PROCEEDINGS, 49th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 12-14, 2024
SGP-TR-227

Power Density Geothermal Resource Estimation Revisited

R. Chadwick Holmes

Chevron Technical Center, a division of Chevron U.S.A., Inc., Houston, TX

chadwick.holmes@chevron.com

Keywords: resource estimation, power density, power per drill length, analogs, data analytics, machine learning

ABSTRACT

The power density method simplifies geothermal resource assessments to just reservoir temperature, resource area, and tectonic environment, using trends observed in a collection of global analog fields to predict power potential in undeveloped areas. For geothermal explorers, the procedural simplicity of power density over other resource estimation methods holds great attraction. However, the degree of variability in the field data used to define power density relationships suggests a non-trivial level of uncertainty in predictions made using traditional power density curves or related resource calculators. Power density also lacks a clear linkage to factors like drilling costs that greatly influence geothermal project economics. This study revisits the power density method by first evaluating its predictive performance in its three-variable form and then expanding the field data using multiple global databases. Representative values for subsurface characteristics like reservoir temperature and target depth are combined with features describing the surface plant design and climate conditions to paint a more complete picture of each field location. The data are analyzed to reveal the most important features for predicting power production in aggregate by field and on an individual power plant basis. New models created from these features maintain a low level of complexity appropriate for exploration. Furthermore, they are easily tuned to predict novel power metrics like power per drill length that incorporate project economic drivers to aid in geothermal strategy and portfolio-building efforts.

1. INTRODUCTION

Geothermal exploration activities serve to reduce the uncertainty in subsurface thermal potential available for direct use or power production. Still, preliminary estimates of the accessible heat-in-place are critical to justifying any budget spent on these activities, particularly investments in costly drilling operations. Explorers must often rely on simplified geothermal resource models to screen opportunities and prioritize which prospects to evaluate further. Of the most common resource estimation methods, the USGS volumetric heat-in-place calculation offers a well-documented and consistent process for determining thermal and electric power potential (Muffler, 1979). The USGS procedure considers the subsurface reservoir as a homogeneous heat store, and calculations combine earth properties and correction factors, each described either deterministically or with a representative value distribution. Depending on the derivation, the list of input variables may include reservoir temperature, reservoir areal extent and thickness, porosity, dual-phase fluid fraction, density and specific heat capacity of reservoir and fluids, recovery factor, heat-to-power conversion efficiency, plant capacity factor, and life of field (Ciriaco et al., 2020; Muffler, 1979; Sanyal & Sarmiento, 2005). However, defining appropriate values for each of these variables can be challenging in the data-limited context of early exploration work. This diminishes the applied value of well-parameterized estimators like the USGS method over lower-complexity measures tied to global analog data.

One example analog-driven alternative is the power density method, which uses power potential per unit area across a range of geothermal reservoir temperatures as a means for determining resource capacity. Depending on the global subset of fields considered, geothermal researchers have described increasing linear, power law, and exponential trends, often fit against a few tens of observations or less due to the difficulty in cataloging field power output and developed resource area (Bertani, 2005; Grant, 2000; Wilmarth et al., 2021). Wilmarth and Stimac (2014) detailed a methodology for consistently calculating field area using just production well paths and a 500 m radius outline buffer, and their follow-on reports include a growing curated data set (Wilmarth et al., 2021; Wilmarth & Stimac, 2015). These data are described to follow two separate trends: a "main sequence" that exhibits roughly exponential growth in power density with reservoir temperature, and a constant power trend centered just under 10 MW/km^2 (Figure 1). The constant trend correlates most with "hot arc" fields, i.e., areas marked by high-temperature, compressive tectonic regimes where volcanism occurs in arcs above a subducting plate. Other tectonic regimes are believed to cluster along different sections of the main sequence curve, leading past authors to suggest a predictive relationship connecting tectonic setting and reservoir temperature to power density.

Without visual guides like color-filled polygons or curves projected on the plot in Figure 1, two characteristics of the data stand out. First, the spread in the power density values increases with reservoir temperature, broadening by a factor of 2 for fields with an average reservoir temperature >250°C compared to lower-temperature geothermal assets. Secondly, the overlap in tectonic setting markers suggests clusters based on tectonic classification are not cleanly separable. Because the data set only includes ≈100 fields, data coverage is both sparse and imbalanced across the temperature spectrum. These observations collectively underscore a degree of uncertainty that accompanies power density-based resource estimation. In this study, the objective is to evaluate the impact of this uncertainty on the predictive ability of power density curves, while also appraising the choice of a field-scale, 3-parameter representation of geothermal resources against alternative methods. The results provide new insights into the efficacy of power density estimation methods for geothermal exploration assessments and highlight the potential for future refinement with additional data engineering and analytics approaches.

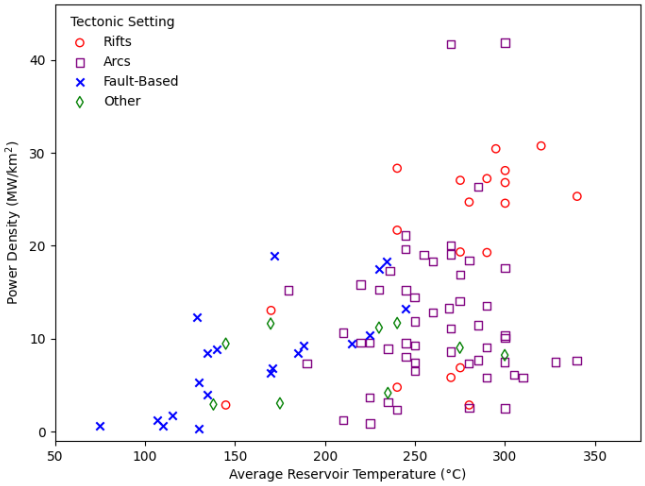

Figure 1: Power density of 102 geothermal fields plotted against reservoir temperature, adapted from Figure 1 in Wilmarth et al. (2021). Colors and shapes depict tectonic setting classifications. The data show an apparent increase in power density with greater reservoir temperatures, paired with increasing variance.

2. METHODOLOGY

2.1 Data Collection

The initial data collection consisted of digitizing 102 identifiable points from the power density figure included in the GeoRePORT Protocol for the Resource Size Assessment Tool (Rubin et al., 2022). This report accompanied the release of the RSAT tool, which directly implements a power density resource estimator based on the two-curve fit of the data. Additionally, both the field name and tectonic setting for the points were manually transcribed from the same GeoRePORT plot. Figure 1 depicts this base data set, which includes 99 labeled geothermal fields and 3 unknown fields that were unlabeled in the source figure.

2.2 Data Expansion

The base field data set was expanded by merging it with several ancillary geothermal power plant databases. The first is a power plant location file uploaded to the Geothermal Data Repository (https://gdr.openei.org) with the Cascades/Aleutian Play Fairway Analysis data submission (Shevenell, 2015). This file contains geocoordinates marking individual power plant facilities around the globe, verified by visual inspection in Google Earth. The Global Power Plant Database (GPPDB) provides greater facility-level detail, including location country, commission year, and nameplate capacity (Byers et al., 2018). Estimates of net power production appear in the GPPDB as well, although these contain a mixture of official reports, modeled approximations, and null values and consequently were dropped from consideration in this study. The last data set incorporated into the analysis comes from Rystad Energy's GeothermalCube (Rystad Energy, 2023). GeothermalCube uniquely provides a total well count, the number of active production wells, and the average drilling depth by power plant location. Rystad captures information on historical geothermal fields, plants under construction, and future development plans spanning hydrothermal, enhanced (EGS), and advanced closed loop (AGS) geothermal systems. However, because EGS and AGS system entries primarily consist of past and future pilot data, only hydrothermal geothermal systems were included in this study. The data was also filtered to only include power plants with commission dates no later than 2022.

All three supplementary data sets contain entries at the power plant level of detail. To merge them with the base data set, power plants were assigned to a designated field and then aggregated up to the field scale. This meant taking the mean for numerical variables like latitude, longitude, and reservoir temperature, using the first value for labeled data like country name, and applying the mode on ordinal data like drilling length group which was already binned in ranges of 500 m. The unknown geothermal fields were dropped from the base data set. Some geothermal fields include power plants of different power system configurations. Because the power system type is tracked as a variable in the expanded data, the final field-scale data set consisted of 125 entries.

2.3 Data Engineering

Both the original power density plot and the Rystad data set include estimates of the reservoir temperature for each geothermal field. A one-to-one scatter plot of the two variables showed that these reported values varied, sometimes by several tens of degrees. The reservoir temperatures of 16 entries that had values with a difference of 50°C or more were replaced with the most recent numbers from studies on those fields based on a search of the literature.

The existing columns in the data set – also known as "features" – were then augmented to provide a richer set of inputs for advanced data analysis. To start with, continuous variables, including reservoir temperature, location (latitude, longitude), and commission year, were binned into ranges to increase the instance count across representative values for each variable. Rather than presuppose effective binning schemes, the Jenks method was applied to determine the natural breaks for each feature based on a specified number of bins (Jenks, 1967). Using a range of bin counts, optimal values for each variable were selected based on the minimum imbalance ratio, i.e., the ratio of the

sample size for the largest majority class to that of the smallest minority class. Next, all binned features were encoded as ordinal variables, preserving the order of value options while replacing values with integer representations. Lastly, new features related to surface conditions were incorporated into the data set. Topography was sampled from the ETOPO1 1 arc-minute resolution model based on the latitude and longitudes for each field (Amante & Eakins, 2009). Climate 30-year normals for annual average temperature, minimum temperature, maximum temperature, and annual precipitation were likewise captured using the WorldClim version 2.1 30-sec resolution gridded models (Fick & Hijmans, 2017). WorldClim features provided as monthly normals were first combined into an annual grid before sampling at field locations, using the mean function for average values, min for minima, and max for maxima. Additionally, the difference between the minimum and maximum temperature values was calculated to define an ambient temperature range feature.

2.4 Exploratory Data Analysis

The original field-scale data set was first examined with the power density curves for the main sequence and hot arc fields illustrated in Wilmarth et al. (2021). The use of statistics on the residuals helped to highlight outlier areas that could be considered anomalies to the overall trend. Additionally, the root mean squared error (RMSE) and coefficient of determination (R^2) were calculated for evaluation and direct comparison of the data fit and predictive ability between the curve-based model and other models.

With the expanded field-scale data set, addressing the question of which features provide the most value began with determining the level of feature independence. If two features vary in a colinear fashion, removing one from the data set can reduce data dimensionality, often with little loss in the predictive ability of the remaining features. Here, the decision to eliminate certain features was based on pairwise feature correlations and the evaluation of mutual information in relation to the predicted variable, power density. The aim was to focus the data set down to the key data features that could improve an explorer's ability to evaluate geothermal prospect potential.

2.5 Modeling

Given that exploration often depends on rapid, reproducible, and justifiable assessment methods in scenarios with sparse data, this study primarily focused on low-complexity modeling algorithms. The first followed past efforts to evaluate power density trends using the algebraic best-fit lines with an assumed character, i.e., exponential for the main sequence and constant value for the hot arc fields. Three machine learning techniques – linear regression, lasso (least absolute shrinkage and selection operator) regression, and decision tree regression – were then applied to the augmented field-scale data set. Alternative models were developed using power plant-scale data to examine the effects of data loss due to the aggregation of the original data to the field scale. Next, a random forest model was implemented to demonstrate how more complex machine learning techniques, such as ensemble models, can enhance predictive capabilities. Lastly, the dependent variable in these predictive models, power density, was reassessed within the framework of suitable exploration metrics, and new models were constructed to predict an alternate metric that combines power with the cumulative drill length.

3. RESULTS

3.1 Analysis of the Original Power Density Model

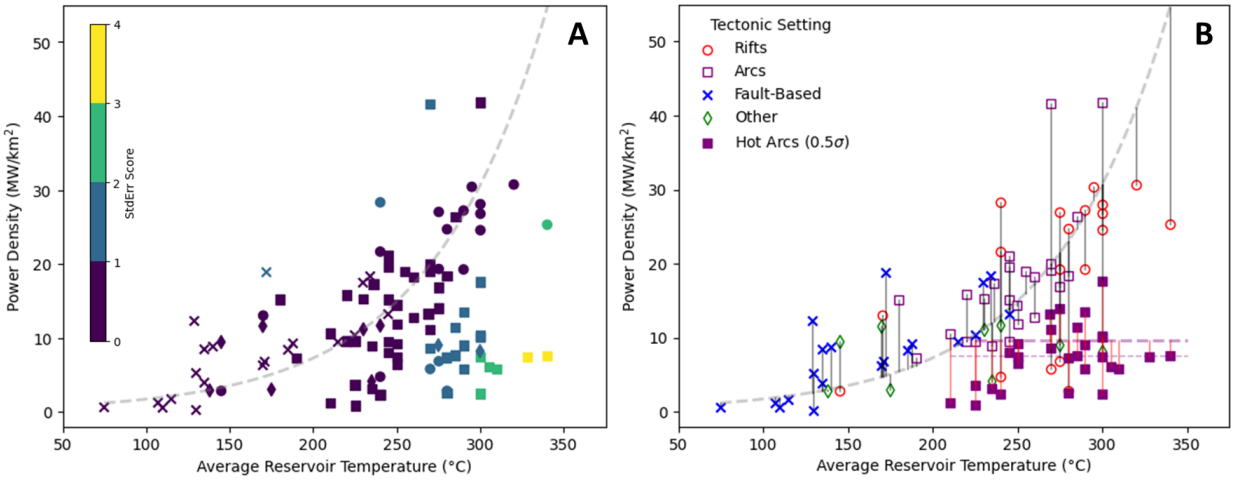

Figure 2: A. Power density plot with the Wilmarth et al. (2021) exponential curve fit as a dashed line. Field locations are colored by the normalized standard error to illustrate the presence of just 2 fields as outliers by the 3-σ rule. B. Residual plot illustrating the increasing deviation of field data from the exponential curve with reservoir temperature. The hot arc constant power line suggested in Wilmarth et al. (2021) is plotted as a bold purple dashed line. The hot arc-designated fields (filled markers) were defined for the plot using a cutoff of 0.5 standard deviations of the residuals from the exponential curve, and their mean value (7.6) is shown in lighter purple dash to demonstrate how the choice of hot arc cutoff can impact the best-fit constant power line.

Figure 2A visualizes the entire set of 102 geothermal fields from the original field-scale data set. The marker shapes match the tectonic classification from Figure 1, color-coded by the normalized standard error derived from fitting the data solely with the main sequence

exponential curve (y = 0.4084 e^0.0144x, Wilmarth et al., 2021). Applying a 3-σ heuristic for outlier identification, only the two fields highlighted in yellow meet the criteria of anomalous power density with temperature. Both are classified as Arc tectonic regimes and demonstrate an estimated power density of under 10 MW/km^2. The large spread of the data about the exponential curve is reflected in the RMSE of 12.2 MW/km^2. In addition, the negative R^2 statistic, which is defined by subtracting from 1 the ratio of model residual sum of squares to the sum of squares about the mean, suggests the curve is a weaker predictor for power density compared to simply using the mean of the data (see Table 1).

Figure 2B illustrates the residuals for each field when the secondary hot arc line is included in the model ($y = 9.7\ MW/km^2$, Wilmarth et al., 2021). Measuring the goodness of fit for this two-curve model requires a judgment on which Arc-labeled fields should be tied to the main sequence line. Table 1 describes the impact of applying different cut-off values to separate the main sequence and hot arc associations for fields assigned to the Arc tectonic setting, each threshold based on the normalized standard error from the single-curve (exponential-only) model. RMSE ranges from 10.6 MW/km^2 at a 3-σ threshold to 7.2 MW/km^2 for a 0.5-σ threshold. Similarly, R^2 values progressively improve to 0.31, indicating some predictive ability for the dual-line model. However, further adjustments offer little model improvement; when all Arc fields beneath the exponential curve are assigned to the hot arc designation, and the constant power line is shifted to match the mean of the hot arc subset (9.0 MW/km^2), the RMSE and R^2 slightly degrade to 7.3 MW/km^2 and 0.29, respectively.

Table 1: Comparison of different power density curve models using root mean squared error (RMSE) and the coefficient of determination (R^2). Negative R^2 indicates the model performs worse than simply using the mean of the data as a predictor.

Model	RMSE (MW/km^2)	R^2 (unitless)
Exponential curve only	12.2	-1.00
Exponential curve + Hot Arc curve ("hot arcs" identified from 3-σ cutoff)	10.6	-0.52
Exponential curve + Hot Arc curve ("hot arcs" identified from 1-σ cutoff)	7.4	0.26
Exponential curve + Hot Arc curve ("hot arcs" identified from 0.5-σ cutoff)	7.2	0.31
Exponential curve + shifted Hot Arc curve ("hot arcs" identified from 0-σ cutoff)	7.3	0.29

3.2 Dimensionality Reduction of the Expanded Field-Scale Data Set

The matrix in Figure 3 shows pairwise correlation values for the continuous and encoded variables in the expanded data set. PDTemp and PDPowerDensity refer to the original reservoir temperature and power density data pairs displayed in Figure 1. Data features that were engineered into binned values predictably show significant correlation with their continuous counterparts (e.g., LatitudeDegree and LatitudeBinned, StartYear and StartYearBinned). NameplateCapacity, ProdWellCount (production wells), and TotWellCount (all wells) exhibit perfect collinearity (+1.0), which follows the logic that the number of wells drilled in a field is intrinsically associated with the capacity of existing power facilities. In addition, surface climate characteristics and spatial location parameters such as Country, Latitude, and Longitude give rise to strong pairwise correlations (>0.5). This means that a selected subset could effectively represent the information gain obtained by using all related features. Based on these observations, and with the goal of simplifying the data set to its lowest-dimensional form, the following features were selected for removal: Country, LatitudeDegree, LongitudeDegree, NameplateCapacity, TotWellCount, StartYear, MinAmbientTemp, MaxAmbientTemp, and AmbientTempRange.

Mutual information (MI) measures the association between two variables, such that fully independent variables will have an MI of zero and stronger variable associations lead to higher MI values (Cover, 1999). To evaluate feature importances, MI analysis pairs each data feature with the variable being predicted. Here, a small MI value implies a feature offers lower predictive value to the power density prediction. The type of power system installed at a power plant (PlantType), the number of production wells (ProdWellCount), and year of commission (StartYearBinned) all exhibit the smallest MI scores (Figure 4). The features with the greatest predictive potential based on MI are geospatial features (e.g., LatitudeBinned, Region), topography (Elevation), reservoir temperature (PDTemp), and tectonic classification (TectonicRegime). This analysis could inform further reductions of the expanded data set, however several of the approaches discussed in Section 3.3 naturally reduce the feature set during model-building. For this reason, the MI results were considered a benchmark for later analysis and the twelve features shown in Figure 4 were kept for machine learning.

3.3 Machine Learning Applied to Field and Plant-Scale Data

Using known algebraic structures like constant and exponential curves as the foundational relationship between two-dimensional data (i.e., $y = f(x)$) is an intuitive choice of techniques. In the context of higher-dimensional data like the expanded data set under consideration here, machine learning methodologies are proficient in fitting a diverse range of models, and the choice of method may depend on the preferred degree of model complexity. Two contrasting examples can be considered: linear regression, which yields a simple weighted sum of features, and deep neural networks, which establish an intricate network of interconnections often characterized as black box in nature.

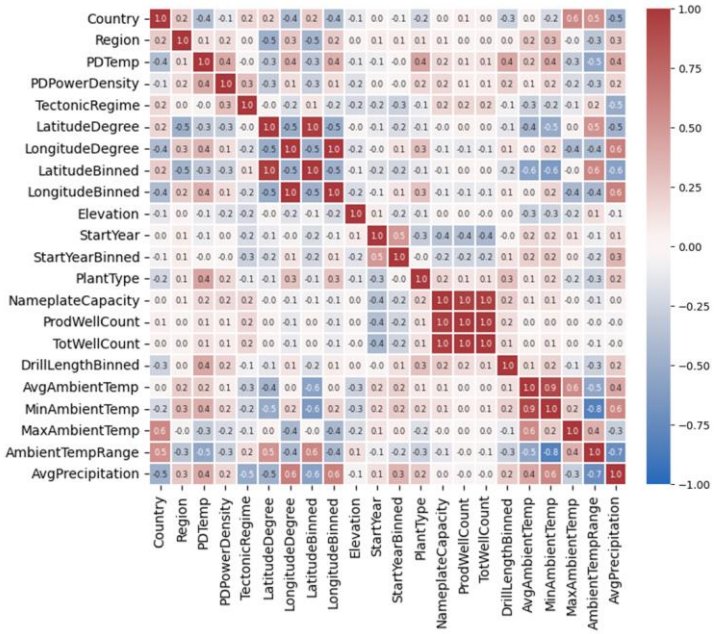

Figure 3: Pearson correlation matrix illustrating pairwise relationships between different features in the expanded field-scale data set. Correlation values of ±1.0 indicate two features are colinear, i.e., they vary linearly with full dependence.

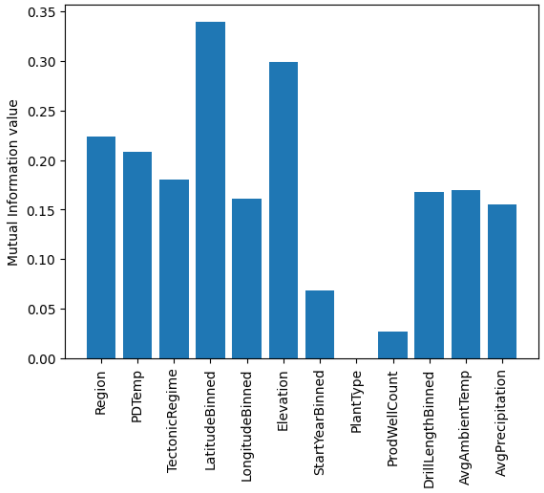

Figure 4: Mutual information values for each listed feature versus power density from the field-scale expanded data set. The smaller the value, the less predictive the feature is for determining power density.

The results shown in Table 2 capture outcomes of applying linear, lasso, and decision tree regression models on the data. Unlike other machine learning approaches, the predictions from these methods are transparent, traceable, and arguably of appropriate complexity for early opportunity screening exercises. Results with the "Field" label were derived using the field-scale expanded data set described in Section 2.2. The second set of results, noted by the label "Plant," were calculated using data from the same fields disaggregated at the power plant scale, totaling 382 unique entries. All models are trained to predict power density, although the Plant models rely on a pseudo-power density value calculated from the nameplate capacity scaled by the sum of 500 m-radius circular areas positioned at each production well. This differs from the power density at the field scale, which Wilmarth et al. estimated using net power production values and a map-based area determined by tracing a 500 m-radius buffer along actual production well paths (2021). Implicit in this substitution are two necessary assumptions: i) the nameplate capacity and net power production values differ in a predictable way, and ii) the additional area introduced by deviated well paths is compensated for by the overlapping area around closely spaced production wells that gets multiply counted. All machine learning cases are trained on 80% of the input data and checked against a separate randomly sampled 20% to allow for out-of-sample testing of model performance. A lightly tuned random forest regression model was also trained on the plant-scale data to demonstrate the improvement in predictive ability achievable using a more complex ensemble model formulation (Table 2).

Table 2: Fit and performance comparison of different power density machine learning models based on root mean squared error (RMSE, in MW/km^2) and the coefficient of determination (R^2, unitless).

PowerDensity model	RMSE (train)	RMSE (test)	R^2 (train)	R^2 (test)
Linear Regression (Field)	1.6	3.4	0.42	0.01
Lasso Regression (Field)	1.7	3.4	0.38	0.00
Decision Tree (Field)	1.2	2.9	0.65	0.26
Linear Regression (Plant)	3.1	3.5	0.25	0.24
Lasso Regression (Plant)	3.1	3.5	0.24	0.24
Decision Tree (Plant)	1.2	2.4	0.88	0.64
Random Forest (Plant)	1.4	2.0	0.84	0.76

Field-scale models show improved data fit but poorer predictive performance than the two-curve model described in Table 1. All three field machine learning models have an $RMSE_{test}$ of roughly 4 MW/km^2 better than the best case (0.5-σ cutoff) two-curve model, but only the decision tree model R^2_{test} comes close to the R^2 of the same two-curve case (0.26 vs. 0.31). However, when power plant-scale data are used to train the models, both performance metrics show significant improvement. Linear and lasso regression models achieve an $R^2 \approx 0.25$, while the decision tree algorithm demonstrates an out-of-sample $R^2_{test} > 0.6$. Hyperparameter tuning for the decision tree used grid-search optimization with 5-fold cross-validation, but over-fitting might be a concern based on the differences in performance metrics between the test and training data (Table 2). A similar tuning methodology was applied to the random forest model, and while the test vs. train results still suggest some overfitting, the test set metrics improve even more; the $RMSE_{test}$ drops to 2.0 MW/km^2 and R^2_{test} reaches 0.76, which is superior to the RMSE of 7.2 MW/km^2 and R^2 of 0.31 for the best 2-curve model in Table 1.

3.3 Feature Importances from Machine Learning Models

Feature importance analysis using trained machine learning models can highlight which data features consistently contribute the most to a prediction. For this study, SHAP analysis (SHapley Additive exPlanations) was preferred for deriving feature importances (Figure 5). This method reveals feature contributions to local point predictions, and the average absolute SHAP magnitude across all points approximates the general significance of the features globally (Lundberg & Lee, 2017). For the plant-scale models, features associated with geography and climate show greater importance than reservoir temperature or tectonic setting, which runs counter to the original choice of variables for the power density method (Figure 5D-G).

3.4 Accounting for Economics in Power Prediction

Well count and average drilling depth strongly influence the overall economics of geothermal projects. In fact, drilling costs can account for 50% or more of a geothermal project budget (Robins et al., 2022). Exploration portfolio managers will naturally focus both on the power potential and economics of heat capture in opportunity prioritization decisions, which suggests metrics tuned to both would be helpful for early screening decisions as well. Rather than simply considering power per developed resource area for analog fields, explorers might instead consider power per drill length. Such a metric could be normalized by the number of wells, making it a useful proxy for a power to cost ratio applicable to new field prediction. Power per Drill Length (*PDL*) is defined by:

$$PDL = \frac{P}{\sum_{w=1}^{W} d_w}, \tag{1}$$

where P is the power production, W is the well count, and d_w is the drill depth for well w. Explorers considering a new geothermal opportunity can start with the estimated reservoir depth and multiply that by a PDL to determine the power potential for a well. Unlike power density, this metric inherently accounts for the subsurface dimension and the related cost component for a project. Variants of the metric might consider measured depth rather than total vertical depth – an important distinction as geothermal drilling moves more toward highly-deviated or horizontal well designs. For the purposes of this study, the hydrothermal-focused historical field data is assumed to largely comprise vertical or near-vertical wells. Furthermore, PDL is approximated using the nameplate capacity for power, the production well count for W, and the average drill length for d_w.

Results shown in Table 3 reflect the outcome of applying the same machine learning algorithms to predict PDL from the plant-level data set. DrillLengthBinned and ProdWellCount were removed from the training data since they already contribute directly to calculating the dependent variable (PDL). The predictive performance of the linear models remains low, but both the decision tree and random forest achieve an $R^2_{test} > 0.4$. In both cases, R^2_{test} differs from R^2_{train} by ≈ 0.3, suggesting the models are overfitting the training data and could benefit from additional hyperparameter tuning and regularization. Reservoir temperature dominates the other features based on the SHAP analysis in Figure 5H-I. Both the decision tree and random forest models weigh the importance of TectonicRegime over Region or Latitude, counter to the trend observed in the previous models. Revisiting the interplay between these variables would be useful future work to clearly determine feature dependencies and differentiate which should be included in future modeling efforts.

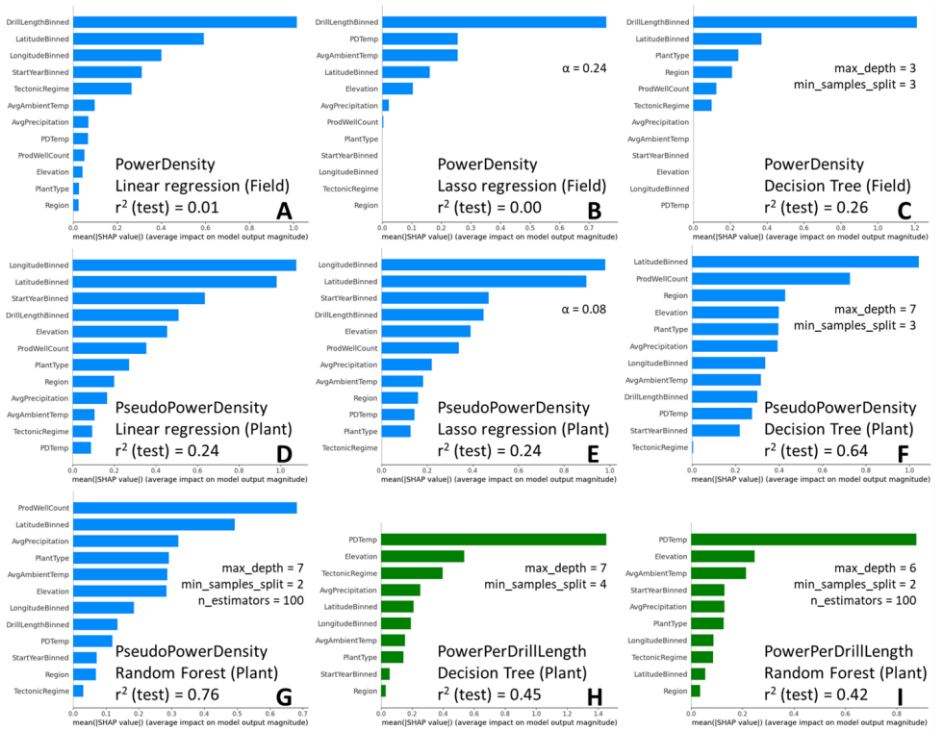

Figure 5: Feature importances by model based on average SHAP magnitude. Features are sorted in descending order such that those contributing the most to model prediction appear near the top of each plot. Large-font annotations indicate the variable being predicted, the machine learning algorithm applied, the data set utilized, and the out-of-sample R^2 outcome. Model hyperparameters tuned using a grid search optimization routine are noted in finer font on each plot as well. A.-C. models trained on the field-scale data set to predict power density. D.-G. models trained on the plant-level data set to predict pseudo-power density. H.-I. models trained on the plant-level data set to predict power per drill length, also shown with green bars to highlight the difference in dependent variable. Models trained on the plant-level data tend to be more performant. Decision trees and random forest models achieve better results than linear and lasso regressors.

Table 3: Performance metrics measured for different machine learning models trained to predict Power per Drill Length (PDL).

PowerPerDrillLength model	RMSE (train)	RMSE (test)	R^2 (train)	R^2 (test)
Linear Regression (Plant)	2.4	2.3	0.16	0.13
Lasso Regression (Plant)	2.4	2.3	0.16	0.13
Decision Tree (Plant)	1.4	1.8	0.73	0.45
Random Forest (Plant)	1.2	1.8	0.79	0.42

4. DISCUSSION

Implications for Geothermal Resource Estimation

One of the fundamental drivers for the geothermal exploration community to use resource estimation methods is to capture a reasonable assessment of different prospects, specifically for the purpose of understanding which opportunities should be actioned further. The upside potential matters because it helps highlight the hopeful best-case scenarios, but low sides determine the viability of a geothermal program, and potentially an exploration business for cost-constrained firms. The residuals observed in Figure 2B tell an important story. Using average reservoir temperature and an interpreted tectonic setting with the 2-curve power density model could result in not insignificant mischaracterizations of resource size. RMSE varies from 7.2-12.2 MW/km^2 depending on the subset of Arc-assigned fields treated as "hot arcs" (Table 1), which equates to a standard error of ≈ 55-95 MW if production area is assumed to match the mean area in the base data set (7.9 km^2, Wilmarth et al., 2021). And because the production area follows a lognormal distribution, some of the fields in the base data are significantly larger than the mean, with an equivalently outsized error in power estimates. Some recent power density plots illustrating the 2-curve model are paradoxically annotated with an R^2 value > 0.70 (see Rubin et al., 2022; Wilmarth et al., 2021), but the findings in

Table 1 indicate this must be incorrect. Alone, the exponential curve achieves no better than a negative R^2, and the two-curve solution does not seem to exceed an R^2 of ≈ 0.3.

Results obtained after applying different machine learning approaches to the data support using a decision tree as a performant alternative. Decision trees also meet the low-complexity objective; they reduce resource estimation to a series of simple binary classification criteria. The true or false boolean decision outcome at each tree node directs the traversal of the tree until a leaf node with a power density prediction has been reached. Decision trees could be implemented using code, nesting *if* statements in an Excel workbook, or even calculated by hand. Random forest results show promise as well, but the mass ensemble of decision trees within a random forest renders the final model less explainable and traceable to users and decision makers than a single decision tree. Paramount to unlocking the predictive potential of these tree-based models was replacing the field summary data with individual power plant information. In fact, the decision tree trained on field data demonstrated an R^2_{test} similar to the two-curve model, but shifting to the plant-scale data resulted in an altogether different model with 2.5× better R^2_{test} results. Figure 6 depicts the structure of the final decision tree for plant-scale data. The shortest traversal requires two decisions before reaching a power density estimate, while the deepest traversal involves seven decisions.

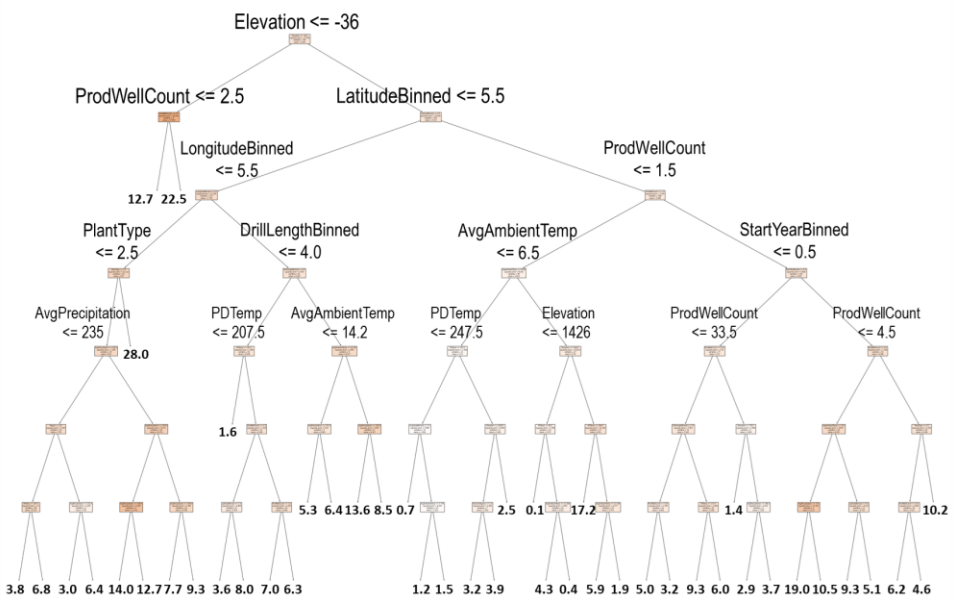

Figure 6: Final decision tree structure after training on the data consisting of individual power plant entries. Leaf nodes are labeled with the associated pseudo-power density prediction in MW/km^2. The boolean decision rules are only annotated for the uppermost layers to reduce visual clutter in the graphic.

SHAP value analysis indicates a strongly dominant influence of latitude and production well count in the plant-scale decision tree and random forest models for pseudo-power density (Figure 5F-G). Interestingly, tectonic setting consistently ranks among the least important two features for all power plant models (Figure 5D-G). One possible explanation for this disconnect with the original power density temperature-tectonic framework might be that tectonic setting is an interpreted feature, derived using observed spatial patterns and a plate tectonic reference framework. Machine learning models trained on location-based features like latitude, longitude, and region can conceivably reconstruct the same patterns without the need of a tectonic label. Commission (start) year also tracks low in the importance ranking, which builds some confidence in using a combination of recent and decades-old power plant information when training models to predict outcomes for new plant locations. Overall, the features that matter most relate to geography and climate conditions, followed by subsurface characteristics like drilling depth and reservoir temperature. This matches trends seen in the MI values in Figure 4, although reservoir temperature and drilling length dominate both the power plant system and the production well count by MI, counter to SHAP importances. Improvements to the decision tree model could be the focus of future work, including additional tuning of the decision tree parameters, further creative data engineering efforts, and expanding the data set further with features not yet captured. Revisiting the feature importances after that work will help validate these trends, which in turn can inform which data explorers should focus on gathering to improve their early screening power assessments.

One barrier to constructing any power density model is the ready availability of net power production values for individual power plants or their associated fields, not to mention well path information needed to calculate a production area. The pseudo-power density calculation used in this study was sufficient to develop proof-of-concept machine learning models, but working with simpler metrics that can be calculated with consistency using easily accessible data would be preferred. Additionally, power density as a metric cannot account for the variances in power plant productivity and economics linked to target reservoir depth. Using PDL partly addresses these issues by removing the need to calculate production area while also accounting for the cumulative borehole lengths drilled for existing fields. This doesn't negate the need or desire for a reliable and complete compendium of net power output for existing power plants. Without this important data, the nameplate capacity can stand in but will mask variances related to oversized facilities, which in turn can skew the models constructed to estimate power production for new areas. Nevertheless, this study helps illustrate the possibility of incorporating well count and drilling lengths into the predicted variable as an evolution of existing power density methodologies. Of course, introducing

this metric effectively removes two contributing features to predicting power density – production well count and average drill length – from the feature set, and the resulting models demonstrate some reduction in data fit and predictive capabilities compared to the plant-scale models for pseudo-power density (Table 2 vs. Table 3). With more power plant data, and potentially additional features not considered here, PDL model improvement seems likely. Plus, the benefits of having a power prediction that links to the drilling requirements of an opportunity naturally allows for prospect ranking and portfolio management practices that geothermal businesses will need to stay competitive.

5. CONCLUSIONS

The power density method that relies on a two-curve fit to a global collection of geothermal fields falls short of demonstrating strong predictive performance based on statistical measures, including root mean squared error (RMSE) and the coefficient of determination (R^2). Lower-complexity machine learning models like linear, lasso, and decision tree regression also struggle with predicting power density, even after including additional information on geothermal field design or its local geography and climate. Expanding the data to separately track individual power plants within the same fields greatly improves machine learning results. The decision tree model performs particularly well, and its transparency and explainability as a predictive model make it a preferred alternative to the two-curve power density representations in published literature. Nevertheless, power density alone may not be sufficient for screening exploration opportunities because it cannot account for variations in the depth domain, which ties to drilling costs and overall geothermal project economics. Instead, the Power per Drill Length (PDL) metric described in this study combines the power output of a geothermal plant with the borehole lengths of its wells. Models built to predict this metric could be used to derive a value that, when combined with a target reservoir depth, describes the power potential for a new opportunity. Furthermore, PDL offers a value over cost proxy, making it an interesting candidate metric for prospect ranking when managing a portfolio of geothermal opportunities to execute. Future work will focus on adding more features to the data set, improving the machine learning models, and using probabilistic methods to provide confidence intervals to better reflect the uncertainty related to each prediction.

REFERENCES

Amante, C., & Eakins, B. W. (2009). *ETOPO1 arc-minute global relief model: Procedures, data sources and analysis*.

Bertani, R. (2005). World geothermal power generation in the period 2001–2005. *Geothermics*, *34*(6), 651–690. https://doi.org/10.1016/j.geothermics.2005.09.005

Byers, L., Friedrich, J., Hennig, R., Kressig, A., Li, X., McCormick, C., & Valeri, L. M. (2018). A global database of power plants. *World Resources Institute*, *18*.

Ciriaco, A. E., Zarrouk, S. J., & Zakeri, G. (2020). Geothermal resource and reserve assessment methodology: Overview, analysis and future directions. *Renewable and Sustainable Energy Reviews*, *119*, 109515.

Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315.

Grant, M. A. (2000). Geothermal Resource Proving Criteria. *Proceedings World Geothermal Congress 2000*, 2581–2584.

Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, *7*, 186–190.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Muffler, L. J. (1979). *Assessment of geothermal resources of the United States, 1978* (No. USGS-CIRC-790). Geological Survey, Reston, VA (USA). Geologic Div.

Robins, J. C., Kesseli, D., Witter, E., & Rhodes, G. (2022). *2022 GETEM Geothermal Drilling Cost Curve Update*. National Renewable Energy Lab.(NREL), Golden, CO (United States).

Rubin, R., Kolker, A., Witter, E., & Levine, A. (2022). *GeoRePORT Protocol Volume VI: Resource Size Assessment Tool* (No. NREL/TP-5700-81820). National Renewable Energy Lab.(NREL), Golden, CO (United States).

Rystad Energy. (2023). *GeothermalCube* [dataset]. https://clients.rystadenergy.com/old-corporate-web/energy-themes/supply-chain/geothermal/geothermal-cube/

Sanyal, S. K., & Sarmiento, Z. (2005). Booking geothermal energy reserves. *Geothermal Resources Council Transactions*, *29*, 467–474.

Shevenell, L. (2015). *Cascades/Aleutian Play Fairway Analysis: Data and Map Files* [dataset]. Geothermal Data Repository. https://doi.org/10.15121/1254444

Wilmarth, M., & Stimac, J. (2014). Worldwide power density review. *Proceedings, Thirty-Ninth Workshop on Geothermal Reservoir Engineering, Stanford University*, 24–26.

Wilmarth, M., & Stimac, J. (2015). Power Density in Geothermal Fields. *Proceedings World Geothermal Congress 2015*, 1–7.

Wilmarth, M., Stimac, J., & Ganefianto, G. (2021). Power Density in Geothermal Fields, 2020 Update. *Proceedings World Geothermal Congress 2020+1*, 1–8.