Proceedings World Geothermal Congress 2026
Calgary, Canada, June 8 – 11, 2026

Refining the INGENIOUS Project Regional Play Fairway Approach: Evaluating Negative Training Site Strategies and Logistic Regression Model Performance in the Great Basin Region, Western USA

Nicole Hart-Wagoner [1]*, Scott Mello [2], Whitney Trainor-Guitton [2], Mark Coolbaugh [1], Cary Lindsey [1], and James Faulds [1]

Great Basin Center for Geothermal Energy, Nevada Bureau of Mines and Geology, University of Nevada [1]*, National Renewable Energy Laboratory [2]

nwagoner@unr.edu

Keywords: Play fairway analysis, logistic regression, probability, prior correction, residual analysis, model error, negative training site, INGENIOUS, Great Basin

ABSTRACT

The INnovative Geothermal Exploration through Novel Investigations Of Undiscovered Systems (INGENIOUS) project aims to facilitate the discovery of new, economically viable hidden geothermal systems in the Great Basin region of the western United States by building on previous play fairway analysis and machine learning work. A key objective of this project is to develop an exploration workflow to reduce exploration risks, particularly for hidden geothermal systems. The workflow was based on the statistical assessment and integration of regional geological, geophysical, and geochemical datasets. This led to the generation of new predictive geothermal fairway maps for the INGENIOUS study area, which encompasses most of Nevada, western Utah, southern Idaho, southeastern Oregon, and easternmost California. The work presented here updates the preliminary INGENIOUS workflow by testing multiple versions of negative training sites and by implementing a systematic evaluation of logistic regression model performance. To more directly compare the resulting models, a prior correction method was applied to adjust model intercepts and align predicted probabilities across training set versions, with differing class balances, with a more realistic prevalence of geothermal systems. To assess logistic regression modeling error, a residual-analysis framework enabling bin-based comparison of predicted and empirical probabilities was applied. Results show that all training set versions have residuals centered near zero and are stable across the well-supported Z range, with only one version displaying a small but systematic calibration trend. These findings demonstrate that geothermal favorability models generally behave consistently and predictably across log-odds space and also suggests that negative-site definitions should be informed by geological context rather than being randomly selected.

1 INTRODUCTION

The INnovative Geothermal Exploration through Novel Investigations Of Undiscovered Systems (INGENIOUS) project is an effort aimed at developing models and workflows that can be implemented to improve the discovery success rate of hidden geothermal resources across the Great Basin region of the western United States (Figure 1). The project integrates regional geological, geophysical, and geochemical datasets to advance play fairway analysis (PFA) methods and develop tools for geothermal exploration. A central objective of INGENIOUS is to generate updated geothermal favorability maps that identify where conditions most strongly support the presence of hidden geothermal systems, thereby reducing early-stage exploration uncertainty and guiding future field investigations.

Previous work established a preliminary regional PFA workflow for the Great Basin region based on statistical relationships between regional datasets and known geothermal systems (Hart-Wagoner et al., 2024). In that study, datasets were evaluated using weights-of-evidence, logistic regression, and feature engineering to identify statistically significant predictors of known geothermal systems. These components were combined within a PFA architecture (Figure 2) to produce new preliminary geothermal favorability maps that improve upon previous models by incorporating higher-resolution datasets and balancing data-driven methods with expert knowledge.

Building on that foundation, this study advances the Great Basin region PFA of Hart-Wagoner et al. (2024) by testing negative training sites as well as developing a workflow to systematically evaluate the predictive performance of logistic regression based geothermal favorability models. While the initial analysis did not include true negative sites, curated negatives have been employed in other studies in the region, such as the

Nevada machine learning project (Smith et al., 2023) and in models using randomly selected negatives (e.g., Caraccioli et al., 2023; Mordensky et al., 2023). Although several geothermal resource assessments have incorporated logistic regression (e.g., Coolbaugh et al., 2003; Williams et al., 2008; Faulds et al., 2021; Mordensky et al., 2023), some employed it primarily to guide the relative weighting of input layers rather than as the core predictive model, and most did not evaluate model error or uncertainty. The preliminary INGENIOUS workflow relied directly on the logistic regression outputs; therefore, a new approach was required to evaluate model error and uncertainty. To address this need, this study adapts the uncertainty quantification framework of Zhan et al. (2023), originally developed for logistic regression-based geospatial hazard modeling. This updated analysis herein focuses on understanding model sensitivity, characterizing prediction errors, and evaluating the patterns of geothermal favorability across the Great Basin region. Specifically, this study isolates the influence of negative site definitions and introduces a bin-based residual framework to quantify model calibration in log-odds space.

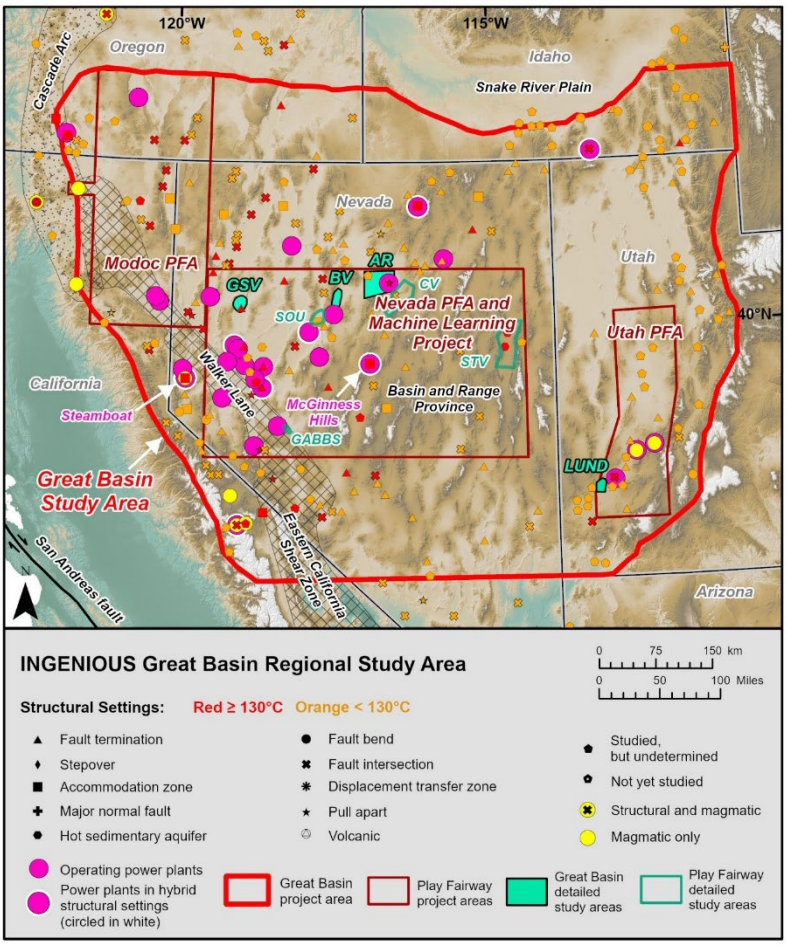


Figure 1: Regional setting of Great Basin study area for the INGENIOUS project, from Faulds and Richards (2023).

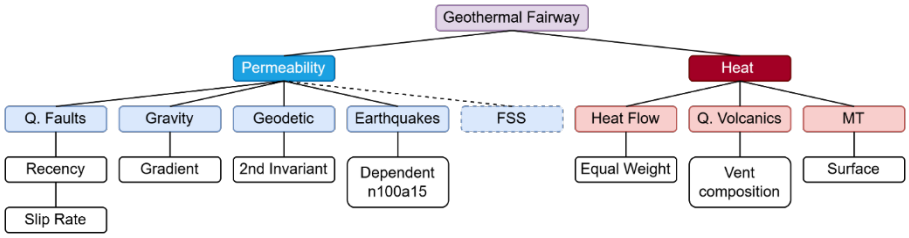


Figure 2: Preliminary play fairway workflow from Hart-Wagoner et al. (2024). Blue boxes are parameters for estimating permeability. The dashed box and line for favorable structural settings (FSS) are not included in the analysis at this time. Red boxes are parameters for estimating heat. Permeability and heat parameters are combined using logistic regression to produce the regional preliminary geothermal play fairway map.

2 DATA AND METHODS

2.1 INGENIOUS Data

In Phase I of the INGENIOUS project, a regional-scale geoscience compilation of 14 datasets was completed for the 494,269 km^2 study area (Ayling et al., 2022). These included: 1) location of Quaternary faults, 2) slip rates on Quaternary faults, 3) age or recency of faulting, 4) slip-dilation tendency (TSTD) on Quaternary faults, 5) active and paleo-geothermal features, 6) Quaternary volcanic distribution, 7) gravity data and models, 8) magnetic data and models, 9) magnetotelluric (MT) data and models, 10) geodetic strain rate, 11) earthquake distribution, 12) regional heat flow/temperatures, 13) temperature-geochemical data from wells and springs, and 14) two-meter temperature data. All datasets are publicly available through the Geothermal Data Repository (https://gdr.openei.org/submissions/1391). The key predictive datasets identified for logistic regression modeling by Hart-Wagoner et al. (2024) include the Quaternary fault super-feature (recency and slip rate), horizontal gravity gradient, second invariant of strain rate, dependent earthquake rate density model (N = 100, α = 0.15), conductive heat flow model (equal-weight), Quaternary volcanic vent composition super-feature (felsic and mafic vents), and near-surface (2-12 km) conductance depth slice.

2.2 Training Sites

A total of 109 known geothermal systems with measured or estimated temperatures >120˚C in the INGENIOUS study area were identified as positive training sites for this analysis (Figure 3A). These systems are either producing geothermal systems (i.e., operating geothermal power plants), undeveloped geothermal systems (proven but without power plants), or geothermal systems defined from measured well temperatures or temperatures estimated using geothermometry. A data value from each of the datasets listed above was extracted at each of the 109 known geothermal systems. These data were then utilized for logistic regression.

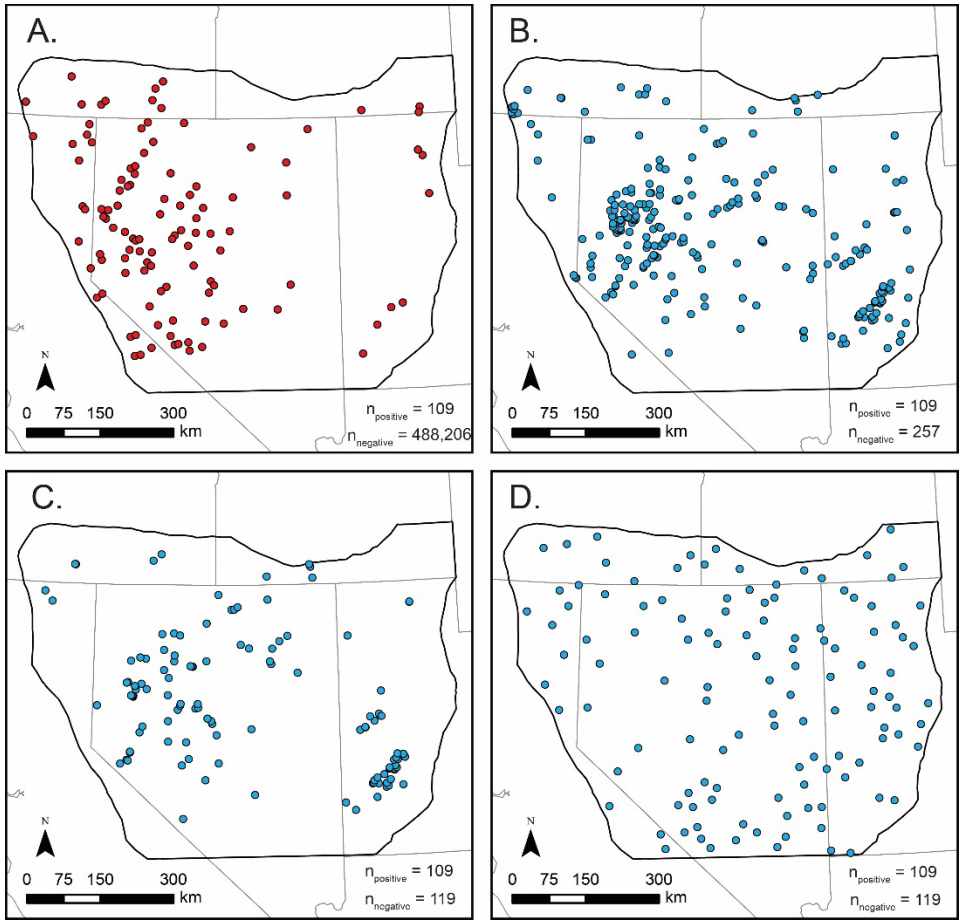


Figure 3. Training sites used in this study. A) Known geothermal systems (positive). B) Heat flow residual negative sites. C) Temperature-filtered heat flow residuals negative sites. D) Spatially random negatives placed at least 10 km from any existing training point. Note: The unlabeled negative sites are not shown due to their continuous 1 km grid spacing across most of the study area.

Four sets of negative sites were used in this logistic regression analysis to assess their impact on geothermal favorability modeling. The first set is the base case that was utilized to develop the preliminary regional

INGENIOUS fairway models (e.g., Hart-Wagoner et al., 2024). In this version, all of the unlabeled 1 km grid cells across the study area were considered negative sites and are referred to in this paper as unlabeled negatives. These unlabeled negatives (n = 488,206) have a severe class imbalance with the positive training sites (n= 109). The remaining three negative training site sets are much closer to being evenly balanced (i.e., similar numbers of positive and negative samples in the training dataset). The second version used a set of negatives defined by Trainor-Guitton et al. (2025), based on heat flow residuals from regional heat flow models (Figure 3B). See Trainor-Guitton et al. (2025) for a detailed description of how these were derived. These are referred to in this paper as heat flow residual negatives (n = 257). The third version was a subset of the second set, further constrained to include only locations with negative temperature residuals (Figure 3C) following Hart-Wagoner et al. (2025). These are referred to in this paper as temperature-filtered heat flow residual negatives (n = 119). The fourth approach generated a set of random points that were not within 10 km of any previously defined positive or negative site (Figure 3D), providing an independent, spatially distributed set of negatives for model evaluation. These are referred to as the spatially random negatives (n = 119) and provide a baseline comparison against the geologically informed negative definitions. Data values from the key predictive INGENIOUS datasets were extracted at each negative site and used in the logistic regression analyses. These four negative datasets enabled systematic evaluation of how different definitions of absent geothermal activity influence model predictive patterns.

2.3 Methods

2.3.1 INGENIOUS PFA Methods

The methods applied in this study follow the stacked logistic regression framework of Hart-Wagoner et al. (2024). The preliminary INGENIOUS workflow separates predictor layers into permeability and heat components. Key permeability datasets include the Quaternary fault super-feature, horizontal gravity gradient, second invariant of strain, and dependent earthquake rate model, which together quantify fault-related permeability, subsurface structural contrasts, regional strain, and fault activity. Key heat predictors include the conductive heat flow model, Quaternary volcanic vent super-feature, and the 2–12 km conductance depth slice, representing regional conductive heat, volcanically driven heat, and crustal thermal properties. Separate regional permeability and heat models were generated using logistic regression to combine the datasets listed above. The permeability and heat component models were then combined, again using logistic regression, to produce a preliminary regional geothermal fairway model.

Logistic regression, a modified form of linear regression, predicts binary outcomes (presence or absence of geothermal systems) as probabilities and can integrate multiple predictor layers (x_1, x_2, …, x_k) without assuming full independence (Wright, 1996). This makes it well-suited for combining geoscience datasets that are partially correlated, as commonly occurs in geothermal exploration (e.g., Coolbaugh et al., 2003).

Data from the known geothermal system dataset were combined with each negative dataset, and the resulting combined datasets were split into training and testing sets using an 80/20 split. The training data were standardized using a standard scaler, which was then applied to the testing data. Logistic regression was trained on the standardized training data.

The logistic regression model produces a linear predictor,

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (1)$$

where β_0 is the intercept, x_1 to x_k are the predictor datasets, β_1 to β_k are the fitted coefficients describing the contribution of each predictor, and Z represents the log-odds of a geothermal system present at a given pixel. The predicted probability (P) of system presence is then obtained by transforming Z through the logistic function:

$$P = \frac{1}{1 + e^{-Z}} \quad (2)$$

2.3.2 Prior correction for rare-event imbalance

Geothermal systems are considered statistically rare, meaning that the true fraction of positives in the population is expected to be very small. However, in model training, the datasets were intentionally balanced (similar numbers of positive and negative sites) in the heat flow residual, temperature-filtered heat flow residual, and spatially random negative site sets to allow the model to better learn the positive class. This balancing artificially inflates predicted probabilities, because the logistic regression intercept (β_0) reflects the prevalence in the training sample, not the true real-world prevalence.

To address this, we applied the prior correction method of King and Zeng (2001), which adjusts the intercept so that model probabilities approximate the estimated true real-world prevalence. The corrected intercept (β_C) is:

$$\beta_C = \beta_0 - \ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right] \quad (3)$$

where τ is the estimation of the true population prevalence of geothermal systems, $\bar{y}$ is the fraction of positives in the training sample, and β_0 is the intercept from the fitted logistic regression model.

To estimate τ, the study area was expressed in units proportional to the area of geothermal systems. The data were sampled on a 1 km^2 grid, and an average Great Basin geothermal system footprint was assumed to be 9 km^2 (≈ 9 pixels). Given the number of pixels in the study area and 109 known systems, this yields $\tau \approx 0.002$. This approximation is conservative, since the true number of geothermal systems in the Great Basin exceeds 109. The intent of this adjustment is to place all models on a comparable probabilistic basis, rather than to determine an exact real-world prevalence. Applying Equation 3 yields the corrected logistic predictor that reflects more realistic population-level probabilities:

$$Z_C = \beta_C + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (4)$$

where Z_C is the corrected log-odds, which can be converted to corrected probabilities using Equation 2. This adjustment places predicted probabilities on a scale that approximates estimated geothermal prevalence and enables direct comparison across models trained with different class balances.

2.3.3 Quantifying Logistic Regression Modeling Error

Zhan et al. (2023) proposed a unified uncertainty framework for geospatial logistic regression models that separates uncertainty into (1) parameter uncertainty, (2) modeling error, and (3) input data uncertainty.

For this study, only the modeling error, quantified through residual analysis, was readily applicable to the INGENIOUS PFA models. Additionally, the fairway model integrates the heat and permeability models that were derived from prior logistic regressions, the reported modeling error reflects the final integration model conditional on those indices and does not include propagated uncertainty from the intermediate stages.

Following the approach of Zhan et al. (2023), model-predicted probabilities *P* are transformed into the log-odds (Z) to evaluate model calibration as a function of the linear predictor.

$$Z = \log\left(\frac{P}{1-P}\right) \quad (5)$$

Samples are then grouped into uniform bins along Z. For each bin, the empirical probability is

$$P_{empirical_i} = \frac{N_{positive}(Z_{predicted_i})}{N_{total}(Z_{predicted_i})} \quad (6)$$

where $N_{positive_i}$ is the number of known geothermal sites whose predicted Z values fall within that bin *i*, and N_{total_i} is the total number of pixels in that bin. To ensure that empirical probabilities are comparable to the prior-corrected probability scale, we adjust the positive count to reflect the difference between the resolution of the full prediction grid and the resolution of the training data. This produces an "effective" $N_{positive_i}$:

- Unlabeled negatives: multiplier = 9
- Heat flow residual, temperature-filtered heat flow residual, and spatially random negatives: multiplier = 144

The multipliers reflect how many prediction-grid pixels correspond to the assumed 9 km^2 geothermal system footprint at the resolution of each model (1 km^2 pixels for the unlabeled negative version; 0.0625 km^2 pixels for the other versions). Scaling the positive counts by these factors corrects for the mismatch in pixel density so that empirical bin probabilities are computed on the same basis as the prior corrected model predictions.

The model-predicted bin probability $P_{predicted_i}$ is the mean predicted probability within that same bin. Per-bin modeling error (calibration) is quantified as the difference between empirical and predicted probabilities (residual) within each Z bin:

$$\epsilon(Z_i) = P_{empirical_i} - P_{predicted_i} \quad (7)$$

To ensure that residuals represent true model behavior rather than sampling artifacts, we enforced thresholds of a minimum number of samples and known geothermal sites for each bin. Sparsely populated bins typically occur near the edges (lowest and highest Z values). These were excluded when their residuals deviated sharply from the well-supported interior bins, consistent with the methods described in Zhan et al. (2023).

3 RESULTS

3.1 Logistic Regression Analysis

The resulting geothermal fairway maps, produced by combining the permeability and heat maps through logistic regression, are shown in Figure 4. Overall patterns of elevated favorability are broadly similar across all versions, but the extent and continuity of predicted favorable areas vary. The fairway from the unlabeled negatives (Figure 4A) exhibits the most widespread high-favorability zones with minimal low-favorability regions. The version from the heat flow residual negatives (Figure 4B) generates a more spatially variable pattern and larger areas of low favorability, particularly in the Nevada-Utah border region and southwest Idaho. The version from the temperature-filtered heat flow residual negatives (Figure 4C) displays a similar pattern but with a more even distribution of predicted favorability across the domain, including broader moderate-favorability zones relative to version from the heat flow residual negatives. The map from the version utilizing the spatially random negatives (Figure 4D) is characterized by more continuous regions of moderate-to-high predicted favorability, particularly in western and central Nevada, with fewer but sharply defined low-favorability areas.

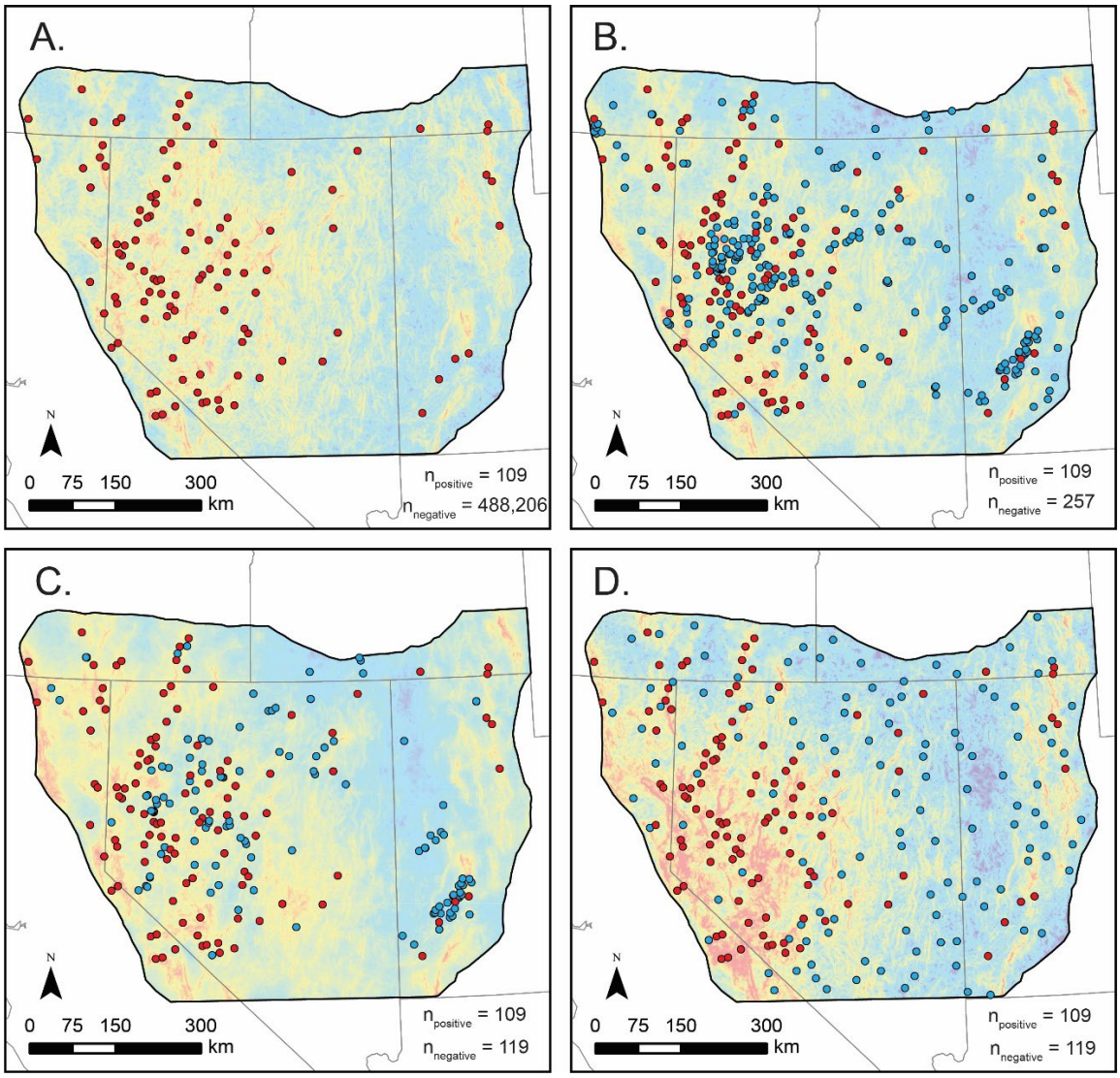


Figure 4. Geothermal fairway model results using different negative sites versions: A) Unlabeled negatives (1 km grid). B) Heat flow residual negatives. C) Temperature-filtered heat flow residual negatives. D) Spatially random negatives. All maps are on the same favorability scale, where warmer colors indicate higher favorability and cooler colors indicate lower favorability. Red points are positive sites (known geothermal systems) and blue points represent different versions of negative sites.

3.2 Residual Analysis

The results of the residual analysis outlined in Section 2.3.2 are summarized in Figure 5, with each plot illustrating model calibration within the Z range, where bins are well supported. Edge bins with low positive counts produced large residual outliers in some cases, which were clipped automatically. Table 2 summarizes the mean model probabilities, class balance, number of bins retained, number dropped, and residual statistics for each model component.

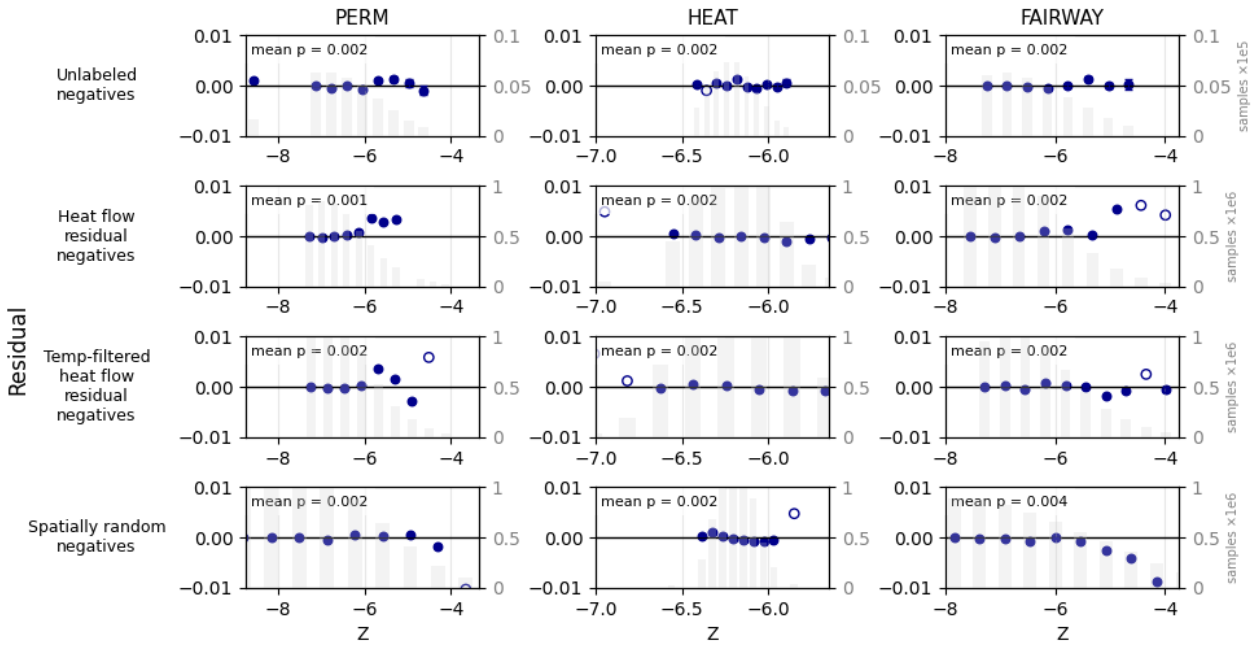


Figure 5. Residuals as a function of Z for all model versions and components (permeability, heat, and fairway). Rows correspond to logistic regression models trained using four different definitions of negative sites. Blue markers with standard error bars show retained bins; empty circles show dropped bins identified as sampling-artifact outliers. Per-bin sample counts are shown as background gray bars on a secondary axis (right).

Table 2. Residual calculation summary statistics for each model version and component. Mean Prob. refers to the mean predicted probability across all prediction-grid pixels.

Version	Training # positives : # negative sites	Component	Mean Prob.*	Max Prob.	N Bins	N Clipped	Std. Residual
Unlabeled negatives	109 : 488,206	Permeability	0.0019	0.0227	9	0	0.00082
		Heat	0.0020	0.0048	10	1	0.00049
		Fairway	0.0019	0.026	8	0	0.00058
Heat flow residual negatives	109 : 247	Permeability	0.0015	0.0124	11	2	0.0040
		Heat	0.0021	0.0087	9	1	0.00044
		Fairway	0.0016	0.0228	9	2	0.0020
Temperature-filtered heat flow residual negatives	109 : 119	Permeability	0.0017	0.106	9	2	0.0020
		Heat	0.0023	0.0227	8	2	0.00053
		Fairway	0.0020	0.227	10	1	0.00077
Spatially random negatives	109 : 119	Permeability	0.0023	0.667	10	1	0.00075
		Heat	0.00205	0.0049	11	3	0.00060
		Fairway	0.0043	0.760	13	2	0.0068

Residuals quantify how much the model over- or underpredicts within each range of predicted favorability. Across the unlabeled, heat flow residual, temperature-filtered heat flow residual negative versions and most components of spatially random negative version, residuals fluctuate around zero without systematic bias across the well-supported Z range. Residual magnitudes are generally on the order of 10^{-3} to 10^{-2} in probability units and remain centered near zero across bins. The spatially random negative fairway is the only case that exhibits a clear directional trend: a gradual increase in overprediction with increasing Z. Even in this component, however, absolute residuals remain small, and the trend appears smooth and well-resolved rather than noisy. Consistent with this, Table 2 shows that the standard deviations for all versions and components fall within narrow ranges, with only the fairway of the spatially random negative site version reaching higher values 0.0068. These results indicate that the models exhibit stable calibration behavior across the log-odds space, with only one component displaying a systematic diverging pattern. While residual magnitudes can be comparable to predicted probabilities in low-Z bins due to rare-event scaling, their proportional impact decreases at higher predicted probabilities.

4 DISCUSSION

Geothermal systems are statistically rare, making them inherently difficult to model accurately. There is a severe class imbalance between known geothermal systems (positive sites) and the rest of the study area, which has historically been treated as negative sites, leading to underestimated probabilities. Conversely, balancing positives and negatives overrepresents the positive sites, resulting in severely overestimated probabilities. Implementing a prior correction is therefore essential, both to adjust predicted probabilities to an estimated realistic prevalence and to enable fair comparison across model versions. By explicitly accounting for estimated real-world prevalence, this workflow moves beyond the relative-probability outputs typical of logistic regression PFAs, paving the way for more interpretable and comparable probability maps of geothermal potential.

Across all the negative site versions, the geothermal favorability patterns show broad regional agreement, indicating that the core predictor datasets, faulting, strain, gravity gradients, earthquake rates, heat flow, volcanism, and crustal conductance, consistently identify similar areas of elevated geothermal potential. However, differences among negative site versions highlight how negative training sites influence the spatial variability of favorability. The differences among these versions suggest that negative site selection is a nontrivial component of geothermal favorability modeling and can influence the identification of both promising and unfavorable regions.

The residual analysis provides an independent evaluation of model calibration and error. Residuals remain centered near zero across all models and components within the well-supported Z-range, with magnitudes that are modest in absolute probability units. This behavior indicates that the modified residual-analysis procedure successfully stabilizes the empirical probabilities and avoids sampling artifacts. As a result, the residuals primarily reveal whether each model maintains consistent calibration across Z, rather than noise induced by extreme class imbalance. Additionally, the magnitudes of the residuals and associated standard deviations overall demonstrate that the models are well calibrated in the central Z-range and that the adapted residual framework provides a stable measure of modeling error. While residual magnitudes are sometimes comparable to typical predicted probabilities in low-Z bins due to rare-event prior correction, they do not exhibit systematic divergence across Z.

The one notable exception is the fairway model trained using spatially random negative sites, which exhibits a gradual trend toward overprediction at higher Z values. Although the absolute magnitude of this bias is small, its presence suggests that randomly generated negatives, despite being spatially independent, do not appropriately represent the true absence of geothermal systems in all geologic contexts. This underscores the importance of negative site definitions that incorporate geological reasoning rather than relying solely on a random spatial distribution.

5 CONCLUSIONS AND NEXT STEPS

This study expands the INGENIOUS project play fairway workflow by testing multiple definitions of negative training sites and implementing a formal residual-analysis framework to evaluate logistic regression model performance. The residual analysis confirms that (1) residual magnitudes are modest in absolute probability units and centered near zero across models, (2) most components from different versions of negative sites display stable, near-zero residual patterns with no meaningful bias, and (3) the fairway of the spatially random negative site version is the only component showing a noticeable calibration trend (increasing overprediction with Z), though the absolute residuals remain modest in probability units. Together, these updates provide new

insight into both the sensitivity of geothermal favorability mapping to training data design and the stability of the resulting model predictions across the Great Basin region.

The uncertainty and error estimations of the INGENIOUS PFA workflow could be further characterized by extending the uncertainty framework beyond modeling error by incorporating the remaining components of Zhan et al. (2023): parameter uncertainty and input-data uncertainty. Quantifying parameter uncertainty would allow us to evaluate how variability in the fitted logistic regression coefficients influences geothermal favorability predictions, while characterization of input-data uncertainty would help assess noise, resolution differences, and measurement error within the input datasets. Together, these components would enable a full model sensitivity analysis that partitions the relative contribution of each uncertainty source to overall output variability. This, in turn, would guide forward uncertainty propagation and highlight which datasets, predictors, or modeling assumptions most strongly control prediction stability across the Great Basin region. Integrating these additional uncertainty pathways would further improve characterization of model reliability, error, and uncertainty, and help guide decision-making in exploration targeting.

Acknowledgements

This project is funded by U.S. Department of Energy - Geothermal Technologies Office under award DE-EE0009254 to the University of Nevada, Reno. Thank you to the entire INGENIOUS team past and present (Great Basin Center for Geothermal Energy, Nevada Bureau of Mines and Geology, University of Nevada Reno, Aprovechar Lab L3C, Cyrq Energy, Inc., Hi-Q Geophysical, Inc., Idaho Geological Survey, Innovate Geothermal Ltd, Lawrence Berkeley National Laboratory, National Laboratory of the Rockies – formerly National Renewable Energy Laboratory, RESPEC – formerly Geothermal Resource Group, Teverra LLC – formerly Petrolern Ltd., U.S. Geological Survey, and the Utah Geological Survey) for their contributions to the INGENIOUS project, which supported this work. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This work was authored in part by the National Laboratory of the Rockies for the U.S. Department of Energy (DOE), operated under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

REFERENCES

Ayling, B., Faulds, J., Morales Rivera, A., Koehler, R., Kreemer, C., Mlawsky, E., Coolbaugh, M., Micander, R., dePolo, C., Kraal, K., Wagoner, N., Siler, D., DeAngelo, J., Glen, J., Peacock, J., Batir, J., Gentry, E., Berti, C., Lifton, Z., Clark, A., Kirby, S., Hardwick, C., and Kleber, E.: INGENIOUS - Great Basin Regional Dataset Compilation. GBCGE, NBMG, UNR, Distributed by Geothermal Data Repository, (2022), https://doi.org/10.15121/1881483.

Caraccioli, P.D., Mordensky, S.P., Lindsey, C.R., DeAngelo, J., Burns, E., and Lipor, J.: Don't Let Negatives Hold You Back: Accounting for Underlying Physics and Natural Distributions of Hydrothermal Systems When Selecting Negative Training Sites Leads to Better Machine Learning Predictions. GRC Transactions, Vol 47, (2023), 1672-1693.

Coolbaugh, M.F.; The Prediction and Detection of Geothermal Systems at Regional and Local Scales in Nevada using a Geographic Information System, Spatial Statistics, and Thermal Infrared Imagery, Ph.D. dissertation, University of Nevada, Reno, USA, (2003), 172 p.

Faulds, J. E., Hinz, N. H., Coolbaugh, M., Ayling, B., Glen, J., Craig, J., McConville, E., Siler, D., Queen, J., Witter, J., and Hardwick, C.: Discovering Blind Geothermal Systems in the Great Basin Region: An Integrated Geologic and Geophysical Approach for Establishing Geothermal Play Fairways: All Phases, (2021), https://doi.org/10.2172/1724080.

Hart-Wagoner, N., Coolbaugh, M., Ayling, B., Faulds, J.E., Lindsey, C., Mlawsky, E., and Lu, C.: Enhancing the Known Geothermal Systems Database Using INGENIOUS Project Data, GRC Transactions, Vol 49, (2025), 1558-1573.

Hart-Wagoner, N., Coolbaugh, M., Faulds, J.E., Mlawsky, E., Lindsey, C., Trainor-Guitton, W., and Brown, S.: Preliminary Regional Play Fairway Workflow for the Great Basin Region, USA, Proceedings, 49th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, USA, (2024), p. 12.

King, G., and Zeng, L.: Logistic regression in rare events data, Political Analysis 9, no. 2, (2001), 137-163.

Mordensky, S. P., Lipor, J.J., DeAngelo, J., Burns, E.R., and Lindsey, C.R.: When less is more: How increasing the complexity of machine learning strategies for geothermal energy assessments may not lead toward better estimates, Geothermics, 110, (2023): 102662.

Trainor-Guitton, W., Menon, K, Pinchuk, P., Min Thomson, S., Hart-Wagoner, N., Lu, C., Mlawsky, E., Coolbaugh, M., Lindsey, C., and Faulds, J.: "Hidden" hydrothermal technical potential & technoeconomics: Revealing permeability & fluids with more data, Geothermics, Volume 133, (2025), 103473, ISSN 0375-6505, https://doi.org/10.1016/j.geothermics.2025.103473.

Williams, C.F., and DeAngelo, J.: Mapping Geothermal Potential in the Western United States, GRC Transactions, V32, (2008), 181–188.

Wright, D.F.: Evaluating volcanic-hosted massive sulphide favourability using GIS-based spatial data integration models, Snow Lake area, Manitoba, Ph.D. Dissertation, University of Ottawa, Ottawa, Canada, (1996), 376 p.

Zhan, W., Baise, L.G., and Moaveni, B.: An uncertainty quantification framework for logistic regression based geospatial natural hazard modeling, Engineering Geology, 324, (2023), 107271.