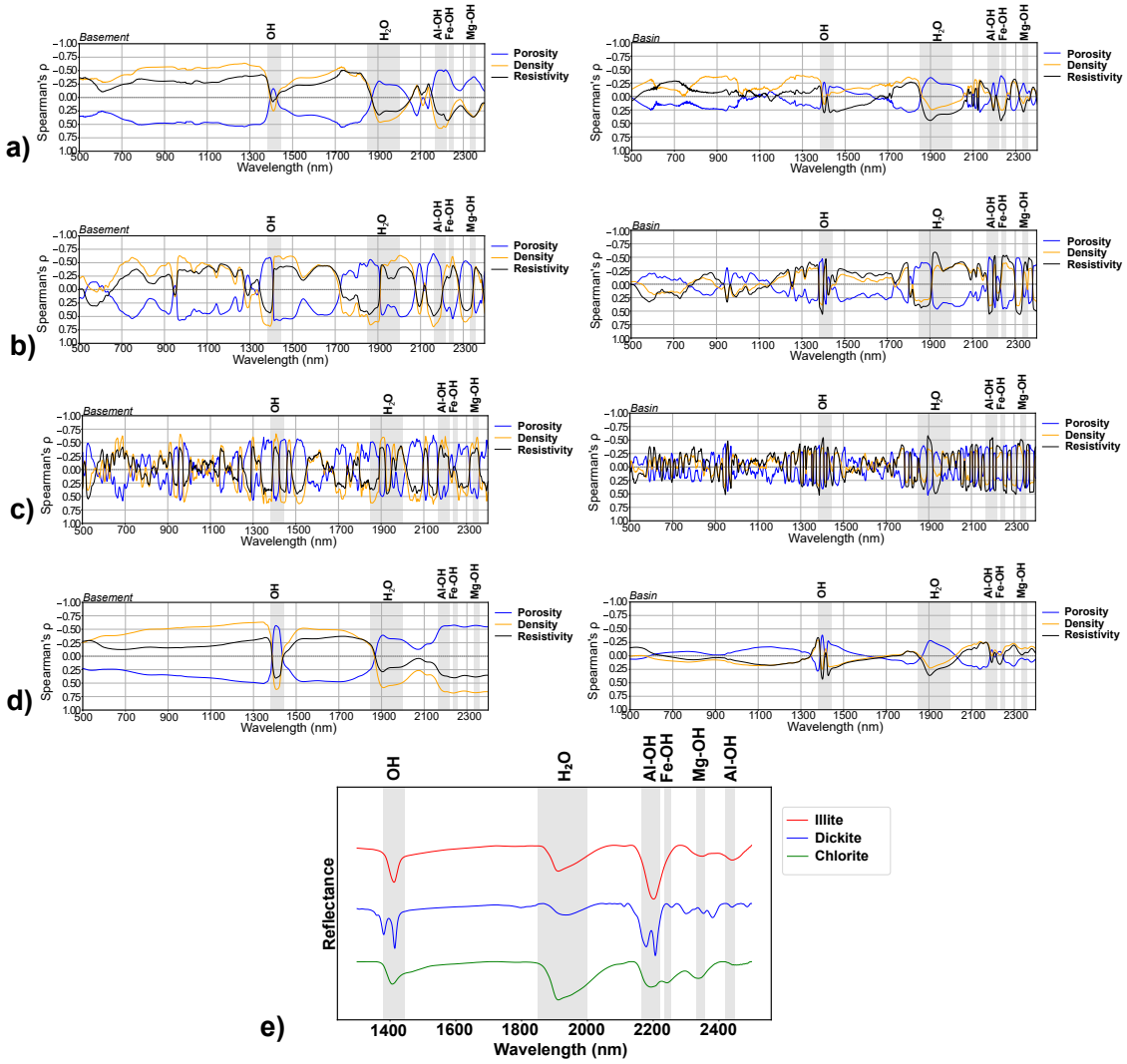

Figure 9: Correlation spectra of physical properties for (a) hull quotient removal, (b) first derivative (c) second derivative and (d) simple Savitzky-Golay smoothing. Y-axis was voluntarily reversed, as it measures correlation with absorption depths, i.e., a strong negative correlation of a property with a given spectral region means that it is positively related to the associated mineral species. (e) Reference of pure continuum-removed IR spectra of the main minerals seen through IR in the Athabasca basin and basement

and optimization for each model are provided in ESM-1. In two specific models (out of 80), SVR with simple smoothing and continuum removal on the basin subset were unable to generalize to unseen data, yielding negative R. Most models yielding R values superior to 0.7 displayed rather small generalization gaps between training and testing datasets, indicating a good capacity to generalize to unseen data.

For predictions using solely IR spectroscopy, the first derivative was the most effective method for most algorithms and was used for data fusion, with simple smoothing yielding the poorest results. In terms of log-ratio transformations for geochemical variables, pwlr appears to slightly outperform other log-ratio transformations for most methods, so pwlr was preferred as the method for data fusion with IR spectroscopy.

In the basin, the best model obtained used geochemistry along with a Light Gradient Boosting Machine (LightGBM) and yielded an $R = 0.80$; $RMSE = 0.16$. In the basement, the best model was a Support Vector Regression and yielded an $R = 0.82$; $RMSE = 0.35$ using data fusion of geochemistry and IR spectroscopy. The best model using IR spectroscopy was obtained using a LightGBM on the first spectral derivative in the basin and a Support Vector Regressor in the basement, giving $R = 0.71$; $RMSE = 0.19$ for the former and R= 0.55; $RMSE = 0.54$ for the latter. Scatterplots of real versus predicted values for these scenarios are given in Figure 10 for the best models. All other scatterplots related to different ML methods in the basement and the basin are also provided as supplementary material in ESM-1. The variance of resistivity values in the basement subset is much higher than that of the basin subset as was shown in Table 3 and Figure 5, explaining higher $RMSE$ given similar R values. Among all the methods, LightGBM appeared to consistently outperform other methods for the basin. In the basement, SVR yielded slightly better results. Prediction using geochemistry only was also found to be consistently better than using IR spectroscopy. Adding IR spectroscopy to geochemistry slightly improved prediction results only in the basement. Gain-based feature importances were computed for the basement and the basin, using the best performing scenario with LightGBM which features an in-built feature importance based on importance gain (Figure 10). The two best models for the sandstone and the basement were fitted on the external dataset (Getty Russell property) and concatenated to obtain complete synthetic logs for comparison with actual resistivity domains and assess the generalizability of the model. Synthetic logs are given in Figure 12. They are plotted next to visual logging of important geological variables performed by exploration geologists and made available by Orano Canada Inc. including graphite, clay and chlorite contents as well as the intensity of silicification noted on a scale from 0 (absent) to 4 (abundant).

5 Discussion

5.1 Features importance

Filler appeared to have a strong influence on the resulting model in the sandstone, as Filler/MgO was by far the most important feature used in terms of gains on prediction (Figure 11). As a reminder, SiO2 was the only non-measured major oxide and is estimated to represent over 98% of the bulk geochemistry of the sandstone [6]. As a consequence, the filler can be taken to approximately represent the SiO_2 content of the rock (the only other major not included component in the basin being LOI), and the most representative variables would thus roughly correspond to SiO_2/MgO, MgO/Ba, SiO_2/K_2O, MgO/Hf, SiO_2/B and K_2O/Ba. These variables all consist of log-ratio containing one immobile element (SiO_2, Hf, Ba) and a mobile element signing hydrothermal alteration, K_2O being primarily a signature of illite [5], B a signature of hydrothermal tourmaline [2] and MgO a signature of Mg-chlorites [46]. These results are in line with the correlations seen with multivariate analysis

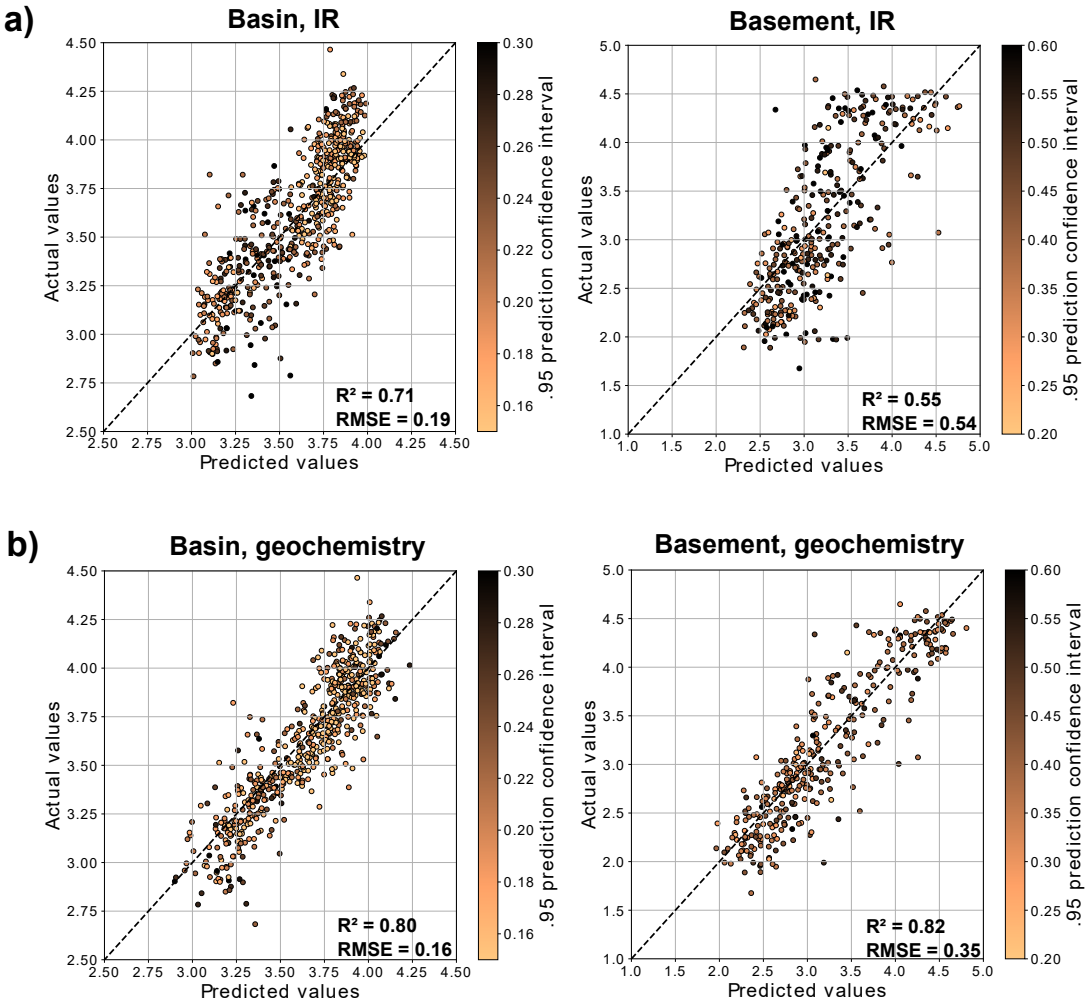

Figure 10: Scatterplots for the best models on the validation dataset for both subsets using a) best models using IR spectroscopy only (basin: LGBR using FOD; basement: SVR using FOD) b) best models using geochemistry (basin: LGBR using PWLR geochemistry; basement: SVR using data fusion of geochemistry and IR)

comparing geochemistry and petrophysical measurements on cores.

In the crystalline basement of the Athabasca, MgO/Na_2O and Al_2O_3/Na_2O were by far the most influent features, followed by MnO/Ni, B/Filler and Eu/Sr, with MgO and Al_2O_3 also translating mainly to the presence of chlorite and illite, respectively. These correlations are in agreement with what was seen through correlation analysis of IR spectroscopy against physical parameters, in which low resistivity values were primarily correlated to the 2250nm Fe-OH absorption bands associated with chlorite and to a lesser extent with the 2200nm Al-OH absorption band associated with illite.

5.2 Potential applications of the method

Synthetic logs obtained allowed to reconstruct resistivity domains well for most drillholes, for both the basin and the basement subsets. Geochemistry consistently outperformed IR spectroscopy for

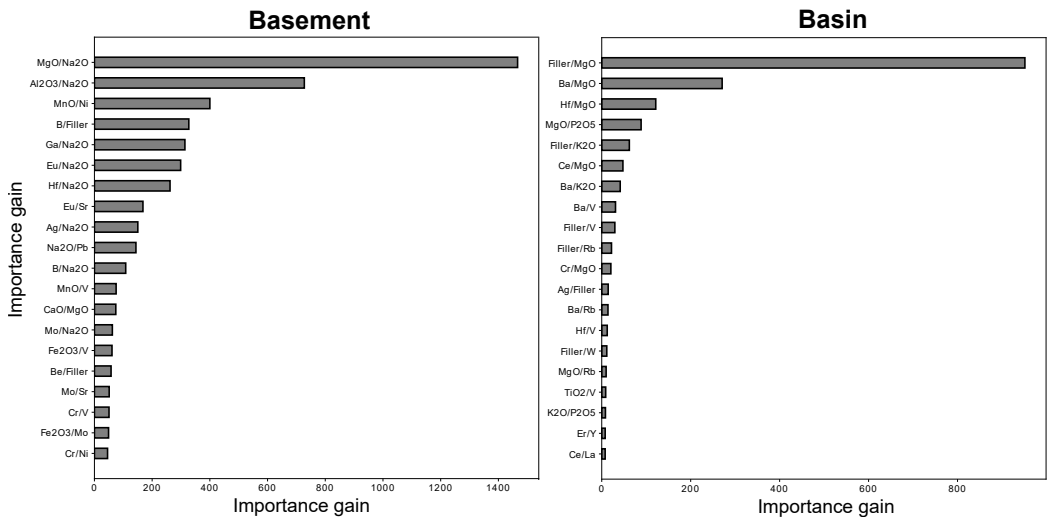

Figure 11: Features importances for LightGBM (only the first 20 features are displayed) using pwlr geochemistry for both subsets

the predictive task, although for the dataset located in the sedimentary part of the studied area the difference was much smaller than for the crystalline part.

Predictive tasks are usually mostly performed the other way around (i.e., predicting geochemistry using petrophysical logs, e.g., [18]) due to the higher cost of the former compared to the latter. However, geochemistry and IR spectroscopy have long been used routinely in several mining districts worldwide, which is not always true for resistivity logs. Such a method could be used for the reassessment of historical prospects that were not logged for electrical properties several decades ago, in order to produce synthetic data usable as entry data for 2D and 3D geophysical models. This also extends to sectors where companies tend to relaunch exploration (e.g., hydrogen, helium, geothermal fields) in prospects where drilling was previously performed for other commodities (such as coal or hydrocarbons; e.g., [30]) and where widespread historical data is already available to perform such work.

5.3 Biases and sources of errors

In the basinal part of the studied drillholes, resistivity can be summarized as the combined effect of three interdependent processes: (i) connected porosity, (ii) clay alteration and type of clay minerals [50] and (iii) groundwater chemistry, especially salinity (e.g., [65]). Groundwater chemistry and resistivity were not available for this study and thus were not considered. A hydrogeochemical study performed by [21] showed that the total dissolved solids (TDS) content of Athabasca sandstone groundwater significantly varies between zones (50 mg/l to 60'000mg/l) but that samples originating from the southern and eastern edges of the basin all displayed TDS values on the lower end of the distribution, inferior to 500mg/l due to higher recharge from meteoric waters. This is confirmed by recent reports [58] on the McClean Lake area which found all groundwater samples displaying TDS lower than 1000mg/l with average values ranging from ca. 40 to 80mg/l. Furthermore, stratigraphic and lithological variations were found to be rather independent of groundwater chemistry, with samples displaying similar ionic compositions independently of the rock they were extracted

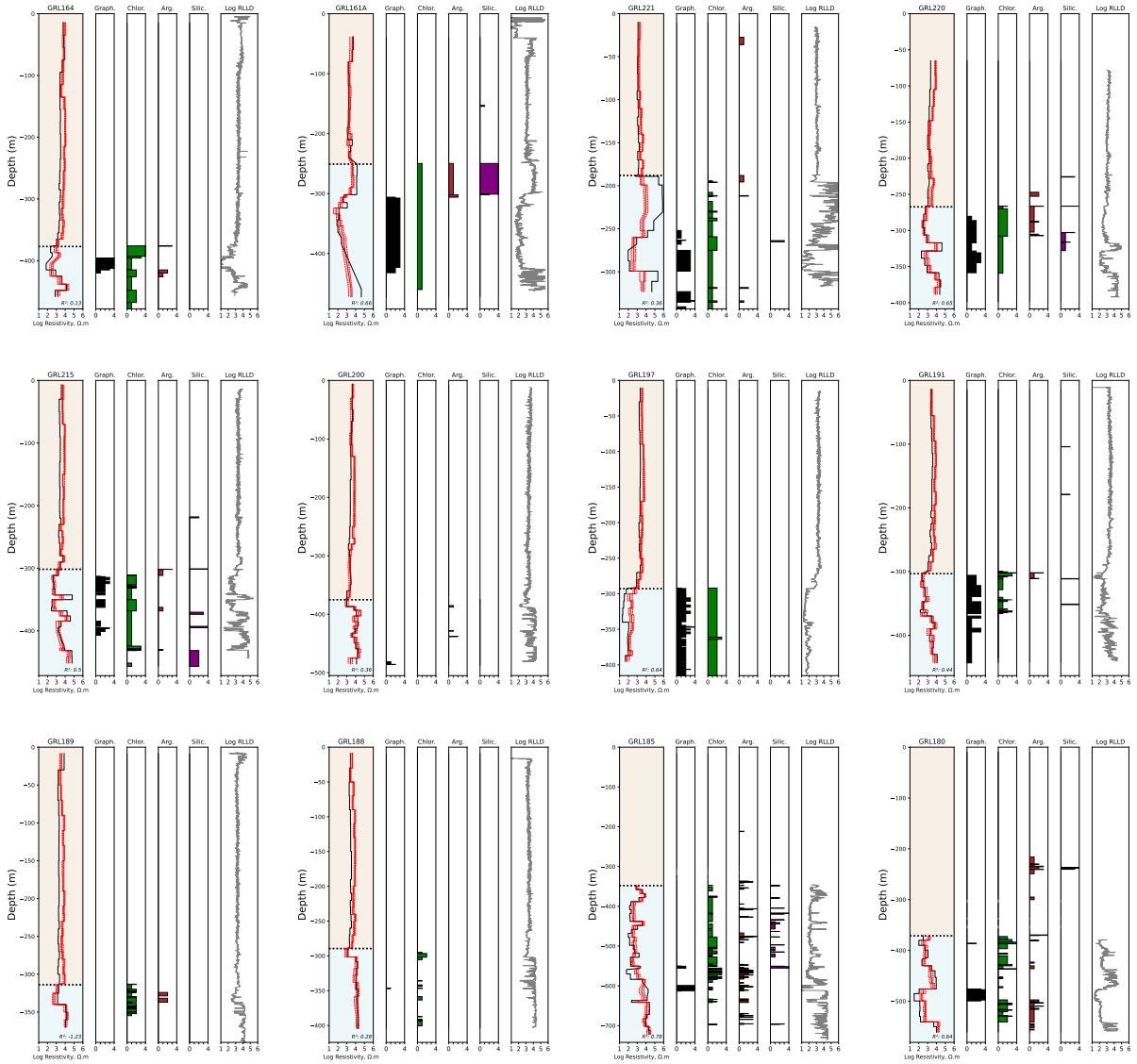

Figure 12: Synthetic logs for 12 drillholes of the Getty Russell property and associated geological logs (noted by Orano Canada logging geologists, on a visual scale from 0 to 4). Horizontal dotted lines correspond to unconformity depth, light orange and light blue domains to the basin and basement, respectively. Black full line corresponds to the ground-truth, red full line to the prediction using LightGBM along with geochemistry, dashed red lines to the upper and lower limits of the prediction intervals using bootstrap. Graph. = graphitization ; Chlor. = chloritization ; Arg. = argilization ; Silic. = Silicification. Log RLLD corresponds to the original deep resistivity probing (before averaging on geochemical composites). Associated R for each drill hole are given on the bottom right of synthetic logs. GRL180 and GRL185 were not run for resistivity in the basin

from [58]. On this basis and by lack of consistent hydrogeochemical measurements across the study area, we considered groundwater chemistry to be uniform between the different drillholes, as all areas considered in the study came from the southern and eastern edges of the basin. One should however be aware that not integrating hydrogeochemical variations in the study could be a potential source of error in the results, in the case of local or regional variations between drilled areas. In the basement, groundwater chemistry can be considered negligible as it has barely any influence on the meso- and micro-scales, and thus on the bulk electrical behavior of rocks at the scale considered.

Another, more significant, source of bias in the present study is due to the widespread presence of graphite in the Athabasca basement. Graphitic zones have been long shown to display anomalous resistivity responses due to the high conductivity of graphite [47][46]. Carbon was not part of the whole-rock geochemistry assays made available for this study and, having no signature in the SWIR domain, the considered dataset did not offer any opportunity to take into account the strong influence graphite has on resistivity. This effect is seen on the synthetic logs (Figure 12) where strongly graphitized intervals appear to be consistently overestimated, for example in GRL164, GRL180, GRL185, GRL197 and GRL221.

A third important potential source of bias in this study may be due to the absence of SiO_2 measurements. Silicification has been shown to result in a significantly increased resistivity signal [32][59]. However, as mentioned above, this information is partially hidden within the filler variable and thus, to some extent, present in the geochemical data. The impact of the lack of SiO_2 values is harder to estimate as it is less widespread in the predicted logs, and often limited to very small intervals whose resistivity response would be hidden after averaging. The only drillhole displaying high silicification values over a large interval is GRL161A in which associated predicted values appeared to be slightly underestimating actual values, hinting that the lack of SiO_2 may still degrade the model. Values associated to another silicified interval in GRL220 are also seen to be underestimated around 330m depth.

However and in spite of a more complex geology, the basement dataset did not yield a lower overall predictivity than the sandstone, being even slightly higher for the best models. This is probably due to the differences in sampling methodology, as basement samples were taken at each lithological break, giving more consistent resistivity intervals based on changes in rock characteristics while sandstone samples were taken arbitrarily every 20 or 10m. Although geochemistry appears to be slightly superior to IR in terms of predictive power, it is important to note that we used geochemistry from samples analyzed using ICP-MS, whilst IR spectroscopy is a cheap and flexible method that can be used on site during exploration and provide immediate results.

5.4 Controls on physical properties

The Bulk-Grain-Texture (BGT) model for petrophysical parameters [19] is a recent conceptual framework for explaining primary controls on physical parameters. According to this model, physical parameters are divided between parameters dominated by the bulk composition of the rock (geochemistry and mineralogy) such as density, parameters driven by grain parameters (volume, size, shape, such as ferromagnetic properties) and parameters driven by textural properties (porosity, pore connectivity and geometry and permeability, such as electrical properties).

According to this model, electrical properties are the less likely features to be predicted using geochemistry and mineralogy, as they are primarily governed by rock texture, whose variations is only to a limited extent due to variations in chemical and mineralogical processes. We argue however that there are reasons why this study was able to provide, at first-order, constraints on electrical properties of rocks using geochemistry and IR spectroscopy:

- In the sedimentary part of the studied area, it should be reminded that our dataset was ex-

tremely homogeneous in terms of lithology, not only because it solely included chemically highly mature Athabasca sandstones, but also because it included samples coming from the very same sub-basin and the very same formation (Manitou Falls Formation), whose members have been shown to display very little lithogeochemical variations (Alexandre, 2021). In this case, most of the lithogeochemical and spectral variations were due to hydrothermal alteration and alternances of conglomeratic and fine-grained layers, both features being related to connected porosity.

- In the basement, all publications that presented a study on resistivity values in the crystalline rocks observed very different values for different lithologies in the basement, suggesting at least some type of lithogeochemical control on the electrical properties of the rock [47] [46].

This study showcases the potential of segmented learning for improving overall ML prediction quality. In this case, segmented learning here corresponds to splitting a dataset into smaller subsets (here, basement and basin) on which all ML tasks are performed separately before rearranging them for obtaining synthetic logs. As resistivity displays very different controls based on a rock texture [19] and thus its lithology, different subsets will display different controls on resistivity (see Figures 7 and 8 for the basin and basement subsets presented in this study). As an example, K_2O and Rb were associated with resistivity lows in the basin because it is associated with illite. In the basement, these elements are associated with resistivity highs, because they are mainly associated with fresh, highly resistive granites less affected by fluid circulations. Therefore, training specific models for each subset instead of one global model will better capture the different controls on resistivity properties.

In practice, this approach has however two limitations: it can become time consuming for complex datasets, and the resulting subsets may not have enough samples for ML models to be trained on sufficient data. It is thus needed to find a balance between the size of the different subsets, the degree of precision one wants to obtain and the effort one is willing to put into the task.

6 Conclusions

The pipeline presented in this paper tested several different scenarios using IR spectroscopy (hull-quotient removal, first and second derivatives and simple smoothing of the original reflectance) and ICP-MS geochemistry (pwlr, ilr, clr transformations). Geochemistry demonstrated consistently higher results than IR spectroscopy, with best models displaying an $R = 0.82$ and $RMSE = 0.35$ in the basement using a Support Vector Regressor and $R = 0.80$ and $RMSE = 0.16$ in the sandstone using a Light Gradient Boosting Regressor along with geochemistry taken as pairwise log-ratios. This level of accuracy allowed us to obtain synthetic resistivity logs for an external dataset that reproduced well the general trends of the resistivity variations. Results were however limited by sources of biases induced by the lack of several important components in the geochemical data, the most important being C (i.e., graphite) and SiO_2 (silicified zones). Significant improvements could be expected in geological contexts where graphite is absent or where C and SiO_2 are measured. This method can prove particularly useful for generating synthetic data for geophysical models in areas where extensive geochemical and infrared information was historically performed before the advent of modern resistivity logging tools.

7 Computer code availability

The code used for the Machine-Learning workflow and generation of synthetic logs presented in this study is made available at: `https://github.com/MehdiSerdoun/ResistivityInversion`. It was

run and tested under Windows 10 with a 11th Gen Intel(R) Core(TM) i9-11950H @ 2.60GHz and 32GB RAM.

8 Acknowledgements

Thanks to Orano Canada Inc. (OCI) for providing all the data displayed in this paper. This research was jointly funded by Orano Mining and the French Agence Nationale de la Recherche (ANR) through grant ANR-21-CHIN-0006.

References

[1] S. J. Adams. "Core-to-LogComparison—What's A Good Match?" In: *SPE Annual Technical Conference and Exhibition?* SPE. 2005, SPE–97013.

[2] E. E. Adlakha and K. Hattori. "Paragenesis and composition of tourmaline types along the P2 fault and McArthur River uranium deposit, Athabasca Basin, Canada". In: *The Canadian Mineralogist* 54.3 (2016), pp. 661–679.

[3] J. Aitchison. "The statistical analysis of compositional data". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 44.2 (1982), pp. 139–160.

[4] T. Akiba et al. "Optuna: A next-generation hyperparameter optimization framework". In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 2623–2631.

[5] P Alexandre et al. "Alteration mineralogy and stable isotope geochemistry of Paleoproterozoic basement-hosted unconformity-type uranium deposits in the Athabasca Basin, Canada". In: *Economic Geology* 100.8 (2005), pp. 1547–1563.

[6] P. Alexandre. "Geochemistry of the Athabasca Basin, Saskatchewan, Canada, and the unconformity-related uranium deposits hosted by it". In: *The Canadian Mineralogist* 59.5 (2021), pp. 847–868.

[7] I. R. Annesley, C. Madore, and P. Portella. "Geology and thermotectonic evolution of the western margin of the Trans-Hudson Orogen: evidence from the eastern sub-Athabasca basement, Saskatchewan". In: *Canadian Journal of Earth Sciences* 42.4 (2005), pp. 573–597.

[8] J. Bergstra et al. "Algorithms for hyper-parameter optimization". In: *Advances in neural information processing systems* 24 (2011).

[9] C. L. Bérubé et al. "Predicting rock type and detecting hydrothermal alteration using machine learning and petrophysical properties of the Canadian Malartic ore and host rocks, Pontiac Subprovince, Québec, Canada". In: *Ore Geology Reviews* 96 (2018), pp. 130–145.

[10] S. A. Bosman and P. Ramaekers. "Athabasca Group+ Martin Group= Athabasca Supergroup? Athabasca Basin multiparameter drill log compilation and interpretation, with updated geological map". In: *Summary of Investigations* 2 (2015), pp. 2015–4.

[11] L. Breiman. "Random forests". In: *Machine learning* 45 (2001), pp. 5–32.

[12] L. M. Campos et al. "The hydrothermal footprint of the Crixás deposit: New vectors for orogenic gold exploration in central Brazil". In: *Ore Geology Reviews* 146 (2022), p. 104925.

[13] J. Carpenter and J. Bithell. "Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians". In: *Statistics in medicine* 19.9 (2000), pp. 1141–1164.

[14] T. Chen and C. Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

[15] H. Chu et al. "Diagenetic and geochemical studies of sandstones from drill core DV10-001 in the Athabasca basin, Canada, and implications for uranium mineralization". In: *Journal of Geochemical Exploration* 148 (2015), pp. 206–230.

[16] D. A. Clark. "Magnetic effects of hydrothermal alteration in porphyry copper and iron-oxide copper–gold systems: A review". In: *Tectonophysics* 624 (2014), pp. 46–65.

[17] R. N. Clark and T. L. Roush. "Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications". In: *Journal of Geophysical Research: Solid Earth* 89.B7 (1984), pp. 6329–6340.

[18] L. A. B. De Oliveira and C. de Carvalho Carneiro. "Synthetic geochemical well logs generation using ensemble machine learning techniques for the Brazilian pre-salt reservoirs". In: *Journal of Petroleum Science and Engineering* 196 (2021), p. 108080.

[19] M. Dentith et al. "Petrophysics and mineral exploration: a workflow for data analysis and a new interpretation framework". In: *Geophysical Prospecting* 68.1-Cost-Effective and Innovative Mineral Exploration Solutions (2020), pp. 178–199.

[20] H. Drucker et al. "Support vector regression machines". In: *Advances in neural information processing systems* 9 (1996).

[21] S. A. Earle and G. L. Drever. "Hydrogeochemical exploration for uranium within the Athabasca Basin, northern Saskatchewan". In: *Journal of Geochemical Exploration* 19.1-3 (1983), pp. 57–73.

[22] B. Efron. "Computers and the theory of statistics: thinking the unthinkable". In: *SIAM review* 21.4 (1979), pp. 460–480.

[23] J. J. Egozcue et al. "Isometric logratio transformations for compositional data analysis". In: *Mathematical geology* 35.3 (2003), pp. 279–300.

[24] R Gatzweiler, B Schmeling, and B Tan. "Exploration of the Key lake uranium deposits, Saskatchewan, Canada". In: *Uranium exploration case histories*. 1981.

[25] GDD. *Manual SCIP Archer2 18_ 09_ 2018 EN*. Instrumentation GDD Inc. Quebec City, Canada, 2018.

[26] P. Geladi and B. R. Kowalski. "Partial least-squares regression: a tutorial". In: *Analytica chimica acta* 185 (1986), pp. 1–17.

[27] GeoVista. *Dual focused resistivity sonde*. Accessed: 2023-12-20. 2019. URL: `https://www.geovista.co.uk/product-listing/dual-focused-resistivitysonde/`.

[28] E. Grunsky and P. d. Caritat. "State-of-the-art analysis of geochemical data for mineral exploration". In: *Geochemistry: Exploration, Environment, Analysis* 20.2 (2020), pp. 217–232.

[29] M. Haest et al. "Quantitative mineralogy from infrared spectroscopic data. I. Validation of mineral abundance and composition scripts at the rocklea channel iron deposit in Western Australia". In: *Economic Geology* 107.2 (2012), pp. 209–228.

[30] M. T. Ireland et al. "Suitability of legacy subsurface data for nascent geoenergy activities onshore United Kingdom". In: *Frontiers in Earth Science* 9 (2021), p. 629960.

[31] P. Jeanneret et al. "Tectono-metamorphic evolution of the pre-Athabasca basement within the Wollaston–Mudjatik Transition Zone, Saskatchewan". In: *Canadian Journal of Earth Sciences* 53.3 (2016), pp. 231–259.

[32] C. Jefferson et al. "Unconformity-associated uranium deposits of the Athabasca Basin, Saskatchewan and Alberta". In: *Bulletin-geological survey of Canada* 588 (2007), p. 23.

[33] Y. Jiao et al. "Preprocessing methods for near-infrared spectrum calibration". In: *Journal of Chemometrics* 34.11 (2020), e3306.

[34] G. Ke et al. "Lightgbm: A highly efficient gradient boosting decision tree". In: *Advances in neural information processing systems* 30 (2017).

[35] T. Kotzer and T. Kyser. "Petrogenesis of the Proterozoic Athabasca Basin, northern Saskatchewan, Canada, and its relation to diagenesis, hydrothermal uranium mineralization and paleohydrogeology". In: *Chemical Geology* 120.1-2 (1995), pp. 45–89.

[36] K. Kyser, M. Bates, and D. Gregoire. "Diagenetic fluids in Paleo- and Meso-Proterozoic sedimentary basins and their implications for long protracted fluid histories". In: *Mineralogical Association of Canada Short Course*. Vol. 28. Mineralogical Association of Canada, 2000, pp. 225–262. DOI: 10.1016/0009-2541(94)00114-N.

[37] C. Laukamp et al. "Mineral physicochemistry underlying feature-based extraction of mineral abundance and composition from shortwave, mid and thermal infrared reflectance spectra". In: *Minerals* 11.4 (2021), p. 347.

[38] E. Laverret et al. "Mineralogy and geochemistry of the host-rock alterations associated with the Shea Creek unconformity-type uranium deposits (Athabasca Basin, Saskatchewan, Canada). Part 1. Spatial variation of illite properties". In: *Clays and Clay Minerals* 54.3 (2006), pp. 275–294.

[39] P. Ledru et al. "The Unconformity-related Uranium Mineral System of the Athabasca Basin (Canada)". In: *Metallic Resources 2: Geodynamic Framework and Remarkable Examples in the World*. 2023, pp. 143–194.

[40] H. Leväniemi and J. Hokka. "Petrophysical target characterization with lithogeochemical clustering: the Metsämonttu Zn–Pb–Cu deposit, southern Finland". In: *Near Surface Geophysics* 20.6-Near-Surface Geophysics for Mineral Exploration and Mining (2022), pp. 637–660.

[41] X. Lu et al. "3D electromagnetic modeling of graphitic faults in the Athabasca Basin using a finite-volume time-domain approach with unstructured grids". In: *Geophysics* 86.6 (2021), B349–B367.

[42] T. Martin, J. Tadla, and Z. Jobe. "Digitalization of Legacy datasets and Machine Learning Regression Yields Insights for Reservoir Property Prediction and Submarine-Fan Evolution: A Subsurface Example From the Lewis Shale, Wyoming". In: *The Sedimentary Record* 20.1 (2022).

[43] A. McDonald. "Data quality considerations for petrophysical machine-learning models". In: *Petrophysics* 62.06 (2021), pp. 585–613.

[44] J. McGaughey. "The common earth model: A revolution in mineral exploration data integration". In: *GIS for the Earth Sciences, SP* 44 (2006), pp. 567–576.

[45] J. M. McKinley et al. "The single component geochemical map: Fact or fiction?" In: *Journal of Geochemical Exploration* 162 (2016), pp. 16–28.

[46] A. Menier et al. "Relationship between rock physical properties and spectral mineralogy applied to exploration for an unconformity-related uranium deposit (Saskatchewan, Canada)". In: *Canadian Journal of Earth Sciences* 57.11 (2020), pp. 1349–1364.

[47] C. Mwenifumbo et al. "Physical rock properties from the Athabasca Group: designing geophysical exploration models for unconformity uranium deposits". In: *Journal of Applied Geophysics* 55.1-2 (2004), pp. 117–135.

[48] R. Ng, P. Alexandre, and K. Kyser. "Mineralogical and geochemical evolution of the unconformity-related McArthur River Zone 4 orebody in the Athabasca Basin, Canada: implications of a silicified zone". In: *Economic Geology* 108.7 (2013), pp. 1657–1689.

[49] M. Pagel, B. Poty, and S. Sheppard. "Contribution to some Saskatchewan uranium deposits mainly from fluid inclusion and isotopic data". In: *Uranium in the Pine Creek geosyncline*. Vienna: International Atomic Energy Agency, 1980, pp. 639–654.

[50] G. Palacky. "Clay mapping using electromagnetic methods". In: *First Break* 5.8 (1987), pp. 295–306. DOI: 10.3997/1365-2397.1987015.

[51] J. Palarea-Albaladejo and J. Martín-Fernández. "Values below detection limit in compositional chemical data". In: *Analytica Chimica Acta* 764 (2013), pp. 32–43.

[52] J. Palarea-Albaladejo and J. Martín-Fernández. "zCompositions—R package for multivariate imputation of left-censored data under a compositional approach". In: *Chemometrics and Intelligent Laboratory Systems* 143 (2015), pp. 85–96.

[53] J. Palarea-Albaladejo, J. Martín-Fernández, and A. Buccianti. "Compositional methods for estimating elemental concentrations below the limit of detection in practice using R". In: *Journal of Geochemical Exploration* 141 (2014), pp. 71–77.

[54] J. Palarea-Albaladejo, J. Martín-Fernández, and J. Gómez-García. "A parametric approach for dealing with compositional rounded zeros". In: *Mathematical Geology* 39 (2007), pp. 625–645.

[55] F. Pedregosa et al. "Scikit-learn: machine learning in Python". In: *The Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[56] J. Percival, K. Bell, and J. Torrance. "Clay mineralogy and isotope geochemistry of the alteration halo at the Cigar Lake uranium deposit". In: *Canadian Journal of Earth Sciences* 30.4 (1993), pp. 689–704. DOI: 10.1139/e93-056.

[57] J. Percival and H. Kodama. "Sudoite from Cigar Lake, Saskatchewan". In: *The Canadian Mineralogist* 27.4 (1989), pp. 633–641.

[58] V. Plagnes et al. "Hydrogeological modelling applied to mineral exploration". In: *EGU General Assembly Conference Abstracts*. 2020, p. 22496.

[59] B. Powell, G. Wood, and L. Bzdel. "Advances in geophysical exploration for uranium deposits in the Athabasca Basin". In: *Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration*. Toronto, Canada, 2007, pp. 771–790.

[60] P. Ramaekers et al. "Revised geological map and stratigraphy of the Athabasca Group, Saskatchewan and Alberta". In: *Bulletin-Geological Survey of Canada* 588 (2007), p. 155.

[61] L. Robb. *Introduction to ore-forming processes*. Hoboken: John Wiley & Sons, 2005.

[62] R. Roy et al. "Three-dimensional gravity modelling applied to the exploration of uranium unconformity-related basement-hosted deposits: the Contact prospect case study, Kiggavik, northeast Thelon region (Nunavut, Canada)". In: *Canadian Journal of Earth Sciences* 54.8 (2017), pp. 869–882. DOI: 10.1139/cjes-2016-0225.

[63] E. Schetselaar et al. "Geologically driven 3D modelling of physical rock properties in support of interpreting the seismic response of the Lalor volcanogenic massive sulphide deposit, Snow Lake, Manitoba, Canada". In: *Geological Society, London, Special Publications* 453.1 (2018), pp. 57–79. DOI: 10.1144/SP453.5.

[64] F. Scott. "Midwest Lake uranium discovery". In: *Uranium Exploration in Athabasca Basin*. Saskatchewan, Canada: Geological Survey of Canada, 1983, pp. 82–11.

[65] P. Sikandar et al. "The use of vertical electrical sounding resistivity method for the location of low salinity groundwater for irrigation in Chaj and Rachna Doabs". In: *Environmental Earth Sciences* 60 (2010), pp. 1113–1129. DOI: 10.1007/s12665-009-0255-6.

[66] R. Stine. "Bootstrap prediction intervals for regression". In: *Journal of the American Statistical Association* 80.392 (1985), pp. 1026–1031. DOI: 10.1080/01621459.1985.10478220.

[67] H. Talebi, H. Grom, and M. Caccetta. "Surficial and deep earth material prediction from geochemical compositions". In: *Natural Resources Research* 28 (2019), pp. 869–891. DOI: 10.1007/s11053-018-9423-2.

[68] D. Thomas, R. Matthews, and V. Sopuck. "Athabasca Basin (Canada) unconformity-type uranium deposits: Exploration model, current mine developments and exploration directions". In: *Geology and Ore Deposits 2000: the Great Basin and Beyond*. Ed. by J. Cluer et al. Reno, Nevada: Geological Society of Nevada Symposium, 2000, pp. 103–126.

[69] R. Tibshirani and B. Efron. *An introduction to the bootstrap*. Monographs on statistics and applied probability. New York: Chapman and Hall, 1993.

[70] R. Tolosana-Delgado and K. Van den Boogaart. "Joint consistent mapping of high-dimensional geochemical surveys". In: *Mathematical Geosciences* 45 (2013), pp. 983–1004. DOI: 10.1007/s11004-013-9485-y.

[71] R. Tolosana-Delgado et al. "On machine learning algorithms and compositional data". In: *Proceedings of the 8th International Workshop on Compositional Data Analysis (CoDaWork2019)*. Terrassa, 2019, pp. 172–175.

[72] J. Vidal et al. "Clay minerals related to the circulation of geothermal fluids in boreholes at Rittershoffen (Alsace, France)". In: *Journal of Volcanology and Geothermal Research* 349 (2018), pp. 192–204. DOI: 10.1016/j.jvolgeores.2017.10.019.

[73] M. Williams et al. "Pyrolite: python for geochemistry". In: *Journal of Open Source Software* 5.50 (2020), p. 2314. DOI: 10.21105/joss.02314.

[74] G. Yeo et al. "Environmental significance of oncoids and crypto-microbial laminates from the Late Paleoproterozoic Athabasca Group, Saskatchewan and Alberta". In: *Bulletin-Geological Survey of Canada* 588 (2007), p. 315.

[75] G. Zhang, K. Wasyliuk, and Y. Pan. "The characterization and quantitative analysis of clay minerals in the Athabasca Basin, Saskatchewan: application of shortwave infrared reflectance spectroscopy". In: *The Canadian Mineralogist* 39.5 (2001), pp. 1347–1363. DOI: 10.2113/gscanmin.39.5.1347.