GRC Transactions, Vol. 44, 2020

Quantifying the Relationship of Porosity and Alteration based on Lithology Classification in Icelandic Rocks

Cari Covell[1], Samuel Scott[1], Ágúst Valfells[1], Juliet Newson[1], Egill Júlíusson[2], Halldór Pálsson[3], Birgir Hrafnkelsson[3], and María Guðjónsdóttir[1]

[1] Reykjavik University

[2] Landsvirkjun

[3] University of Iceland

Keywords:

Classification, measured helium porosity, thin section porosity, total alteration, k nearest neighbors (kNN).

ABSTRACT

Porosity and alteration play an important role in the mechanical behavior of volcanic rocks. The measurement of these properties is done several ways. We analyze four cases of quantifying the relationship of porosity and alteration data for Icelandic rocks using thin section porosity, measured porosity, microporosity, and total alteration. We use the k nearest neighbors supervised learning approach to test a portion of the data and compare the lithology class prediction to the observed class, and evaluate the sensitivity of k for each case. We determine the best accuracy measurements for each lithology. Thin section porosity is best to use for intrusions while measured porosity is best for hyaloclastites, when comparing to alteration. Lava flows are ambiguous and need more evaluation of the observed data. This study is part of a larger project to use these methods and results for probabilistic reservoir modeling, where a prior distribution is necessary.

1. Introduction

The mechanical behavior of volcanic rocks in Iceland is largely governed by porosity and alteration. Hyaloclastites host major fluid reservoirs and are highly susceptible to hydrothermal alteration (Sigurdsson et al., 2000; Franzson et al., 2001; and Frolova et al., 2005). In addition, there are sub-aerially erupted lava flows, which tend to have lower porosity than hyaloclastites, as well as intrusive rocks, which have the lowest primary porosity. The relationship has been analyzed in several works (Pola et al., 2012; Anovitz & Cole, 2015; Heap et al., 2020). However, limited use of a machine learning approach has been made to quantify the relationship of these properties (Kiran & Salehi, 2020).

There are several kinds of porosity that effect the alteration of minerals. Measured porosity is commonly associated with He-gas expansion, while thin section porosity is done by point counting or x-ray diffraction (XRD) (Franzson et al., 2011). A misconception is that the two porosities are equal, when in reality measured porosity is slightly higher than thin section porosity (Sigurdsson et al., 2000). Microporosity is termed as the measured porosity minus the thin section porosity (Franzson et al., 2001). Additionally, alteration is defined as the abundance of altered primary minerals (mostly plagioclase, pyroxene, and olivine) plus secondary minerals (mostly smectite, zeolite, and calcite) filling primary pore space as a proportion of solid rock (Franzson et al., 2011, Levy et al., 2018).

In this study, the aim is to classify the different types of porosity and alteration data to quantify the accuracy of their relationship. We use the k-nearest neighbors algorithm; a supervised machine learning approach, with lithology as the class criteria (hyaloclastite, lava flow, and intrusion). The purpose is to compute an accuracy measurement for each case of porosity and alteration prediction to input into future probabilistic reservoir modeling.

2. Porosity and Alteration Data

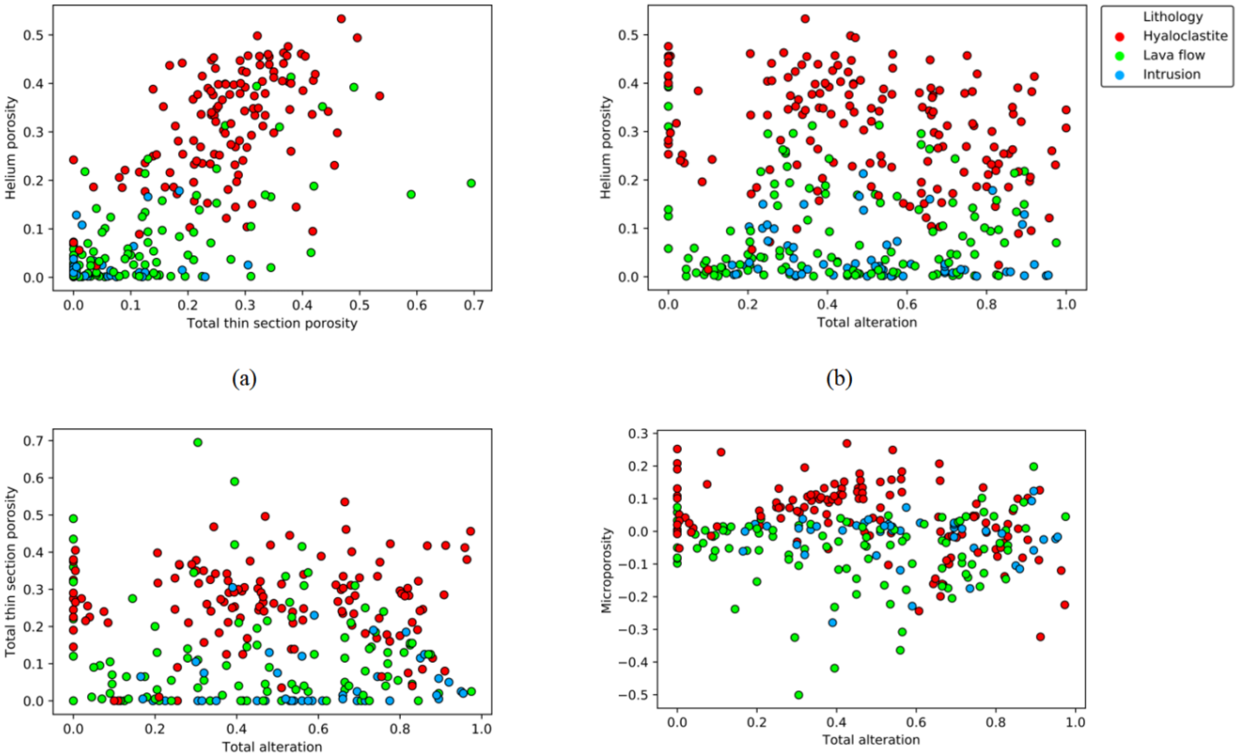
2.1 Data collection

The Valgardur database was established through the collaboration of Iceland GeoSurvey (ISOR), The National Energy Authority (Orkustofnun), Reykjavik Energy (Orkuveita Reykjavkur), and Landsvirkjun. The compilation of a database of Icelandic rock properties began in the early 1990s and has been documented in numerous publications (Sigurdsson & Stefansson, 1994; Gudmundsson et al., 1995; Stefansson et al., 1997; Sigurdsson et al., 2000; Franzson et al. 2001; Frolova et al., 2005; Franzson et al. 2008; Franzson et al., 2010; and Franzson et al., 2011). In addition to the data contained in Valgardur, we extracted a subset of data from the PetroPhysical database P^3 (Bär et al., 2017; Bär & Reinsch, 2020).

The Levy samples are from the Krafla high-temperature geothermal field, where cores are available in four boreholes: KH-1, KH-3, KH-5, and KH-6 (Levy et al., 2018). This field, where a geothermal powerplant is operated, is located within the caldera of the Krafla central volcano in northeast Iceland (Sigmundsson, 2006). For the purposes of this study, we extract the measured He porosity and the total alteration that was tabulated using XRD. The texture of the samples varies between fully glassy basalt (tuff hyaloclastite) to highly crystalline basalt (intrusions) (Levy et al., 2018).

2.2 Data relationships

In this work, thin section porosity, measured porosity, microporosity, and total alteration for each rock sample characterized by lithology were collectively extracted from the databases. The data were classified into three general rock types: hyaloclastites, lava flows, and intrusions. A series of 2-dimentional scatter plots comparing the rock properties are shown in Figure 1.



(a) (b)

(c) (d)

Figure 1: Porosity and alteration data from the Valgardur, P^3, and Levy databases for Icelandic rocks.

Observations of thin section porosity and helium measured porosity (Figure 1a) indicate that hyaloclastites show clarity of having the highest porosities (> 0.2) and intrusions have good representation at the lowest porosities (< 0.05). The porosity relationship of lava flows shows ambiguity, as there are differences in the two measurements where thin section porosity is higher than measured porosity.

We take a closer look at the thin section and helium measured porosities separately by comparing each to total alteration (weight %) in the sample (Figures 1b and 1c, respectively). In each case, there are various degrees of alteration regardless of porosity and lithology. The expectation is to have consistency with lithology based on rock mechanics. Hyaloclastites are more porous, glassy, and easier to alter, while intrusions have very low porosity, higher crystallinity, and are more difficult to alter (Levy et al., 2018). Most notably, we observe this contradiction in intrusions where more samples have a higher measured porosity than thin section porosity (values > 0.05) at alteration values < 0.4.

When observing the alteration against microporosity (Figure 1d), many lava flows have higher thin section porosity than measured porosity and are therefore tabulated at microporosity values < 0. Despite this, vagueness of the microporosity relationship to alteration for lava flows is consistent with observations from Figure 1a. Hyaloclastites with alteration between 0.2 and 0.6 have clear distinction of microporosity from about 0.1 to 0.2. Contrarily, many samples with alteration > 0.6 are mixed with lava flows and intrusions, with several samples tabulated at microporosity < 0. Henceforth, we use machine learning classification techniques to more thoroughly understand underlying trends of porosity and alteration as well as the quantification of associated errors.

3. kNN Classifier

The kNN classifier is a non-parametric supervised learning method for classification (Cover and Hart, 1967), and is implemented with prior domain knowledge where the rule classifies each unlabeled sample set by a majority among its k-nearest neighbors (Bishop, 2006). kNN identifies the k nearest neighbors of the test samples where the label sets of its neighboring samples are obtained. Each classified sample neighbor is considered as a piece of evidence corresponding to the class of that pattern. In this study, a simple approach has been adopted to train a portion of the data set and then predict the accuracy of the remaining data.

The robustness of the classifier depends on the metric distance used to identify the nearest neighbors (Cover and Hart, 1967). We use a uniform weight for the data where all points in each neighborhood are weighted equally. This is to avoid a bias for the data that would otherwise be classified relative to a constrained range of values. In addition, the Euclidean distance is applied to identify the nearest neighbor sample at each iteration of the classifier. The optimal number of k nearest neighbors was decided based on a GridSearchCV algorithm (Pedregosa et al., 2011), where up to k = 25 neighbors were trained over a 10-fold cross validation.

The samples were randomly divided into a train set and a test set. With approximately 300 samples for each case of porosity and alteration data, the training set and testing set are regulated at a 66:33 ratio. Table 1 shows the size of the train set and test set as well as the optimal number of k nearest neighbors. Ideally, k nearest neighbors should be an odd number in order to prevent ties during classification. However, we use an odd number of classes for lithology (3), therefore ties are less common, and an even k is acceptable (Bishop, 2006).

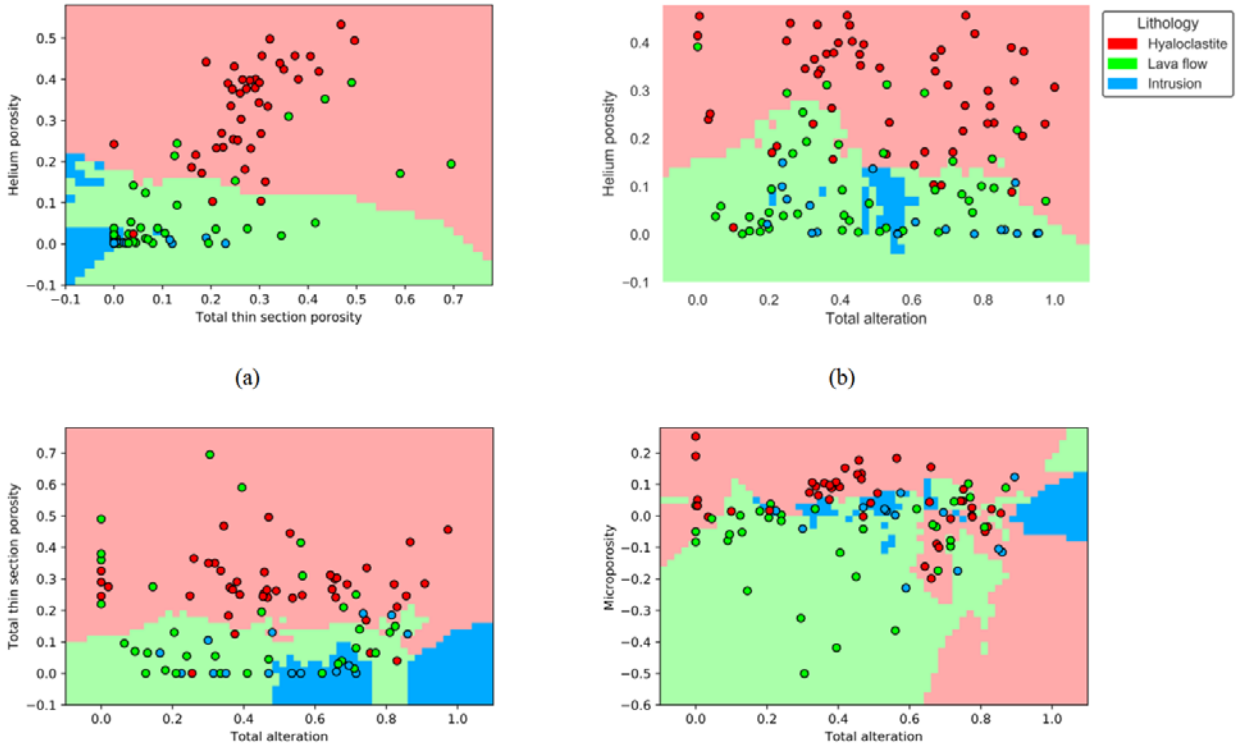
Table 1: Sample size for each comparison of porosity and alteration. The optimal k nearest neighbors is also shown based on results from GridSearchCV.

Parameters	Train sample size	Test sample size	Total sample size	k nearest neighbors
Thin section porosity & Measured porosity	181	90	271	10
Total alteration & Measured porosity	231	115	346	14
Total alteration & Thin section porosity	184	92	276	8
Total alteration & Microporosity	180	89	269	4

4. Results

4.1 Evaluation of lithology classification trends

We now evaluate the results from kNN of the four cases of porosity and alteration on a sample-by-sample basis. Figure 2 shows the position of class boundaries predicted from the kNN classifier, populated by the observed test data.



(a)

(b)

(c)

(d)

Figure 2: Classification of the test data. Samples that stayed in original lithology are plotted with samples that classified to a different lithology.

The predictions for measured and thin section porosity (Figure 2a) show clear distinction of the hyaloclastites, where predicted and observed data have higher porosities > 0.2. Several data for lava flows and intrusions are close to 0 porosity, which result in some misclassification between the two lithologies as there is a prediction boundary placed roughly at 0.025 porosity. In Figure 2b, lava flows with measured porosity > 0.1 have alteration measures up to 0.4; and as porosity decreases, alteration increases. Yet, there are several lava flow observations that are predicted as hyaloclastites, like what is seen in Figure 2c. Interestingly, intrusions have different prediction boundaries when comparing measured (Figure 2b) and thin section (Figure 2c) porosities < 0.1. Measured porosity ranges from about 0.45 to 0.6 alteration and thin section porosity ranges from about 0.5 to 0.75 alteration. Figure 2d shows that the prediction boundaries are close for all three lithology classes for most alteration > 0.35 and microporosity close to 0. This indicates sensitivity of the kNN classifier and proves difficult to correctly predict lithology from observed data.

When using the kNN, data is tabulated based on correct and incorrect classification between classes. The confusion matrix is a performance measure that allows for specific identification of confusion between classes; e.g. when one class is mislabeled as another class (Kiran & Salehi, 2020). The number of correct and incorrect predictions are summarized with count values and broken down by each class. This provides insight into the types of errors being made by a classifier. Each row of the matrix represents the instances in a predicted class, while each column represents the instances in an observed class. Table 2 shows the confusion matrix for each comparison of porosity and alteration data.

Table 2: Confusion matrix for each comparison of porosity and alteration data.

Parameters	**Confusion matrix**			
		Hyaloclastite	*Lava flow*	*Intrusion*
Thin section porosity & Measured porosity	*Hyaloclastite*	38	3	0
	Lava flow	8	19	7
	Intrusion	0	6	9
Total alteration & Measured porosity	*Hyaloclastite*	45	8	0
	Lava flow	10	31	2
	Intrusion	1	16	2
Total alteration & Thin section porosity	*Hyaloclastite*	39	4	0
	Lava flow	12	18	3
	Intrusion	3	8	5
Total alteration & Microporosity	*Hyaloclastite*	34	5	3
	Lava flow	13	18	2
	Intrusion	5	4	5

For each case of parameters, hyaloclastites are well predicted and don’t show much variation in false observations. This is most likely due to having the largest number of samples, high porosity, and the ability to distinguish alteration minerals. In contrast, intrusions overall show more samples predicted outside of the true observation than those samples correctly classified. As seen from Figure 2, the prediction boundaries are the smallest due to low sample size and therefore are sensitive to the other classes. Lava flows have about a 50/50 correct prediction to deviation in other classes.

4.2 Classification report

A classification report is used to measure the quality of predictions using a precision, recall, and F_1 score. Precision is the division of the number of correctly classified samples by the total number of actual samples in the respective class (columns of confusion matrix). Recall is defined as the ratio of the number of correctly classified samples divided by the total number of predicted samples, relative to class (rows of confusion matrix). The F_1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0. Table 3 shows the classification report for each comparison of porosity and alteration data.

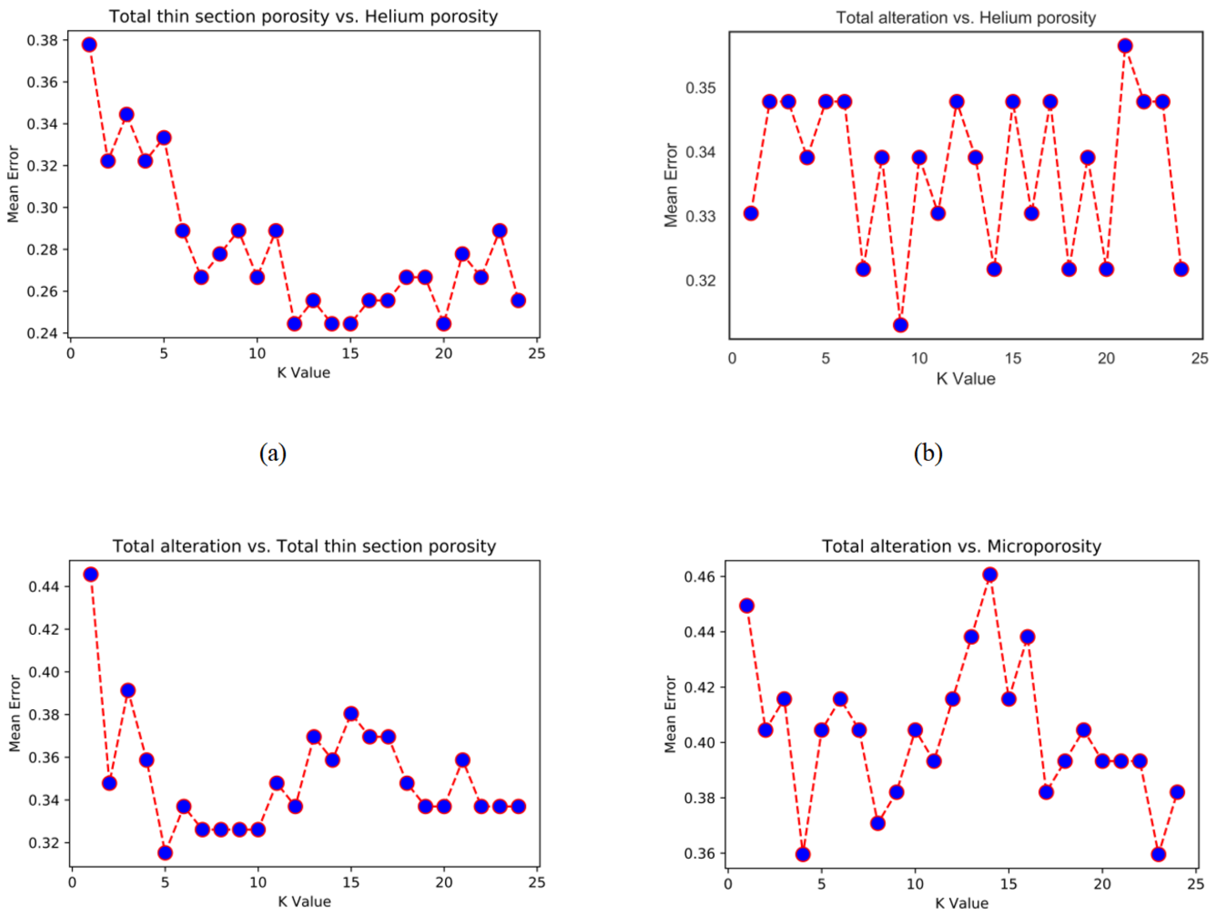
Table 3: Classification report for each comparison of porosity and alteration data.

Parameters	**Classification report**			
		Precision	*Recall*	*F_1-score*
Thin section porosity & Measured porosity	*Hyaloclastite*	0.83	0.93	0.87
	Lava flow	0.68	0.56	0.61
	Intrusion	0.56	0.6	0.58
	Average	0.73	0.73	0.73
Total alteration & Measured porosity	*Hyaloclastite*	0.80	0.85	0.83
	Lava flow	0.56	0.72	0.63
	Intrusion	0.50	0.11	0.17
	Average	0.66	0.68	0.65
Total alteration & Thin section porosity	*Hyaloclastite*	0.72	0.91	0.80
	Lava flow	0.60	0.55	0.57
	Intrusion	0.62	0.31	0.42
	Average	0.66	0.67	0.65
Total alteration & Microporosity	*Hyaloclastite*	0.65	0.81	0.72
	Lava flow	0.67	0.55	0.60
	Intrusion	0.50	0.36	0.42
	Average	0.63	0.64	0.63

As a rule of thumb, the F_1-score is used to compare classifier models. Hyaloclastites have a respectable average class prediction of above 0.8, but intrusions vary greatly depending on the case of parameters being analyzed. For both hyaloclastites and intrusions, the thin section porosity and measured porosity case of kNN classification performs best. This is a valid result as the two porosity measurements are the same in theory. When understanding the relationship between porosity and alteration, hyaloclastites are best using the measured porosity (0.83). Intrusions could be classified using either thin section porosity or microporosity as there is a tie in the F_1-score (0.42), however differences in precision and recall would have the thin section porosity win. For lava flows, class prediction is best using measured porosity and total alteration (0.63) likely due to having the most amount of data to predict lithology for this case. Due to much ambiguity in the observed data, the general results for lava flows are middling. When considering alteration, overall kNN classification quality is irrelevant of the type of porosity measurement (approx. 0.65). However, the quality of classification with alteration is about 0.08 less than using strictly porosity.

5. Sensitivity of kNN

The sensitivity of the kNN can be analyzed by understanding the mean error for each selected value of k, as shown in Figure 3.



(a) (b)

(c) (d)

Figure 3: Mean error of selection for k nearest neighbors.

Ideally, the least sensitive kNN model is one where the mean error is consistently the lowest. The total alteration and thin section porosity case (Figure 1c) shows the least sensitivity where mean error is relatively consistent for 6 to 10 k-nearest neighbors. The most sensitive is total alteration and microporosity (Figure 1d) as k is difficult to optimally determine, which is why the selected k from GrodSearchCV was the lowest at k = 4. The total alteration and helium porosity case is also quite sensitive as k is only consistent at rather high values.

6. Conclusions

We analyzed four cases of Icelandic rock data to quantify the relationship between parameters: thin section porosity & measured porosity, measured porosity & total alteration, thin section porosity and total alteration, and microporosity & total alteration. We used the kNN machine learning approach to identify trends in the prediction relative to the observation, and discussed sensitivities of k selection and effects on classification quality.

In all cases, many observed hyaloclastites are predicted correctly. Lava flows are ambiguous because many observed data classify either on the border with another lithology or have high porosities that trend with hyaloclastites. Intrusions vary by case of porosity and alteration parameters. Hyaloclastites are easier to classify because they are more susceptible to alteration, and we have the most data on them given that we are studying Icelandic systems where hyaloclastite alteration is common.

Most of the time the GridSearchCV will pick the k with the lowest mean error, but a low k does not fully govern the quality of class prediction. Additionally, the best case from the sensitivity analysis is only in line with intrusions. The Icelandic data shows that intrusions have the lowest porosities but varying alterations, therefore thin section analysis is best to differentiate the samples as best as possible.

Now that we understand the data more clearly, porosity and alteration constraints can be defined knowing the quantification of correctly classified lithology. This study is part of a larger project to use these methods and results for probabilistic reservoir modeling, where a prior distribution is necessary. We would incorporate the quality of the data respective to lithology in order to more accurately measure uncertainty in rock properties that govern fluid mechanics.

Acknowledgements

The authors work together under the Operations Research in Subsurface Modeling (ORSM) group based at Reykjavik University. The work presented here is part of a larger project entitled Data Fusion for Geothermal Reservoir Characterization, funded by RANNIS under the Technology Development Fund. Use of the open source Valgardur Database is with guidance from Orkustofnun, and clarifications were specifically made by Hjalti Franzson of Iceland GeoSurvey (ISOR). Access to the IMAGE or PetroPhysical database P^3 was with permission from Kristian Bär of Technische Universitat Darmstadt.

REFERENCES

Anovitz, L. M., & Cole, D. R. (2015). Characterization and analysis of porosity and pore structures. *Reviews in Mineralogy and geochemistry*, 80(1), 61-164.

Bär, K., Reinsch, T., Sippel, J., Strom, A., Mielke, P., & Sass, I. (2017). P^3-International PetroPhysical Property Database. *Proceedings, 42nd Workshop on Geothermal Reservoir Engineering*. Stanford University. Stanford, CA (pp. 1-6).

Bär, K., & Reinsch, T. (2020). P^3–PetroPhysical Property Database–a global compilation of lab measured rock properties. *Earth System Science Data Discussions*.

Bishop, C. M. (2006). Pattern recognition and machine learning. New York, New York: Springer Science Business Media, LLC.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.

Franzson, H., Gudfinnsson, G., Helgadottir, H., & Frolova, J. (2010). Porosity, density, and chemical composition relationships in altered Icelandic hyaloclastites. *Birkle & Torres-Alvarado (ed), Water-Rock Interaction* (pp. 199-202). London, England, United Kingdom: Taylor & Francis Group.

Franzson, H., Gudfinnsson, G. H., Frolova, J., Helgadottir, H. M., Pauly, B., Mortensen, A. K., & Jakobsson, S. P. (2011). Icelandic Hyaloclastite Tuffs, Petrophysical Properties, Alteration and Geochemical Mobility. Reykjavik: Iceland GeoSurvey (ISOR-2011/064).

Franzson, H., Gudlaugsson, S., & Fridleifsson, G. O. (2001). Petrophysical Properties of Icelandic Rocks. *6th Nordic Symposium on Petrophysics*. Trondheim, Norway.

Franzson, H., Zierenberg, R., & Schiffman, P. (2008). Chemical transport in geothermal systems in Iceland: Evidence from hydrothermal alteration. *Journal of Volcanology and Geothermal Research*, 217-229.

Frolova, J., Ladygin, V., Franzson, H., Sigurdsson, O., Stefansson, V., & Shustrov, V. (2005). Petrophysical Properties of Fresh to Mildly Altered Hyaloclastite Tuffs. *World Geothermal Congress*. Antalya, Turkey.

Gudmundsson, A., Franzson, H., & Fridleifsson, G. (1995). Fordafraedistudlar: sofnun syna. OS-9517-JHD-11 B, 73 pp. In Icelandic.

Heap, M. J., Gravley, D. M., Kennedy, B. M., Gilg, H. A., Bertolett, E., & Barker, S. L. (2020). Quantifying the role of hydrothermal alteration in creating geothermal and epithermal mineral resources: The Ohakuri ignimbrite (Taupō Volcanic Zone, New Zealand). *Journal of Volcanology and Geothermal Research*, 390, 106703.

Kiran, R., & Salehi, S. (2020). Assessing the Relation between Petrophysical and Operational Parameters in Geothermal Wells: A Machine Learning Approach. *Proceedings 45th Workshop on Geothermal Reservoir Engineering*. Stanford University. Stanford, CA, USA.

Lévy, L., Gibert, B., Sigmundsson, F., Flóvenz, Ó. G., Hersir, G. P., Briole, P., & Pezard, P. A. (2018). The role of smectites in the electrical conductivity of active hydrothermal systems: electrical properties of core samples from Krafla volcano, Iceland. *Geophysical Journal International*, 215(3), 1558-1582.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 2825-2830.

Pola, A., Crosta, G., Fusi, N., Barberini, V., & Norini, G. (2012). Influence of alteration on physical properties of volcanic rocks. *Tectonophysics*, 566, 67-86.

Sigurdsson, Ó., Gudmundsson, A., Fridleifsson, G. O., Franzson, H., Gudlaugsson, S., & Stefansson, V. (2000). Database on Igneous Rock Properties in Icelandic Geothermal Systems: Status and Results. *World Geothermal Congress*. Kyushu-Tohoku, Japan.

Sigurdsson, O., & Stefansson, V. (1994). Fordafraedistudlar. Maelingar a synum. (Reservoir parameters. Measurements of rock samples). Report Orkustofnun OS-94049/JHD-28 B (in Icelandic), 35 pp.

Stefansson, V., Sigurdsson, O., Gudmundsson, A., Franzson, H., Fridleifsson, G.O., & Tulinius, H. (1997). Core measurements and geothermal modelling. *Second Nordic Symposium on Petrophysics*. Fractured reservoir. Nordic Petroleum Series, vol.1, pp. 198-220.