GRC Transactions, Vol. 47, 2023

Machine Learning Applications for Predicting Heat Flow and Thermal Conductivity in Northwest U.S.

Dale E. Bird[1], Naila Dowla[2] and William E. Cathey[3]

[1]Bird Geophysical, Houston, Texas

[2]Dowla Analytics LLC, Houston Texas

[3]Earthfield Technology LLC, Richmond, Texas

Keywords

Heat Flow, Thermal Conductivity, Magnetics, Machine Learning, Northwest U.S.

ABSTRACT

Machine learning algorithms may be trained to predict data in areas where coverage is sparse. We make use of a Stacking Regressor, which is a supervised machine learning algorithm consisting of individual Decision Tree based ensemble estimators (Tuned Decision Tree, Tuned Random Forest, and Tuned Gradient Boost), to predict heat flow and thermal conductivity in parts of Oregon, Idaho, California, Nevada, and Utah. Decision Tree based ensemble methods are chosen for their ability to work on data without scaling, dimensionality reduction, or normalizing. After required corrections and processing, measured data are direct inputs, without modification, making this workflow adaptable to almost any combination of datasets.

For this study, well and station data include heat flow, thermal conductivity, thermal gradient, bottom hole temperature, and heat production. Gridded data include topography, Curie point depth, and magnetic susceptibilities computed from a 3D inversion of magnetic anomalies. Two suites of machine learning solutions are calculated: 1) heat flow is targeted, with remaining variables used to train algorithms, and 2) thermal conductivity is targeted, with remaining variables used to train algorithms. Predicted heat flow and thermal conductivity values are sampled onto a 5 km mesh, spanning over 1,000,000 km^2. Note that these results can be predicted at smaller mesh increments if desired.

Our results improve heat flow and thermal conductivity coverage by predicting anomalies where well and station sampling are sparse. Regionally, high-resolution results are comparable to lower resolution interpolated well and station data. Examples include broad geologic terranes such as the Coast Ranges, volcanic regions, and areas dominated by high heat flow throughout northwest U.S. More locally, the Snake River Plain (SRP) volcanic region in southern Idaho lies over a thermal anomaly that may extend down to the upper mantle. It is one of the highest heat flow areas in the U.S. even though thermal gradients are suppressed by the Snake River aquifer. Four wells were

drilled at three sites in SRP, *HOTSPOT: The Snake River Plain Scientific Drilling Project* (Sep. 2010 to Jan. 2012), to understand the compositional and eruptive history of SRP volcanism. *HOTSPOT* results confirmed a deep hydrothermal system in SRP, which was then used to characterize geothermal play fairways. We find that high geothermal resource probability fairways generally correlate with our predicted results; that is, with higher heat flow values, but with lower thermal conductivity values. We note, however, that predicted heat flow and thermal conductivity anomalies do not precisely coincide with play fairway geometries; therefore, we integrate our results with published fairways to high-grade prospective geothermal plays.

1. Introduction

High heat flow values in northwestern United States, related to extensional tectonic forces that produced the Great Basin and surrounding region, underpin geothermal exploration (Figure 1). Geothermal play fairways are characterized by three critical parameters: 1) heat source, 2) reservoir permeability, and 3) seal quality (Shervais et al., 2020). In this study, we address the first two of these parameters. Heat flow may be directly linked to heat source, however thermal conductivity may be indirectly linked to permeability; that is, although the relationship is complex, fractured rocks in reservoirs may provide conduits for fluid flow that may in turn decrease thermal conductivity (Surma and Geraud, 2003; Garcia and Santamarina, 2021).

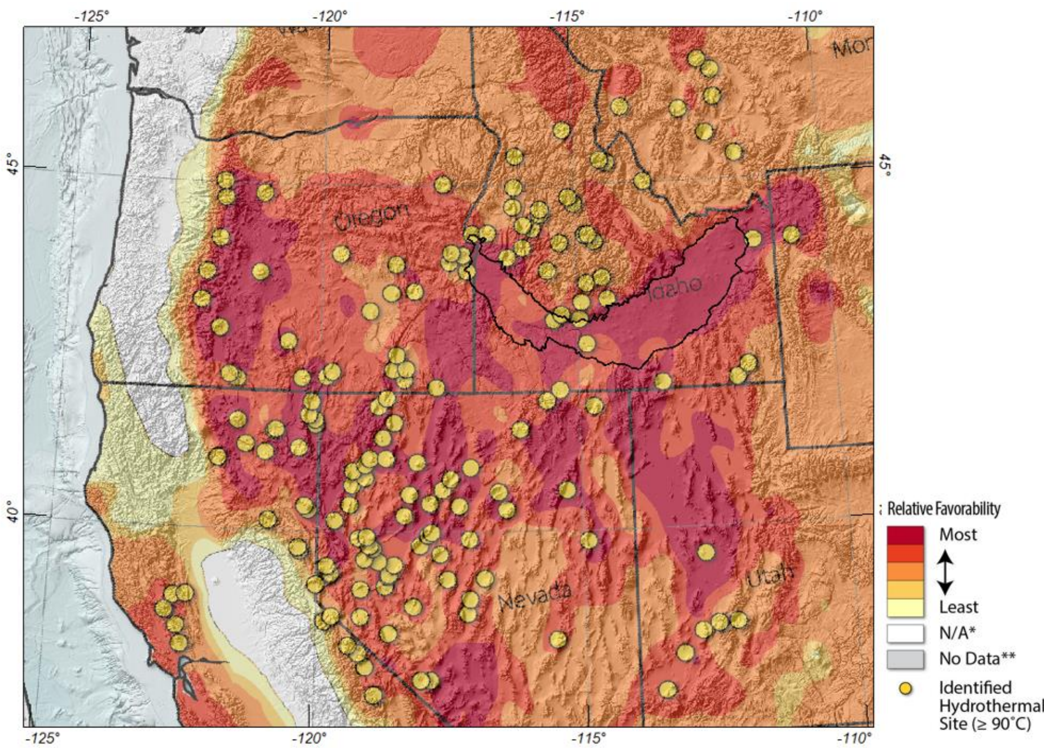

Figure 1: Geothermal resources (after Roberts, 2018) draped over shaded, gray-scale imaged topography (GEBCO Compilation Group, 2020); Snake River Plain region outlined in southern Idaho.

Temperature is a critical variable for both heat flow and thermal conductivity, and thermal gradients of course require temperature measurements. But other data, such as magnetic, may reflect broad temperature variations. The Curie point is the temperature at which rocks gain or lose magnetization as they are cooled or heated, respectively. Curie point temperatures vary due to rock composition, but the range is generally between 550° to 600°C. The depth to Curie point may be estimated by spectral analyses of magnetic anomalies where the longest wavelengths are thought to be produced by Curie point depth variations (Bouligand et al., 2009; Bansal et al., 2011; Li et al., 2017).

We invert the earth layer above Curie point depth for magnetic susceptibility and combine these susceptibilities with measured heat flow, thermal conductivities, and heat production, as well as temperature horizons (borehole and Curie), to predict heat flow and thermal conductivity on regularly spaced grid locations using open-source computer programs. Easy and friendly access to machine learning applications, via *Scikit-Learn* Python-based algorithms in *Jupyter Notebooks* computing platform (Kluyver et al., 2016; Pedregosa et al., 2011), allow use of this powerful technology for the greater science community.

2. Data

All data used in this study are derived from open-file sources. Heat flow, thermal conductivity and thermal gradient data, measured in wells, were extracted from Geothermal Resources Council and Geothermal Service of Canada compilations (Jessop et al., 1976; Blackwell and Richards, 2004; Blackwell et al., 2006). Heat production station data were extracted from Hasterok and Webb (2017), and the Geothermal Service of Canada, compilations. Uneven, or sparse, distributions of heat flow and thermal conductivity measurements produce spatial biases that may complicate interpretations of these data (Figure 2). Gridded data include topography, total magnetic intensity anomalies, and Curie point depth (Finn et al., 2001; Li et al., 2017; GEBCO Bathymetric Compilation Group 2020, 2020).

3. Methods

3.1 3D Magnetic Model

A single-layer 3D model is constructed from topography and Curie point depth horizons. Modeling magnetic susceptibilities is a linear inverse problem (Blakely, 1995) and we use Seequent's *Oasis Montaj* software package to calculate susceptibilities following Parker (1973) formulas.

3.2 Supervised Machine Learning

Supervised machine learning regression algorithms are chosen to train input data for predicting output target values. Real world data values are generally quite variable with an enormous variety of dimensions and ranges. Therefore, decision tree-based algorithms are chosen for their ability to train data without scaling, dimensionality reduction, or normalizing the input data (Pedregrosa et al., 2011). For our experiments, like all machine learning experiments, the data are divided into training and testing portions so that predictions (training) can be compared to data that is not included in the analysis (testing). We assigned 70% of the input data to training, and 30% to testing. Model results are assessed by examining performance indicators such as root mean square errors between training and testing data.

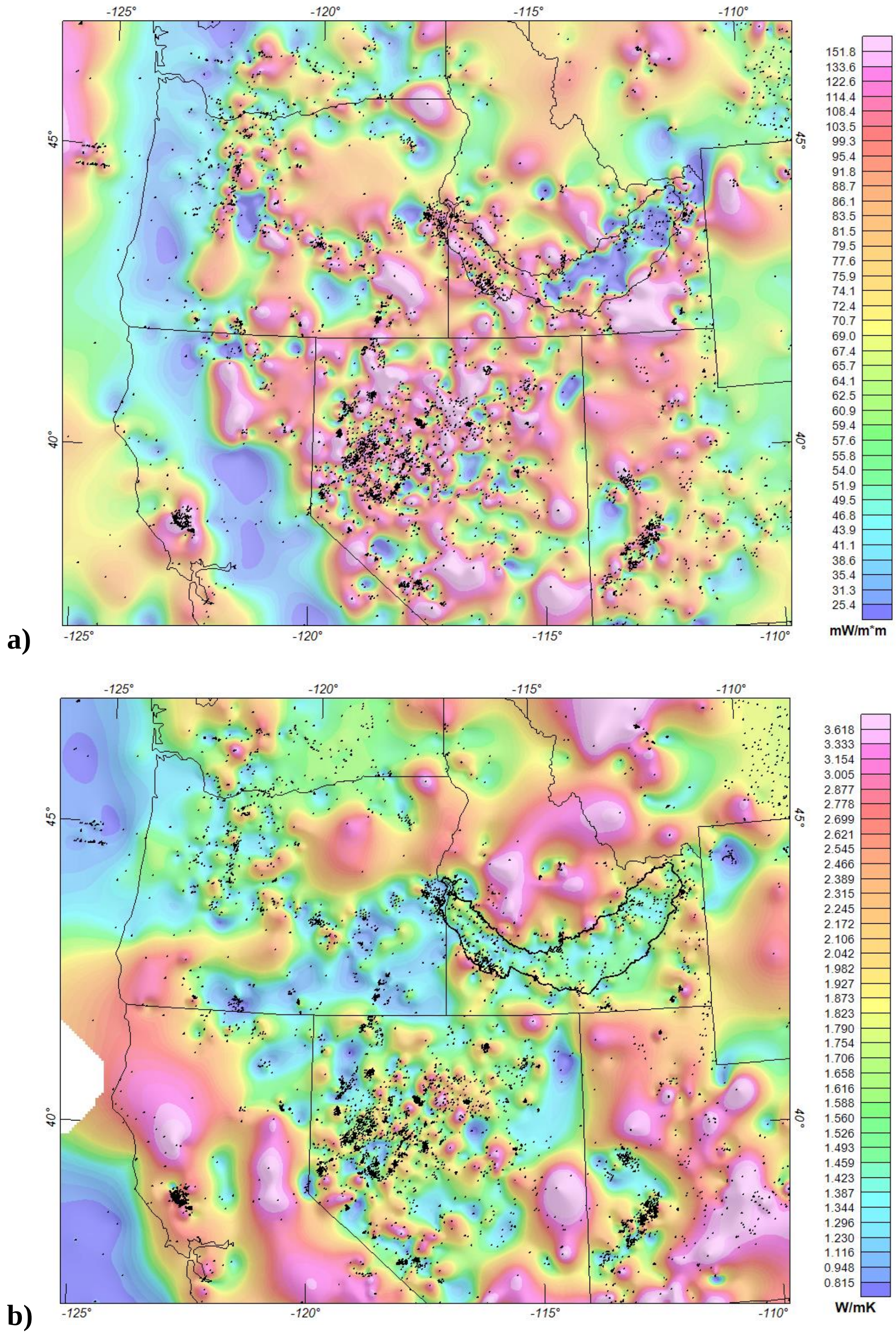

Figure 2: a) Heat flow grid of measured locations (black dots); b) Thermal conductivity grid of measured locations (black dots); Snake River Plain region outlined in southern Idaho.

Decision tree algorithms work by splitting the predictor space into “branches”, like a tree, formed from rules defined by data characteristics (James et al., 2023). However, decision trees sometimes overfit the data (too many branches are created) and predicted trained data accuracy decreases when compared with test data (Pandey et al., 2020). Overfitting can be addressed by tuning or pruning (James et al., 2023). Tuning modifies splitting (or segmenting) rules to prevent trees from growing too large, for example by minimizing the number of decision nodes. Pruning works by allowing trees to grow large, and then overly complex or minute segmentation divisions are removed in favor of more generalized segmentation rules (James et al., 2023).

Ensemble decision tree-based algorithms, such as Random Forest and Gradient Boost, produce more robust solutions than single decision tree algorithms. They work by growing multiple trees that are then amalgamated into a single prediction, thus avoiding problems produced by a single tree such as overfitting (Pandey et al., 2020). Extending this idea, Stacking Regressor is a composite of ensemble decision tree methods, built from the results of two or more ensemble estimators (Pedregrosa et al., 2011). We use a Stacking Regressor constructed by combined Tuned Decision Tree, Tuned Random Forest and Tuned Gradient Boost results. In our workflow we use measured location data for training and testing, and then apply the prediction model to a 5 km grid mesh.

4. Results

Predicted heat flow and thermal conductivity models, sampled on 5 km grids (Figure 3), are consistent with minimum curvature interpolations, also sampled on 5 km grids (Figure 2). Regions characterized by broad highs and lows map onto each other, such as the Coast Ranges or the basin and range morphology. However predicted data shows a higher level of detail throughout, especially where measured location data are sparse.

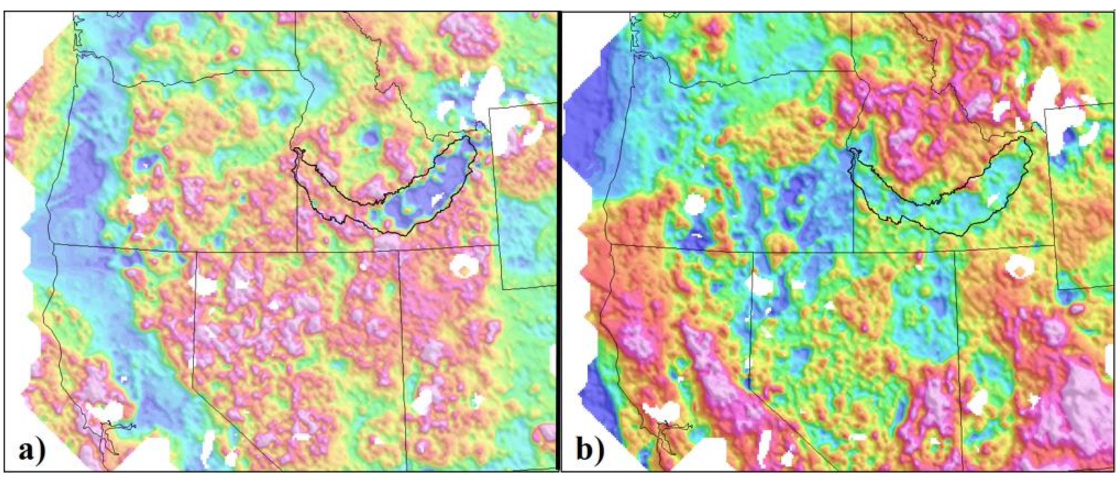

Figure 3: Supervised machine learning results. a) Predicted heat flow, b) Predicted thermal conductivity. Snake River Plain region outlined in southern Idaho.

Qualitatively, mapped results are interesting and suggest a significant improvement in heat flow and thermal conductivity coverage. Quantitative estimates of prediction results can be described by performance indicators: root mean square error (RMSE), maximum absolute error (MAE), R-squared (R^2) or coefficient of determination, and adjusted R-squared (Tables 1 and 2). RMSE is the square root of the mean squared error between predicted and actual values. MAE is the average error between measured and predicted values. RMSE and MAE calculations are consistent with input data ranges with lower values being the best results. However, RMSE is more sensitive to outliers because errors are initially squared.

R^2 indicates how well predicted results match the measured data, also called goodness of fit. If $R^2 = 1$, then the data are 100% predicted by the model, and if $R^2 = 0$, then the model has zero predictive power. In this study, R^2 is a measure of how well predicted and measured data fit a straight-line regression. However, R^2 either increases or remains unchanged even as additional predictors are added to the model, which means that some predictors do not contribute to the solution. Adjusted R^2 attempts to correct this optimistic result by disregarding non-contributing predictors. It is always equal to or less than R^2.

Table 1: Performance indicators of heat flow prediction.

	RMSE	**MAE**	**R^2**	**Adj. R^2**
Decision Tree	177.485	14.78641	0.89115	0.89068
Decision Tree Tuned	177.2521	16.20521	0.89144	0.89096
Random Forest Estimator	162.0735	12.99258	0.90923	0.90884
Random Forest Tuned	162.932	13.19822	0.90827	0.90787
Gradient Boost Estimator	156.4881	34.61037	0.91538	0.91501
Gradient Boost Tuned	168.7884	12.58656	0.90156	0.90113
Stacking Classifier	177.9082	15.51112	0.89063	0.89016

Table 2: Performance indicators of thermal conductivity prediction.

	RMSE	**MAE**	**R^2**	**Adj. R^2**
Decision Tree	0.30609	0.11856	0.75819	0.75721
Decision Tree Tuned	0.2903	0.14375	0.78249	0.78161
Random Forest Estimator	0.19742	0.10393	0.89941	0.899
Random Forest Tuned	0.19739	0.10377	0.89944	0.89903
Gradient Boost Estimator	0.33881	0.23587	0.70373	0.70254
Gradient Boost Tuned	0.17626	0.08728	0.91982	0.91949
Stacking Classifier	0.20566	0.10554	0.89083	0.89039

5. Discussion

Plate tectonic theory describes a global system where earth is capped by rigid lithospheric plates that are in relative motion with each other, and these relative motions produced broad regions of deformation (often over hundreds of km) along the plate boundaries. Regardless of reference frame, the relative motions of the Pacific and North American Plates are roughly to the northwest and west respectively, with Pacific Plate velocity being about three times that of the North

American Plate (Gripp and Gordon, 2002; Kreemer et al., 2003). Workers sometimes confuse structuring within plate boundary deformation zones with actual plate motions. For example, subparallel structuring along the San Andreas Fault is often thought to reflect North American – Pacific Plate motions, but it is actually structuring within the plate boundary deformation zone.

The net WNW-oriented divergent tectonic force between North American and Pacific Plates is oblique to a large part of the plate boundary between them (i.e., the San Andreas Fault), and it has produced the basin and range morphology of the Great Basin, broad dextral deformation in the Walker Lane, and the NW-oriented rift basin that lies beneath the western Snake River Plain (SRP). Further complicating this regional structuring, the Yellowstone mantle plume has produced a line of ENE-oriented felsic and mafic volcanic eruptions that lie beneath the central to east SRP (Figure 4).

5.1 Snake River Plain

Even though they note exceptions, Nielson et al. (2015) explained that basaltic terranes are not generally considered to be viable geothermal exploration targets, because deep-sourced basalt intrusives cool too quickly to be dependable heat sources. However, the blind Mountain Home hydrothermal system was discovered during their *HOTSPOT* play fairway analysis, and it has since been thoroughly studied for its geothermal potential (Varriale, 2016; Lachmar et al., 2019; Batir et al., 2020; Shervais et al., 2017 and 2020)

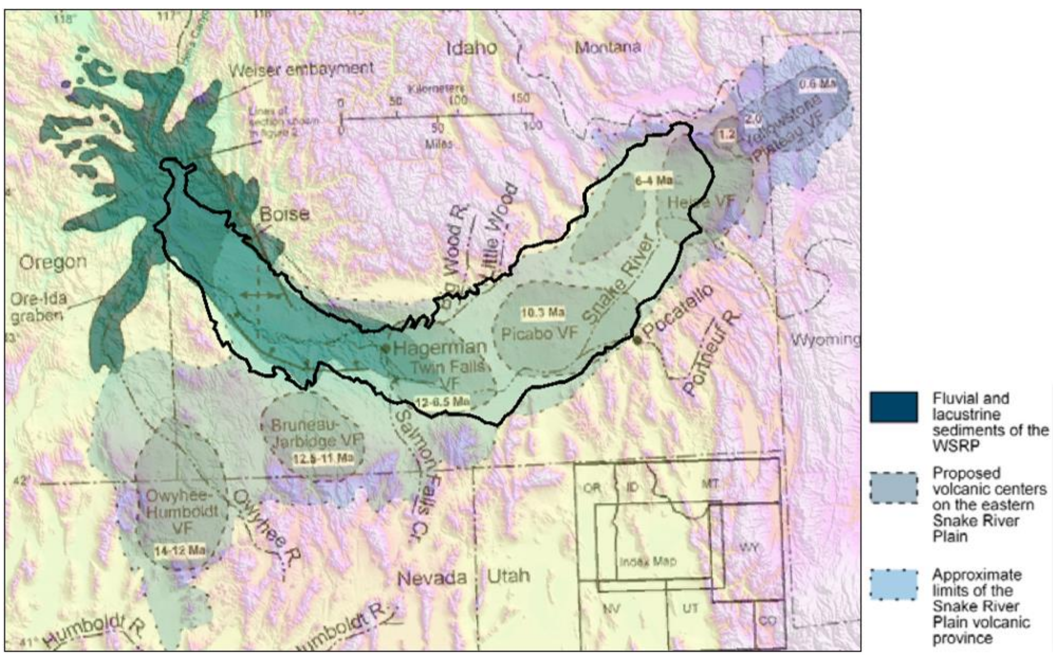

Figure 4: Snake River Plain (SRP) volcanic province including age-chronologic volcanic centers, decreasing from 14.6 to 0.6 Ma northeastward through the eastern SRP to the Yellowstone Caldera, and fluvial / lacustrine deposition in the western SRP (after Sant, 2012) draped over topography (GEBCO Compilation Group, 2020).

A prominent predicted thermal conductivity low corresponds closely with an outline of SRP (Figure 3b) and may be correlated with fractured basaltic flows. Comparing predicted heat flow in SRP (Figure 3a), the field changes abruptly, with higher heat flow over the western SRP, but low heat flow over the eastern SRP. However, older calderas that extend southwestward beyond SRP do not appear to produce low heat flow, which suggests that the prominent heat flow low in the east SRP might be related to the Snake River Aquifer.

Geothermal exploration play fairways, developed by Shervais et al. (2017 and 2020), are used to guide outlined regions of low and high exploration probability (Figure 5). The rift basin and younger fluvial and lacustrine sedimentary rocks in the western SRP are high potential areas, while the central to eastern SRP include smaller regions of high and low probability. Areas over the buried calderas are mostly associated with low probability and may be a complicating factor beneath central and eastern SRP. Predicted thermal conductivity cannot be directly correlated with exploration play fairways, suggesting a more complex relationship.

a)

b)

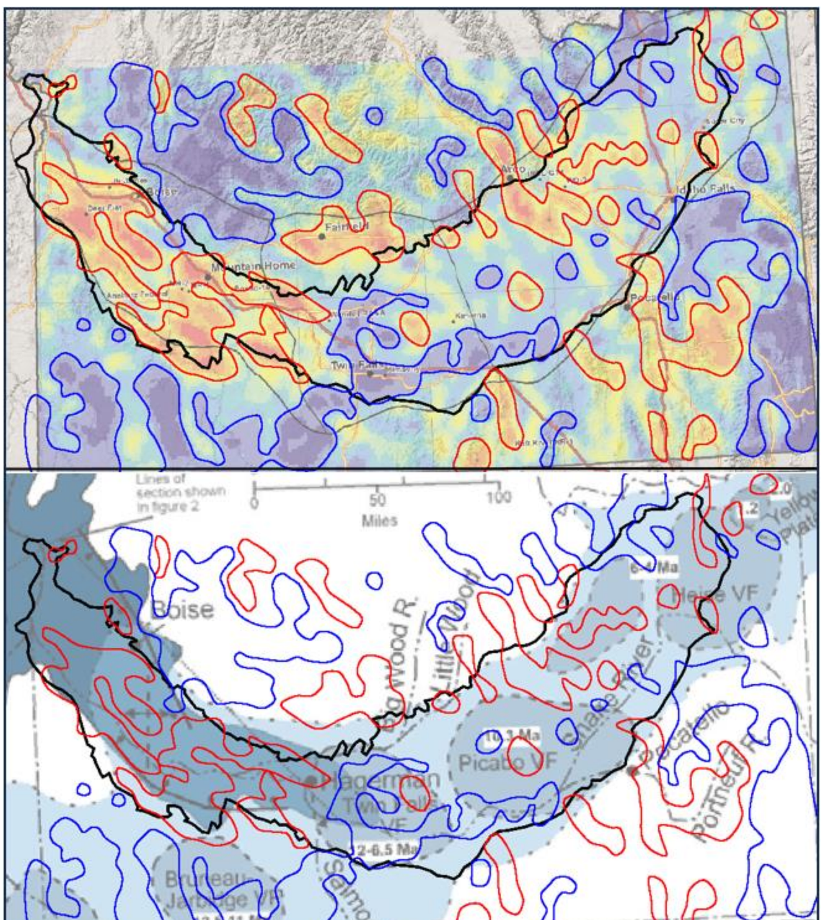

Figure 5: Snake River Plain play fairways. A) map of play fairway analysis results (after Shervais et al., 2020) with high and low probabilities outlines in red and blue respectively; b) high and low play fairway probability outlines over Sant's (2012) Yellowstone volcanic fields map.

5.2 Basement depth and terranes

Our workflow can be improved by including a basement depth horizon and basement terranes, which can be the basis for assigning heat production values. We have labeled this layer, between the top of the crystalline crust and Curie point depth, The Magnetic Layer (Bird et al., 2022 and 2023). Adding these components to the present study was not feasible because basement geometries and terranes are not yet well defined. We integrated inverted magnetic susceptibilities with mapped basement terranes (Whitmeyer and Karlstrom, 2007), and basement depths interpreted from aeromagnetic data, to predict heat flow in the Denver-Julesburg and Powder River Basins (Figure 6). We are planning similar work in parts of this northwest U.S. region.

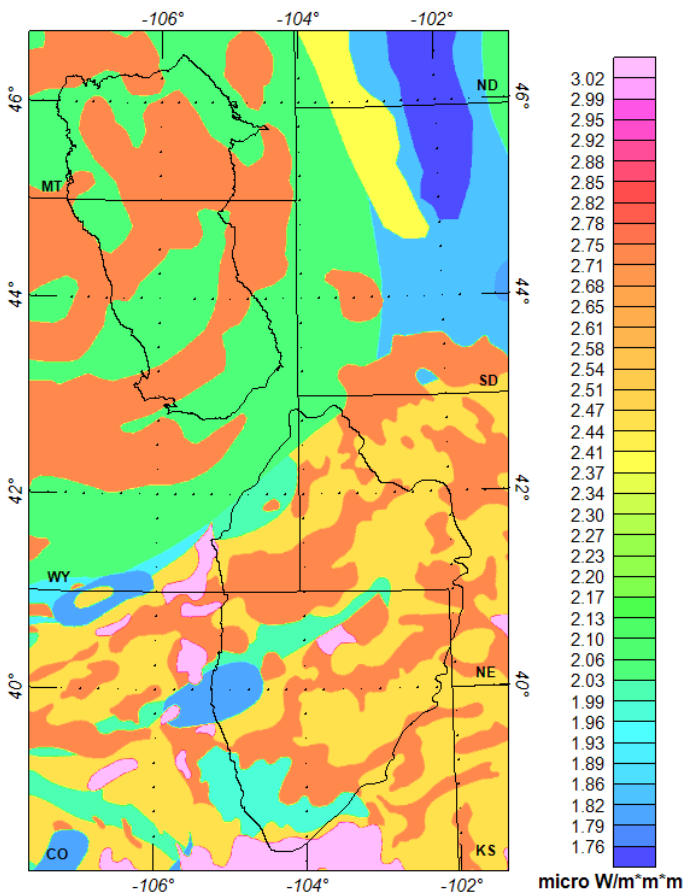

Figure 6: Outlines of Powder River (north) and Denver-Julesburg (south) Basins, located on the eastern limit of the Laramide deformation front. Gridded heat production follows Archean and Proterozoic basement terrane interpretations from an integration of 3D inverted magnetic susceptibilities and published terrane maps (Whitmeyer and Karlstrom, 2007; Bader, 2018 and 2019).

6. Conclusions

We successfully applied open-source supervised machine learning algorithms to predict heat flow and thermal conductivity in a large area of northwestern United States spanning over 1,000,000 km^2, essentially the Great Basin and surrounding region. Both predicted results are consistent with existing measured data as well as regional geology in the Snake River Plain.

All performance indicators of the decisions trees tested demonstrate acceptable results. This is true for the consistency of all RMSE and MAE results; R^2 and Adjusted R^2 values over 0.60 are usually acceptable, so over 0.80 is excellent, again for all outcomes. Hence, our preference of Stacked ensembles was based on qualitative examinations of mapped predictions. These promising results demonstrate that heat flow and thermal conductivity data are especially amenable to machine learning methods.

Improved estimates of heat flow and thermal conductivity can support geothermal play fairway analyses. A regular grid mesh of heat flow data may directly reinforce heat source estimates, and although not as straight-forward, a regular grid mesh of thermal conductivity may be useful for permeability modeling along prospective play fairways.

REFERENCES

Bader, J. W. “An Ancient Everest: Precambrian Basement Terranes of the Williston Basin”. *Geo News* (2019) January.

Bader, J. W. “Structural Inheritance and the Role of Basement Anisotropies in the Laramide Structural and Tectonic Evolution of the North American Cordilleran Foreland, Wyoming”. *Lithosphere* 11 (2108) 129-148.

Bansal, A. R., Gabriel, G., Dimri, V. P., and Krawczyk, C. M. “Estimation of Depth to the Bottom of Magnetic Sources by a Modified Centroid Method for Fractal Distribution of Sources: An Application of Aeromagnetic Data in Germany”. *Geophysics* 76 (2011) L11-L22.

Batir, J., Richards, M., Hornbach, M., Blackwell, D., Kolker, A., and Waibel, A. “Shallow Geothermal Potential of the Snake River Plain”. *Geothermal Resources Council Transactions* 44 (2020) 504-534.

Bird, D. E., Cathey, W. E., and Dowla, N. “The Magnetic Layer in Powder River and Denver-Julesburg Basins: A New Crustal Temperature Constraint for Estimating Thermal Properties in Sedimentary Basins”. *SEG International Exposition and Annual Meeting* (2022).

Bird, D. E., Cathey, W. E., and Dowla, N. “Heat Flow and Supervised Learning in the Permian Basin: Applying The Magnetic Layer (TML) to Control Thermal Property Calculations”. *SEG International Exposition and Annual Meeting* (2023).

Blackwell, D. D., Negraru, P. T., and Richards, M. C. “Assessment of the Enhanced Geothermal System Resource Base of the United States”. *Natural Resources Research* 15 (2006) 283-308.

Blackwell, D. D., and Richards, M. C. “The 2004 Geothermal Map of North America Explanation of Resources and Applications”. *Geothermal Resources Council Transactions* 28 (2004) 317-320.

Blakely, R. J. Potential Theory in Gravity & Magnetic Applications. *Cambridge University Press* (1995).

Bouligand, C., Glen, J. M. G., and Blakely, R. J. "Mapping Curie Temperature Depth in the Western United States with a Fractal Model for Crustal Magnetization". *Journal of Geophysical Research* 114 (2009) B11104.

Finn, C. A., Pilkington, M., Cuevas, A., Hernandez, I., and Urrutia, J. "New Digital Magnetic Anomaly Database for North America". *The Leading Edge* 20 (2001) 870-872.

Garcia, A. V., and Santamarina, J. C. "Heat Flow in Fractured Rocks: Stress and Moisture-Dependent Thermal Contact Resistance". *Geothermics* 95 (2021) 102113.

GEBCO Bathymetric Compilation Group 2020 "The GEBCO 2020 Grid – a Continuous Terrain Model of the Global Oceans and Land". *British Oceanographic Data Centre, National Oceanography Centre*, NERC, UK. doi:10.5285/a29c5465-b138-234d-e053-6c86abc040b9 (2020).

Gripp, A. E., and Gordon, R. G. "Young Tracks of Hotspots and Current Plate Velocities". *Geophysical Journal International* 150 (2002) 321-361.

Hasterok, D., and Webb, J. "On the Radiogenic Heat Production of Igneous Rocks". *Geoscience Frontiers* 8 (2017) 919-940.

James, G., Witten, D., Hastie, T., Tibshirani, R., and Taylor, J. *An Introduction to Statistical Learning: with Applications in Python*. Springer (2023)

Jessop, A. M., Hobart, M. A., and Sclater, J. G. "The World Heat Flow Data Collection – 1975". *Energy Mines and Resources Canada, Earth Physics Branch, Geothermal Service of Canada, Geothermal Series* 5 (1976).

Kluyver, T., Ragan-Kelly, B., Perez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., and Jupyter Development Team. "Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows". In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt. IOS Press (2016) 87-90.

Kreemer, C., Holt, W. E., and Haines, A. J. "An Integrated Global Model of Present-Day Plate Motions and Plate Boundary Deformation". *Geophysical Journal International* 154 (2003) 8-34.

Lachmar, T. E., Freeman, T. G., Kessler, J. A., Batir, J. F., Evans, J. P., Nielson, D. L., Shervais, J. W., Chen, X., Schmitt, D. R., and Blackwell, D. D. "Evaluation of the Geothermal Potential of the Western Snake River Plain Based on a Deep Corehole on the Mountain Home AFB Near Mountain Home, Idaho". *Geothermal Energy* 7 (2019) 1-15.

Li, C. -F., Lu, Y., and Wang, J. "A Global Reference Model of Curie-point Depths Based on EMAG2". *Scientific Reports* 7 (2017) 45129.

Pandey, Y. N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., and Luigi, S. *Machine Learning in the Oil and Gas Industry*. Apress L. P., Berkeley, California (2020).

Parker, R. L. "The Rapid Calculation of Potential Anomalies". *Geophysical Journal International* 31 (1973) 447-455.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michael, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanerplas, J., Passos, A., and Cournapeau, D.

"Scikit-Learn: Machine Learning in Python". *Journal of Machine Learning Research* 12 (2011) 2825-2830.

Roberts, B. J. "Geothermal Resources of the United States: Identified Hydrothermal Sites and Favorability of Deep Enhanced Geothermal Systems (EGS)". *National Renewable Energy Laboratory* (2018).

Sant, C. J. "Geothermal Alteration of Basaltic Core from the Snake River Plain, Idaho". *Utah State University* (2012).

Shervais, J. W., Glen, J. M. G., Nielson, D. L., Garg, S., Liberty, L. M., Siler, D., Dobson, P., Gasperikova, E., Sonnenthal, E., Neupane, G., DeAngelo, J., Newell, D. L., Evans, J. P., and Snyder, N. "Geothermal Play Fairway Analysis of the Snake River Plain: Phase 2". *Geothermal Resources Council Transactions* 41 (2017) 2328-2345.

Shervais, J. W., Glen, J. M., Siler, D., Liberty, L. M., Nielson, D., Garg, S., Dobson, P., Gasperikova, E., Sonnenthal, E., Newell, D., Evans, J., DeAngelo, J., Peakock, J., Earney, T., Schermerhorn, W., and Neupane, G. "Play Fairway Analysis in Geothermal Exploration: The Snake River Plain Volcanic Province". In *Proceedings, 45th Workshop on Geothermal Reservoir Engineering: Stanford University, Stanford California,* (2020) SGP-TR-216.

Surma, F., and Geraud, Y. "Porosity and Thermal Conductivity of the Soultz-sous-Forets Granite". *Pure and Applied Geophysics* 160 (2003) 1125-1136.

Varriale, J. A. "The MH-2 Core from Project HOTSPOT: Description, Geologic Interpretation, and Significance to Geothermal Exploration in the Western Snake River Plain, Idaho". *Utah State University* (2016).

Whitmeyer, S. J., and Karlstrom, K. E. "Tectonic Model for the Proterozoic Growth of North America". *Geosphere* 3 (2007) 220-259.