

Exploration and Production Risk Mitigation for Geothermal Adoption in the Energy Transition

by

Robert Chadwick Holmes
B.S., Duke University (2000)
M.A., Columbia University (2004)
M.Ph., Columbia University (2006)
Ph.D., Columbia University (2009)

Submitted to the System Design & Management Program
in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

©2021 Robert Chadwick Holmes. All rights reserved.

The author hereby grants to MIT permission to reproduce and to
distribute publicly paper and electronic copies of this thesis document
in whole or in part in any medium now known or hereafter created.

Author
System Design & Management Program
August 6, 2021

Certified by
Aimé Fournier
Research Scientist, Earth and Planetary Sciences
Thesis Supervisor

Certified by
Bryan R. Moser
Academic Director, System Design & Management Program
Thesis Supervisor

Accepted by
Joan Rubin
Executive Director, System Design & Management Program

Exploration and Production Risk Mitigation for Geothermal Adoption in the Energy Transition

by

Robert Chadwick Holmes

Submitted to the System Design & Management Program
on August 6, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science in Engineering and Management

Abstract

Geothermal provides a continuous, low-emissions source of energy with enormous potential in the United States, both singularly or as part of a broader energy mix. Although a small contributor to the current national energy grid, geothermal electricity generation dates back nearly a century for natural hydrothermal systems. More recently, enhanced geothermal systems (EGS) promise a broader reach with engineered solutions for extracting subsurface heat from a wider variety of locations. The potential synergy between the oil & gas and geothermal offers an opportunity for building a lower-carbon energy portfolio that requires compatible skills and expertise. Nevertheless, the risks involved at multiple stages of the field lifecycle remain a hurdle to adoption of geothermal.

In this thesis, risk-mitigation strategies for geothermal target two phases of the lifecycle: exploration and production. The first strategy uses a diverse set of measurements spanning multiple interrelated earth systems to collectively determine geothermal potential at the play scale. Analytical workflows integrate geologic, geochemical, and geophysical data to estimate subsurface geothermal gradient, with quantitative uncertainty estimates associated with the measurements, the models, and the solution space. These uncertainty estimates provide a measure of risk, as well as decision tools for investments in additional data-gathering activities before the first well is drilled. The second strategy applies flexibility in engineering design to a hypothetical EGS expansion of an existing power facility. Specifically, key uncertainties are integrated into a cost model with operational decision rules to create an ensemble of possible outcomes. Tailoring the model and decision rules to a particular facility concept allows for a rapid feasibility testing and optimization of project actions that limit downside risk while capturing upside potential. Both of these strategies use uncertainty characterization to reduce the threat of high-consequence geothermal risks. And by including them in a broader risk management approach, oil & gas companies can make data-driven decisions on investing in geothermal during the energy transition.

Thesis Supervisor: Aimé Fournier

Title: Research Scientist, Earth and Planetary Sciences

Thesis Supervisor: Bryan R. Moser

Title: Academic Director, System Design & Management Program

Acknowledgments

The past year will be remembered for the impact a global pandemic had on society at large. This thesis is a product of that time. Those mentioned below played a significant role in keeping things on track in spite of the many months of mask-wearing, virus-testing, quarantining, remote classes and conversations, and eventual vaccination. Some I have even met in person, although not all. That is part of the legacy of this most unusual and memorable year.

First and foremost, I want to thank my advisor, Aimé Fournier for his good humor, guidance, and willingness to advise under remote conditions. Aimé showed early interest in the thesis before a proposal was even drafted, and his perspective and ideas helped shape what it became. I greatly appreciate his time, input, and willingness to take a chance on a stranger with only one year to produce results.

I wish to thank the System Design & Management program for an intense and incredible educational experience. Remote learning is nothing new to the program, and it was obvious from the quality of instruction they provided. Special thanks to my thesis reader and instructor, Bryan Moser, whose passion for learning and teaching is contagious. Bryan once commented that “research is a lifestyle,” which rang true in his lectures for the SDM foundations courses, his agent-based modeling course, his Global Teamwork Lab meetings, his IEEE World Forum session, and basically any other time I saw him in action. Thanks also to Joan Rubin for leading the program and supporting each of us throughout the year. On a personal level, Joan kept an open door to communication by email, by phone, and eventually a handful of in-person meetings in support of my thesis journey.

A separate and heartfelt thank you goes to Elizabeth Baker for taking the giant leap from being my teaching assistant in a class of eighty-some students to becoming a very dear friend. Our time spent both online and in-person together made the highs and lows of thesis writing so worth it.

Thanks also to the MIT Energy Initiative for welcoming me among your ranks. Thanks especially to Diane Rigos for the regular check-ins, kind offers of help, and a

couple of in-person meals earlier this summer that helped bring the Chevron students together for brief but memorable moments in time.

The Earth Resources Lab was one of my first connections to MIT thanks to their brief visit to Houston in early Spring 2020 (pre-Covid). Thanks go to Laurent Demanet, Director of ERL, for responding to my trial balloon emails and connecting me with both Aimé Fournier and Mike Fehler. Mike in turn connected me with Steve Brown, Chen Gu, John Queen, Bill Rodi, Connor Smith, Sven Treitel, and the rest of the Great Basin Machine Learning project team under Jim Faulds. I feel honored to have met and conversed with so many brilliant people whom I now consider friends. Perhaps unsurprisingly, our group conversations were a strong influence on the machine learning investigation in this thesis.

I wish also to thank my company sponsor, Chevron, for taking a chance on me. I specifically want to thank those who believed in my potential and supported my year-long absence to pursue this degree: Sebastien Bombarde, John Moore, and Janet Yun for being my champions; Mason Edwards, Kenn Ehman, Kellen Gunderson, Ash Harris, Fabien Laugier, Rhonda Welch, and Khryste Wright for keeping up with me during the year away; Brendan Horton for being my Digital Scholar mentor; and so many others in the greater Chevron community. On the program side, special thanks go to Shana Bolen and Margery Connor for a year of conversations, encouragement, and a lot of behind-the-scenes efforts that I may not have been fully aware of but certainly appreciate very much.

Being a graduate student at MIT would be daunting in a normal year, but with a pandemic raging and real-life connections to faculty, staff, and fellow students reduced to a laptop screen and webcam, I have the 15 other Chevron scholars to thank for making the year a truly positive and life-changing experience. I feel so honored to have met such a diverse and wonderful group, and I look forward to many years of friendship and collaboration to come: Robert Andrais, Louis Catalan, Gloria Bahl Chambi, Christian Dowell, Matthew Hernandez, Matthew Kieke, Hemant Kumar, Alessandro Lucioli, Elias Machado, Monthep Parimontonsakul, Allison Polly, Kelsey Prestidge, Bagdat Toleubay, John Ward, and my thesis buddy, Surge Yemets.

Last but not least, thank you to my wonderful husband of nearly 8 years, Hans, for putting up with my absence for a full 12 months. You bravely held down the fort on your own, caring for two senior dachshunds and a puppy while balancing a full-time job, commitments to the community, extreme Texas heat, and even the winter storm blackouts that make a cameo in Chapter 1. I dedicate this thesis to you in honor of the impact it had on your life as well as my own.

Contents

List of Figures	17
List of Tables	23
List of Acronyms	25
1 Introduction	29
1.1 Energy Trends	30
1.2 Upstream Commercial Pressures	31
1.3 Net Zero Ambitions	33
1.4 Geothermal Energy	33
1.5 Research Questions	36
1.6 Thesis Outline	37
2 Geothermal Background	39
2.1 Geothermal Systems	39
2.1.1 Origins of Heat	39
2.1.2 Measuring Heat	41
2.1.3 System Fundamentals	43
2.2 Geothermal Exploration	50
2.2.1 Background	50
2.2.2 Regional Exploration	50
2.2.3 Play-Scale Exploration	51
2.2.4 Integration Strategies	54

2.2.5	Uncertainties	58
2.3	Power Generation	59
2.3.1	Direct Steam Power Plant	61
2.3.2	Flash Power Plant	61
2.3.3	Binary-Cycle Power Plant	63
2.4	Geothermal Cost Modeling	64
2.4.1	GEOPHIRES	65
2.4.2	GETEM	66
2.4.3	SAM	67
2.4.4	CREST	67
2.4.5	Cost Model Insights	68
2.5	Case Study: Southwestern New Mexico	68
2.5.1	Southern Basin and Range	69
2.5.2	Rio Grande Rift	69
2.5.3	Colorado Plateau	71
2.5.4	Mogollon-Datil Volcanic Field	71
2.5.5	Lightning Dock	72
2.6	Recap	74
3	Geothermal Exploration with Machine Learning	77
3.1	Data Sources	77
3.2	Data Preparation	80
3.2.1	Fundamentals	80
3.2.2	Data Conditioning	84
3.3	Data Modeling	99
3.3.1	Assessing Performance	100
3.3.2	Hyperparameter Tuning	102
3.3.3	Logistic Regression	103
3.3.4	Decision Trees	105
3.3.5	Tree Ensembles (XGBoost)	108

3.3.6	Neural Networks	111
3.4	Uncertainty Analysis	113
3.4.1	Classification Uncertainty Measures	114
3.4.2	Structural Uncertainty	115
3.4.3	Parameter Uncertainty	115
3.4.4	Measurement Uncertainty	117
3.5	Recap	118
4	Cost Modeling for an	
	EGS Power Plant Expansion	121
4.1	EGS Expansion Concept	121
4.1.1	Lightning Dock EGS	121
4.1.2	New Mexico Electricity Demand	123
4.1.3	Modular Geothermal	124
4.1.4	Flexible Cost Models	125
4.2	Static Cost Model	126
4.2.1	NPV Model	126
4.2.2	Rate Calculations	132
4.2.3	Model Parameters	134
4.3	Probabilistic Cost Model	136
4.3.1	Model Uncertainties	137
4.3.2	Sensitivity Testing	143
4.3.3	Probability Density Functions	145
4.4	Flexibility with Design Options	149
4.4.1	Redevelop Only Case	150
4.4.2	Redevelop & Grow Case	151
4.4.3	Full Flexibility Case	152
4.5	Recap	153
5	Geothermal Exploration Machine-Learning Results	155
5.1	Logistic Regression	155

5.1.1	Hyperparameter Tuning	155
5.1.2	Feature Selection	159
5.1.3	Optimized Model Results	159
5.2	Decision Trees	162
5.2.1	Hyperparameter Tuning	162
5.2.2	Feature Selection	166
5.2.3	Optimized Model Results	166
5.3	Tree Ensembles (XGBoost)	170
5.3.1	Hyperparameter Tuning	170
5.3.2	Feature Selection	174
5.3.3	Optimized Model Results	176
5.4	Neural Networks	178
5.4.1	Network Architecture	178
5.4.2	Hyperparameter Tuning	180
5.4.3	Optimized Model Results	183
5.5	Uncertainty Analysis	186
5.5.1	Structural Uncertainty	186
5.5.2	Parameter Uncertainty	188
5.5.3	Measurement Uncertainty	193
5.6	Comparative Study Insights	200
5.6.1	Southwestern New Mexico PFA	200
5.6.2	Southwestern New Mexico PCA	207
5.7	Recap	209

6 EGS Power Plant Expansion

Cost Model Results	211
6.1 Static Model	211
6.1.1 Model Selection	211
6.1.2 Construction Optimization	213
6.1.3 Summary Statistics	213

6.2	Probabilistic Model Metrics	214
6.3	Base Case Model	214
6.3.1	Model Results	214
6.3.2	Summary Statistics	216
6.4	Redevelop Case Model	217
6.4.1	Model Results	217
6.4.2	Summary Statistics	219
6.5	Redevelop & Grow Case Model	219
6.5.1	Model Results	220
6.5.2	Summary Statistics	222
6.6	Full Flexibility Case Model	222
6.6.1	Model Results	222
6.6.2	Summary Statistics	224
6.7	Combined Model Comparison	225
6.8	Full Flexibility Case Sensitivity Testing	225
6.8.1	Reduction Amount	226
6.8.2	Expansion Amount	226
6.9	Recap	228
7	Discussion	229
7.1	Field Lifecycle	229
7.2	Role of Uncertainty	232
7.3	Risk Analysis	233
7.4	Great Opportunity	238
8	Conclusions	241
9	Future Work	245
9.1	Machine-Learning Applications	245
9.2	Cost Modeling	247
9.3	Related Research Topics	249

A	Exploration Data Layers	251
A.1	Average Air Temperature	251
A.2	Average Precipitation	253
A.3	Basement Depth	254
A.4	Boron Concentration	256
A.5	Crustal Thickness	258
A.6	Drainage Density	260
A.7	Earthquake Density	262
A.8	Gamma-Ray Absorbed Dose Rate	264
A.9	Geodetic Strain Rate	266
A.10	Gravity Anomaly	268
A.11	Gravity-Anomaly Gradient	269
A.12	Heat Flow	270
A.13	Lithium Concentration	272
A.14	Magnetic Anomaly	274
A.15	Magnetic-Anomaly Gradient	275
A.16	Quaternary Fault Density	276
A.17	Silica Geothermometer Temperature	278
A.18	Spring Density	280
A.19	State Map Fault Density	282
A.20	Surface Topography (DEM)	284
A.21	Topographic Gradient	286
A.22	Volcanic-Dike Density	287
A.23	Volcanic-Vent Density	288
A.24	Water-Table Depth	290
A.25	Water-Table Gradient	292
A.26	Geothermal Gradient Points	294
A.27	Geothermal Gradient Layer	296

B Cost Model Spreadsheets	297
B.1 Static NPV Model	297
B.2 Probabilistic NPV Model	299
C Risk Mitigation Log	303
References	308

List of Figures

1-1	U.S. EIA projections based on the AEO2021 reference case	30
1-2	Lazard Levelized Cost of Energy 2021 projections	31
1-3	Thesis structure flow chart	37
2-1	Earth mechanical and compositional structure	40
2-2	Heat flow across the continental U.S.	44
2-3	Terminal Geyser, Lassen National Park	45
2-4	Mud pot, Lassen National Park	45
2-5	Country rankings for installed geothermal capacity	47
2-6	Enhanced Geothermal Systems schematic	48
2-7	Intent to concept for geothermal power plants	60
2-8	Geothermal applications by temperature and depth	60
2-9	Direct steam power plant schematic	62
2-10	Single flash power plant schematic	63
2-11	Binary cycle power plant schematic	64
2-12	Power plants in the United States	65
2-13	Physiographic provinces of SW New Mexico	70
2-14	Lightning Dock power plant location map	73
3-1	Mesh grid point set	82
3-2	Data conditioning workflow	85
3-3	Data imputation strategy	86
3-4	Unconditioned FDS histograms	88
3-5	Unscaled FDS scatter plots	89

3-6	Conditioned FDS histograms	92
3-7	Scaled FDS scatter plots	93
3-8	Feature correlation matrix	95
3-9	Machine-learning modeling workflow	99
3-10	Example four-class confusion matrix	100
3-11	Receiver Operating Characteristic diagram	101
3-12	Decision tree schematic	106
3-13	Neural network schematic	111
3-14	Uncertainty analysis workflow	113
4-1	New Mexico energy consumption	124
4-2	Modular power plant schematic	125
4-3	Electricity price forecast	136
4-4	Carbon tax price impact	137
4-5	Electrification price impact	138
4-6	GeoVision drilling cost curves	141
4-7	Sensitivity testing tornado diagram	144
4-8	Cost model probability distributions	145
4-9	Cost model probabilistic price forecasts	148
4-10	Reservoir temperature PDF	149
5-1	Logistic regression hyperparameter tuning	156
5-2	Logistic regression feature coefficients	158
5-3	Logistic regression feature selection	159
5-4	Logistic regression confusion matrix	160
5-5	Logistic regression ROC curves	161
5-6	Logistic regression prediction map	162
5-7	Decision tree max depth tuning	163
5-8	Decision tree min samples tuning	164
5-9	Decision tree max features tuning	165
5-10	Decision tree alpha tuning	165

5-11	Decision tree feature importances	167
5-12	Decision tree confusion matrix	168
5-13	Decision tree ROC curves	169
5-14	Decision tree prediction map	169
5-15	Decision tree visualization	171
5-16	XGBoost hyperparameter tuning	173
5-17	XGBoost SHAP variable importance	175
5-18	XGBoost confusion matrix	177
5-19	XGBoost ROC curves	177
5-20	XGBoost prediction map	178
5-21	Neural network structural flow	179
5-22	Neural network structural schematic	179
5-23	Neural network hyperparameter tuning	181
5-24	Neural network training loss	182
5-25	Neural network confusion matrix	184
5-26	Neural network ROC curves	184
5-27	Neural network prediction map	185
5-28	Combined machine learning results	187
5-29	Combined model prediction map	188
5-30	Structural uncertainty map	189
5-31	Structural uncertainty mask on prediction map	190
5-32	Bayesian neural network structural flow	191
5-33	Bayesian neural network structural schematic	191
5-34	Bayesian neural network training loss	192
5-35	Bayesian neural network class PDFs	193
5-36	Bayesian neural network parameter uncertainty map	194
5-37	Parameter uncertainty mask on prediction map	195
5-38	SiGT measurement uncertainty map	197
5-39	SiGT measurement uncertainty mask on prediction map	198
5-40	SiGT standard error and model entropy	199

5-41	SW New Mexico PFA prospective areas	201
5-42	Machine learning models with prospective areas	203
5-43	Structural uncertainty map with prospective areas	204
5-44	Parameter uncertainty map with prospective areas	205
5-45	Measurement uncertainty map with prospective areas	206
6-1	Static cost-model comparison	212
6-2	Base Case histogram	215
6-3	Base Case CDF	216
6-4	Redevelop Case histogram	218
6-5	Redevelop Case CDF	218
6-6	Redevelop & Grow Case histogram	220
6-7	Redevelop & Grow Case CDF	221
6-8	Full Flexibility Case histogram	223
6-9	Full Flexibility Case CDF	224
6-10	All Cases CDF	225
6-11	Reduction Amount sensitivity test	227
6-12	Expansion Amount sensitivity test	227
7-1	Geothermal field lifecycle	230
7-2	Geothermal project risk matrix	237
A-1	Average air-temperature data layer	252
A-2	Average precipitation data layer	253
A-3	Basement depth data layer	255
A-4	Boron concentration data layer	257
A-5	Crustal thickness data layer	259
A-6	Drainage density data layer	261
A-7	Earthquake density parameter tuning	262
A-8	Earthquake density data layer	263
A-9	Absorbed dose rate data layer	265

A-10 Geodetic strain rate data layer	267
A-11 Gravity anomaly data layer	268
A-12 Gravity anomaly gradient data layer	269
A-13 Heat flow data layer	271
A-14 Lithium concentration data layer	273
A-15 Magnetic anomaly data layer	274
A-16 Magnetic-anomaly gradient data layer	275
A-17 Quaternary fault density data layer	277
A-18 Silica geothermometer temperature data layer	279
A-19 Spring density parameter tuning	280
A-20 Spring density data layer	281
A-21 State fault density data layer	283
A-22 Surface topography (DEM) data layer	285
A-23 Topographic gradient data layer	286
A-24 Volcanic dike data layer	287
A-25 Volcanic-vent density parameter tuning	288
A-26 Volcanic-vent data layer	289
A-27 Water-table depth data layer	291
A-28 Water-table gradient data layer	293
A-29 Geothermal gradient well data	295
A-30 Geothermal gradient data layer	296
 B-1 Static cost model spreadsheet (part 1)	 298
B-2 Static cost model spreadsheet (part 2)	299
B-3 Probabilistic cost model spreadsheet (part 1)	300
B-4 Probabilistic cost model spreadsheet (part 2)	301

List of Tables

1.1	Annual renewable energy fluxes	34
2.1	Data collection methods for geothermal subsystems	52
3.1	Southwestern New Mexico feature list	79
3.2	Yeo-Johnson transformation lambda values	91
3.3	Geothermal gradient classes	96
3.4	Data set class distribution	97
3.5	Class count for test-train-validate split	99
4.1	Power plant labor costs	132
4.2	Cost model parameters for resource recovery	134
4.3	Cost model parameters for operations	135
4.4	Cost model parameters for economics	135
4.5	Projected regional temperature changes	139
4.6	Lightning Dock well data	142
4.7	Sensitivity testing results	144
5.1	Logistic regression hyperparameter tuning results	156
5.2	Multi-class logistic regression classifier coefficients	157
5.3	Decision tree hyperparameter tuning results	166
5.4	XGBoost hyperparameter tuning results	174
5.5	Neural network hyperparameter tuning results	182
5.6	Bayesian neural network hyperparameters and single-run metrics . . .	192
5.7	Feature standard error estimation	196

6.1	Static model module installation schedule	213
6.2	Probabilistic Base Case statistics	216
6.3	Probabilistic Redevelop Case statistics	219
6.4	Redevelop & Grow Case statistics	222
6.5	Full Flexibility Case statistics	224
7.1	Risk likelihood	233
7.2	Risk consequence	234
7.3	Geothermal risk log	235
7.4	Geothermal mitigation log	236
C.1	Risk-likelihood scores and explanations	304
C.2	Risk consequence scores and explanations	305
C.3	Updated risk likelihood scores and explanations	306
C.4	Updated risk consequence scores and explanations	307

List of Acronyms

AEO Annual Energy Outlook. 30

ANN Artificial Neural Network. 57

AOI Area of Interest. 80

BHT Bottom Hole Temperature. 50

BNN Bayesian Neural Network. 116

CAPEX Capital Expenditures. 126

CDF Cumulative Distribution Function. 214

CP Colorado Plateau. 68

CV Cross Validation. 103

DEM Digital Elevation Model. 53

DOE United States Department of Energy. 46

EBK Empirical Bayes Kriging. 84

EGS Enhanced Geothermal Systems. 47

EIA United States Energy Information Administration. 30

ENPV Expected Value of NPV. 214

FDS Full Data Set. 85

FN False Negative. 100

FP False Positive. 100

FPR False Positive Rate. 100

GIS Geographic Information System. 80

GP Great Plains. 69

GTO Geothermal Technologies Office. 46

KDE Kernel Density Estimation. 81

KGRA Known Geothermal Resource Area. 46

LANL Los Alamos National Laboratory. 48

LCOE Levelized Cost of Electricity. 30

LR Logistic Regression. 103

Ma Million Years Ago. 69

MDVF Mogollon-Datil Volcanic Field. 68

MW·h Megawatt-Hour. 124

NaN Not a Number. 87

NF-EGS Near Field EGS. 49

NGDS National Geothermal Data System. 294

NM New Mexico. 57

NN Neural Network. 111

NPV Net Present Value. 126

NREL National Renewable Energy Laboratory. 49

OPEX Operating Expenses. 126

OvO One-versus-One. 104

OvR One-versus-Rest. 104

PCA Principle Component Analysis. 57

PDF Probability Density Function. 145

PFA Play Fairway Analysis. 55

PNM Public Service Company of New Mexico. 72

PPA Power Purchase Agreement. 72, 124

PPI Producer Price Index. 128

RFE Recursive Feature Elimination. 105

RGR Rio Grande Rift. 68

ROC Receiver Operating Characteristic. 101

RPS Renewable Portfolio Standard. 72, 123

SBR Southern Basin and Range. 68

SHAP Shapley Additive Explanation. 110

SiGT Silica Geothermometer Temperature. 196

STEO Short Term Energy Outlook. 135

TN True Negative. 100

TP True Positive. 100

TPR True Positive Rate. 100

USGS United States Geological Survey. 46

VAG Value at Gain. 214

VAR Value at Risk. 214

WDS Well Data Set. 85

WDS4 Well Data Set plus 4. 87

WDS8 Well Data Set plus 8. 87

Chapter 1

Introduction

The pursuit of energy has shaped the history of mankind from its very beginning. And while the image of ancient humans huddled around fires for warmth, protection, and meal preparation is an archetype of our ancestral past, modern human needs remain much the same. Lighting to extend day into night, heating and cooling for residential comfort, cooking of the food we eat, access to the advanced technologies of our time — these all require energy from one source or another. Choices abound, from animal and plant-based fuels, to buried hydrocarbon resources, to alternatives like solar, wind, hydro, nuclear, and geothermal. The balance and utilization of these resources can shape societal growth on the geopolitical stage and influence the very future of the habitable Earth.

This thesis examines risk-mitigation strategies to help advance the role of one source, geothermal, in addressing ever-growing energy needs in a viable way. This chapter reflects on the extent of those needs and the conditions that may uniquely support an increased focus on geothermal as part of a commercial energy portfolio in the near-term. Opportunities and challenges associated with geothermal also lay the foundation for research questions motivating the remainder of this body of work.

1.1 Energy Trends

The United States Energy Information Administration (EIA) publishes annual forecasts on U.S. energy generation and consumption in the Annual Energy Outlook (AEO) report. Based on the 2020 reference case, the AEO model predicts a 70% increase in U.S. energy usage by 2050, driven primarily by the industrial and power sectors (EIA, 2021a). Electricity generation also grows by a third, largely due to renewables and natural gas as coal, nuclear, and oil experience reductions (Figure 1-1). These predictions are offered with the caveat of greater uncertainty in the wake of the COVID-19 pandemic, although the EIA suggests a return to normal will occur by 2025 and broader, decadal trends will remain unchanged (EIA, 2021a). International forecasts show similar growth in consumption and production, but traditional sources of energy like coal and natural gas also increase in capacity to meet the needs of India, China, and other rapidly developing nations (EIA, 2020b).

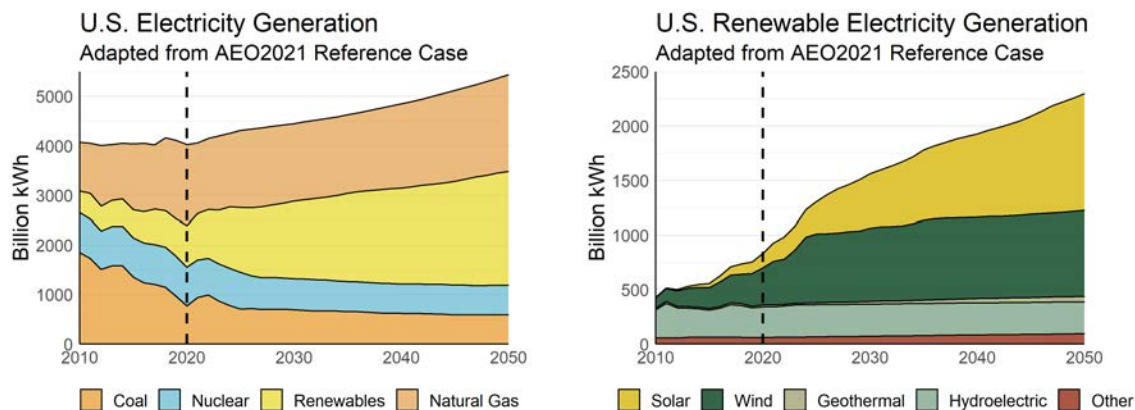


Figure 1-1: U.S. EIA projections of (Left) U.S. electricity generation by fuel source and (Right) individual contributions by renewables based on the AEO2021 reference case (EIA, 2021a). Vertical dashed line marks where historical records end and projections begin.

Lazard Asset Management breaks renewables down by Levelized Cost of Electricity (LCOE) in U.S. dollars/MWh, where LCOE is the estimated lifetime average net cost per unit energy of an electricity-generating plant. In their 2020 analysis, intermittent energy sources like wind and utility-scale solar are already cost-competitive with fossil fuel-derived sources (Figure 1-2) (Lazard, 2020). Geothermal, an “always

on” source of power, ranges from \$59-\$101/MWh LCOE, making it second-tier in cost-competitiveness but comparable to community and rooftop solar installations (Lazard, 2020). Overall, the transition away from fossil fuel super-dominance is in progress, and the demand for energy in support of population growth and country development will remain an industry driver over the next 30 years.

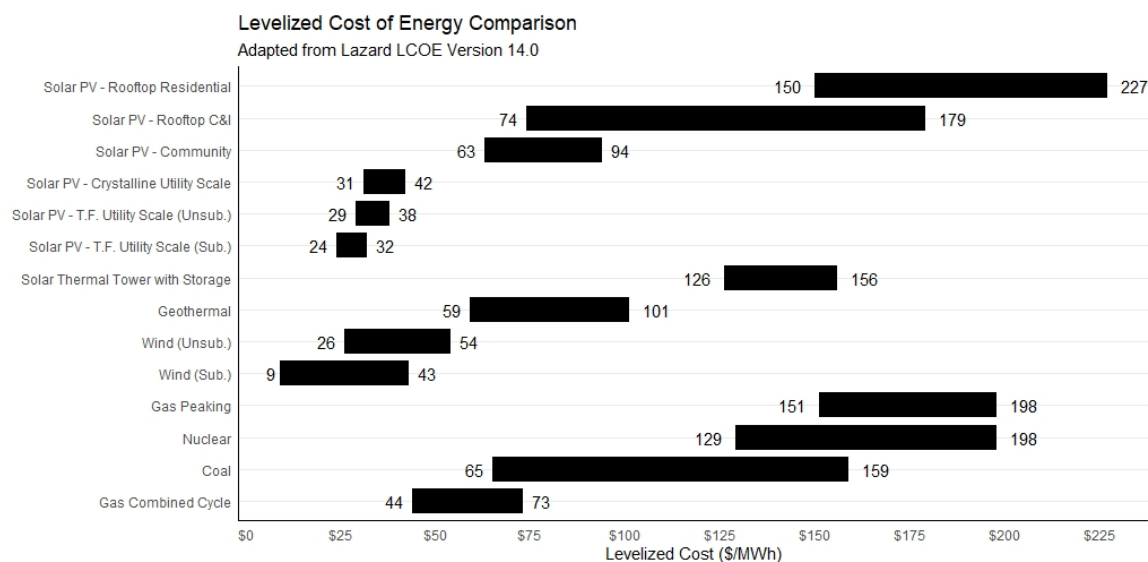


Figure 1-2: Cost comparison between different energy sources based on Lazard Asset Management LCOE analysis. C&I: Commercial & Industrial, T.F.: Thin Film, Unsub.: Unsubsidized, Sub.: Subsidized. For specific assumptions and caveats related to the analysis, see (Lazard, 2020).

1.2 Upstream Commercial Pressures

Businesses focused on exploration and production of oil & gas face a growing list of pressures influencing future corporate strategy. On one hand, the increase in energy consumption predicted by the AEO and IEO supports a steady increase in production to meet global demand; however, geopolitical tensions, state-ownership of oil companies, and breakthrough technologies create a volatile landscape unforgiving of an unsophisticated production approach. In just the past 15 years, major downturns in oil prices were triggered by a mixture of factors: the financial crisis, tied to banking practices and housing market instability in 2008 (Singh, 2021); increased production

from U.S. unconventional plays and supply decisions from the Organization of the Petroleum Exporting Countries (OPEC) in 2014 (Lioudis, 2021); and a price war between Russia and Saudi Arabia coinciding with a global pandemic in 2020 (Blessing, 2021). Additional uncertainty comes from National Oil Companies (NOCs) that control the majority of the world’s petroleum reserves, production, and rights for exploration and development. NOCs operate under different business drivers than International Oil Companies (IOCs), with ramifications on the stability of IOC investments and access to reserves as seen in Russia and Venezuela (Bremmer, 2010; Pirog, 2007). Meanwhile, disruptive technologies like precise directional drilling and efficient hydraulic fracturing have freed access to previously cost-prohibitive reserves, changing the balance of power as countries like the U.S. and China become less reliant on foreign hydrocarbon imports (Shuen et al., 2014).

Layered on top of these macroeconomic influences are unexpected events that have enormous impact on energy production and distribution operations. The 2020 outbreak of COVID-19 acted as an accelerator on longer-term trends of digital transformation and decarbonization in the oil & gas industry. In the wake of a 25% decrease in global demand, companies responded with massive layoffs and restructuring, a heightened focus on digitalization, and portfolio rationalizations that include shale write-downs and asset divestments (Deloitte, 2020). Also, extreme weather events consistent with global warming predictions are shining a critical spotlight on how energy is managed now and in the future. Blackouts, water outages, and surge pricing on electricity impacted millions of Texans in February 2021 when a winter storm brought record cold temperatures, exposing systematic weaknesses in energy infrastructure and generator preparedness for low-probability but feasible working conditions (HARC, 2021; Lazard, 2020). And additional threats loom in the cyber-world as malicious hacking activities have rippling social and financial implications. One such attack on Colonial Pipeline, which handles almost 50% of the liquid fuels supplied to the U.S. East Coast, led to gasoline shortages, price spikes, chemical factory shut-downs, and worldwide news coverage until the nearly \$5 million in ransom was paid in May 2021 (Sanger & Perlroth, 2021).

1.3 Net Zero Ambitions

The 2015 Paris climate agreement set a target of $< 2^{\circ}\text{C}$ on the rise in global temperatures above the pre-industrial average (i.e., the period from 1850-1900) to prevent the most extreme impacts of climate change ([UNFCCC, 2015](#)). The more commonly-ascribed 1.5°C target may be unachievable given current trajectories, and international calls to action are focusing on reducing anthropogenic carbon dioxide emissions to “net zero” as quickly as possible ([IPCC, 2018](#)). Greater public awareness about the environmental impacts of global warming and alignment with these targets is putting pressure on the energy industry to revise their traditional business models. Beginning in 2020, top-tier oil and gas companies started issuing press releases outlining energy transition targets for 2025, 2030, 2050, and beyond ([BP, 2020](#); [Chevron, 2021](#); [ConocoPhillips, 2020](#); [Equinor, 2020](#); [ExxonMobil, 2021](#); [Shell, 2020, 2021](#); [Total, 2020, 2021](#)). The proposed strategies vary but generally focus on (i) reducing company stake in fossil-fuel exploration and production activities, (ii) setting a net zero target applicable to emissions from operations, product carbon intensity, and carbon offsets, and (iii) dedicating investments in low- and no-carbon energy alternatives to replace hydrocarbons. Nevertheless, the pressure to do more, faster reached a new peak in May 2021 when a court decision in The Netherlands and approved shareholder proposals for two U.S. majors demanded an accelerated push toward emissions reductions and low-carbon energy options ([McWilliams, 2021](#)).

1.4 Geothermal Energy

Several factors must be met for an energy source to be considered a viable and sustainable alternative to the carbon-based fuels (coal, natural gas, oil) that currently meet much of the world’s energy needs. [Glassley](#) outlines a succinct set of criteria for such an energy resource: i) self-replenishment, ii) adequate abundance to meet energy demands, iii) low- to no-greenhouse gas emissions, and iv) cost competitiveness compared to accessible alternatives ([2015](#), p. 7). The first criterion precisely

characterizes renewable sources of energy like solar, gravitational, and geothermal. Solar energy includes both direct (sunlight) and indirect (wind, waves, water cycle) resources (Hohmeyer, 2008). Gravitational energy drives the tides and supports energy-storage solutions like pumped-storage hydropower (EERE, 2021; Hohmeyer, 2008). And geothermal energy is derived from accretionary heat and natural decay of radioactive isotopes within the Earth (Hohmeyer, 2008). Table 1.1 specifically addresses criterion ii; the annual abundances of most renewable sources exceed energy demand, with geothermal scaling several orders of magnitude greater in annual flux/demand than other options (Hohmeyer, 2008). Geothermal also meets the third criterion by providing a no-carbon source of energy that could support net zero aspirations when used in place of fossil fuels. The fourth criterion is less clear-cut. Geothermal costs depend strictly on use case, but as noted in Section 1.1, LCOE analysis currently ranks geothermal below wind and solar in the list of renewable options (Lazard, 2020). Does this make geothermal non-viable as an alternative? Perhaps no, as geothermal comes with several unique opportunities and benefits.

Renewable Source	Annual flux (EJ/yr)	Annual Flux/Demand*
Solar	3,900,000	8,700
Wind	6,000	13
Hydro	149	0.33
Bioenergy	2,900	6.5
Ocean	7,400	17
Geothermal	140,000,000	31,000

Table 1.1: Annual renewable energy fluxes, adapted from Table 1 of (Hohmeyer, 2008). The (*) indicates flux/demand ratio is derived from global demand estimates.

Most fundamentally, geothermal energy offers a reliable, nearly inexhaustible resource accessible anywhere around the world. Unlike wind and solar, which depend on favorable locations and vary with both season and time of day, geothermal is ubiquitous and continuous. Furthermore, it can provide baseload power for regional electrical grids without the additional need of assistive energy-storage technology (Tester et al., 2006). Based on history, one might assume geothermal only works under conventional hydrothermal conditions, e.g., near active volcanoes (e.g., Iceland,

Indonesia) or within major rift zones (e.g., East African Rift). However, additional opportunities lie in low-temperature direct-use geothermal for heating and cooling of buildings, industrial processing, agricultural activities, and manufacturing ([Glassley, 2015](#), p. 9). And technology supporting Enhanced Geothermal Systems (EGS) provides access to subsurface heat in areas without hydrothermal conditions ([Tester et al., 2006](#)). Natural radioactive decay takes place throughout the Earth’s crust, contributing to the global rise in temperature with depth known as the geothermal gradient ([Fowler, 2005](#), p. 279). Where conditions permit access to suitable depths, there is the potential for geothermal energy capture.

For all its benefits, the use of geothermal energy comes with a set of challenges that are unique among renewable energy options. Many mirror issues faced by oil & gas producers, like failures in drilling equipment or borehole integrity. Others include risk of low resource quality, poor reservoir productivity, unexpected structural and stratigraphic complexity, and undesirable fluid chemistry ([Beckers, 2016](#); [Hadi et al., 2010](#)). The similarities extend into un conventionals/EGS, where hydrofracturing risks comprise inadequate permeable rock volume within the stimulated fracture zone, short-circuiting of fluid flow between injection and productions wells, fluid losses within the subsurface, and induced seismic activity ([Jelacic et al., 2008](#); [Pan et al., 2019](#)). As in oil & gas projects, some of these risks can be mitigated through appropriate subsurface characterization, others from high-resolution reservoir and fracture modeling. Collectively, the overlap in operational challenges and the skill sets required to tackle them defines a unique compatibility between oil & gas and geothermal. Knowledge transfer between the two domains could directly benefit geothermal operations and risk-mitigation strategies ([Petty et al., 2009](#)). And investing in geothermal assets could take upstream companies a step closer to meeting net-zero commitments while retaining and utilizing existing talent.

One of the most significant uncertainties for geothermal is cost. The scope of an LCOE assessment considers all costs in a geothermal project, from early exploration, through development drilling and power plant construction, to operations and maintenance over a 25- to 30-year lifetime ([Beckers et al., 2013](#); [Entingh et al., 2006](#);

Tester & Herzog, 1990). Studies show drilling of exploration, appraisal, injection, and production wells can account for 60-75% of the total EGS project expenses (Lukawski et al., 2016; Petty et al., 2009). New technologies being developed may help address future drilling costs (Lowry, Finger, et al., 2017; NREL, 2020). However, substantial project savings could also be achieved through better characterization prior to drilling the first exploration well; increasing the probability of well success without multi-million dollar well failures would certainly reduce LCOE. In addition, choices in the design of the geothermal power plant may also have significant cost implications over the life of a geothermal field. Lessons can be learned from power plants built to design specifications incompatible with actual production conditions (e.g., Manente et al., 2011). Improving the economics of geothermal start with recognizing uncertainties in the system and using those uncertainties to make better decisions on *where to drill* and *how to produce*. Providing the appropriate tools and methods to make these decisions could also help oil & gas companies manage the risk of embracing lower-carbon energy production with geothermal as part of their energy mix.

1.5 Research Questions

Based on the above-mentioned opportunity to define methods that incorporate uncertainty characterization for risk management in geothermal exploration and production, this thesis will address the following research questions:

1. Can geothermal exploration risk be mitigated by insights derived from readily-available data sources with little change in project cost? Can the data establish additional actions for further risk reduction before drilling a geothermal well?
2. How can characterization of present and future uncertainties influence the development and production strategy for geothermal facilities? In what ways will such an approach mitigate risk, if at all?

1.6 Thesis Outline

This thesis follows the flow shown in Figure 1-3 and is structured as follows:

- Chapter 2 provides relevant background material on geothermal systems, a literature review on geothermal exploration and cost modeling, and a primer on case study areas investigated in later chapters.
- Chapters 3–4 describe the thesis data sources, data preparation strategies, and the chosen methods for favorability prediction with machine learning (Ch. 3) and cost modeling of geothermal project designs (Ch. 4).
- Chapters 5–6 describe the machine learning model results (Ch. 5) and cost modeling results (Ch. 6).
- Chapter 7 frames learnings from the previous chapters as risk mitigation actions in an overall risk management workflow.
- Chapters 8 and 9 conclude the thesis with a summary of thesis insights (Ch. 8) and a list of future work opportunities (Ch. 9).

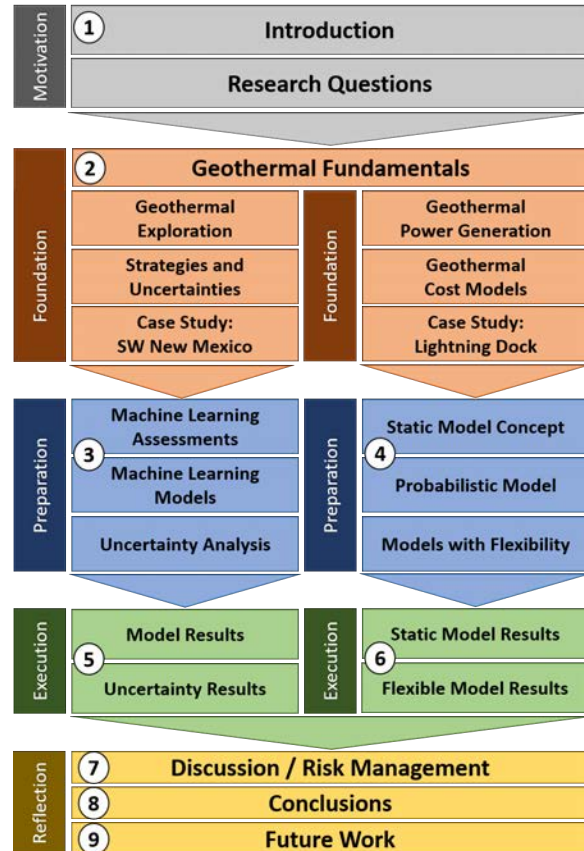


Figure 1-3: Flow chart of thesis structure. Chapter numbers are circled.

Chapter 2

Geothermal Background

2.1 Geothermal Systems

2.1.1 Origins of Heat

Accretion

The story of geothermal energy begins with the birth of planet Earth. Approximately 4.56 billion years ago ([Allegre et al., 1995](#); [Patterson, 1956](#)), the Earth coalesced as a molten body heated by repeated impacts with other objects in the early solar system, like the planetesimal collision responsible for the formation of the Moon ([Stevenson & Halliday, 2014](#)). Over tens of millions of years, the Earth compacted, cooled, and differentiated, settling into the now familiar layered structure of a solid inner core, liquid outer core, viscous mantle, and outermost brittle crust ([Press et al., 2004](#), p. 7) (Figure 2-1). The intense heat from that early accretionary history remains concentrated in the core, where temperatures fall in the range — a matter of continued debate — of 6000 ± 500 K ([Fowler, 2005](#), p. 372). Of the heat reaching the surface of the earth, $\approx 60\%$ flows through conductive and convective pathways from the lower crust or below ([Stein, 1995](#)). Diffuse conductive heat transfer occurs everywhere across the Earth’s surface, but narrow zones of high heat flow follow crustal plate boundaries. In fact, the subduction-sourced volcanoes that ring the Pacific Ocean, divergent zones at the mid-ocean ridges and East African rift, and

major strike-slip boundaries like the San Andreas fault zone all mark locations where focused heat anomalies are being tapped by geothermal installations (DiPippo, 2012, p. 16).

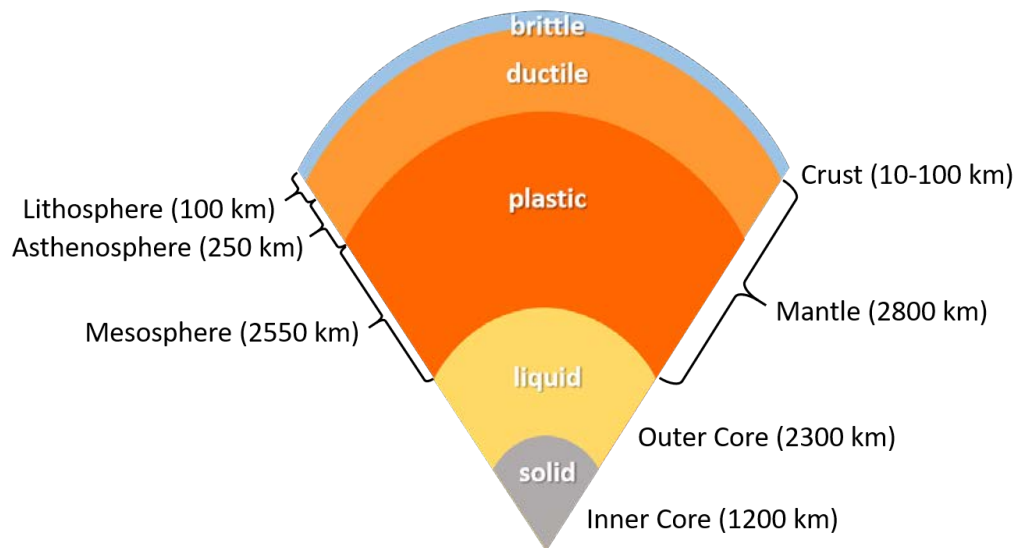


Figure 2-1: Inner structure of planet Earth by mechanical properties (left) and composition (right). Approximate layer thicknesses are noted in parentheses.

Radioactive Decay

The second major source of heat within the Earth is the decay of radioactive isotopes. Early radioactive heating included radioisotopes with short half-lives like Aluminum-26 and Hafnium-182, which are now no longer present (Glassley, 2015, p. 16). Among the radioactive elements most influential to crustal heat today are uranium (U), thorium (Th), rubidium (Rb), and potassium (K) (Glassley, 2015, p. 17). The decay of these and other elements accounts for 40% of the crustal thermal budget (Stein, 1995). But element abundances are not distributed uniformly throughout the crust. On average, continental crust, particularly the upper continental crust, has significantly higher concentrations of U, Th, and K radioactive elements compared to oceanic crust, and both types of crust are 1–2 orders of magnitude more enriched than the mantle (Fowler, 2005, p. 276). This relationship holds for representative igneous rock types; granite generates more heat than basalt, and both out-produce

ultramafic rocks like peridotites ([Fowler, 2005](#), p. 276).

2.1.2 Measuring Heat

Geothermal Gradient

Subsurface conditions show an increase in temperature with depth sustained by the flow of original accretionary heat and generated radioactive heat. This is commonly referred to as the geothermal gradient and averages ≈ 30 K/km for continental crust ([Press et al., 2004](#), p. 209). Horizontal thermal gradients exist and can be quite high (e.g., at the contact between country rock and intruded magma), but they are mathematically required to average to 0 K/km globally. Deviations from the average geothermal gradient are common and reflect the complexity of the rock record. The crust comprises distinct layers, or strata, that vary in composition and rock type. Unlike igneous formations that can be relatively homogeneous, sedimentary rocks derive from surface processes that mix sediments from a variety of original source rocks with different degrees of sorting ([Press et al., 2004](#), p. 164–168). Alteration from fluids, heat, and pressure can then modify the composition of these rocks, causing constituent minerals to change form and arrangement to create metamorphic rocks ([Press et al., 2004](#), p. 195–205). The spatial and depth variations in these formations, enhanced by structure processes like faulting, create subsurface compositional heterogeneity, directly reflected in rock properties. Thermal conductivity, specifically the ability to move deep-sourced heat to shallower depths, and radioactive element abundance, or the ability to generate additional heat in situ, can therefore vary in all directions in the subsurface. Thermal heterogeneity can be further compounded by anomalies created from salt movement ([Press et al., 2004](#), p. 164–168), magmatic intrusions, or global tectonic processes. Geology and geologic history therefore play an important role in defining the geothermal gradient of an area.

Heat Flow

Fundamentally, heat moves in the direction from hot to cold (Second Law of Thermodynamics) at a rate that linearly scales with the thermal gradient. Simple one-dimensional thermal conduction can be characterized by the relationship (Fourier’s Law, [Fowler, 2005](#), p. 270):

$$\mathbf{q} = -k\nabla T, \quad (2.1)$$

where $\mathbf{q}$ is heat flux, or heat flow per unit time per unit area, with S.I. units of $\text{W}\cdot\text{m}^{-2}$. Heat flux depends on the gradient of temperature (T) and the thermal conductivity (k), that is, the ability of material to conduct heat. Different rock types have different values of k , e.g., sandstone varies from 1.60–2.10 $\text{W}/\text{m}\cdot\text{K}$, granite has higher values of 1.73–3.98 $\text{W}/\text{m}\cdot\text{K}$ ([DiPippo, 2012](#), p. 30). The most abundant minerals in crustal rocks, feldspars and quartz, can differ in k by up to $3 \times$ in value, so their relative fractions strongly influence the thermal conductivity of a formation ([Glassley, 2015](#), p. 22). Regardless, these and other common crustal minerals tend to be poor thermal conductors compared to metals like aluminum (210 $\text{W}/\text{m}\cdot\text{K}$) and iron (73 $\text{W}/\text{m}\cdot\text{K}$), making crustal conduction a slow means of heat transfer ([DiPippo, 2012](#), p. 23).

Conduction is the primary method of heat transfer in the crust, while convection dominates on global, tectonic scales. The equation governing convection can be arranged to highlight an internal conduction term, illustrating the greater complexity of convection by comparison ([Turcotte & Schubert, 2002](#), p. 267):

$$\frac{\partial T}{\partial t} = \frac{1}{\rho c_P} \nabla \bullet (k \nabla T) - \mathbf{u} \bullet \nabla T, \quad (2.2)$$

where $\mathbf{u}$ is the velocity of the fluid, ρ is the material density, and c_P is the specific heat, which defines the amount of heat necessary to raise 1 kg of that material by 1 K. Convection thus combines heat transfer from conduction with mass movement. Since mantle minerals are poor conductors of heat, the mantle insulates and traps heat from the core near the core-mantle boundary ([Glassley, 2015](#), p. 25). The combined effects of lower viscosity, thermal expansion, and buoyancy forces near that

boundary all drive convective mantle flow ([Glassley, 2015](#), p. 25). Mantle convection is responsible for the high heat-flow values at crustal plate boundaries like mid-ocean ridges, as well as intraplate hot spots underlying Hawaii, Yellowstone, and several other locations around the world. Smaller-scale convection also takes place at subduction zones where the water-rich material from the down-diving plate experiences low-temperature melting, migrates upwards, and forms volcanic arcs on the surface as observed in Japan, Indonesia, and the U.S. Pacific Northwest ([Press et al., 2004](#), p. 31–33) — all locations with geothermal potential.

Heat-flow measurements capture the flux of heat through the Earth’s surface as a result of these and other complex subsurface processes. In this respect, heat flow serves as a simpler, more accessible geothermal metric than the more sparsely-measured geothermal gradient. Today, high-quality heat flow measurements can be obtained in marine conditions, along continental margins, on mid-ocean ridges, and from the multitude of wells drilled by the oil & gas industry, supporting large aggregate data sets like the New Global Heat Flow Database ([Lucazeau, 2019](#)). As [Figure 2-2](#) shows, data from these collections can be gridded to create spectacular maps of heat flow variations around the world. These maps offer a good starting point for quickly targeting regional-scale geothermal potential, which can be further refined through other methods (see [Section 2.2](#)).

2.1.3 System Fundamentals

The conventional concept of a geothermal system consists of five key entities ([DiPippo, 2012](#), p. 9):

- i Heat source of significant size and temperature
- ii Permeability, typically in the form of a fracture network within crystalline rock
- iii Ample volume of working fluids e.g., water from precipitation and drainage
- iv An impermeable sealing layer
- v Consistent, reliable fluid recharge

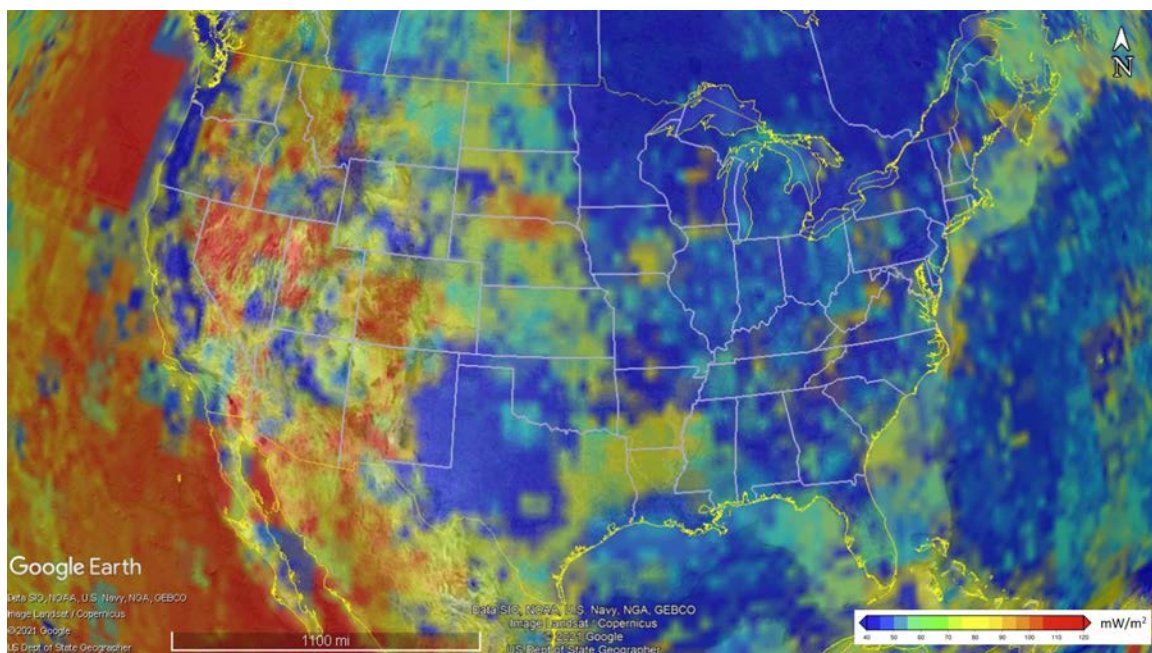


Figure 2-2: Heat-flow estimates for the continental United States, plotted in Google Earth with data layer from (Lucazeau, 2019). Units are mW/m^2 .

Hydrothermal Systems

If naturally present, the combination of these five elements defines a hydrothermal system. As water percolates down, captures heat from the permeable thermal reservoir, and gets trapped beneath the sealing caprock, a small fraction of the resource can escape to the surface to produce distinctive geothermal manifestations like fumaroles, hot pools, geysers, mud pots, and discolored or altered rocks (Figures 2-3 and 2-4). These features are strong indicators of hydrothermal activity at depth.

Hydrothermal resources have been exploited by humans for many millennia. Artifacts show proto-Native American use of hydrothermal waters for cleaning and health restoration over 10,000 years ago (DOE, 2021a). The importance of geothermal hot springs for Roman, Japanese, Chinese, and Ottoman baths is also well-established in the historical record (Lund, 2007). Industrial use began in the 1800s with chemical extraction from geothermal steam, pools, and deposits in Larderello, Italy (DiPippo, 2012, p. 251). Geothermal district heating, or large-scale heating of residences and businesses using geothermal-produced fluids, was pioneered in Chaudes-Aigues, France in the 1300s and first introduced to the United States in 1892 with an instal-



Figure 2-3: Terminal Geyser, Lassen National Park, California. Photo credit: Author



Figure 2-4: Active mud pot and ground staining on the bank of Boiling Springs Lake, Lassen National Park, California. Photo credit: Author

lation in Boise, Idaho ([Lund, 2007](#)).

These few examples show some of the many opportunities for low-temperature ($T < 90^{\circ}\text{C}$) and medium-temperature ($90^{\circ}\text{C} < T < 150^{\circ}\text{C}$) geothermal resource use, which extends beyond just hydrothermal systems. District heating can help meet building and water temperature needs, and agriculture, textiles, chemicals, and even the food industry can benefit from access to low-temperature geothermal ([DOE, 2021b](#); [Liu et al., 2015](#)). Growing interest has led to dedicated funding opportunities; for example, the Geothermal Technologies Office (GTO) within the United States Department of Energy (DOE) recently awarded a grant to Cornell University to pilot a deep direct-use geothermal project providing baseload heating for the university campus during cold New York winters ([Hamm et al., 2021](#); [Tester et al., 2015](#)).

The topic of this thesis instead concerns the use of moderate- to high-temperature geothermal for generating electricity. The first example of geothermal power production came from Italian experiments in 1904, and the first commercial plant went online in Larderello, Italy in 1914 ([DiPippo, 2012](#), p. 251). Geothermal power made its debut in the United States with the development of The Geysers field beginning in 1960 ([Tester et al., 2006](#)). Hydrothermal plants quickly appeared in New Zealand, Japan, Iceland, Indonesia, Kenya, the Philippines and elsewhere throughout the 1970s-1980s, with continued growth through to present-day ([Lund, 2007](#)). 2020 statistics from the International Renewable Energy Agency (IRENA) place the United States as world leader in geothermal installed capacity (2587 MW), followed by Indonesia (2131 MW), and the Philippines (1928 MW) ([IRENA, 2021](#)) (Figure 2-5).

A comprehensive assessment of moderate and high-temperature Known Geothermal Resource Areas (KGRAs) by the United States Geological Survey (USGS) determined the U.S. has conventional geothermal (hydrothermal) power generation potential of $\approx 9,000$ MW, and an additional $\approx 30,000$ MW potential exists in undiscovered resources ([Williams et al., 2008](#)). The recent DOE GeoVision study notes hydrothermal potential in Alaska and Hawaii alone are $\approx 8,000$ MW due to their unique tectonic environments (Aleutian subduction zone and Hawaiian hot spot, respectively), and high-case model estimates for the continental U.S. forecast an installed capacity

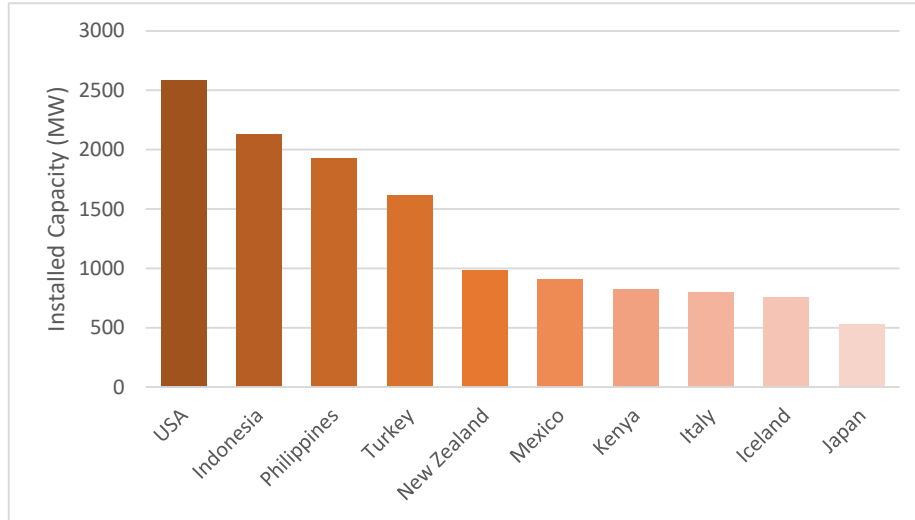


Figure 2-5: Countries ranked by installed geothermal capacity based on 2020 statistics, adapted from ([IRENA, 2021](#))

of ≈ 16.5 GW by 2050 from known and unknown hydrothermal resources ([Augustine et al., 2019](#); [Hamm et al., 2019](#)).

Enhanced Geothermal Systems

Geothermal potential exists even when one or more of the fundamental elements listed in Section 2.1.3 are missing. These unconventional geothermal systems, often referred to as Enhanced Geothermal Systems (EGS), contain a significant heat source but lack either the adequate permeability or sufficient rechargeable working fluids to meet the requirements of a hydrothermal system. A broader definition for EGS includes thermal production from sedimentary and crystalline tight rock, poorly-performing hydrothermal systems, co-production from oil & gas operations, and even thermal recovery directly from magma ([Tester et al., 2006](#)). Focusing on the more common tight-rock scenario, EGS systems work by artificially creating or improving reservoir permeability and ensuring sustained fluid flow through the reservoir ([Glassley, 2015](#), p. 281). Fluids pumped down an injection well pass through a stimulated fracture network. Heat from the thermal reservoir warms the fluid before it returns to the surface via one or more producing wells and enters a power plant to produce electricity (Figure 2-6). Reservoir stimulation generally involves hydraulic fracturing

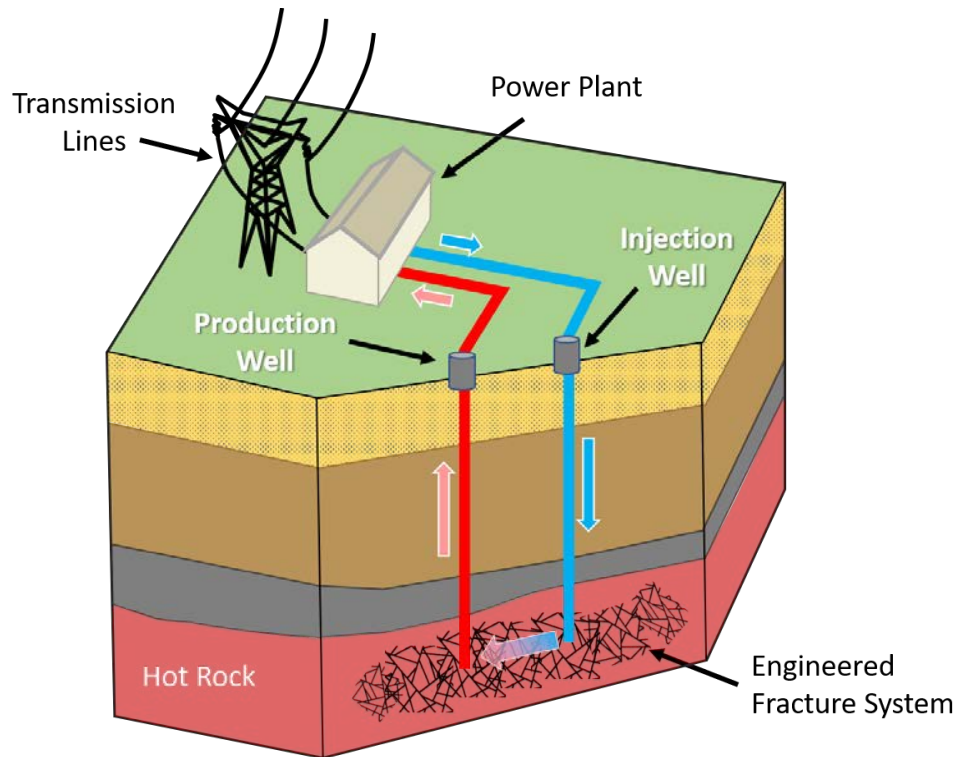


Figure 2-6: Schematic diagram illustrating the EGS concept with a single injection-production well pair.

or “fracking” techniques to create pathways connecting injector-producer pairs. The technology and field capabilities are thus well-aligned with unconventional oil & gas operations (Petty et al., 2009).

EGS could help propel geothermal beyond niche hydrothermal environments to become a more significant contributor to electricity production in the United States. The sources of subsurface heat exist everywhere (see Section 2.1.1), so accessing temperatures that could support power production fundamentally rely on drilling deep enough and creating the artificial conditions necessary for heat capture. Los Alamos National Laboratory (LANL) validated the EGS approach in crystalline rock at Fenton Hill beginning in 1974 (Tester et al., 2006). Feasibility studies followed soon thereafter in Japan, Germany, the U.K., and France (Breede et al., 2013). Among the most notable EGS projects providing key lessons learned are The Geysers (U.S.), Soultz-sous-Forêts (France), and Cooper Basin (Australia). The Geysers stands out as the largest geothermal power-generating complex in the world, delivering $\approx 1,000$

MW even after 60 years of steam production (Jelacic et al., 2008; Williams et al., 2008). Although drilling issues and public concern over induced seismicity halted efforts (E. Larson, 2009), rock-stimulation experiments in the NW Geysers proved distinct reservoirs can be developed adjacent to active hydrothermal operations (Pan et al., 2019) — a critical concept revisited in Chapters 4 and 6. The Soultz project commenced in 1987 with one injector and two producers drilled to 5 km depth, eventually supporting a small 1.5 MW power plant built in 2007-2008 (DiPippo, 2012, p. 463). In addition to experimental lessons learned for drilling, hydrofracturing, chemical stimulation, scaling, and corrosion, Soultz today supports both power production and district heating (Durst, 2013). Cooper Basin, the largest EGS demonstration project of its time, showed great promise after a 6-year proof-of-concept phase (Stephens & Jiusto, 2010). However, the project halted in 2015 after a previously-unrecognized intersecting fault zone derailed further progress, highlighting the importance of robust structural appraisal in assessing geothermal prospects (Holl, 2015).

As the fate of these projects might suggest, the long-standing promise of commercial EGS has not yet been fully realized. However, several active projects supported by the National Renewable Energy Laboratory (NREL) and GTO (e.g., EDGE, EGS Collab, FORGE) are providing insights necessary to mature subsurface models, drilling technologies, and stimulation methods for more widespread EGS adoption (Hamm et al., 2021). And recent technology advances and partnerships involving start-ups like Deep Earth Energy (GeoEnergy, 2021), Eavor Technologies (Ross, 2020), Eden GeoTech (Daso, 2020), and Fervo Energy (Moss, 2021; Shieber, 2021), show there is a growing fervor to overcome technology roadblocks currently holding EGS back. Projections show the size of the prize with success in EGS. Assuming a maximum cut-off depth of 7 km and a minimum reservoir temperature of 150°C, the EGS resource potential for electricity production in the continental United States might be at least 5,150 GW, with an additional $\approx 1,500$ MW from Near Hydrothermal Field-EGS (NF-EGS) (Augustine et al., 2019). To put this opportunity into context: the total utility-scale electricity generation capacity from all sources in the United States was $\approx 1,200$ GW in 2019, over $4\times$ less in magnitude than predicted

EGS potential ([EIA, 2020a](#)).

2.2 Geothermal Exploration

2.2.1 Background

The identification of geothermal resources traditionally relied on surface expressions of hot fluids circulating at depth. Quite simply, if a bubbling hot spring or a geyser was present, you had a working geothermal system. Some of the first sites for geothermal power production, like Larderello and The Geysers, were (unsurprisingly) targeted because of their tell-tale surface characteristics ([Glassley, 2015](#), p. 111). However, this prospecting method only applies to fully-functioning hydrothermal systems, and not all such systems have surface manifestations if fluids remain trapped beneath a subsurface impermeable seal. These hidden or “blind” geothermal resources require more sophisticated exploration methods to identify and assess accurately.

2.2.2 Regional Exploration

Regional evaluation techniques historically depended on sparse borehole data to map the geothermal potential of the U.S. ([Kehle et al., 1970](#)). Successive efforts led to progressively more comprehensive collections of heat flow and bottom hole temperature (BHT) measurements ([Blackwell et al., 1990, 2011](#); [Muffler, 1979](#); [Sorey et al., 1983](#); [Wisian et al., 1999](#)), now widely-available with other geothermal data through the DOE NGDS platform ([Anderson et al., 2013](#)). These efforts supported a better understanding of broad trends that tie directly to tectonic provinces in the United States. Subduction and transform plate boundaries to the west, combined with extension in the Basin and Range, make this area of the U.S. more susceptible to high heat flow, high geothermal gradients, and greater potential for hydrothermal activity ([Mariner et al., 1983](#)). Passive margins along the Atlantic and Gulf of Mexico sides of the country make those regions less likely to host significant hydrothermal systems, although surveys have identified sedimentary basins with elevated heat flow

in the South (Louisiana-Arkansas), Central (Iowa-Illinois, Nebraska-South Dakota), and Eastern (Appalachians) United States ([Blackwell et al., 1995](#); [Sorey et al., 1983](#)).

2.2.3 Play-Scale Exploration

Continental maps provide a super-regional view of geothermal prospectivity, but further refinement is required to progress an exploration program. Specifically, the determination of semi-regional plays and local prospects must come before deciding where and how to develop a geothermal field. Characterization activities can be decomposed into the evaluation of four earth subsystems: hydrologic, stratigraphic, structural, and thermal. The main types of surveys associated with each subsystem are listed in Table 2.1 and discussed briefly in the following section. Note that survey methods typically provide information on multiple subsystems. The associated non-uniqueness of subsurface interpretations based on these surveys highlights the need to integrate multiple lines of evidence for exploration activities.

Geologic Data Collection

Geologic field mapping can provide crucial direct evidence supporting the presence of geothermal systems. Surface manifestations like geysers, vents, and mud pots may also be coincident with mappable mineralogic indicators of the subsurface chemistry and style of geothermal activity. Volcanic-based geothermal systems tend to have acid-sulfate waters with hydrogen sulfide-rich brines that leave behind sulfur deposits ([Glassley, 2015](#), p. 123). Bicarbonate geothermal waters can produce distinctive travertine terraces or subaqueous tufa deposits, as well as a unique variety of potassium feldspar called adularia ([Glassley, 2015](#), p. 125). And chloride geothermal fluids are known to precipitate sinter or geyserite deposits composed of opal or amorphous silica ([Glassley, 2015](#), p. 125). Elevated abundances of trace elements like boron and lithium typically occur in chloride brines compared to meteoric (i.e., derived from precipitation) waters, so their presence in mineral assemblages is also diagnostic ([Bielicki et al., 2015](#); [Milot & Négrel, 2007](#)). Other valuable products of a

Hydrologic	
Earthquake records	movement of magma or fluids
Geologic survey	surface expressions (vents, geysers, deposits), drainage
Hydrologic survey	fluid geochemistry, recharge and rate, water table
Magnetic survey	hydrothermal alteration
Precipitation records	water cycle inputs, recharge rate
Resistivity survey	subsurface fluids or hydrothermal alteration
Satellite survey	surface drainage patterns, distribution of deposits
Seismic survey	presence and location of subsurface fluids
Stratigraphic	
Geologic survey	stratigraphic successions, seal and reservoir
Gravity survey	density anomalies, stratigraphic variations
Magnetic survey	igneous formations, stratigraphy and reservoir
Radiometric survey	mineral abundances, source and reservoir
Resistivity survey	thermal conductivity, lithology
Satellite survey	distribution of outcrops and formations
Seismic survey	stratigraphy and rock properties
Structural	
Aerial survey	surface fault traces
Earthquake records	fault location, fault recency
Geodetic survey	active deformation or faulting
Geologic survey	surface fault traces, fault recency
Gravity survey	subsurface faults, plutons, salt
Resistivity survey	fractured zones
Satellite survey	topography, structural patterns
Seismic survey	subsurface faults, folds, other structural features
Thermal	
Aerial survey	thermal anomalies in shallow subsurface (IR)
Air Temperature records	near-surface thermal conditions
Geologic survey	surface manifestations (dikes, vents, deposits)
Gravity survey	presence of high-T anomalies (e.g., magma)
Hydrologic survey	geothermometry, dominant geofluid liquid phase
Radiometric survey	radioactive heat generation
Resistivity survey	thermal conductivity, temperature gradient
Seismic survey	depth to mantle, intrusive igneous features
Temperature survey	heat flow, geothermal gradient

Table 2.1: Data collection methods useful for characterizing Earth subsystems that influence geothermal favorability (DiPippo, 2012, p. 19-33; Doughty et al., 2018; Glassley, 2015, p. 154-155).

geologic survey include maps of surface fault patterns to better constrain the structural history, as well as volcanic intrusive (dikes, sills) and extrusive (flows) features for understanding the thermal history and potential deeper reservoir potential.

Direct geochemical analysis of springs, pools, and samples collected from wells provides additional insights into the hydrologic characteristics of an area. Water chemistry offers information on the dominant resource fluid phase (vapor vs. liquid), the temperature of the subsurface formations encountered by the fluids, and the nature of the original water source (DiPippo, 2012, p. 25). The concentration or equilibria of different elements, e.g., quartz, chalcedony, sodium, potassium, and calcium, can be compared to empirically-derived trends for reservoir temperature estimates (Glassley, 2015, p. 157). These geothermometry methods offer insights into the deep thermal regime, although uncertainty around fluid migration pathways disallows any clear designation of the exact location and depth of a thermal reservoir.

Field methods like water sampling and geologic mapping provide local insights that can be aggregated for a bigger picture understanding of an area, with the caveat that field data are often limited in quantity and spatial distribution. Aerial and satellite surveys gather regionally-extensive measurements without the spatial sampling bias implicit in field activities. High-resolution topography captured in Digital Elevation Models (DEMs) and gradient (slope) maps can reveal morphology patterns tied to surface water drainage and recharge potential for deeper geothermal systems. Other optical products provide additional information of value; infrared imagery captures thermal anomalies in the shallow subsurface, stereographic images emphasize fault offsets missed in the field, and hyperspectral imaging can discriminate between different mineral assemblages, including geothermally-sourced boron-rich accumulations (DiPippo, 2012, p. 22; Glassley, 2015, p. 154-155).

Geophysical Data Collection

Geophysical surveys target variations in the subsurface, revealing how conditions and properties vary with depth. Magnetic surveys detect the fields imprinted on rocks with susceptible minerals that have experienced appropriate thermal conditions (Lowrie,

2007, p. 248-249). Magnetic anomalies, calculated by computationally subtracting the regional magnetic field and non-geologic signals, can indicate the presence of intrusive volcanic bodies or hydrothermally-altered formations (Glassley, 2015, p. 146). Gravity surveys similarly require several corrections to reveal local anomalies of interest (Lowrie, 2007, p. 59-62). Gravity anomalies highlight differences in subsurface density, which may be diagnostic of mineral alteration from hydrothermal processes, the presence of fractures, or pore fluid changes (e.g., replacement of meteoric water with hydrocarbons, hydrothermal fluids, or steam) (Glassley, 2015, p. 150). Resistivity surveys measure electrical resistivity (or its inverse, conductivity) within an instrumented area — a property sensitive to entrained fluids and the variations in mineralogy associated with alteration zones (Glassley, 2015, p. 147). However, poor resolution beyond shallow (≈ 1 km) depths strongly limits the reach of traditional resistivity studies. Magnetotellurics (MT), measurements of currents induced by natural electromagnetic waves originating in the ionosphere, can extend conductivity insights much deeper, even into the upper mantle (Lowrie, 2007, p. 225). And seismic surveys, which measure acoustic wave propagation in the subsurface, can be processed and modeled to image stratigraphy, faults, fluids, and rock properties to a range of depths. Seismic refraction data can constrain whole crustal thickness (e.g., Holmes, 2009), while seismic reflection data is useful for defining the prospect geometry and dimensions of the reservoir and seal (e.g., Cappetti et al., 2005).

2.2.4 Integration Strategies

Joint Inversion

As powerful as geophysical methods are at remotely detecting earth properties, each method represents an inherently underconstrained problem. Complex mathematical routines can invert data collected by aerial survey (e.g., gravity, magnetics) or acquired on the surface (resistivity, MT, seismic) to create 2-D or 3-D subsurface models. Still, unlike highly precise medical imaging technologies like Magnetic Resonance Imaging (MRI) that completely surround the target, geophysical techniques

have a limited top-down or cross-well view of the earth and must contend with noisy environments and many unknown parameters. Solutions to geophysical problems are thus non-unique, and uncertainty increases with depth.

Joint approaches to mathematical inversion for subsurface models address this ambiguity by constraining solutions to match the observations from multiple geophysical methods at once (Vozoff & Jupp, 1975). The complexity of a joint inverse problem applied to geothermal evaluation rapidly grows as more data sets are incorporated, particularly when the different data are sensitive to different earth properties (Moorkamp et al., 2011). In addition, geothermal model results can meaningfully differ depending on the selected additional assumptions made to make an under-determined problem into a well-posed one (Rosenkjaer et al., 2015). One alternative approach avoids the mathematical and computational demands by combining data semi-quantitatively, either by visually correlating individual model results or by cascading constraints from one geophysical model to the next to generate an integrated solution (Jousset et al., 2011; Lichoro et al., 2019). Absent a tightly-coupled formulation of the relationship between all data inputs, the weighted influence of each geophysical data source must be chosen by the analyst, which can be a significant source of uncertainty, analogous to assigning prior mean and covariance in multivariate Bayesian inference. Integrating sparse or qualitative geologic data also becomes an issue in this already difficult problem of data integration.

Play Fairway Analysis

Regional or play scale exploration methods adapted from oil & gas companies include geospatial risk assessments known as Play Fairway Analysis (PFA). Conceptually, PFA breaks risk down into the constituent elements of a successful play before summarizing them into an overall favorability prediction (Garchar et al., 2016). For hydrocarbons, risk elements include reservoir, seal, and charge (Fraser, 2010), and sometimes structure or trap (Doust, 2010). Maps are generated for each element based on any available data, including literature reviews, point data like wells or field sampling, and modeling results. Taking the collective evidence (or lack thereof) as

input, subject matter experts provide a perception of chance as a probability, and statistical approaches consolidate the probability maps into a cumulative favorability map in addition to yet-to-find resource volume calculations ([Lottaroli et al., 2018](#)).

The GTO recently supported projects applying the PFA technique to identify geothermal plays across the United States, including blind and EGS geothermal systems ([EERI, 2014](#)). Each study developed its own methodology for defining the primary geothermal play risk elements, quantifying uncertainty, and generating a favorability map ([Faulds et al., 2019](#); [Jordan et al., 2016](#); [Nash et al., 2017](#); [Wannamaker, 2016](#)). Final numerical favorability scores were defined by a combination of risk elements, most often heat and permeability, with weights determined from data confidence and/or expert opinion ([Garchar et al., 2016](#)).

Machine Learning

Both joint inversion and PFA attempt to identify patterns from sometimes disparate data sets to identify and characterize geothermal resources. And both require expert guidance on the weighting of data inputs to create an integrated final product. Machine learning methods can instead determine the appropriate relative weights directly from the data, making results repeatable and open to continuous improvement as additional data become available. Advances in data-driven machine learning approaches for pattern recognition and prediction are a major part of a “digital transformation” in the earth sciences beginning in the late 2010s, driving significant change in geoscience training and application ([Gunderson et al., 2020](#)). National labs and academic programs are embracing the opportunity to apply machine learning to a variety of geothermal problems, with many federally-funded projects currently underway, e.g., image analysis for production-related ground deformation ([Cavur et al., 2021](#)), real-time prediction of induced seismic events ([Small, 2019](#)), and identification of faults from seismic data ([Gao et al., 2021](#)).

Supervised learning methods like regression, tree-based ensemble methods, and neural networks all need labeled example data, e.g., from wells or KGRA studies, for training predictive models. Unsupervised learning approaches like cluster analy-

sis can learn directly from the structure of unlabeled input data. Studies applying both machine learning methodologies are revisiting PFA investigations that already have curated and archived data sets. For example, the PFA for the Great Basin region of Nevada originally combined nine data sets (or “features”) by a grouping-and-weighting workflow to determine favorability for blind geothermal systems (Faulds et al., 2017). As more data were acquired and previous features transformed or refined, the data progressively grew to over 20 feature layers (Brown et al., 2020; Faulds et al., 2019). A proof of concept Artificial Neural Network (ANN) successfully reproduced the original PFA favorability map (Brown et al., 2020), and further efforts illustrated value in applying more advanced algorithms like Principle Component Analysis (PCA) paired with k -means clustering (Smith et al., 2021) and a probabilistic neural network for prediction with parameter uncertainty (Brown, 2021).

In another example, Bielicki et al. (2015) defined play fairways in Southwestern New Mexico (NM) using a combination of 12 geologic, geophysical, and geochemical features to describe fluid, heat, and permeability risk elements. A subsequent project expanded this data set to 20 features and used a semi-supervised PCA and K-means clustering framework to define KGRA-associated groupings (Pepin, 2019). The study found each KGRA cluster correlated strongly with four regional physiographic provinces, i.e. regions of unique physical geography, in Southwestern NM: the Basin and Range, Colorado Plateau, Mogollon-Datil Volcanic Field, and Rio Grande Rift (Pepin, 2019). A separate effort led by LANL tested an unsupervised learning method, non-negative matrix factorization with k -means clustering, using a 22-feature data set (Vesselinov et al., 2020). This method determines feature signatures for each cluster, and results suggest each physiographic province may have its own unique set of features that signal the presence of hidden geothermal resources.

In all studies conducted thus far for Southwestern New Mexico, geothermal favorability models provide a deterministic view of problem. This thesis reinvestigates the NM study area with a focus on the variety of uncertainties involved in a machine learning approach, as well as how those uncertainties can impact the final model results and choices made by geothermal project decision-makers.

2.2.5 Uncertainties

Machine learning methods typically create mathematical models of a system based on empirical evidence rather than a formalized physics-based approach. Three main types of uncertainty impact these models, and each should be assessed when weighing model results for project decisions in either exploration or production scenarios.

Measurement Uncertainty

Every data point is a measurement of an object or phenomenon susceptible to multiple sources of error. The environmental conditions, instrument calibration, resolution limitations, and human skill can all impact the final value obtained (Baird, 1962, p. 11–14). Measurement uncertainty defines the range within which the true measurement value lies. Expressed mathematically, $y = \hat{y} \pm ku_c$ where y is the true measurement value, $\hat{y}$ is the measured value, and ku_c is some factor times the estimate of the standard deviation of $\hat{y}$, also called the standard error (u_c). Under the assumption of a Gaussian distribution, $k = 2$ corresponds with a 95% confidence level and is a typical choice for reporting measurement uncertainty (NIST, 2021).

Parameter Uncertainty

Fitting a model to data fundamentally involves estimating the values for a set of model parameters b_i , $i = 1, \dots, n$, where the total number of parameters can vary from one (e.g., the average value) to over one million for weights in deep neural networks. The degree with which the $\hat{b}_i$ values match the true parameter values, b_i , depends on the quality and amount of input data used for model training (James et al., 2013, p. 81). This type of uncertainty is evaluated using probabilistic methods and can be effectively reduced with the addition of more data.

Structural Uncertainty

Models represent simplified approximations of real systems, which respond to and interact with a myriad of other systems. Reducing a system down to its essential

complexity keeps it within the bounds of human cognition while also delivering an objective level of descriptive or predictive ability (Crawley et al., 2015, p. 306). But even the most elaborate system model does not capture a fully accurate or complete depiction of real-world system behavior. Instead, a model choice is a trade-off between the validity of the model results and the effort required to build and interpret the model (Morgan, 2009, p. 23). Fundamentally, the uncertainty in model structure requires examining how results change as the structure changes.

This thesis considers an approach where all three types of uncertainty are directly evaluated to understand their impact on geothermal predictions. Forecasts are a product of the data, model parameters, and the variety of model techniques applied, so examining where and to what degree these factors influence results fundamentally establishes the value of the forecasts for a geothermal project.

2.3 Power Generation

Geothermal facilities share similarities with many other methods of energy capture. The primary value function for a facility is producing power, further specified as a turbine generating electricity (Figure 2-7). This solution-neutral function could be specified into a variety of concepts that involve different operands — wind, solar, water, hydrocarbons — to deliver the same result. For geothermal, the common concepts include steam plants, flash plants, and binary cycle plants. Each concept has a range of temperatures and typical depths over which it is best suited to operate.

Additional applications for geothermal are illustrated in Figure 2-8, including heat pumps and direct use, however those concepts abstract to a different functional intent than energy production. Also shown are specific types of geothermal systems like hot sedimentary basins, co-production from non-geothermal wells, and EGS. Note that EGS could extend shallower than depicted for tight reservoirs, as proposed in Chapter 4, but deeper wells ($> 3\text{--}5$ km) become more challenging to justify primarily due to cost of drilling (J. N. Moore & Simmons, 2013).

Although the thermodynamics of the geothermal system go beyond the intended

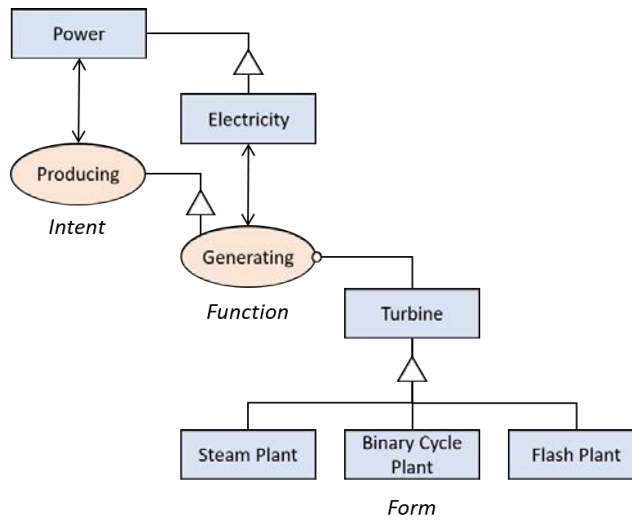


Figure 2-7: Mapping solution-neutral functional intent to geothermal power plant concept.

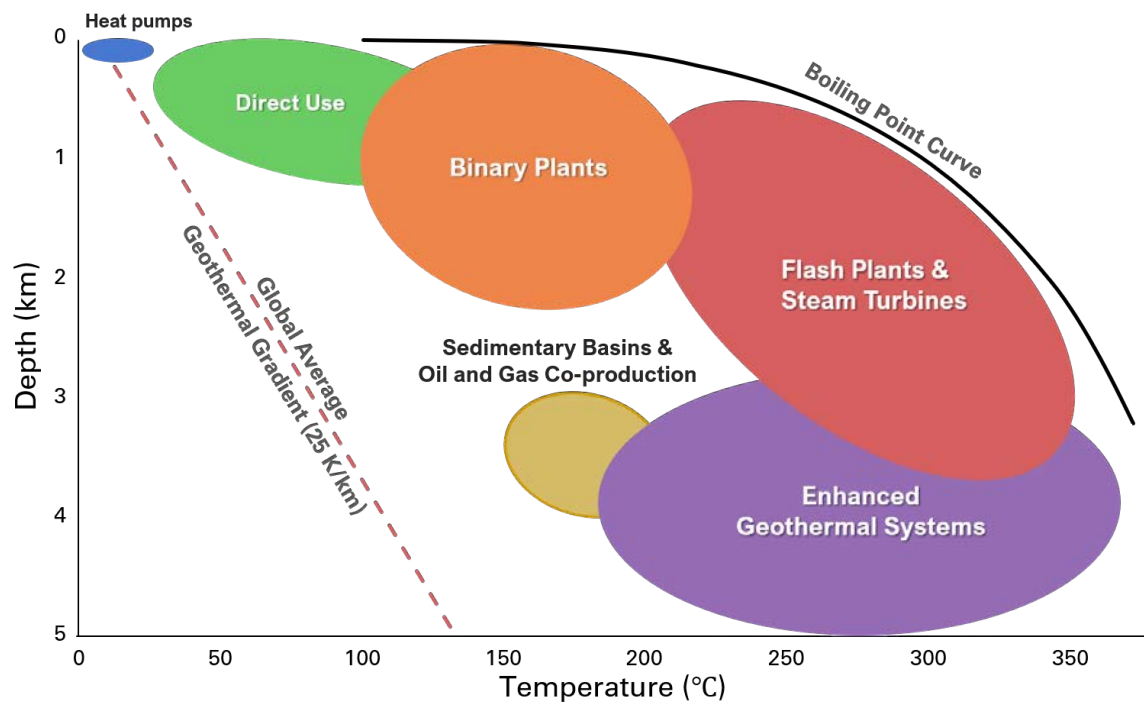


Figure 2-8: Traditional temperatures and depths associated with different geothermal applications. Boiling point defines the upper temperature limit for subcritical systems. Ellipses are approximate. Figure adapted from (J. N. Moore & Simmons, 2013).

scope of this thesis, the basics for how heat becomes electricity are worth visiting. Geothermal plants use a working fluid in vapor form to spin the turbine that generates electricity. The enthalpy of the working fluid defines the energy available to perform work. Fluid enthalpy in the surface plant is less than that of the reservoir due to losses in the system (Glassley, 2015, p. 204). Produced fluid loses heat energy when traveling up the production well and can also suffer frictional losses, resulting in a small drop in enthalpy, oftentimes ignored due to its relatively low impact (Glassley, 2015, p. 204). Other sources of enthalpy loss depend on the specifics of the system, as discussed in the following overview.

2.3.1 Direct Steam Power Plant

Direct steam power production dates back to the first power plant in Larderello, Italy and is the method for energy capture at The Geysers field as well (DiPippo, 2012, p. 131-132). This type of system involves dry steam with no fluid secondary phase that could otherwise remove enthalpy from the vapor as it separates at lower pressures. As a result, dry-steam power plants are the simplest geothermal plants to engineer, and the produced vapor delivers the highest energy per kilogram of the different geothermal systems (Glassley, 2015, p. 205).

Power generation performance of a steam turbine primarily depends on the difference in enthalpy between the steam entering and exiting the turbine, discounted by the efficiencies of the turbine (generally $> 85\%$, Glassley, 2015, p. 206) and the generator ($\approx 98\%$, Augustine, 2009, p. 116). Steam exiting the turbine is cooled in a condenser to liquid form before reinjection back into the ground. Figure 2-9 illustrates a simple schematic of how a dry-steam system works.

2.3.2 Flash Power Plant

Reservoirs that produce wet steam with sustained wellhead temperatures of over 200°C commonly use flash plants for power generation (J. N. Moore & Simmons, 2013). Wet steam refers to steam with some component of entrained fluid, different

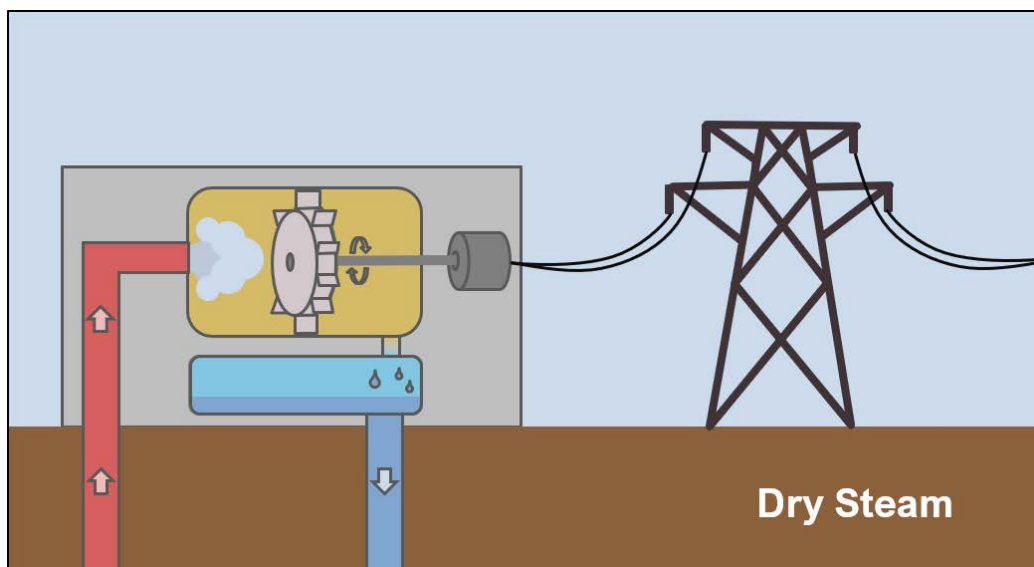


Figure 2-9: Direct steam power plant schematic. Steam is produced directly from the subsurface, passed through a turbine to generate electricity, cooled and condensed to a fluid, and reinjected. The condenser usually connects with a cooling tower for sufficient cooling before injection (not shown).

from the dry steam scenario with 100% vapor. Fluids in these systems “flash” to vapor when produced and thus require both fluid and vapor management.

The steam quality, or steam-liquid ratio, strongly influences power production efficiency; for every percentage increase in liquid mixed in with the steam, turbine efficiency degrades by 0.5–1.0% (Baumann Rule, [Glassley, 2015](#), p. 207). In addition, the separation of the produced fluid into liquid and vapor also partitions the enthalpy. Therefore, maximizing the amount of steam produced and removing the liquid phase before steam enters the turbine are key factors for operating a flash system ([Glassley, 2015](#), p. 215–216). A cyclone separator is commonly installed between the wellhead and turbine inlet to manage the latter issue. This component siphons off the fluid phase such that power production beyond the separator works the same as a dry-steam plant ([DiPippo, 2012](#), p. 88).

Dual- and triple-flash systems operate the same way as single-flash but add sequential secondary and tertiary flashing and separation at incrementally lower temperatures and pressures. This results in greater energy extraction — 20–30% above single-flash for dual, and even more for triple ([Glassley, 2015](#), p. 216). Figure 2-10

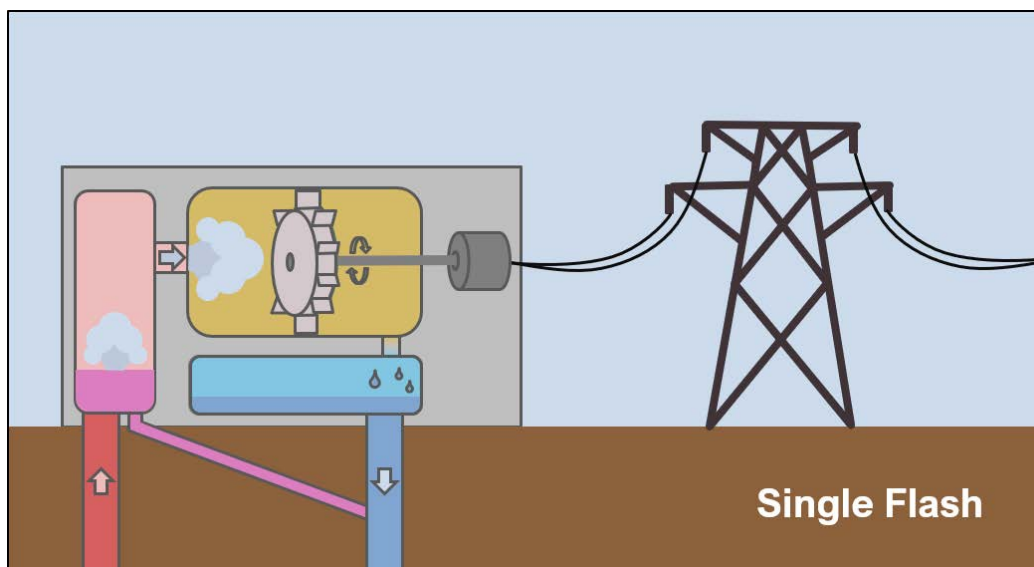


Figure 2-10: Single flash power plant schematic. Fluids produced from the subsurface flash to vapor in a cyclone separator. Residual fluid is diverted for reinjection, while the vapor enters the turbine to generate electricity, cools in the condenser, and is also reinjected. The condenser usually connects with a cooling tower and/or an air-cooling system for sufficient cooling before injection (not shown).

illustrates the mechanics of a flash system.

2.3.3 Binary-Cycle Power Plant

When production temperatures fall below 200 °C, systems relying on steam generation from the produced fluids alone will no longer perform with reasonable efficiency. Binary systems fill this gap by adding a heat exchanger and secondary organic working fluid with a lower boiling point than water ([J. N. Moore & Simmons, 2013](#)). The second fluid cycle is closed loop, meaning the subsurface “brine” — here, referring to produced fluids with any entrained particulates or corrosive chemistry — never interacts with the turbine. Heat from produced brine instead flashes the working fluid to vapor and the secondary fluid drives the turbine. This process is known as an Organic Rankine Cycle (ORC) and commonly involves fluids like isobutane (boiling point −12 °C) and isopentane (boiling point of 28 °C) ([Glassley, 2015](#), p. 219). ORC facilitates vapor creation at lower production temperatures, but it also introduces another point of enthalpy loss to the power generation process. Figure 2-11 depicts

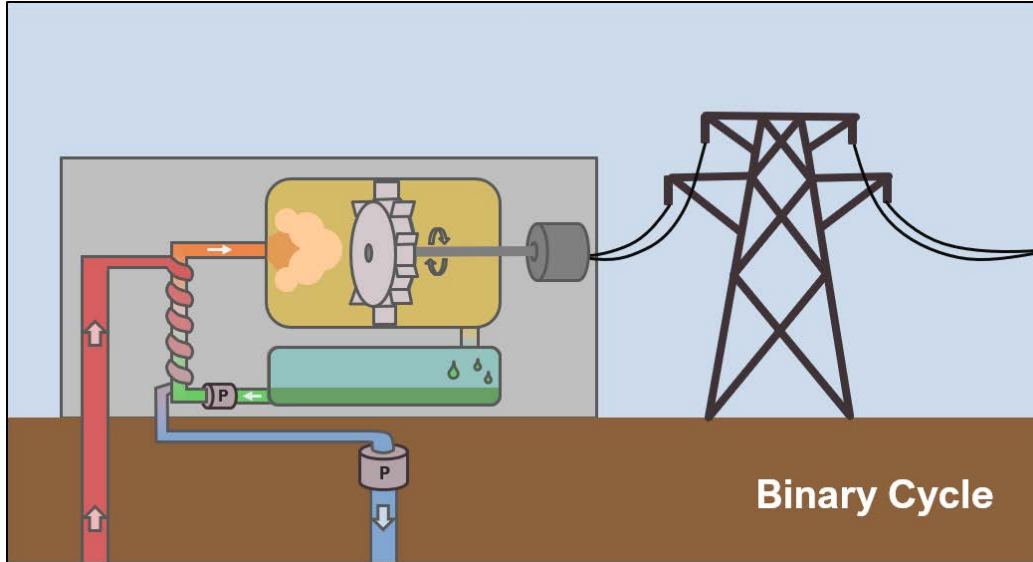


Figure 2-11: Binary cycle power plant schematic. Brine produced from the subsurface flashes a secondary working fluid to vapor, which enters the turbine to generate electricity, condenses back to fluid, and is re-flashed in a closed loop. Produced brine is similarly cooled and reinjected. Pumps are used to maintain flow in both cycles. Cooling is typically facilitated by a fan array (not shown).

the mechanics of a binary cycle geothermal system.

Because the natural drive of fluids from subsurface to surface diminishes at lower temperatures, and mineral deposition in the wells (scaling) can also be an issue, pressure must be managed with downhole pumps (DiPippo, 2012, p. 153). A pump also regulates the flow of the secondary fluid. These pumps add another efficiency factor on the system while acting as parasitic loads on power generation, reducing the net power output of the plant (Lowry, Foris, et al., 2017, p. 4).

2.4 Geothermal Cost Modeling

One might argue that cost estimates for future power plants could be derived from plants already in operation. After all, analog databases serve as powerful tools for deriving group estimates or empirical relationships for other complex processes like drilling wells (Lukawski et al., 2014; Tester et al., 2006). But data-driven estimates require a reasonable amount of data to serve as constraints. Consider a list of 96 U.S.-based geothermal power plants accessed through NREL’s Geothermal Prospector

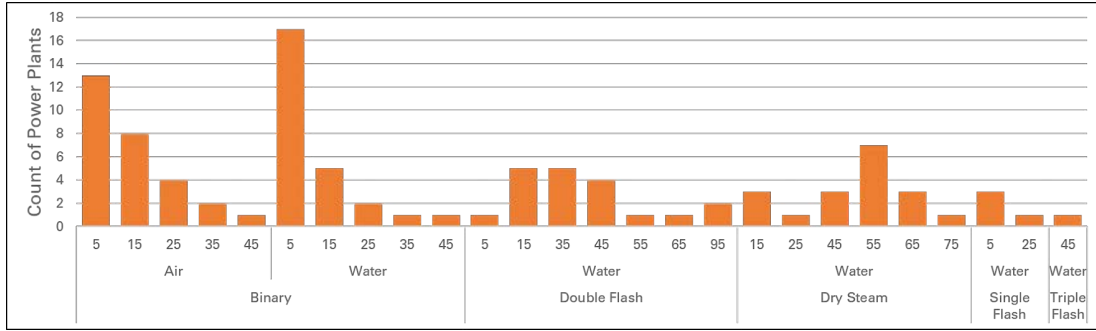


Figure 2-12: Count of U.S. power plants aggregated by conversion type (binary, steam, flash), cooling system (air, water), and binned plant nameplate capacity rounded to the nearest 5 MW. Data from NREL Geothermal Prospector (NREL, 2021a).

(NREL, 2021a). Figure 2-12 shows the results of binning the plants by conversion type, cooling type, and capacity. Adding additional filters like reservoir temperature, well depth, number of wells, and pump installations, and the number of valid analogs drops to either single digits or none for any particular plant definition. No commercial EGS plants are currently in operation within the U.S., so operational EGS analogs do not exist. Geothermal production planning must therefore rely on model-based estimates. The rest of this section will focus specifically on cost modeling of EGS as this is the topic of the case study described in Chapters 4 and 6.

Prior to the seminal report on “The Future of EGS” (Tester et al., 2006), cost models for EGS followed simple approaches in determining order of magnitude estimates for a project. Investigators chose not to directly model profitability, focusing instead on the relative impact model parameters had on the feasibility of an EGS project (Augustine, 2009). Importantly, models were able to identify optimal reservoir depths and design temperatures for a given geothermal gradient range by balancing the costs of drilling with costs for constructing the power plant (Tester & Herzog, 1990). More sophisticated models were built off of this early work.

2.4.1 GEOPHIRES

The EGS model first developed by Tester & Herzog (1990) and refined for “The Future of EGS” study (Tester et al., 2006) became known as the MIT EGS model, a

Windows application written in FORTRAN for economic analysis. In developing the code, the authors defined built-in correlations for various cost elements (e.g., drilling, plant construction, reservoir stimulation) based on empirical relationships they derived from available global data (Tester et al., 2006). After an additional upgrade to model direct-use geothermal and combined heat & power (cogeneration), the application was re-branded “GEOthermal energy for the Production of Heat and electricity (*IR*, standing for current $\times$ resistance) Economically Simulated” or GEOPHIRES (Beckers et al., 2013). Users can perform an analysis for a power plant with EGS or optimize on the design and drilling depth for minimized levelized cost of electricity (LCOE). GEOPHIRES was revamped in 2019 with the code refactored to Python, open source distribution, and extendibility to pair with external simulators like those for surface equipment or reservoir performance (Beckers & McCabe, 2019). Fundamentally, GEOPHIRES operates from a deterministic set of input parameters grouped into seven categories: resource, engineering, reservoir, financial, capital costs, operations & maintenance costs, and optimization settings (Beckers et al., 2013). Incorporating uncertainty in the modeling process requires incrementally re-parameterizing the input and running the model again. Strategic choices during the modeled lifecycle of the power plant are not supported.

2.4.2 GETEM

The primary alternative to GEOPHIRES is the Geothermal Electric Technology Evaluation Model (GETEM), a spreadsheet model that determines LCOE for commercial geothermal power production. GETEM was created for NREL by Princeton Energy Resources International and released in 2005 as a tool for the DOE to prioritize geothermal projects and test the economic impact of technology improvements (Entingh et al., 2006). GETEM offers a tremendous number of user inputs, but default values can reduce user interaction to defining resource temperature, depth, and a conversion system choice of binary or flash. An upgrade in 2011 included support for EGS reservoirs (EERE, 2012). Users can target a specific production amount for power sales or set a fixed count of production wells for the calculation (G. Mines, 2008). The

main project components include: cost for exploration, drilling and stimulation, well and reservoir management, and power plant construction and maintenance (Entingh et al., 2006). The design of GETEM as an Excel-based tool makes it transparent and configurable, however casual users could find its many worksheets and complex cross-references overwhelming. In addition, password protections lock down many of the sheets from full visibility or editing. As with GEOPHIRES, GETEM does not natively support probabilistic modeling or dynamic decision-making in the cost simulation.

2.4.3 SAM

NREL elected to include GETEM logic in their System Advisor Model (SAM) for multiple renewable energy systems, available both online and as a downloadable application (NREL, 2021b). SAM goes beyond power sales to a utility, modeling both residential projects to offset electricity needs and third-party ownership arrangements (Blair et al., 2018). SAM is also open source, but since the majority of the code was written in PowerBuilder and C/C++, incorporating custom logic or strategic decisions into cost calculations requires at least a moderate level of software development proficiency. Uniquely, SAM supports Monte Carlo simulation with multi-valued input variables (Blair et al., 2018). None of the other models reviewed here incorporate uncertainty into a geothermal economic model in this way.

2.4.4 CREST

On the other side of the spectrum is the Cost of Renewable Energy Spreadsheet Tool (CREST) developed in 2010 for NREL by Sustainable Energy Advantage (Gifford & Grace, 2013). This tool targeted state policy-makers interested in crafting renewable energy policies for solar, wind, and geothermal, aiming for ease of use over heavily-parameterized solutions like GETEM (Gifford et al., 2011). Rather than diving deeply into power plant performance, the model focuses primarily on financing, market factors, taxes, and incentives. Key cost buckets include exploration, confirma-

tion (appraisal) well drilling, well field and power plant construction, and operations & maintenance (Gifford & Grace, 2013). The spreadsheet form and standardized design makes CREST highly accessible, but the model simplifies geothermal to such a degree that conversion system or resource type are fully abstracted from the user.

2.4.5 Cost Model Insights

Several geothermal cost-model options are widely available and accessible from researchers and U.S. government sources. While they differ in target audience and capabilities, the models capture similar inputs and produce levelized cost estimates for electricity generation from hydrothermal and/or EGS reservoirs. Uncertainty is rarely accounted for in the models. Instead, the user parameterizes the resource and power generation scenario and typically receives a single cost estimate. Sensitivity testing or use of parameter ranges must be handled manually (except with SAM), and the models are incompatible with dynamic strategic decision-making over the lifetime of a field. The absence of these features defines an opportunity for a different modeling approach: one that accounts for uncertainty, allows for strategic flexibility, and also comes packaged in a familiar and customizable form (see Chapter 4).

2.5 Case Study: Southwestern New Mexico

The area of interest (AOI) for the geothermal exploration work in Chapters 3 and 5 is a 37,600 square mile region of Southwestern New Mexico covered by nine counties: Cibola, Valencia, Catron, Socorro, Grant, Sierra, Luna, Dona Ana, and Hidalgo (Figure 2-13). This region marks the juxtaposition of four significant geologic provinces. The Southern Basin and Range (SBR) extends across the lower third of the AOI. To the east lies the Rio Grande Rift (RGR), traced today by the course of the Rio Grande river. The Colorado Plateau (CP) covers the north of the study area, and the central-west region is blanketed by the Mogollon-Datil Volcanic Field (MDVF).

The following is an overview of each province, followed by an in-depth look of the only active commercial power plant in NM, Lightning Dock. Cost modeling work in

Chapters 4 and 6 reference Lightning Dock as the parent facility to a proposed EGS expansion project.

2.5.1 Southern Basin and Range

Plate tectonic activity along the western edge of the United States transitioned ≈ 30 Ma from widespread subduction to the present-day split between transform motion along the San Andreas Fault and subduction off the Pacific Northwest (Fowler, 2005, p. 81). This transition created a broad extensional regime in the Southwestern U.S. believed to be responsible for the alternating narrow, fault-bounded mountain-and-valley signature of the Basin and Range (Henry & Aranda-Gomez, 1992). Successive north-south striking normal faults in the province level out with depth, creating asymmetric graben structures (Frisch et al., 2011, p. 28-29). Cumulative extension has reduced crustal thickness in SBR to 30–35 km, with associated enhanced volcanism, geothermal gradient, and heat flow throughout the province (Lerch et al., 2007).

2.5.2 Rio Grande Rift

Even greater extension has taken place within the RGR province, a ≈ 1000 km long zone separating the Great Plains (GP) to the east and Colorado Plateau to the west. Rifting occurred in at least three stages: initiation ≈ 36 Ma, rapid increase in extension ≈ 28 Ma as part of the Basin and Range formation, and localized thinning between ≈ 3 –10 Ma (Bielicki et al., 2015; Mack et al., 2008; Seager et al., 1984). Mini-basins chained together along the rift show an alternating asymmetry, with transfer faults and accommodation zones separating successive basins. The faults bounding and connecting these basins could create favorable structural settings for geothermal systems (Faulds & Hinz, 2015). High heat-flow measurements in the RGR suggest geothermal gradients that, upon extrapolation, would exceed the solidus at the crust-mantle boundary (Olsen et al., 1987). This can be explained by a thermal anomaly with asthenospheric convection beneath the rift center (Olsen et al., 1987). Additionally, seismic and gravity data show crustal thinning to ≈ 30 km, with even

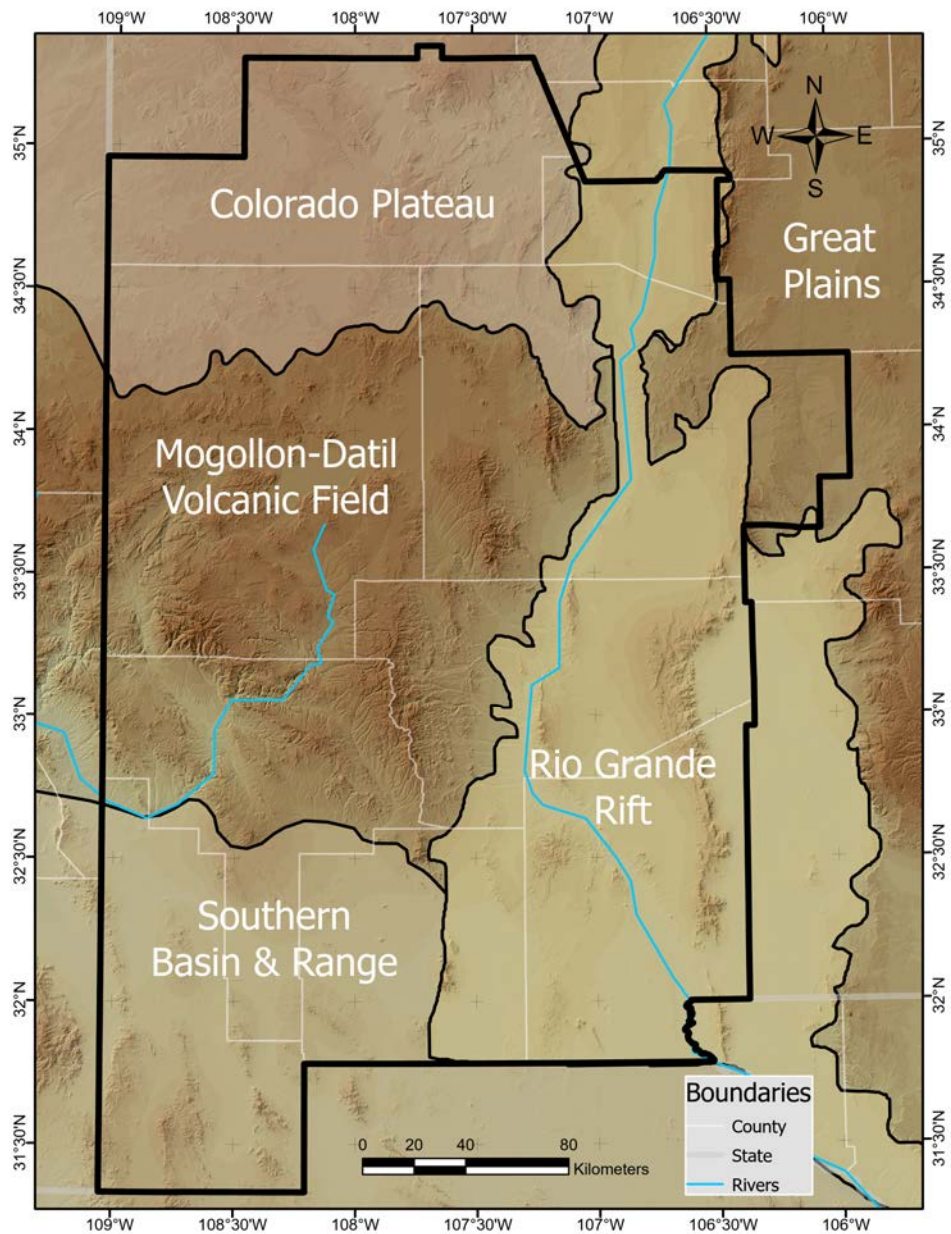


Figure 2-13: Physiographic provinces in the Southwestern New Mexico study area. The thick black line defines the AOI. Thinner black lines outline the province boundaries. County boundaries shown in light white lines. Province outlines from ([Bieliński et al., 2015](#), Figure 2-2).

greater thinning to the south (Keller & Baldrige, 1999). In summary, geologic and geophysical observations collectively support the likelihood of heat and permeability risk elements being met in the RGR province.

2.5.3 Colorado Plateau

The CP province presents a very different geologic picture, one of stability and lack of significant deformation for around 600 million years (Leighty, 1997). Uplift of the province took place over several different phases, beginning with the Laramide orogeny (40–80 Ma) and totaling more than 2 km of vertical offset relative to sea level based on exposed outcrops (Moucha et al., 2009). Unlike the surrounding provinces, the CP acted as a cohesive block and still maintains a significantly greater crustal thickness (≈ 45 km) compared to the SBR or RGR (D. Wilson et al., 2005). Recent models suggest CP uplift continues today, as complex replacement interactions occur between the denser, brittle lithosphere and more buoyant underlying asthenosphere (Levander et al., 2011). However, lower heat-flow values compared to the surrounding provinces (Thompson & Zoback, 1979) suggest these crust-mantle dynamics have little effect on the relatively low CP geothermal potential.

2.5.4 Mogollon-Datil Volcanic Field

On the western side of the AOI lies the MDVF, a 15,000 square mile outpouring of rhyolitic flows as part of a super-eruptive volcanic episode preceding rift initiation in the RGR province (Keller & Baldrige, 1999). The timing indicates the thermal source for MDVF magmas likely originated from a Farallon subduction-related event rather than onset of extension, and extrusive activity only represents a fraction of the total magma volume in the underlying composite pluton (Olsen et al., 1987; Schneider et al., 1994). MDVF is just one of several Late Eocene-Oligocene volcanic fields in a chain from Colorado through central Mexico, and isotopic dating defines four pulses of surface activity beginning 36 Ma near Las Cruces, NM and ending conclusively 24 Ma after a general westward migration (W. C. McIntosh et al., 1992). Lack of consistent

trends in current heat-flow measurements across the field results from the heterogeneous distribution of volcanic features and extrusive volumes (W. C. McIntosh et al., 1992). Nevertheless, evidence for greater water availability and high geothermometry values make MDVF worth considering for geothermal exploration (Pepin, 2019).

2.5.5 Lightning Dock

Present-day commercial geothermal activity in New Mexico is limited to a single hydrothermal power plant in Hidalgo County, the southwestern “boot-heel” of the state (Figure 2-14). The power plant, traditionally known as Lightning Dock and recently renamed the Dale Burgett Geothermal Facility, sits atop a KGRA documented by the USGS in 1974, $\approx$ 20 miles southwest of Lordsburg, NM (Dahal et al., 2012). Although the hydrothermal resource was originally blind, boiling water was encountered in 1948 by drilling operations for water wells (Elston et al., 1983). Commercial use of the resource dates back to 1977 to support greenhouse facilities and later, in the 1990s, for aquaculture (Crowell & Crowell, 2014). Power production commenced in 2013 under the ownership of Cyrq Energy with a 4 MW air-cooled binary cycle power plant (Goodman & Smiley, 2013). Public Service Company of New Mexico (PNM) signed a long-term Power Purchase Agreement (PPA) for the generated electricity (Dahal et al., 2012), as encouraged by the New Mexico Renewable Portfolio Standard (RPS) carve-out for 5% non-wind, non-solar renewable technologies (DSIRE, 2021). Lightning Dock was fully revamped in 2018 by Turboden to its present 11.2 MW (net) single-turbine binary cycle architecture (Bonafin & Dickey, 2019). The 20-year PPA with PNM was reset and secures power sales through 2038 (O’Connell, 2018).

The Lightning Dock facility is on the eastern side of the Animas Valley, which lies at the northern extent of the Mexican Highland as part of the SBR (Cunniff & Bowers, 2005). The bounding ranges are the Pyramid Mountains to the east and Pelocillo Mountains to the west (Figure 2-14). In typical SBR fashion, this surface geography corresponds with a fault-bounded graben subsurface architecture. Lightning Dock employs the bounding Animas Valley fault to access deep-sourced hydrothermal waters for power production. Interestingly, this fault ties to the ring

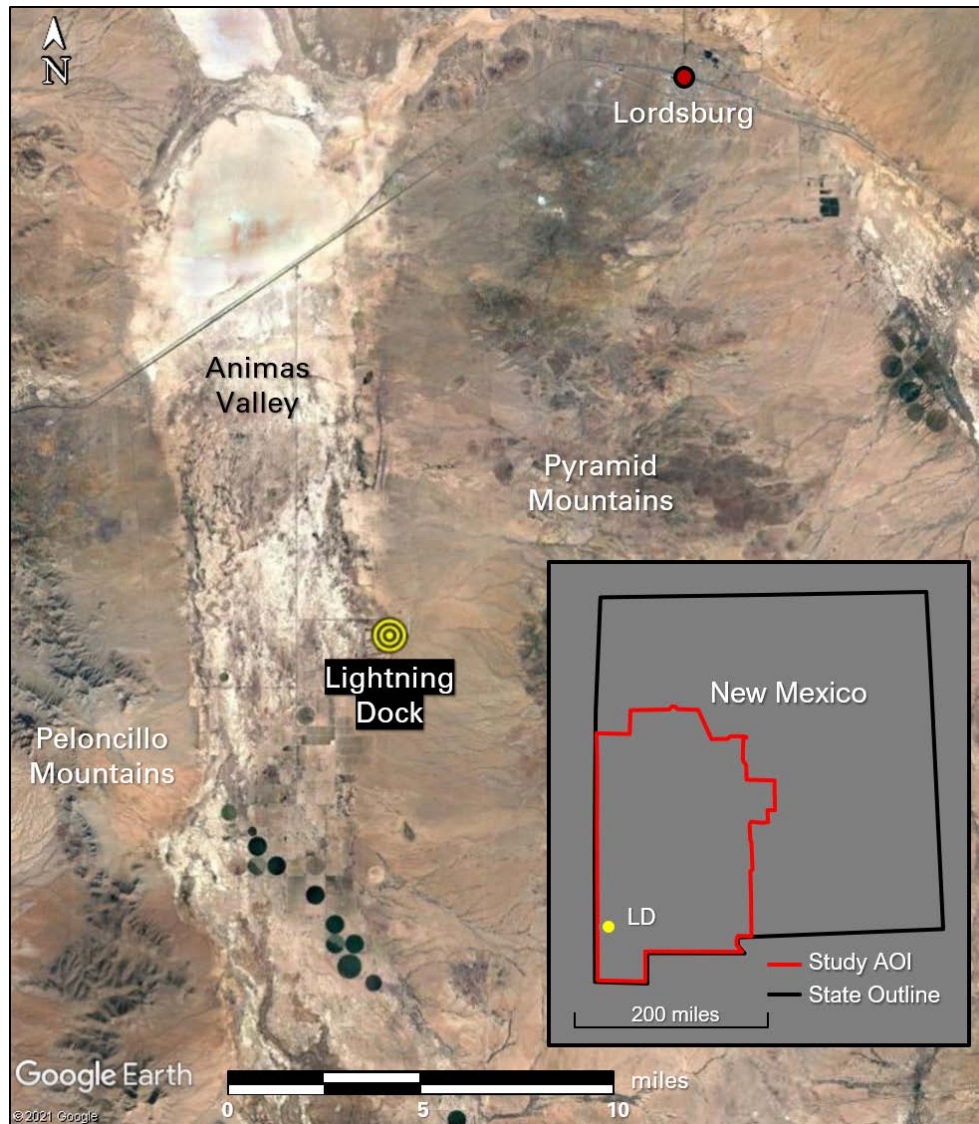


Figure 2-14: Map location of the Lightning Dock power plant. Inset map shows the facility in relationship to the state of NM and study area of interest for Chapters 3 and 5 of this thesis. Map created using Google Earth.

fracture zone of a 20-km wide Oligocene volcanic feature known as Muir Cauldron (Elston et al., 1983). The exact heat source responsible for hydrothermal activity remains in question, but geochemical analysis suggests circulation of 250 °C waters from depths of 6–8 km (Schochet & Cunniff, 2001). These fluids upwell within the fault zone and intermingle with cool groundwater fed by runoff from the neighboring highlands, creating a lower-temperature 150–170 °C brine targeted by Lightning Dock (Crowell & Crowell, 2014).

Prior to the construction of Lightning Dock, Schochet & Cunniff (2001) submitted a proposed development plan for a hybrid hydrothermal-EGS project to the DOE. Similar to the Lightning Dock design, hydrothermal fluids would be produced from shallow depths (< 1 km). In addition, the ≈ 600 m thick Horquilla Limestone formation could be treated as an EGS reservoir to supplement the hydrothermal power production and serve as a proof of concept for commercial EGS (Schochet & Cunniff, 2001). The proposal provided a cost analysis for two possible power-plant sizes using most-likely values for variables and a single deterministic estimate of total plant and field cost (Table 3, Schochet & Cunniff, 2001). The proposal was never realized, but the concept of combining the Lightning Dock power plant with an EGS expansion is an intriguing modern-day extension of Schochet and Cunniff’s vision. However, such a project would come with many financial risks tied to operations, resource management, and external market factors. This thesis considers how that risk might be mitigated through cost modeling that accounts for uncertainty in the model parameters and the potential for just-in-time strategic decision-making. Such an approach goes beyond the cost models described in Section 2.4 and the simple cost analysis performed by Schochet & Cunniff (2001), treating the project as a system with its own dynamics and emergent financial results.

2.6 Recap

This chapter provided relevant background information and a literature review on the topics of geothermal energy, geothermal exploration, and geothermal cost modeling.

Key insights from this chapter include:

1. Geothermal energy exists everywhere around the world and is primarily sourced from accretionary heat and radioactive decay.
2. Conventional geothermal (hydrothermal) systems consist of a heat source, permeability, circulating fluids, a top seal, and source of fluid recharge.
3. Unconventional geothermal (Enhanced Geothermal Systems) requires artificial creation of permeability and/or fluid circulation to capture subsurface heat.
4. Geothermal exploration historically relies on geologic field mapping, geochemical indicators, and geophysical measurements.
5. Play Fairway Analysis integrates disparate observations and measurements related to key geothermal risk elements into an overall map of geothermal favorability.
6. An opportunity exists to apply supervised and unsupervised machine learning methods to geothermal exploration, particularly as an extension of the PFA concept. Existing approaches have under-explored the influence of uncertainties on model prediction and reliability.
7. Geothermal power-plant conversion types include high-temperature resource steam plants, moderate to high-temperature resource flash plants, and low-to moderate-temperature resource binary-cycle plants.
8. Existing geothermal economic models for power production estimate costs for a specified resource and power plant scenario. An opportunity exists to combine management of uncertainties, dynamic decision-making, and deployment using an open and adaptable platform.
9. Southwestern New Mexico consists of four unique physiographic provinces that have been well-documented in the past. This is the chosen study area for testing geothermal-exploration risk-mitigation strategies in Chapters [3](#) and [5](#).

10. Lightning Dock, a USGS-designated KGRA and active binary cycle hydrothermal power plant, lies within the Southwest New Mexico study area. This is the selected area for testing risk mitigation with cost models in Chapters [4](#) and [6](#).

Chapter 3

Geothermal Exploration with Machine Learning

The first of the two research questions presented in Chapter 1 addressed the topic of risk in geothermal exploration. As discussed in Chapter 2, exploration has traditionally relied on geologic, geochemical, or geophysical surveys, analyzed individually or collectively through qualitative means, joint inversion methods, or play fairway analysis (PFA). Recent applications of machine-learning methods to exploration show promise in considering a range of data sources all at once, for collective insights on geothermal favorability (see Section 2.2.4). This chapter describes a methodology for applying machine learning and uncertainty analysis to the study area in Southwestern New Mexico introduced in Section 2.5, as an investigation in data-driven risk-mitigation for geothermal exploration.

3.1 Data Sources

This investigation brings together a total of twenty-five (25) data sets covering the Southwestern NM study area (Table 3.1). Data were collected from previously published works, open-access databases, or derived from those original sources as secondary products. The form of the data varies between pre-gridded raster files, point data sets with repeat or overlapping measurements, non-overlapping point sets, and

polyline data. Previous research efforts produced the raster files or raster-ready gridded data (also called “features”) that comprise ten of the data sets. Three features are generated by running procedures on one of the existing rasters. The remaining layers were created from polylines (5), co-located points (4), and distinct points (3). Although complex interactions between earth systems should be expected, these layers represent the “independent” or *predictor* variables for analysis purposes. Section 3.2.2 explains how evaluating feature correlations allows for pre-screening before modeling, and further analysis of feature importance in Chapter 5 helps hone in on the most influential features for simpler prediction models.

As discussed in Section 2.1.3, viable geothermal systems require permeability, heat, fluids, seal, and recharge. Following the more simplified PFA risk elements of Bielicki et al. (2015), an explorationist will want to first identify where fluids, heat, and permeability together define a favorable setting, then consider seal as a final screening factor. The chosen feature inputs collectively address the first three elements, as noted in Table 3.1. Rather than defining a “dependent” or *response* variable that describes a total favorability score, this thesis focuses on a proof-of-concept predictive workflow for a physically-measurable quantity associated with the heat risk element: geothermal gradient. Geothermal gradient estimates come from temperature gradient logs, a common component of well log suites. Gradient values are determined from regularly-sampled thermistor readings as a wireline tool is lowered down a borehole following a post-drill thermal recovery period (Prensky, 1992).

The choice of geothermal gradient was made because a) combined favorability scores are less straightforward to define and calibrate than geothermal gradient values, b) gradient point data are available from suitable compilations of well measurements covering the study area, c) geothermal gradient provides a direct proxy for accessible heat content, and d) for EGS applications, the crucial element that must be naturally present is heat. Heat flow might be a reasonable alternative response variable, however heat flow values in the available well database were derived from geothermal gradient and would introduce uncertainty stemming from conductivity estimation. Another alternative could be geothermometer measurements (see Section

No.	Name	Type	Risk Element	Source	Source Location
1	Average Air Temperature	Raster	H	PRISM (Oregon State University)	https://prism.oregonstate.edu/normals/
2	Average Precipitation	Raster	F	PRISM (Oregon State University)	https://prism.oregonstate.edu/normals/
3	Basement Depth	Raster	H	OpenEI Submission	https://gdr.openet.org/submissions/597
4	Boron Concentration	Overlapping Points	F,H	OpenEI Submission	https://gdr.openet.org/submissions/597
5	Crustal Thickness	Lines	P,H	Keller et al., 1991, Figure 4	https://www.sciencedirect.com/science/article/abs/pii/0040195191900503
6	Drainage Density	Lines	F	OpenEI Submission	https://gdr.openet.org/submissions/597
7	Earthquake Density	Overlapping Points	F,P	NM Bureau of Geology and Mineral Resources, USGS	https://geoinfo.nmt.edu/repository/index.cfm?rid=20020001 https://geoinfo.nmt.edu/repository/index.cfm?rid=20060001 https://geoinfo.nmt.edu/repository/index.cfm?rid=20130001 https://earthquake.usgs.gov/earthquakes/search/
8	Gamma-Ray Dose Rate	Raster	H	USGS	https://pubs.usgs.gov/of/2005/1413/datafiles.htm
9	Geodetic Strain Rate	Raster	P	Global Strain Rate Map (UNV Reno)	http://geodesy.unr.edu/GSRM/
10	Gravity Anomaly	Raster	P,H	OpenEI Submission, USGS	https://gdr.openet.org/submissions/597 http://cybershare.utep.edu/dataset/gravity-dataset-united-states-lower-48-states
11	Gravity-Anomaly Gradient	Raster	P	N/A	Generated in ArcGIS
12	Heat Flow	Nonoverlapping Points	H	Lucazeau, 2019	https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GC008389
13	Lithium Concentration	Overlapping Points	F,H	OpenEI Submission	https://gdr.openet.org/submissions/597
14	Magnetic Anomaly	Raster	F,P	OpenEI Submission, USGS	https://gdr.openet.org/submissions/597 https://mrdata.usgs.gov/magnetic/
15	Magnetic-Anomaly Gradient	Raster	F,P	N/A	Generated in ArcGIS
16	Quaternary Fault Density	Lines	P	OpenEI Submission	https://gdr.openet.org/submissions/597
17	Si Geothermometer Temperature	Overlapping Points	H	OpenEI Submission	https://gdr.openet.org/submissions/597
18	Spring Density	Nonoverlapping Points	F	USGS	https://waterdata.usgs.gov/nwis/inventory?search_criteria=state_cd&search_criteria=site_tp_cd&submitted_form=introduction
19	State Map Fault Density	Lines	P	OpenEI Submission	https://gdr.openet.org/submissions/597
20	Surface Topography (DEM)	Raster	P,H	OpenEI Submission, USGS	https://gdr.openet.org/submissions/597 https://viewer.nationalmap.gov/basic/#/
21	Topographic Gradient (Slope)	Raster	P	N/A	Generated in ArcGIS
22	Volcanic-Dike Density	Lines	H	OpenEI Submission, USGS	https://gdr.openet.org/submissions/597 https://my.usgs.gov/eerma/
23	Volcanic-Vent Density	Nonoverlapping Points	H	OpenEI Submission	https://gdr.openet.org/submissions/597
24	Water-Table Depth	Raster	F	OpenEI Submission	https://gdr.openet.org/submissions/597
25	Water-Table Gradient	Raster	F,P	OpenEI Submission	https://gdr.openet.org/submissions/597
D	Geothermal Gradient	Nonoverlapping Points	H	Southern Methodist University	http://geothermal.smu.edu/static/DownloadFilesButtonPage.htm

Table 3.1: List of data sets included in this analysis. Data type, source, and source location are noted. Suggested feature-sensitive risk elements include fluids (F), heat (H), and structure/permeability (P). Numbered features are treated as predictor variables. ‘D’ indicates the dependent or response variable. See Appendix A for details on how each feature GIS layer was constructed for modeling.

2.2.3) that characterize reservoir temperature, but the uncertainty in fluid pathways leading to the measurement location means these values suffer from less spatial and depth certainty than does geothermal gradient.

Regarding the remaining risk elements, permeability, fluids, and seal could be separately predicted using the methods described in the present study. A final favorability score could then be derived by combining the risk-element maps as is done in PFA risk assessments. This extended methodology is outside the scope of this thesis and thus appears in the list of future work opportunities in Chapter 9.

3.2 Data Preparation

Before experimenting with a variety of machine-learning methods, all input data sets must first be transformed into fully-complete Geographic Information System (GIS) layers such that any location within the study area has a corresponding set of predictor values. Steps taken to condition and process each layer are introduced in this chapter and detailed in Appendix A. The following section reviews several fundamental concepts and algorithms utilized in the preparation of the data layers.

3.2.1 Fundamentals

Extents

The data sets imported into ArcGIS and Python scripts for feature preparation required cropping, gridding, or less frequently, extrapolation to match each other in coverage of the Southwestern New Mexico study area. Two polygons were used for this purpose:

- **Regional Polygon:** this is a simple polygon capturing the broader Southwestern NM region, defined by the following corner points in latitude and longitude: $(-31.3^\circ, -109.1^\circ)$, $(31.3^\circ, -105.9^\circ)$, $(35.4^\circ, -105.9^\circ)$, $(31.3^\circ, -109.1^\circ)$.
- **Area of Interest (AOI):** this polygon appears in most map figures in this thesis (e.g., Figure 3-1) and is the perimeter of a nine-county block in Southwest

New Mexico that includes Cibola, Valencia, Catron, Socorro, Grant, Sierra, Luna, Dona Ana, and Hidalgo counties.

Point Mesh Grid

Machine-learning models generally treat data as a matrix of observations. The concept of a geospatial dataframe follows this paradigm, where each row represents a discrete location specified in latitude and longitude columns, and all other columns contain feature values for that location. Since the area outlined by the study AOI covers 97,469 km², and each location is described by 25 features and 2 map coordinates, the matrix of the study area at 1 km² resolution would be over 2.6 million values in size. This could be problematic since the time complexity of some machine-learning methods shows non-linear growth with data size, e.g., decision trees (Section 3.3.4) have $O(mn \log_2 n)$ complexity, where n is the number of observations and m is the number of features (Sani et al., 2018). As such, a coarser resolution of 0.025° ($\approx$ 2.5 km on average across the AOI) was selected as a manageable sampling interval for the present study. To support downsampling operations, the ArcGIS *Create Fishnet* tool was used to generate a 0.025° $\times$ 0.025° grid that, when constrained to the AOI polygon, consists of $n = 15,137$ point locations (Figure 3-1). Data-preparation steps use this grid where noted in Appendix A, and final model predictions are made on these points to enable direct comparison between different model results.

Density Estimation

One technique for converting a discrete set of points into a continuous field of values uses the concept of point density. The Kernel Density Estimation (KDE) method fits a smooth function (kernel) to each point with the constraint that the volume under a kernel surface is 1.0. Density values assigned to cells of a gridded surface or raster image represent the sum over the kernels crossing those cells. The kernel density tool can provide an estimate of density anywhere within the AOI, but it requires a kernel radius that sets the distance over which a point influences the density value. If left undefined, ArcGIS defaults to an optimal radius value based on a bandwidth formula

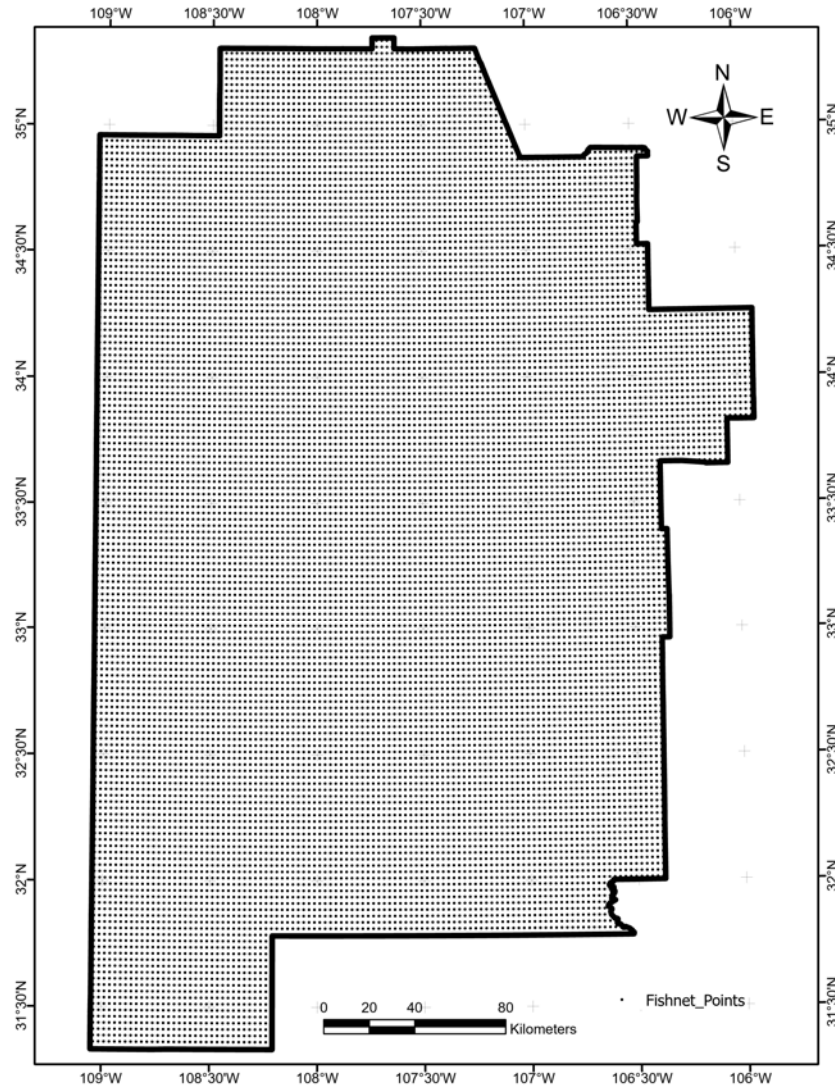


Figure 3-1: Mesh grid of points spaced 0.025° apart in latitude and longitude. The grid point set is used for data preparation and model predictions.

([ESRI, 2021b](#)). ArcGIS uses Silverman quartic kernel as the functional basis in the *Kernel Density* operation ([ESRI, 2021b](#); [Silverman, 1986](#)).

Polyline data can also be fit with smoothly-curved surfaces using the density concept. The maximum value of the elongate surface follows the trace of the polyline, and the surfaces decays in value away from the line over a distance determined by the radius, which defaults to the bandwidth estimate. The volume under the surface scales with line length. For more details on the ArcGIS implementation of polyline density estimation, refer to the *Density toolset* documentation on *Kernel Density* ([ESRI, 2021b](#)).

Surface Fitting

A multitude of algorithms exist for the purpose of fitting a set of values with a surface, thereby providing a means for interpolating missing values. Methods used in this thesis include the following four operations:

- **Splines** - Spline functions have a well-defined mathematical formulation that minimizes curvature while exactly fitting the input points. Regularized splines add a third derivative term, controlled by a weight parameter, that enforces a higher degree of smoothness with the trade-off of some data misfit. This is helpful if outliers in the input point set would otherwise distort the spline fit. For details on the ArcGIS implementation for splines, refer to the *Interpolation toolset* documentation on the *Spline* function ([ESRI, 2021d](#)).
- **Topo to Raster** - The ArcGIS *Topo to Raster* operation creates surfaces that honor the input data, typically contour sets, while ensuring i) correct representation of abrupt morphological features like rivers and ridges and ii) connectedness of drainage patterns. Specifically, *Topo to Raster* uses an iterative finite difference method and an algorithm to remove local minima not supported by the input data, assuming natural sinks are rare and erroneous features in an interpolated surface. For details on the ArcGIS implementation, refer to the *Interpolation toolset* documentation on the *Topo to Raster* function ([ESRI,](#)

[2021e](#)).

- **Ordinary Kriging** - Kriging methods are geostatistical algorithms that take the correlation distance and directional bias into account when creating a surface. The two tasks involved in kriging include first estimating the functions that characterize these spatial relationships, then using these functions to generate predictions for interpolation (or extrapolation). For the first task, semivariograms are calculated by binning the semivariance of points based their distance from each other, then fitting a model curve (e.g., linear, spherical, exponential) to these values. Semivariograms can be anisotropic, meaning they vary with direction. The second task uses the semivariogram to define a weighted average of the input data when estimating new values. For more details on the ArcGIS implementation of kriging operations, refer to the *Interpolation toolset* documentation on the *Kriging* function ([ESRI, 2021c](#)).
- **Empirical Bayes Kriging (EBK)** - The EBK algorithm extends ordinary kriging by automating parameter selections and accounting for uncertainty in semivariogram estimation. Ordinary kriging generates a single variogram and treats it as ground truth while EBK generates an ensemble of semivariograms that can more accurately estimate standard errors. As a computationally heavy method, EBK takes much longer to apply than other curve-fitting operations. For more details on the ArcGIS implementation of EBK, refer to the documentation on the *Empirical Bayes Kriging* function ([ESRI, 2021a](#)).

3.2.2 Data Conditioning

After building the various GIS data layers, several data conditioning steps were taken to explore variable relationships, rationalize feature choices, and prepare data values for modeling. Figure [3-2](#) outlines the general workflow followed, detailed more extensively below.



Figure 3-2: Workflow for conditioning data prior to predictive modeling.

Dataframe Creation

Feature GIS maps spanning the full AOI provide the primary data resource, but predictions of geothermal gradient require ground-truth observations for training a predictive model. Initially, two separate data sets could be constructed for further evaluation. The first is a Full Data Set (FDS) that consists of feature values extracted using the pre-defined point mesh grid (Figure 3-1), resulting in 15,137 records for the 25 predictors. This data set does not include values for the response variable. Although there is a geothermal-gradient feature layer described in Section A.27, this is just a derived layer for comparison and not ground truth. The second data set was built from actual geothermal-gradient observations in the SMU database described in Section A.27, combined with 25 feature values extracted from the GIS maps at the observation locations. This data set is hereafter referred to as Well Data Set (WDS). Both FDS and WDS were stored in “geodataframes” consisting of the feature values and their associated latitude and longitude coordinates. These structured matrices are compatible with machine-learning techniques like those found in the scikit-learn

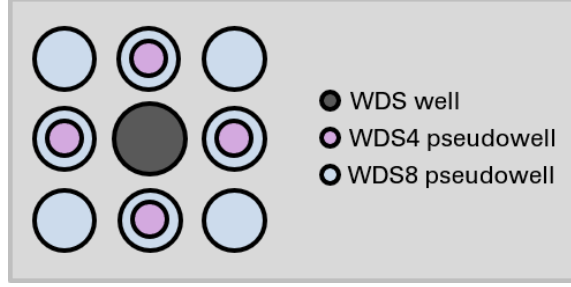


Figure 3-3: The data imputation strategy creates neighboring well locations a short distance away from each original well in the WDS (dark gray) and uses kriging to assign a geothermal gradient to these “pseudowells.” For WDS4, pseudowells (purple) are placed to the N, S, E, and W. For WDS8, pseudowells (blue) are placed at eight locations around the central well. Latitude and longitude offsets are $\pm 0.01^\circ$ for pseudowell placement.

([Pedregosa et al., 2011](#)) and TensorFlow ([Abadi et al., 2016](#)) Python packages.

Data Imputation

Noting the AOI under investigation spans over 97,000 km², the relatively small size of WDS (≈ 600 observations) raises concern over whether enough data are available to obtain data-driven insights using supervised learning methods. If predictions are reduced to just locations in the AOI mesh grid, there are still 2 orders of magnitude difference between ground-truth observations and the points being predicted, exacerbated further by the need to partition the input data into one subset for training and two others for model validation and testing as a machine-learning best practice (e.g., [Hastie et al., 2009](#), p. 222) (see Section 3.2.2).

Data-imputation methods can increase the size of a sparse data set by filling in for missing values using basic assumptions, heuristics, or even complex imputation models applied to the existing data ([Hastie et al., 2009](#), p. 332-333). The present study applies the concept of spatial autocorrelation at the heart of variography and kriging methods; in geography, all things are related, but the correlation usually increases as the spatial distance decreases ([Gimond, 2021](#), Chapter 13). For each well location in the WDS, an additional four points were placed to the north, south, east, and west by adding or subtracting a constant 0.01° to each well’s geographic coordinates (Figure 3-

3). Feature values were extracted from the layer maps in ArcGIS at these “pseudowell” locations. For the response variable, geothermal gradient, an interpolation layer was created from the WDS observations using the ArcGIS *Kriging* function with spherical variogram, auto-determined lag size of 0.097, and variable search radius with 12-point requirement. Gradient values were extracted from this layer for each pseudowell. The use of an interpolated gradient layer avoided conflict in pseudowell values for close-proximity real-well locations since step-out pseudowells for neighboring original wells could (and do) overlap. Although this method may introduce additional spatial correlation than present in the original data, the small step-out interval constrains that added correlation to a short distance from each well location. This overall workflow generated a new data set with 2,995 observations within the study AOI, referred to as WDS4.

Extending this method further, a second imputed data set placed pseudowells to the NE, SE, NW, and SW as well, resulting in eight pseudowells for every original well in the WDS (Figure 3-3). Restricting the results to the AOI, this produced a data set with 5,386 observations (WDS8) for use in training and testing of machine-learning models.

Data Exploration

The comprehensive coverage of FDS makes it an appealing data set to use for exploring the attributes and relationships of the 25 data layers. Although care was taken to ensure each GIS layer fully spanned the AOI, a search for missing values identified 163 NaNs (i.e., Not a Number, unassigned values) among the features. The corresponding data rows plot along the study area boundary and likely represent places where one or more data layers ended just short of the point locations in the AOI mesh grid. These rows were dropped from FDS, reducing its size to 15,007 records.

Histograms can offer insights on the data distributions of predictor variables. Based on Figure 3-4, only the magnetic-anomaly layer has the appearance of a zero-mean Gaussian distribution. All other variables are offset and skewed to some extent. Many statistical tools rely on the assumption of normally-distributed random vari-

ables, so non-normality can be problematic for modeling (Montgomery, 2012, p. 85). These results suggest that variable scaling and transformations will be a useful part of data preparation (Montgomery, 2012, p. 221).

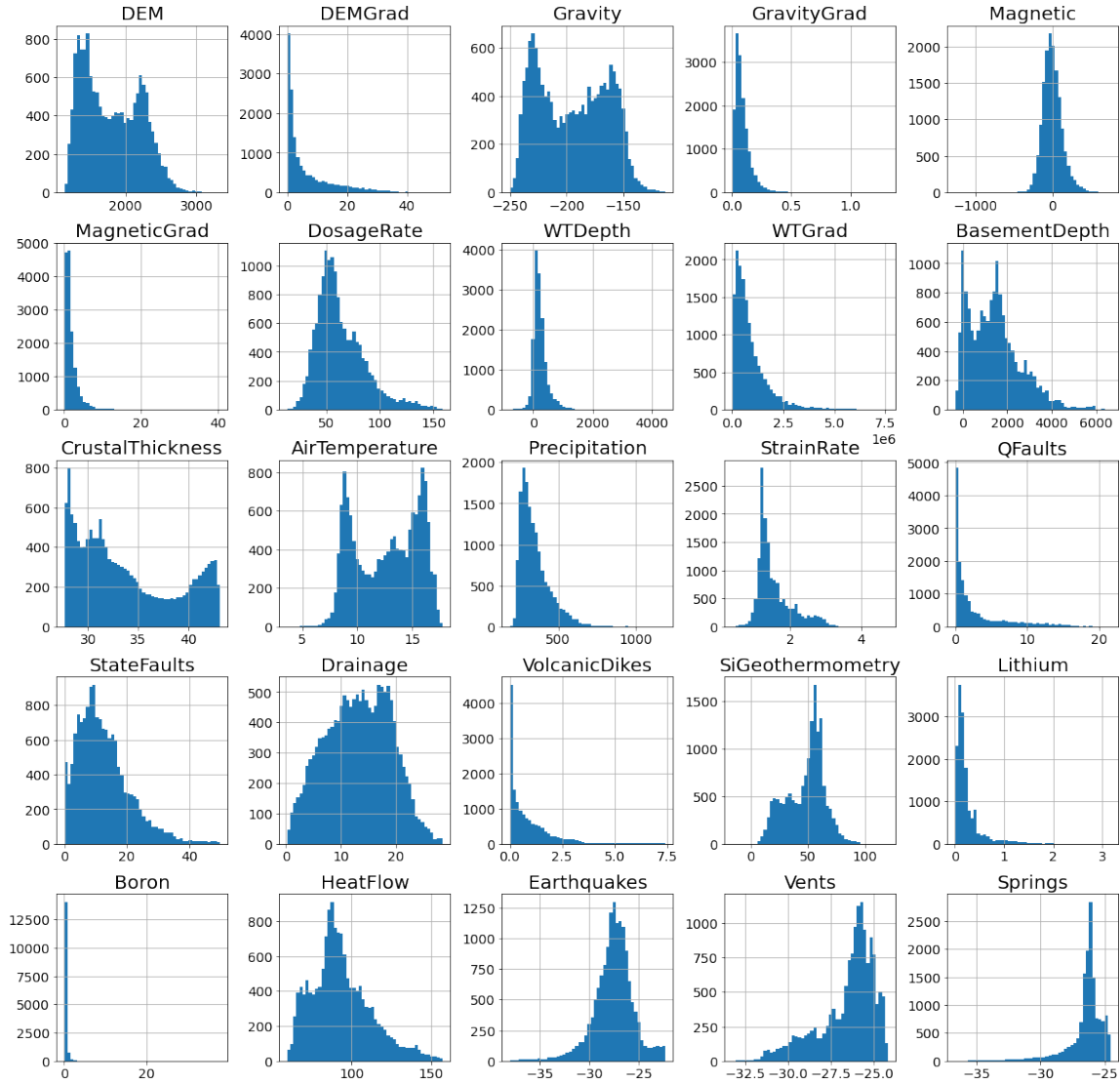


Figure 3-4: Histograms of the 25 features using 50 bins and FDS data. No scaling or transformations were applied to the data.

Scatter plots between variables can highlight collinear behavior where a close relationship between two predictors creates uncertainty in their balance, that is, their individual contributions to the response variable (James et al., 2013, p. 99). This can reduce the accuracy of model parameters like regression coefficients, impact the statistical significance of predictors, and lead to overly complex models (James et al., 2013,

p. 100-101). Figure 3-5 illustrates all permutations of feature pairwise relationships. Although individual plots are too small to appreciate in detail, the overall shapes of the plots suggests some linear behavior between a handful of variables. The inset maps illustrate two examples of collinearity, and the third shows how non-normality in variable distributions makes it difficult to discern some feature relationships.

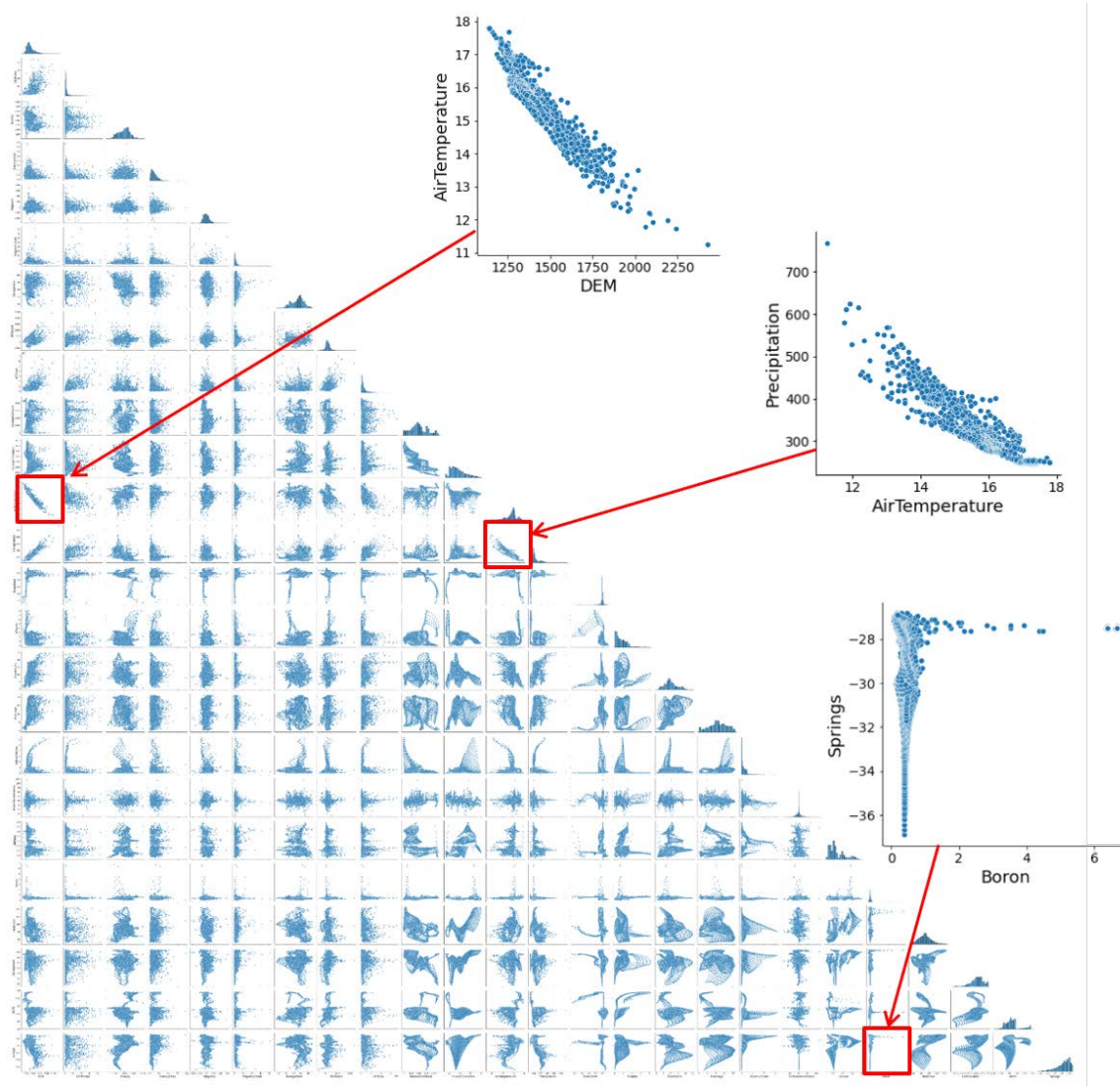


Figure 3-5: Scatter plots between all possible pairs of the 25 features. The upper two plot call-outs illustrate collinear relationships. The lowermost highlighted plot shows the impact of skewed distributions. Note the difference in axis ranges depending on the variable. All plots show the first 2,000 points of the 15,007-point FDS.

Feature Scaling

Large differences in the ranges and average values of the predictor variables are also evident in the scatter plots in Figure 3-5. For some machine-learning algorithms, variables with larger value ranges can have an out-sized effect on the model, so scaling variables to comparable ranges and removing variable bias is an important step in data conditioning. Scaling also makes a predictor closer to the standard normal in appearance, i.e., $Z \sim N(\mu = 0, \sigma = 1)$, as implicitly required by some statistical methods. The scikit-learn *StandardScaler* function transforms data using the Z-score formulation (scikit learn, 2021):

$$Z = \frac{x - \mu}{\sigma}, \quad (3.1)$$

where μ and σ^2 are the sample mean and variance of x . This data scaling can be directly paired with non-linear data transformations that alter the shape of variable distributions, replacing skewness with more Gaussian-like symmetry. One such transformation is the Yeo-Johnson method, which can handle both positive and negative data. The Yeo-Johnson power transformation actually represents a family of transformations, the choice of which depends on a single parameter, λ (Yeo & Johnson, 2000):

$$x_i^{(\lambda)} = \begin{cases} [(x_i + 1)^\lambda - 1]/\lambda & \text{if } \lambda \neq 0, x_i \geq 0, \\ \ln(x_i + 1) & \text{if } \lambda = 0, x_i \geq 0, \\ -[(-x_i + 1)^{2-\lambda} - 1]/(2 - \lambda) & \text{if } \lambda \neq 2, x_i < 0, \\ -\ln(-x_i + 1) & \text{if } \lambda = 2, x_i < 0. \end{cases} \quad (3.2)$$

Scikit-learn supports Yeo-Johnson through the *PowerTransformer* preprocessing tool that automatically estimates the λ parameter using maximum likelihood (scikit-learn, 2021f). Figure 3-6 shows the same histograms after applying both standard scaling and Yeo-Johnson transformation to the predictor variables. Many of the distributions appear much less skewed, and all have zero-mean and unit variance. The feature-specific values for λ derived from the FDS and used for the transformation are listed in Table 3.2.

Feature	λ
Average Air Temperature	1.19
Average Precipitation	0.05
Basement Depth	0.50
Boron Concentration	-1.44
Crustal Thickness	0.42
Depth to Water Table	0.44
Drainage Density	1.04
Earthquake Density	1.25
Gamma-Ray Dose Rate	0.36
Geodetic Strain Rate	0.04
Gravity Anomaly	0.92
Gravity-Anomaly Gradient	-0.03
Heat Flow	0.57
Lithium Concentration	-0.84
Magnetic Anomaly	0.86
Magnetic-Anomaly Gradient	-0.53
Quaternary Fault Density	-0.83
Si Geothermometer Temperature	1.29
Spring Density	2.03
State Map Fault Density	0.38
Surface Topography (DEM)	0.70
Topographic Gradient	-0.63
Volcanic-Dike Density	-0.42
Volcanic-Vent Density	1.86
Water-Table Gradient	-0.30

Table 3.2: Yeo-Johnson transformation λ parameter values determined using the scikit-learn *PowerTransformer* method ([scikit-learn, 2021f](#)).

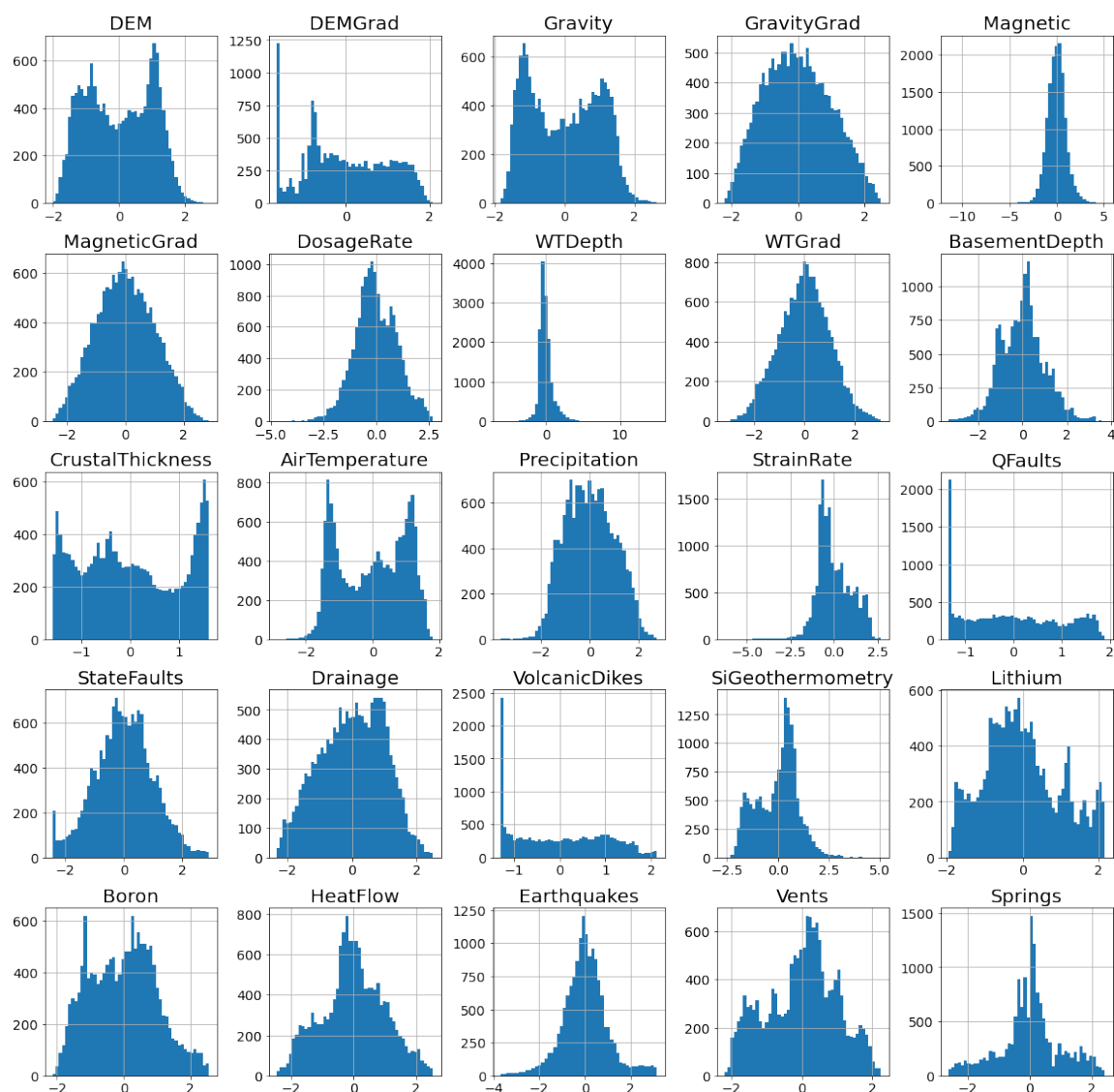


Figure 3-6: Histograms of the 25 features after standard-scaling and Yeo-Johnson transformation of FDS. Plots use 50 bins.

Scaling and transformation also impact the variable pairwise scatter plots. Figure 3-7 illustrates the improvement in scatter plot appearances as a result of the feature conditioning. With greater spread in the variable distributions, relationships are more readily apparent.

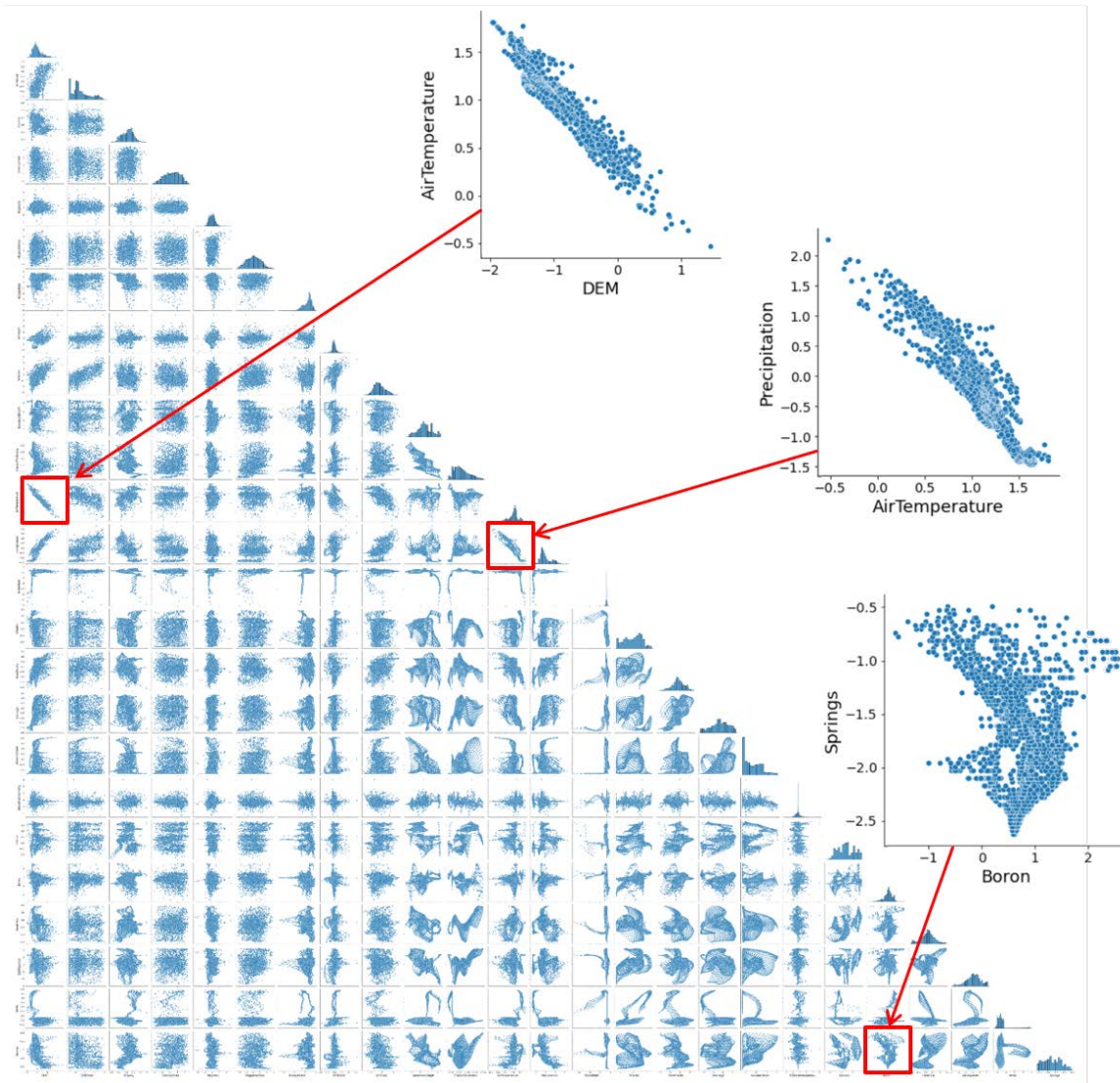


Figure 3-7: Scatter plots between all possible pairs of the 25 features after standard scaling and Yeo-Johnson transformation for FDS. The upper two plot call-outs illustrate collinear relationships. The lowermost highlighted plot shows the impact of Yeo-Johnson transformation on revealing variable relationships that were hidden by skewed distributions. All plots show the first 2,000 points of the 15,007-point FDS.

Feature Correlation

Another way to evaluate collinearity between two features is to calculate their correlation coefficient. The standard Pearson correlation coefficient (r) is strictly defined as the covariance between two variables, scaled by the product of their standard deviations. On a per-sample basis, this becomes (James et al., 2013, p. 70):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3.3)$$

where $\bar{x} = \sum_{i=1}^n x_i/n$ and $\bar{y} = \sum_{i=1}^n y_i/n$. When r is close to zero, no significant covariance takes place between the two variables. Values close to 1 or -1 suggest the two variables are linearly related, where the sign indicates direction of the relationship. A lower triangular matrix of pairwise correlation coefficients was calculated using the scaled, transformed version of FDS (Figure 3-8). Average Air Temperature stands out as highly collinear with multiple variables: DEM (-0.97), Gravity Anomaly (0.89), and Crustal Thickness (-0.89). Crustal Thickness also shows some collinearity with Gravity Anomaly (-0.88) and DEM (0.80). The same is true for Gravity and DEM (-0.83).

Focusing on the related Earth systems, the logic behind these relationships makes sense. High surface elevations recorded in the DEM layer will have correspondingly lower average air temperatures, hence snow caps appearing on mountains. In addition, if the crust is assumed to be in isostatic equilibrium with the mantle — like an iceberg floating in the ocean — topographic highs will be supported by thick roots and DEM will directly covary with crustal thickness. And since crustal densities are less than those of the underlying mantle, displacement of upper mantle by crustal roots (high crustal thickness) results in a lower average density and negative gravity anomaly values. The reverse is true as well; thinner crust will correspondingly tie to positive gravity anomaly values. Of these variables, only Air Temperature and DEM have a correlation value over 0.9. In fact, a 0.97 r -value suggests Air Temperature and DEM are nearly interchangeable in the value of information they provide to a predictive model. Since air temperature does not directly relate to subsurface characteristics

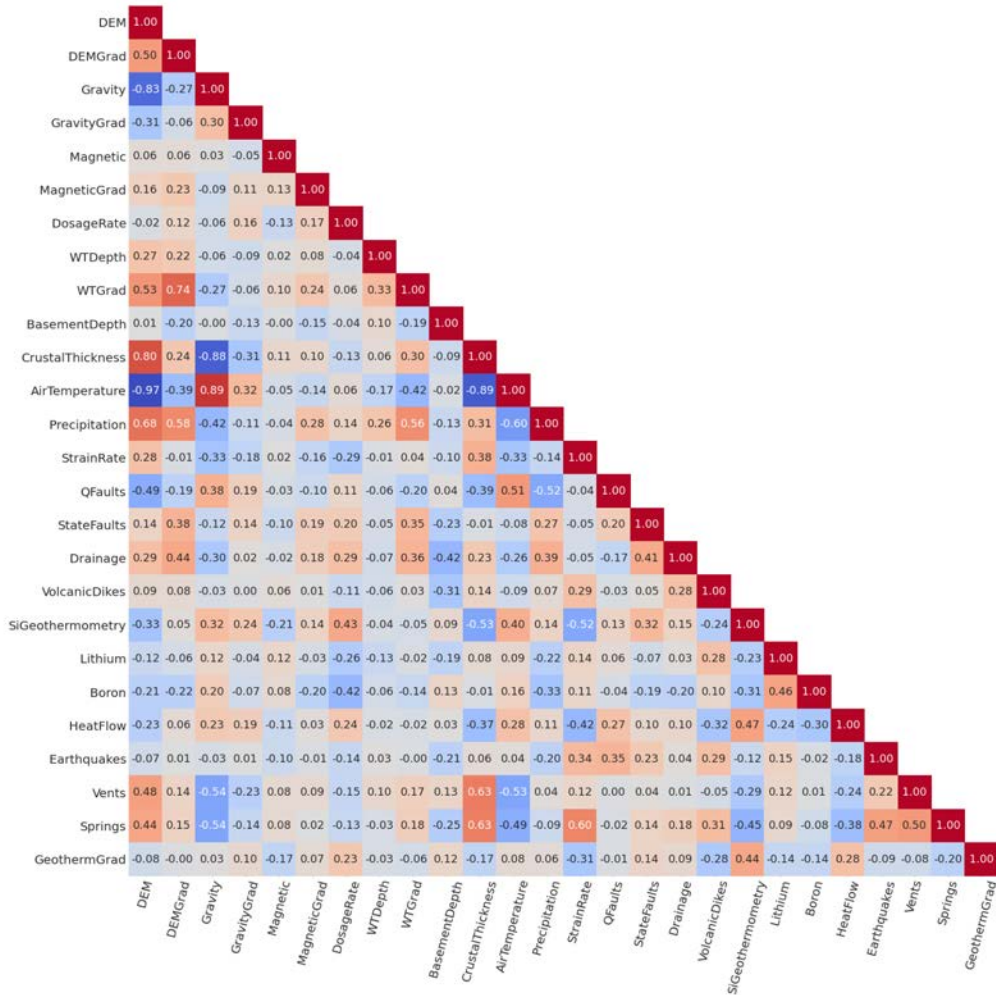


Figure 3-8: Correlation matrix with Pearson correlation scores for each feature pair based on the scaled and transformed version of the FDS.

except for thermal conditions at zero-depth, Average Air Temperature can be removed from the main set of predictors to reduce overall collinearity in the data set. Crustal Thickness and Gravity Anomaly both similarly show non-ideal r -values, but feature ranking while modeling will provide another opportunity to consider which of them might need to be removed (e.g., Section 3.3.5).

Classification Framework

Geothermal gradient can be treated as a continuous variable and predicted directly using regression methods. Alternatively, binning the geothermal gradient into discrete ranges changes the approach into a classification problem. In the context of exploration, geospatial classifications have a direct corollary in the typical traffic-light coloration of PFA favorability maps and other simplified displays of complex risk. Furthermore, regression model results provide exact geothermal-gradient estimates, which could easily be mistaken for certainty in a largely under-constrained problem. The binning approach is adopted in this thesis, largely based on previously published work.

The *Geothermal Gradient Map of the Conterminous United States*, published in 1991, separates geothermal gradient into five 15 K/km bins that cap out with 60–75 K/km (LANL et al., 1991). Armstead & Tester (1987) instead defined non-thermal gradients as 20–25 K/km, thermal gradients as ≥ 38 K/km, and hydrothermal gradients as 60–80 K/km on average. Tester & Herzog (1990) reframed this model with discrete representative values for three EGS grades: high = 80

Gradient Range	Class
[0 K/km, 30 K/km)	0
[30 K/km, 40 K/km)	1
[40 K/km, 60 K/km)	2
[60 K/km, 999 K/km)	3

Table 3.3: Geothermal gradient ranges and assigned class values using set notation. Ranges are left-inclusive.

K/km, mid = 50 K/km, and low = 30 K/km. In a later iteration, Herzog et al. (1997) confirmed ranges for EGS resources: high-grade for > 60 K/km, mid-grade for 40–60 K/km, and low-grade for < 40 K/km. This thesis extends the Herzog model by adding a non-thermal range as < 30 K/km, recognizing

the average geothermal gradient ranges from 25–30 K/km and anything below that would be ill-suited for geothermal exploration and development (see Table 3.3).

Preparation of well data sets WDS, WDS4, and WDS8 first involved removing records with NaN values or negative geothermal gradients. The remaining records were split into the gradient categories defined in Table 3.3. Distributions of class values for all data sets are shown in Table 3.4. For FDS, the class break-down is based on the extrapolated Bielić et al. (2015) geothermal-gradient layer (Figure A-30) sampled using the AOI mesh grid (Figure 3-1). Note that class imbalance exists in all data sets; mid-grade values dominate in the FDS, but the well data sets show a bias toward high-grade examples with few non-thermal examples. This is a common conundrum when using well data for characterization and analysis: drilling campaigns tend to target areas to drill based on chance of success, rather than drilling at random, so low-side under-representation tends to be ubiquitous.

	FDS	WDS	WDS4	WDS8
Class 0	2,394	20	101	184
Class 1	4,432	99	499	905
Class 2	6,243	232	1,144	2,029
Class 3	1,938	245	1,229	2,226
Total	15,007	596	2,973	5,344

Table 3.4: Distribution of geothermal gradient classes for each data set.

Stratified Sampling

Supervised machine-learning methods tune model predictions based on the input data provided during training. In approximating the target function, i.e., the response variable, these models seek to minimize an objective function as much as possible. The pitfall here lies in the representativeness of the training data set; if a model exactly matches the input data, it may not generalize well to other unseen observations. This would be a high-variance model, meaning the model predictions are strongly coupled to the characteristics of the training data. Variance trades off with model bias, which includes simplifying assumptions on the form of the target function. The bias-variance trade-off is a core concept in all statistical modeling, including machine

learning ([James et al., 2013](#), p. 33-36) and drives many of the model decisions made within this thesis.

One method of reining in the high-variance overfitting problem involves splitting the input data set into training and testing subsets. Model fitting uses the training set, while model evaluation relies on the test set. An additional split of the testing subset into separate validation and testing sets allows a modeler to validate model parameter choices with validation data not seen during model training, then conduct a final model evaluation using completely unseen test data ([Hastie et al., 2009](#), p. 222). Creating such firm boundaries between seen and unseen data helps prevent the issue of data leakage. Fundamentally, if a model trains or is tuned on data used to evaluate its performance, that evaluation no longer reflects real world predictive ability because information from the evaluation data has already “leaked” into the model ([Kaufman et al., 2012](#)). When data leakage occurs, model performance after deployment might not match that seen during testing. In other words, it may not be as good a model as the modeler thinks it is.

For classification problems, randomly splitting input data set into 3 subsets will violate the balance between class proportions in the original input data. Fortunately, sampling techniques exist that preserve the relative fractions of each class in the derivative subsets. Scikit-learn provides the *StratifiedShuffleSplit* method for randomly sampling from each class subgroup to generate training, validation, and test sets that look like one another and the original complete data set ([scikit-learn, 2021e](#)). The FDS, WDS, WDS4, and WDS8 data sets were each partitioned into 70% training, 15% validation, 15% testing using this method. Table [3.5](#) lists raw counts of observations associated with each class bucket and illustrates the consistency in class proportions across the different data sets.

Data leakage can be a concern when scaling or transforming split data sets. For example, the mean and standard deviation used for Z -score normalization should be derived from the training subset before being applied to the validation and testing subsets. This maintains separability between seen and unseen data. For this reason, the train-validate-test splits of the four different data sets listed in Table [3.5](#) were

	%	WDS train	WDS validate	WDS test	WDS4 train	WDS4 validate	WDS4 test	WDS8 train	WDS8 validate	WDS8 test
Class 0	3	14	3	3	71	15	15	129	27	28
Class 1	17	69	15	15	349	75	75	633	136	136
Class 2	38	162	35	35	801	171	172	1,420	305	304
Class 3	42	172	36	37	860	185	184	1,558	334	334
Total	100	417	89	90	2,081	446	446	3,740	802	802

Table 3.5: Raw observation counts for each geothermal gradient class across the different data sets after splitting each into training, validation, and testing subsets.

performed using the unscaled, untransformed versions of those data sets. Feature scaling and the Yeo-Johnson transformation discussed in Section 3.2.2 take place immediately before predictive modeling in a multi-step pipeline approach supported by scikit-learn (see [scikit-learn](#), 2021c).

3.3 Data Modeling

Supervised learning methods for classification come in a variety of shapes and sizes. Rather than settle on one for predicting geothermal gradient, four different methods are applied to the Southwestern NM data set. Figure 3-9 illustrates the high-level modeling flow, where model complexity increases with successive steps. The method descriptions below only briefly delve into important model mechanics and key *hyperparameters* (i.e., parameters not learned from data) that impact model performance. Other sources can provide a deeper review of machine-learning algorithms and their mathematical underpinnings. This investigation should instead be considered an applied case study that uses these algorithms as tools for generating insights on geothermal potential.

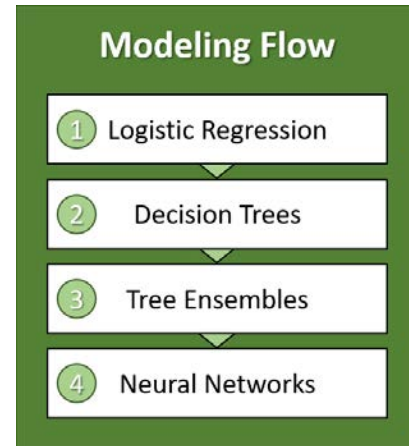


Figure 3-9: Workflow for predicting the class of geothermal gradient in the Southwestern NM study area using a variety of common machine-learning methods.

3.3.1 Assessing Performance

Building an intuition for the differences in predictive ability of various models first requires a clear definition of the scoring metric(s) used to compare those models. The characterization of classifier performance typically begins with a confusion matrix. In its simplest 2×2 form, the confusion matrix evaluates binary class predictions as True Positive (TP; predicted 1, actually 1), True Negative (TN; predicted 0, actually 0), False Positive (FP; predicted 1, actually 0), and False Negative (FN; predicted 0, actually 1). For the multi-class problem, the confusion matrix expands to include all correct classification and misclassification options. Figure 3-10 illustrates the elements of a 4×4 four-class matrix.

		Predicted			
Actual	A	0	1	2	3
	0	True 0,0	False 0,1	False 0,2	False 0,3
	1	False 1,0	True 1,1	False 1,2	False 1,3
	2	False 2,0	False 2,1	True 2,2	False 2,3
	3	False 3,0	False 3,1	False 3,2	True 3,3

		Predicted			
Actual	B	0	1	2	3
	0	True Negatives		False Positives	TN
	1				
	2	False Negatives		TP	FN
	3	TN		FP	TN

Figure 3-10: Confusion matrix diagram for a 4-class scenario. A. Each cell represents a pairing between an actual class label (rows) and the predicted label (columns). True positives for each class are down the diagonal. B. Example of matrix interpretation using class 2 as a point of reference. Elements associated with TP, FP, TN, and FN values are labeled.

Several statistical measures can be defined using combinations of elements in the confusion matrix. Of significance to the present study are Accuracy, True Positive Rate (TPR) and False Positive Rate (FPR) (Tharwat, 2020):

- **Accuracy:** the fraction of predictions that were correct:

$$(TP + TN) / (TP + TN + FP + FN)$$

- **True Positive Rate:** the count of correctly-predicted positives scaled by the actual positives: $TP/(TP + FN)$
- **False Positive Rate:** the count of incorrectly-predicted positives scaled by the actual negatives: $FP/(FP + TN)$

Classification relies on a probability threshold for assigning a class label. As the threshold lowers, the chance that the classifier believes it has a label match increases. By varying this threshold, it becomes possible to map out the discriminating ability of a classifier by plotting a curve in TPR vs. FPR space (Figure 3-11). This is commonly referred to as the Receiver Operating Characteristic (ROC) curve (Fawcett, 2006). A classifier that cannot discriminate between classes performs no better than random guessing, with a curve that plots along the diagonal from the origin to the upper right of the plot. On the other hand, a perfect classifier has a TPR of 1.0 for all thresholds, so it plots up along $FPR = 0.0$ then horizontally along $TPR = 1.0$. Typical ROC curves appear in the super-diagonal space between these two extremes.

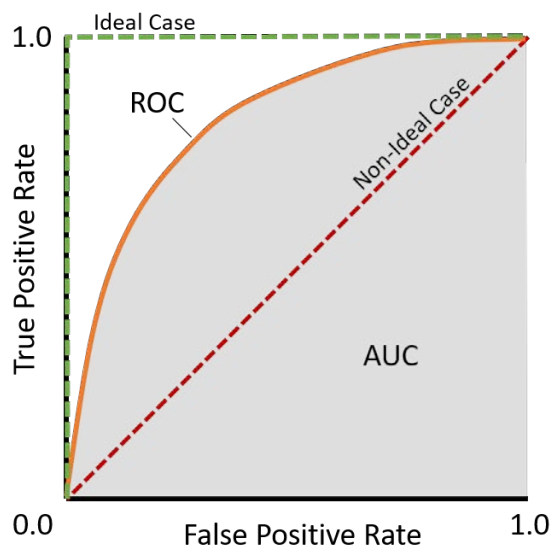


Figure 3-11: ROC curve diagram in TPR vs. FPR space. Perfect classifiers will plot along the ideal case line (green), poor classifiers plot along the diagonal (red). AUC (gray) characterizes the quality of the classifier, usually with values between 0.5 and 1.0.

Area Under the ROC Curve (AUC) defines a summary statistic for the ROC

function (see Figure 3-11) (Fawcett, 2006). Since the ideal classifier has a TPR of 1.0 at all times, the ideal AUC also equals 1.0. In the non-ideal case where a classifier performs no better than a random guess, the AUC drops to 0.5. If a classifier routinely mis-predicts and $\text{TPR} < \text{FPR}$, AUC drops below 0.5. AUC values and ROC curves provide a standardized means of comparing classifiers and are the primary performance measures used in this thesis.

With multi-class classification, defining single-class performance using the definitions of TP, FP, TN, and FN as shown in Figure 3-10B is relatively straightforward. Overall classifier performance across all classes can also be characterized with a single ROC curve using macro, weighted, or micro averaging (scikit-learn, 2021d).

- **Macro averaging:** matches the unweighted arithmetic mean of metric values.
- **Weighted averaging:** follows the procedure of macro averaging but adds a weight for each class contribution based on the fraction of total observations that fall within that class.
- **Micro averaging:** considers class results in aggregate, so statistics are calculated across the entire confusion matrix. TPR becomes the accuracy and FPR becomes the error rate.

For imbalanced data sets, a micro-average ROC curve will indicate better performance than the macro-average ROC curve due to the impact of the dominant class. Both micro- and macro-average ROC curves are included in the classification analysis for each machine-learning model in this thesis.

3.3.2 Hyperparameter Tuning

For a machine-learning model to perform at its best, the hyperparameters controlling model behavior must first be optimized or “tuned.” Assuming a large enough set of data is available, tuning simply involves training a series of classifiers with different values for a hyperparameter, assessing their performance against the validation data

subset, and choosing the value with the best results. One statistically stable alternative approach useful for sparser data sets involves k -Fold Cross Validation (CV). By this method, the input data is split into k subsets, or folds. The model is repeatedly trained on the aggregate of all but one fold, then assessed against the remaining fold (James et al., 2013, p. 181). This leave-one-out strategy cycles through all k permutations of splitting the data, and the scores are averaged to produce a summary statistic (e.g., AUC). For imbalanced class data, folds can be stratified-sampled such that class proportions of the unpartitioned data are preserved within each fold (Brownlee, 2020a).

When tuning a specific hyperparameter, the k -Fold CV process defines a set of average scores for the range of hyperparameter values under consideration, and the optimal parameter value can be determined from a plot of those scores. In some circumstances, a clear maximum in cross-validation results indicates the best value to use for modeling. In others, the CV curve levels off to form a corner or “elbow.” Choosing a hyperparameter value near this corner position balances the trade-off between overfitting and underfitting the training data.

3.3.3 Logistic Regression

Algorithm Details

The classic Logistic Regression (LR) model is a binary classifier that predicts one of two labels based on the input data. Linear regression treats the problem as a linear combination of the input observations (Bertsimas et al., 2016, p. 369):

$$g = \boldsymbol{\theta}^T \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n, \quad (3.4)$$

where x_i are the n feature observations, θ_i are $n + 1$ coefficients or weights for those features, and g is the log-odds of $\mathbf{x}$. Logistic regression adjusts the problem such that predictions define the probability of belonging to class 1. This is done by using a

non-linear logistic response function:

$$P(y = 1) = h_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{1 + e^{-g}}. \quad (3.5)$$

Equation 3.5, also known as the *sigmoid* function, converts the weighted sum from equation 3.4 to values between 0 ($g \rightarrow -\infty$) and 1 ($g \rightarrow \infty$) (Bertsimas et al., 2016, p. 369). Solving for the weights (θ_i) in this equation requires an iterative optimization procedure like gradient descent. This procedure minimizes an objective function ($J(\boldsymbol{\theta})$) based on the negative log likelihood (Ng, 2011a):

$$\begin{aligned} J(\boldsymbol{\theta}) &= -\frac{1}{n} \sum_{i=1}^n \text{Cost}(h_{\boldsymbol{\theta}}(\mathbf{x}_i), y_i) \\ &= -\frac{1}{n} \sum_{i=1}^n (y_i \log h_{\boldsymbol{\theta}}(\mathbf{x}_i) + (1 - y_i) \log(1 - h_{\boldsymbol{\theta}}(\mathbf{x}_i))) \end{aligned} \quad (3.6)$$

Regularization is added to logistic regression to avoid overfitting, specifically by penalizing the sum of the squared weights (L^2 -regularization). A constant (λ) determines the trade-off of influence between the magnitude of the weights and negative log likelihood in the minimization (Ng, 2011c):

$$\text{regularized } J(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^n \text{Cost}(h_{\boldsymbol{\theta}}(\mathbf{x}_i), y_i) + \frac{\lambda}{2m} \sum_{j=0}^m \theta_j^2, \quad (3.7)$$

where m is the number of features. The scikit-learn *LogisticRegression* function used in this thesis applies a hyperparameter $\mathcal{C}$ to the negative log-likelihood term, which acts like the inverse of λ . Larger values of $\mathcal{C}$ result in less regularization (scikit-learn, 2021a).

Multi-Class Heuristics

The formulation of logistic regression defines a strictly binary classification problem without multi-class support. Two heuristic methods allow LR to extend to multi-class classification: One-versus-One (OvO) and One-versus-Rest (OvR) (Brownlee, 2020b; scikit-learn, 2021b). Both split the problem into multiple binary classifications. OvO

considers every class versus every other class. In the 4-class geothermal gradient problem, this amounts to six classifications: (0 vs. 1), (0 vs. 2), (0 vs. 3), (1 vs. 2), (1 vs. 3), (2 vs. 3). OvR simplifies the problem by combining class alternatives so the number of classifiers matches the number of classes: (0 vs. [1, 2, or 3]), (1 vs. [0, 2, or 3]), (2 vs. [0, 1, or 3]), (3 vs. [0, 1, or 2]). For both methods, the class with the greatest score or sum of scores wins, where the score is akin to the probability of class membership. This thesis uses the OvR strategy.

Recursive Feature Selection

By consequence of Equation 3.5, logistic regression assumes a linear relationship between the weighted sum of independent predictors and the log-odds, i.e., $g = \log\left(\frac{P(y=1)}{1-P(y=1)}\right)$. However, including all possible predictors x_i will not necessarily improve the model. Feature selection can lead to simpler models with the same predictive power but reduced risk of collinearity, which is important when managing data from naturally-integrated earth systems.

One method for selecting the features to keep involves an iterative process called Recursive Feature Elimination (RFE) (Brownlee, 2020c; scikit-learn, 2021d). The concept is relatively simple: RFE recursively selects and removes the feature with the smallest-magnitude coefficient θ_i in the logistic-regression model, then refits the model to the data and repeats until a user-defined number of features is reached. A plot of AUC vs. number n of features can be constructed by looping over different feature limits, where the logistic-regression model is fit on the training subset and evaluated on the validation subset. Much like in the CV process, the maximum will define the optimal number of features, and the elbow provides guidance when results show no clear maximum.

3.3.4 Decision Trees

A decision tree, sometimes known as a Classification and Regression Tree (CART), classifies observations using a cascading set of evaluations, each on individual pre-

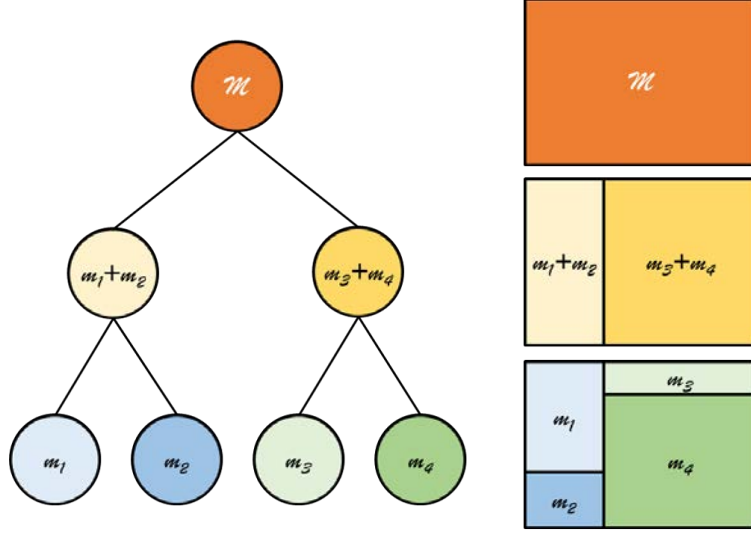


Figure 3-12: Decision-tree diagram illustrating the binary split process partitioning an initial data set M into data subsets $m \in \{m_1, \dots, m_4\}$. Each node represents a fraction of the data. The leaf nodes define the most granular subsets.

dictor variables. There is no assumption of a linear response in the system. In fact, decision trees are uniquely suited to representing non-linear behavior in a highly-explainable way; once constructed, the decision tree describes a flowchart-like roadmap for each label assignment (Bertsimas et al., 2016, p. 373–375). Not every predictor needs to appear in the decision tree, just those found to be significant during tree construction. And the most significant variables tend to be associated with early decision splits, placing them higher in the tree (Bertsimas et al., 2016, p. 376). Because of this, decision trees naturally provide insights into feature importance.

Algorithm Details

Decision trees are constructed by recursively performing binary splits on the training data set (see Figure 3-12). Each split defines two new nodes in the tree, which correspondingly partitions a group within the training data into two subgroups. These subgroups represent new terminal leaf nodes on the decision tree. The classification decision for each leaf will be the most commonly occurring class from the training data observations that are assigned to that leaf (James et al., 2013, p. 311).

There are three metrics that can play a role in evaluating the quality of a node in

the decision tree:

- **Classification error rate:** the proportion of training samples that don't match the dominant class in a leaf node (James et al., 2013, p. 312):

$$E = 1 - \max_{k=1}^K (\hat{p}_{mk}), \quad (3.8)$$

where m is the subset of the training data associated with a tree node, k is a class among K possible classes, and $\hat{p}_{mk}$ is the fraction of all training observations in m that are of class k .

- **Gini index:** measures variance across all K classes. Gini is sometimes known as a purity measurement because low values correspond with a strongly dominant class (James et al., 2013, p. 312):

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk}). \quad (3.9)$$

- **Entropy:** an alternative form of purity measure, which also shows low values when the proportion of one class dominates the others within a node (James et al., 2013, p. 312):

$$D = - \sum_{k=1}^K \hat{p}_{mk} \log \hat{p}_{mk}. \quad (3.10)$$

Tree construction takes place in two passes. In the forward pass, the tree will iteratively grow by selecting nodes in the tree, a predictor to split on, and a threshold value defining the split. These choices are made to maximize the purity of the child nodes, typically by using Gini index or entropy (James et al., 2013, p. 307). The tree will grow until a stopping condition is met, like reaching a maximum tree depth or minimum number of observations allowed per node. Tree clean-up or “pruning” then takes place in a backward pass, where the following decision tree objective governs whether a tree branch is kept or removed (James et al., 2013, p. 309):

$$J(\boldsymbol{\theta}) = E + \alpha_T |T|, \quad (3.11)$$

where $|T|$ refers to the number of terminal nodes in the decision tree and the classification error rate (E) is used for measuring quality. The objective could rely on Gini index or entropy as well, but classification error rate will maximize prediction accuracy (James et al., 2013, p. 312). If E increases by less than α_T when a branch is removed, that branch will remain removed from the decision tree. α_T acts as a regularization parameter, balancing prediction accuracy with model complexity; greater values of α_T result in simpler decision trees.

3.3.5 Tree Ensembles (XGBoost)

As simple and effective as decision tree classifiers may be, they only demonstrate a single model solution. And since random selection can play a role in their construction (e.g., in the scikit-learn implementation), a different tree structure may arise if the tree is rebuilt on the same data set. The random forest algorithm embraces these random variations and generates a large number of decision trees. To increase randomness, only a subset of the predictors are used when building each tree, and trees are trained on a data subset selected at random with replacement from the full training set (Bertsimas et al., 2016, p. 376–377). Through aggregation, the collection of trees acts as a single, more performant ensemble model. However, greater potential accuracy in predictions trades off with less interpretability; as the forest grows larger, the number of constituent decision trees quickly exceeds the limits of human comprehension. As such, ensemble tree methods like random forests are considered black-box algorithms.

Algorithm Details

A variation on this ensemble method uses “gradient boosting,” where the trees are chained in succession and train on the residual error of the preceding tree. The trees are weak learners that underfit the data, giving them low variance, but high bias. Yet when they connect together through residual prediction, the final boosted model (Equation 3.12) can out-perform conventional random forests (James et al., 2013, p.

323). A gradient-boosting model takes the form of:

$$\hat{f}(\mathbf{x}) = \alpha_s \sum_{b=1}^B \hat{f}_b(\mathbf{x}), \quad (3.12)$$

where $\hat{f}(\mathbf{x})$ is the boosted model, $\hat{f}_b(\mathbf{x})$ are the individual trees in the chained ensemble totaling B in number, and α_s is the shrinkage parameter or learning rate.

XGBoost, a variant of gradient-boosting tree algorithms, has gained notoriety from a history of machine-learning competition wins (Chen, 2021). The objective function governing XGBoost model construction balances two influences (Chen & Guestrin, 2016):

$$\begin{aligned} J(\boldsymbol{\theta}) &= \mathcal{L} + \Omega \\ &= \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{b=1}^B \omega(\hat{f}_b) \\ &= \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{b=1}^B \left(\gamma |T|_b + \frac{1}{2} \lambda \sum_{t=1}^{|T|_b} \theta_{b,t}^2 \right), \end{aligned} \quad (3.13)$$

where the first part ($\mathcal{L}$) expresses how poorly the model fits the data, while the second term (Ω) describes the complexity of the model. $\mathcal{L}$ is the sum of individual loss calculations ($l(\hat{y}_i, y_i)$) on the n predicted and observed response variable values. Tree-specific complexity ($\omega(\hat{f}_b)$) calculations balance the number of leaves in a tree ($|T|_b$) with the L2 norm of leaf weights ($\theta_{b,j}$), which are unique to XGBoost decision trees. Both γ and λ serve as regularization factors.

Adding a new tree to the ensemble during training is an additive operation, optimized using a quality score of a specific tree structure based on Equation 3.13. Chen & Guestrin (2016, Equations 3–6) step through the derivation, which include a second-order approximation to simplify to the following relationships:

$$\begin{aligned} \theta_{b,t} &= -\frac{\sum_{i \in m_t} g_i}{\sum_{i \in m_t} h_i + \lambda}, \\ \tilde{J}(\boldsymbol{\theta})_b &= -\frac{1}{2} \sum_{t=1}^{|T|_b} \frac{\sum_{i \in m_t} g_i}{\sum_{i \in m_t} h_i + \lambda} + \gamma |T|_b, \end{aligned} \quad (3.14)$$

where $\theta_{b,t}$ represents the optimized weight of leaf t for tree b , and $\tilde{J}(\boldsymbol{\theta})_b$ scores the quality of the tree. Here, g defines the first-order gradient statistics for the training data subset (m_t) assigned to leaf t , and h represents the second-order gradient statistics (Chen & Guestrin, 2016). XGBoost is clearly a complex algorithm, but it comes with many optimizations that make it extremely efficient, scalable, and popular among machine-learning practitioners.

Shapley Analysis

Being a tree-based machine-learning method, XGBoost naturally provides feature importances as a product of model-fitting. In fact, the XGBoost package offers five kinds of global feature-importance calculations (xgboost developers, 2020), but each can give slightly different results in feature ordering or relative feature impact on model predictive behavior. At issue here are two concepts in importance definition of “feature attribution” methods: *consistency*, or the independence of a feature-importance value and the reliance of a model on that feature, and *local accuracy*, or the idea that the sum of the importances is equivalent to the original model output for a given input (Lundberg, 2020). A study of six different approaches to interpreting models through feature attribution showed that only one method meets both of these properties: Shapley Additive Explanation (SHAP) (Lundberg et al., 2019).

Shapley values derive from cooperative gain theory (Shapley, 1997), but have gained traction in the machine-learning community partly because they predict importances without assuming complete feature independence (Lundberg & Lee, 2017). In addition to managing collinearity, Shapley values have several desirable attributes. For example, if two features impact a model prediction equally, they will have equivalent Shapley values. And a Shapley value of zero means the feature has no predictive impact. The SHAP method produces values that have global significance for general feature importance, but also local significance for an individual prediction; the sum of SHAP values is equivalent to the deviation of the model prediction from the average value (baseline), meaning SHAP values describe the individual feature contributions to a prediction value (Lundberg & Lee, 2017). This thesis uses SHAP values di-

rectly provided by the XGBoost package (xgboost developers, 2020) for evaluating importances and for feature selection.

3.3.6 Neural Networks

The original concepts behind Artificial Neural Networks (ANNs or just NNs) come from simplified models of neuron activity in the brain (Hastie et al., 2009, p. 394). At a high level, multiple inputs are fed into neuron cells from many branches on one end, and given the right combination of those inputs, these cells will fire and propagate a signal to the next group of connected neurons.

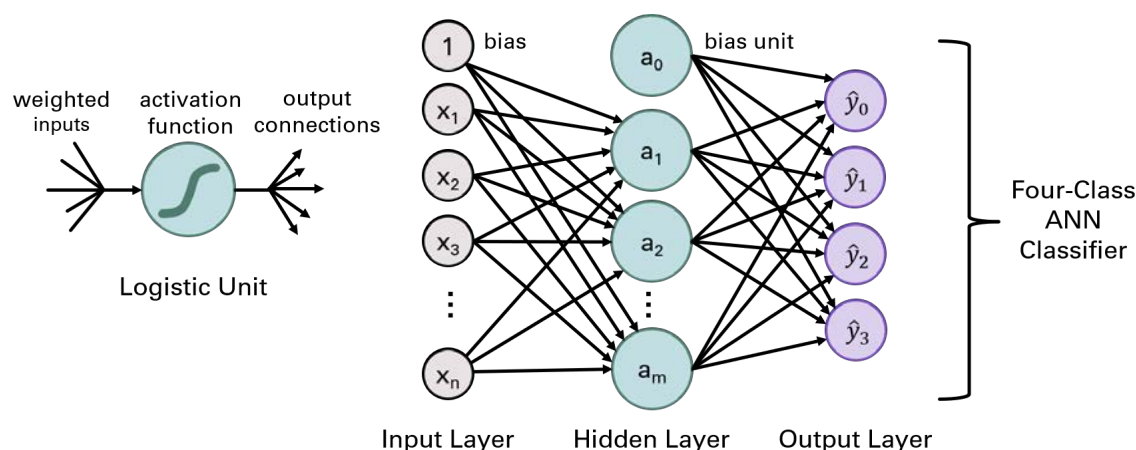


Figure 3-13: Schematic of a logistic unit (Left) and a fully-connected neural network (Right) with a single hidden layer made up of logistic units a_1 through a_m . Bias units a_0 traditionally have a value of 1, but are weighted in linear combinations like other units. Four units in the output layer make this a four-class classifier.

In neural networks, cells are represented by logistic units (Figure 3-13). Each unit acts like a logistic regression operation, where inputs are scaled by weights, and the linear sum of the weighted inputs are passed through an *activation function* to determine if the output is a 0 or a 1. As in logistic regression, the sigmoid function (Equation 3.5) appears in many network architectures. However, several limitations of sigmoid functions in the context of neural networks, e.g., limited sensitivity and vanishing gradients, have driven machine learning practitioners towards alternatives like the Rectified Linear Unit (ReLU) (Brownlee, 2019a; Nair & Hinton, 2010). The

present study uses ReLu activation functions for hidden layers in the network.

Algorithm Details

Each logistic unit performs the following operation:

$$\mathbf{a} = h(\mathbf{z}) = h(\mathbf{\Theta}^T \mathbf{x}), \quad (3.15)$$

where h is the activation function and $\mathbf{\Theta}$ is a matrix of weights. For the first hidden layer, $\mathbf{x}$ will be the predictor values from the input layer as shown in Figure 3-13. For any additional hidden layers, $\mathbf{x}$ consists of the outputs of units from the preceding layer. $\mathbf{z}$ is analogous to g in Equation 3.4.

The cost function for a neural network takes on the form of a loss term and a complexity term (Ng, 2011b):

$$J(\mathbf{\Theta}) = -\frac{1}{n} \sum_{i=1}^n l(\hat{\mathbf{y}}_i, \mathbf{y}_i) + \frac{\lambda}{2n} \sum_{\zeta=1}^{L-1} \sum_{i=1}^{s_{\zeta}} \sum_{j=1}^{s_{\zeta+1}} \left(\Theta_{ji}^{(\zeta)} \right)^2, \quad (3.16)$$

$$l(\hat{\mathbf{y}}_i, \mathbf{y}_i) = \sum_{k=1}^K (y_{i,k} \log (h_{\mathbf{\Theta}}(\mathbf{x}_i))_k + (1 - y_{i,k}) \log (1 - (h_{\mathbf{\Theta}}(\mathbf{x}_i))_k)),$$

Note the similarity in appearance to the regularized negative log likelihood cost function for logistic regression (Equation 3.7). Here, L is the number of layers in the network, s_{ζ} defines the number of units in layer ζ , and $(h_{\mathbf{\Theta}}(\mathbf{x}_i))_k$ is the k th output of the network. $\Theta_{ji}^{(\zeta)}$ represents the weight assigned to the connection between unit i in layer ζ and unit j in layer $\zeta + 1$. Both i and j start at 1 because the bias terms (index 0) are unregularized. λ controls the balance between the loss and complexity terms.

Minimizing this objective function is generally accomplished by applying the back-propagation algorithm to compute gradients needed for a gradient descent optimization (Hastie et al., 2009, p. 396). The Adam (adaptive moment) optimization method is used in the present study due to its computational efficiency and current status as the recommended general-purpose optimizer (Brownlee, 2017). A description of Adam is beyond the scope of this thesis, but more information is widely available

elsewhere (e.g., [Kingma & Ba, 2017](#)).

The training process updates network weights using gradients calculated from the entire training data in each training epoch ([Hastie et al., 2009](#), p. 397). Successive epochs incrementally adjust the ANN to match the training data, so evaluating loss, AUC, or other metrics on both the training and validation sets is important for identifying the best number of epochs to avoid overfitting. Training can also proceed with mini-batches, that is, small subsets of the training set, rather than the whole set at once. This method makes the training results noisier, but generally speeds up learning while adding a regularization effect to changes in the weights ([Brownlee, 2019b](#)).

3.4 Uncertainty Analysis

In the case study considered in this thesis, each of the four supervised machine-learning methods discussed so far will generate different predictions for geothermal gradient at each point across the Southwestern NM AOI. But presenting these results to an explorationist will invariably elicit two important questions: 1) which of these model results is the right one; and 2) what actions should I take based on these models.

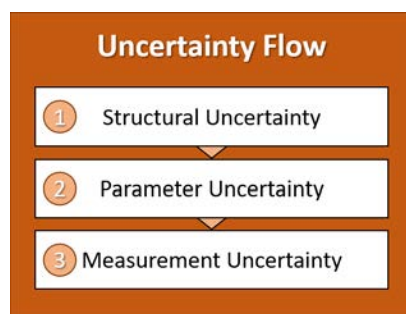


Figure 3-14: Workflow for analyzing uncertainties in model results for geothermal gradient prediction across the Southwestern NM study area.

George Box famously remarked, “Essentially, all models are wrong, but some are useful” ([Box & Draper, 1987](#)), which serves as an effective answer to the first

question. But knowing that none of the prediction maps is individually correct falls short of providing impactful and actionable information for geothermal-exploration decision makers. Value lies in describing the sources of uncertainty and where those uncertainties manifest in a spatial prediction of geothermal gradient or other modeled risk elements. To this end, the following analysis considers uncertainties from the three sources described in Section 2.2.5 and illustrated in Figure 3-14: model type (structural), model calibration (parameter), and model input (measurement).

3.4.1 Classification Uncertainty Measures

Unlike standard regression problems that result in continuous variable predictions, classifiers are inherently limited to selecting among K discrete class values. Variability in model results cannot be measured by standard deviation. For geospatial classifications, the simplest comparison between maps is visual; plotting two displays side-by-side can effectively communicate qualitative differences. Quantifying the uncertainty of categorical predictions based on multiple realizations defines a more difficult problem, with a variety of different statistical approaches to consider. Among these are performance measures based on the class assignments, including percent correctly classified (accuracy), confusion matrix, AUC, Kappa (Cohen, 1960), and weighted Tau (Ma & Redmond, 1995). Other methods consider the probabilities that precede the final classification choice, including Brier score (Brier, 1950) and Shannon entropy (Shannon, 1948).

Rather than perform an exhaustive search through different metrics, this thesis follows the recommendation of Beaudette (2020) in characterizing classification uncertainty using entropy. Although the standard entropy calculation (Equation 3.10) does not account for class similarity, it does show good discrimination ability for model results with low, medium, and high stand-out probability for the majority class (Beaudette, 2020). A normalized version of the entropy calculation is as follows:

$$H = -\frac{1}{\log_2 n} \sum_{i=1}^n p_i \log_2 p_i, \quad (3.17)$$

where p_i represents the probability assigned to class label i . Entropy maps can be constructed with the output probabilities of a single classification model or from the results of an ensemble of model predictions by first averaging across class probabilities for all models at each map location.

3.4.2 Structural Uncertainty

The specific choice of classification algorithm implicitly applies constraints on the solution space for the problem. For example, logistic-regression solutions assume there is a linear relationship between predictor variables and the log-odds of the response variable. And decision trees use hard thresholds that can create step-like model behavior. Model formulations thus influence the individual results and range of results that a model is capable of producing.

In order to examine the uncertainty introduced by structural aspects of the chosen machine-learning models, the predicted class probabilities for multiple models can be averaged together to create an ensemble prediction of geothermal gradient classes. Shannon entropy can then be calculated for the class probabilities of this average model and displayed as a map, with normalized values between 0 to 1. High entropy locations mark regions where there is no clear differentiation between most-likely gradient class labels. Low entropy locations indicate the opposite — multiple models agree on a single dominant class label for that area.

3.4.3 Parameter Uncertainty

The fitting procedure for machine-learning models generally involves minimizing a cost function as updates are iteratively made to model parameters, e.g., feature weights. However, the final trained models treat these parameters as deterministic values with no uncertainty. Using neural networks as an example, the complexity of the model architecture and the potential for different parameterizations as randomness in techniques like dropout and mini-batching do influence the training process, suggesting that a deterministic view insufficiently represents the model space. By

defining the range of uncertainty on model parameters, it is possible to generate many solutions from the same model — an ensemble of valid predictions. Uncertainty estimates derived from this ensemble can introduce a measure of confidence in a model’s output, allowing the model to reveal where predictions are well-constrained and, more importantly, where they are not.

The Google DeepMind team introduced a method called “Bayes by Backprop” in 2015 where single-value weights are replaced by probability distributions in a neural network architecture (Blundell et al., 2015). Making the simplifying assumption that the weight distributions are Gaussian, each weight parameter in a standard ANN layer is replaced by a mean and a standard deviation in a probabilistic layer. A fully-trained Bayesian Neural Network (BNN) samples from the weight distributions for just-in-time determination of weight values as data are fed through to produce a prediction. Therefore, each prediction of the BNN varies depending on the selected weights, and running the network many times on the same data set will produce a set of different results. An analysis of the variation in this solution ensemble can characterize how parameter uncertainty impacts BNN classification performance.

At a fundamental level, BNNs operate using Bayes theorem for training the network (Webster, 2021):

$$P(\mathbf{w}|D) = \frac{P(D|\mathbf{w})P(\mathbf{w})}{P(D)}, \quad (3.18)$$

where $\mathbf{w}$ are the model weights and D are the observed data. $P(\mathbf{w})$ define the *prior*, or the distribution assigned to the weights before obtaining D . $P(D|\mathbf{w})$ defines the likelihood of D given $\mathbf{w}$. $P(\mathbf{w}|D)$ is the *posterior* or the distribution after taking D into account.

In practice, BNNs use a Variational Bayes method, which approximates $P(\mathbf{w}|D)$ with a variational posterior, $q(\mathbf{w}|\theta)$. The desired difference between these two functions is zero (perfect approximation), so this defines the basis of a loss function defined

using the Kullback-Leibler divergence ([Webster, 2021](#)):

$$\begin{aligned} D_{KL}(q(\cdot|\theta) || P(\cdot|D)) &= \int q(\mathbf{w}|\theta) \log \frac{q(\mathbf{w}|\theta)}{P(\mathbf{w}|D)} d\mathbf{w} \\ &= \log P(D) + D_{KL}(q(\cdot|\theta) || P) - \mathbf{E}_{q(\cdot|\theta)} [\log P(D|\cdot)]. \end{aligned} \quad (3.19)$$

Dropping constant terms yields the objective to minimize ([Blundell et al., 2015](#)):

$$J(\theta) = D_{KL}(q(\cdot|\theta) || P) - \mathbf{E}_{q(\cdot|\theta)} [\log P(D|\cdot)]. \quad (3.20)$$

This two-term objective describes a trade-off between the complexity as controlled by deviation from the prior ($P(\mathbf{w})$) and the negative log-likelihood term measuring the fit to the data.

In this thesis, the TensorFlow Probability package is used to transform the ANN from Section 3.3.6 into a BNN for uncertainty estimation ([Dillon et al., 2017](#)). Running a trained BNN multiple times will produce a collection of different results due to the randomness in the model. Geothermal-gradient class probabilities from these realizations can be averaged by class for each point location, and entropy values calculated from the ensemble-averaged probabilities to examine parameter uncertainty.

3.4.4 Measurement Uncertainty

Data modeling and analytics generally begin with the collection, engineering, and conditioning of features of interest. Building the perfect dataframe for machine learning may be the first priority, but capturing standard error estimates for its constituent features is fundamental to understanding how much uncertainty they bring to the prediction problem.

With appropriate measures of standard error, the values for each feature in a data set can be perturbed to create a range of statistically-similar derivative data sets. Variability in the classifications from these data sets highlights model sensitivity to uncertainties in the feature measurements. As with the other uncertainties, the predicted class probabilities for model realizations using the different data sets can

be ensemble-averaged at each point within the AOI. And summary statistics like Shannon entropy calculated from the class probabilities indicate where the predictive power of the model is most sensitive to measurement uncertainty. This procedure can apply to a single feature, group of features, or all features at once to reveal how individual features or combinations of features impact model uncertainty.

3.5 Recap

This chapter discussed the data sources and conditioning steps followed to prepare a data set for exploration-scale predictive modeling of the Southwestern NM study area. Detailed information on each GIS data layer is presented in [Appendix A](#). The machine-learning models and uncertainties under consideration were also outlined. Key take-away messages from this chapter include the following:

1. Data sets gathered from geothermal archives, academic and agency repositories, and directly from the literature require extensive preparation to create consistent GIS layers as input features for modeling.
2. The predictor variables total 25 features that collectively measure aspects of the fluids, heat, and permeability among the geothermal risk elements. The response variable, geothermal gradient, represents a proxy for the heat risk element and is binned into 4 ranges representing non-thermal, low-grade, mid-grade, and high-grade gradient values for prediction.
3. Collinearity is an issue for several features. Average Air Temperature is very tightly coupled with Surface Elevation and was preemptively removed from the set of predictors.
4. The relatively sparse well data set (WDS) used to train the supervised machine-learning models is augmented through a data imputation strategy that creates pseudowells for each original well location. WDS4 adds pseudowells in the N, S, E, and W directions. WDS8 extends WDS4 with NE, SE, SW, and NW pseudowells.

5. Data sets are split into training, validation, and testing subsets using a stratified sampling technique, to preserve the distribution of geothermal gradient classes.
6. Machine-learning algorithms investigated in this thesis include logistic regression, decision trees, tree-based ensembles (XGBoost), and a neural network.
7. Assessments of classifier performance span multiple dimensions. Here, models are assessed and compared using confusion matrices, classifier accuracy, AUC, and ROC plots.
8. Feature importances describe the relative influence different features have on the prediction from a classifier. Some models provide this directly, others require routines like RFE to rank the features.
9. Structural uncertainty refers to classification uncertainty tied to model structure. This is evaluated both by visual results comparison and entropy calculations from an ensemble of model results.
10. Parameter uncertainty refers to classification uncertainty from the model parameterization. This is examined through entropy analysis of a results ensemble generated using a Bayesian Neural Network.
11. Measurement uncertainty refers to classification uncertainty related to the standard errors of input feature values. This is evaluated by generating statistically-similar variations of the input data and comparing model results through entropy analysis.

Results from applying these methods and models in the context of a risk-mitigation strategy for geothermal exploration are explored in Chapter 5.

Chapter 4

Cost Modeling for an EGS Power Plant Expansion

The second of the two research questions presented in Chapter 1 addressed the topic of risk in geothermal development and production. As discussed in Chapter 2, production planning generally includes an economic analysis of the subsurface conditions, development plan, and power-plant concept projected over the anticipated lifespan of a field. This chapter describes a methodology for modeling the value of an EGS power generation project applied to the Lightning Dock KGRA in Southwestern New Mexico. The method combines uncertainties and variable operational strategies to mitigate risk in geothermal production.

4.1 EGS Expansion Concept

4.1.1 Lightning Dock EGS

Lightning Dock is presently the only commercial power plant operating in the state of New Mexico (see Section 2.5.5). The net generating capacity after its first phase of development was 4 MW in 2013. An expected second-phase upgrade to 10 MW never came to fruition. Instead, the facility underwent a significant refit in 2018, resulting in a net capacity of 11.2 MW generated entirely from hydrothermal brine production

([Bonafin & Dickey, 2019](#)).

DOE-funded efforts to characterize the geothermal resources of the Animas Valley—where Lightning Dock is located—revealed two different thermal reservoirs: the hydrothermal resource targeted by Lightning Dock, where deep geothermal fluids ascend along the Animas Valley Fault complex to ≈ 365 –1000 m depth, and a secondary interval at ≈ 900 –1200 m depth that requires permeability enhancement for production ([Schochet & Cunniff, 2001](#)). The Horquilla limestone formation defines the second reservoir, estimated to span a minimum volume of 6 km³, based on conservative figures. By one proprietary study completed in 2001 for Ormat International, the Horquilla has a most-likely production potential of 9.3 MW and an 88% probability of exceeding 6 MW ([Schochet & Cunniff, 2001](#)).

[Schochet & Cunniff \(2001\)](#) proposed the construction of a 6 MW hybrid power plant combining hydrothermal and EGS-sourced power generation a decade before operations commenced at Lightning Dock. In their development plan, they noted several benefits of pursuing EGS in this location:

- Relatively shallow resource equates to lower drilling costs
- EGS water requirements attainable from paired hydrothermal operations
- Low/no assessed environmental impact from geothermal operations
- Direct access to in-place transmission lines
- Opportunity for direct electricity sales to local users
- Purchase agreements with regional utilities incentivized by NM legislation

As suggested by this list, conditions at Lightning Dock offer a nearly ideal test case for an EGS proof-of-concept on a manageable scale. In addition, land utilization in the area is historically agricultural with few residences, so the risk is low for adverse impact on an existing population. And the use of binary cycle generation as proposed by [Schochet & Cunniff \(2001\)](#) supports power production with zero greenhouse-gas emissions.

In this thesis, the [Schochet & Cunniff](#) concept is revisited with the existing geothermal production at Lightning Dock kept in mind; rather than building a new hybrid facility, the revised concept involves targeting the deeper reservoir as an NF-EGS development that ties back to the current Lightning Dock facility. Stepping out from the hydrothermal zone in proximity to the Animas Valley Fault complex, thermal conditions settle to a high background geothermal gradient between $\approx 80\text{--}120$ K/km, based on boreholes TG 56-14 and TG 12-7 ([Cunniff & Bowers, 2003](#)) —high enough to support geothermal capture. These conditions make for an interesting case study on risk-mitigation options for EGS production planning founded on an NF-EGS concept that was already proven at The Geysers ([Pan et al., 2019](#)).

Public records regarding power generation at Lightning Dock provide guidance on the appropriate size of such an EGS expansion. After its initial phase 1 development, the plant produced 4 MW. An additional 6 MW was slated for phase 2, but re-powering of the plant added over 7 MW to the capacity after several years of development stasis ([Think GeoEnergy, 2020](#)). [Schochet & Cunniff](#) originally proposed a 6 MW hybrid plant for the site, but they also noted 6 MW was likely understating the full reservoir potential of the Horquilla alone ([2001](#)). In consideration of the step-wise trajectory of plant improvements and the assessment of available thermal resources, the present case study targets 5 MW as an expansion goal.

4.1.2 New Mexico Electricity Demand

Pursuing the expansion of a power plant requires sufficient demand to ensure total revenue offsets project expenses. Fortunately, New Mexico regulations support further development of geothermal power production in the state. Specifically, the Energy Transition Act signed in 2019 updated the New Mexico Renewable Portfolio Standard (RPS) to go zero-carbon by 2050, with milestone targets along the way ([Lillian, 2019](#)). The RPS dates back to the Renewable Energy Act passed in 2004 and comes with several carve-outs, including a 30% requirement for wind energy, 20% for solar, and 5% for other renewables like geothermal ([DSIRE, 2021](#)). Public Service Company of New Mexico (PNM) is the state’s largest energy provider and services the Lordsburg

area where Lightning Dock is located. PNM and Cynq Energy currently share a 20-year Power Purchase Agreement (PPA) for electricity generated at Lightning Dock. The PPA has gone through amendments over time to update both the electrical power supplied to PNM and the pricing structure per MW·h (e.g., [PNM, 2014](#); [Stanfield, 2017](#)). This indicates a willingness to revisit a PPA if conditions change, which is an important aspect to consider when modeling project financials.

In addition to the RPS requirement for a diversified portfolio, coal power plants across the state face mandated shut-downs as a consequence of the Energy Transition Act. Coal currently supplies a large fraction ($\approx 37\%$) of in-state electricity generation ([EIA, 2021b](#)) and nearly 20% of consumed energy in New Mexico (Figure 4-1). The supply gap introduced as coal-based production drops to zero could more than compensate for a 5 MW addition of no-emissions energy to the New Mexico grid.

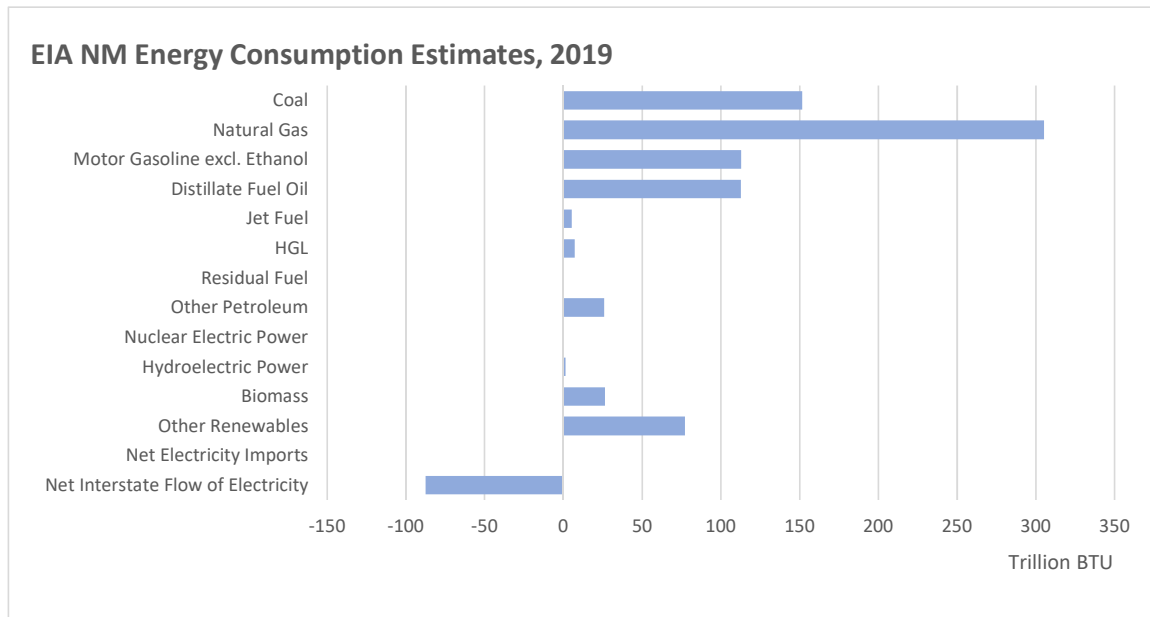


Figure 4-1: Energy consumption by source for New Mexico. Adapted from data and graphics reported by the EIA ([EIA, 2021b](#)).

4.1.3 Modular Geothermal

Limiting the expansion to a single 5 MW facility represents one design alternative, but others exist as well. One flexible option uses modular technology that recently cap-

tured the attention of high-stakes investors across the world (Shieber, 2019). Climeon has engineered a compact binary-cycle unit capable of 150 kW of generated electricity using inlet fluid temperatures rated up to 120 °C and flow rates of up to 35 kg/s (Climeon, 2021). These units can be combined into a larger deployable “Power Block” for 1050 kW of electric capacity (Winther, 2018) (Figure 4-2). Using this technology, power plants can now be treated like multi-unit assemblages, installed all at once or over an extended period based on operator needs (Climeon, 2018).

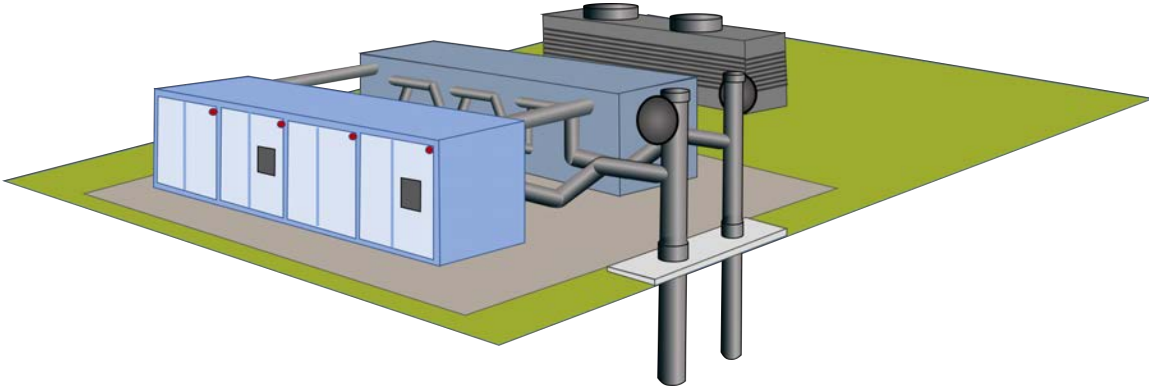


Figure 4-2: Modular binary cycle power plant concept, adapted from Climeon PowerBlock schematic diagram (Climeon, 2021). Each block consists of seven active units chained together to sum to ≈ 1 MW of generating capacity.

4.1.4 Flexible Cost Models

As discussed in Section 2.4, cost models can provide insights into the potential value gained or lost by a proposed facility before construction even begins. Well-established geothermal cost models like GETEM (Entingh et al., 2006) present a highly parameterized but deterministic view of cost and investment opportunity given a defined geothermal resource and development concept. Other models may apply different assumptions or mathematical treatments for various facets of the system; however, they uniformly offer a single-track aspect to how the project unfolds over its lifecycle. Users can test ideas, but the solution space remains under-explored due to implicit assumptions of variable trends or stases for what is actually a highly dynamic system.

In the cost model outlined below, the economic analysis accounts for uncertainty by replacing single value estimates with distributions for model variables. This enables

the model to produce a representative range of possible outcomes when simulated many times over. In addition, the model flexibly adapts by executing design options, where model updates triggered by changing conditions allow the system to realize upside potential or characterize the extent of downside risk. Designs need not be static, and flexibilities can greatly increase the expected value of a project by exploring execution strategies otherwise missed by more traditional modeling approaches (de Neufville & Scholtes, 2011, Chapter 6).

4.2 Static Cost Model

Geothermal cost models typically report Levelized Cost of Electricity (LCOE) for direct comparison with other renewable energy sources. However, LCOE summarizes the total lifetime costs of a power plant normalized by the total power generation from start-up to plant decommissioning. It is thus not well-suited for communicating projected net gains or losses under different plant designs or scenarios, which are the focus of the present analysis. Instead, the model described here relies on Net Present Value (NPV), a simple measure of project worth that accounts for the time value of money by applying a single interest rate, the discount rate, for both borrowing and deposits (de Neufville & Scholtes, 2011, p. 195-215). Here, “present value” refers to a 2020 cost basis. For power generation over a 30-year lifespan – the default for geothermal models like GETEM (Entingh et al., 2006) – this basis takes the model out to 2050, a common benchmark year for future projections.

4.2.1 NPV Model

Following the general outline for geothermal cost modeling from previous work (e.g., Augustine, 2009; Beckers et al., 2013; Tester et al., 2006), this thesis considers revenue (R), operating & maintenance costs (OPEX or OM), and capital expenditures (CAPEX or C) as the primary components defining annual cash flow (see Equation 4.1). Capital expenses can be further decomposed into five sub-components associated with exploration, drilling, reservoir stimulation, fluid distribution, and power plant

costs. Likewise, operating expenses subdivide into subsidiary costs for the power plant, wells, and water management:

$$NPV = \sum_{t=1}^T D_t \cdot (R_t - C_t - OM_t),$$

where:

$$C_t = [C_{expl} + C_{dc} + C_{stim} + C_{dist} + C_{pp}]_t,$$

$$OM_t = [OM_{pp} + OM_{well} + OM_{water}]_t.$$
(4.1)

Revenue and expenses are treated on an annual basis, meaning shorter-term fluctuations like price and production seasonality are not explicitly modeled. D_t in Equation 4.1 defines the time-based conversion factor between cash flow for a specific year and discounted cash flow for the basis year (see Equation 4.14).

Revenue

Annual revenue calculations rely on an estimate of power production within a year (W) and the power purchase agreement pricing (p_{PPA}) for that electricity (Entingh et al., 2006):

$$R = W p_{PPA} = (b_e \dot{m}) p_{PPA}.$$
(4.2)

Brine effectiveness (b_e) describes the electricity output per unit flow ($\dot{m}$) of produced brine and depends on the production temperature of the brine. The GETEM model uses an empirically-defined relationship with brine temperature ($^{\circ}\text{C}$) to determine brine effectiveness ($\text{W}\cdot\text{h}/\text{kg}$) (Entingh et al., 2006, p. 62):

$$b_e = C_0 + C_1 T_{prod} + C_2 T_{prod}^2 + C_3 T_{prod}^3 + C_4 T_{prod}^4,$$

$$C_0 = 9.41376,$$

$$C_1 = -0.182542,$$

$$C_2 = 0.0001765735,$$

$$C_3 = 0.000012204486,$$

$$C_4 = -0.0000000335559.$$
(4.3)

In the GETEM interface, users choose to either determine electricity output for a specified input temperature and flow rate or derive the flow rate required to meet a pre-determined power sales capacity for the same fluid temperature (Entingh et al., 2006). Since the Climeon analog has a known net capacity of 150 kW per unit or 1.05 MW for each Power Block (Climeon, 2021), and standard flow rates are provided in models like GETEM, both options are explored in Chapter 6.

Exploration Capital Expenses

Costs for exploration activities are estimated by the same method defined for the 2012 GETEM model (Equation 4.4) (EERE, 2012):

$$C_{expl} = PPI \cdot [1.12(\$1M + 0.6C_{dc})] . \quad (4.4)$$

This relationship assumes slim hole (3–6" diameter) drilling for exploration at a 60% discounted cost compared to standard-sized ($\geq 8.5''$ diameter) geothermal wells (EERE, 2012). The constant \$1M term accounts for pre-drilling costs, including fieldwork, geophysical surveys of field structure, and interpretation of results (EERE, 2012). Technical and office support is covered by an additional 12% applied to the estimate (EERE, 2012). Total exploration costs are converted to a 2020 cost basis using the Producer Price Index (PPI) for electric power generation from the U.S. Bureau of Labor and Statistics (U.S. BLS, 2021b).

Drilling Capital Expenses

Geothermal drilling costs differ from traditional oil & gas wells due to differences in hole diameter, thermal and geochemical conditions, and the strength and abrasiveness of the target formations (Lowry, Finger, et al., 2017). Here, drilling capital expenditures rely on an empirical cost curve described by Beckers et al. (Eq. 4, 2013):

$$C_{dc} = PPI \cdot [1.65 \times 10^{-5} MD^{1.607}] , \quad (4.5)$$

where C_{dc} is measured in \$M and MD refers to well measured depth in meters. Each power-plant module will require an injector-producer pair, so this represents one-half of the drilling cost per module. Drilling costs are converted to a 2020 cost basis using the PPI for electric power generation (U.S. BLS, 2021b).

Note that Equation 4.5 was derived for well depths of 1600–9000 m. Assuming an average geothermal gradient of 100 K/km (Table 4.2), the wells considered for the present study could extend slightly shallower than this range, so this should be viewed as a minimum drilling estimate. The probabilistic model considered later in the present study includes variability in geothermal gradient and drilling costs for a more comprehensive treatment of both variables.

Simulation Capital Expenses

EGS at Lightning Dock requires stimulation of the Horquilla reservoir to create fluid pathways for thermal extraction. The stimulation cost estimate used in the present study comes from the recent GeoVision analysis (Lowry, Finger, et al., 2017):

$$C_{stim} = \$1,250,000. \quad (4.6)$$

Since this represents a recent ballpark estimate, no cost basis conversion was applied in the model. In fact, the value in Equation 4.6 may be high since it includes the cost of water, which may not be a factor at Lightning Dock with the availability of hydrothermal brine from adjacent power-plant operations. The model assumes stimulation is only performed for the injection well in each injector-producer pair, so this represents a *per module* value.

Distribution Capital Expenses

Fluid-distribution costs include the entire surface piping system between the wells and power-plant modules. The present study uses the same estimate included in the

GEOPHIRES model (Beckers et al., 2013):

$$C_{dist} = PPI \cdot [\$50,000q_{in}], \quad (4.7)$$

where q_{in} is the heat input from the produced brine and PPI converts this 2012 value to a 2020 cost basis. The 2nd Law efficiency or thermodynamic efficiency (η) governs how much heat from the input fluid can be converted to work. Beckers (2016, pp. 39–41) provides an estimate for sub-critical ORC power plant thermodynamic efficiency as follows:

$$\eta = 0.002713T_{prod} - 0.0918. \quad (4.8)$$

Combining Equations 4.7 and 4.8 allows distribution costs to be calculated in terms of electricity production:

$$C_{dist} = PPI \cdot [\$50,000W \cdot (0.002713T_{prod} - 0.0918)], \quad (4.9)$$

where W is the electricity output and T_{prod} is the temperature of the produced brine. Under the scenario where modular power plant units are pre-fabricated and directly provided by a company like Climeon, fluid distribution may be included in the installation fees. Distribution capital expenditures would therefore be subsumed by power plant costs and C_{dist} would reduce to zero. However, without confirmation of the fee break-down structure from Climeon, the model described here relies on Equation 4.9.

Power Plant Capital Expenses

Power plant costs for a modular installation remain a source of significant uncertainty for this cost model. The GEOPHIRES model implements a temperature-variable cost estimate first described by Tester et al. (2006) for a binary-cycle power plant (Beckers et al., 2013). Schochet & Cunniff (2001) predicted produced fluid temperatures of 280–320°F (137–160°C) for the Lightning Dock EGS reservoir, which equates to \$1565–\$1694 per kW by the GEOPHIRES estimate. Converted to a 2020 cost basis (U.S. BLS, 2021b), this amounts to \$2230–\$2415 per kW.

If power plant capacity is modularized with pre-fabricated units like the Climeon Power Block concept, economies of scale should reduce the cost of construction and installation. Unanswered inquiries to that company left this rationale unconfirmed. Nevertheless, the author chose to assume a round-number estimate accounting for a modularity discount (Equation 4.10). This estimate should be replaced by more accurate numbers when those values become available:

$$C_{pp} = \$2,000W, \quad (4.10)$$

where W is the electricity output of the plant in kW and C_{pp} is measured against a 2020 cost basis. Pump costs are assumed to be included in this expense.

Well Operating Expenses

Operations and maintenance costs per geothermal well combines labor with a fraction of the drilling expenses (Equation 12, Beckers et al., 2013).

$$OM_{well} = 0.25C_{labor} + 0.01C_{dc} \quad (4.11)$$

where C_{labor} refers to labor costs in Table 4.1. Equation 4.11 covers expenses for a single well and must be doubled for injector-producer pairs associated with each power plant module.

Power Plant Operating Expenses

Power-plant operating expenses follow the relationship used by the GEOPHIRES model based on a previous GETEM formulation (Equation 9, Beckers et al., 2013):

$$OM_{pp} = 0.75C_{labor} + 0.015C_{pp}. \quad (4.12)$$

Here, C_{labor} refers to labor costs scaled to power plant production. The values in Table 4.1 follow Beckers et al. (Equation 10, 2013), updated to a 2020 cost basis using the Employment Cost Index (ECI) for total compensation for private industry

utilities workers ([U.S. BLS, 2021a](#)).

Electricity Output (MW)	Labor Costs (2020 \$)
< 5	326,000
[5, 10)	1,073,000
[10, 20)	1,460,000
[20, 40)	2,167,000
40+	2,581,000

Table 4.1: Power-plant labor costs by plant capacity ([Beckers et al., 2013](#)).

Water Operating Expenses

Water expenses refer to make-up water that replaces subsurface losses to the reservoir. The value applied here comes directly from the GETEM model ([EERE, 2012](#)):

$$OM_{water} = PPI \cdot [\$300V_{loss}] , \quad (4.13)$$

where V_{loss} is water loss in units of acre-feet and PPI converts this estimate to a 2020 cost basis. This operating cost could be alleviated by directly using excess water from the Lightning Dock hydrothermal operations. The cost model includes it for a more conservative cost estimate, but there is an argument to remove this cost entirely.

4.2.2 Rate Calculations

The cost model considers four rates when performing the NPV calculation.

Discount Rate

Discount rate defines the time value of money and is held constant throughout the 30-year time period being modeled. Equation 4.14 describes how discount rate re-scales cash flow to a present “discounted” value for the basis year ([de Neufville & Scholtes, 2011](#), p. 199):

$$DCF = \frac{CF}{(1 + r)^t} , \quad (4.14)$$

where DCF is discounted cash flow, CF is the cash flow for a specific year, r is the discount rate, and t represents the number of years between the modeled year and the basis year. Combining this relationship with Equation 4.1, $(1 + r)^{-t}$ replaces D_t as the discount term needed to calculate NPV .

Learning Rate

Learning rate defines the improvement in cost as a result of accumulated knowledge and experience from repeatedly performing an action. In this model, a learning rate only applies to the drilling costs for EGS wells in the expansion project area. Drilling costs progressively decrease based on the following relationship (de Neufville & Scholtes, 2011, p. 213):

$$U_i = U_1 i^\beta, \quad (4.15)$$

where U_1 and U_i are the costs to drill the first and i^{th} wells, respectively, i is the total well count, and β is the slope of the empirical $(\log i, \log U_i)$ curve.

Thermal Drawdown Rate

The thermal drawdown rate defines the progressive cooling of the stimulated geothermal reservoir over time. In the model, the reservoir temperature, and hence the temperature of the produced geothermal brine, decreases with each year of continued production by the relationship defined for GETEM (Entingh et al., 2006):

$$T_n = T_0(1 - d)^n, \quad (4.16)$$

where T_n is reservoir temperature at year n , d is thermal drawdown rate, and n is the number of years since drilling and stimulation activities last took place.

Capacity Factor Degradation Rate

The NREL Cost of Renewable Energy Spreadsheet Tool (CREST) incorporated an additional capacity-factor degradation rate, separate from thermal degradation of

the resource when modeling geothermal LCOE (Gifford & Grace, 2013). This rate accounts for natural long-term production degradation of plant performance over the lifetime of the asset. In this cost model, capacity factor degradation is modeled by reducing the capacity factor as the plant ages by applying a relationship similar to Equation 4.16 for thermal drawdown:

$$C_n = C_0(1 - a)^n \quad (4.17)$$

where C_n is power-plant capacity at year n , a is the degradation factor, and n is number of years since the power plant commenced operations.

4.2.3 Model Parameters

In order to estimate the values for the NPV model components, several parameters related to resource recovery, field and plant operations, and key economic factors were chosen for the cost model. The selected parameters are representative of the Lightning Dock area and limits on components of the system to the best of the author's knowledge.

Resource recovery parameters

Parameter	Value	Source
Ambient surface temperature	15.8 °C	(Dahal et al., 2012)
Average geothermal gradient	100 K/km	(Crowell & Crowell, 2014)
Initial average reservoir temperature	149 °C	(Schochet & Cunniff, 2001)
Cooling in production well	7.5%	(Lowry, Finger, et al., 2017)
Flow rate per producer	40 kg/s	(Entingh et al., 2006)
Thermal drawdown rate	0.5%	(Entingh et al., 2006)
Water loss rate	2%	(Blair et al., 2018)

Table 4.2: Parameters related to resource recovery in the cost model

Field and plant operations parameters

Parameter	Value	Source
Well redevelopment factor	0.85	(Prestidge, 2021)
Plant capacity factor	95%	(Glassley, 2015 , p. 309)
Plant degradation factor	0.5%	(Augustine et al., 2019)

Table 4.3: Parameters related to field and plant operations in the cost model

Economic factors

Parameter	Value	Source
Discount rate	7%	(EERE, 2012)
Drilling cost learning rate	9%	(Lukawski et al., 2014)
Contract rate above wholesale	50%	(PNM, 2014)
Price trigger for flexibility	20%	for Sections 4.4.2-4.4.3
Expansion amount	25%	for Section 4.4.2
Reduction amount	25%	for Section 4.4.3

Table 4.4: Parameters related to economic factors in the cost model

Electricity Price

Electricity prices are referenced from the industrial electricity price forecast for the Mountain region (including New Mexico) provided by the EIA in their Short Term Energy Outlook (STEO) projections out to 2023 ([EIA, 2021c](#)). While industrial pricing differs slightly from wholesale, it more closely mimics wholesale prices than residential or commercial rates and was therefore selected as a wholesale proxy for the cost model. The Forecast Tool in Excel projected prices out to 2050 with 95% confidence bounds (Figure [4-3](#)) using the Exponential Triple Smoothing algorithm for time series data ([Microsoft, 2021](#)). For the static cost model, electricity prices are directly sampled from the forecast for any year when capacity increases and then multiplied by the PPA *Contract rate above wholesale* value listed in Table [4.4](#). This

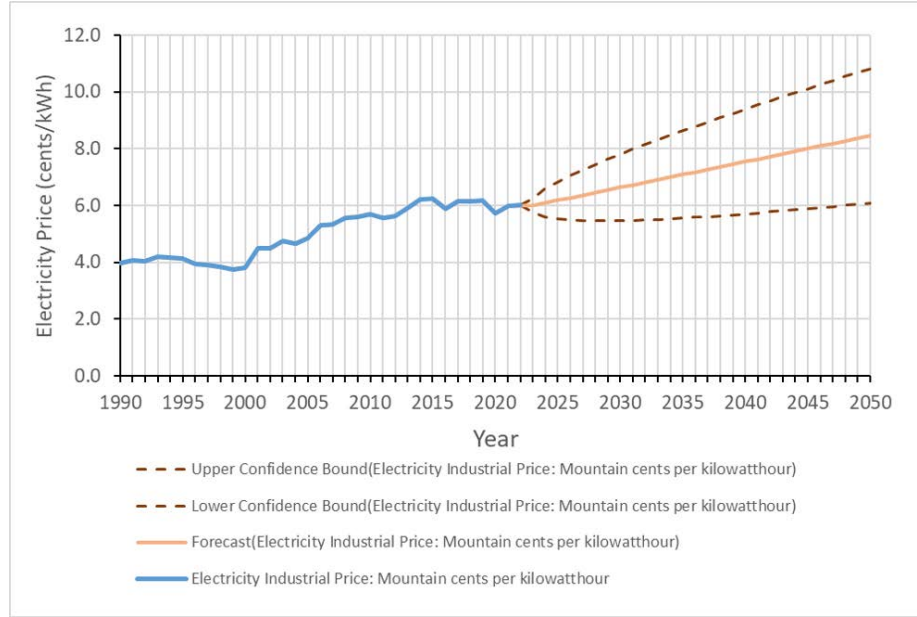


Figure 4-3: Price of electricity from the EIA Short-Term Energy Outlook (EIA, 2021c), forecast out to 2050.

simulates amending the PPA with a local utility whenever new capacity is available for power sales. Electricity pricing is held flat compared to the previous year when no capacity change occurs.

4.3 Probabilistic Cost Model

The model described thus far takes a deterministic approach; parameter values are fixed to their most-likely values when performing the NPV calculation. A probabilistic approach replaces these static values with distributions and repeatedly samples from those distributions to capture an ensemble of results. This Monte Carlo-style simulation can provide a more realistic assessment of system performance. However, all variables in the model have some underlying uncertainty, and defining distributions for every variable would add significant complexity to the model. Variable selection can be performed using sensitivity testing to target the most impactful variables for uncertainty characterization. This helps balance model complexity with representativeness of the physical system.

Recognizing the full probable range of variable values and the scenarios that trigger

them requires a deep understanding of the scientific, engineering, and socio-technical elements influencing a system. For geothermal, subsurface-characterization uncertainties play an important role, but so do uncertainties tied to public policy and market dynamics. The limited focus on the issues listed below should be considered fit-for-purpose for this thesis. Further analysis and discussion with subject matter experts on the local, state, and national levels is advised for similar analysis applied to an active geothermal project.

4.3.1 Model Uncertainties

Carbon Taxation

One proposal for advancing the transition to more renewable and sustainable energy solutions involves a carbon tax levied on fossil fuels. The SIPA Center on Global Energy Policy at Columbia University recently studied three scenarios based on federal agency benchmark taxation rates of \$14/ton, \$50/ton, and \$73/ton CO₂-equivalent with annual percentage-rate increases of 3, 2, and 1.5%, respectively (J. Larson et al., 2018)

(Figure 4-4). Their analysis forecasts the impact on electricity pricing out to 2030, with relatively steady-state implications that depend on the specified carbon tax rate. In all taxation cases, electricity prices increase over the present-day, no-tax scenario, likewise boosting the value of a zero-emissions geothermal power relative to fossil fuel-based options. The selected value range for sensitivity testing was a 0–28% increase in wholesale price, which matches Figure 4-4.

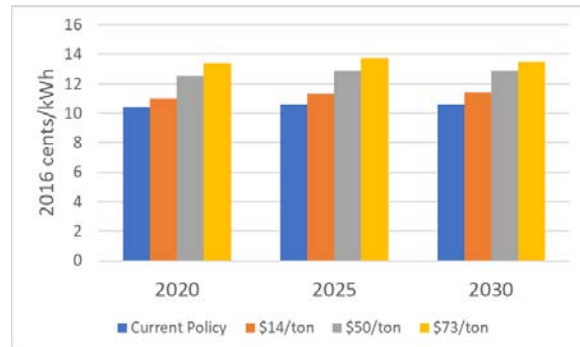


Figure 4-4: National average retail electricity price changes with benchmark levels of carbon taxation, after (J. Larson et al., 2018, Figure 30).

Future Electrification

NREL published a report earlier in 2021 outlining the impact of heightened public trends away from non-electric sources of consumed energy, otherwise known as widespread electrification (Murphy et al., 2021). Some key findings include: (i) end-use natural gas consumption decreases, but so do natural gas prices, which can lead to an increase in natural gas-fueled power plants if no curtailments are imposed by fossil fuel policies, (ii) deployment of renewables will intensify overall, and (iii) local resources, potentially including new renewable power generation facilities, will mitigate the need for long-distance electricity transmission (Murphy et al., 2021).

The issue of national electrification is quite complex, particularly in predicting the interplay between the natural gas market and renewables. Additional dependencies include infrastructure upgrades and development to handle growing capacity, as well as local effects (e.g., permitting, water or electrical transmission, community support) that act as enablers or hurdles to building a new renewable-fueled power plant or expanding on existing power facilities. One way to simplify a model representation

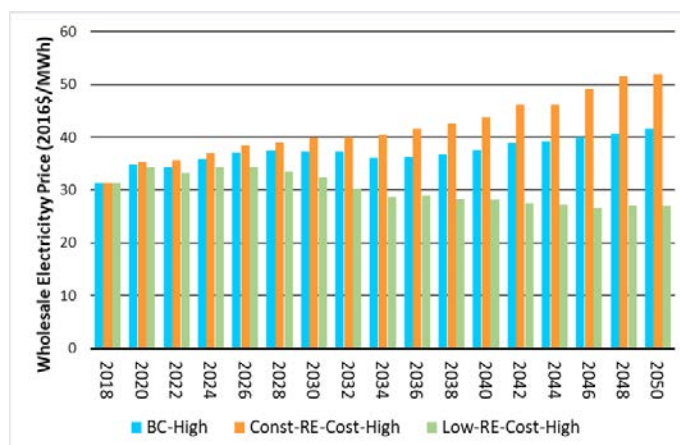


Figure 4-5: Wholesale electricity price forecasts for high future electrification scenarios: base case (blue), constant renewable technology cost (orange), and low renewable technology cost (green). Cases are from the NREL Electrification Futures Study (Murphy et al., 2021). Figure is adapted from interactive plots at <https://cambium.nrel.gov/?project=fc00a185-f280-47d5-a610-2f892c296e51>.

of widespread electrification is to incorporate swings in electricity prices similar to the scenarios shown in Figure 4-5 with the caveat that other related factors (e.g., federal and state-level incentive programs or infrastructure improvements) can also

influence the bottom line for a geothermal project. Based on NREL projections for High Future Electrification cases, wholesale electricity prices in 2050 could vary from 23% lower than baseline for the Low Renewable Technology Costs case to 50% greater for the Constant Renewable Technology Costs case (Figure 4-5). Therefore, -23% and $+50\%$ define the range of price factors used for sensitivity testing.

Climate Change

The 1.5°C climate change goal described in the 2018 IPCC special report (IPCC, 2018) refers to a global average, so more extreme temperature changes are expected to occur on a local scale even if this target gets met. New Mexico, a state already known for semi-arid conditions, is at risk of encountering warming far in excess of 1.5°C by 2050 (Table 4.5). The North Carolina Institute for Climate Studies (NCICS) reports annual average temperatures in NM have already increased 1.1°C since

Region	RCP4.5	RCP8.5
Northeast	2.21	2.83
Southeast	1.89	2.39
Midwest	2.34	2.94
Great Plains North	2.25	2.83
Great Plains South	2.01	2.56
Southwest	2.07	2.67
Northwest	2.03	2.59

Table 4.5: Projected *average* temperatures in $^{\circ}\text{C}$ for mid-century (2036-2065) relative to the 1976-2005 average baseline under lower emissions (RCP4.5) and higher emissions (RCP8.5) scenarios, adapted from (Vose et al., 2017, Table 6.4). New Mexico is included in the Southwest region, highlighted in bold.

the 1970s, and the observed number of days with maximum temperatures of 100°F (37.8°C) or higher is rapidly climbing (Frankson et al., 2019).

Geothermal plant performance is sensitive to the temperature difference between the hot and cooled states of the working fluid. For air-cooled binary plants, changes in ambient temperature could impact overall power plant generation potential. In fact, geothermal power output typically shows seasonality, sometimes with variances of several percentage points in thermodynamic efficiency between winter and summer (Glassley, 2015, p. 52).

[Frankson et al. \(2019, Figure 1\)](#) present the range of model scenarios for temperature changes in New Mexico that ties to [Table 4.5](#). The high-emissions case predicts an average of $\approx 2.7^{\circ}\text{C}$ but goes to a maximum of a $\approx 4.2^{\circ}\text{C}$ increase in state-wide temperatures by 2050. In order to explore a broad range for this variable, an adjustment of $0 - 4.2^{\circ}\text{C}$ was selected for cost-model sensitivity testing.

Drilling Costs

Studies consistently show drilling-related costs are the primary contributor to overall geothermal project expenses — up to 60-75% of the total cost of an EGS project ([Lukawski et al., 2016](#)). According to annual benchmark standards published by [NREL \(2020\)](#), future advances in geothermal drilling technology must address several factors for cost-reduction, including but not limited to: efficiencies in penetration rate and bit life, number of casing intervals, and consumption of drilling materials. All aspects of stimulation also must show improved economics to drive down costs ([NREL, 2020](#)).

Multiple scenario-based drilling cost curves were derived in association with the 2017 GeoVision study as potential updates for the GETEM model ([Figure 4-6](#)) ([Lowry, Foris, et al., 2017](#)). The following list covers the geothermal drilling technology advancements required to justify these curves ([Augustine et al., 2019](#)):

- Bit life and rate of penetration scale from 2–4 $\times$ faster than the base case.
- Number of casing intervals incrementally reduces to just one for the ideal case.
- Mud costs decline, as greater fractions of the well use air-drilling techniques.
- Logging while drilling (LWD) replaces wireline drilling for up to the entire well length in the ideal case.
- Contingency costs related to unexpected or adverse conditions drop from 15% to 0% across the four cases.

In consideration of the cost curves in [Figure 4-6](#) and the anticipated depths for the target reservoir in the Lighting Dock expansion, the selected range of drilling costs

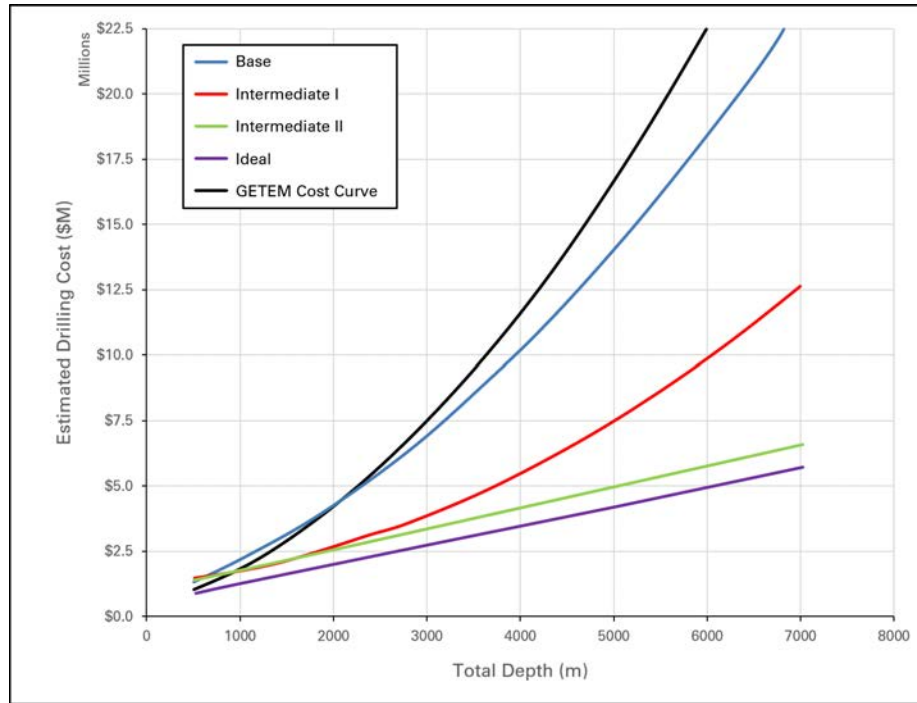


Figure 4-6: Drilling cost curves in \$M US per meters depth for a large-diameter open-hole vertical well, adapted from (Augustine et al., 2019, Figure 8)

to test for model sensitivity is \$1–\$3 million.

Thermal Drawdown Rate

Much like wells for water access or oil & gas production, geothermal wells create drawdown effects from extended operations. The thermal drawdown rate defines how quickly the heat content (enthalpy) accessible within the reservoir fracture network declines over time. As thermal drawdown increases, the temperature of produced fluids decreases, as does the amount of electricity generated by the binary cycle process.

Recent EGS studies suggest 0.5–0.6%/year is an appropriate drawdown rate for EGS (Augustine et al., 2019), although more pessimistic assessments range from 1.5%/year (Beckers, 2016), to 3.3%/year (Augustine et al., 2006) and 4%/year (Tester & Herzog, 1990). End-cap values of 0.5% and 4% are used for sensitivity analysis.

Well Name	Depth (m)	BHT (°C)	Gradient (K/km)	Reported (K/km)
TG12-7	305	69	177	120
TG56-14	381	36	55	80
TG36-7	305	90	246	-
TG57-7	278	108	335	-
TG52-7	771	137	158	-

Table 4.6: Examples of Lightning Dock geothermal gradients from bottom hole temperatures (BHT). Gradient is a linear approximation assuming 15 °C at the surface. The two Reported gradients use temperature log trends near TD. Table adapted from (Table 1, [Cunniff & Bowers, 2005](#)).

Geothermal Gradient

As a blind geothermal system with no original surface expression, the Lightning Dock discovery only occurred after anomalously high temperature gradients (and boiling water) were found in local agricultural wells ([Crowell & Crowell, 2014](#)). Table 4.6 lists bottom hole temperatures for wells drilled within a 0.5–4.0 km distance from the field-central TFD55-7 well during the 2001–2004 Geothermal Resource Evaluation and Definition (GRED) program ([Cunniff & Bowers, 2005](#)). Note that the Gradient column in the table describes a linear fit from an assumed surface temperature of 15°C to BHT, potentially over-simplifying complex temperature relationships with depth. The gradients in the Reported column come from more reliable assessments near well total depth (TD) as documented by [Cunniff & Bowers \(2003\)](#).

Thermal models calibrated to these wells show local gradients in excess of 300 K/km near the field center, and temperature inversions occur on the flanks of the main Lightning Dock thermal anomaly (see Figs. 23–24, [Cunniff & Bowers, 2005](#)). Away from this fault-centered hydrothermal plume —where an EGS expansion project would be targeted— the thermal field settles into a more traditional monotonically-increasing depth trend. Wells TG12-7 and TG56-14, located 1 km and 4 km away from TFD55-7, respectively, have reported gradients of 80–120 K/km. This range is used for testing model sensitivity to thermal-gradient variations.

4.3.2 Sensitivity Testing

Selecting which of the uncertainties discussed in Section 4.3 should be treated as probabilistic values in the cost model requires testing the sensitivity of NPV to model adjustments bounded by parameter uncertainty ranges. Specifically, NPV is recalculated after changing a single model variable at a time to match the extremal values outlined in the previous discussion. The full NPV calculation follows the model descriptions given in Sections 4.2.1–4.2.3, with the exception of price-related uncertainties, which are modeled by changing the price annually to mimic the most price-sensitive scenario, where PPAs are market-based rather than fixed. No flexibility is assumed for this exercise, so the full power-plant expansion takes place at the start of the 30-year timeline. Also, the sensitivity analysis uses the version of the static model where production flow rate is fixed at 40 kg/s (see Table 4.2) and the capacity per module depends only on the temperature of the produced brine. The reasons for this choice of model structure over one where power output is strictly capped at 5 MW are discussed further in Chapter 6.

The tornado diagram in Figure 4-7 provides a simple visualization of model sensitivity based on NPV calculation results for the different uncertainties, sorted in order of descending importance. Results also appear in Table 4.7. The baseline static model predicts a NPV of $\approx$ \$91,000. Results deviate from baseline most significantly for thermal drawdown rate; when no measurable drawdown takes place, NPV reaches \$18 million, but predicted losses top \$47 million if the drawdown rate is as high as 4%. Uncertainties related to drilling costs, future electrification, and the geothermal gradient all show moderate importance for project NPV. Changes to ambient surface temperature (i.e., due to climate change) and pricing from carbon taxation both have an order of magnitude less influence on project NPV than other uncertainties. Based on this sensitivity test, variables tied to the top 4 uncertainties are treated as random variables for a probabilistic NPV model.

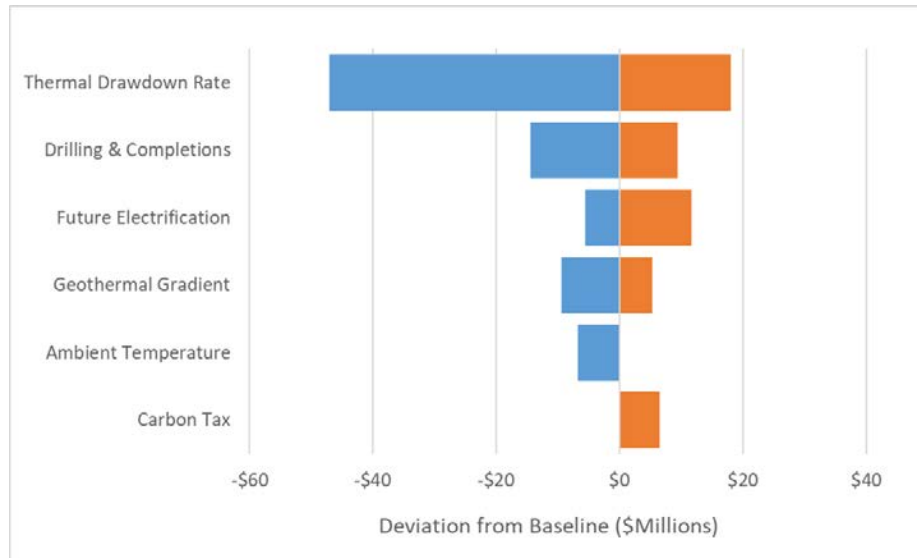


Figure 4-7: Tornado diagram showing NPV model sensitivity to different system uncertainties for the proposed Lightning Dock expansion. X-axis measures deviation from base case when all plant construction takes place at project start and the model parameterization matches Tables 4.2-4.4. Values are in \$M US, where M indicates millions.

	Low (\$M)	High (\$M)	Range (\$M)
Thermal Drawdown Rate	−47.0	18.0	65.0
Drilling Costs	−14.5	9.4	23.9
Future Electrification	−5.5	11.7	17.2
Geothermal Gradient	−9.4	5.4	14.7
Ambient Temperature	−6.7	−0.1	6.6
Carbon Tax	−0.1	6.5	6.6

Table 4.7: Results of NPV model sensitivity testing for different system uncertainties associated with the Lightning Dock expansion. NPV values are listed in \$M US, where M indicates millions.

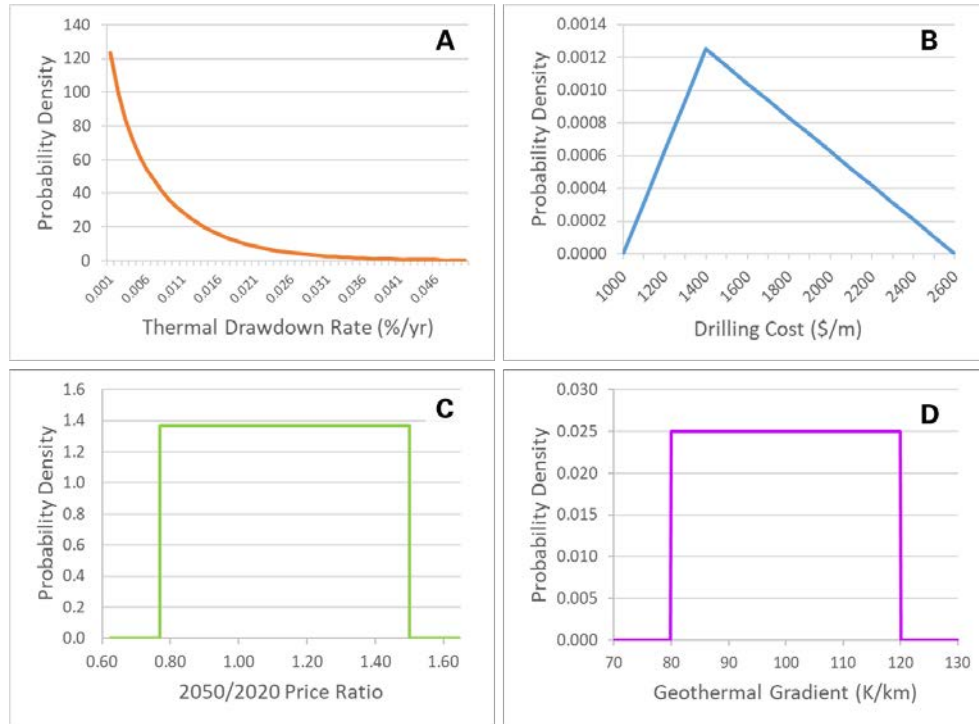


Figure 4-8: Probability distribution functions for A. thermal drawdown rate, B. drilling costs, C. electricity pricing, and D. geothermal gradient.

4.3.3 Probability Density Functions

Having established the key uncertainties, Probability Density Functions (PDFs) can be assigned to the related model parameters for use as part of a stochastic NPV assessment. Running the model multiple times in succession creates a Monte Carlo ensemble of NPV solutions, each representing the model response to a different sampling of the parameter PDFs. The ensemble can be evaluated using a combination of metrics for individual model analysis or comparison with alternative models. Chapter 6 outlines the results of the Monte Carlo approach applied to the Lightning Dock expansion cost model using the PDFs in Figure 4-8 and described below.

Thermal Drawdown Rate

The latest version of GETEM (G. L. Mines, 2016) and its variant in the NREL System Advisor Model (Blair et al., 2018) apply 0.5% as a default value for thermal drawdown rate. Higher decline rates tend to be associated with older references;

1–2% for GEOPHIRES (Beckers, 2016), 3% in a thesis by Augustine (2009), and 4% for work done in the 1990s by Tester & Herzog (1990). The probability density function for thermal drawdown was designed using a beta function such that the P_{50} value aligns with 0.5% annual drawdown rate, and 4.0% represents the $P_{97.5}$ case (Figure 4-8A). Note that the beta function was slightly altered to follow a linear trend from P_{95} to P_{100} to ensure rare extremely high rates in the distribution function do not asymptotically approach over 10% per year. The highest rate supported by the distribution is 5.6% per year.

Drilling Costs

Drilling costs for geothermal wells remain a topic of debate due in part to the small number of direct analogs, particularly for EGS wells, and the documented differences with oil & gas drilling operations. This discrepancy was noted in the 2006 MIT study on EGS, which promoted the use of a dedicated geothermal drilling cost index as a solution (Tester et al., 2006). Nevertheless, numerous and sometimes quite disparate relationships have appeared in the years since; for the 1.0–1.5 km drilling depths considered in the present study, recent estimates range from a low $\approx$ \$500/m (Lukawski et al., 2016) to a very high \$2,800/m (Lowry, Foris, et al., 2017). In order to capture a reasonable spread while recognizing the uncertainty in even defining a distribution shape, geothermal drilling costs are modeled as a triangular distribution (Figure 4-8B). The midpoint value of \$1400/m comes from the predicted well depth and cost in the static model (see Section 4.2.1). The extreme values of \$1000/m and \$2800/m approximate the range shown for depths of 1.0–1.5 km among the drilling cost curves from the recent GeoVision study (Figure 4-6).

Electricity Pricing

Electricity prices in the static cost model are determined by the EIA STEO price forecast for the Mountain region (Figure 4-3) (EIA, 2021c). Two variable price components are superimposed on this trend for the probabilistic model. First, a disruption to the cost curve is simulated by randomly selecting a year between 2020–2050 and

introducing a step change in price to capture the sudden nature of energy-transition events. The magnitude of the step change is determined from a uniform distribution bounded by the range of 2050 High Future Electrification prices relative to the 2020 HFE base case in the Electrification Futures Study (Figure 4-8C) (Murphy et al., 2021). An example of how this randomly-timed, randomly-sampled step change affects the price curve is shown in Figure 4-9A.

Second, volatility could also influence the spot price used for setting a power purchase agreement for a given model year. Using the 95% confidence bounds on the price curve to derive standard deviation, each point in the forecast is replaced by a normal distribution and randomly sampled to produce different price model realizations (Figure 4-9B). This curve will regenerate as a unique price projection for each Monte Carlo realization of the cost model.

Geothermal Gradient

Local spatial variations in geothermal gradient are difficult to characterize with only a sparse sampling of the Lightning Dock area by predominantly shallow boreholes. Subsurface models like those shown in Figures 22–24 of (Cunniff & Bowers, 2005) generally predict a smoothly-varying thermal field in areas without direct observational data. But the complex temperature structure associated with the Lightning Dock hydrothermal plume suggests thermal heterogeneity can exist away from the Animas Valley fault. Uncertainty in geothermal gradient is therefore represented in the cost model by a uniform probability distribution with end points determined by measured gradients from wells TG12-7 and TG56-14 (Figure 4-8D).

Reservoir Temperature

Although not included in the sensitivity testing exercise in Section 4.3.2, the original proposal for EGS production at Lightning Dock by Schochet & Cunniff (2001) noted a range of likely reservoir temperatures in the Horquilla limestone formation. Geothermal power production relies first and foremost on the subsurface temperatures being “mined” by circulating fluids. Uncertainty in initial reservoir temperature is therefore

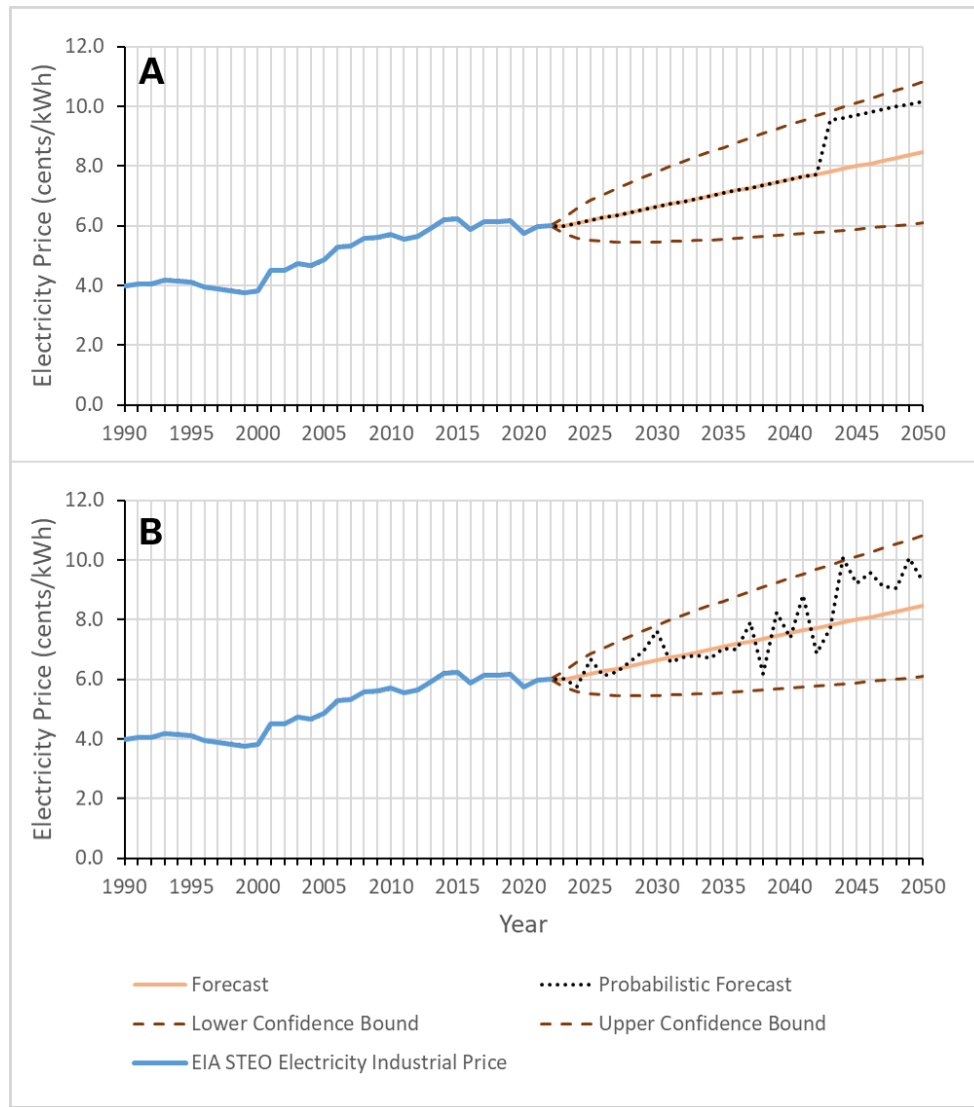


Figure 4-9: EIA STEO electricity industrial prices for the Mountain region (blue), forecast to 2050 (orange) as in Figure 4-3, with A. a randomly-defined step change in pricing and B. added annual volatility using the forecast 95% confidence intervals.

included in the cost model, represented by a uniform probability distribution with bounds determined from the temperatures proposed by [Schochet & Cunniff \(2001\)](#) (Figure 4-10).

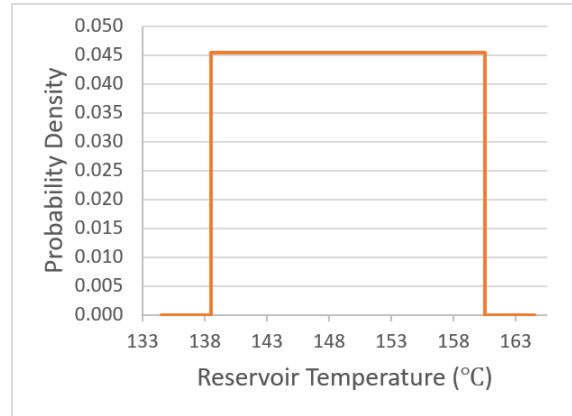


Figure 4-10: Reservoir temperature probability density function. The value range is based on values originally proposed by [Schochet & Cunniff \(2001\)](#).

4.4 Flexibility with Design Options

The addition of probability density functions to the cost model for Monte Carlo simulation provides a means of testing the model response to uncertainties in the system. But this probabilistic Base Case model still remains inflexible in the face of emergent conditions that would trigger actions in a real-life scenario. These actions are sometimes characterized as design options that, like financial options, can be exercised in the future if doing so might benefit system stakeholders ([de Neufville & Scholtes, 2011](#), p. 270–272).

Design options define decision rules for how a model behaves based on past observations. Decision rules may act independently or be chained together to mimic complex system flexibilities that can reveal otherwise hidden financial value. The following scenarios extend the Base Case model with one or more decision rules. Chapter 6 examines how implementing these rules impacts predicted model ENPV, target curves, and other forecast performance measures described in Section 6.2.

4.4.1 Redevelop Only Case

Sensitivity testing revealed thermal-drawdown rate as the most important uncertainty governing cost-model performance (Figure 4-7). Over time, cooling of the reservoir by injected fluids results in declining input temperatures to the binary cycle plant and hence lower electricity production. If the latter drops below a certain level, redrilling or restimulation of the reservoir is required to ensure generation rates remain within a reasonable (or profitable) range. The GETEM model tracks thermal decline and discounts power plant performance until the accessible reservoir temperature reaches a certain threshold defined by (Entingh et al., 2006):

$$\Delta T = (T_{threshold} - T_i) = 0.21T_i - 12.2. \quad (4.18)$$

Combining this equation with a harmonic decline curve assumption results in the following relationship for the time before the maximum acceptable decline is reached:

$$\begin{aligned} t &= \frac{1}{d} \left(\frac{T_i}{T_{threshold}} - 1 \right) \\ &= \frac{1}{d} \left(\frac{T_i}{1.21T_i - 12.2} - 1 \right). \end{aligned} \quad (4.19)$$

To counteract the negative impact of this decline, a full field re-drill campaign is triggered in cost models like GETEM. This may occur several times over the lifespan of a geothermal power plant depending on the drawdown rate (d), although GETEM freezes re-drills in the final 5 years to ensure no redevelopment cost is incurred just prior to end of life for the facility (Entingh et al., 2006). This methodology is applied here using the following decision rule.

Redevelopment Decision Rule

1. Determine the temperature threshold for viable power production using Equation 4.18 and the initial reservoir temperature.
2. Calculate the number of years until the temperature threshold is reached based on the thermal drawdown rate and Equation 4.19. This defines the redevelop-

ment interval for the field.

3. In the annual cash flow analysis, determine if time since installation of any power plant modules is a multiple of the redevelopment interval. If so and the year being evaluated does not fall within the final 5 years of the project lifespan:
 - (a) Identify how many modules need to be redeveloped. Multiply this by 2 to define the number of wells being sidetracked (or redrilled). This assumes the wells are in pairs for each module.
 - (b) Calculate CAPEX for redevelopment by multiplying the drilling costs per well by the number of wells being reworked, then discount by the pre-determined redevelopment factor. Scale this by the learning rate discount based on the number of wells already drilled since field operations began.
 - (c) Update the running tally of wells drilled or redrilled to include wells from this redevelopment effort.
 - (d) Reset the produced brine temperature to the initial reservoir temperature.

4.4.2 Redevelop & Grow Case

Redevelopment of the geothermal field is primarily a mitigation against loss of accessible resource as thermal drawdown impacts the flow paths between wells. Capturing upside potential is equally important. The Redevelop & Grow case recognizes that up-swings in wholesale electricity prices may signal a comprehensive shift in long-term energy pricing due to influences like societal shifts toward electrification. To take advantage of the opportunity, this case considers a price change threshold (*Price trigger for flexibility*, see Table 4.4) as the trigger for installing additional geothermal power plant modules and renegotiating the PPA with the local utility company. The scenario assumes a flat percentage increase in capacity (*Expansion amount*, see Table 4.4) and universal success in establishing new power agreements at a set mark-up percentage above wholesale (*Contract rate over wholesale*, see Table 4.4). The field redevelopment decision rule outlined for the Redevelop Only case remains intact, and

another decision rule for design flexibility in modular growth is as follows.

Capacity Growth Decision Rule

1. In the annual cash-flow analysis, look up the predicted wholesale electricity price for the current year and determine the deviance between this price and the wholesale price used in the last PPA contract (i.e., when the last capacity change took place). Here, deviance is defined as: $\frac{\text{current price} - \text{past price}}{\text{past price}}$.
2. If the deviance exceeds the pre-set price trigger and the year being evaluated does not fall within the final 5 years of the project lifespan:
 - (a) Multiply the number of operating power plant modules in the field by the pre-set expansion parameter to determine the number of modules to add.
 - (b) Calculate CAPEX for drilling an injector-producer pair for each added module. Scale this value by the learning-rate discount based on the number of wells previously drilled or redrilled in the field.
 - (c) Update the tallies for the number of modules in the field and the number of wells drilled or redrilled to include the added modules and their wells.
 - (d) Determine the new PPA contract price by multiplying the predicted wholesale electricity price for the current year by the pre-determined contract rate above wholesale factor.

4.4.3 Full Flexibility Case

Price swings can go the opposite direction as well. The NREL Electrification Futures Study (Murphy et al., 2021) identified scenarios where electricity prices fall between 2020 and 2050, so having a means of addressing a future with tighter margins would be a useful flexibility. In the Full Flexibility Case, field redevelopment with thermal degradation and capacity increases in response to price surges remain in effect. In addition, a sudden drop in electricity prices (*Price trigger for flexibility*, see Table 4.4) serves as a trigger for the power plant operator to remove or decommission a number of binary cycle modules (*Reduction amount*, see Table 4.4). Since modules operate

independently with their own injector-producer couplet, they can be individually decommissioned with no impact on other installed modules in the aggregate facility. Additional cost savings might be realized if the modules are leased and equipment can be returned early to the vendor when no longer in use, although for the sake of simplicity, this option has not been included in the cost model. The decision rule for price-based decommissioning of active modules is as follows.

Capacity Reduction Decision Rule

1. In the annual cash-flow analysis, look up the predicted wholesale electricity price for the current year and determine the deviance between this price and the wholesale price used in the last PPA contract (i.e., when the last capacity change took place). Here, deviance is defined as: $\frac{\text{current price} - \text{past price}}{\text{past price}}$.
2. If the deviance is negative and exceeds the pre-set price trigger (in magnitude), and the current year does not fall within the final 5 years of the project lifespan:
 - (a) Multiply the number of operating power plant modules in the field by the pre-set reduction-amount parameter to determine the number of modules to decommission.
 - (b) Reduce the count of operating modules in the field to account for taking these modules offline.
 - (c) Make sure OPEX is only calculated for still-operating power plant modules.
 - (d) Do not reduce the running tally of wells drilled or redrilled. Shutting down modules does not negate the learning experience of drilling the wells associated with those modules.

4.5 Recap

This chapter covered the methodology for using cost models to mitigate the risk of expanding an existing power facility with geothermal production. The Lightning Dock KGRA and present-day power plant are the subject of a hypothetical 5 MW

EGS expansion project. The model assumes a 30-year useful life for the expansion, and construction is based on the deployment of pre-fabricated binary-cycle modules with one injector-producer pair per module.

The modeling strategy follows a step-wise increase in model complexity:

1. Start with a static model that calculates NPV based on estimates for Revenue, CAPEX, and OPEX. The model includes thermal drawdown of the reservoir, power-plant degradation, a learning rate for drilling costs, and a discount rate for the time value of money. All parameters are pre-defined.
2. Replace the static model with a probabilistic one by assigning probability density functions to key model parameters. Sensitivity testing identifies the key parameters to treat as uncertain in the model. Results are obtained through Monte Carlo sampling to build a solution ensemble, evaluated by multiple measures like ENPV, NPV percentiles, and target curves. Variables defined with PDFs in this analysis include: thermal-drawdown rate, drilling costs, electricity pricing, geothermal gradient, and reservoir temperature.
3. Incorporate flexibility with design options as decision rules in the probabilistic model. The decision rules being evaluated in this analysis include: field redevelopment due to thermal drawdown, growth in capacity when prices surge, and capacity reductions when prices decline.

Appendix [B](#) presents the cost model spreadsheets created to apply this strategy to the Lightning Dock case study. The results generated from using these spreadsheets as risk assessment and mitigation tools are explored in Chapter [6](#).

Chapter 5

Geothermal Exploration

Machine-Learning Results

Chapter 3 outlined the machine-learning and uncertainty-analysis strategy for characterizing geothermal gradient, a proxy for the heat-risk element, across the South-western NM study area. This chapter reviews the results of applying that strategy with the curated data set described in Appendix A. In addition, this chapter places insights from the model results and uncertainty evaluation in context with mitigating risks associated with geothermal exploration.

5.1 Logistic Regression

5.1.1 Hyperparameter Tuning

The logistic regression (LR) model used in this analysis (Pedregosa et al., 2011) includes a single tunable hyperparameter, C . Rather than rely on the pre-split training and validation subsets defined in Section 3.2.2 to tune this hyperparameter, the subsets were re-combined and stratified-sampled as part of a 10-fold CV process. ROC AUC OvR was used as the scoring metric. Results vary for the different input data sets; CV results for WDS show a clear maximum AUC marking the optimal value for C , whereas CV for WDS4 and WDS8 demonstrate a leveling-off trend, and the

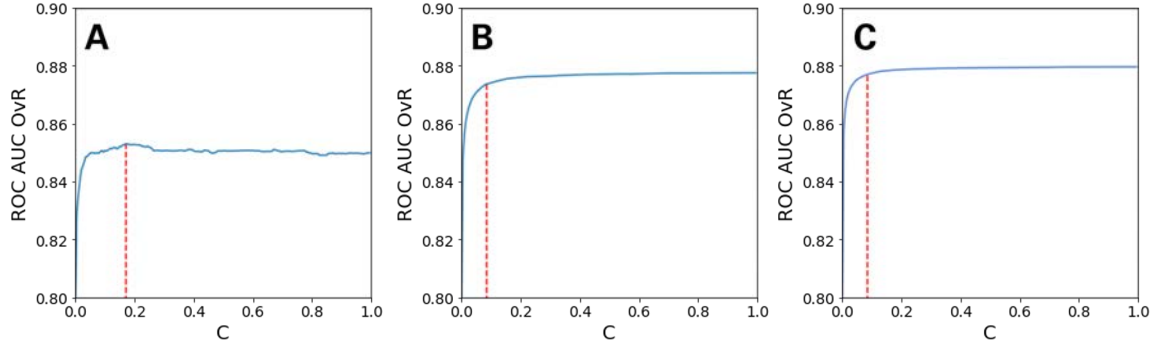


Figure 5-1: Tuning plots for logistic regression C hyperparameter, based on ROC AUC OvR values for A. WDS, B. WDS4, and C. WDS8. WDS value selected from the plot maximum. Selections for WDS4 and WDS8 target the elbow of the curve.

optimal value must be selected near the elbow (Figure 5-1). The chosen C values for WDS, WDS4, and WDS8 are listed in Table 5.1.

Out-of-sample AUC values calculated on the testing subset indicate that WDS4 has the best performance of the three data sets. Table 5.2 lists the feature coefficients for each of the four OvR classifiers that make up the multi-class LR model. Figure 5-2 illustrates these coefficients as a stacked bar chart, where

each color bar depicts the magnitude of a coefficient for one of the OvR classifiers. The absolute length of

the stacked bar for each feature, composed of the coefficient bars for all four classifiers, indicates the relative influence that feature has on the model prediction. Sorting the features by total stacked bar length quickly communicates the most important features for the logistic regression classifier: Si Geothermometer Temperature, Basement Depth, Drainage Density, Spring Density, and Volcanic-Dike Density.

	WDS	WDS4	WDS8
C	0.170	0.085	0.085
$\text{Accuracy}_{\text{train}}$	0.703	0.701	0.709
$\text{Accuracy}_{\text{test}}$	0.611	0.709	0.701
$\text{AUC}_{\text{train}}$	0.892	0.877	0.882
AUC_{test}	0.785	0.891	0.875

Table 5.1: Logistic regression hyperparameter tuning results for each data set. Accuracy and AUC model statistics are split into train (in-sample) and test (out-of-sample) values.

	Class0 OvR	Class1 OvR	Class2 OvR	Class3 OvR
SiGeothermometry	-0.81	-0.17	0.33	0.65
BasementDepth	-0.75	-0.19	0.47	0.48
Drainage	-0.61	0.19	0.66	-0.25
Springs	0.28	0.02	-0.83	0.53
VolcanicDikes	-0.39	0.31	0.41	-0.33
Boron	-0.37	0.67	-0.14	-0.16
DEM	0.03	0.61	-0.44	-0.21
DosageRate	-0.35	0.21	-0.27	0.40
HeatFlow	-0.08	0.31	-0.49	0.26
StateFaults	0.19	-0.54	0.09	0.26
CrustalThickness	0.14	0.40	-0.30	-0.24
Vents	0.54	-0.15	-0.29	-0.10
StrainRate	0.14	-0.50	0.27	0.08
Earthquakes	-0.09	0.35	-0.40	0.14
QFaults	0.13	0.34	0.00	-0.47
Lithium	0.30	-0.47	0.05	0.12
Precipitation	0.21	0.20	-0.19	-0.22
Gravity	0.12	0.24	-0.29	-0.08
WTGrad	0.02	-0.26	0.25	-0.01
GravityGrad	0.14	-0.14	0.10	-0.10
WTDepth	0.10	-0.10	0.13	-0.13
Magnetic	0.16	0.03	0.00	-0.19
MagneticGrad	-0.04	-0.03	-0.01	0.09
DEMGrad	-0.02	-0.07	0.03	0.05

Table 5.2: Feature coefficients for each of the four OvR classifiers that make up the multi-class logistic regression classifier for WDS4. Values are rounded to the nearest hundredth for display purposes.

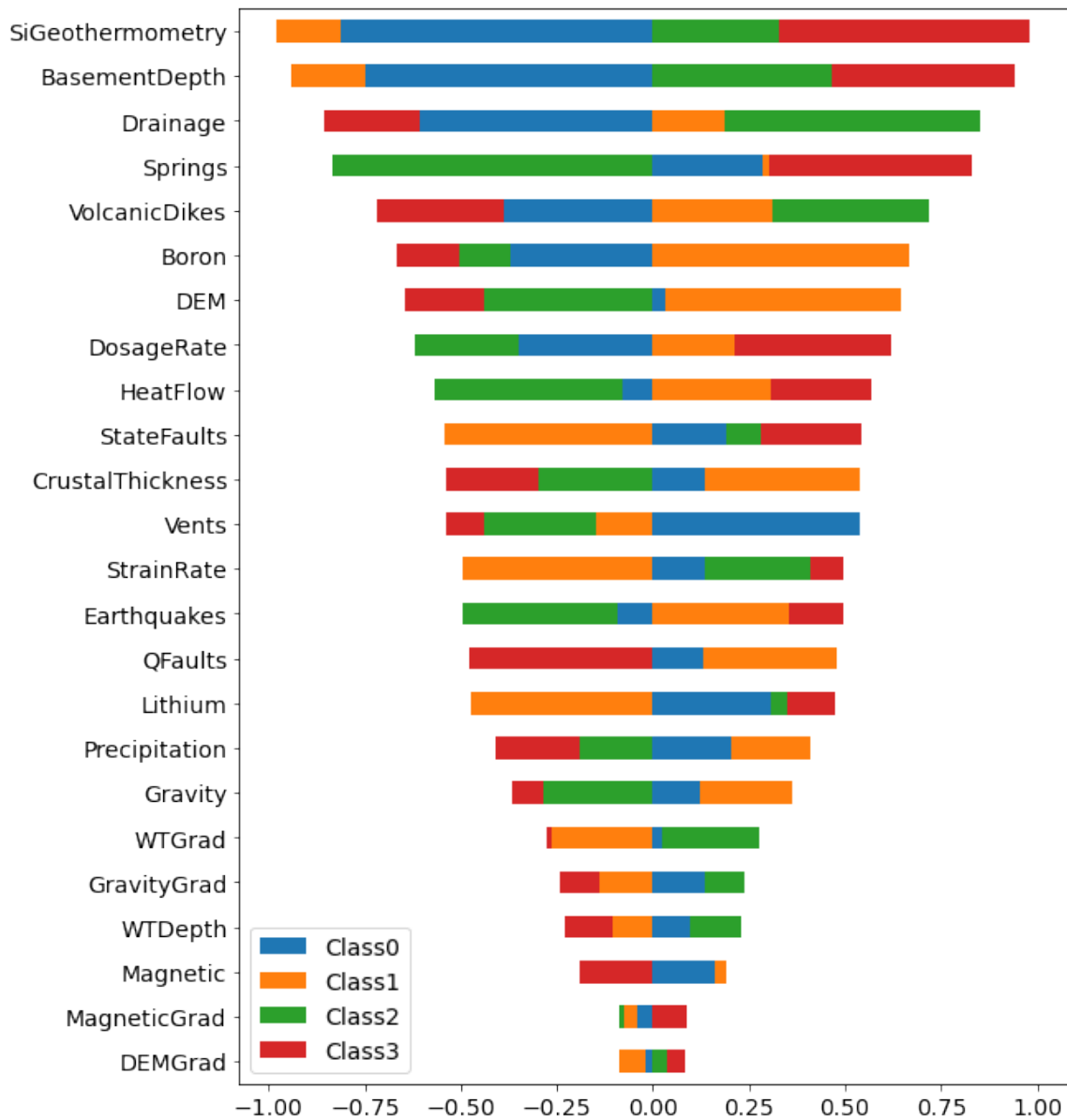


Figure 5-2: Stacked bar chart of logistic regression coefficient values for OvR classifiers trained on WDS4. Values are also listed in Table 5.2. Coefficient bars are color-coded by individual classifier. Total stacked bar length for a feature is a proxy for overall importance of that feature to the classification.

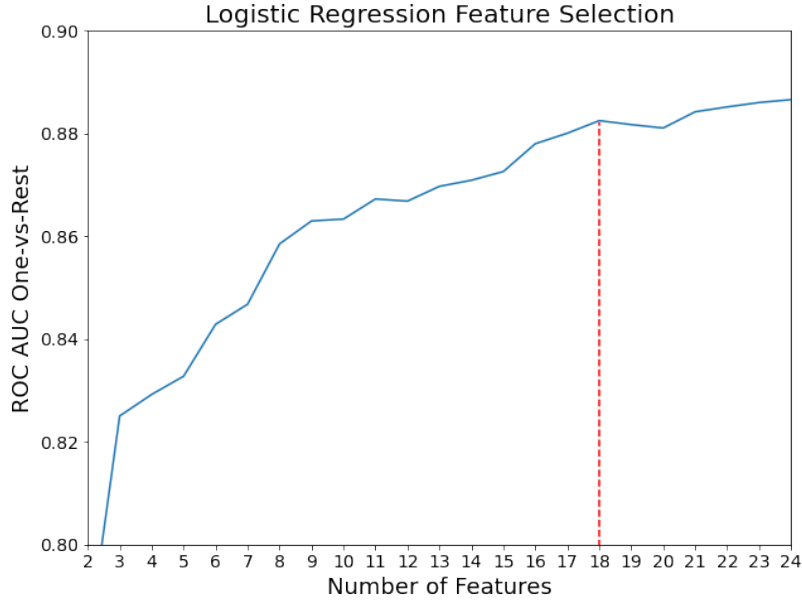


Figure 5-3: Logistic regression feature selection using RFE and WDS4. Red dashed line indicates the chosen number of features to use for the LR model.

5.1.2 Feature Selection

Figure 5-3 presents the results of Recursive Feature Selection (RFE, see Section 3.3.3) applied to the LR model for WDS4. Based on the plot, a local peak in AUC occurs when 18 features are used. Adding the remaining features results in small gains in AUC, but with diminishing returns for six additional features of complexity. Using this threshold, the data layers removed from the model include: DEM Gradient, Gravity Gradient, Magnetic Anomaly, Magnetic-Anomaly Gradient, Water-Table Depth, and Average Precipitation. Note that Average Precipitation appears higher on the coefficients plot (Figure 5-2) than other features that were not removed. Since RFE iteratively removes predictors and refits the model, relative coefficients can change, particularly if there was collinearity with a removed variable.

5.1.3 Optimized Model Results

A final LR model trained on WDS4 was constructed using the tuned `C` hyperparameter and reduced feature set from RFE. The confusion matrix shows moderately good

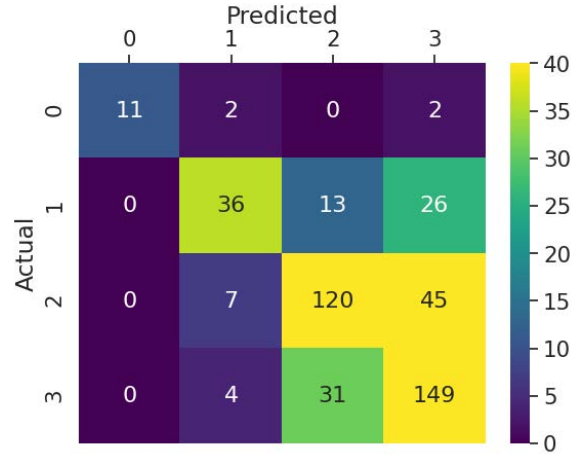


Figure 5-4: Confusion matrix for the tuned LR model trained on WDS4.

model performance (Figure 5-4). Correct predictions (TP) for each of the four classes of geothermal gradient outnumber the misclassifications (FP) for those classes. The model appears to struggle most with differentiating between class-2 mid-grade (40–60 K/km) and class-3 high-grade (> 60 K/km) gradient locations (31 + 45 misclassifications), although there are also a large number (26) of false high-grade assignments to class-1 low-grade gradient points.

Figure 5-5 plots the macro average, micro average, and individual class ROC curves. Class-0 (non-thermal) predictive ability is quite high, pulling the micro-average AUC up to 0.88. The macro AUC value of 0.85 is more aligned with the performance for other classes, which range over an AUC of 0.79–0.83. The trade-off between class 2 and class 3 is apparent in how the curve shapes mirror each other approximately, with respect to the anti-diagonal line $\text{TPR} = 1 - \text{FPR}$.

Model predictions for the study area are generated by passing the FDS through the final trained model. Class predictions are plotted in Figure 5-6. High-grade geothermal-gradient patches are concentrated to the southeast and through the center of the AOI. Smaller high-grade regions are observed along the southwest state boundary, following the Rio Grande to the northeast, and a smaller patch directly to the north. Comparing this result to the Southwestern NM PFA geothermal-gradient data layer from Bielicki et al. (2015), high-grade predictions match in general spatial

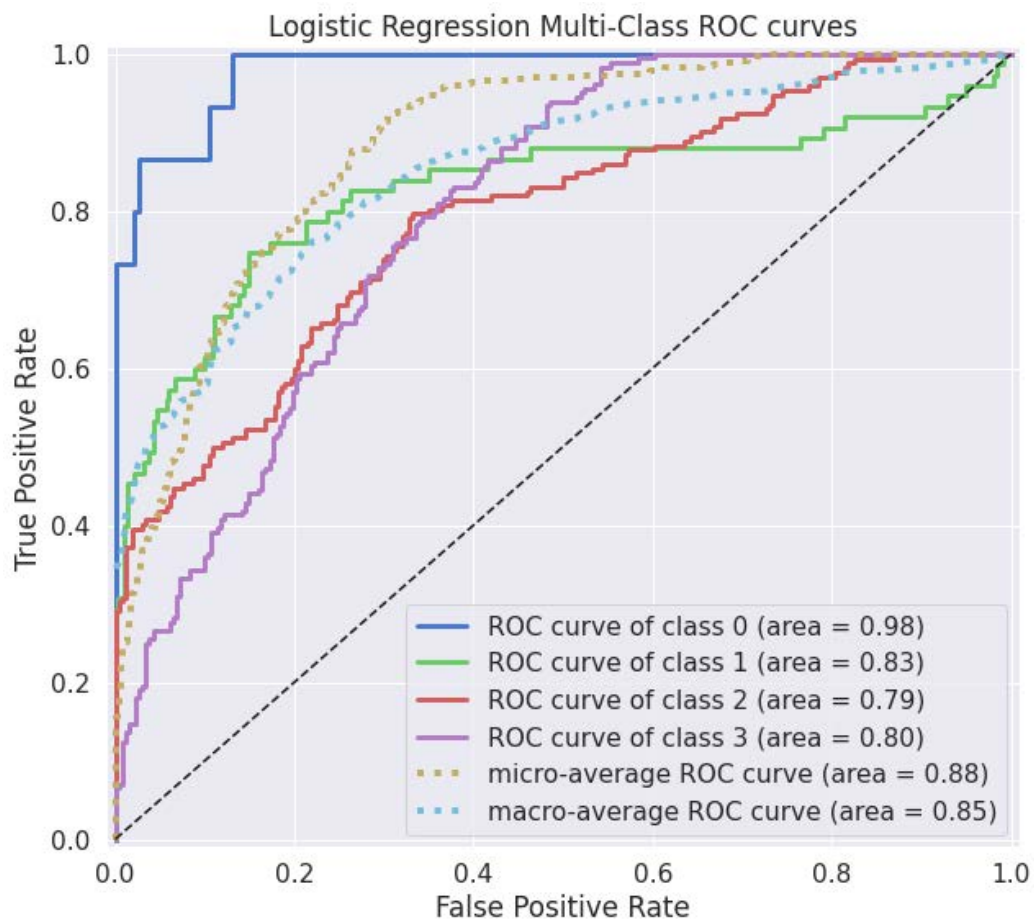


Figure 5-5: ROC curves for the tuned LR model trained on WDS4.

location except for the predicted patches to the north. The LR model tends to predict more widespread and spatially-continuous high-grade regions, while under-predicting lower gradient regions to the north (Colorado Plateau), east (Great Plains), and mid-AOI near the Rio Grande, compared to the PFA layer. Recall the PFA geothermal gradient layer is an interpolated layer from well data and not ground-truth (see Section A.27). Here it simply serves as a convenient reference for comparison.

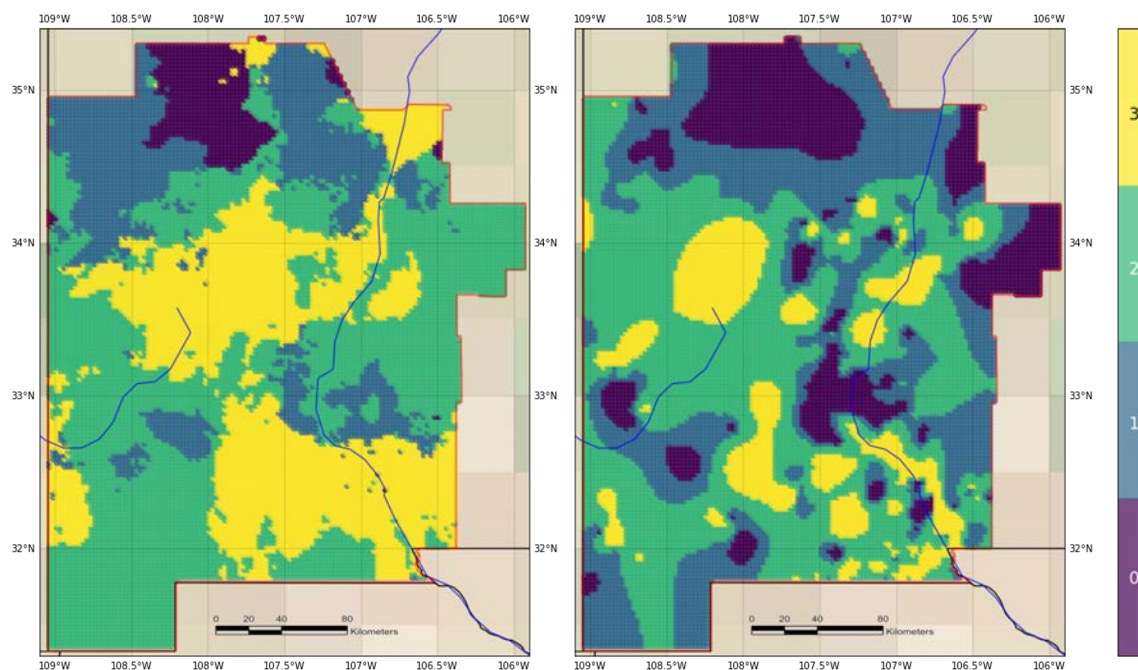


Figure 5-6: Left: Map predictions of geothermal-gradient class from the tuned LR model trained on WDS4. Right: geothermal-gradient feature layer from Southwestern NM PFA study (Bielicki et al., 2015).

5.2 Decision Trees

5.2.1 Hyperparameter Tuning

Stepping up in complexity, the scikit-learn version of the decision tree (DT) classifier has over ten adjustable hyperparameters for tuning performance (Pedregosa et al., 2011). Here, six hyperparameters are tuned using the stratified k -fold CV method described in Section 3.3.2, with 10 folds and the multi-class OvR ROC AUC scoring

metric. The tuning process relied on defaults for the scikit-learn *DecisionTreeClassifier* to define initial values until a specific hyperparameter was tuned. Figures below illustrate the tuning results for the WDS4 data set.

First, `max_depth` and `criterion` were tuned together. `max_depth` limits tree expansion by capping the number of parameter evaluations (tree nodes) considered before a classification label assignment. `criterion` refers to the quality metric used for tree construction, i.e., Gini index or Entropy. The similarity of AUC curves in Figure 5-7 illustrates a relative insensitivity to `criterion`, while `max_depth` plays a stronger role in classifier performance. A maximum AUC score is observed with a `max_depth` of 8 and `criterion` choice of Entropy.

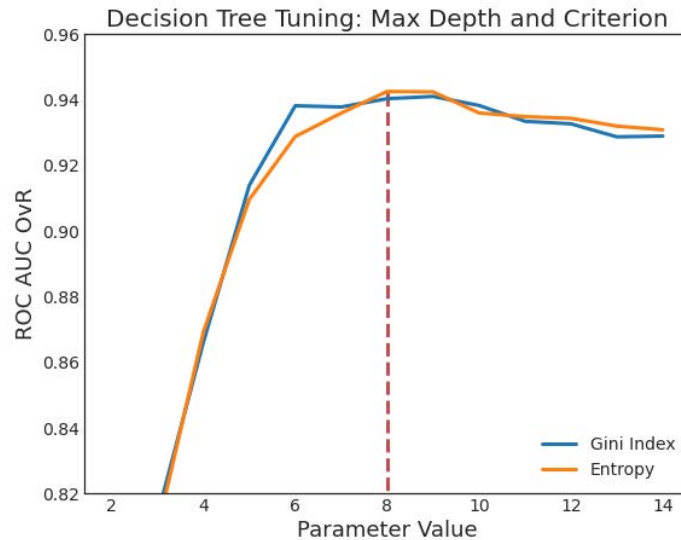


Figure 5-7: Results from stratified k -fold cross validation tuning of the `max_depth` and `criterion` hyperparameters using WDS4. The red dashed line indicates the selected `max_depth` value.

Next, `min_samples_leaf` and `min_samples_split` were tuned in succession. The former defines the minimum number of samples from the training set that must be assigned to a leaf node for that leaf to remain in the tree. The latter sets a minimum number of training-set samples that must be assigned to a node before that node can be considered for a split. Figure 5-8 illustrates the selected hyperparameter values, defined by maxima in the AUC vs. hyperparameter value plots. The insensitivity of the classifier to low values of `min_samples_split` demonstrates the cascading

influence of the hyperparameters in this tuning flow. When `min_samples_leaf` is set to 8, a split can only occur when the node being split has at least 16 observations, so any `min_samples_split` value under 16 does not influence tree construction.

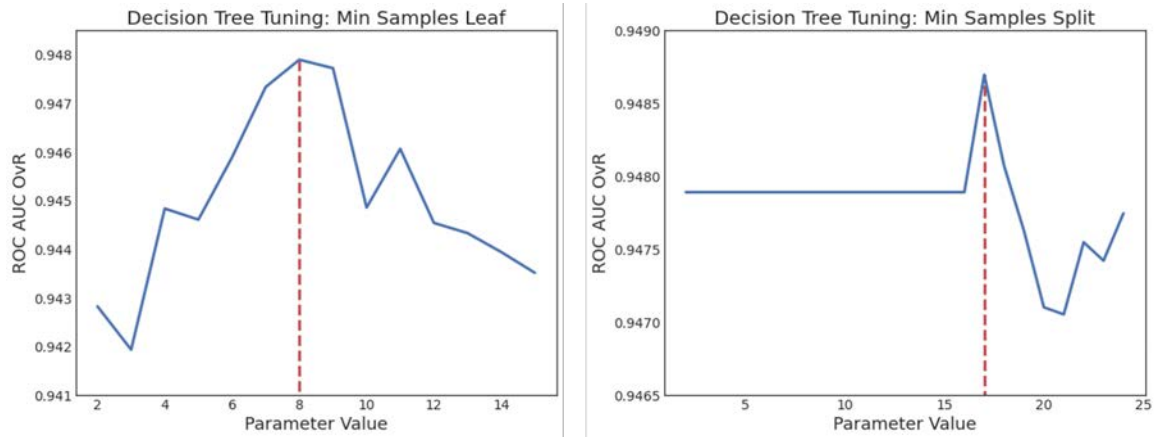


Figure 5-8: Results from stratified k -fold cross validation tuning of the `min_samples_leaf` (Left) and `min_samples_split` (Right) hyperparameters using WDS4. The red dashed lines indicate the selected values.

One optimization trick when training DT models is to consider only a subset of the features when splitting decision tree nodes. This also adds an element of randomness to tree construction because decision trees can differ when constructed on the same training data depending on which features were considered for each split. No clear maximum appears in the AUC plot for `max_features` (Figure 5-9), so a value of 8 was selected using the elbow criterion. `max_features` can also cause problems when performing feature selection if the feature count drops below the `max_features` value, so care must be taken in using this hyperparameter.

The final hyperparameter, `ccp_alpha`, controls the trade-off between model fit and complexity during the tree-pruning backward pass of tree construction. As the alpha value increases from zero, the difference between in-sample (training) and out-of-sample (validation) performance decreases, but so does the overall performance of the classifier on the validation subset. Figure 5-10 shows a clear minimum in train-validate AUC difference, however the validation-set performance drops from 0.95 to under 0.85 when using this value for `ccp_alpha`. The default value of `ccp_alpha` = 0.0 is chosen instead to maximize out-of-sample AUC.

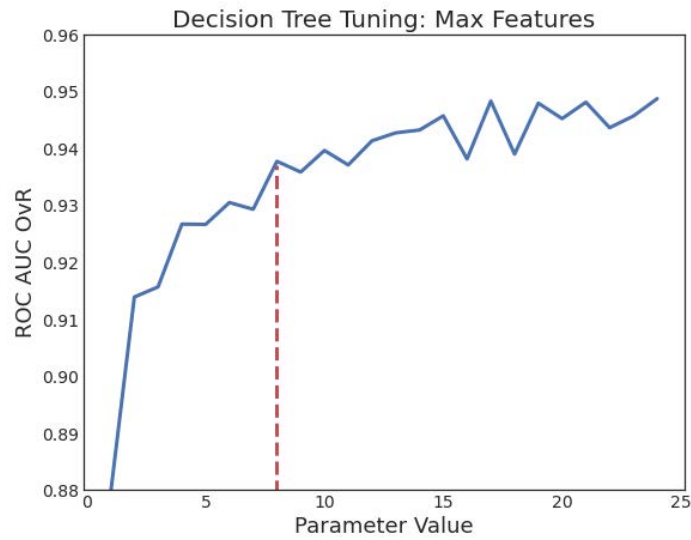


Figure 5-9: Results from stratified k -fold cross validation tuning of the `max_features` hyperparameter using WDS4. The red dashed line indicates the selected value, conservatively selected at the high end of the “elbow” in the plot.

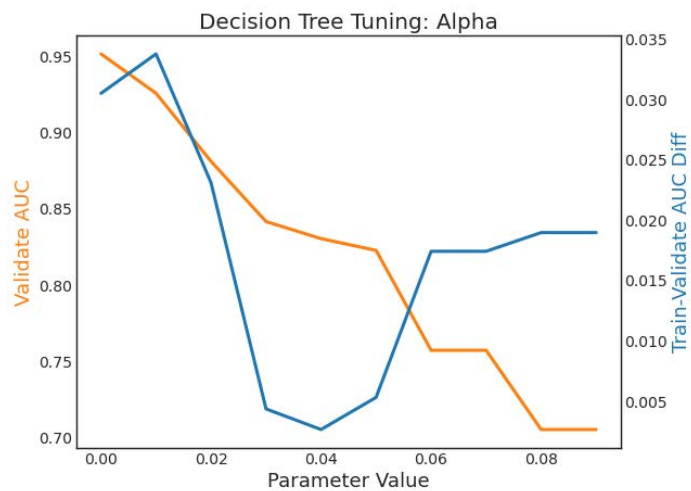


Figure 5-10: Results from stratified k -fold cross validation tuning of the `ccp_alpha` hyperparameter. The orange line plots the validation subset AUC. The blue line plots the difference in AUC between training and validation subsets.

Final hyperparameter values and performance results for WDS, WDS4, and WDS8 are listed in Table 5.3. The data imputation strategy used to create WDS4 and WDS8 results in a significant improvement in classifier performance over the original WDS.

	WDS	WDS4	WDS8
criterion	Gini	Entropy	Entropy
max_depth	5	8	10
min_samples_leaf	7	8	10
min_samples_split	21	17	24
max_features	5	8	6
ccp_alpha	0	0	0
Accuracy_{train}	0.767	0.881	0.880
Accuracy_{test}	0.600	0.818	0.838
AUC_{train}	0.909	0.982	0.985
AUC_{test}	0.814	0.944	0.961

Table 5.3: Tuned hyperparameter selections and resulting decision-tree model Accuracy and AUC for training and testing subsets of WDS, WDS4, and WDS8.

5.2.2 Feature Selection

Figure 5-11 shows the feature importances determined by models constructed using WDS, WDS4, and WDS8. Features are sorted on the sum of importance values across all 3 data sets. Although differences exist, Si Geothermometer Temperature tops the list as the most important predictor in all cases. And the bottom six features are also remarkably consistent across data sets: Water-Table Depth, Water-Table Gradient, Gravity-Anomaly Gradient, Magnetic Anomaly, Magnetic-Anomaly Gradient, and DEM Gradient. Dropping these features reduces the count to 18 features in total.

5.2.3 Optimized Model Results

A final DT model trained on WDS4 was constructed using the tuned hyperparameters and 18-predictor reduced feature set. The confusion matrix (Figure 5-12) demonstrates an improvement in model results over the logistic-regression method. Low-grade gradient locations are correctly predicted for almost double the number of

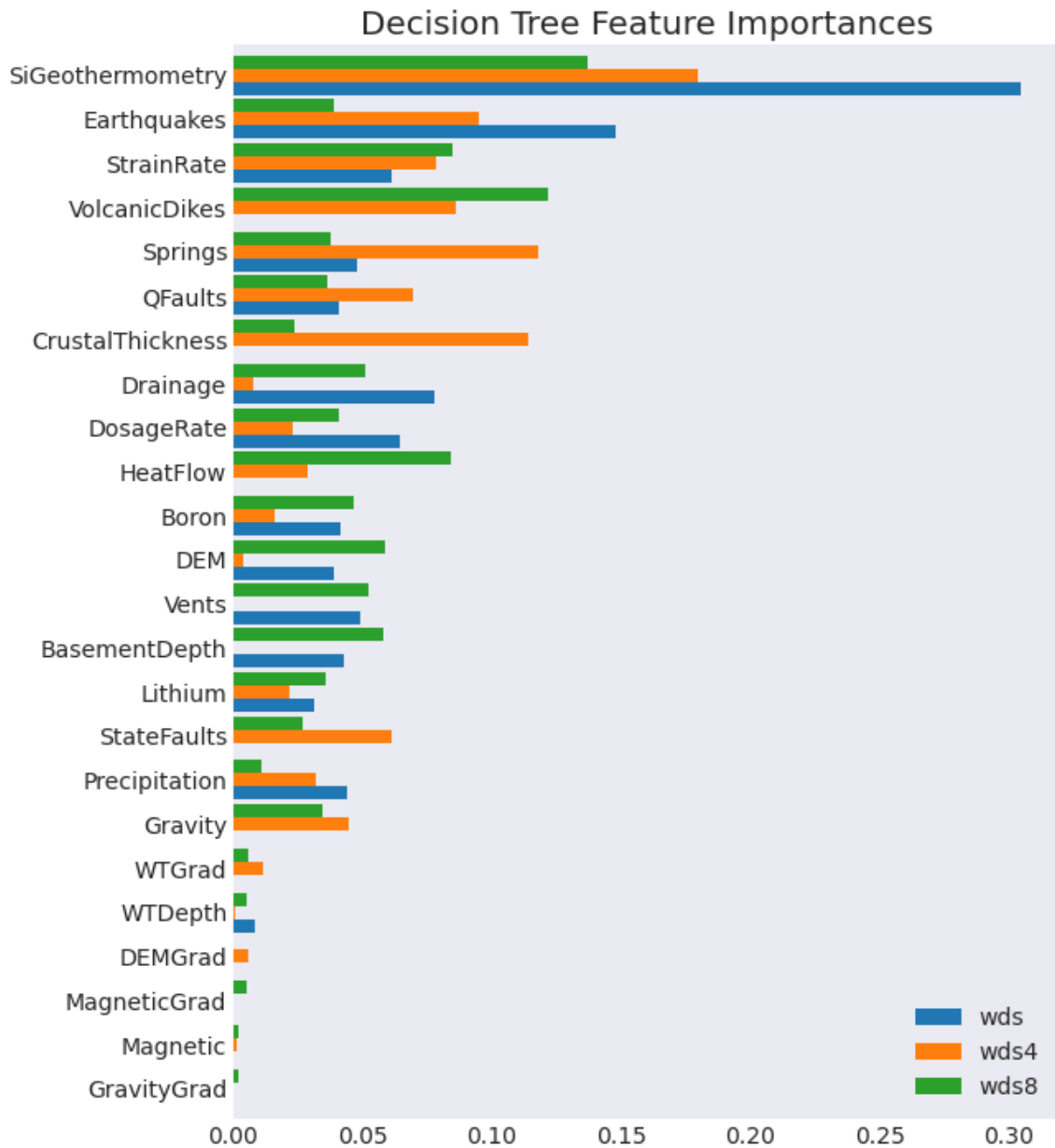


Figure 5-11: Decision tree feature importances for WDS, WDS4, and WDS8, sorted on the sum total importance across the 3 data sets.

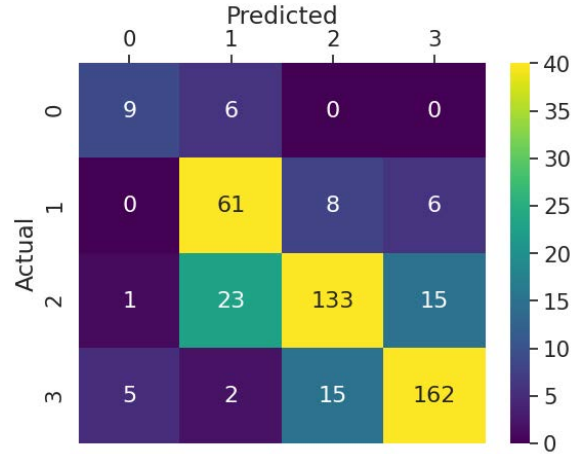


Figure 5-12: Confusion matrix for the tuned DT model trained on WDS4.

sites, and half as many misclassifications are observed between class-2 (mid-grade) and class-3 (high-grade) geothermal gradients than with the LR model.

Figure 5-13 shows the macro average, micro average, and individual class DT ROC curves. All individual class AUC values exceed 0.90. Class 0 (non-thermal) continues to demonstrate the highest predictive performance ($AUC = 0.97$), but class-3 (high-grade) predictive ability boosts the micro-average AUC at higher decision thresholds, i.e., where the class-0 ROC curve steeply drops in TPR for $FPR < 0.1$. Class-2 (medium-grade) classification performance lags behind all other classes for the DT classifier.

Model predictions for the study area are generated by passing the FDS through the DT model. Class assignments are plotted in Figure 5-14. The high-grade geothermal gradient patches to the southeast and central regions of the AOI are not as broad and continuous as in the LR model (Figure 5-6). Predictions for low-grade gradient or non-thermal areas are concentrated to the NW (Colorado Plateau) and central-east (Rio Grade Rift-Great Plains transition). The overall distribution of geothermal gradient classes is similar to the Bielicki et al. (2015) PFA layer (Figure 5-14) but with a greater apportionment of high-grade gradient areas and fewer low-grade or non-thermal locations in the DT model.

Note that if the random seed (fixed in the present study for solution repeatability)

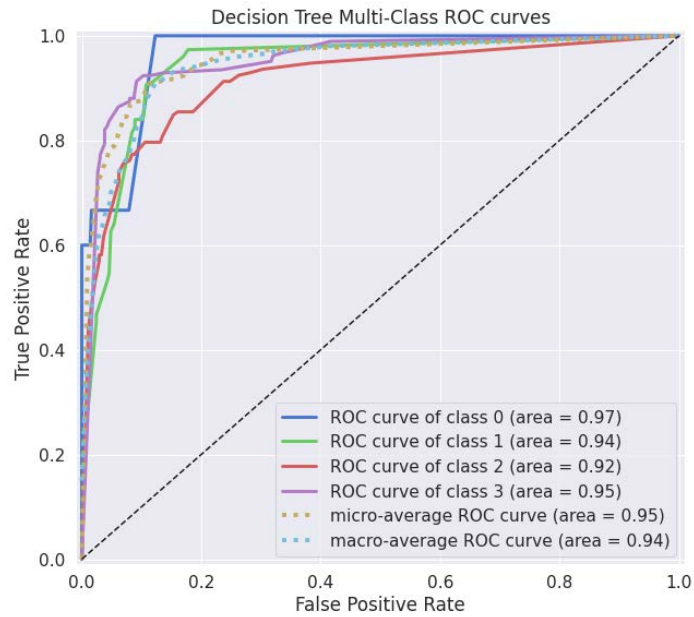


Figure 5-13: ROC curves for the tuned DT model trained on WDS4.

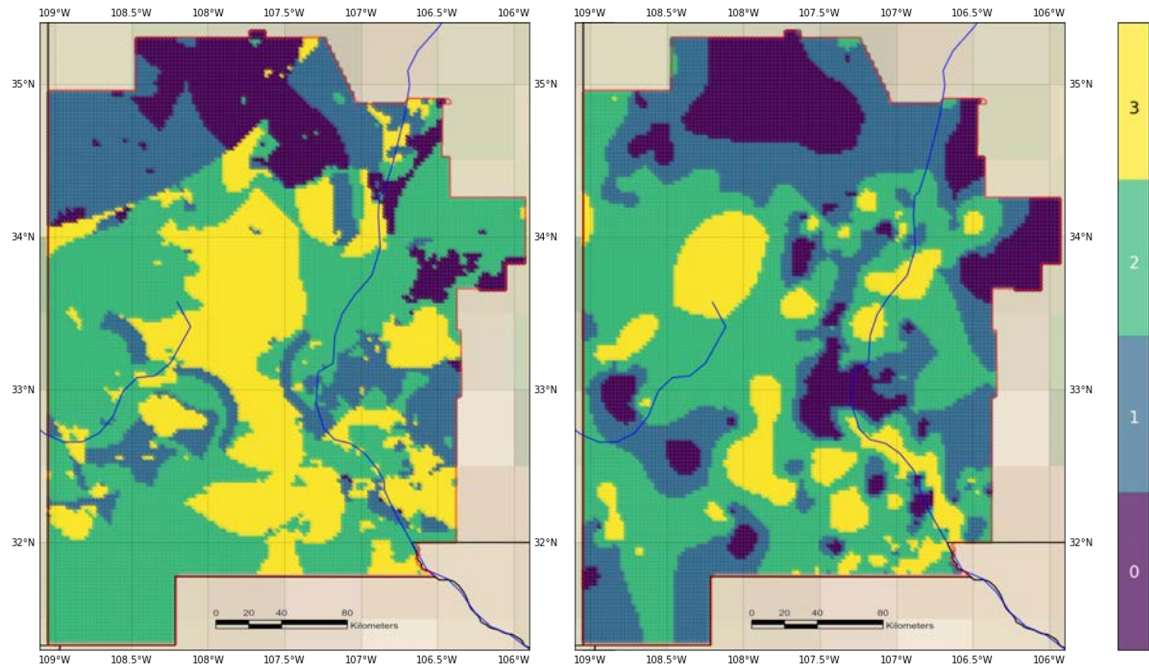


Figure 5-14: Left: Map predictions of geothermal gradient class from the tuned DT model trained on WDS4. Right: geothermal gradient feature layer from Southwestern NM PFA study (Bielicki et al., 2015)

is free to change and new DT models are constructed, some of the characteristics of this DT prediction will also change. In fact, one downside of the DT model is this randomness; decision trees will structurally rearrange each time the algorithm is run, even on the same training data, due to randomness in the node splitting process. Nevertheless, tree performance remains relatively stable overall. And the ability to plot and visually step through the model makes it one of the most accessible and interpretable methods to use (see Figure 5-15).

5.3 Tree Ensembles (XGBoost)

5.3.1 Hyperparameter Tuning

XGBoost comes with many of the same hyperparameters as decision trees, plus additional parameters related to boosting and optimization. With so many hyperparameters, tuning the model becomes a time-consuming and complex process. The method followed here was adapted from an online tutorial covering the topic (Jain, 2016).

The following hyperparameters were tuned for the final model:

- **Maximum depth:** same as the `max_depth` hyperparameter for decision trees, restricts how deep each tree can grow during construction. Initially set to 5.
- **Minimum child weight:** sets a minimum weight requirement for a leaf node during the backward-pass pruning process. Similar to `min_samples_leaf` in decision trees, except this uses the XGBoost-specific weights (θ) noted in equation 3.13. Initially set to 1.0.
- **Gamma:** defines a minimum reduction in the loss function necessary for a split to be preserved during tree pruning. Initially set to 0.2.
- **Learning rate:** a.k.a. shrinkage factor, scales the impact of each tree on the boosted model prediction, i.e., the α_s in equation 3.12. Initially set to 0.01.
- **Lambda:** L2 regularization term on the leaf scores, i.e., the λ in equations 3.13 and 3.14. Initially set to 1.0.

- **Subsample:** defines the fraction of observations in the full training set that are randomly selected as a limited training set for each tree. Initially set to 1.0.
- **Column sample by tree:** sets the fraction of predictors in the full training set that are randomly selected as a limited feature set for constructing each tree. Initially set to 1.0.
- **Number of estimators:** controls the number of sequential trees in the model, i.e., B in equation 3.12. Initially set to 200.
- **Scale positive weight:** scales the gradient for the positive class to influence model corrections during training, helping manage imbalanced classification. Initially set to 1.0.

The preferred tuning method for XGBoost is the cross-validation strategy described in Section 3.3.2. Since this method uses the input data to both train and validate, the pre-generated training and validation subsets (Table 3.5) were re-combined before using CV. An initial trial-and-error testing of different hyperparameter combinations found the best starting values for `learning_rate` and `n_estimators` were 0.01 and 200, respectively. Higher learning rates like those suggested by Jain (2016) led to a nearly perfect-fitting model before hyperparameters could be tuned, and more estimators created undesirably long training times. In the figures and discussion that follow, results are described for WDS4. Full results for all three data sets are provided in Table 5.4.

XGBoost hyperparameters were tuned in succession using stratified 10-fold CV and a grid search across potential hyperparameter values. Figure 5-16A shows the results for `max_depth`. As noted when tuning the DT model (Section 5.2.1), `max_depth` does not exhibit a clear maximum in the AUC vs. hyperparameter value plot. A preferred value of 6 was selected to ensure sufficient tree performance while balancing tree complexity.

Next, tuning was performed on the `min_child_weight` hyperparameter. Results of the stratified 10-fold CV method revealed a clear maximum AUC when `min_child_weight` is 2 (Figure 5-16B).

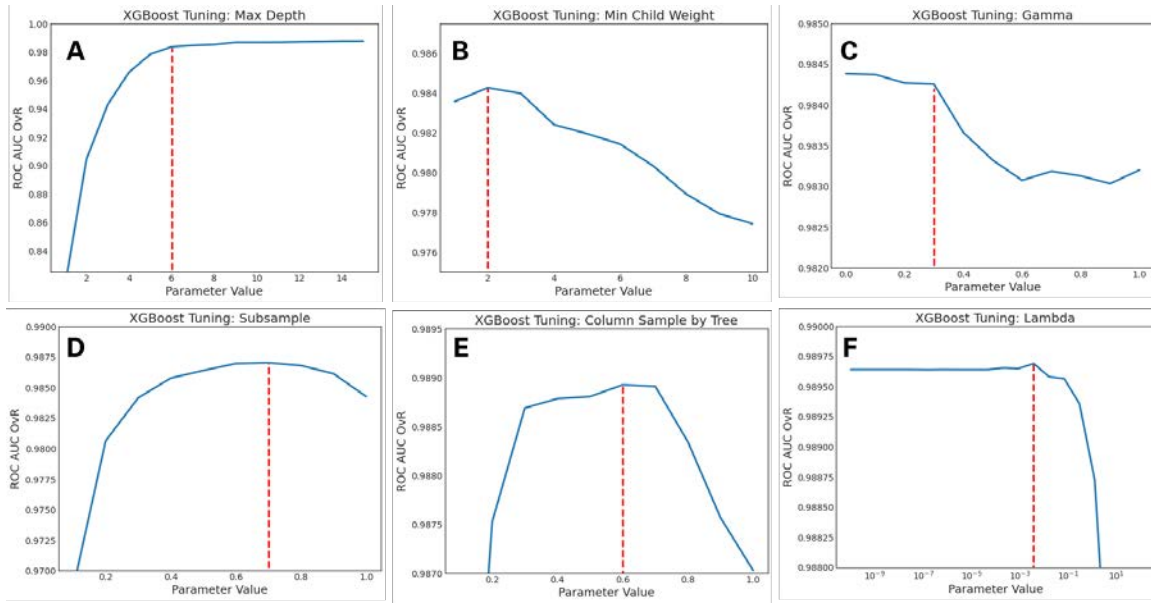


Figure 5-16: Hyperparameter tuning results for XGBoost modeling. A. `max_depth`, B. `min_child_weight`, C. `gamma`, D. `subsample`, E. `colsample_bytree`, F. `lambda`. Red dashed lines indicate values selected for the final model.

XGBoost uses a default value of 0 for the `gamma` hyperparameter, which controls when tree partitioning should stop based on loss reduction. Cross-validation shows nearly level AUC values out to `gamma` = 0.3, after which the AUC score quickly drops. This threshold value was selected for the model (Figure 5-16C).

Trees in the XGBoost model can be trained on random subsets of training data observations and individual feature columns. Tuning of `subsample` identified a broad maximum in AUC when using 60–80% of the training observations to build decision trees (Figure 5-16D). The selected `subsample` value is 70%. For `colsample_bytree`, the best AUC results occur when 60% of the features are included in the training (Figure 5-16E). A value below 100% for `colsample_bytree` suggests that removing features based on importance estimates could be beneficial, as previously discussed for both logistic regression (Section 5.1.2) and decision trees (Section 5.2.2). A more robust method for feature attribution and selection using Shapley values is considered in Section 5.3.2.

Tuning results show model AUC remains flat, then quickly drops for increasing values of the `lambda` hyperparameter (Figure 5-16F). This behavior shows how higher

	WDS	WDS4	WDS8
max_depth	6	6	6
min_child_weight	2	2	5
gamma	0.5	0.3	0.5
subsample	0.8	0.7	0.7
colsample_bytree	0.4	0.6	0.6
reg_lambda	0.30	0.0038	1.3
scale_pos_weight	0.0	0.0	0.0
learning_rate	0.005	0.005	0.005
n_estimators	1000	1000	1500
Accuracy_{train}	0.995	0.991	0.988
Accuracy_{test}	0.767	0.944	0.955
AUC_{train}	1.000	1.000	1.000
AUC_{test}	0.938	0.995	0.995

Table 5.4: Tuned hyperparameter selections and resulting XGBoost model Accuracy and AUC for training and test subsets of WDS, WDS4, and WDS8.

levels of regularization can cause data underfitting. The largest `lambda` value just short of this drop-off was selected for the final model.

`scale_pos_weight` was similarly tuned, but AUC results remained unchanged for all hyperparameter values tested. The multi-class XGBoost classifier appears to be insensitive to this hyperparameter.

As a final step in model tuning, `learning_rate` was decreased to 0.005 and `n_estimators` increased to 1000. Using the slower learning rate reduces the chance of overfitting the training data, while the increased number of sequential trees in the model ensures a good fit can be learned. Final hyperparameter values and performance results for WDS, WDS4, and WDS8 are listed in Table 5.4. Once again, WDS4 and WDS8 out-perform the original WDS based on AUC values for the test set.

5.3.2 Feature Selection

Section 3.3.5 noted several desirable attributes of feature importances derived from Shapley analysis. Figure 5-17 plots the results of SHAP value calculations for the WDS4 XGBoost classifier. Recall that SHAP values have local significance on a per-observation basis (i.e., local accuracy). This plot illustrates the mean absolute

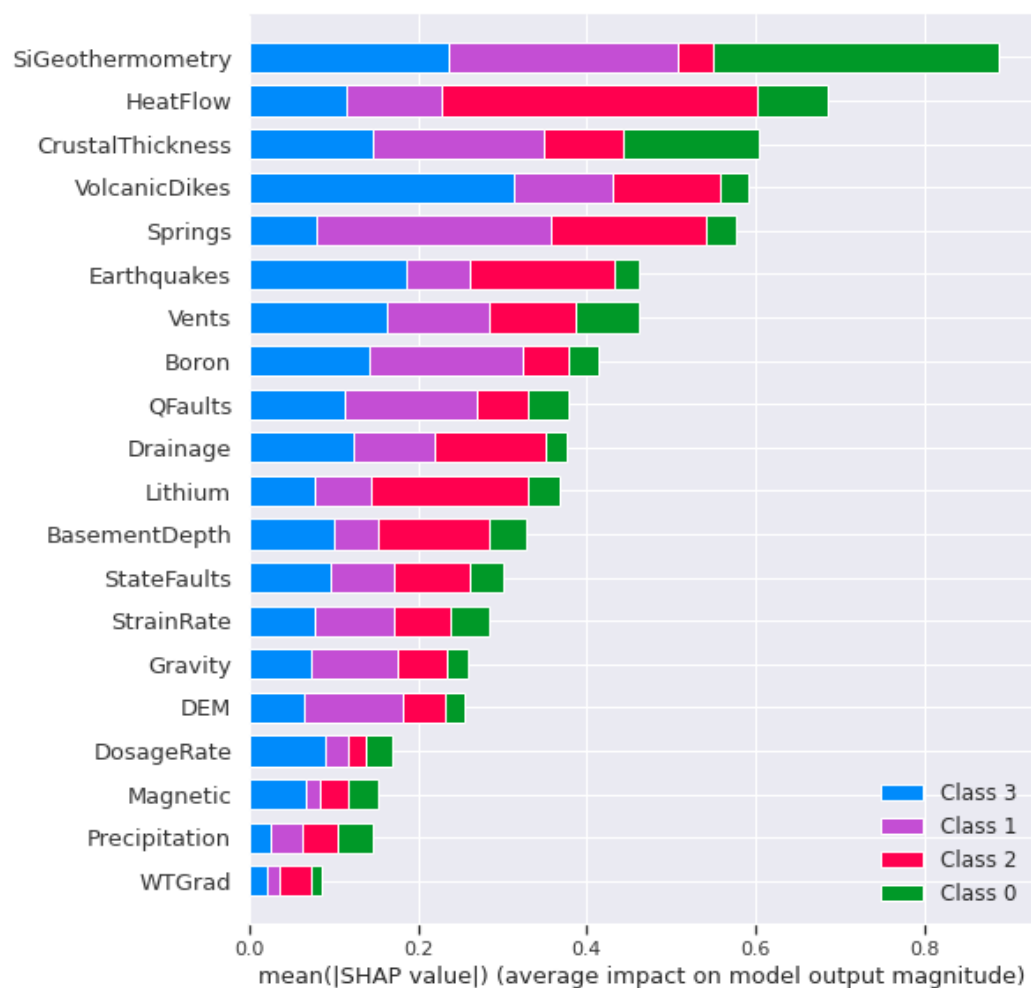


Figure 5-17: SHAP variable importance plot for the XGBoost classifier, derived using the testing data subset of WBS4. Colors illustrate feature importances for specific classes of geothermal gradient. The sum of colored bar lengths indicates overall feature importance for the model. Nearby values in sorted stacked bar length suggest the model has five key predictive features, eleven features with moderate influence on the model, and four features with low predictive value.

SHAP value of each feature across all instances in the test data subset, giving an average global impact. The colored bars illustrate relative importance of that feature in the prediction for the respective geothermal gradient class. Global importance is indicated by the sum of colored bar lengths (stacked length) for each feature. The order of the features is sorted on stacked length to highlight both the most and least important features. The top five most important features for the WDS4 model include Si Geothermometer Temperature, Heat Flow, Crustal Thickness, Volcanic-Dike Density, and Spring Density. Features not shown in the plot were deemed to have zero predictive value (i.e., DEM Gradient, Gravity Gradient, Magnetic-Anomaly Gradient, and Water-Table Depth). The lowermost four features on the plot — Gamma Ray Absorbed-Dose Rate, Magnetic Anomaly, Average Precipitation, and Water-Table Gradient — are also selected for elimination, reducing the final feature set for the XGBoost model to sixteen predictors.

5.3.3 Optimized Model Results

A final XGBoost model parameterized with the tuned hyperparameter values (Table 5.4) was trained on the reduced sixteen-feature version of WDS4. The resulting confusion matrix (Figure 5-18) demonstrates why XGBoost receives best-in-class praise as a predictive model. Test set accuracy for WDS4 is 94%. The largest number of misclassifications (6 + 6) occur between high-grade (class 3) and medium-grade (class 2) geothermal gradient. Ten additional locations are mistakenly labeled low-grade (class 1) for medium-grade or vice-versa. Only two test set locations were off by more than one consecutive class.

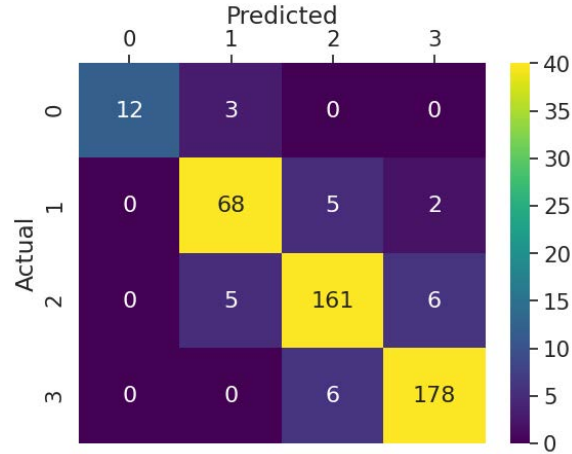


Figure 5-18: Confusion matrix for the tuned XGBoost model trained on WDS4.

Figure 5-19 shows the macro average, micro average, and individual class XGBoost ROC curves. All individual class AUC values are at or above 0.99. This is close to an ideal ROC plot for a classifier.

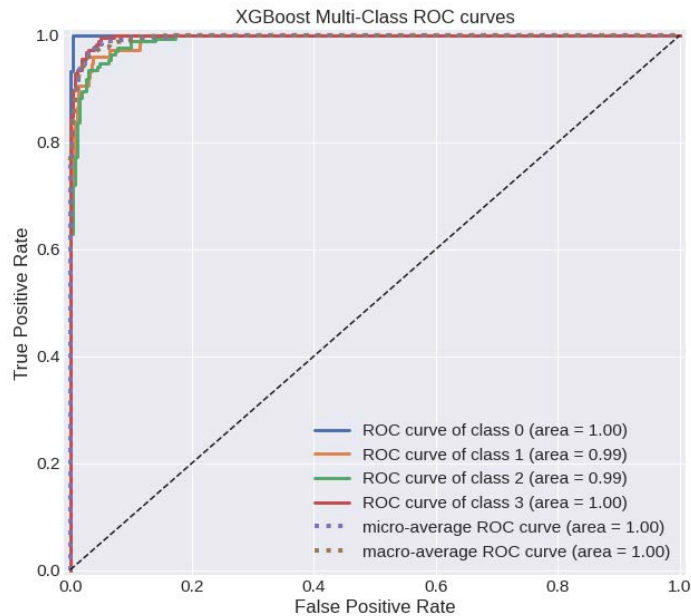


Figure 5-19: ROC curves for tuned XGBoost model trained on WDS4.

Geothermal gradient predictions for the study area are generated by passing the FDS through the XGBoost model, as shown in Figure 5-20. High-grade gradient areas to the southeast and central regions of the AOI align with the class 3 locations

in the [Bielicki et al. \(2015\)](#) Southwestern NM PFA map. XGBoost predicts more spatially continuous and connected high-grade regions, with a limited number of isolated patches to the southwest and on the northern section of the Rio Grande. Both maps have a similar non-thermal class 0 region to the north (Colorado Plateau), but they differ most significantly midway along the Rio Grande, to the southwest (Basin and Range), and in the eastern panhandle (Great Plains) where the PFA map predicts low-grade gradient or non-thermal classifications.

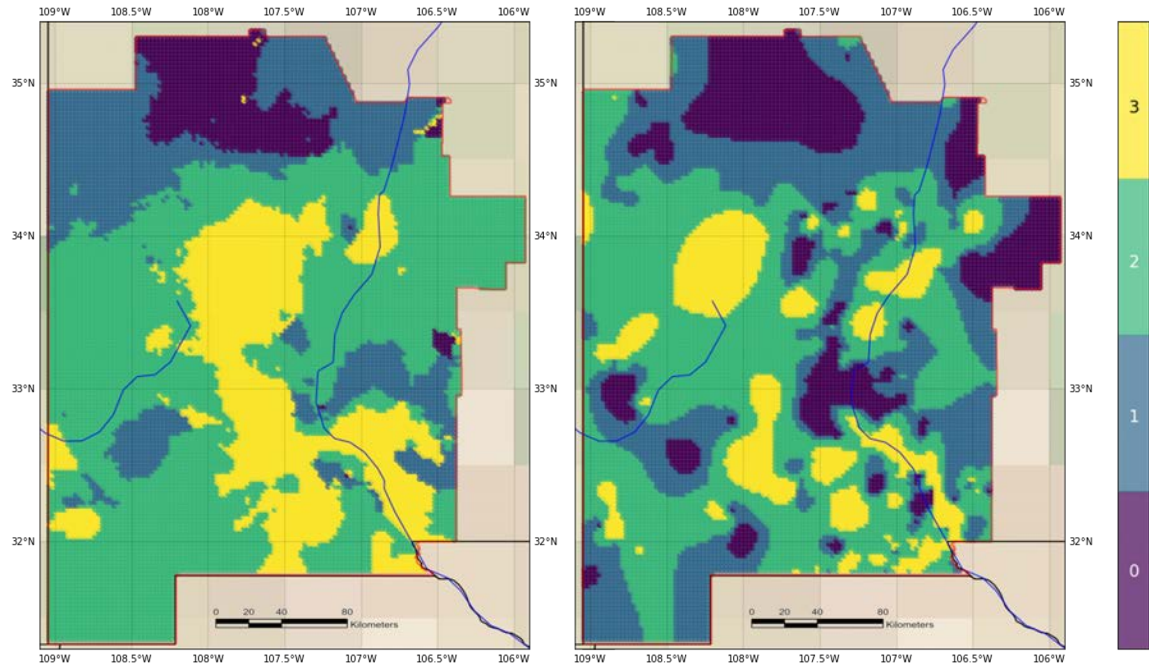


Figure 5-20: Left: Map predictions of geothermal gradient from the tuned XGBoost model trained on WDS4. Right: geothermal gradient feature layer from Southwestern NM PFA study ([Bielicki et al., 2015](#)).

5.4 Neural Networks

5.4.1 Network Architecture

This thesis uses a neural network constructed using TensorFlow ([Abadi et al., 2016](#)) to test geothermal gradient classification performance for the Southwestern NM study area. The network design is a 4-layer structure starting with an input layer of size

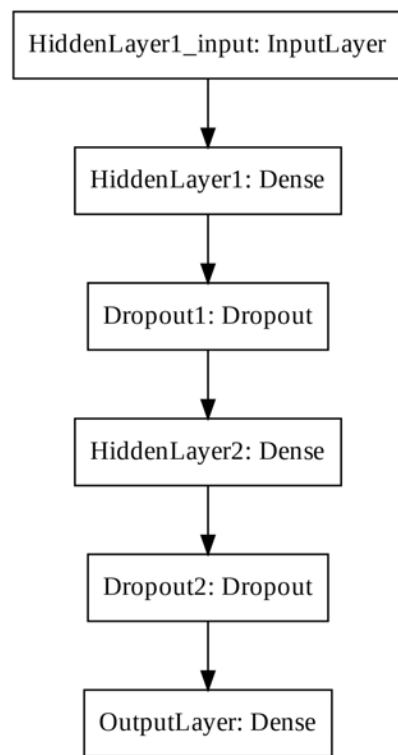


Figure 5-21: Structural flow chart for the TensorFlow-based neural networks.

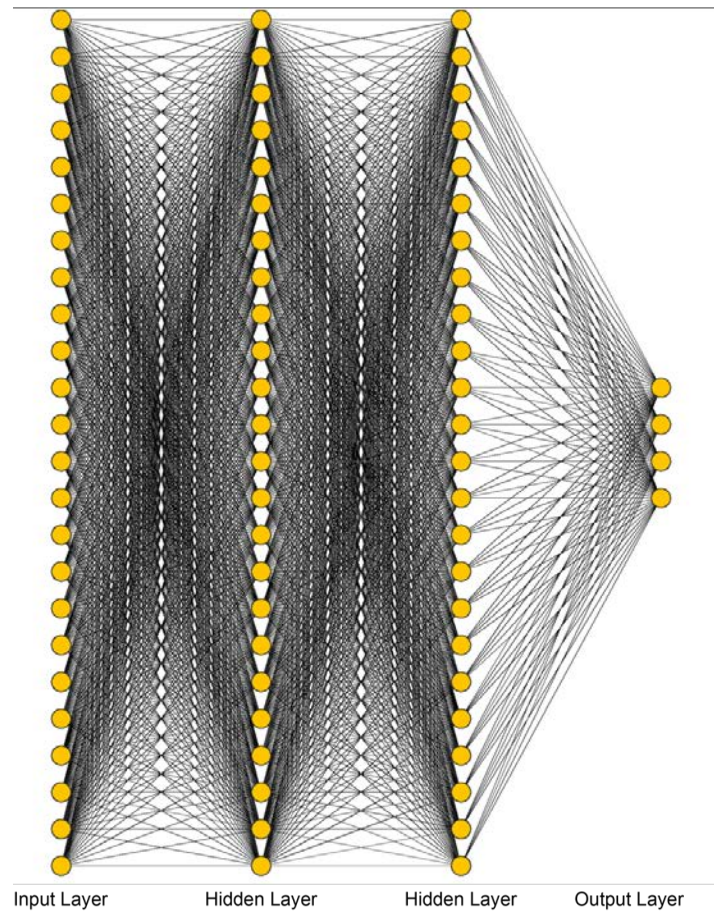


Figure 5-22: Schematic diagram of the artificial neural networks. Not shown are the dropout operations after each hidden layer.

(1×24) , followed by two hidden layers sized (1×24) , and ending in a four-class output layer (1×4) (Figures 5-21, 5-22). Layer sizes assume all original features except Average Air Temperature (see Section 3.2.2) are included in the training. In addition, the network applies dropout after each of the hidden layers. Dropout helps prevent overfitting by randomly assigning inputs to zero, temporarily severing a defined percentage of node connections (a hyperparameter) during each step of the training process.

5.4.2 Hyperparameter Tuning

Tuning of the network focuses on five key hyperparameters:

- **Learning rate:** controls the magnitude of adjustments to the network weights during each step of the training process. The Adam optimizer adjusts this rate during training, so the value here defines a starting learning rate. Initially set to 0.001.
- **Lambda (λ):** the L2 weight regularization term, i.e., the λ in equation 3.16. Initially set to 1×10^{-4} .
- **Batch size:** the number of training samples included in each mini-batch when calculating training gradients for updating the network weights. Initially set to one tenth the number of samples.
- **Dropout rate:** the fraction of network connections zeroed-out during each step of the training process. Initially set to 0.2.
- **Number of epochs:** the number of training repetitions when fitting the model to the data. Initially set to 100.

The stratified k -fold CV strategy described in Section 3.3.2 was used for tuning ANN hyperparameters. Training continued for 100 epochs, and since TensorFlow models easily track metric values for both the training and validation sets at each epoch, the two sets were kept separate during the 10-fold CV process. In the figures

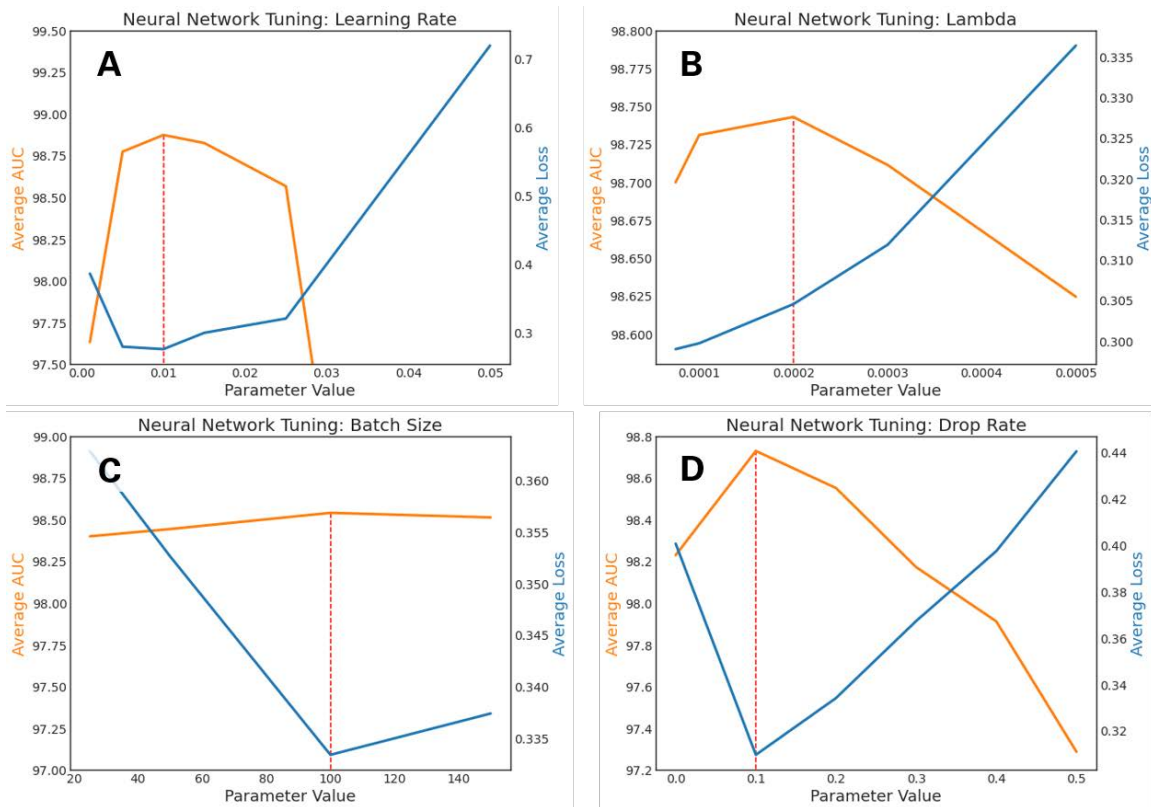


Figure 5-23: ANN hyperparameter tuning results for A. learning rate, B. lambda (λ), C. batch size, and D. dropout rate, using WDS4. The orange lines track AUC after 100 training epochs, averaged over 10-folds of stratified cross-validation for each hyperparameter value. The blue lines show average loss after 100 epochs. Red dashed lines mark the selected values for the final model.

and discussion that follow, results are described for WDS4. Full results for all three data sets are provided in Table 5.5.

	WDS	WDS4	WDS8
learning rate	0.005	0.010	0.010
lambda (λ)	0.0001	0.0002	0.0001
batch size	50	100	150
dropout rate	0.3	0.1	0.1
epochs	25	100	200
Accuracy_{train}	0.777	0.959	0.965
Accuracy_{test}	0.742	0.949	0.957
AUC_{train}	0.952	0.998	0.999
AUC_{test}	0.863	0.992	0.992

Table 5.5: ANN hyperparameter tuning results for each data set. Accuracy and AUC metrics are split into train (in-sample) and test (out-of-sample) values.

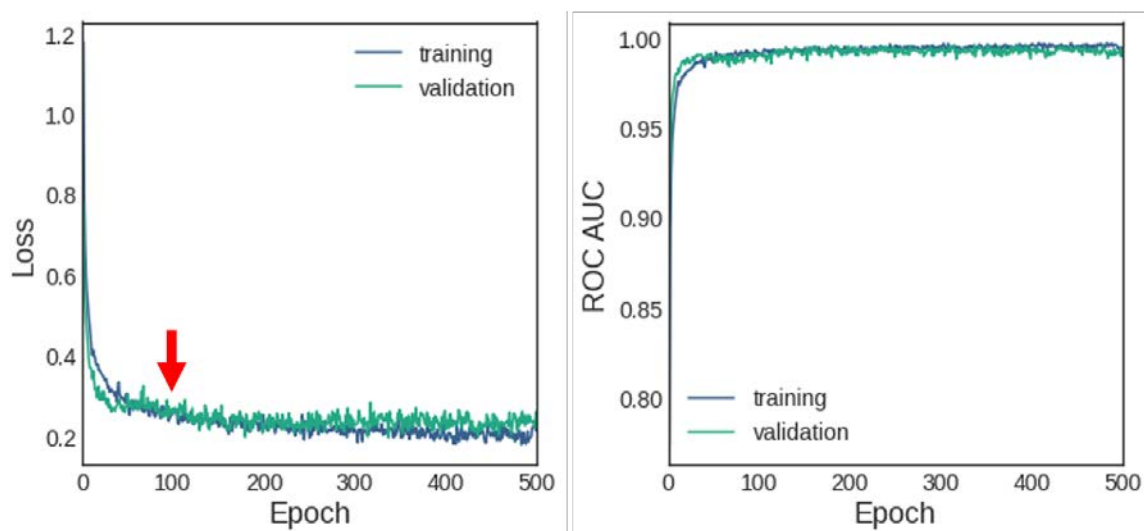


Figure 5-24: ANN training results as a function of epoch count for WDS4. Calculated loss (Left) and AUC (Right) for the training (blue) and validation (green) subsets converge and start to separate near the 100-epoch mark (red arrow). Mini-batch related noise obscures the exact cross-over location.

Tuning was conducted in the same order as listed in Figure 5-23. Distinct maxima in CV-averaged AUC are observed for each hyperparameter. The selected values for the final WDS4 model include: `learning_rate` of 0.01, `lambda` value of 2×10^{-4} , `batch_size` of 100, and `dropout_rate` of 0.1. Minima in the average loss curves

align with the selected hyperparameter values for all except `lambda`, the regularization parameter on model weights. This matches the expected behavior described in Equation 3.16, where `lambda` controls the balance between loss and sum of squared weights in the cost function. As Figure 5-23B shows, the loss term decreases as `lambda` approaches zero.

Figure 5-24 illustrates the WDS4 training progress as a time series of loss and AUC. The variance in both plots reflects noise introduced by mini-batch training, but convergence takes place quickly. The growing separation between the training and validation loss at greater epoch counts indicates the model is overfitting the training data. The cross-over of the training and validation lines, which corresponds with the tuned value for the `epoch` hyperparameter, occurs somewhere close to 100 epochs for WDS4.

5.4.3 Optimized Model Results

The ANN model was re-trained for each of the three data sets using the hyperparameter values listed in Table 5.5. Data sparsity is problematic for the WDS model; even after careful tuning, the out-of-sample accuracy does not exceed 75%, and test set AUC is just $\approx 86\%$. Models trained and tested on respective subsets of WDS4 and WDS8 demonstrate $\approx 95\%$ accuracy and AUC values $> 99\%$ (Table 5.5). Although there is the concern of added spatial correlation through the kriging approach to geothermal gradient imputation described in Section 3.2.2, the imputed data sets provide much-needed constraints for training the many ANN model parameters.

Figure 5-25 depicts the confusion matrix constructed from WDS4 model results. Similar to XGBoost model performance, nearly all of the misclassifications are off by only a single sequential class assignment, the most prevalent being between medium-grade (class 2) and high-grade (class 3) geothermal gradient. Overall, these are very strong results for a predictive classifier.

Figure 5-26 shows the macro average, micro average, and individual class ANN ROC curves. All individual class AUC values are at or above 0.99, as was the case for the XGBoost model. This is close to an ideal ROC plot for a classifier.

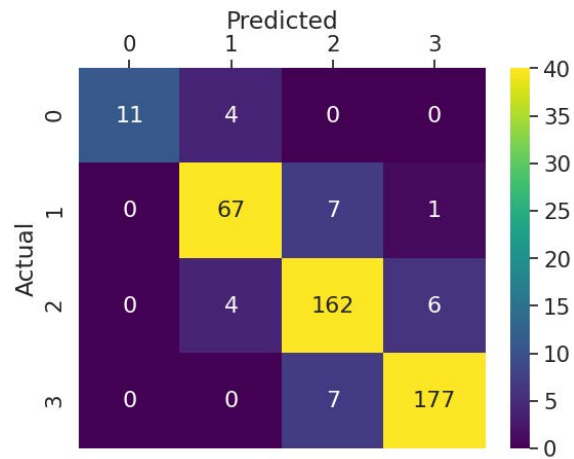


Figure 5-25: Confusion matrix for the tuned ANN model trained on WDS4.

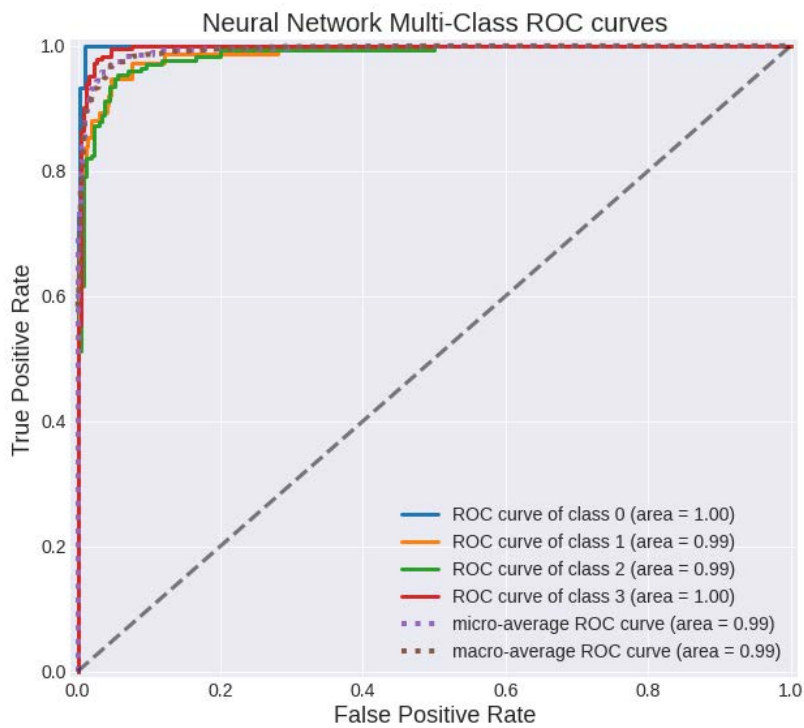


Figure 5-26: ROC curves for tuned ANN model trained on WDS4.

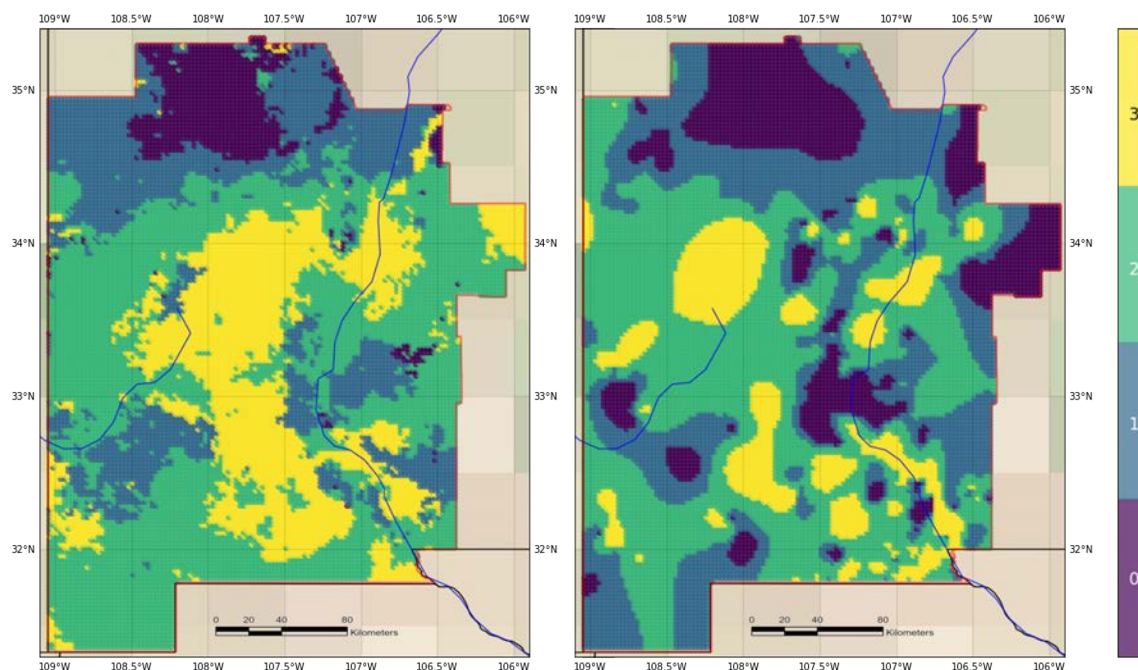


Figure 5-27: Left: Map predictions of geothermal gradient from the tuned ANN model trained on WDS4. Right: geothermal gradient feature layer from Southwestern NM PFA study (Bielicki et al., 2015).

The FDS was passed through the final ANN model to generate predictions for the full study area, as shown in Figure 5-27. Class-3 high-grade geothermal-gradient regions to the southeast in the Bielicki et al. (2015) PFA layer are well captured by the ANN model. The class-3 swath though the center of the AOI describes a broader, more continuous region of gradient potential than suggested by the PFA map. Additional high-gradient patches unique to the ANN map appear along the study-area edges — in the southwest corner (Basin and Range), eastern panhandle (Great Plains), and to the northeast, following the Rio Grande. The low-gradient zone to the north (Colorado Plateau) matches between the two maps, but most other class-0 areas in the Bielicki et al. map are not observed in the ANN map. Interestingly, there appears to be greater overall similarity between the ANN map in Figure 5-27 and the XGBoost map in Figure 5-20 than with the PFA map.

5.5 Uncertainty Analysis

5.5.1 Structural Uncertainty

Figure 5-28 shows all four machine learning model predictions for the SW NM study area. The models differ in how individual areas are classified, as well as overall character of the predictions, suggesting each model has a signature predictive style.

Figure 5-29A illustrates an ensemble average map, where each location is assigned the class with the maximum average probability across the four model predictions. Aspects of all four models in Figure 5-28 can be identified in this result, but the overall effect is a simpler model with a more spatially-conservative high-grade class-3 predictions. Also missing are many of the probably spurious high-gradient patches along the AOI boundaries and within the eastern panhandle or southern boot-heel. The Bielicki et al. gradient-feature map (Figure 5-29B) shows much more structure in geothermal-gradient predictions by comparison, with a pock-marked appearance typical of interpolation bulls-eye patterns.

Shannon entropy is calculated on this same averaged model using the underlying class probabilities. The result is shown in Figure 5-30. Normalized entropy values vary from 0 to 1, with high entropy defining locations where there is no clear differentiation between gradient-class label probabilities. Regions colored red thus identify locations that are difficult for any model to classify, due to inconclusive feature inputs, or alternatively, locations where the predicted class probabilities from the different models varied enough that, upon averaging, they converged to similar values and could not be disentangled.

Figure 5-31 illustrates an alternative way of presenting these results. Regions of high entropy (> 0.7) are masked in dark gray because of the uncertainty in their classifications. All other points have a color transparency that scales with entropy. Low uncertainty/entropy areas are uncommon, focused primarily in the southwest and in small patches to the central west and central east areas. High uncertainties trace the boundaries between areas of consistent gradient classifications to the north and southeast. For example, the non-thermal region (class 0) in the Colorado Plateau

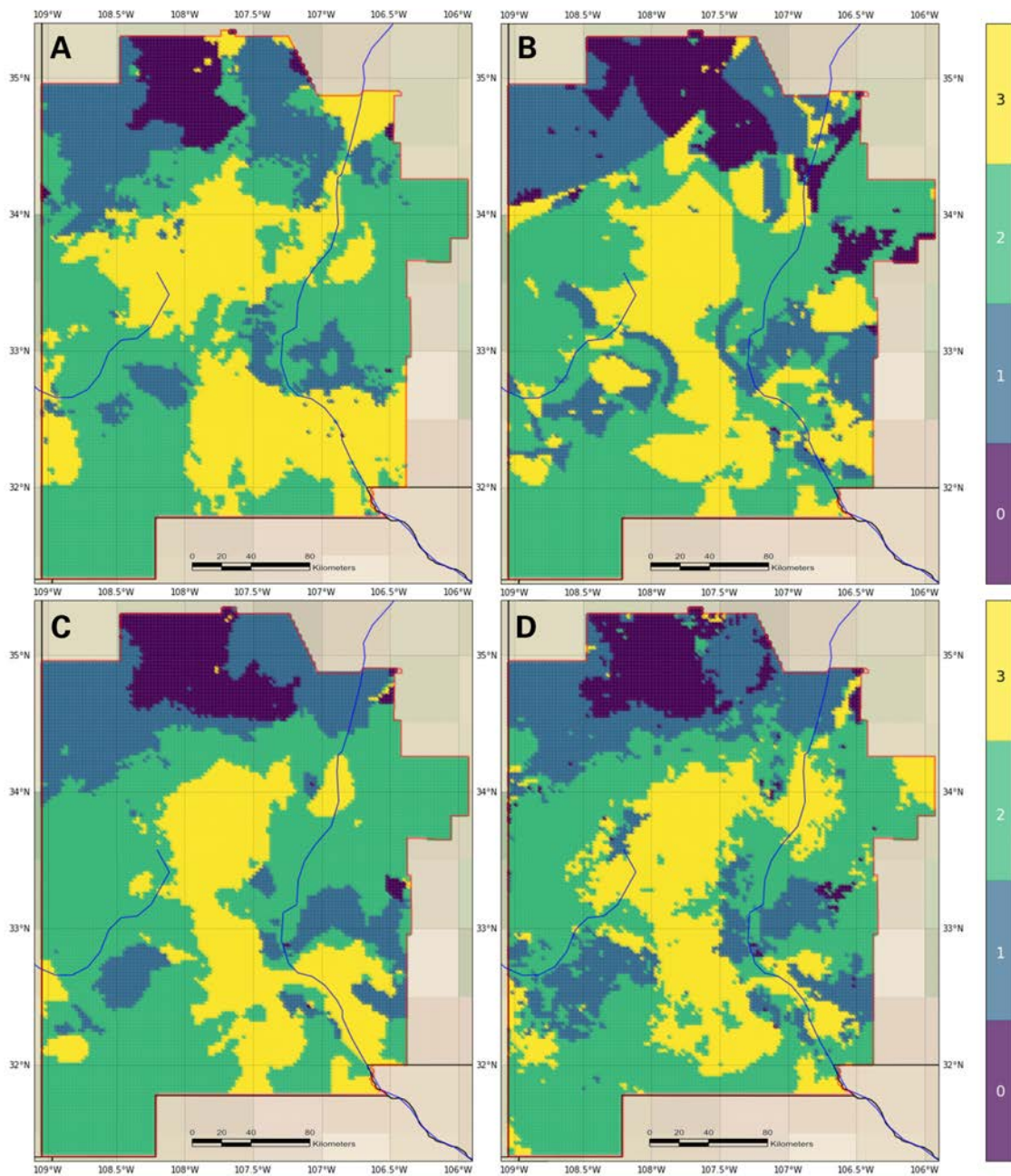


Figure 5-28: Machine learning results for WDS4 using A. logistic regression, B. a decision tree, C. XGBoost, and D. an artificial neural network. Results match those already shown in Figures 5-6, 5-14, 5-20, and 5-27.

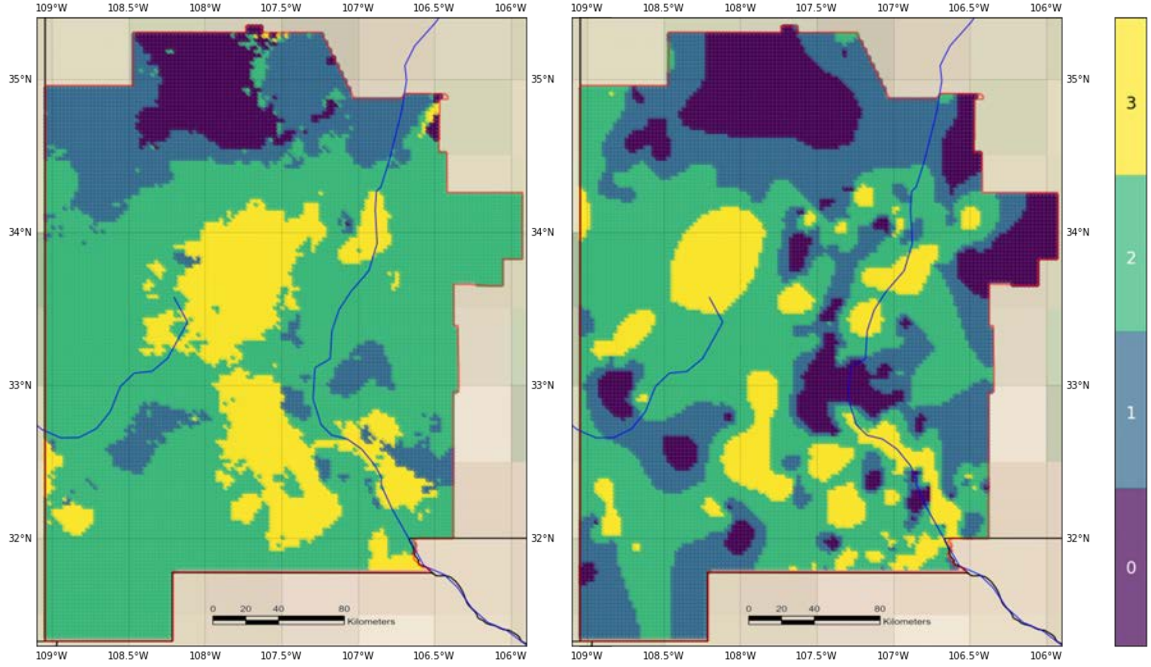


Figure 5-29: Left: Geothermal gradient classification map from combining the four predictive models shown in Figure 5-28. Results are for WDS4. Right: geothermal-gradient feature layer from Southwestern NM PFA study (Bielicki et al., 2015).

to the north is ringed by high uncertainties. This highlights model inconsistencies on exactly where the class 0-class 1 boundary should appear based on trends in the input feature data.

5.5.2 Parameter Uncertainty

The TensorFlow Probability package (Dillon et al., 2017) was used to transform the ANN from Section 5.4 into a Bayesian Neural Network (BNN) for uncertainty estimation. Given the sparsity of training data available and the multiplier effect that probabilistic layers have on the number of trainable parameters in a BNN, only the second hidden layer was converted to a TensorFlow *DenseVariational* layer (Figures 5-32 and 5-33). Standard normal ($N(\mu = 0, \sigma = 1)$) distributions were used as priors for the layer nodes. The same tuned hyperparameter values applied to the ANN (Table 5.6) were also applied to the BNN with one exception: the number of epochs was increased by a factor of 3-4. Figure 5-34 shows the training loss and AUC curves for WDS4, which justify this larger number of epochs for training. Table 5.6 notes

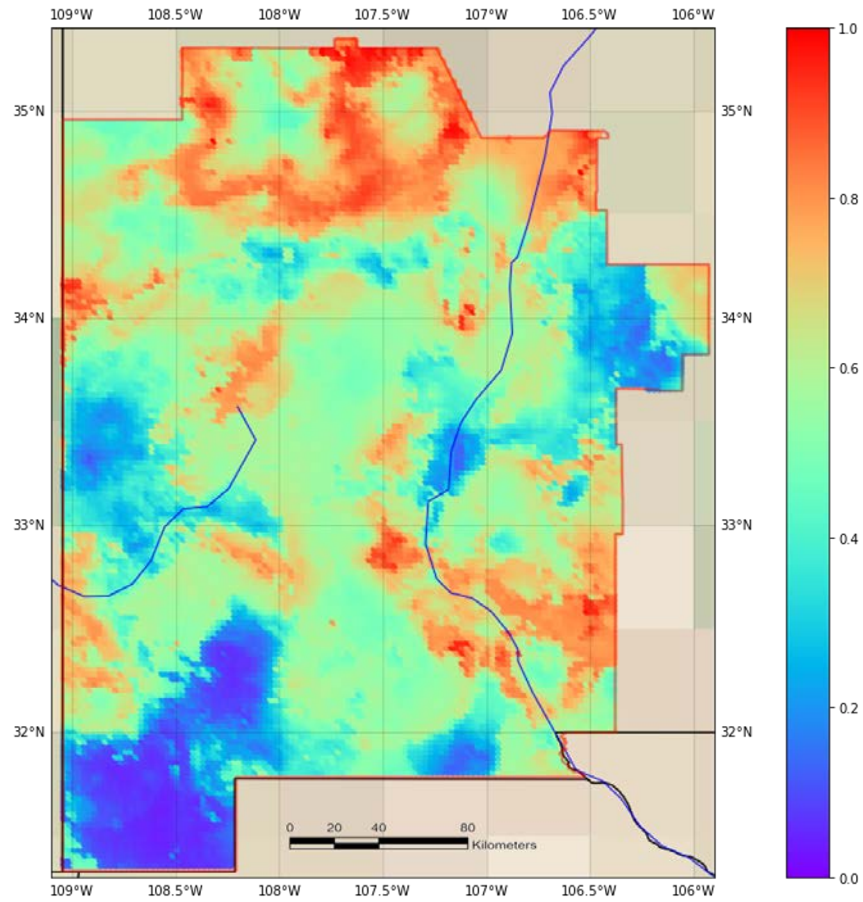


Figure 5-30: Structural uncertainty from the choice of models, measured using Shannon entropy. Values are normalized to range from 0 for low entropy, low uncertainty (blue) to 1 for high entropy, high uncertainty (red).

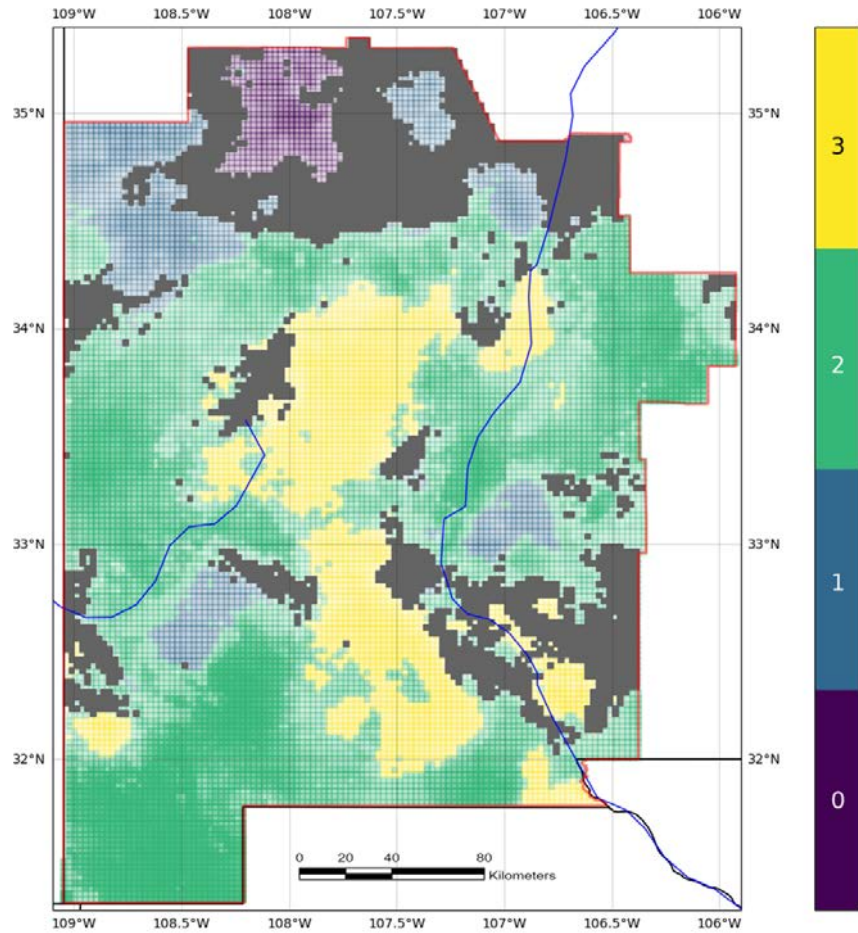


Figure 5-31: Combined-model prediction map with uncertainty masking. Normalized entropy values > 0.7 are grayed out and values ≤ 0.7 determine transparency of the colored scatter plot. Transparency increases from none at entropy values close to 0 to full for entropy values close to 1. The background topographic raster has been removed to better visualize the transparency effect.

the accuracy and AUC scores for the BNNs calculated from a single predictive run on the respective test data subsets.

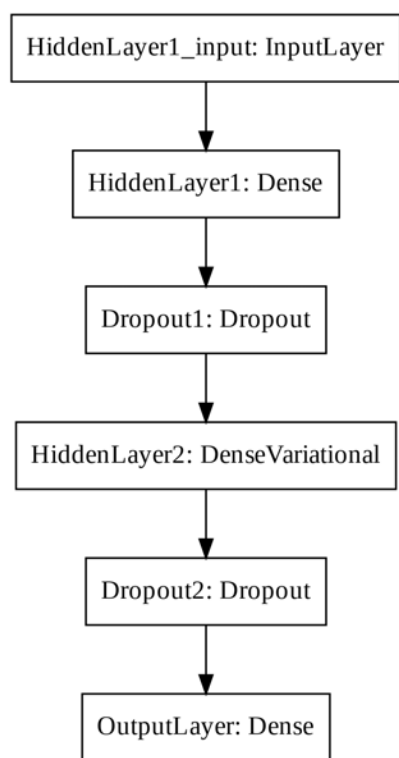


Figure 5-32: Structural flow chart for the Bayesian neural networks.

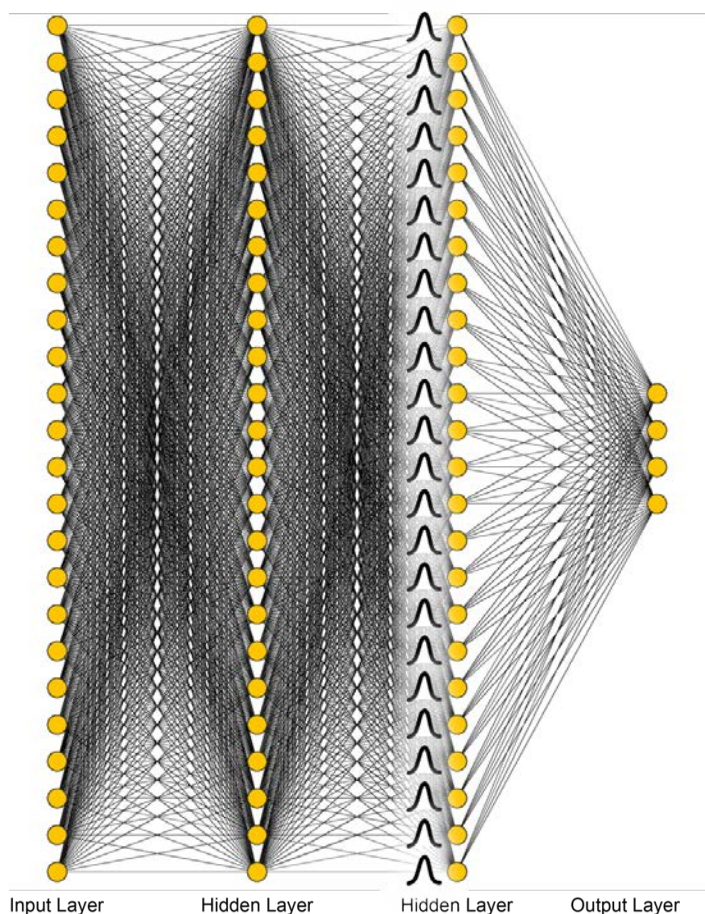


Figure 5-33: Schematic diagram of the Bayesian neural networks. Not shown are the dropout operations after each hidden layer.

Running the BNN multiple times will produce a collection of different model solutions. Figure 5-35 shows the range of class label probabilities after predicting the classification of three well records 100 times using a BNN trained on the WDS4 training subset. Here, the relationship between entropy values and the overlap in class-label probabilities is apparent. A low-entropy scenario (Figure 5-35A) has strong stand-out between the maximum-probability class label and others, while high-entropy situations (Figure 5-35C) have less certainty on the class-label assignment.

After re-training the BNN on the full WDS4 data set, predictions for the entire

	WDS	WDS4	WDS8
learning rate	0.005	0.010	0.010
lambda (λ)	0.0001	0.0002	0.0001
batch size	50	100	150
dropout rate	0.3	0.1	0.1
epochs	100	300	600
Accuracy_{train}	0.516	0.867	0.902
Accuracy_{test}	0.483	0.869	0.894
AUC_{train}	0.798	0.976	0.990
AUC_{test}	0.762	0.961	0.980

Table 5.6: BNN hyperparameters and results for each data set. Accuracy and AUC metrics are split into train (in-sample) and test (out-of-sample) values. Results reflect a single prediction from the BNNs and will vary due to the stochastic nature of the feed-forward BNN operation.

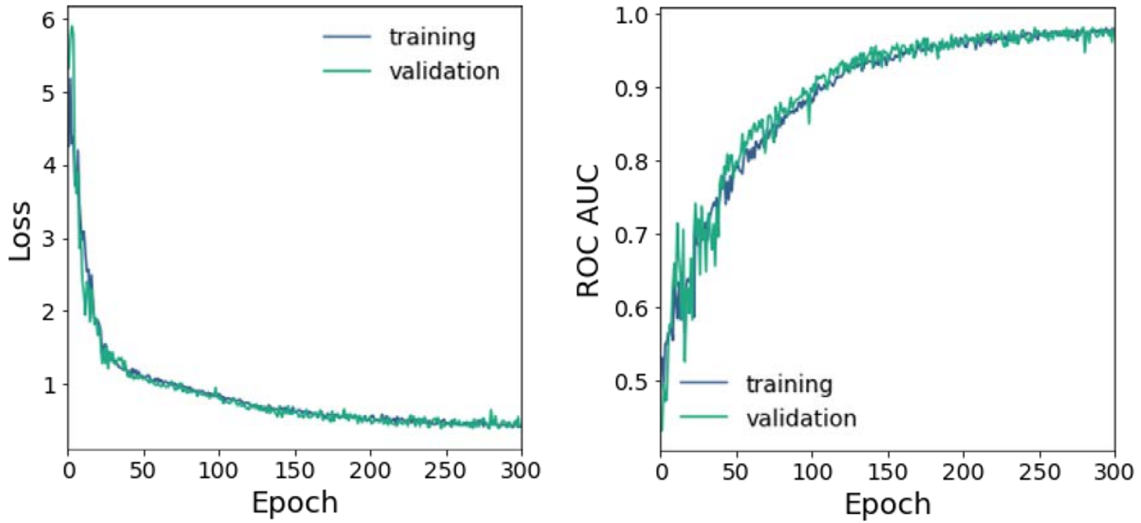


Figure 5-34: BNN training results as a function of epoch count for WDS4. Calculated loss (Left) and AUC (Right) for the training (blue) and validation (green) subsets show training convergence by ≈ 300 epochs. The stochastic nature of the BNN makes the AUC/epoch curve noisier than for the deterministic ANN (Figure 5-24)

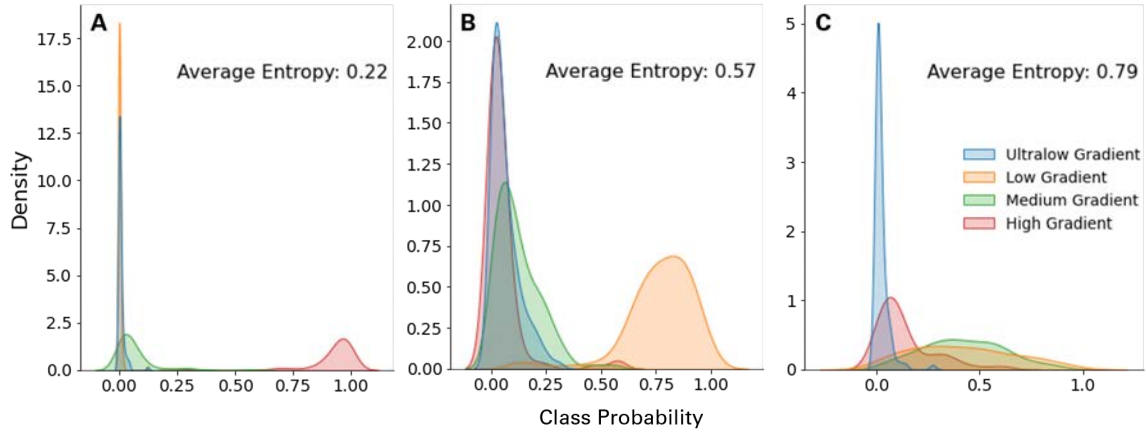


Figure 5-35: Example class label probability density functions for A. well record 34, B. well record 20, and C. well record 35 from WDS4. Distributions are constructed from 100 passes through the WDS4 BNN. Entropy values scale with label distribution entanglement. The predicted label is the one with the highest mean probability: A. High gradient, B. Low gradient, C. Medium gradient.

AOI were generated 1000 times. Geothermal-gradient class probabilities from these realizations were averaged by class for each point location, and the maximum probability class was selected as the predicted label. Entropy values calculated from the ensemble-averaged probabilities are shown in Figure 5-36. Figure 5-37 combines the 1000-run average prediction map with parameter uncertainties using a layer mask and transparency. Concentrated areas of high entropy/uncertainty appear to the southeast to either side of the Rio Grande, and to the north near the edge of the Colorado Plateau (Figure 5-37). Overall, there appear to be fewer locations with high entropy in the parameter uncertainty assessment compared to the structural uncertainty analysis for the four model architectures (Figure 5-31).

5.5.3 Measurement Uncertainty

In the present study, the sources of data, types of data, and options for error estimation vary greatly. Several features were acquired as pre-gridded raster files without accompanying error assessments or access to the original source data (Table 3.1), e.g., air temperature, precipitation, strain rate, and layers from the Bielski et al. OpenEI submission (Kelley, 2015). Table 5.7 shows the top ten features identified

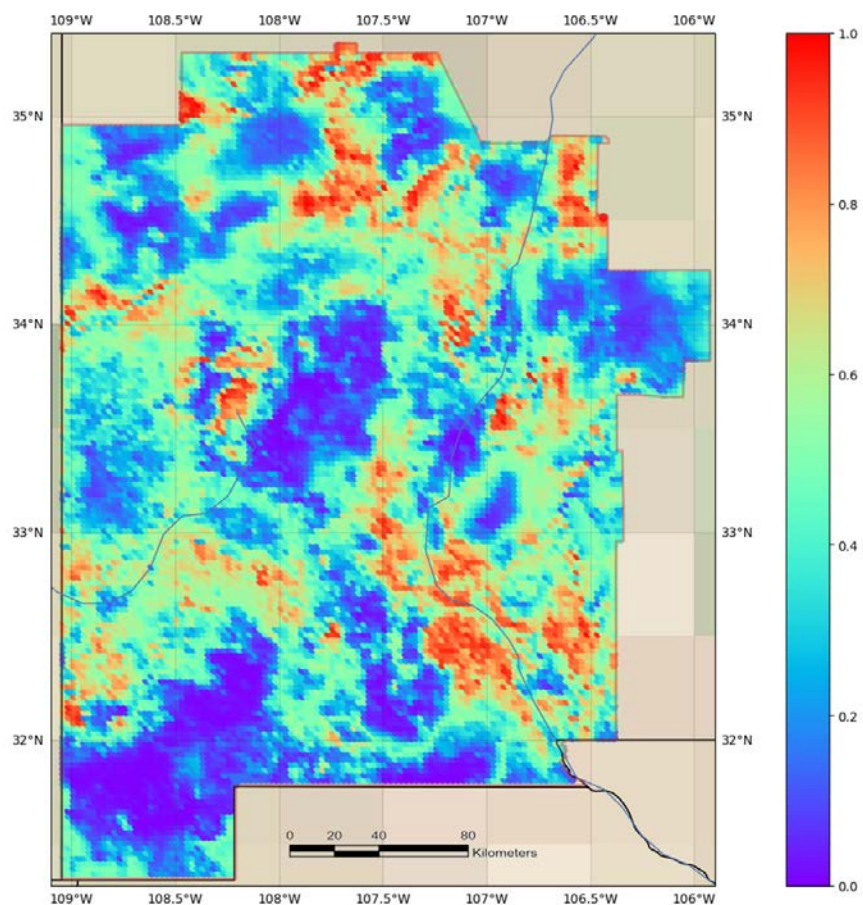


Figure 5-36: BNN parameter uncertainty determined from class probability averages after 1000 runs of the WDS4 predictive model. Uncertainty is measured with normalized Shannon entropy values ranging from 0 for low entropy, low uncertainty (blue) to 1 for high entropy, high uncertainty (red).

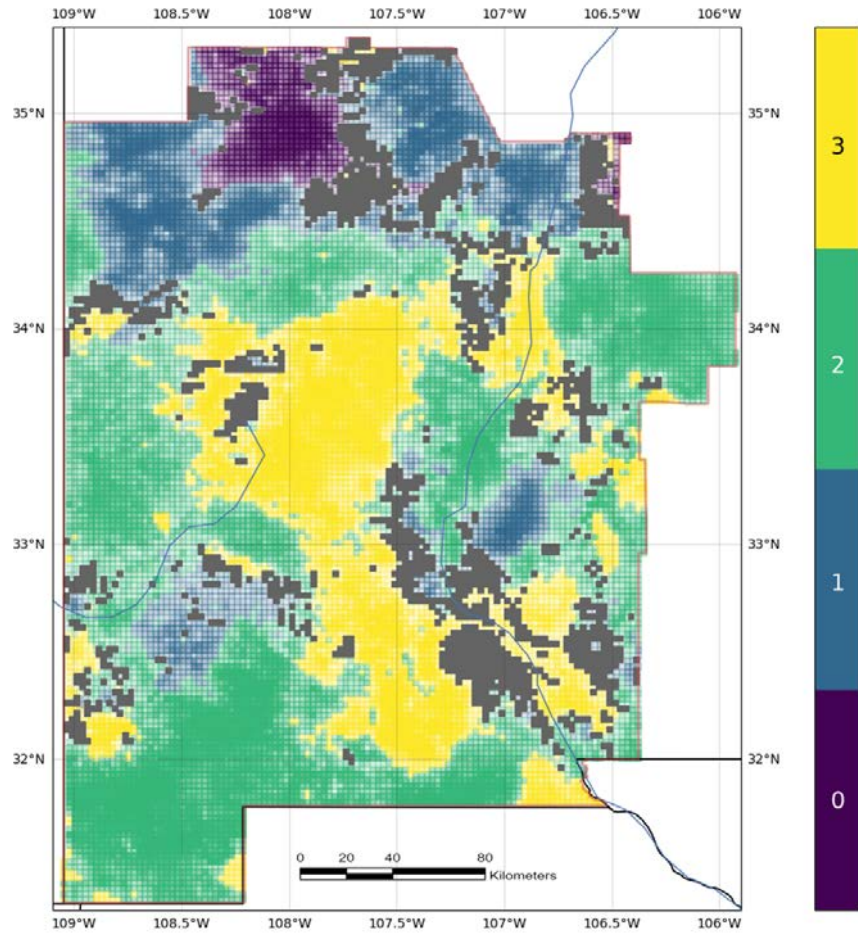


Figure 5-37: Ensemble-averaged WDS4 BNN prediction map with uncertainty masking. Normalized entropy values > 0.7 are grayed out, values ≤ 0.7 determine transparency of the colored scatter plot. Transparency increases from none at entropy values close to 0 to full for entropy values close to 1. The background topographic raster has been removed to better visualize the transparency effect.

by Shapley analysis for the XGBoost model in Section 5.3.2. Among those variables, Silica Geothermometer Temperature (SiGT), Heat Flow, and Boron Concentration have readily-available standard error estimates either from the original source or data preparation process. Crustal Thickness standard error is both unknown and difficult to ascertain; the feature grid comes from contours fit to 2-D seismic models (Keller et al., 1991), which are unavailable for further analysis. The remaining top-ten features consist of point or line data converted to density layers using KDE. Standard errors could be estimated for these features by performing bootstrap or jackknife resampling, applying KDE on the derived data sets, and calculating standard errors from that sample population (James et al., 2013, p. 187–190; A. McIntosh, 2016).

Name	Type	Std Err Estimation	Comments
Si-Geotherm. Temperature	Overlapping points	Kriging	Standard error estimates directly from kriging.
Heat Flow	Non-overlapping points	Directly provided	Standard error estimates provided for each grid point.
Crustal Thickness	Lines	Unknown	Contours based on 2-D seismic lines. Uncertainty in interpretation (e.g., velocity model) unavailable.
Volcanic Dike Density	Lines	Resampling	Jackknife/bootstrap resample and generate density estimates.
Spring Density	Non-overlapping points	Resampling	Jackknife/bootstrap resample and generate density estimates.
Earthquake Density	Overlapping Points	Resampling	Jackknife/bootstrap resample and generate density estimates.
Volcanic Vent Density	Non-overlapping points	Resampling	Jackknife/bootstrap resample and generate density estimates.
Boron Concentration	Overlapping points	Kriging	Standard error estimates directly from kriging.
Quaternary Fault Density	Lines	Resampling	Jackknife/bootstrap resample and generate density estimates.
Drainage Density	Lines	Resampling	If extracted from DEM, relates to DEM slope error. However, can probably use jackknife/bootstrap method.

Table 5.7: Methods of determining standard error estimates for the top ten most important features, identified from Shapley analysis of the WDS4 XGBoost classifier (see Figure 5-17).

Shapley results show that SiGT values influence the XGBoost model predictions more than any other feature (Figure 5-17). Uncertainty in the values for this feature should therefore translate into uncertainty in the classification results. To examine this relationship further, the SiGT values assigned to well locations in WDS4 were perturbed to create a range of statistically-similar derived data sets. Variability in the classification results after training on these data sets highlights model sensitivity

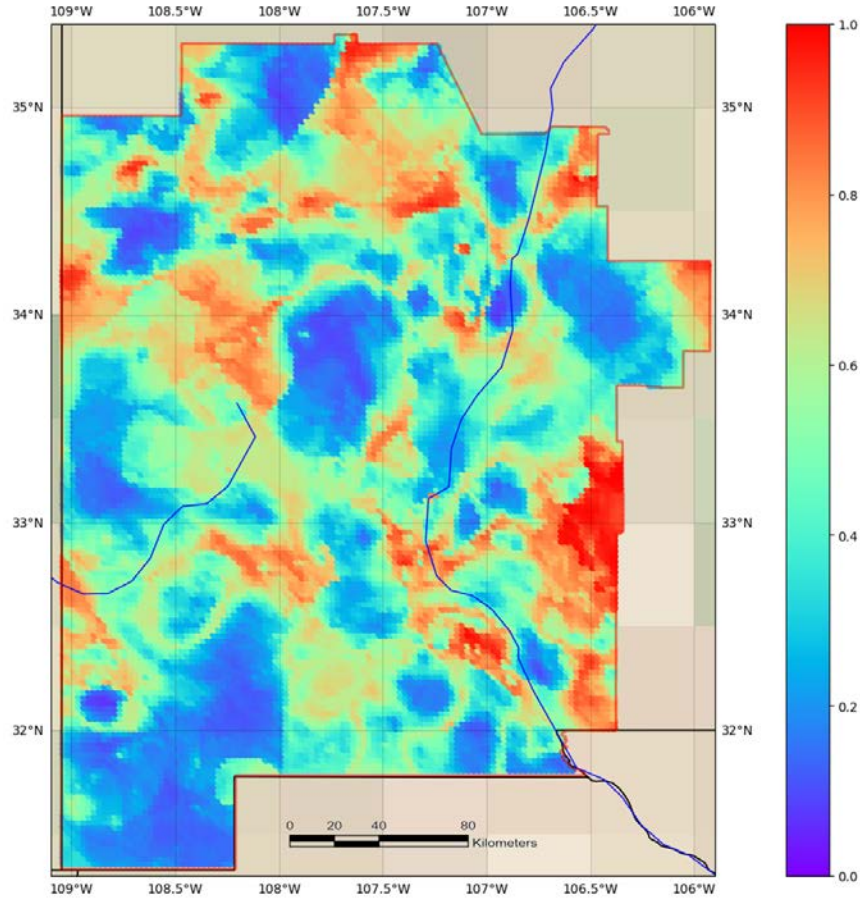


Figure 5-38: SiGT measurement uncertainty based on normalized Shannon entropy. Calculations were made on class probability averages after modeling 100 randomly-perturbed realizations with the WDS4 XGBoost classifier. Values range from 0 for low entropy, low uncertainty (blue) to 1 for high entropy, high uncertainty (red).

to uncertainties in the feature measurements.

A SiGT standard error estimate map (Figure 5-40A) was derived using the ArcGIS *Empirical Bayes Kriging* method as described in Section 3.2.1. All SiGT observations were included in the algorithm, so variability in coincident points due to repeated field measurements influences the estimate. Standard errors were sampled from this map at WDS4 well locations and used to construct a normal distribution ($N(0, \sigma)$) for each location. A total of 100 variants of WDS4 were created by applying perturbations to the WDS4 SiGT values sampled from these distributions.

XGBoost models, parameterized as described for WDS4 in Table 5.4, were trained on the perturbed data sets and used to predict geothermal gradient classifications for

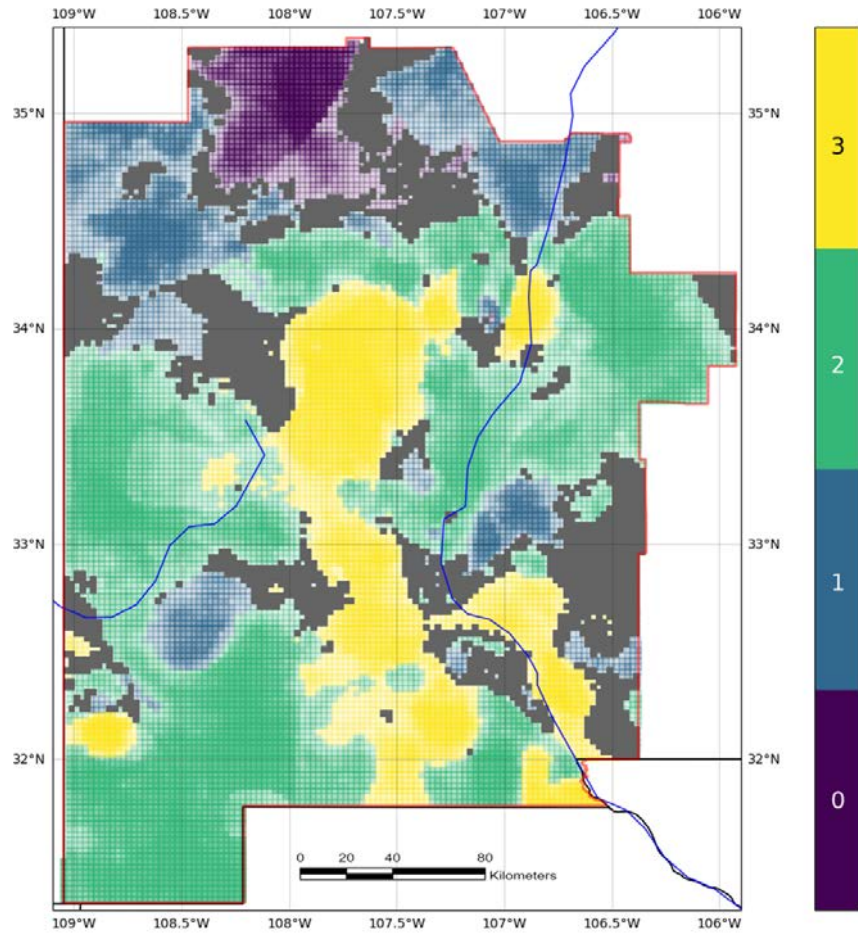


Figure 5-39: Ensemble-averaged WDS4 XGBoost prediction map with SiGT measurement uncertainty masking. Normalized entropy values (> 0.7) are grayed out, and values ≤ 0.7 determine transparency of the colored scatter plot. Transparency increases from none at entropy values close to 0, to full for entropy values close to 1. The background raster has been removed to better visualize the transparency effect.

the FDS covering the full AOI. After ensemble-averaging the class probabilities at each point, the maximum probability class was selected as the class assignment for the average geothermal-gradient classification map. Shannon entropy values calculated from the average class probabilities are shown in Figure 5-38. Figure 5-39 combines both maps using entropy values > 0.7 as a layer mask and assigning increasing transparency with greater entropy for the remaining points.

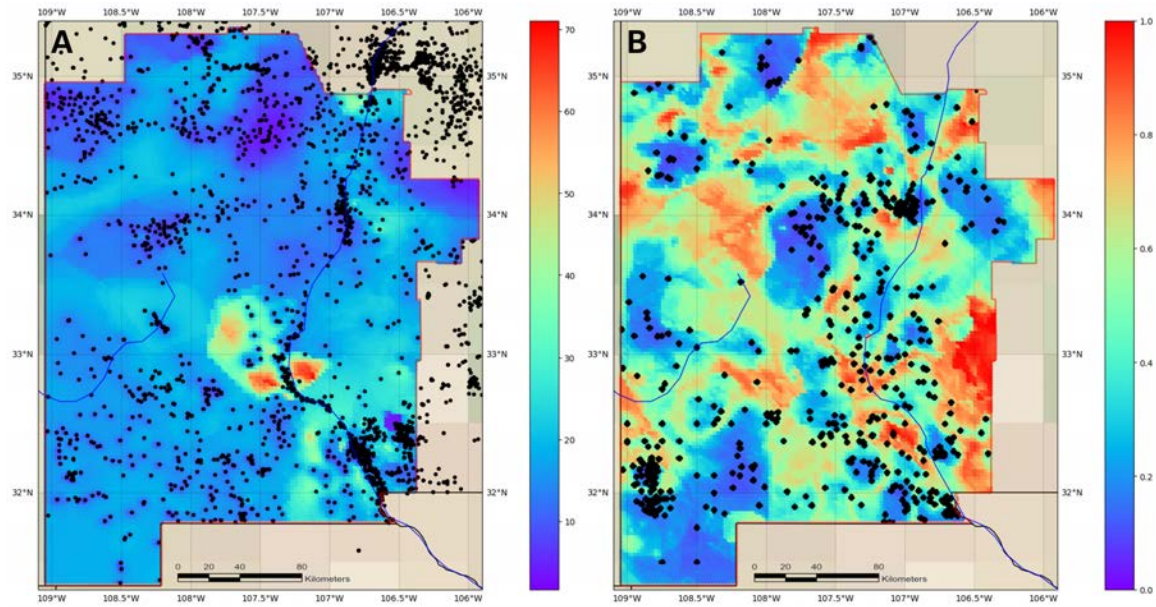


Figure 5-40: A. SiGT standard-error map derived from kriging operation in ArcGIS. Block dots show the locations of the silica concentration samples that are the feature source data. B. WGS4 XGBoost measurement-uncertainty entropy map. Black circles indicate well locations in WDS4 used for training the final XGBoost classifier.

Interestingly, variations in SiGT values result in high classification uncertainty in many locations across the AOI (Figure 5-39). Concentrated areas of uncertainty are observed in the east-southeast—an area where silica samples were collected but not where many WDS4 well observations are located (Figure 5-40A-B). Similarly, large patches of higher entropy appear in the north and just west of the AOI center. These are also under-sampled regions in WDS4 (Figure 5-40B). It appears that as training data values for SiGT change due to perturbations, thresholds for XGBoost decision trees shift enough to significantly impact the stability of model predictions away from wells. SiGT-related splits will appear near the root node of decision

trees based the dominance of SiGT over other features for classification importance. Changes to the decision thresholds will thus have a strong cascading effect on the final classification choices for XGBoost models. The problem lies in both the magnitude of SiGT standard errors as well as the heterogeneous sampling of the study area by WDS4 well locations. The degree of uncertainty seen here would likely be reduced if WDS4 had more comprehensive coverage of the study area.

5.6 Comparative Study Insights

5.6.1 Southwestern New Mexico PFA

In the Southwestern NM PFA project, [Bielicki et al. \(2015\)](#) focused on hydrogeologic windows, defined as areas where erosion, faulting, or volcanic intrusions breach regional sealing layers and allow heated groundwater to reach the surface. Fundamentally, their work targeted hydrothermal systems when assigning geothermal favorability scores, treating heat, fluid volume, and flow path/reservoir as the key risk elements. The last risk element combines particle tracking of geochemical signatures (lithium, boron) with fluid recharge-discharge pathways, effectively replacing permeability and seal risk elements with a model of the subsurface water cycle. Although interesting in its own right, the integrated nature of the PFA favorability map makes it a poor benchmark for the machine-learning results here which focus on the heat-risk element alone. Instead, the geothermal-gradient feature map described in [Figure A.27](#), which [Bielicki et al. \(2015\)](#) derived from NGDS data paired with additional oil & gas well measurements, serves as the best reference for comparison.

The PFA study also honed in on specific locations in the study area due to perceived geothermal favorability. [Figure 5-41](#) highlights four regions of interest combined with the SW NM physiographic province outlines and areas noted for high geothermal gradients (> 60 K/km) in the [Bielicki et al.](#) gradient-feature map. To the southwest in the BR province lies Lightning Dock, the focus of the cost model case study in [Chapters 4 and 6](#) and a USGS-designated hydrothermal KGRA since

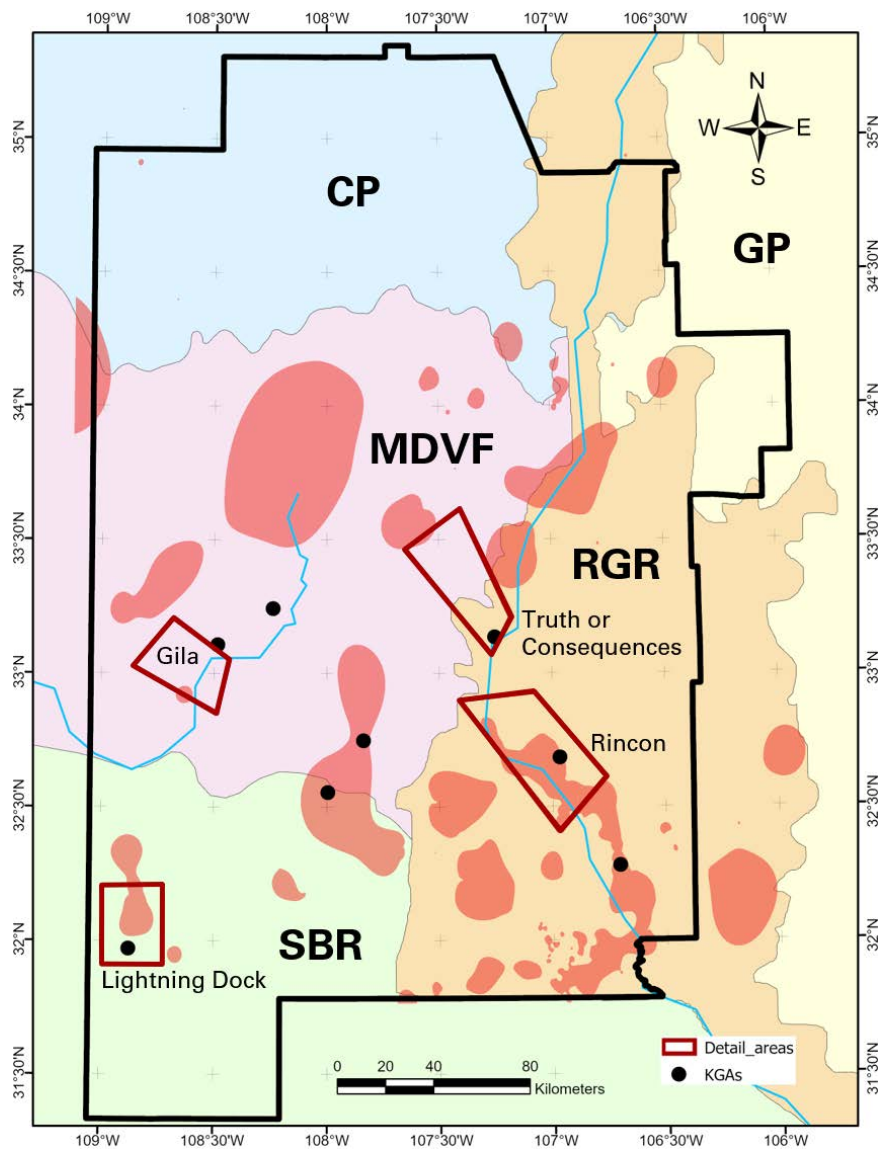


Figure 5-41: Four prospective areas (dark red quadrangles) selected among several in the Southwestern NM PFA study (Bielicki et al., 2015), including the Gila region, Lightning Dock, Rincon, and Truth or Consequences. Filled red regions contain areas with high-grade geothermal gradients in the Bielicki et al. feature map. Black dots mark locations of USGS Known Geothermal Areas with most-likely resource temperatures > 60 °C. Labeled physiographic regions include Colorado Plateau (CP), Great Plains (GP), Mogollon-Datil Volcanic Field (MDVF), Rio Grande Rift (RGR), and Southern Basin and Range (SBR).

1974. To the south in the RGR province, the Rincon geothermal region is marked by high boron concentrations, high heat flow, and shallow geothermal gradients exceeding 300 K/km (Witcher, 2002). North of Rincon is Truth or Consequences, an area locally known for low-temperature hot springs, high permeability, and elevated concentrations of both boron and lithium (Pepin et al., 2015). Finally, a region near the Gila River within the MDVF province was flagged as prospective based on known hot springs, high heat flow, and positive well geochemistry (Bielicki et al., 2015) .

In Figure 5-42, the same prospect quadrangles are superimposed on the geothermal gradient solutions for the four supervised machine-learning models. All models identify high-grade geothermal gradients in the Lightning Dock (LD) and Rincon (RC) areas. Interestingly, the decision tree model predicts a relatively lower gradient at the Radium Springs KGA (marker in RC polygon, Figure 5-42B) where high gradient values have been directly measured. This serves as a good example of where individual models can deviate from a multi-model consensus. All models predict high gradients in the NW part of the Truth or Consequences (TC) polygon, but not at the marked TC spring (Figure 5-42). Gila (GL) also has mixed signals; the decision tree (Figure 5-42B) and neural network (Figure 5-42D) models suggest isolated patches of high gradient potential, but logistic regression (Figure 5-42A) and XGBoost (Figure 5-42C) make Gila look fairly unremarkable.

The uncertainty analysis from Section 5.5 reveals more on how to manage the four areas as exploration prospects. Figure 5-43 quickly confirms some of the observations just outlined. There is low to moderate-low structural uncertainty (i.e., model agreement) for mid-grade gradients in GL and high-grade gradients at LD. High structural uncertainties appear in the TC and RC polygons. For TC, models disagree on the exact boundary between mid- and high-gradient zones. RC uncertainty is more pervasive, so additional modeling or model rationalization would be advised before executing expensive field operations (e.g., drilling) in that area.

Figure 5-44 illustrates parameter uncertainty based on repeated runs of the BNN model. Gradient predictions for LD, GL, and TC all share moderate to high certainty. However, low parameter certainty is observed throughout the RC area. This

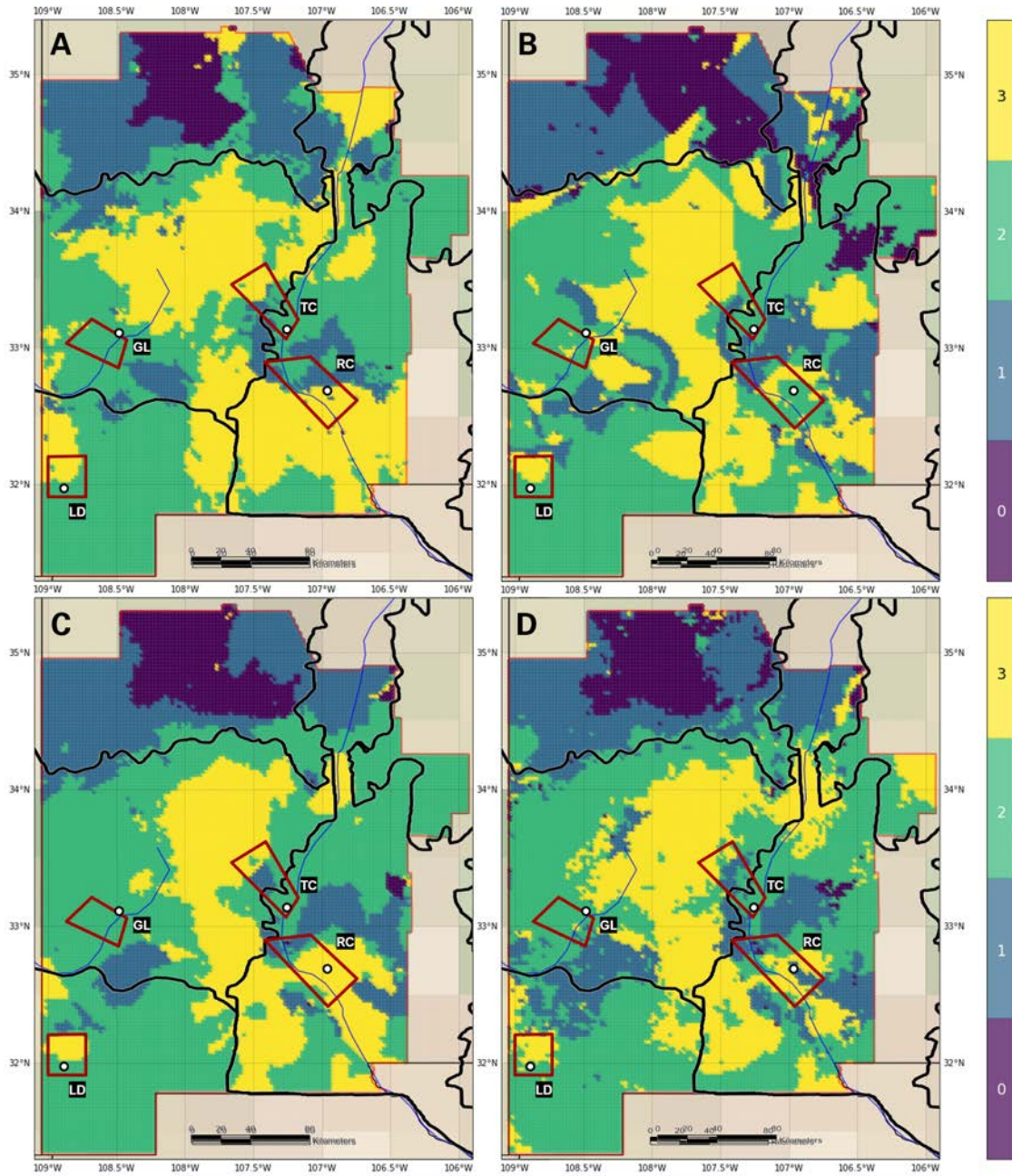


Figure 5-42: Machine learning results for WDS4 from A. logistic regression, B. decision tree, C. XGBoost, and D. neural network models. Dark red quadrangles and labels identify prospective locations discussed in the text. Bold black lines depict the boundaries between the main physiographic regions in the study area. White markers plot KGAs related to the prospects with most-likely resource temperatures $> 60^{\circ}\text{C}$.

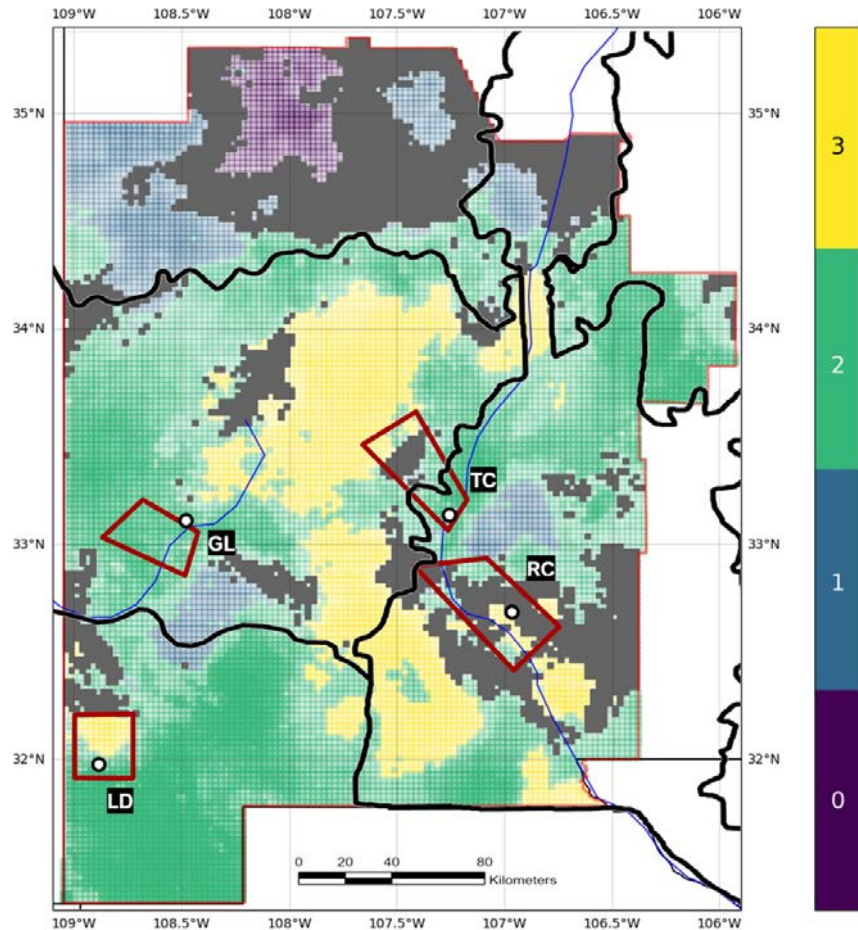


Figure 5-43: Combined-model prediction map with uncertainty masking as described for Figure 5-31. Dark red quadrangles and labels identify prospective locations discussed in the text. Bold black lines depict the boundaries between the main physiographic regions in the study area. White markers plot KGAs related to the prospects with most-likely resource temperatures $> 60^{\circ}\text{C}$.

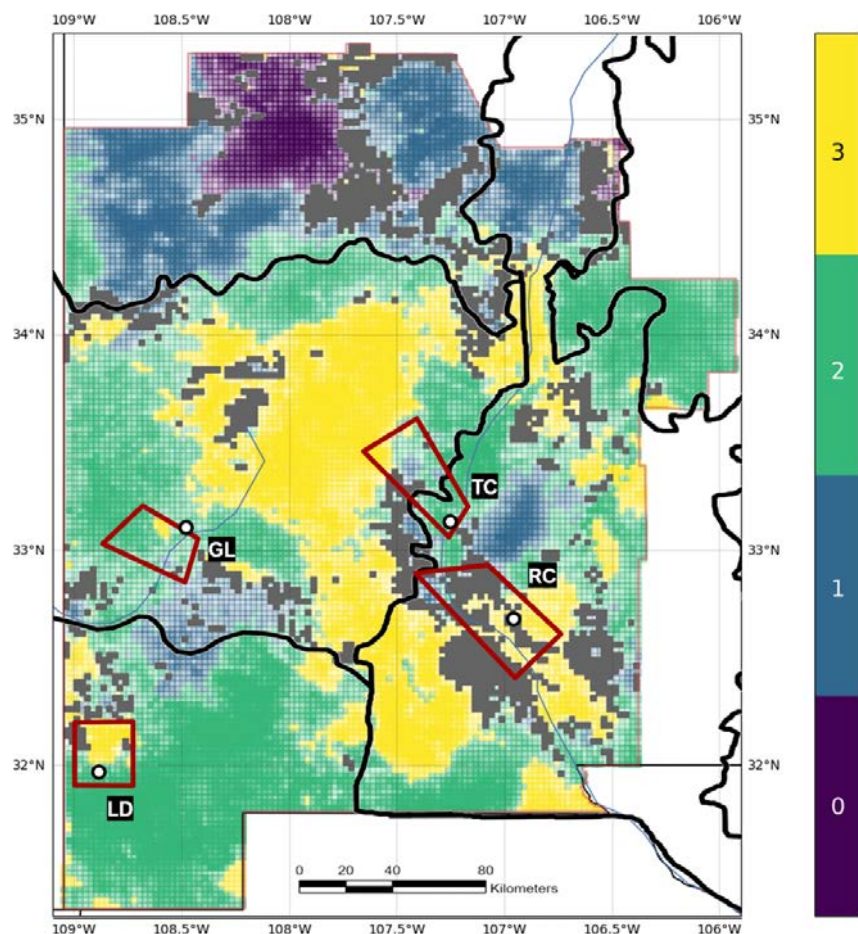


Figure 5-44: BNN ensemble-averaged prediction map with parameter-uncertainty based masking, as described for Figure 5-37. Dark red quadrangles and labels identify prospective locations discussed in the text. Bold black lines depict the boundaries between the main physiographic regions in the study area. White markers plot KGAs related to the prospects with most-likely resource temperatures $> 60^{\circ}\text{C}$.

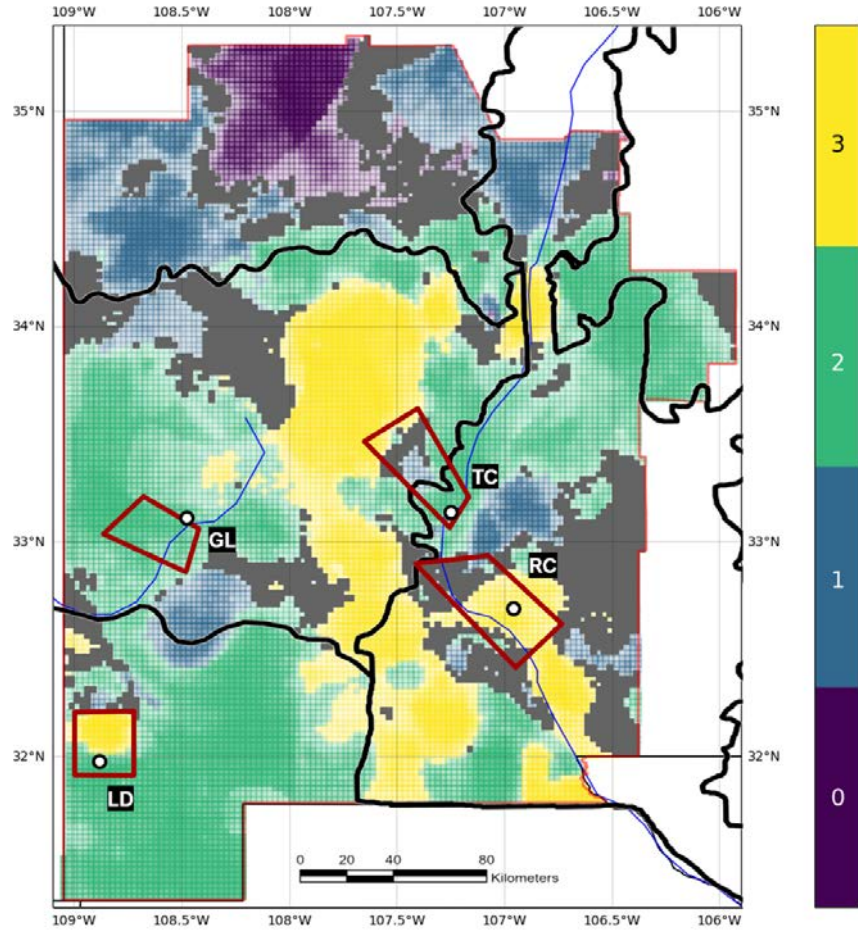


Figure 5-45: Ensemble-averaged WDS4 XGBoost prediction map with SiGT measurement uncertainty masking as described for Figure 5-39. Dark red quadrangles and labels identify prospective locations discussed in the text. Bold black lines depict the boundaries between the main physiographic regions in the study area. White markers plot KGAs related to the prospects with most-likely resource temperatures $> 60^{\circ}\text{C}$.

uncertainty may at least partly account for the diversity of geothermal gradient class assignments within the RC quadrangle by the ANN model (Figure 5-42D). If RC was a target prospect, using predictions from the neural network alone to guide next steps would not be advised given the level of uncertainty. Explorationists would do well to choose a different, more certain model, particularly if no additional data are available for training the neural network further.

Measurement uncertainty depicted in Figure 5-45 considers Si Geothermometer Temperature (SiGT) in isolation from the rest of the input feature set, noting that

SiGT dominates in feature importance for the XGBoost model (see Section 5.3.2). The level of measurement certainty for all four prospective areas is striking. A strip of low certainty appears at the gradient class boundary at TC, and likewise to the NW for Rincon. However, these results indicate no strong need to supplement or replace the SiGT data for the prospect areas, as might be the case for other locations within the AOI.

In summary, the uncertainty analysis shows:

- All models agree LD has the best high-grade gradient potential, while GL and TC are less prospective. The models are collectively inconclusive on RC.
- Neural-network parameter uncertainty makes it an unreliable model for assessing the RC area.
- SiGT measurements are relatively certain for the prospect areas. Additional expenditures for SiGT data may not be necessary.

Treating this analysis as part of a hypothetical exploration program, the overall heat-element favorability and low uncertainty would make LD a good prospect to progress to the well-planning stage. High uncertainty at RC would need to be addressed through additional pre-screening efforts before committing the resources and capital to drilling there.

5.6.2 Southwestern New Mexico PCA

In a more recent study, [Pepin \(2019\)](#) applied unsupervised PCA with *K*-means clustering to assess the geothermal potential of the Southwestern NM study area. [Pepin](#) noted the physiographic provinces in NM exert strong controls on the presence and type of USGS-identified KGRAs observed across the state. The cluster analysis identified two high-potential zones with different characteristics. The first defines a predominantly high-temperature hydrothermal zone stretching east-west from LD to the eastern AOI boundary and up along the Rio Grande to just above RC ([Pepin, 2019](#), Figure 3.5A). A second high-potential region surrounds the narrow zone where CP,

MDVF, RGR, and GP provinces all converge (Pepin, 2019, Figure 3.5C), although this second group largely comprises systems with low-temperature hot springs.

The supervised-learning results presented in Figures 5-42 and 5-43 clearly display a correlation between high-grade geothermal gradient and physiographic province, but the relationship differs from that described by Pepin (2019). Models consistently predict high-grade gradients in the southern RGR and eastern MDVF provinces, with an additional “hot spot” where the RGR narrows to the north. This pattern repeats for all four machine learning models, and is present, albeit less obvious, in the geothermal-gradient feature map from Bielicki et al. (2015) (see Figure 5-41). Pepin (2019) places the MDVF province in a low-potential cluster, driven in part by risk elements not considered here like structure/permeability. Nevertheless, the PCA model shows relatively strong overlap between clusters (Pepin, 2019, Figure 3.4), suggesting additional data engineering or alternative dimensionality-reduction methods could be of value before fully-discounting the prospectivity of the MDVF province.

Cluster analysis shows great promise, particularly in identifying how the influence of different features changes across a study area. Pepin (2019), Smith et al. (2021), and Vesselinov et al. (2020) all illustrate this point with significantly different dominant predictors for clusters correlated to individual physiographic provinces. Similar variance in feature importances was observed in the Shapley analysis by class (Figure 5-17); e.g., Si Geothermometer Temperatures top the importance list for class 0 and 1, but come in 2nd to Volcanic Dikes for class 3, and 4th from last in importance for class 2.

Perhaps more interesting is the relationship between high-grade geothermal gradients and complex physiographic province *junctions* (i.e., physical geography interfaces) as seen in Figure 5-42. Specifically, the two boundary zones between 3 or 4 provinces within the AOI both have consistent high-grade geothermal gradient classifications. These zones also correspond with the two prospective clusters in the PCA analysis (Pepin, 2019). Whether greater geothermal potential at complex province junctions is unique to this study area or more broadly applicable requires further

research.

5.7 Recap

This chapter presented the results from applying the four supervised machine-learning models described in Chapter 3 to the curated data (Appendix A) for the Southwestern NM study area. Uncertainties associated with the choice of models, the model parameters, and data measurements were also examined in the context of next-steps decisions for an explorationist. The chapter concluded with a comparison with other geothermal prospectivity studies for the region.

Key insights from the work include:

1. Logistic regression models are simple and easy to tune, however other model techniques have much stronger predictive performance.
2. Decision trees are highly-explainable and directly identify feature importances. A notable downside is results can vary based on random structural variations during construction.
3. XGBoost models demonstrate very strong predictive performance and support best-in-class Shapley feature attribution analysis. However, XGBoost model tuning can be complex and time-consuming.
4. The highest degree of model complexity comes from neural network architectures, which show great performance if sufficient training data is available to constrain the enormous number of model parameters.
5. Structural uncertainty analysis highlights areas where models collectively differ or agree on predictions. This is useful for rationalizing model selection and identifying spatial areas that require further study before taking costly actions.
6. Parameter uncertainties indicate the “known unknowns” for a model, i.e., locations where model predictions are poorly constrained. Corrective actions include training on more data or relying on a different model technique.

7. Measurement uncertainties characterize the impact of data reliability on model predictions. If high uncertainty exists in an area of interest, additional data acquisition may be needed, especially from sources with lower standard errors.
8. The value of this approach to exploration risk mitigation is demonstrated for prospective areas identified in the Southwestern NM PFA study. High favorability, low uncertainty prospects (e.g., Lightning Dock) could progress to a well-planning stage, but areas with mixed favorability and/or high uncertainty (e.g., Rincon) would require additional pre-screening first.
9. A correlation between physiographic provinces and geothermal favorability established in other studies is also observed in the Southwestern NM study area. In addition, high-grade geothermal gradients (positive heat risk-element values) align with complex convergence zones between 3 or more physiographic provinces.

Chapter 6

EGS Power Plant Expansion

Cost Model Results

Chapter 4 outlined the cost-modeling strategy for a hypothetical 5-MW expansion project of the Lightning Dock power plant in Animas Valley, NM. This chapter reviews the results of the different model approaches, explores insights gained from those models, and describes how this approach mitigates risks associated with geothermal development and production.

6.1 Static Model

6.1.1 Model Selection

Section 4.2.1 described the use of brine effectiveness in the cost model for determining the power output of a binary cycle plant for a given production temperature and flow rate. This formulation provides a choice of how to manage the cost-model mechanics due to a trade-off between plant capacity and flow rate for a given brine effectiveness (Equation 4.2).

In addition, installation of the Lightning Dock expansion can take place over a variety of different deployment schedules due to the modularity of the system. Rather than drill ten wells and install five binary-cycle plants all at once, deciding to delay

aspects of the installation can be financially beneficial and less of an initial risk for the project.

Figure 6-1 shows the results for the pre-set capacity and pre-set flow-rate static models for sixty installation schedule permutations. The pre-set capacity model results in project losses of \$20 million or more for all tested installation options. By contrast, the fixed flow-rate model only drops below \$0 NPV for a handful of project plans, achieving \$3.7 million NPV for the case where three modules are installed up front and two additional ones go live after a year of operation (red diamond, Figure 6-1). Based on these results, all cost models used in the rest of this analysis apply a fixed flow rate per production well and derive the aggregate electricity generation numbers based on the temperature of the produced brine.

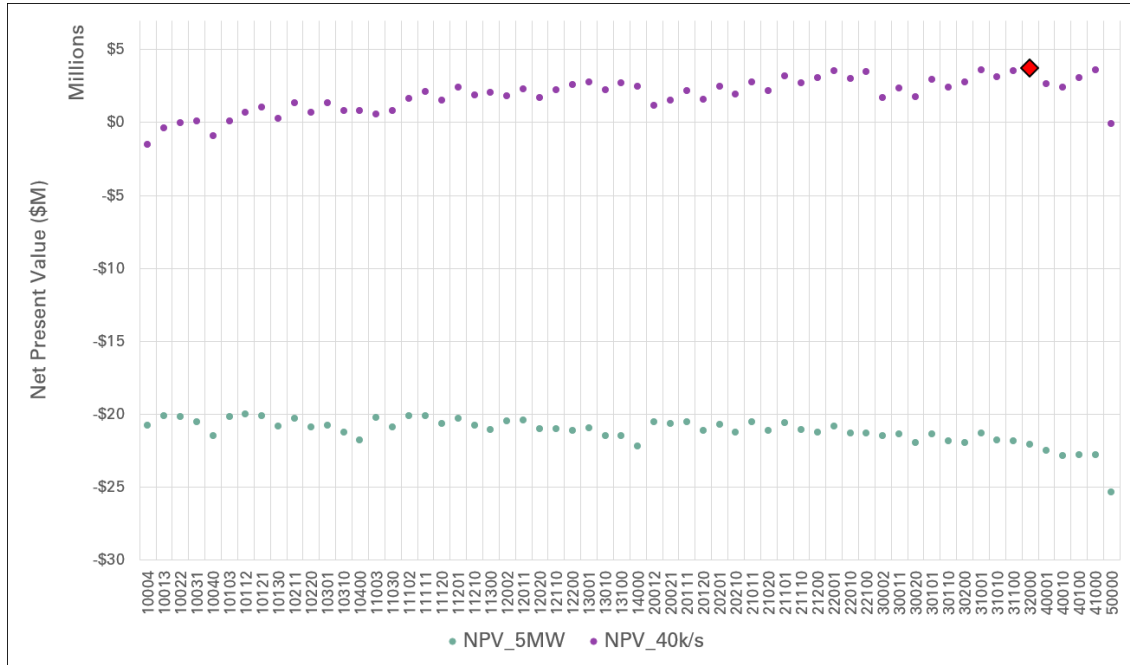


Figure 6-1: Static cost model comparison between pre-set aggregate capacity (5 MW target, green) and pre-set flow rate per production well (40 kg/s, purple), plotted against module installation schedule. Digit d_i in schedule $d_1d_2d_3d_4d_5$ along the x -axis defines that $d_i \in \{0, 1, 2, 3, 4, 5\}$ modules are installed in year i . A total of $\sum_{i=1}^5 d_i = 5$ modules are installed in each of the sixty schedules. The red diamond marks the optimal scenario where the flow rate is fixed and the deployment includes $d_1 = 3$ modules in year 1, $d_2 = 2$ in year 2.

6.1.2 Construction Optimization

Lifting the fixed-capacity requirement changes the production potential for each power-plant module. Using the parameters defined in Section 4.2.3, each module can now generate 2.1 MW. Since this is a hypothetical case study with Climeon modules used as an analog only, further modeling uses this value with the caveat that future studies should confirm its viability as the modular power plant technology continues to evolve.

The updated power production per module reduces the required module count to a total of three (3) modules based on the original expansion target of 5 MW. Table 6.1 revisits the installation schedule grid search exercise to determine the optimal project plan under these circumstances. At an NPV of \$1.0 million, the best option deploys two (2) modules initially and adds an additional one (1) at the end of the first year. In order to standardize cost models for direct comparison, this installation plan is used for all cost models throughout the rest of this analysis.

Year 0	Year 1	Year 2	NPV (\$M)
3	0	0	−1.1
1	0	2	−0.3
1	1	1	0.5
1	2	0	0.6
2	0	1	0.6
2	1	0	1.0

Table 6.1: Grid search for the optimal power-plant installation schedule based on the static cost model. Schedule options are sorted on NPV in \$M, where M is million.

6.1.3 Summary Statistics

As a deterministic cost model, the static model performance is measured strictly on calculated NPV: **\$1.0M**. This value serves as a benchmark for the other cost models explored in the rest of this chapter.

6.2 Probabilistic Model Metrics

Monte Carlo simulation is applied to all of the probabilistic models to estimate the range of model behavior. For each model, results represent 2000 simulated runs or *realizations* ($\mathcal{R}$), where each realization defines a unique combination of variable values sampled from the PDFs reviewed in Section 4.3.3.

Common methods for evaluating a Monte Carlo ensemble include building a NPV histogram, constructing a Cumulative Distribution Function (CDF or target curve), and averaging the results together for Expected Value of NPV (ENPV). Other interesting metrics for model comparison include standard deviation, individual percentiles, and direct comparison with the static model NPV (NPV_s). Standard deviation is the measure of how tightly the results cluster around the mean value. Distributions with low standard deviations are sometimes referred to as robust distributions. P_{50} marks the the median, and P_{05} and P_{95} define marginal percentiles for 5% Value at Risk (VAR) and Value at Gain (VAG), respectively. Each of these measures is reported for the probabilistic model cases described below.

6.3 Base Case Model

The Base Case model mimics the static model in form, but incorporates uncertainties in drilling costs, electricity pricing, geothermal gradient, reservoir temperature, and thermal drawdown rate to provide a more realistic forecast. No decision rules are included in this scenario.

6.3.1 Model Results

At $-\$4.8$ million ENPV, the Base Case model predicts over 560% lower project value than predicted with the static model (Table 6.2). This result alone illustrates how probabilistic approaches can significantly differ from deterministic models that use most-likely or average values. Skewed system performance occurs even when variable distributions are balanced, which makes deterministic results both unrealistic and

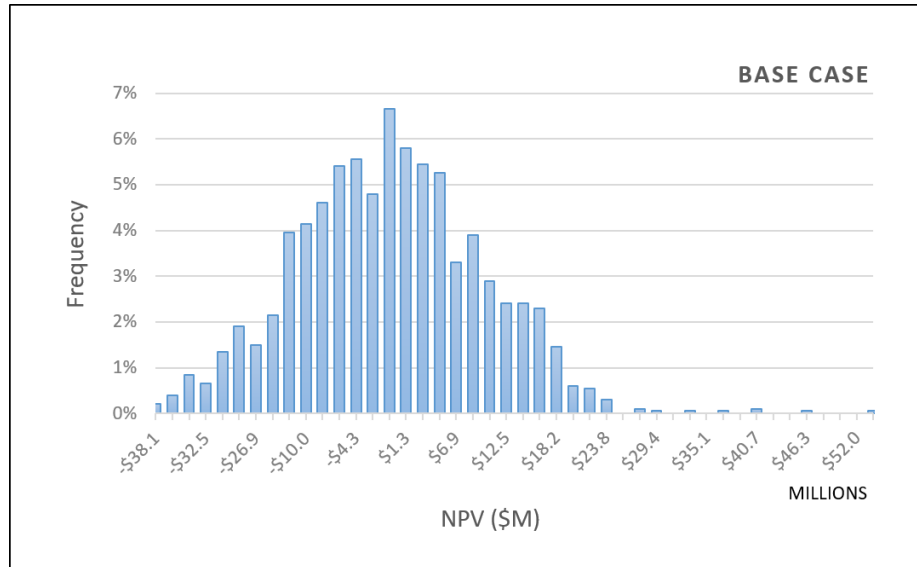


Figure 6-2: Base Case probabilistic cost-model histogram illustrating the distribution of 2000 model realizations. NPV is reported in \$M, where M is million.

unreliable measures for decision-making (de Neufville & Scholtes, 2011, p. 48–49). Here, unanticipated high-side (P_{95} value of \$15.9 million) does exist, but the influence of the low-side (P_{05} value of −\$28.2 million) dominates overall (Table 6.2).

The Base Case ensemble shows a symmetric, pseudo-Gaussian distribution of NPV results, with the exception of a long right tail capturing rare but very positive project outcomes (Figure 6-2). Figure 6-3 illustrates the target curve tracing cumulative probabilities for observed NPV values. The P_{50} value of −\$3.9 million suggests the highly-negative values in the lower half of results pull the average (ENPV) further into negative project value territory. Cumulatively, $\approx 60\%$ of the realizations end in a net loss for the project (Figure 6-3). And at $\approx 3\times$ both the median and ENPV, standard deviation of NPV indicates this solution is not robust.

All told, the Base Case results display both a negative ENPV and a high likelihood of project financial loss. Without a clear strategy for mitigating risk, this project would and should be rejected by a responsible portfolio manager.

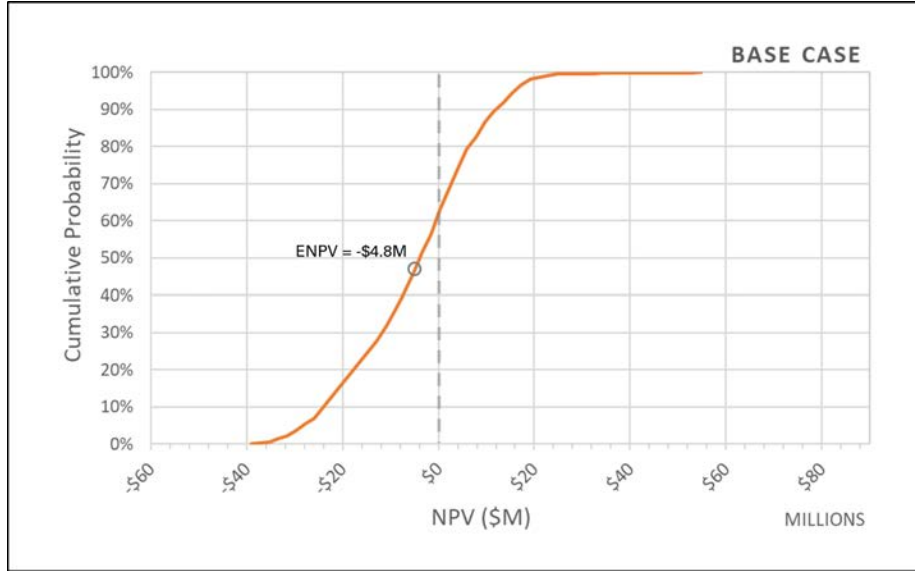


Figure 6-3: Cumulative distribution function for the Base Case probabilistic cost model. The curve summarizes results from 2000 model realizations. NPV is reported in \$M, where M is million.

6.3.2 Summary Statistics

Table 6.2 outlines key statistical measures summarizing the performance of the Base Case probabilistic cost model.

Base Case Statistics	$\mathcal{R} = 2000$
ENPV (\$M)	−4.8
STD(NPV) (\$M)	13.2
P05 NPV (\$M)	−28.2
P50 NPV (\$M)	−3.9
P95 NPV (\$M)	15.9
% Difference from NPV_s	−565%

Table 6.2: Base case probabilistic model statistics for 2000 model realizations. NPV is reported in \$M, where M is million. NPV_s refers to the static model NPV reported in Section 6.1.3.

6.4 Redevelop Case Model

The Redevelop Case model extends the Base Case with a redevelopment plan for the geothermal field to counter thermal decline in the fluid pathways through the reservoir. The Redevelopment decision rule (Section 4.4.1) is triggered when thermal drawdown causes the produced brine to drop below a threshold temperature. As with the Base Case, model uncertainties include drilling costs, electricity pricing, geothermal gradient, reservoir temperature, and thermal drawdown rate.

6.4.1 Model Results

Adding the Redevelopment rule improves the project ENPV by \$2.3 million over the Base Case, but it still remains negative at $-\$2.5$ million, 340% below the static model NPV. Redevelopment drilling costs come into play as the VAG (P_{95} value of \$15.1 million) decreases slightly relative to the Base Case. But VAR (P_{05} value of $-\$21.3$ million) shows a larger difference, improving by nearly \$7 million over the Base Case (Table 6.3). This clearly reflects the improved production and power sales possible by managing reservoir conditions.

Standard deviation of the ensemble distribution is less than observed for the Base Case, likely due to a narrower overall distribution without as many outliers in the tails (Figure 6-4). The balance in distribution shape is reflected in the small difference (\$0.2 million) between ENPV and the P_{50} result (Table 6.3). Comparing the target curves between the Redevelop Case (Figure 6-5) and Base Case (Figure 6-3), it becomes clear that the redevelopment flexibility does not address upside potential. Instead, it acts as a partial stop-gap on the worst-case realizations of the model. The lower half of the curve tightens up, but there is little overall curve movement to the right to make the project more profitable.

As a brief caveat: the idea of periodic redevelopment for a geothermal field is not novel. In fact, many geothermal cost models include it as default behavior (e.g., Entingh et al., 2006; Blair et al., 2018). Nevertheless, the analysis above illustrates why this design option should be included in geothermal operations to help mitigate

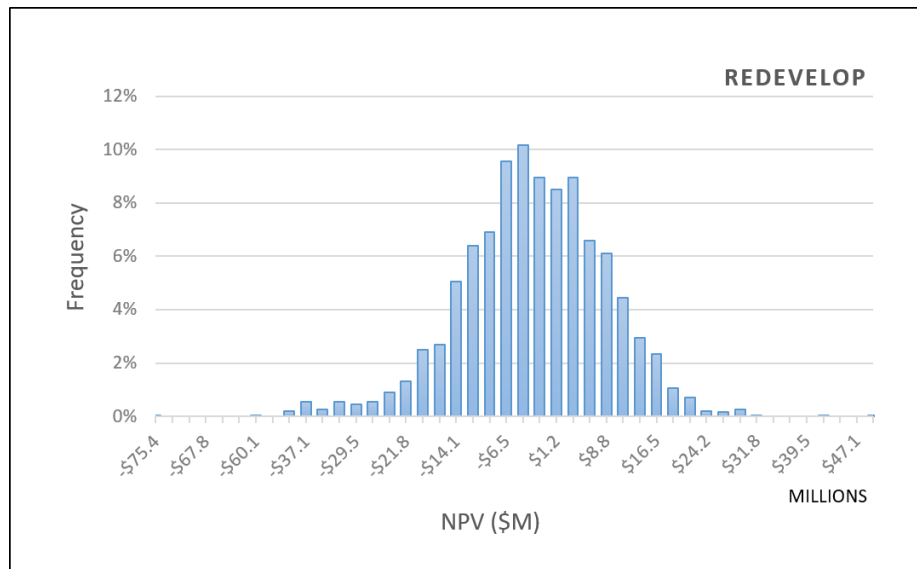


Figure 6-4: Redevelop Case probabilistic cost model histogram illustrating the distribution of 2000 model realizations. NPV is reported in \$M, where M is million.

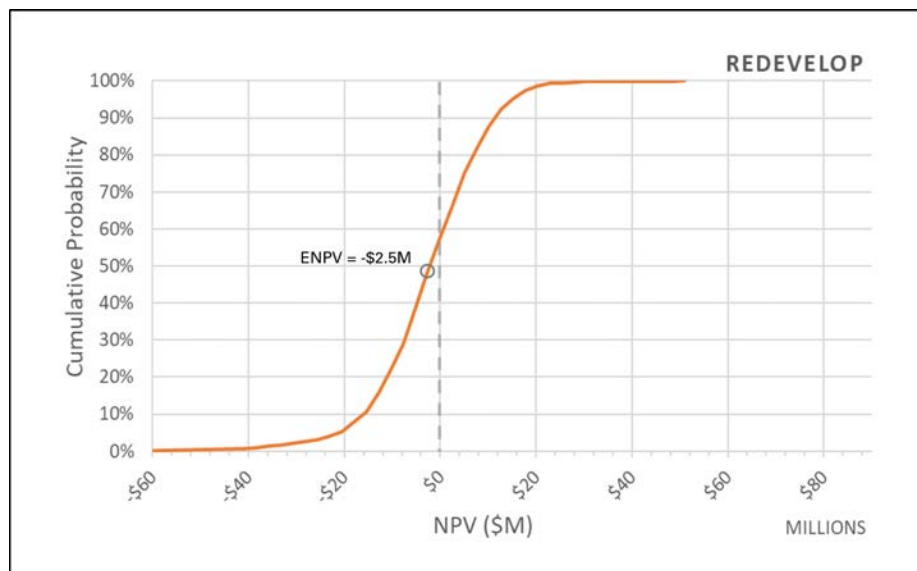


Figure 6-5: Cumulative distribution function for the Redevelop Case probabilistic cost model. The curve summarizes results from 2000 model realizations. NPV is reported in \$M, where M is million.

the risk of high thermal drawdown rates — as long as drilling costs are low enough to make it attractive.

6.4.2 Summary Statistics

Table 6.3 outlines key statistical measures summarizing the performance of the Redevelop Case probabilistic cost model.

Redevelop Case Statistics	$\mathcal{R} = 2000$
ENPV (\$M)	−2.5
STD(NPV) (\$M)	11.7
P05 NPV (\$M)	−21.3
P50 NPV (\$M)	−2.3
P95 NPV (\$M)	15.1
% Difference from NPV_s	−344%

Table 6.3: Redevelop case probabilistic model statistics for 2000 model realizations. NPV_s refers to the static model NPV reported in Section 6.1.3.

6.5 Redevelop & Grow Case Model

The Redevelop & Grow Case model builds on the Redevelop Case with a decision rule around increasing capacity (Section 4.4.2). Specifically, when electricity prices rise by an amount larger than the monitored threshold, more power plant modules are installed to capitalize on the increased prices and inferred demand. Like the other probabilistic models, values for drilling costs, electricity pricing, geothermal gradient, reservoir temperature, and thermal drawdown rate are directly sampled from probability distribution functions (Section 4.3.3) as part of the Monte Carlo simulation.

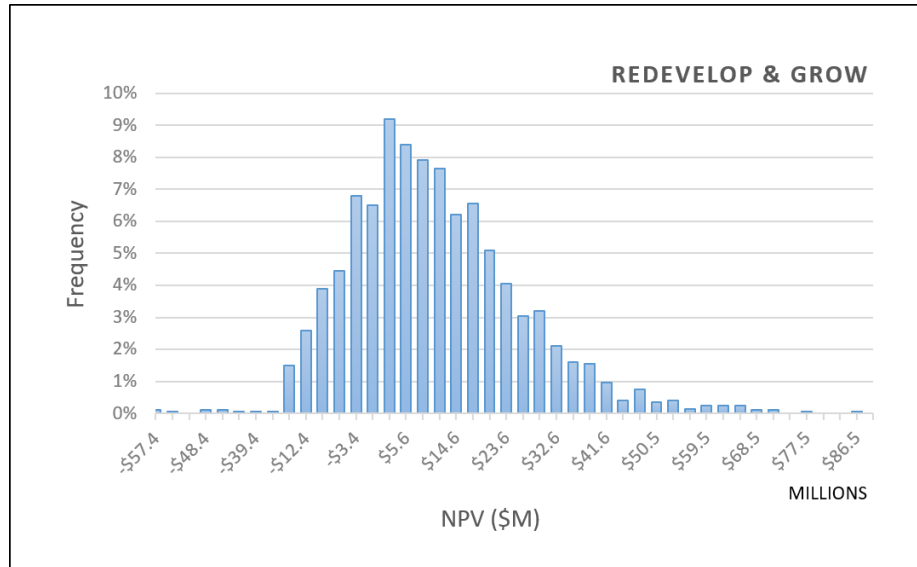


Figure 6-6: Redevelop & Grow Case probabilistic cost-model histogram illustrating the distribution of 2000 model realizations. NPV is in \$M, where M is million.

6.5.1 Model Results

Redevelop & Grow model results use a field expansion amount of 25% when the Capacity Growth decision rule is triggered. Assuming the PPA with the purchasing utility company is successfully renegotiated at the time of the expansion, this model results in a dramatic change in project value. ENPV is \$9.7 million, over \$12 million better than the Redevelop Case and over 800% greater than static model NPV (Table 6.4). The case shows notable improvement in both Value at Risk and Value at Gain; P_{05} shifts by more than \$7 million to $-\$14.2$ million, and P_{95} jumps to \$38.2 million as power-plant growth captures market potential. Project losses occur for 23% of model realizations, compared to 55–60% for the Base and Redevelop Cases.

The histogram for Redevelop & Grow skews noticeably to the right with a long tail of simulation runs marking high-value realizations of the model. Standard deviation of NPV increases compared to the Redevelop Only case, making this model less robust by definition. Robustness measured by standard deviation is a useful metric because it inhibits overconfidence in desirable outcomes. However, robustness alone does not serve as a good criterion for maximizing value; the act of minimizing downside outcomes and expanding upside opportunities may also increase the standard deviation,

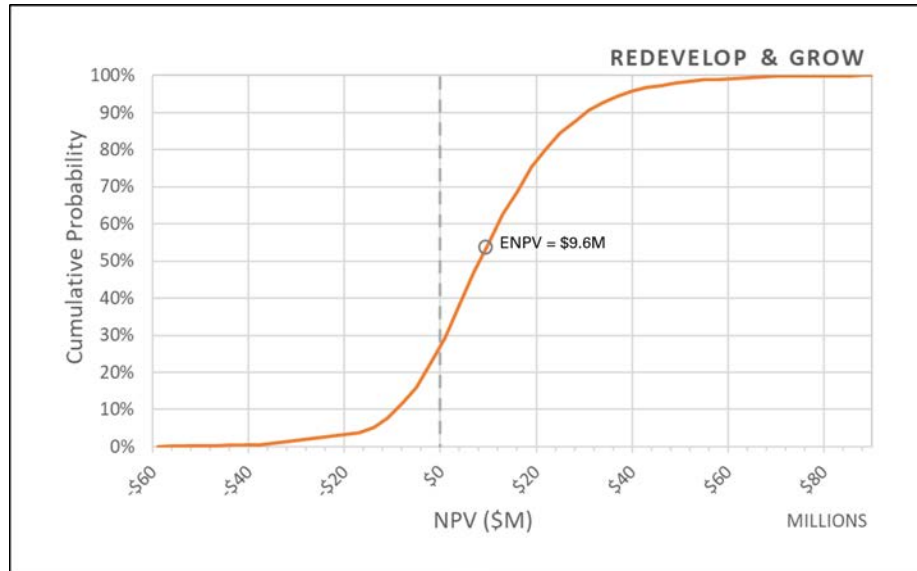


Figure 6-7: Cumulative distribution function for the Redevelop & Grow Case probabilistic cost model. The curve summarizes results from 2000 model realizations. NPV is reported in \$M, where M is million.

as happens with this case. Importantly, the rightward shift of the target curve in Figure 6-7 compared to the previous cases indicates Redevelop & Grow dominates the other scenarios, as suggested by improvements across all NPV metrics. This shift is also illustrated in Figure 6-10.

Some caution should be taken in applying learnings from this model to a real-world geothermal project. Results here depend on a few important assumptions, including the willingness of a partner utility company to accept capacity increases above the market rate for electricity on any given year, the direct relationship between large price increases and electricity demand, and the overall monotonic increase in electricity prices over time. With regard to the latter, the price model used in this analysis includes a single step change and annual volatility (see Figure 4-9), but some high future-electrification forecasts depict continuous decline trends not modeled here (Murphy et al., 2021). As discussed in Section 4.3.1, the interactions between different energy markets and dynamics of major events like future electrification are complex, worth considering, and beyond the scope of this thesis.

6.5.2 Summary Statistics

Table 6.4 outlines key statistical measures summarizing the performance of the Redevelop & Grow Case probabilistic cost model.

Redevelop & Grow Case Statistics	$\mathcal{R} = 2000$
ENPV (\$M)	9.6
STD(NPV) (\$M)	16.5
P05 NPV (\$M)	-14.2
P50 NPV (\$M)	8.2
P95 NPV (\$M)	38.2
% Difference from NPV_s	832%

Table 6.4: Redevelop & Grow case model statistics for 2000 model realizations. NPV_s refers to the static model NPV reported in Section 6.1.3.

6.6 Full Flexibility Case Model

Full Flexibility adds a decision rule around reduction of aggregate plant capacity (Section 4.4.3). If electricity prices drop by a threshold amount, power plant modules are proactively decommissioned to reduce electricity production and operating expenses. Redevelopment and Capacity Growth decision rules remain in effect, as does the random-sampling treatment for drilling costs, electricity pricing, geothermal gradient, reservoir temperature, and thermal drawdown rate based on PDFs defined in Section 4.3.3.

6.6.1 Model Results

In a somewhat surprising outcome, the simulation for Full Flexibility results in an ENPV of \$6.7 million (Table 6.5). Although this value is 545% greater than the static model NPV, it falls short of the Redevelop & Grow case by nearly \$3 million (see Table 6.4). Both VAR and VAG show less attractive results as well; P_{05} is -\$16.3M and P_{95} is \$36.2 million, both $\approx$ \$2 million worse than Redevelop & Grow.

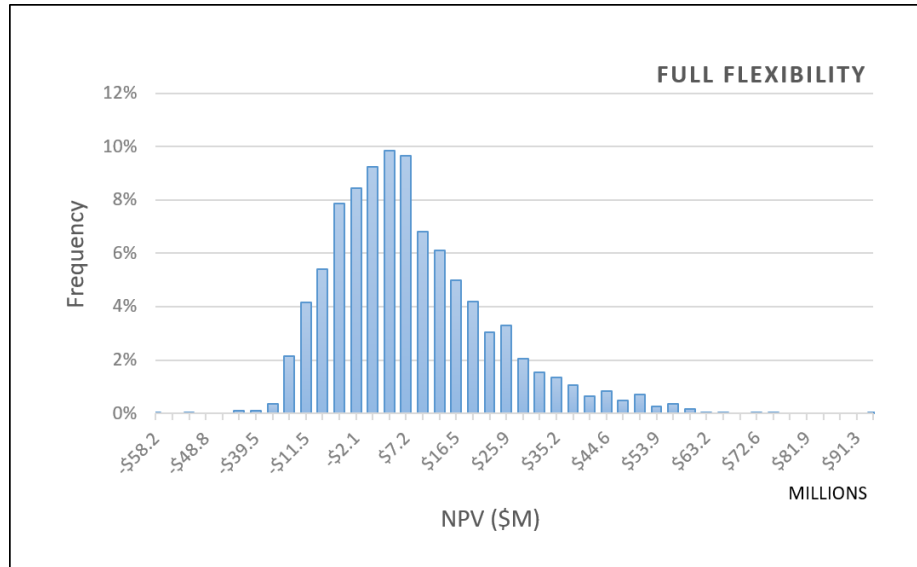


Figure 6-8: Full Flexibility Case probabilistic cost model histogram illustrating the distribution of 2000 model realizations. NPV is reported in \$M, where M is million.

The results histogram (Figure 6-8) shows the same right skew as the Redevelop & Grow model but with a slightly more compact form based on the lower standard deviation (Table 6.5). The difference in median values between the two cases amounts to $\approx$ \$3 million, further confirming the Full Flexibility case is entirely dominated by the Redevelop & Grow case. Figure 6-9 depicts the case target curve. There is a steep climb beginning at $-\$14$ million such that 33% of model realizations result in project losses—that is, 10% more than for Redevelop & Grow.

It is interesting to note that the modeled electricity forecast includes enough volatility to generate cases with both +20% and -20% price deviations (Figure 4-9), potentially (and perhaps unrealistically) triggering both a capacity expansion and reduction within the same 30-year project lifespan. In addition, since modeled prices increase over time, except briefly when the randomly-placed step-change is negative, lost revenue from the utility forcing PPA renegotiations if electricity pricing went into a multi-year decline (e.g., Low Renewable Technology Cost case in Figure 4-5) is not simulated here. These factors may play a role in the dominance of the Redevelop & Grow case over Full Flexibility.

Nevertheless, there is an additional and simple explanation for the model behavior.

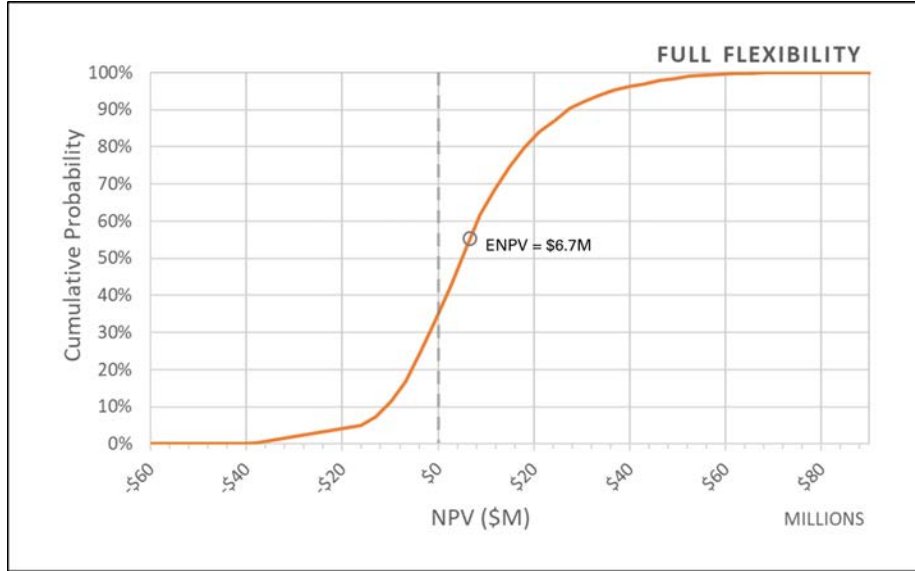


Figure 6-9: Cumulative distribution function for the Full Flexibility Case probabilistic cost model. The curve summarizes results from 2000 model realizations. NPV is reported in \$M, where M is million.

If power-plant modules are decommissioned, the model is more sensitive to the drop in revenue from reduced electricity generation than any benefit realized by discounted OPEX. The cost savings is just not significant enough to make up for less income.

6.6.2 Summary Statistics

Table 6.5 outlines key statistical measures summarizing the performance of the Full Flexibility Case probabilistic cost model.

Full Flexibility Case Statistics	$\mathcal{R} = 2000$
ENPV (\$M)	6.7
STD(NPV) (\$M)	16.0
P05 NPV (\$M)	−16.3
P50 NPV (\$M)	4.9
P95 NPV (\$M)	36.2
% Difference from NPV_s	545%

Table 6.5: Full Flexibility case model statistics for 2000 model realizations. NPV_s refers to the static model NPV reported in Section 6.1.3.

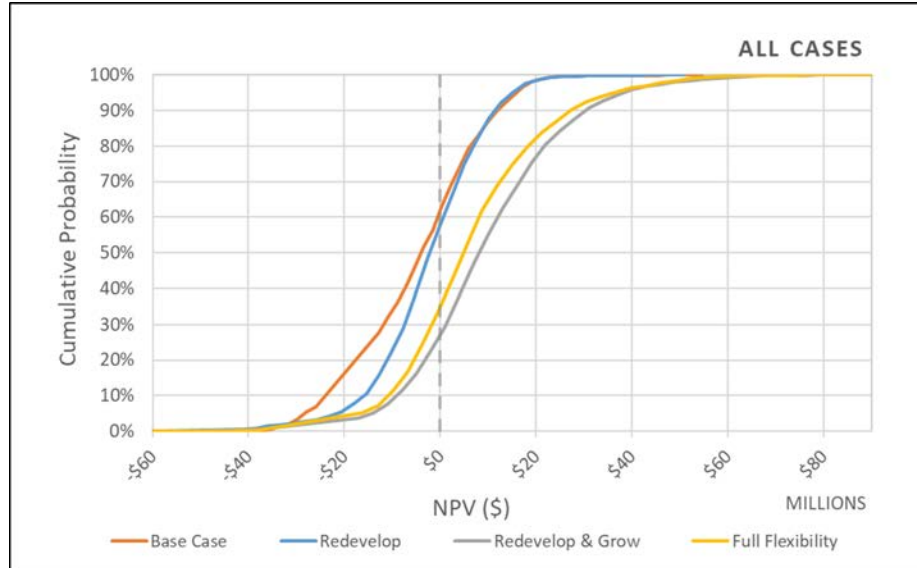


Figure 6-10: Cumulative distribution functions for all probabilistic cases evaluated in Sections 6.3 to 6.6. Each curve summarizes results from 2000 model realizations.

6.7 Combined Model Comparison

Figure 6-10 illustrates the target curves for the four probabilistic models discussed in the previous sections. The Base Case model is dominated by all other models on the low-side and merges with the Redevelop Case on the high-side. The Full Flexibility Case greatly improves on the Base and Redevelop Cases. Redevelop & Grow appears farthest to the right but with the smallest slope, indicating it has the highest standard deviation and hence least confidence. Nevertheless, Redevelop & Grow exceeds all other cases on the high-side, low-side, and expected value, and thus would be the recommended operational strategy based on these model results.

6.8 Full Flexibility Case Sensitivity Testing

In order to explore the hypothesis that the Full Flexibility Case model is responding to revenue losses, a series of sensitivity tests considered adjustments to the *Reduction amount* parameter controlling how many modules are removed by the Capacity Reduction decision rule (Table 4.4, Section 4.4.3). The default value applied throughout this analysis has been 25% to match the *Expansion amount* parameter used by the

Capacity Growth decision rule (Table 4.4, Section 4.4.2). Sensitivity testing with the *Expansion amount* parameter is also conducted.

6.8.1 Reduction Amount

The gap between the Full Flexibility (yellow) and Redevelop & Grow (gray) curves broadens for higher *Reduction amount* values. However, the curves begin to overlap for *Reduction amount* values of $\leq 15\%$. The difference is subtle, but the 5% case shows a slight rightward step-out in the high-side values for Full Flexibility compared to the Redevelop & Grow. This is not seen in the 4% or 6% scenarios, suggesting 5% is the sweet spot for *Reduction amount* when a significant electricity price drop is detected. For values above 5%, the gain from OPEX reductions cannot overcome revenue lost from reduced capacity, which likely reflects the influence of PPA contracts that stabilize the year-to-year price paid for generated electricity. For values less than 5%, the Full Flexibility Case merges with the Redevelop & Grow Case as the number of decommissioned modules trends toward zero.

6.8.2 Expansion Amount

A similar sensitivity test is executed by varying the *Expansion amount* parameter while holding *Reduction amount* at 5%. Both the Full Flexibility and Redevelop & Grow target curves respond to this parameter since both cases include the Capacity Growth decision rule (Section 4.4.2). Figure 6-12 illustrates the results for step-wise increases to *Expansion amount* by an increment of 20%.

Full Flexibility and Redevelop & Grow target curves closely match each other except in the high-side model realizations, i.e., those above \$20 million NPV, and in the low side near $-\$20$ million NPV. The low-side variability relates to the sparse count of model realizations at $-\$20$ million NPV in each Monte Carlo simulation. For the better-constrained high side, Full Flexibility shows a slight advantage when *Expansion amount* is 25% and 65%. However, the separation is more pronounced for a 45% *Expansion amount*. In this sweet-spot case, Full Flexibility ENPV amounts

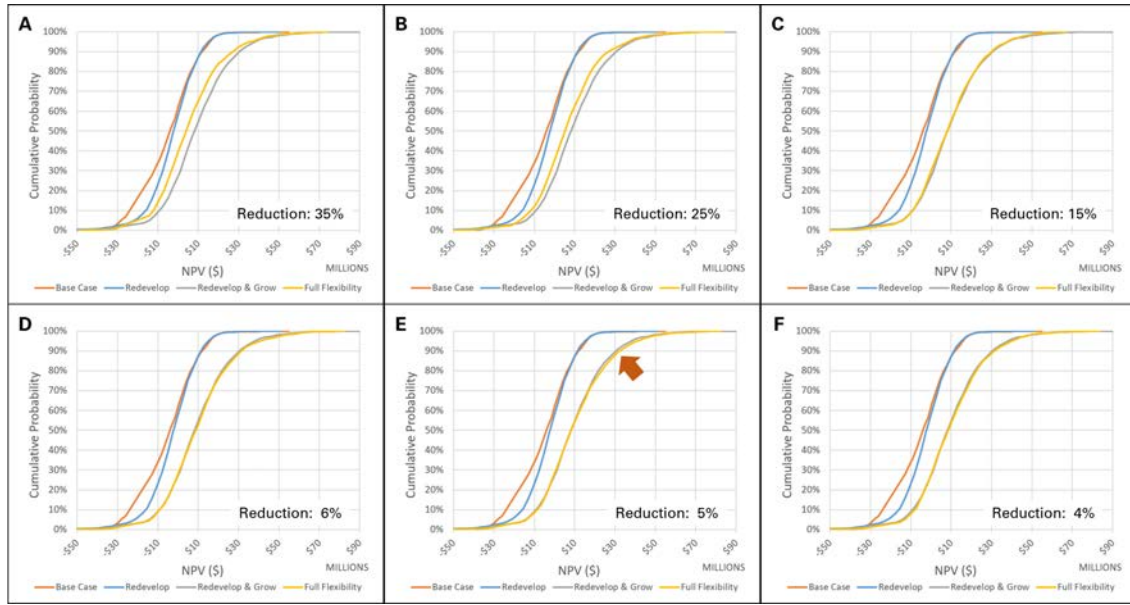


Figure 6-11: *Reduction amount* sensitivity test using model target curves. Each curve illustrates results for a Monte Carlo simulation with 2000 runs of the indicated models. *Reduction amount* only impacts Full Flexibility Case (yellow), so Base Case (orange), Redevelop Case (blue), and Redevelop & Grow Case (gray) remain unchanged. *Reduction amount values* include A. 35%, B. 25%, C. 15%, D. 6%, E. 5%, and F. 4%. NPV is reported in \$M, where M is million.

to \$13.7 million, $\approx$ \$1 million greater than that for Redevelop & Grow. This optimized potential would be missed without testing and refining the project growth and reduction strategies through sensitivity testing and parameter tuning.

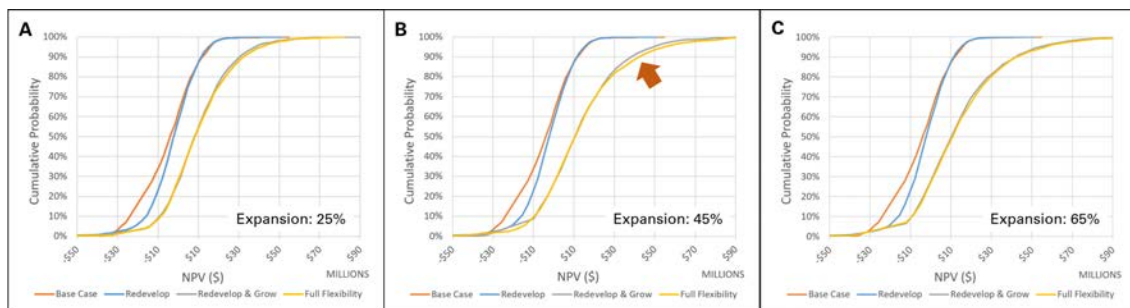


Figure 6-12: *Expansion amount* sensitivity test using model target curves. Each curve illustrates results for a Monte Carlo simulation with 2000 runs of the indicated models. *Expansion amount* impacts both Redevelop & Grow (gray) and Full Flexibility (yellow) target curves. Base Case (orange) and Redevelop Case (blue) remain unchanged. *Expansion amount values* include A. 25%, B. 45%, and C. 65%. NPV is reported in \$M, where M is million.

6.9 Recap

This chapter covered the execution of cost models described in Chapter 4. Results for a deterministic model and four probabilistic models illustrate how incorporating uncertainties and decision rules into a modeling approach can provide important insights into potential project profit and loss. By comparing summary metrics and target curves, this methodology allows a decision-maker to test strategies for mitigating the risk of an unprofitable EGS expansion to an existing power plant.

Key insights from cost modeling include:

1. Static (deterministic) cost models that use average or most-likely parameter values are inherently flawed and limited. They distort estimates of project value and may lead to poor project decisions based on overconfidence stemming from inaccurate raw NPV numbers.
2. Probabilistic model results from Monte Carlo simulation are best compared using target curves (CDFs) and summary statistics like distribution robustness, ENPV, Value at Risk, and Value at Gain in a multi-dimensional analysis. Collectively, these measures can reveal operational strategies that simultaneously protect against project losses and offer opportunities for project gains that may otherwise not be realized.
3. Redeveloping geothermal wells to counter reservoir thermal decline reduces downside risk, but does not improve upside potential for the project.
4. Installing additional power-plant modules in response to greater demand results in significantly greater project value with reduced downside risk.
5. Decommissioning power-plant modules in response to plummeting prices generally does not capture greater value, likely because the lost revenue from reduced power generation outpaces savings in OPEX, at least in this model formulation.
6. Sensitivity testing of decision-rule parameters is a means of optimizing operational strategy and can reveal hidden project value.

Chapter 7

Discussion

Risk acts as a significant barrier to the adoption of geothermal as part of a larger energy portfolio for commercial oil & gas companies. Here, the word *risk* refers to the potential for shortfalls in performance with respect to established requirements, formally defined by the product of probability of occurrence and consequence of failure (Malone & Moses, 2004; NASA, 2017). Companies considering investments in geothermal want to minimize risk exposure, so strategies to mitigate this risk will naturally act as enablers to geothermal growth during the ongoing energy transition.

7.1 Field Lifecycle

Maturing a geothermal asset from initial concept through site decommissioning represents a complex project lifecycle spanning up to several decades in length. Figure 7-1 illustrates the decomposition of a geothermal field lifecycle into a level 1 process flow that mimics that of a typical hydrocarbon field. The level 2 decomposition describes a work breakdown structure, each step with its own inherent risks. Here, the primary play risk elements for geothermal introduced in Section 2.1.3 have been reframed as four components: heat, permeability, fluids, and seal. Each appear in both the Exploration and Appraisal phases of the project.

The red dotted outline in Figure 7-1 illustrates activities in the Exploration and Appraisal stages, where machine-learning methods described in Chapters 3 and 5 can

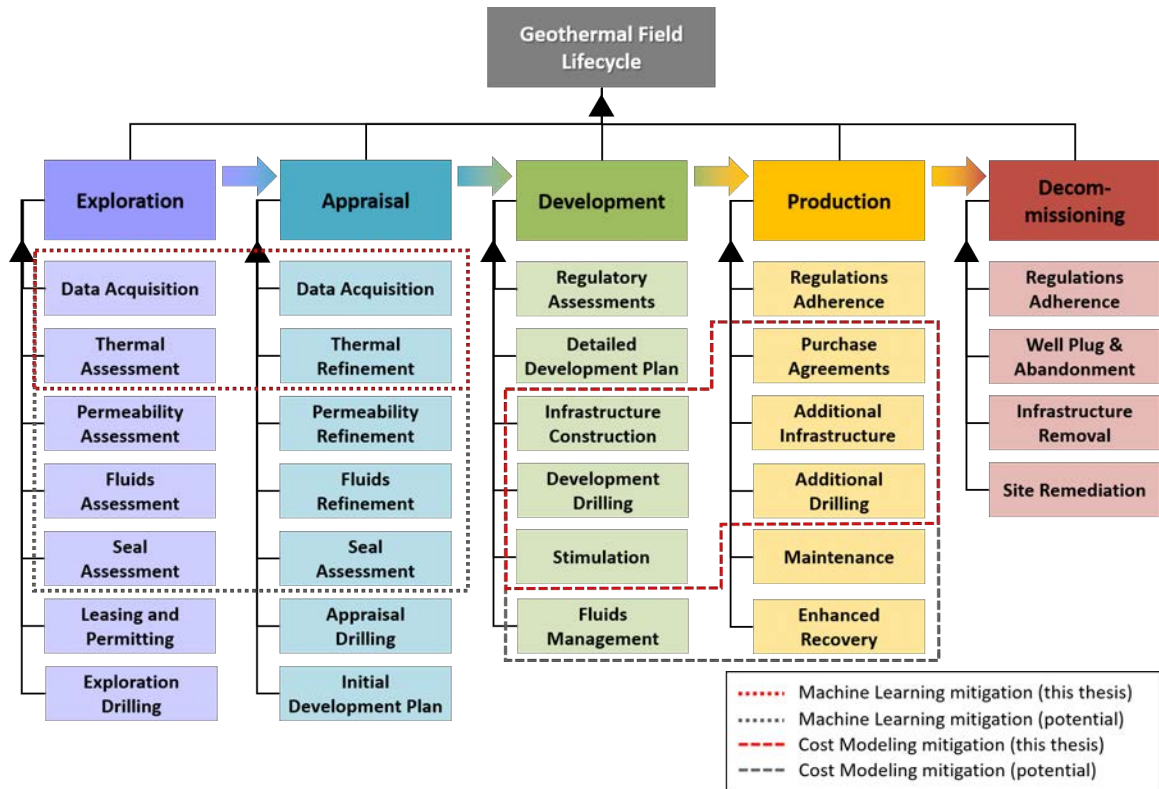


Figure 7-1: Proposed geothermal field lifecycle, and two levels of decomposition, defining major project phases and a high-level WBS. Dotted lines indicate where machine-learning methods could mitigate risk in exploration and appraisal. Dashed lines depict where cost models might mitigate risk during the development and production phases.

reduce the overall risk profile. Geothermal exploration commonly focuses on areas where data and known resources are already present. Reviewing available data to identify feature relationships suggestive of favorable locations is a well-established pre-screening activity for mitigating the risk of costly exploration failures (Doughty et al., 2018). Machine-learning algorithms described in Chapter 5 provide data-driven methods for uncovering these complex feature relationships, and generating resource favorability maps, rapidly and at low cost. Furthermore, feature importances derived from recursive feature elimination (Section 5.1.2), impurity measures like Gini index or entropy (Section 3.3.4), or Shapley attribution analysis (Section 3.3.5) directly rank different data sources by their value for predictive modeling. These measures could also guide exploration and appraisal spending on additional data purchases or acquisition efforts. For example, recognizing that silica geothermometer temperatures, heat-flow measurements, crustal thickness, and density of volcanic dikes and springs all highly influence the geothermal gradient classification model (see Section 5.3.2), an exploration team could focus time and budget on 1) field surveys for silica concentration sampling, 2) field or remote-sensing mapping of springs and dikes, and 3) seismic acquisition for improved crustal-thickness estimates where those estimates are most uncertain. As suggested by the black dotted line, the machine-learning techniques applied to assess geothermal heat content in Chapter 5 are transferable to assessments for the other risk elements.

Cost modeling similarly offers benefits for risk mitigation in the geothermal project lifecycle, as illustrated with the dashed lines in Figure 7-1. Surface plant construction and drilling activities take place during the Development phase and continue into Production, as thermal decline or market forces trigger field-management responses. Rather than treat the extent of these activities as known unknowns, characterizing and including them in flexible economic models offers the opportunity to assess their impact and test different scenarios for field-strategy optimization. As the analysis in Chapter 6 showed, models can include local uncertainties, e.g., geothermal gradient or decline rate, as well as broader risks like a carbon tax or national electrification. The red long-dashed line in Figure 7-1 surrounds factors considered by the cost model

in Chapters 4 and 6. The gray dashed line surrounds additional aspects of the Development and Production phases that could also be characterized in a cost model with distributions and decision rules to determine project viability or refine field strategy.

7.2 Role of Uncertainty

Uncertainty exists, as does the opportunity to include it in a larger decision-making process for geothermal adoption. In the exploration phase, feature standard errors and maps of entropy—or another measure of collective uncertainty—can influence project choices. Observing pervasively high standard errors for a data layer (see Section 3.4.4) raises the question of whether those data should be re-acquired using different tools or survey methods, or if better-quality data might be available in-house or for purchase. And zones of high entropy in measurement uncertainty highlight areas that need additional attention. Is the entropy a result of insufficient data to train machine learning models, leading to poor discrimination ability for the predicted classes? Or are the data in areas of high entropy simply inconclusive or poorly conditioned? Pursuing these questions helps frame a refined project plan for the early phases of the field lifecycle. In this scenario, time and resource allocations to data science and data engineering (using existing data), field studies and data acquisition (supplementing existing data), or exploration of more certain areas, are all expressly driven by the data.

If an ensemble of models is considered, structural or model uncertainty (see Section 3.4.2) can direct efforts on how to approach machine-learning prediction. For example, when entropy appears high throughout the area of interest (AOI), and most of the models show reasonable agreement except for one or two, this could justify down-selecting those models and re-evaluating from the reduced ensemble. But if high variance is observed across many models, this might indicate that the input data insufficiently describe the system. Likewise, predictions from probabilistic models that show high parameter uncertainty in the AOI (Section 3.4.3) should be treated as suspect, and the model redesigned or retrained on a larger data set. Insights like these

mitigate the risk of over-confidence in under-performing models, potentially avoiding a poorly-placed exploration well further down the line, based on those models.

For cost models, directly incorporating uncertainty serves to counter the classic misconception that applying average values to elements of a complex system will lead to an average system result (Flaw of Averages, [de Neufville & Scholtes, 2011](#), p. 17–19). Instead, using variable probability distributions and generating expected values from multiple model realizations provides reliable median estimates and describes the spread of potential results. As discussed in Section 6.2, target curves and percentiles defining Value at Risk and Value at Gain offer a richer set of metrics for model comparison. Using these metrics can reveal the combination of strategic choices for a geothermal project timeline and execution, choices that mitigate the risk of project losses and target the greatest upside opportunity.

7.3 Risk Analysis

One approach to project risk assessment evaluates risks and mitigation actions with a scorecard tracking likelihood (Table 7.1) and consequence (Table 7.2) of individual risks before and after actions are taken ([Malone & Moses, 2004](#)). The process begins with creating a risk log as shown in Table 7.3. This table can be constructed through a variety of risk identification methods, including formal hazard analysis, models and simulations, or group brainstorming ([NASA, 2017](#)). Even the act of populating this table adds value to a project by aligning the deciding group in regard to judgment and assumptions about risk relevance and potential impact.

Likelihood		
Score	Probability (p) of Occurrence	
5	Near certainty	(0.8,1.0]
4	Highly likely	(0.6,0.8]
3	Likely	(0.4,0.6]
2	Low likelihood	(0.2,0.4]
1	Not likely	[0.0,0.2]

Table 7.1: Likelihood score based on probability (p) of occurrence. Adapted from NASA S3001 Guidelines for Risk Management, v.G ([Malone & Moses, 2004](#)).

	Consequence				
	1	2	3	4	5
Performance	Minimal to no impact on meeting project goals.	Minor impact on meeting project goals.	Unable to meet a specific project goal, but remaining goals can be achieved.	Unable to meet multiple project goals, but minimum project success still achievable.	Unable to meet multiple project goals, minimum project success not achievable.
Schedule	Minimal to no impact on project schedule.	No change to critical milestones; at least 10-day buffer between any delay and impact on milestone timing.	No change to critical milestones; less than 10-day buffer between any delay and impact on milestone timing.	One or more critical milestones slip, impacting overall project schedule.	One or more critical milestones slip; one or more milestones cannot be achieved.
Cost	Minimal to no impact on cost.	Minor impact on cost. Deviation < 5% of total approved budget.	Impact on cost. Deviation > 5% but $\leq$ 10% of total approved budget.	Impact on cost. Deviation > 10% but $\leq$ 15% of total approved budget.	Major impact on cost. Deviation > 15% of total approved budget.

Table 7.2: Consequence score defining the level of impact a risk might have if it becomes a reality. Adapted from NASA S3001 Guidelines for Risk Management, v.G (Malone & Moses, 2004).

Having captured risks and ordered them by their risk score, the project team next defines mitigation plans for addressing those risks (e.g., Table 7.4). The resulting *mitigated* risks are analyzed and assigned updated scores for likelihood and consequence, which also quantify the remaining risk. Different mitigation options for the same risk can be directly compared by the final risk scores or individual likelihood or consequence scores if addressing one over the other is preferred.

The risk scorecard methodology tabulates risks before and after mitigation in a 5×5 matrix, where high-likelihood, high-consequence risks fall near the upper-right, and the lower-left represents the ideal low-likelihood, minimal-consequence region. The risk matrix serves two main functions: prioritization and selection. First, it quickly differentiates between risks by using assigned risk priorities for each matrix cell. Following NASA guidelines, priority values across the matrix increase toward the upper-right, are non-symmetric, and skew higher for high-consequence cells (NASA,

ID	Description	Likelihood	Consequence	Risk
EXP1	Insufficient exploration budget	4	3	12
EXP2	Poor subsurface characterization	4	4	16
EXP3	Permitting delays	3	3	9
DRL1	Rig unavailability	4	3	12
DRL2	Drilling cost overruns	3	3	9
DRL4	Downhole equipment failure	3	2	6
PRD1	Insufficient production budget	3	4	12
PRD4	Rapid thermal decline	3	4	12
PRD5	Demand variability	2	4	8
PRD6	Wrong-sized infrastructure	3	4	12

Table 7.3: List of geothermal project risks, each assigned a likelihood of occurrence (see Table 7.1) and consequence (see Table 7.2). Risk is likelihood $\times$ consequence. This list is non-exhaustive and intended to support discussion of risk matrix use.

2017). Secondly, different mitigation strategies for the same risk can be evaluated and plotted in the same 5×5 . This helps quickly communicate ranked outcomes of different strategies, and supports rapid decision-making on which to select.

Although simple to perform, quantifying risk as the product of ranked values can introduce distortion into the final risk numbers. One potential area for improvement would be to instead characterize consequence as a physical value, e.g., a cost estimate, before multiplying it against likelihood. To do so, consequences associated with schedule or performance impacts must be translated into their value equivalents. This translation would require a set of analogous projects for calibration and an evergreen process for validation using present and future project data. Oil & gas companies typically rely on analog databases like those provided by Wood Mackenzie (Wood Mackenzie, 2019) and IHS (IHS, 2021) during risk evaluations, so this adjustment to the NASA method seems achievable.

Figure 7-2 illustrates a risk matrix based on the highlighted rows from Table 7.4. Arrows map the original risk to mitigated risk. Each of these examples utilizes mitigation strategies described in the earlier chapters of this thesis. Appendix C provides additional detail on the choices of likelihood and consequence scores for these risks, both before and after applying the proposed strategies. The scores are also listed in Tables 7.3 and 7.4. Although many other geothermal-project risks exist

ID	Mitigated Likelihood	Mitigated Consequence	Mitigated Risk	Risk Reduction	Mitigation Action
EXP1	2	2	4	8	Use ML model predictions to reduce costs in exploration
EXP2	2	3	6	10	Use ML for baseline model, acquire data based on importances
EXP3	2	3	6	3	Engage with regulators early to fast-track permitting
DRL1	3	3	9	3	Secure rig early and pre-plan for future drilling needs
DRL2	2	3	6	3	Use ML to identify high-gradient areas faster, shallower drilling
DRL4	2	2	4	2	Use service companies with high-Temp equipment track record
PRD1	1	3	3	9	Cost modeling for optimized spending and return
PRD4	3	2	6	6	Regularly re-drill wells or re-stimulate reservoir
PRD5	2	3	6	2	Flexibility in power generation based on market
PRD6	1	3	3	9	Flexible design for demand-triggered capacity changes

Table 7.4: Proposed mitigations for risks listed in Table 7.3 and the change in risk associated with those mitigations. Risk values are likelihood $\times$ consequence. Shaded mitigations include one of the options described in this thesis and appear in Figure 7-2. Adapted from NASA S3001 Guidelines for Risk Management, v.G (Malone & Moses, 2004).

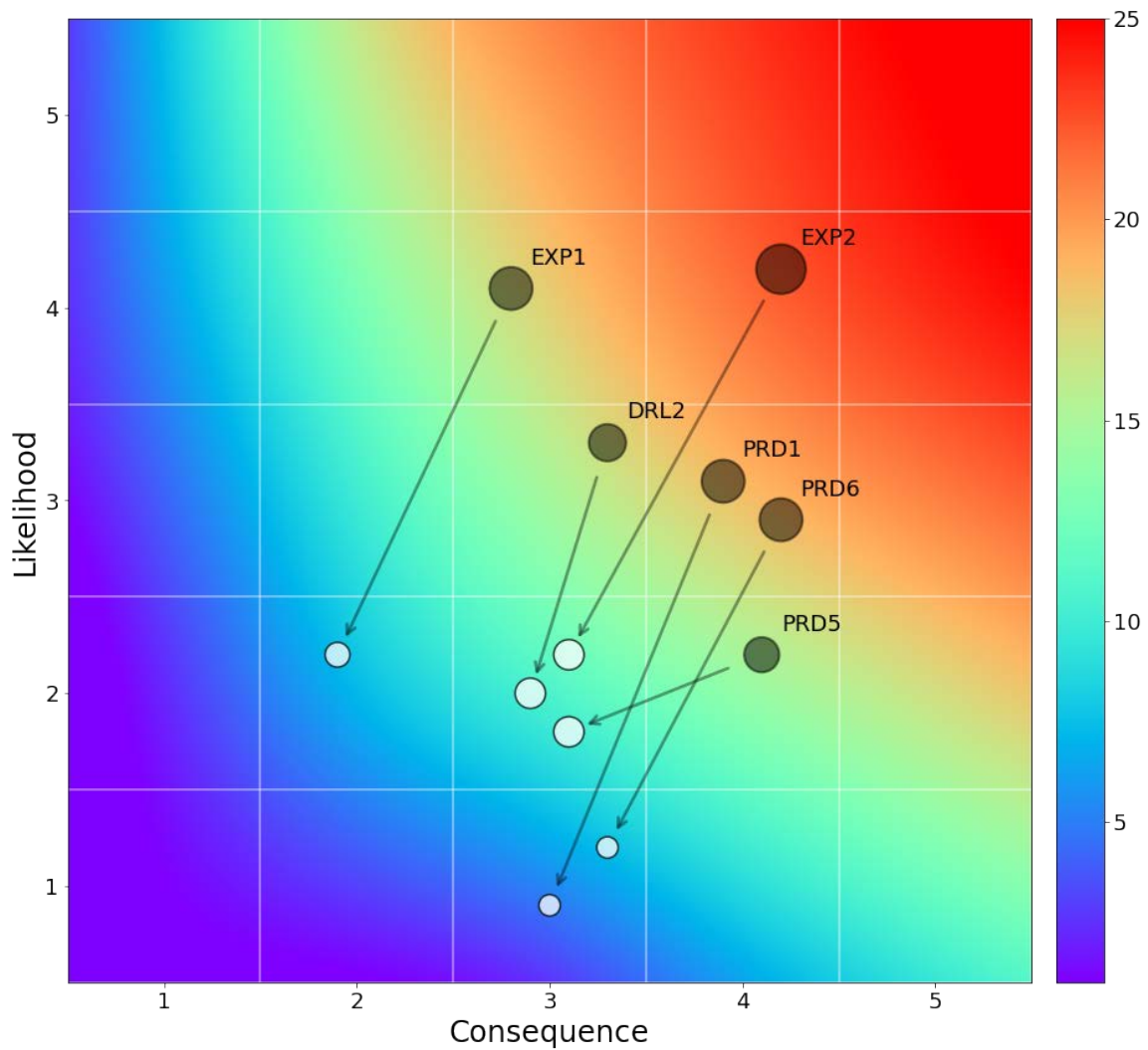


Figure 7-2: Risk matrix for categorizing and prioritizing project risks (dark gray) and charting risk-mitigation strategies (white). Arrows point from the original risk to the risk after performing a specific mitigation action. Marker size corresponds with risk value (likelihood $\times$ consequence). Markers are spread out within each cell for visualization purposes. Risk labels match those in Table 7.3. Background color map depicts risk priority ranging from 1 to 25.

than those described here, what this work demonstrates is the ability *to* reduce risk *by* rapidly applying low-cost assessments throughout the lifecycle of a geothermal field *using* readily-available data. Emphasis is on the to-by-using construct of a classic system problem statement.

At the Exploration stage, machine-learning PFA assessments feed directly into risking processes already common in oil & gas companies (Nash & Bennett, 2015). This reduces barriers around deployment and adoption, and the risk reduction in Figure 7-2 clearly demonstrates the potential value gained. In the Development and Production stages, use of scenario-testing with flexible cost models mitigates risk while possibly revealing new, more profitable means of operating a geothermal field. And doing so with spreadsheet-based tools means petroleum-industry project managers already have the technology and capability to apply these models for decision-support and risk-mitigation today.

7.4 Great Opportunity

Oil & gas companies are uniquely positioned to embrace geothermal, including EGS, as a low-carbon baseload option in the energy transition. Consider some present-day barriers to broad commercial development of EGS described in the GeoVision study:

1. *Companies advancing EGS are small, lacking significant operating capital and the ability to tackle large-scale (> 100 MW) projects* (Doughty et al., 2018). Oil & gas companies have the working capital to take on larger, more expensive projects if those projects have the potential for long-term profitability.
2. *Subsurface characterization with advanced methods like seismic imaging and reservoir modeling requires capabilities not generally found in the geothermal industry* (Doughty et al., 2018). Use of the most advanced geophysical data processing techniques, cutting-edge interpretation platforms, and 3-D earth-modeling methods are standard practice in the petroleum industry.
3. *Geothermal practitioners lack standardization or best practices for drilling com-*

plex wells (Doughty et al., 2018). Oil & gas companies follow standard operating procedures founded on years of experience drilling wells all across the globe, including in environments challenged by extreme depths, complex geology, and high temperatures and pressures.

4. *EGS requires complex subsurface operations like directional drilling and multi-zonal isolation for fracture stimulation that are not typical of traditional hydrothermal operations* (Augustine et al., 2019). Technology improvements and years of expertise from developing unconventional reservoirs have created a wealth of experience in directional drilling and fracking within the oil & gas industry.

Interest in collaboration runs high as well. The U.S. Geothermal Technologies Office regularly promotes cross-over technology transfer between the oil & gas and geothermal communities, funding programs such as GEO out of the University of Texas Austin that specifically target transitioning industry capabilities (Hamm et al., 2021). And the operational efficiencies that companies like Unocal previously brought to geothermal operations in the United States, Philippines, and Indonesia in the past (Melosh, 2017; Palma, 2014) resonate with the needs of today.

Given the promise of change in the energy transition, the cross-sector synergies in fundamental skill sets, and strong interest in partnership from those already committed, geothermal may be the shortest-path solution for building a lower-carbon business portfolio. Getting there requires analysis of the risks, as well as mitigation strategies needed to reduce the likelihood and consequence of those risks, as described in Section 7.3. Based on the results of this thesis, the path from great opportunity to profitability may lie in combining available data and digital technologies to create useful predictive models with uncertainty. The last part is fundamental—uncertainty analysis will define where measurements are meaningful, where models are predictive, and which strategy offers the greatest potential gain. Lessons learned here apply to geothermal exploration and production, but embracing uncertainty for better decision making and risk mitigation will likely benefit every stage of the geothermal lifecycle.

Uncertainty characterization may thus be the key to building greater certainty in the future of geothermal.

Chapter 8

Conclusions

As the transition toward lower-carbon energy solutions continues to progress, geothermal uniquely offers a zero-emissions, continuous source that can fulfill the baseload needs of energy consumers. A geothermal-field lifecycle closely follows that of a hydrocarbon field, from exploration and appraisal, through field development, to production and eventual decommissioning. The expertise with subsurface data, reservoir modeling, complex drilling, and field-management skills valued in oil & gas are similarly of great value to the geothermal industry. However, geothermal without EGS only has limited reach, and EGS comes with high risk. Without clear risk-mitigation strategies, oil & gas will likely discount or delay adding geothermal to their energy portfolios despite the clear synergies between the two domains.

Play fairway analysis, a common tool in the oil & gas industry, has gained traction for reducing risks associated with geothermal exploration. However, PFA requires integration of disparate data sets to define chance of success for geothermal risk elements. This process lacks standardization and requires judgment calls from subject matter experts. Machine learning offers data-driven forecasts that are both quantitative and repeatable, and results from this work shows great promise in the predictive ability of several varieties of machine-learning models. Perhaps more importantly, uncertainty characterization of spatially distributed data delivers invaluable information on where the data should be trusted, where predictive capacity of individual models varies, and where multiple models agree. Furthermore, machine-learning models

determine which data sets provide the most discriminatory value, which can steer exploration decision makers to spend prudently on additional data acquisition activities. Collectively, use of machine learning for play-risking or prospecting reduces the risk of making poor decisions on unreliable favorability maps, allocating budget on low-impact data sets, or missing important signals within the data that make the difference between a productive or an uneconomic geothermal well. And by relying on available data up-front to build these models, machine-learning based risk-mitigation comes with quick results at low cost.

Geothermal cost modeling with existing tools delivers leveled cost estimates from a set of pre-determined resource, surface-plant, field, and financial parameters for the production phase of a field. Yet most of these models do not enable defining input parameters with a distribution of values to capture parameter uncertainty. In addition, the models assume static operational conditions over the lifetime of the field (typically 30 years), which leaves no room for the strategic decision-making that takes place under real-life conditions. This thesis shows economic modeling that includes parameter uncertainties can produce easily-comparable probabilistic distributions as results. Tailoring the model and decision rules to the geothermal field design of interest allows rapid testing of project feasibility and optimization of project actions to limit downside risk while capturing upside potential. Risks are addressed transparently and with quantifiable project impact at little cost in resources or capital.

Mitigating the risks of adopting geothermal energy should be handled systematically through a risk-management process. Working within a geothermal project team, risks are cataloged, assigned likelihood and consequence values, and prioritized based on those values using a risk matrix. Choosing among mitigation plans comes down to comparing the results of executing each proposed plan of action. As shown by the work presented here, mitigation plans that combine available data with digital technologies to create predictive models with uncertainty can significantly reduce the threat of high-consequence geothermal exploration and production risks. Careful uncertainty characterization and evaluation may thus be the key to making geothermal

a commercially-viable, low-carbon investment for oil & gas companies navigating an evolving energy future.

Chapter 9

Future Work

In this chapter, directions of further inquiry are suggested as both extensions of methods already reviewed and new avenues not yet explored by the author. The topic of geothermal energy is a rich one with many great opportunities for research.

9.1 Machine-Learning Applications

This thesis primarily focused on supervised classification models in its survey of machine-learning methods for geothermal exploration. Maintaining a limited scope in modeling – as well as in data-gathering and preparation — was intentional, but doing so set aside a number of topics worthy of follow-up analysis:

- Machine-learning methods in Chapter 5 framed the prediction problem as a classification with four distinct labels for assignment (see Section 3.2.2). Although convenient, this choice is problematic for continuous response variables like geothermal gradient. Classifiers treated each gradient class as independent, with no inherent ordered relationship. This assumption is of course false. Binning, a.k.a., enumerating or quantizing, continuous variables makes sense when training data are sparse, but care must be taken for cases near the bounds between those bins. Further research on applying regression methods would help answer whether similar machine-learning tools can perform well at predicting

those edge-case gradients where even the best classifiers (XGBoost, ANN) appeared to have difficulty.

- Using point data extracted from geothermal feature maps as input to machine-learning models (see Section 3.2.1) was expedient, but it also ignored the spatial correlations inherent to geologic data. The presence of features like faults or volcanic dikes at one location naturally raises the probability of finding the same at a nearby location. This kind of spatial relationship might be better preserved by including map coordinates in the feature set. Alternatively, modeling efforts could apply convolutional neural networks or other more complex architectures that go beyond the fully-connected ANN described in Section 5.4.1.
- Management of data sparsity is a common issue in exploration. Here, the original data was augmented by imputing geothermal gradient values for adjacent pseudo-wells (Section 3.2.2) in an inelegant but effective approach to expanding the training data set. Data imputation via more advanced methods would be a worthwhile topic of study, with the end goal of setting standards to guide future geothermal data engineers. Of particular interest would be identifying methods that minimize spurious correlation imposed on the data as can occur with methods like kriging or k -nearest neighbors imputation.
- Autoencoders, Principle Component Analysis, and Non-negative Matrix Factorization are all effective tools for dimensionality reduction. Evaluating these and other methods can help bridge the gap between the supervised studies in this thesis and unsupervised efforts described in recent literature (Pepin, 2019; Smith et al., 2021; Vesselinov et al., 2020). Of particular interest might be consolidating many features associated with the same geothermal risk element into “super features” before applying tree-based ensemble methods or neural networks for classification.
- Uncertainty estimation for secondary data products, e.g. raster files generated by others without paired standard errors or original primary-source observa-

tions, will be necessary to evaluate measurement uncertainty (see Section 5.5.3) for all modeled features. Until the day that reporting uncertainties becomes an expectation, if not a requirement, there will be the need for clear guidance on deriving error estimates for pre-gridded data.

- The topic of ensemble models was raised in this thesis but not rigorously pursued. Ensemble models apply a model-of-models paradigm to machine learning and show success in other applications (e.g., [O. Wilson, 2020](#)). They naturally tie to Play Fairway Analysis where final favorability maps represent combinatory insights from individual risk-element maps (see Section 2.2.4). An ensemble model approach could streamline the creation of a final geothermal favorability map while also preserving the prediction of specific risk components.

9.2 Cost Modeling

Although a number of economic models for geothermal power production already exist with varying degrees of maturity (see Section 2.4), the contribution of this thesis to incorporating uncertainty and flexibility into cost models merely scratches the surface on what can still be done.

- Using well-known, popular platforms like Microsoft Excel greatly lessens the burden around deployment and adoption of new tools. However, spreadsheets come with limitations, some of which impacted the capabilities of the cost model defined in Section 4.1. Most significantly, the individual power plant modules and injector-producer well couplets were difficult to track and manage as their count dynamically changed throughout the lifespan of a field. One suggested improvement is to expand the existing model with more complex code that treats the modules and wells as objects, each having individually-tracked attributes like age, efficiencies, decline rates, and maintenance records. Uncertainty definitions, decision rules, and the overall user experience can remain Excel-based, but functionality enhancements via an Excel plug-in or VBA code could over-

come calculation limitations.

- Electricity-pricing forecasts and Power Purchase Agreements (PPAs) likely require a more nuanced treatment than conducted in this thesis. Additional research into the best representation of wholesale electricity price changes, including reversing trends, would be a good first step. But perhaps more fundamentally, the relationship between price changes and the likelihood and scope of a PPA update with utility partners must be characterized and modeled.
- Additional research and sensitivity testing on the best distribution functions to use for different variables in geothermal cost models could help improve the reliability of the model presented in this thesis. For example, in the present study, both geothermal gradient and reservoir temperature are represented by uniform distributions with bounds set to the range of values observed at Lightning Dock. Expanding those distributions to capture a broader range of potential (unobserved) values, e.g. with a Gaussian distribution, would be a good first refinement.
- The Electrification Futures Study ([Murphy et al., 2021](#)) and the SIPA study on carbon taxation ([J. Larson et al., 2018](#)) both note complexities of modeling renewable-energy demand when the growth and impact of the natural-gas market remains uncertain. Furthermore, the role of targeted subsidies for technologies like wind energy further disrupt an already tilted playing field (see [Lazard, 2020](#)). As cost models for geothermal continue to be refined, the dynamics associated with the natural-gas market and incentives (including subsidies) for other competing renewables must be incorporated into an overall demand equation. The latter can help determine field-expansion strategies and influence the agreed-upon pricing for PPA updates.
- A unique feature of the cost model described in this thesis is the treatment of power-plant installation and expansion as modular in nature. This concept builds on existing technology, but the companies leading the way with that

technology treat aspects of its performance and their financial terms of service as trade secrets. More accurate data regarding module up-front costs, performance limits for higher-temperature production, and fee structures for leasing, expansion, and decommissioning would all improve the existing model.

- Other opportunities for enhancing the geothermal cost model include: split pricing for electricity sales in addition to the original PPA with a utility; identifying and capturing efficiencies (e.g., lower fluid costs) from the expansion nature of the plant being modeled; incorporating local sales to nearby businesses or towns as separate revenue streams; treating the model as a facility portfolio instead of a single location; and investigating how hybrid power (paired solar, wind) and storage (battery) project options influence the bottom line.

9.3 Related Research Topics

- Throughout the present study, machine learning and cost-modeling approaches are treated as separate opportunities. The combination of the two in hybrid methods defines an additional opportunity for managing risk in a geothermal project. In early project phases, a tradespace methodology balancing benefit from machine-learning feature attribution analysis with costs modeled for data acquisition, processing, and interpretation tasks can help optimize exploration activities. And economic models applied throughout the geothermal lifecycle can incorporate machine learning elements for predicting key inputs (e.g., price or demand forecasts) or automating the search across different strategies to determine truly-optimal recommendations.
- The original scope of this thesis included a section on evaluating existing and future drilling technologies. One possible roadmap for this research effort follows a systems approach. System architectural analyses of innovative techniques like millimeter-wave drilling ([Woskov, 2017](#)) or spallation drilling ([Augustine, 2009](#)) could be compared to expected improvements to traditional rotary drilling

([Lowry, Finger, et al., 2017](#)). Feedback from geothermal stakeholders would calibrate the benefit side of a cost-benefit analysis that includes constructing a tradespace to select preferred drilling architectures. As an added bonus, detailed cost estimates derived in the process could help reduce drilling-cost uncertainty for geothermal cost models.

- Machine-learning methods used to generate map-based assessments of favorability lack the specificity of a 3-D subsurface model. Many of the geothermal data features described in Chapter 3.2 refer to surface observations. However, geophysical techniques (seismic, resistivity, gravity, magnetic, MT measurements) and interpretation products (faults, horizons, rock properties) typically extend into the third (i.e., depth) dimension. Research into transitioning a 2-D PFA into a 2.5-D depth-slice or full 3-D analysis would help bridge the gap between the needs of mapping plays and those of full prospect characterization and detailed field planning.
- Significant efforts have already been made in gathering geothermal-related data for access through the NREL OpenEI ([Hallett, 2010](#)) or Geothermal Prospector ([NREL, 2021a](#)) data portals. With more data available, including those from large-scale integrated EGS projects like Utah FORGE ([J. Moore et al., 2019](#)), research on effective reservoir-characterization methods, early field operations, and sustaining field management will serve to increase efficiencies across the geothermal lifecycle. This research will be a fundamental factor toward broadly commercializing EGS.

Appendix A

Exploration Data Layers

This appendix steps through each of the data layers (or “features”) used as predictor and response variables for the machine-learning approach to exploration risk mitigation described in Chapters 3 and 5. For each feature, the source of the data, specific conditioning steps applied to the data, and an image of the final feature map are provided. Refer to Chapter 3 for details on methods and routines mentioned by name in the descriptions below.

A.1 Average Air Temperature

The University of Oregon PRISM Climate Group maintains regularly updated spatial data sets of climate-related observations, including 30-year normals describing average annual conditions (Daly et al., 2008; PRISM, 2021). The 800-m resolution average air-temperature grid was downloaded and imported into ArcGIS, then cropped using the Regional Polygon (Figure A-1). This layer required no further processing.

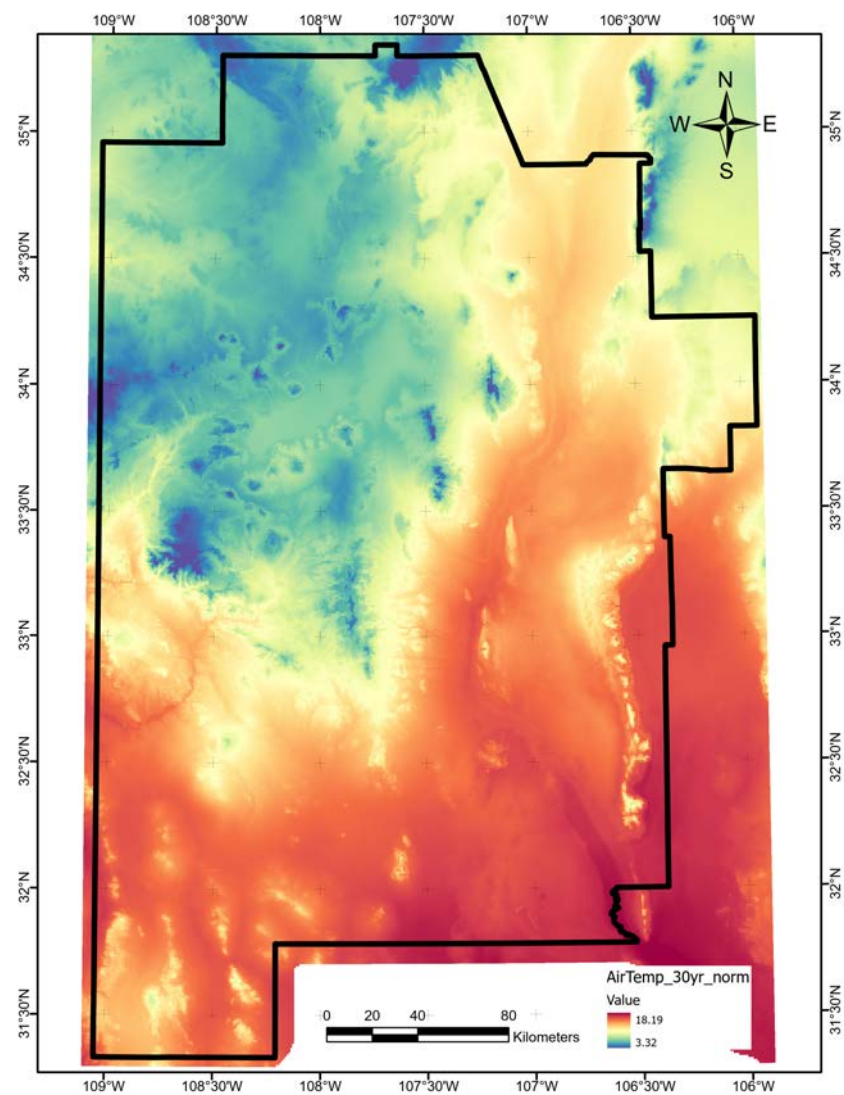


Figure A-1: Average air temperature data layer. Units are °C. Data were retrieved from the PRISM website ([PRISM, 2021](https://prism.climatecentral.com/)).

A.2 Average Precipitation

The University of Oregon PRISM Climate Group also compiles 30-year normals for precipitation ([Daly et al., 2008](#); [PRISM, 2021](#)). The 800-m resolution average-precipitation grid was downloaded and imported into ArcGIS, then cropped to the Regional Polygon boundary (Figure A-2). This layer required no further processing.

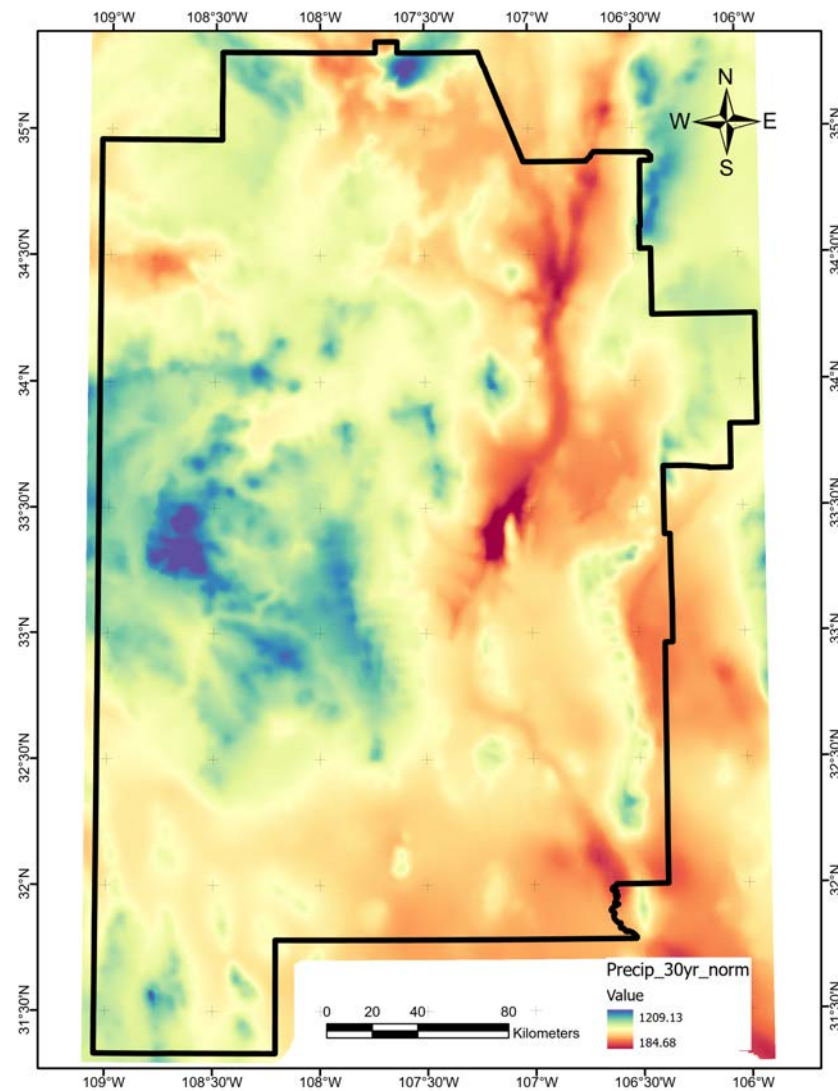


Figure A-2: Average-precipitation data layer. Units are millimeters. Data were retrieved from the PRISM website ([PRISM, 2021](#)).

A.3 Basement Depth

Following the procedure of [Pepin \(2019\)](#), the basement-elevation raster generated by [Bielicki et al. \(2015\)](#) was downloaded, imported into ArcGIS, and processed to calculate depth to basement. Specifically, a unit conversion from feet to meters was applied to the basement elevation surface. Then, values were extracted using the AOI mesh grid (see [3.2.1](#)), which highlighted missing data patches in the data. The ArcGIS *Kriging* function interpolated values across these patches using a spherical semivariogram model, auto-determined lag size of 0.097 degrees, and a variable search radius with a 4-point requirement. Basement depths were then calculated by subtracting the interpolated elevation layer from the Surface Topography (DEM) layer. However, the higher resolution of the DEM layer caused an imprint of detailed surface morphologies to appear on the calculated basement depth layer. To correct for this, the DEM layer was low-pass filtered using the ArcGIS *Filter* method, which averages a 3×3 neighborhood around each point in the data set. The final basement elevation layer was generated from the difference between the filtered DEM and the interpolated basement elevation layers (Figure [A-3](#)).

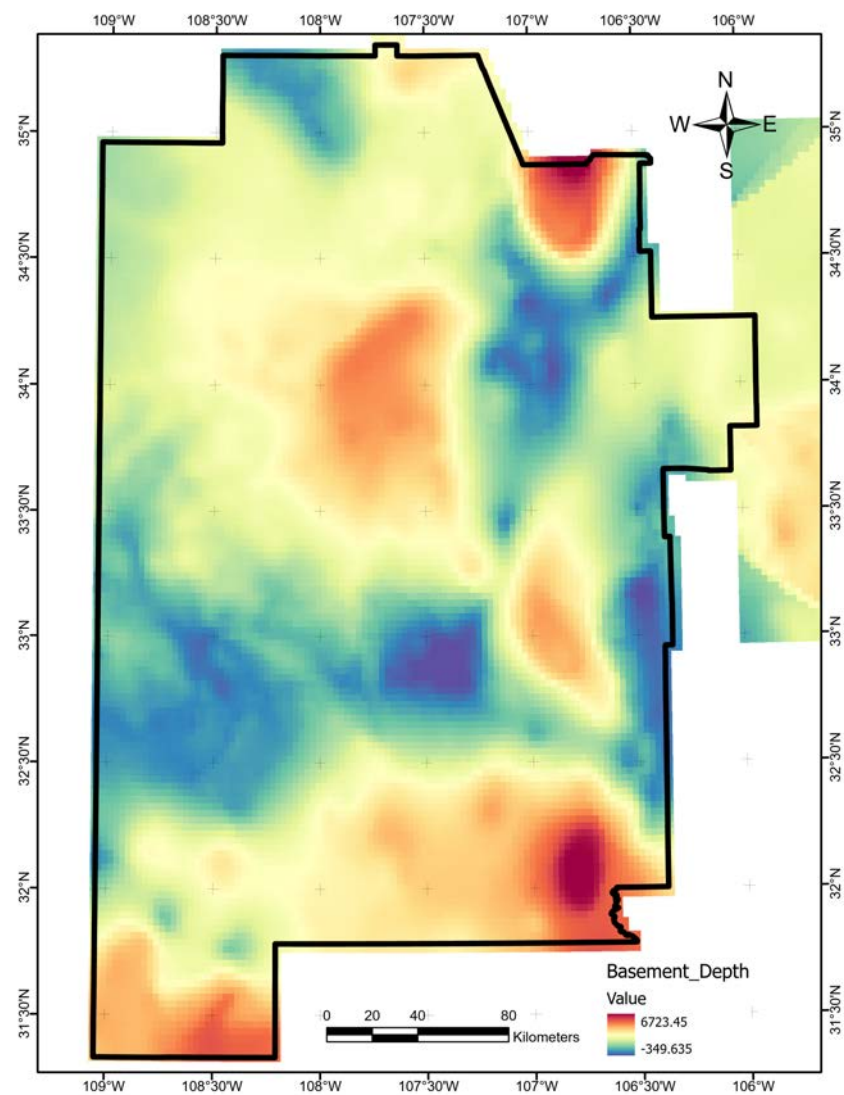


Figure A-3: Basement-depth data layer. Units are meters. Layer is based on basement-elevation raster from [Bielicki et al. \(2015\)](#).

A.4 Boron Concentration

Measurements of boron concentration were assembled by [Bielicki et al. \(2015\)](#) from USGS records, student dissertations, and other sources. These data were downloaded from the NM PFA OpenEI submission ([Kelley, 2015](#)) and merged using ArcGIS and Python to create a single dataframe of 5,686 measurements within the broader Regional Polygon bounds. Restricting data input to the tighter AOI bounds led to artifacts in the interpolation process. Uneven spatial distribution of the data, and sometimes significant variation among overlapping values from different measurement years, created a unique challenge for building a GIS layer. An initial attempt to fit and interpolate the data using tuned Gaussian Process models made feature layers with too much local structure and little character away from the input data points. The ArcGIS *Empirical Bayes Kriging* (EBK) routine was selected instead due to its ability to manage coincident data and its high accuracy with smaller data sets compared to ordinary kriging methods ([ESRI, 2021a](#)). For the final layer, EBK was applied with the Empirical data transformation, a maximum of 100 points in each local model, 100 simulated semivariograms with K-Bessel model type, a Standard Circular search neighborhood, and output cell size of 0.01 degrees. Of important note: the calculation option to include all coincident data were selected, so overlapping measurements were considered in generating the final layer (Figure [A-4](#)).

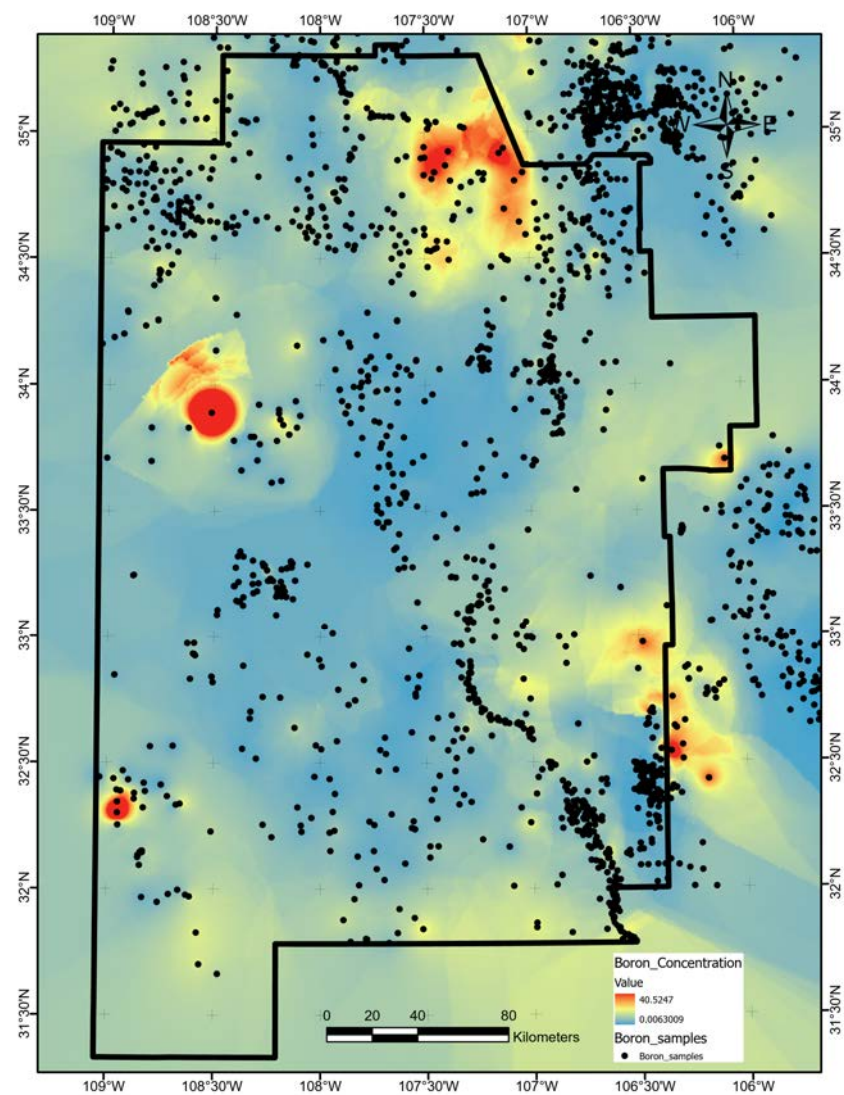


Figure A-4: Boron-concentration data layer. Units are mg/L. Black dots indicate sample locations in the complete data set compiled by [Bielicki et al. \(2015\)](#).

A.5 Crustal Thickness

In the absence of a more recent seismic study constraining variations in crustal thickness across the study area, the regional map published by [Keller et al. \(1991\)](#) was used to construct the crustal-thickness feature layer. Similar to the procedure described by [Pepin \(2019\)](#), the Keller map was georeferenced in ArcGIS, and thickness contours were manually digitized as polylines. These polylines continued slightly beyond the AOI boundary to ensure proper constraints for surface creation without artifacts near the AOI edges. The ArcGIS function *Feature to 3D by Attribute* converted the polylines into 3-D features, and *Topo to Raster* interpolated these features (contours) into a continuous final grid. Since the Keller map was derived from low-resolution seismic lines from the 1960s-1980s, the result is a very low-frequency approximation for crustal-thickness variations associated with the Colorado Plateau and Rio Grande Rift provinces. As such, a slightly larger cell size of 0.025 degrees was used than for other layers. Additional parameter choices included: margin in the cells of 20, smallest z value for interpolation of 25 km, largest z value for the interpolation of 55 km, drainage enforcement, and maximum iterations of 20. The final layer is shown in Figure [A-5](#).

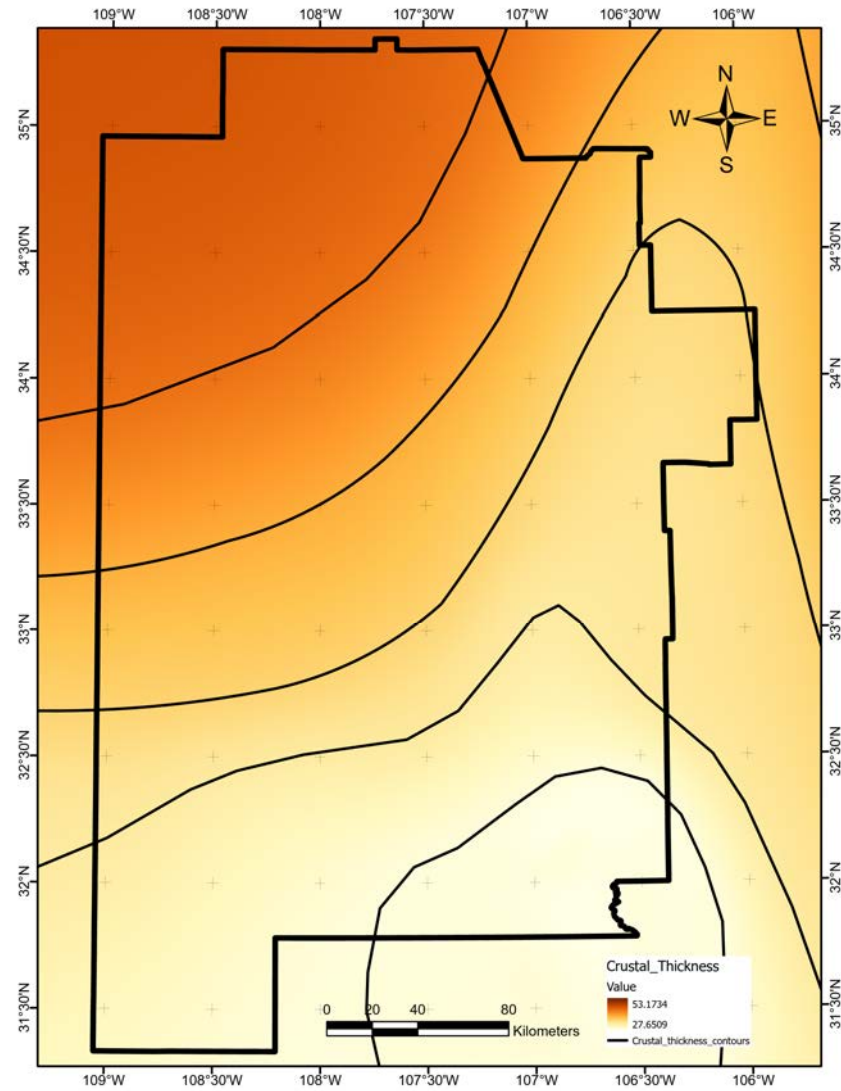


Figure A-5: Crustal-thickness data layer. Units are kilometers. Black lines trace the contours digitized from [Keller et al. \(Figure 4, 1991\)](#).

A.6 Drainage Density

Drainage polyline data come from the [Bielicki et al. \(2015\)](#) PFA OpenEI submission ([Kelley, 2015](#)). The data were downloaded and imported into ArcGIS, then compared to the DEM layer for quality control. A couple of methods were attempted to transform this feature into a continuous-valued layer with full map coverage. First, the polylines were converted to points with 500 m sampling. This point set was loaded into a Python script, which used a grid search routine to determine the best radius for a Gaussian KDE routine available in the scikit-learn package ([Pedregosa et al., 2011](#)). Ten-fold cross-validation was employed, which splits the data into 10 subsets and interchangeably trains on 9, tests on 1 to get an average performance score. Based on a calculation of the log-likelihood, the best radius was found to be 45,600 m. However, when the kernel density operation is applied to the data with this radius, the map shows a central blob of high density, which falls off toward the sides of the survey. With such a large kernel radius, edge effects come into play since no drainage polylines were available outside of the AOI boundary. Furthermore, the conversion of line data to points for this method disregards the spatial relationships of the connected line data. The ArcGIS *Kernel Density* operation, which handles line data and suggests a kernel radius, produced a layer with more reasonable density relationships by visual inspection. The final drainage density layer used an output cell size of 0.0025 degrees and an auto-determined search radius of 0.272 degrees (Figure [A-6](#)).

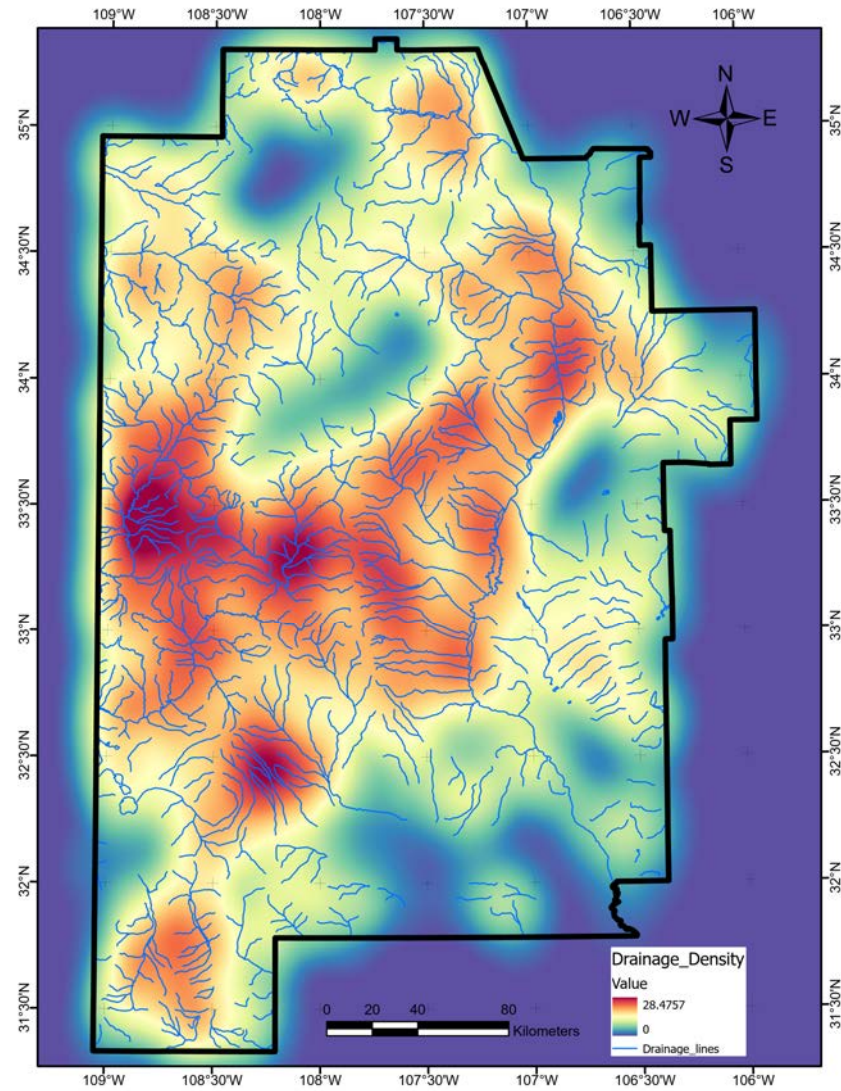


Figure A-6: Drainage-density data layer. Units are degree/degree². Blue lines show the drainage polyline data set from [Bielicki et al. \(2015\)](#).

A.7 Earthquake Density

Following the procedure outlined by [Pepin \(2019\)](#), an earthquake catalog for Southwestern New Mexico was created by combining historical earthquake catalogs for 1869-1998 ([Sanford et al., 2002](#)), 1999-2004 ([Sanford, 2006](#)), and 2005-2009 ([Pursley, 2013](#)) with data pulled from the USGS Earthquake catalog ([USGS, 2021a](#)) through to January 2021. All events were combined into a single dataframe in Python, and event duplicates were removed. The final catalog, cropped to the Regional Polygon boundary, consists of 2,539 events spanning 1962-2020. This point set was loaded into a KDE Python script, which used a grid search to determine the best radius for the scikit-learn *KernelDensity* routine ([Pedregosa et al., 2011](#)). Ten-fold cross-validation was employed, which splits the data into 10 subsets and then interchangeably trains on 9 and tests on 1 to get an average performance score. The maximum log-likelihood indicates a best radius value of 11,600 m (Figure [A-7](#)).

KDE values calculated at each AOI grid-point location were loaded into ArcGIS, and the *Kriging* function created a final surface for plotting purposes. *Kriging* parameters included: Spherical semivariogram model, lag size of 10^{-6} degrees, variable search radius with 12-point requirement, and output cell size of 0.01 degrees. The final layer is shown in Figure [A-8](#).

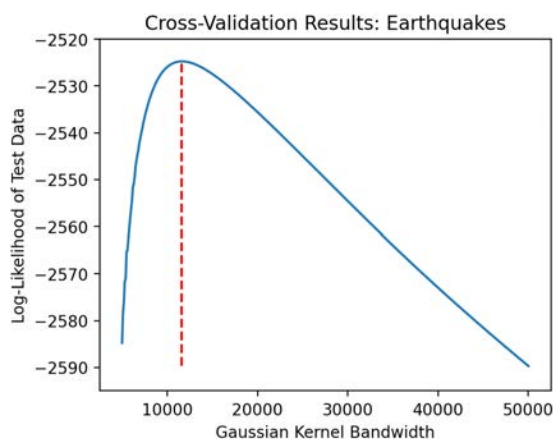


Figure A-7: Cross-validation results for earthquake KDE. Red dashed line indicates maximum log-likelihood value identifying the best kernel radius.

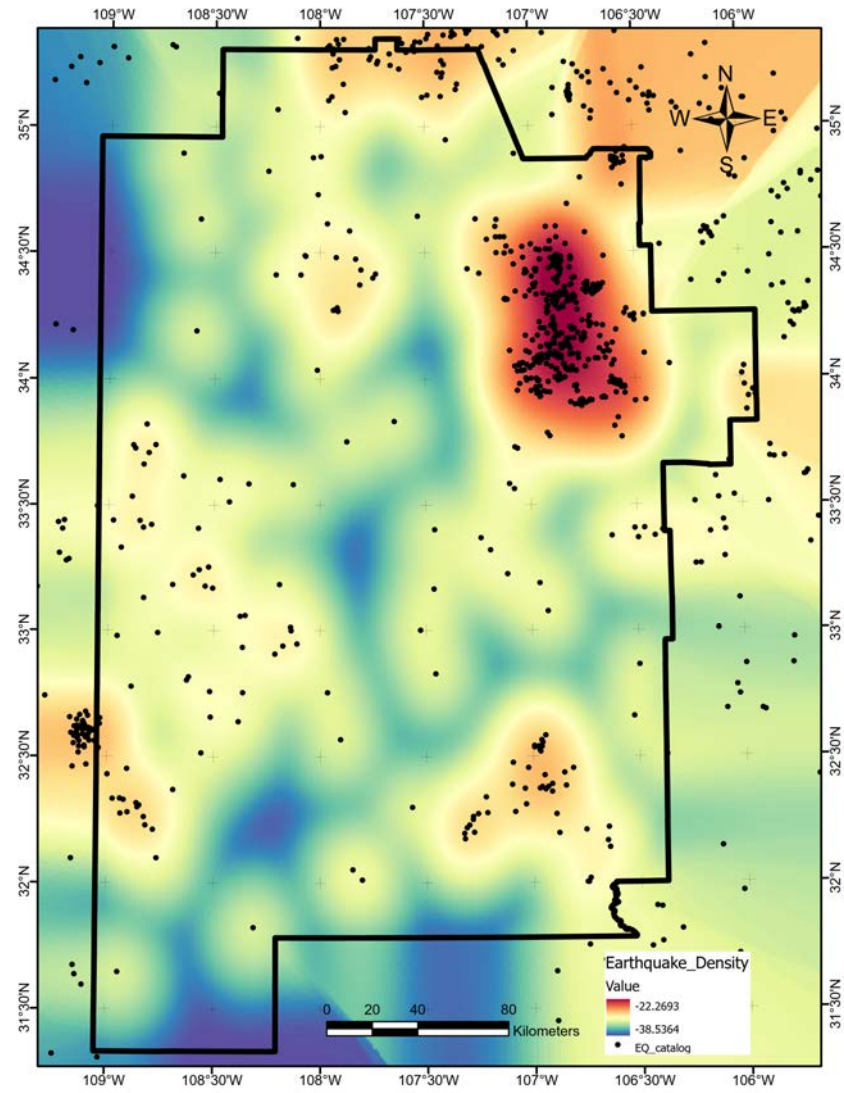


Figure A-8: Earthquake-density data layer. Units are $\log(\text{points}/\text{km}^2)$. Black dots indicate earthquake event point locations.

A.8 Gamma-Ray Absorbed Dose Rate

Aerial gamma-ray surveys conducted across the United States in the late 1970-1980s allowed for the construction of Potassium (K) concentration (in percent K), equivalent Uranium (eU) concentration (in ppm), and equivalent Thorium (eTh) concentration (in ppm) maps, which tie back to mineralogy and hence are a proxy for stratigraphy. These measures collectively define the absorbed dose rate, which can be calculated from the following equation: $D = 13.2K + 5.48eU + 2.72eTh$ (Duval et al., 2005).

The absorbed dose rate for West Central USA was downloaded from the USGS Open-File Report 2005-1413 website (Duval et al., 2005), loaded into ArcGIS, and cropped to the Regional Polygon bounds. A data gap in the vicinity of the White Sands Missile Range to the southeast of the study area necessitated layer interpolation using kriging. Grid values were extracted using the AOI mesh grid, then passed through the ArcGIS *Kriging* function to create the final layer based on the following preferred parameters: Spherical semivariogram model, auto-determined lag size of 0.097 degrees, and a variable search radius with a 4-point requirement (Figure A-9).

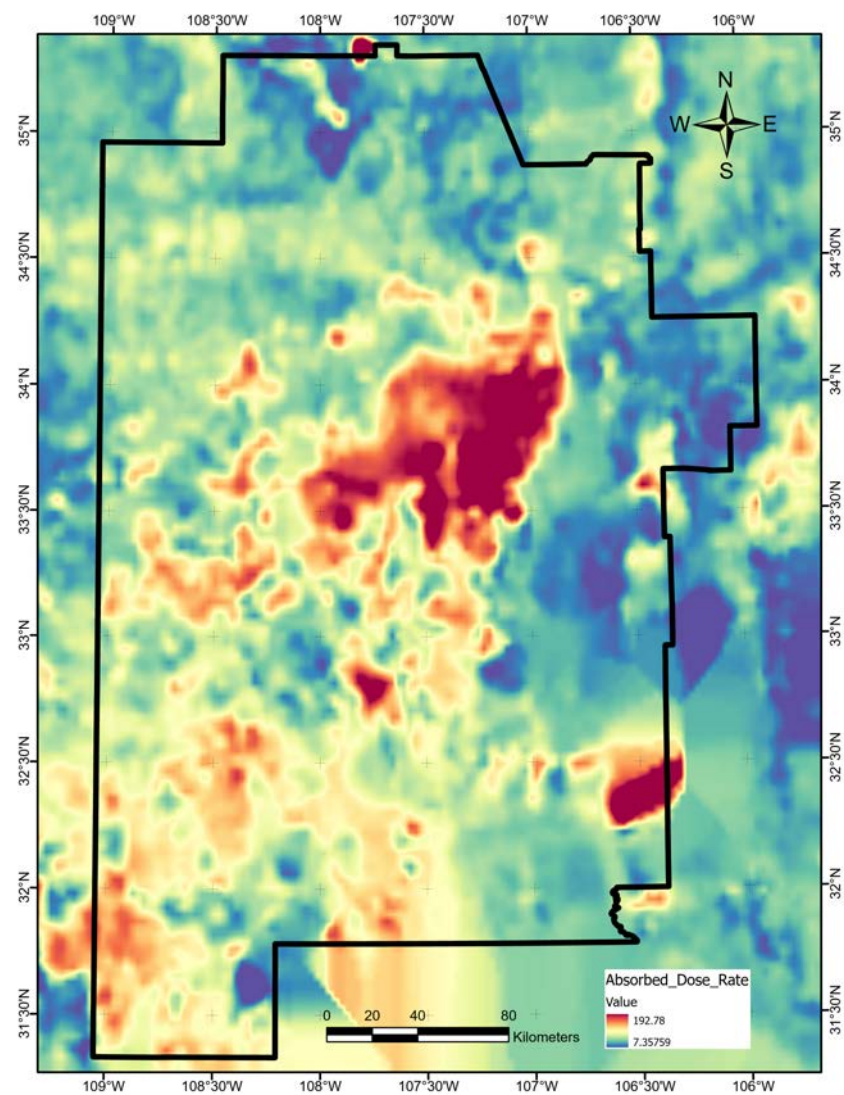


Figure A-9: Absorbed dose-rate data layer. Units are nanograys/hour (nGy/hr). Original data from USGS Open-File Report 2005-1413 (Duval et al., 2005).

A.9 Geodetic Strain Rate

GPS stations worldwide record local movements in the crust. These movements can highlight inflation or subsidence of the surface, fault motions, or plate tectonic activity. The symmetric part of the gradient of crustal-velocity vector is the strain-rate tensor $\dot{\epsilon}$, which indicates the accumulation of strain in an area. More concretely, it defines the speed with which the crust is deforming, and it can be treated as a proxy for earthquake potential since slip occurs due to the accumulation of strain (GEM, 2014). The Global Strain Rate Model (GSRM) v.2.1 provides a model for strain rate based on over 22,000 measurements from over 18,000 locations around the world (Kreemer et al., 2014). The output of this model was downloaded from the University of Nevada Reno Geodetic Laboratory host site (Kreemer, 2020). GSRM describes elements of the full strain tensor at a 0.1° resolution. The magnitude or second invariant of the strain tensor can combine these elements into a single value (Kreemer et al., 2014):

$$\|\dot{\epsilon}\| = \sqrt{\text{tr}(\dot{\epsilon} \bullet \dot{\epsilon})} = \sqrt{\sum_{i,j} \dot{\epsilon}_{ij} \dot{\epsilon}_{ij}}. \quad (\text{A.1})$$

Due to the size of the GSRM model file and the complexity of this calculation, the data were first loaded into Python, cropped to the Regional Polygon bounds, and the strain-rate magnitude was calculated for each point. These data were then loaded into ArcGIS and gridded using the *Spline* function for a smooth interpolation of the coarser GSRM grid. The final layer was created using the following *Spline* parameters: Regularized type, weight of 0.1, 4-point requirement, and cell size of 0.025 degrees (Figure A-10).

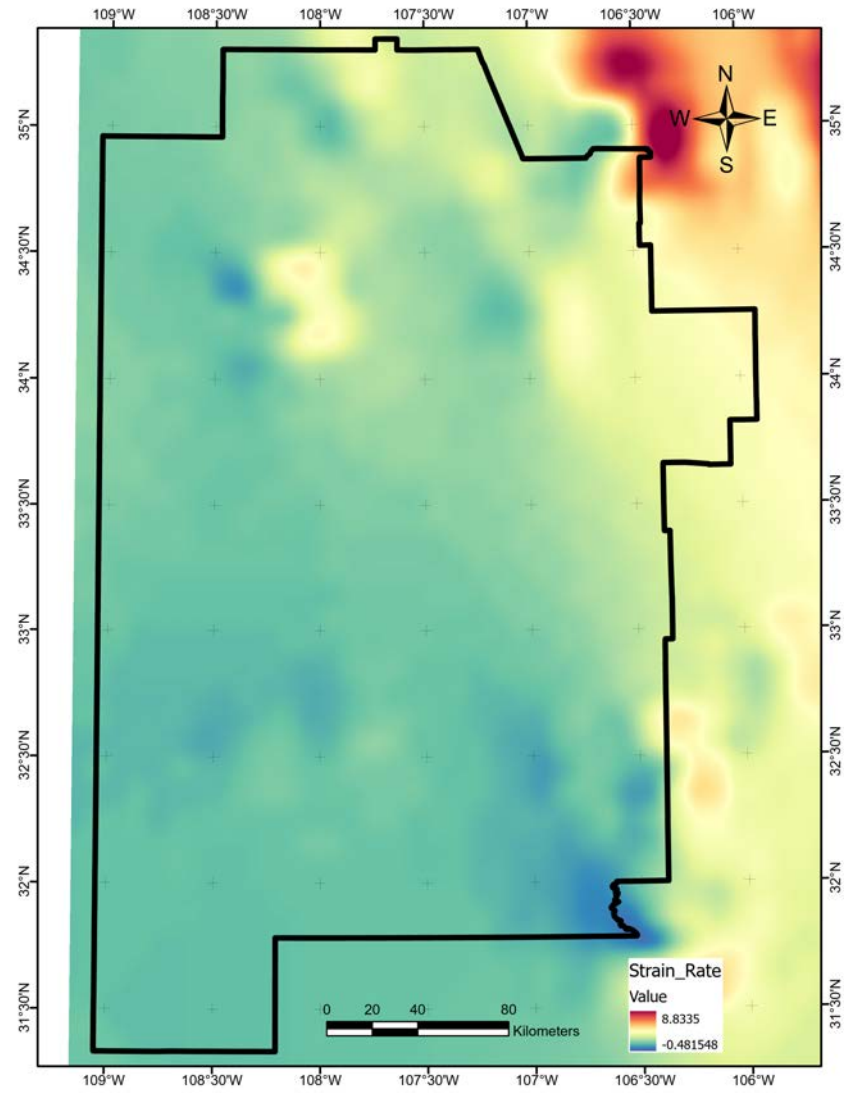


Figure A-10: Geodetic strain-rate data layer. Units are 10^{-9} yr^{-1} . Layer is based on data from [Kreemer et al. \(2014\)](#).

A.10 Gravity Anomaly

Terrain-corrected gravity-anomaly data available from the University of Texas El Paso (UTEP, 2011) were used in both the PFA analysis (Bielicki et al., 2015) and cluster analysis (Pepin, 2019) for Southwestern NM. The data layer from Bielicki et al. (2015) was downloaded from their OpenEI submission (Kelley, 2015) and loaded into ArcGIS. This layer required no further processing (Figure A-11).

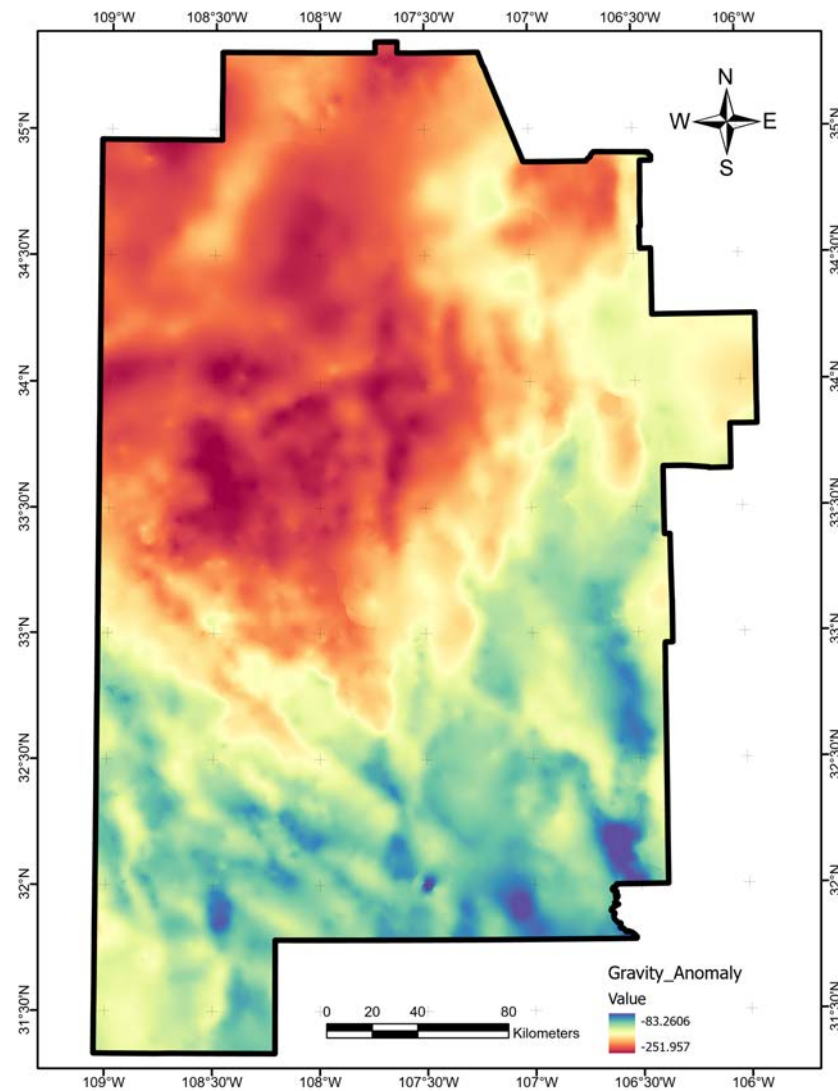


Figure A-11: Gravity-anomaly data layer. Units are milligals (mGal). Raster originally created by Bielicki et al. (2015).

A.11 Gravity-Anomaly Gradient

Gravity-anomaly gradient was calculated using the ArcGIS *Slope* function on the final Gravity Anomaly raster. Parameters used to create the final layer include geodesic method, z -unit of meters, and output measurement of degrees (Figure A-12).

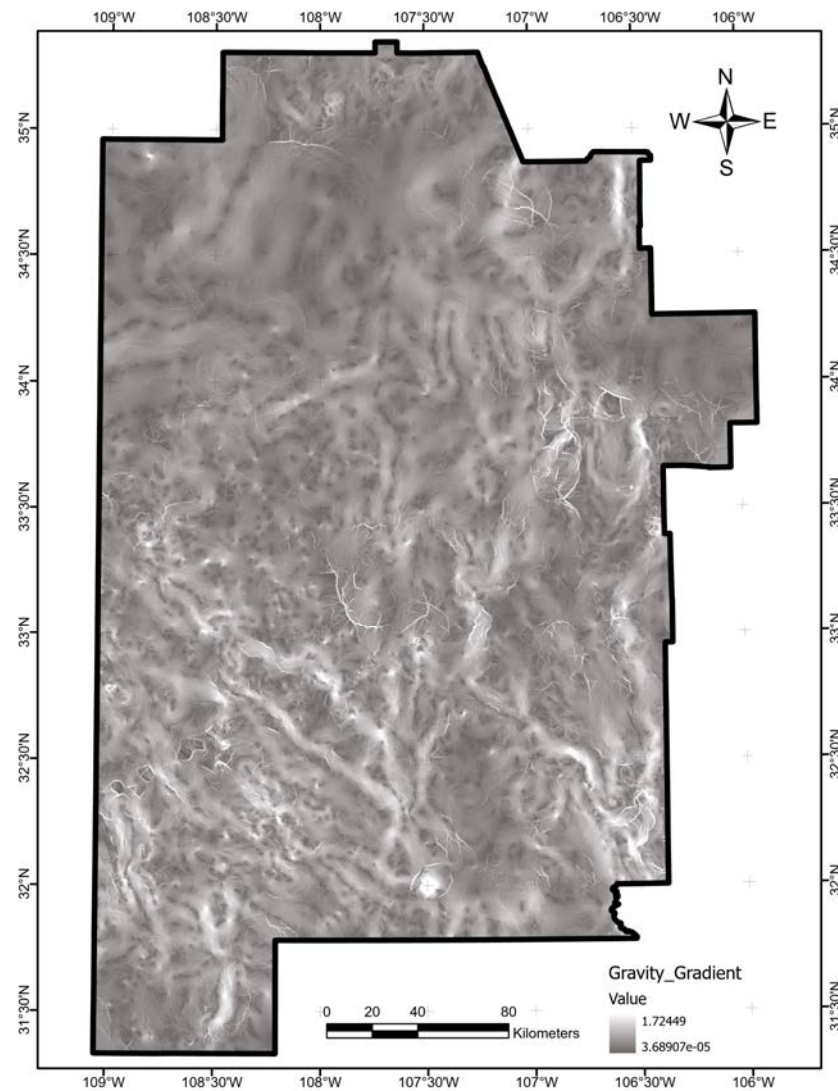


Figure A-12: Gravity-anomaly gradient data layer. Units are mGal/degree.

A.12 Heat Flow

The $0.5^{\circ} \times 0.5^{\circ}$ -resolution heat-flow model from [Lucazeau \(2019\)](#) offers coarse coverage across the Southwestern NM AOI. This model was downloaded directly from the supporting information section of the publication page ([Lucazeau, 2019](#)), imported into ArcGIS, and cropped to the Regional Polygon boundaries. After testing several gridding algorithms for a smooth representation of these sparse data, the ArcGIS *Topo to Raster* function produced the best results. The parameters for creating the final layer include: tolerance-1 of 2.5, tolerance-2 of 100, Enforce drainage setting, Contour input data, and output cell size of 0.01 degrees (Figure [A-13](#)).

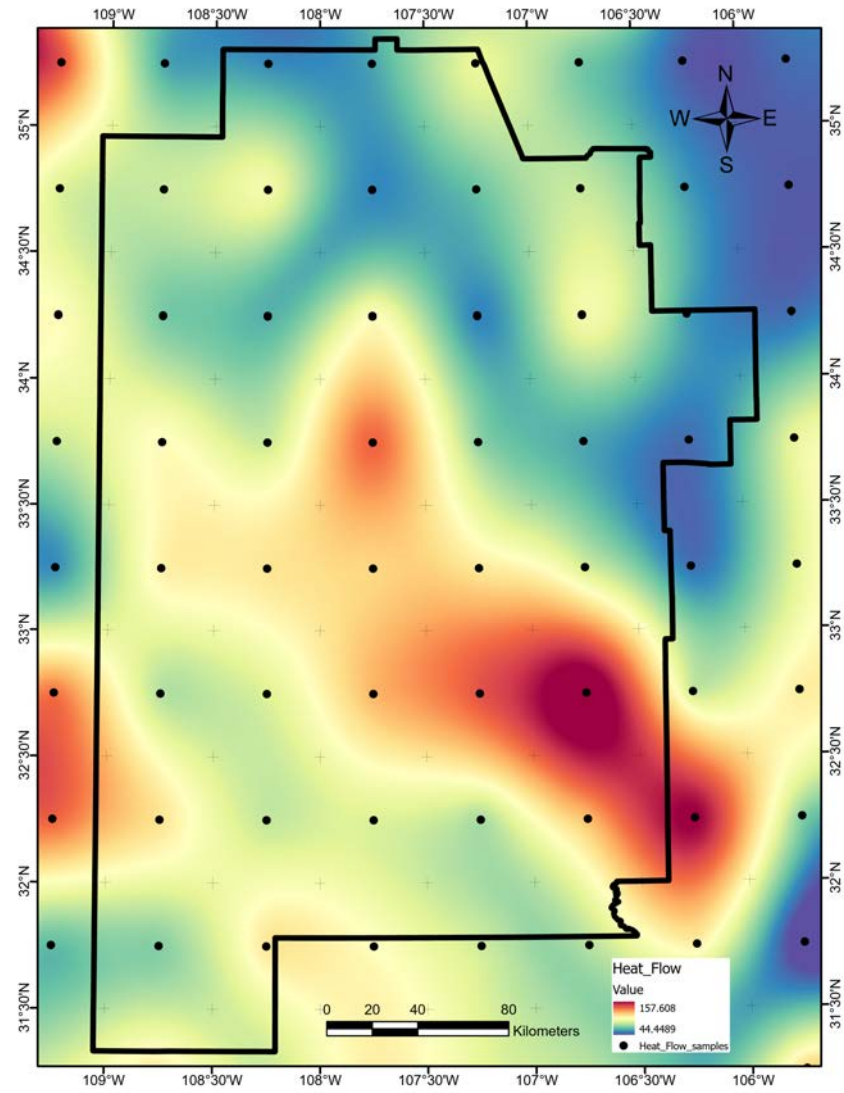


Figure A-13: Heat-flow data layer. Units are mW/m^{-2} . Black dots mark the original source data points from [Lucazeau \(2019\)](#).

A.13 Lithium Concentration

Measurements of lithium concentration were assembled by [Bielicki et al. \(2015\)](#) from USGS records, student dissertations, and other sources. These data were downloaded from the NM PFA OpenEI submission ([Kelley, 2015](#)) and merged using ArcGIS and Python to create a single dataframe of 3,595 measurements within the broader Regional Polygon bounds. Restricting data input to the tighter AOI bounds led to artifacts in the interpolation process. As described for the Boron Concentration data layer, attempts to model lithium concentration using Gaussian Processes provided unsatisfactory results. Instead, the final layer was generated using the ArcGIS *Empirical Bayes Kriging* routine. Selected parameters include Empirical data transformation, a maximum of 100 points in each local model, 100 simulated semivariograms with K-Bessel model type, a Standard Circular search neighborhood, and output grid cell size of 0.01 degrees. All coincident data were included in the calculation, so any overlapping measurements were considered in generating the final layer (Figure [A-14](#)).

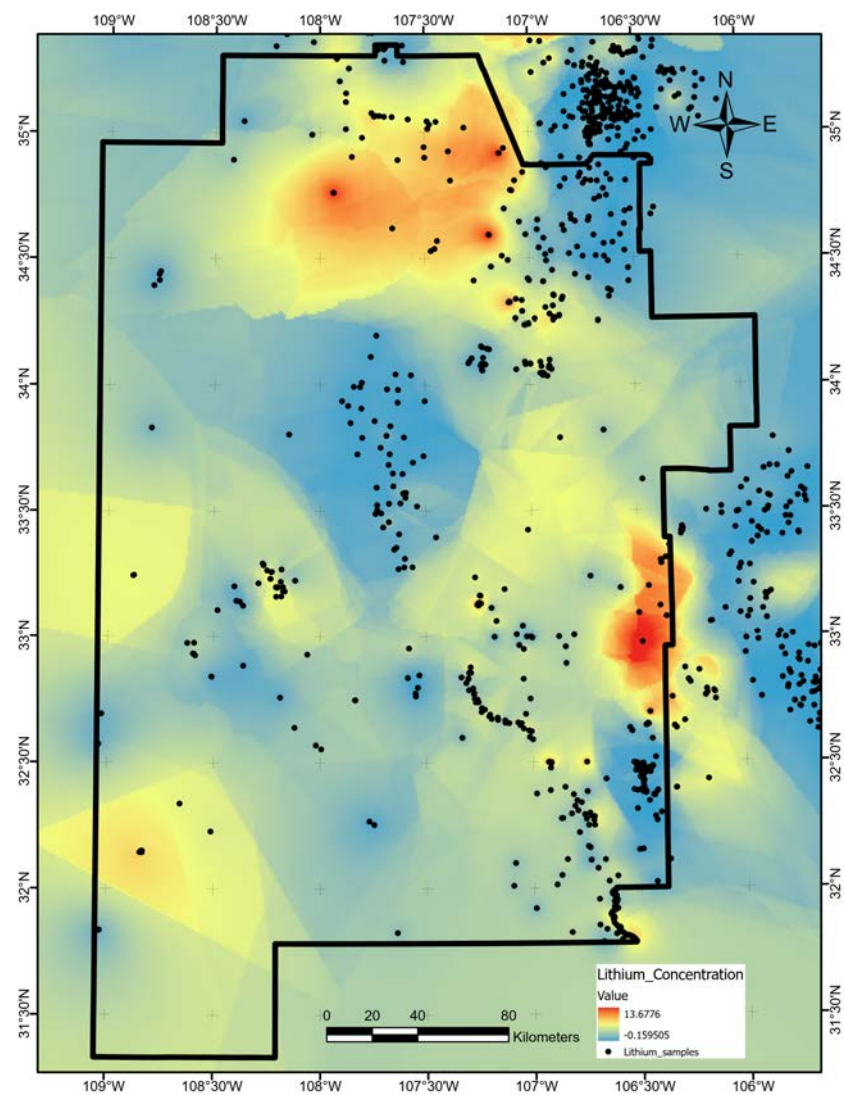


Figure A-14: Lithium-concentration data layer. Units are mg/L. Black dots mark sample locations in the complete data set from [Bielicki et al. \(2015\)](#).

A.14 Magnetic Anomaly

USGS magnetic-anomaly data from merged aerial surveys ([Bankey et al., 2002](#)) were used in both the Southwestern NM PFA analysis ([Bielicki et al., 2015](#)) and cluster analysis ([Pepin, 2019](#)). After downloading the raster from the PFA OpenEI submission ([Kelley, 2015](#)), it was imported directly into ArcGIS. No further processing was required (Figure A-15).

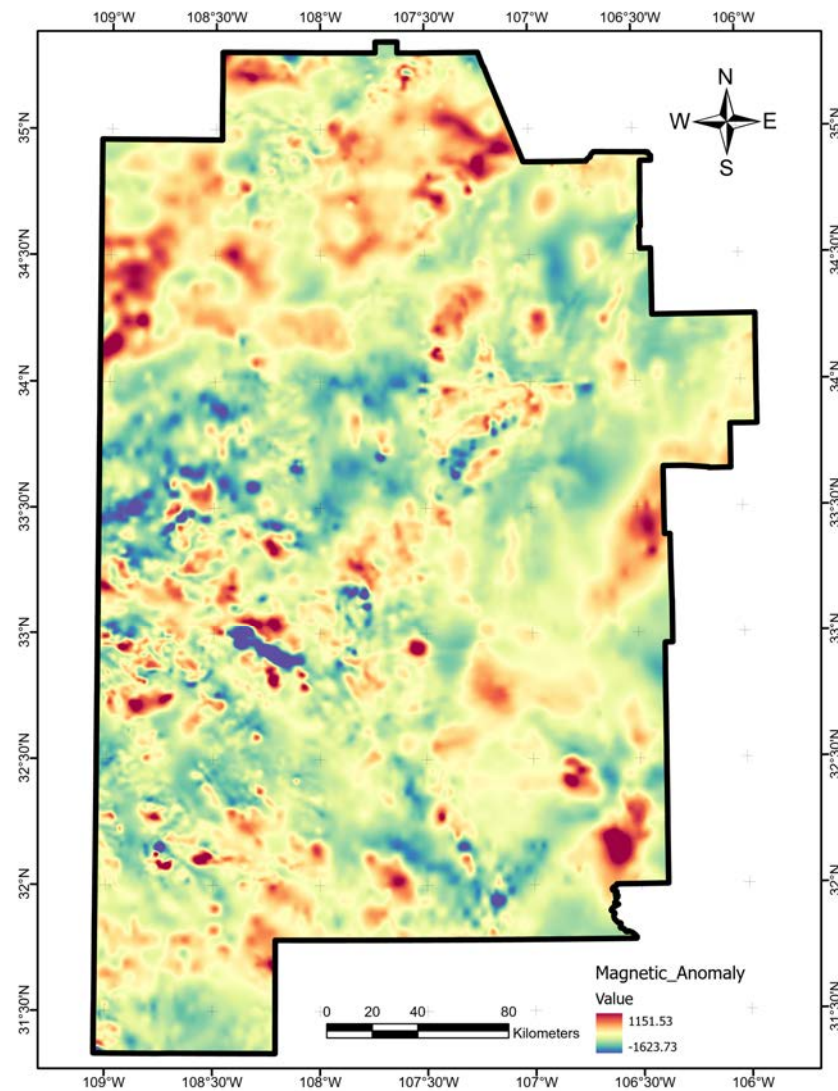


Figure A-15: Magnetic-anomaly data layer. Units are nanoteslas (nT). Raster originally created by [Bielicki et al. \(2015\)](#).

A.15 Magnetic-Anomaly Gradient

Magnetic-anomaly gradient was calculated using the ArcGIS *Slope* function on the final Magnetic Anomaly raster. Parameters selected to create this layer include geodesic method, z-unit of meters, and output measurement of degrees (Figure A-16).

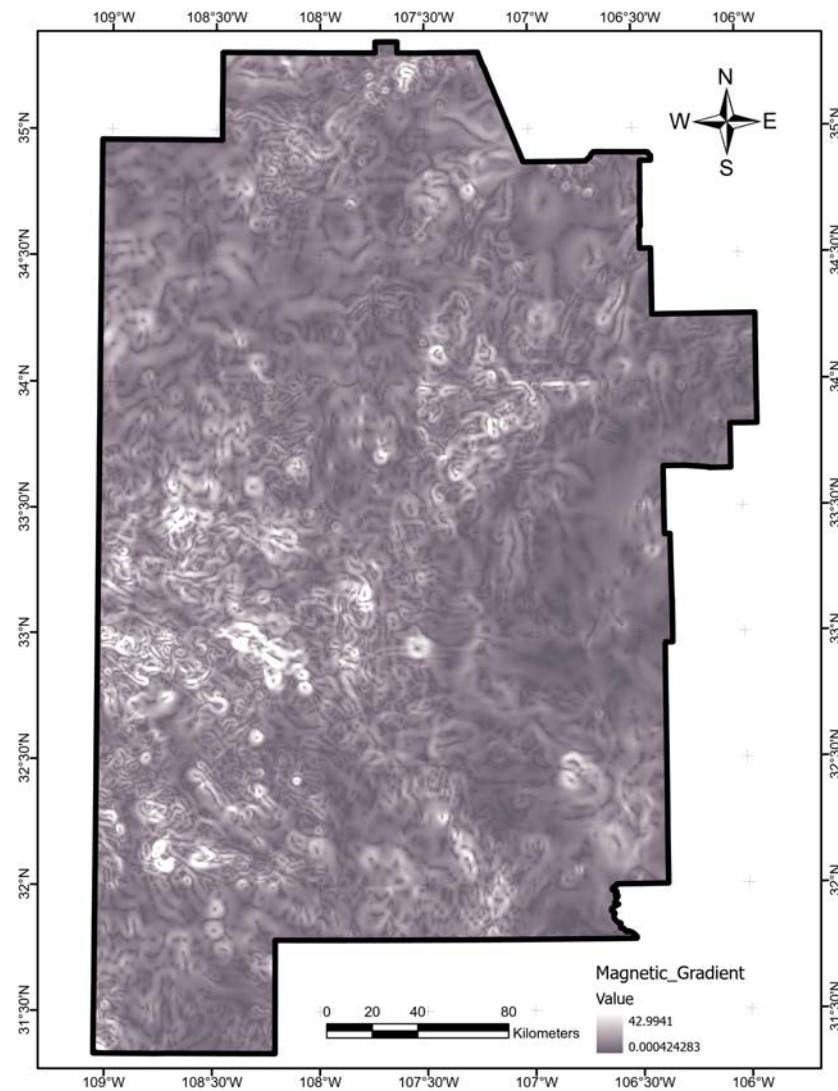


Figure A-16: Magnetic-anomaly gradient data layer. Units are nT/degree.

A.16 Quaternary Fault Density

Faults showing Quaternary displacement were digitized at the 1:24,000 scale by the New Mexico Bureau of Geology and Mineral Resources and provided to [Bielicki et al. \(2015\)](#) and [Pepin \(2019\)](#) in support of their investigations. The associated polyline features were downloaded from the PFA OpenEI submission ([Kelley, 2015](#)) and loaded into ArcGIS. As discussed for the Drainage Density layer, a Python-based kernel-density workflow, using extracted points from these polylines, failed to produce satisfactory results. Instead, the ArcGIS *Kernel Density* function was applied to create the final layer map (Figure [A-17](#)). Selected parameters for this function include an output cell size of 0.0025 degrees and an auto-determined search radius of 0.367 degrees.

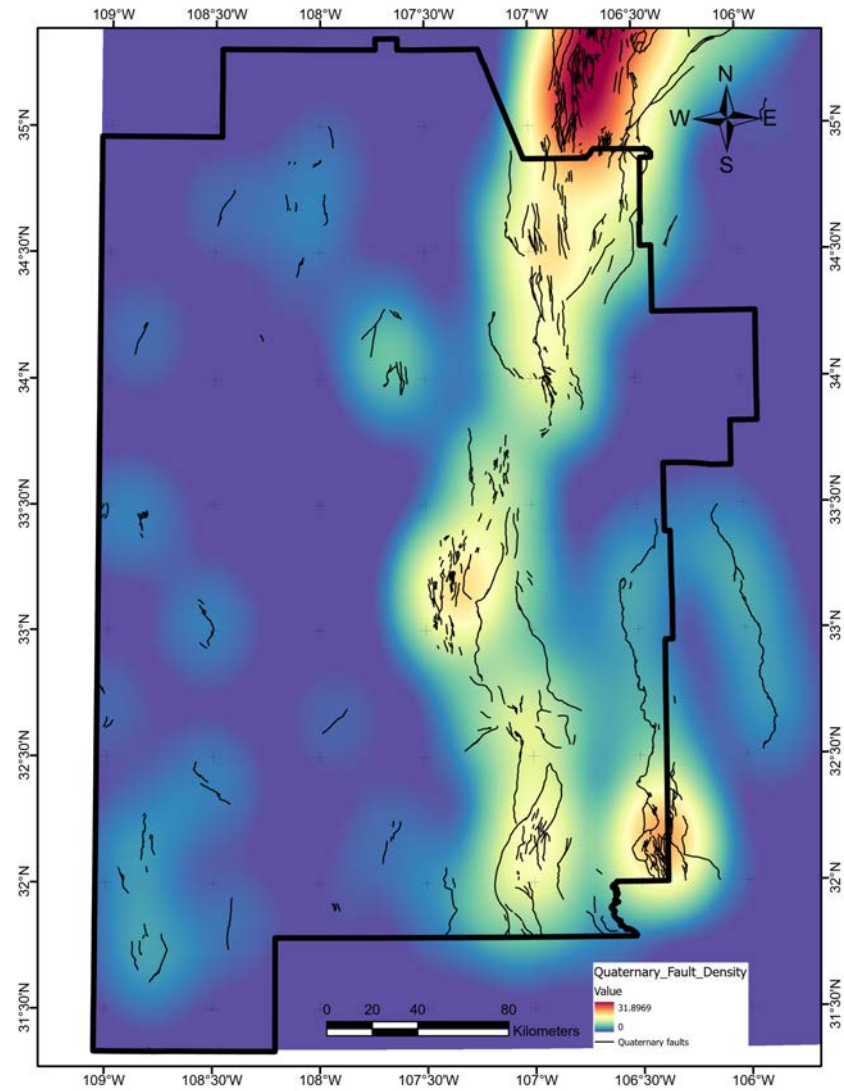


Figure A-17: Quaternary fault-density data layer. Units are degree/degree². Black lines show the fault polyline data set archived by [Bielski et al. \(2015\)](#).

A.17 Silica Geothermometer Temperature

Silica-concentration data from across the study area were compiled by [Bielicki et al. \(2015\)](#) and converted to reservoir temperatures using the Fournier chalcedony geothermometer relationship ([Fournier, 1977](#)). These data were downloaded from the Southwestern NM PFA OpenEI submission ([Kelley, 2015](#)) and merged using ArcGIS and Python to create a single dataframe of 7,259 measurements, all within the broader Regional Polygon bounds to avoid surface-creation edge effects within the tighter AOI. As described for the Boron Concentration data layer, attempts to model Si-geothermometer estimates using Gaussian Processes provided unsatisfactory results. Instead, the final layer was created using the ArcGIS *Empirical Bayes Kriging* routine. Selected parameters include Empirical data transformation, a maximum of 100 points in each local model, 100 simulated semivariograms with K-Bessel model type, a Standard Circular search neighborhood, and output grid cell size of 0.01 degrees. All coincident data were included in the calculation, so any overlapping measurements were considered in generating the final layer (Figure [A-18](#)).

Note that low groundwater silica concentrations can lead to physically unrealistic negative temperatures with the Fournier chalcedony relationship. These values are preserved here to capture relative variation from high to low silica concentrations. The machine learning methods described in Chapter [3](#) focus on differences in value, not absolute magnitudes, so negative Si geothermometer temperatures can be tolerated.

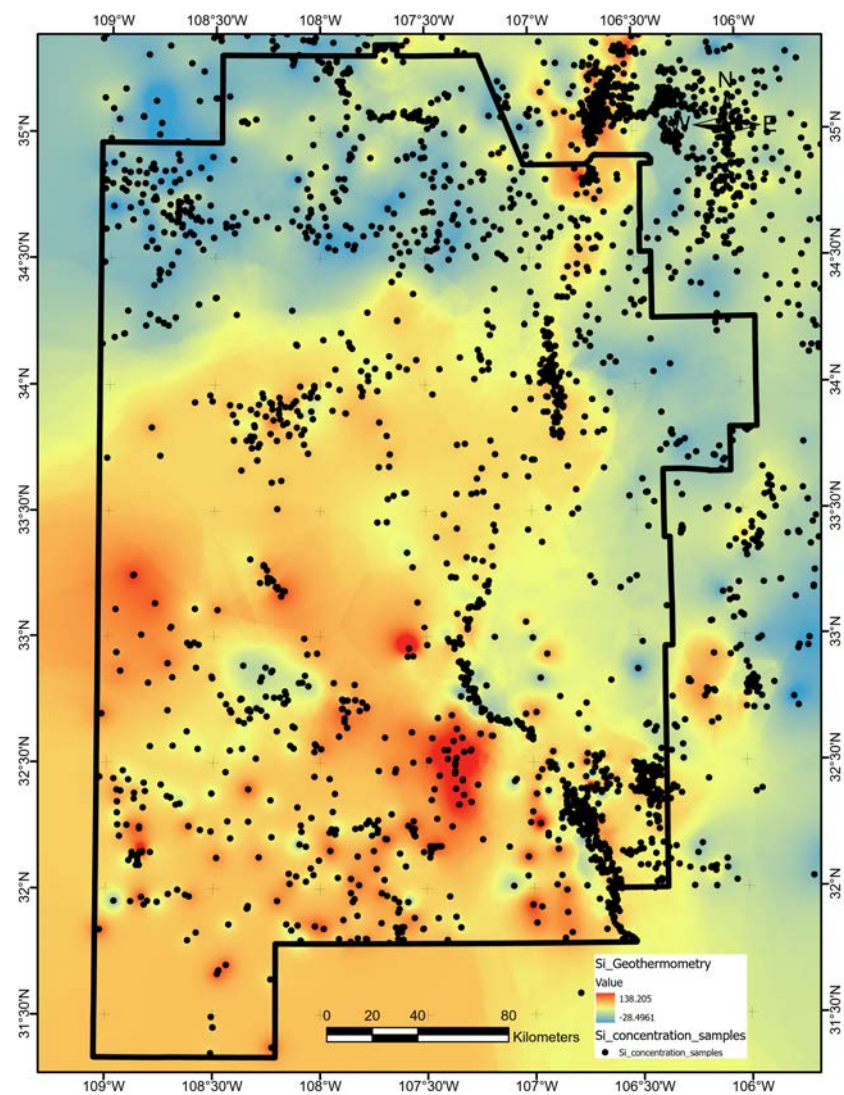


Figure A-18: Chalcedony geothermometer data layer. Units are °C. Black dots indicate locations where silica concentration was sampled, as collected by [Bielicki et al. \(2015\)](#).

A.18 Spring Density

The locations of springs in the study area were downloaded from the USGS National Water Information System (USGS, 2021c). A total of 2,565 springs were recorded within the bounds of the Regional Polygon. As with the Earthquake Density layer, this point set was loaded into a KDE Python script, which used a grid-search routine to determine the best kernel radius for a Gaussian kernel density operator. Ten-fold cross-validation identified the best radius value of 31,400 m (Figure A-19).

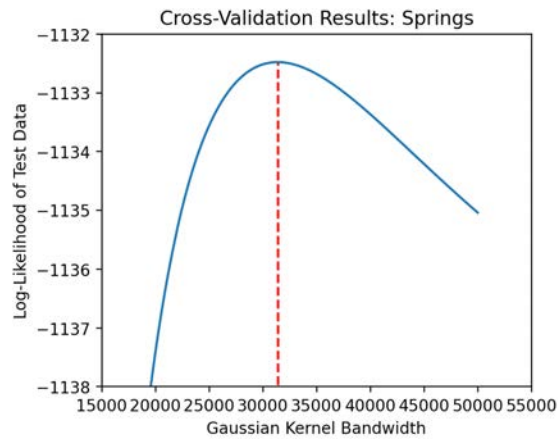


Figure A-19: Cross-validation results for the springs KDE. Red dashed line indicates maximum log-likelihood value identifying the best kernel radius.

KDE values determined at each AOI grid-point location were loaded into ArcGIS, and *Kriging* was used to generate a final layer for plotting purposes (Figure A-20). Selected *Kriging* parameters included a Spherical semivariogram model, lag size of 10^{-6} degrees, variable search radius with 12-point requirement, and output cell size of 0.01 degrees.

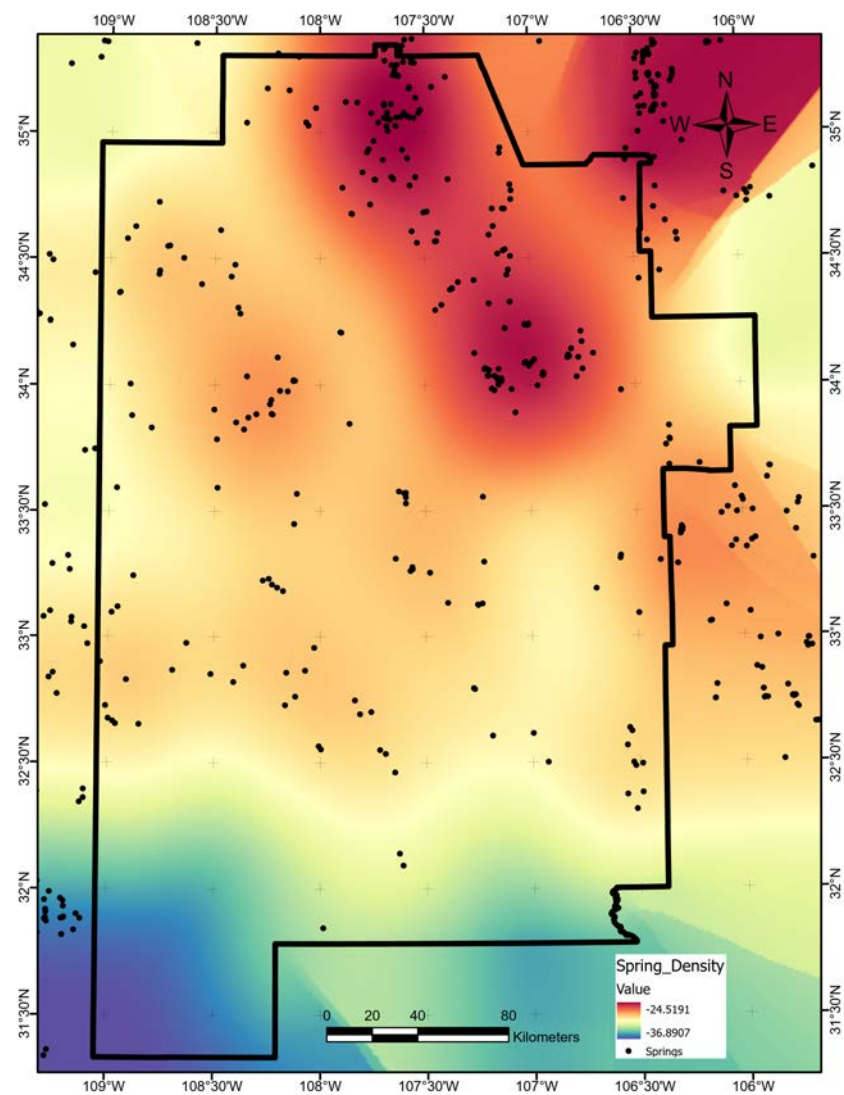


Figure A-20: Spring-density data layer. Units are $\log(\text{points}/\text{km}^2)$. Black dots indicate spring locations from the [USGS \(2021c\)](#).

A.19 State Map Fault Density

Digitized fault outlines from state geologic maps are available for download from the USGS Energy and Environment in the Rocky Mountain Area data portal ([USGS, 2021b](#)), including New Mexico state faults ([Stoeser et al., 2005](#)). These fault polylines were loaded into ArcGIS and, like the Quaternary faults, converted to fault density using the *Kernel Density* operation (Figure [A-21](#)). Selected parameters for this function include an output cell size of 0.0025 degrees and an auto-determined search radius of 0.252 degrees.

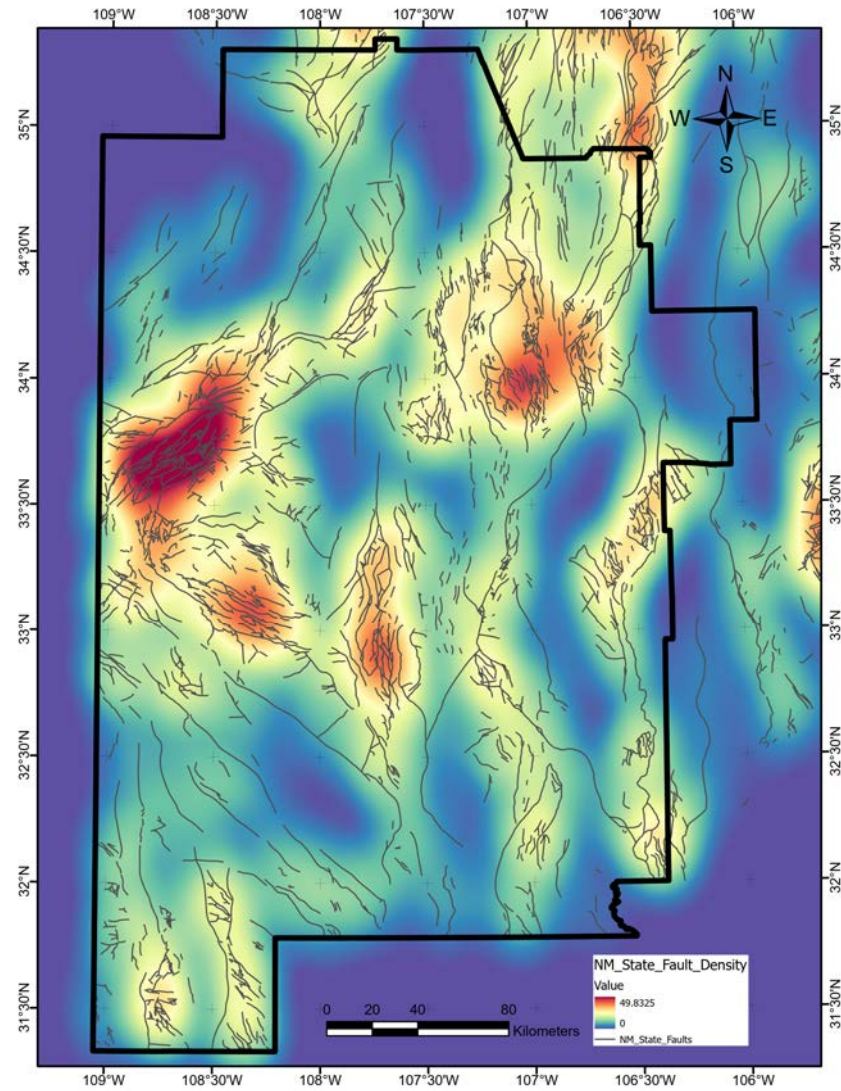


Figure A-21: State fault density data layer. Units are degree/degree². Dark gray lines trace the fault polyline data set obtained from USGS Open-File Report 2005-1351 (Stoeser et al., 2005).

A.20 Surface Topography (DEM)

The Southwestern NM PFA data archive ([Kelley, 2015](#)) includes a Digital Elevation Model (DEM) layer with surface elevations at one arc-sec resolution. This layer was downloaded and imported into ArcGIS; however, a data gap along the easternmost section of the AOI required the addition of two $1^{\circ} \times 1^{\circ}$ DEM tiles at the same resolution to fully complete the layer. These tiles were downloaded from the USGS National Map website ([USGS, 2021d](#)) and merged with the original DEM. The final layer required no further processing (Figure [A-22](#)).

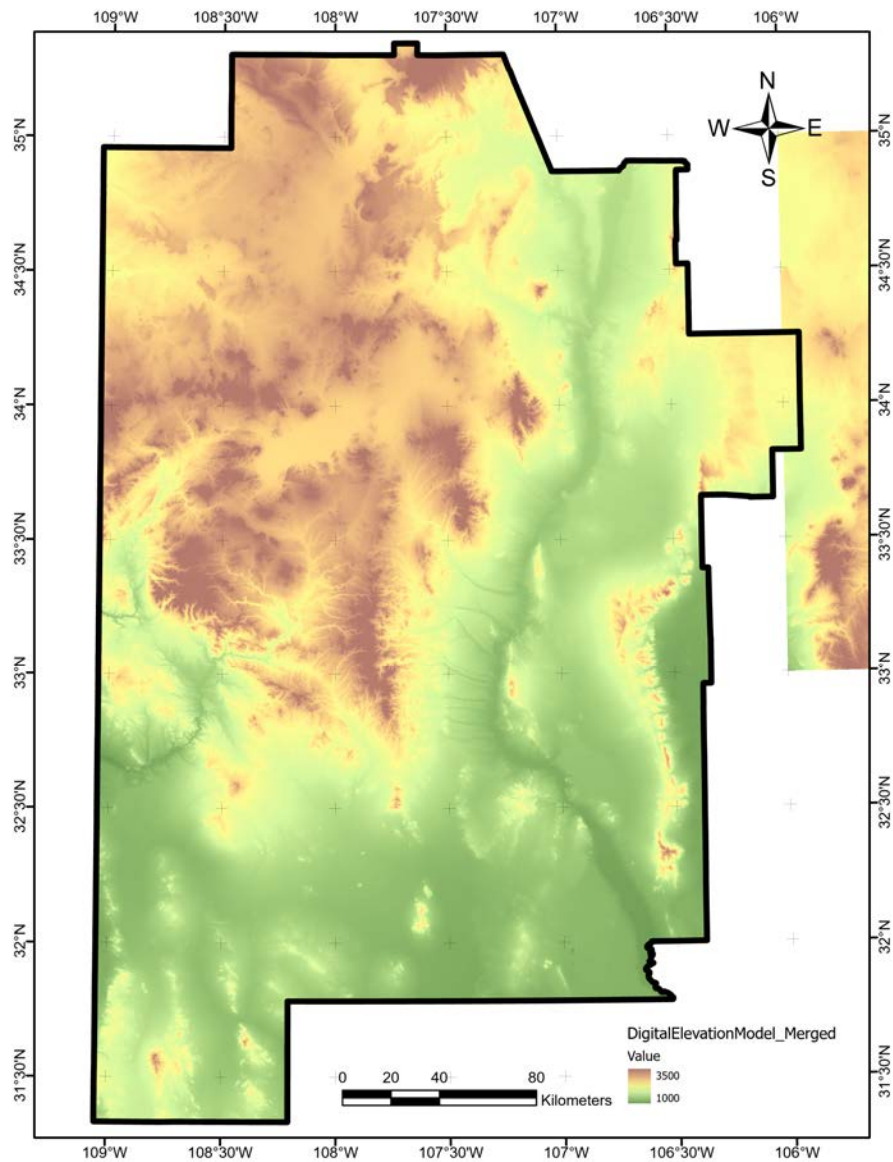


Figure A-22: Surface-topography (DEM) data layer. Units are meters. Layer combines the DEM raster from [Bielicki et al. \(2015\)](#) with data from The National Map online ([USGS, 2021d](#)).

A.21 Topographic Gradient

Topographic-gradient magnitude was calculated using the ArcGIS *Slope* function on the final DEM layer. This method calculates the maximum horizontal gradient from each raster cell to each of its eight neighbors and reports this value in degrees. Parameters selected to create this layer include Geodesic method and a z -unit of meters (Figure A-23).

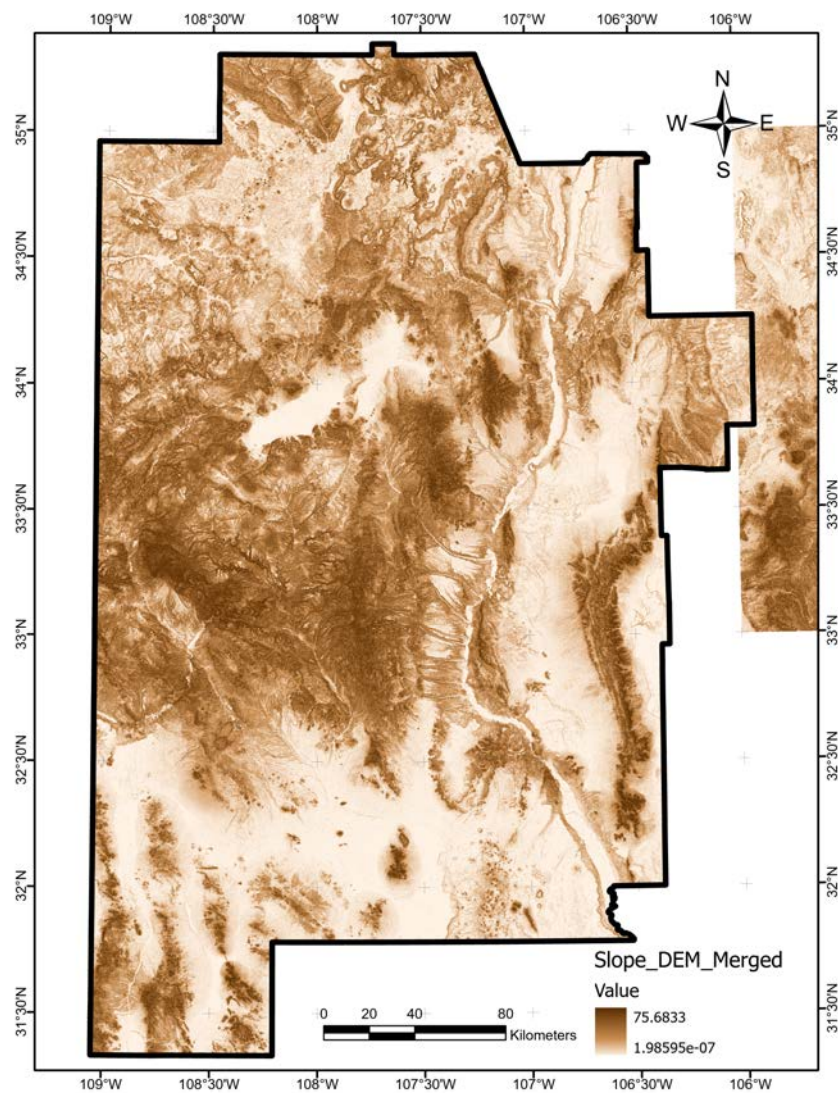


Figure A-23: Topographic gradient-magnitude data layer. Units are meter/degree.

A.22 Volcanic-Dike Density

The USGS Energy and Environment in the Rocky Mountain Area data portal ([USGS, 2021b](#)) also includes digitized volcanic-dike outlines from the New Mexico state geologic map ([Stoeser et al., 2005](#)). These polylines were imported into ArcGIS and, like the Quaternary fault data set, converted to density using the *Kernel Density* operation (Figure A-24). Selected parameters for this function included an output cell size of 0.0025 degrees and an auto-determined search radius of 0.252 degrees.

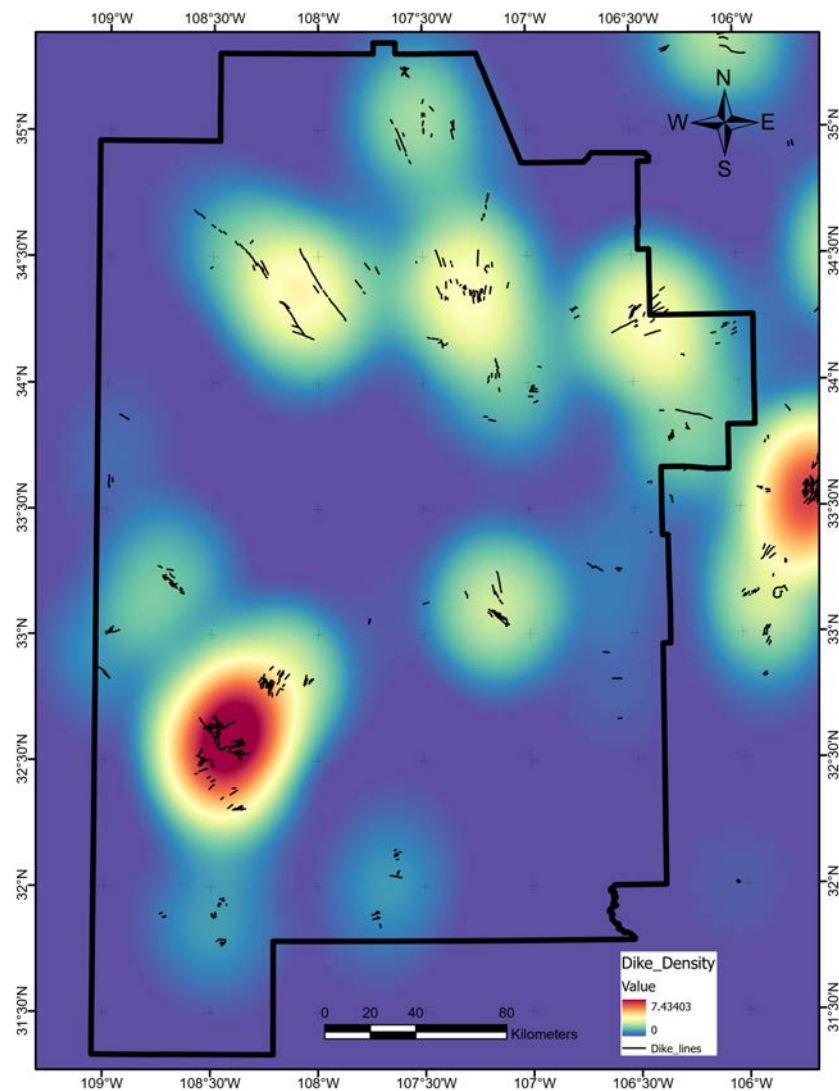


Figure A-24: Volcanic dike-density data layer. Units are in degree/degree². Black lines trace the dike polyline data set obtained from USGS Open-File Report 2005-1351 ([Stoeser et al., 2005](#))

A.23 Volcanic-Vent Density

Volcanic vents identified in the study area were retrieved from the New Mexico Bureau of Geology and Mineral Resources using the NMBGMR Interactive Map (NMBGMR, 2021). A total of 811 volcanic vents are observed within the geographic bounds of the Regional Polygon. As with the Earthquake Density layer, the point set was loaded into a KDE Python script, which used a grid-search routine to determine the optimal kernel radius. Ten-fold cross-validation suggested a radius of 28,300 m (Figure A-25).

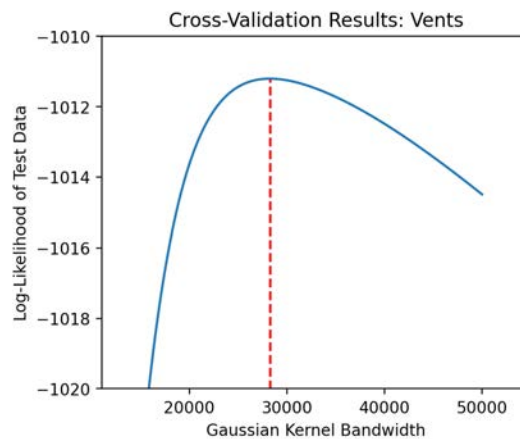


Figure A-25: Cross-validation results for volcanic-vent KDE. Red dashed line indicates maximum log-likelihood value identifying the best kernel radius.

KDE values calculated at each AOI grid-point location were loaded into ArcGIS, and *Kriging* was used to generate a final layer for plotting purposes (Figure A-26). *Kriging* parameters included spherical semivariogram model, lag size of 10^{-6} degrees, variable search radius with 12-point requirement, and output cell size of 0.01 degrees.

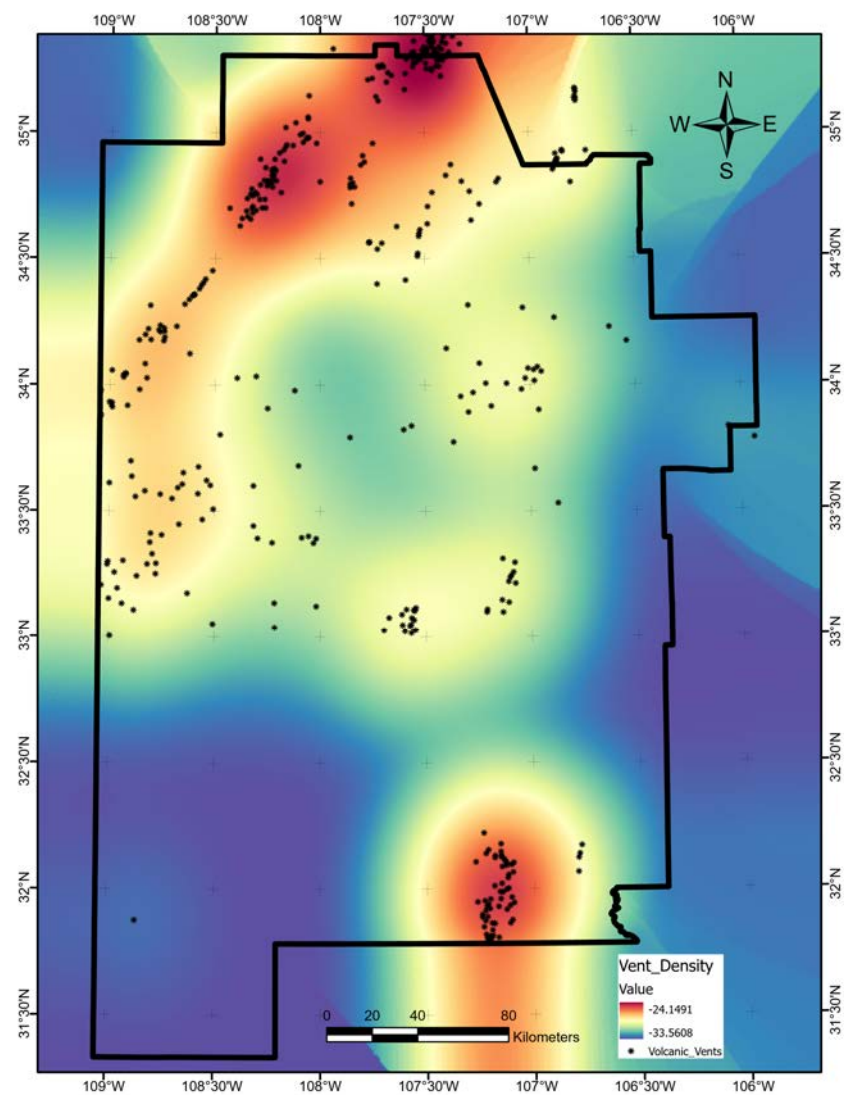


Figure A-26: Volcanic-vent density data layer. Units are in $\log(\text{points}/\text{km}^2)$. Black dots indicate vent locations from the [NMBGMR \(2021\)](#).

A.24 Water-Table Depth

[Bielicki et al. \(2015\)](#) mapped depth to the water table using data from the USGS and several additional sources. This raster was downloaded from their OpenEI submission ([Kelley, 2015](#)) and imported into ArcGIS. Data gaps between the raster extent and AOI polygon to the south and the east necessitated extrapolation of the layer, so ArcGIS *Empirical Bayes Kriging* was applied to fill in the missing values. After some trial-and-error, the chosen parameter values include: Empirical data transformation, maximum of 100 points in each local model, 100 simulated exponential semivariograms, a Standard Circular search neighborhood, and an output cell size of 0.01 degrees. Figure [A-27](#) shows the final layer.

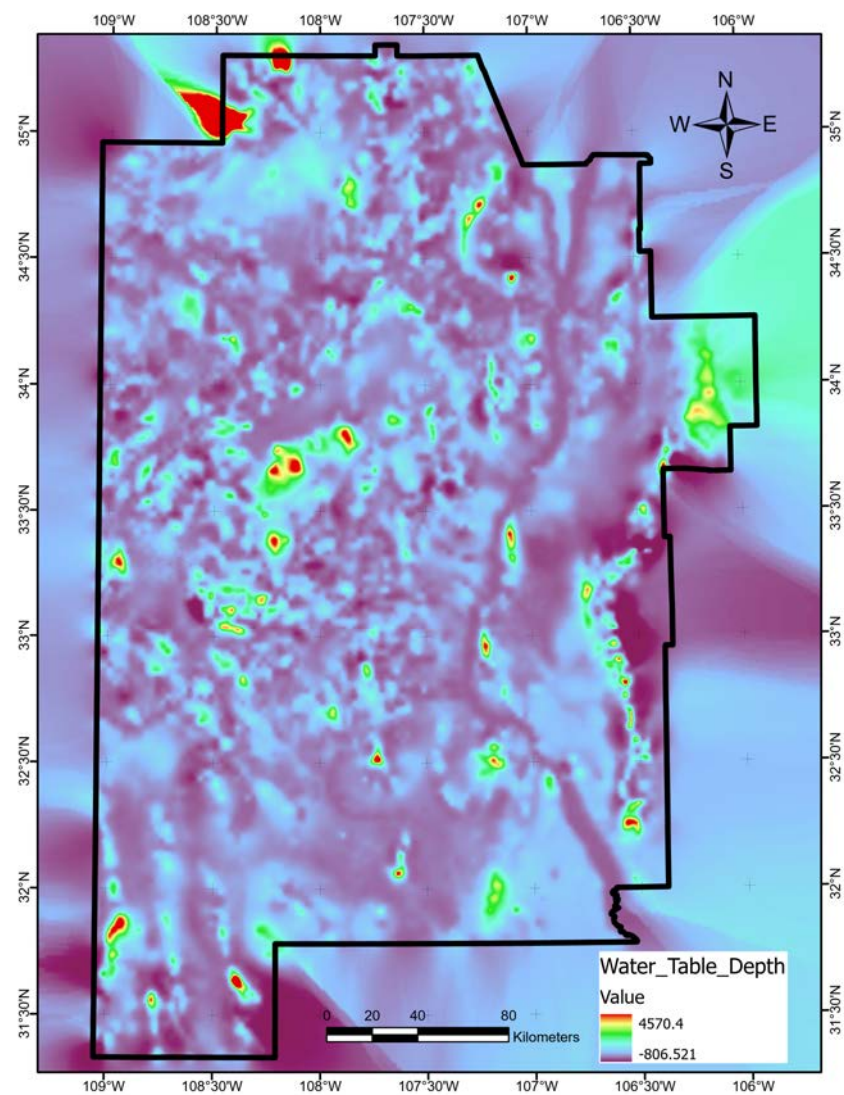


Figure A-27: Water-table depth data layer. Units are in feet. Adapted from raster created by (Bielicki et al., 2015).

A.25 Water-Table Gradient

A pre-calculated grid for the water-table gradient is among the layers included in the Southwestern NM PFA archive ([Kelley, 2015](#)). This raster was downloaded and imported into ArcGIS. In order to fill data gaps between the raster extent and the AOI polygon to the south and the east, the ArcGIS *Empirical Bayes Kriging* process was applied. After some trial-and-error, the final layer was generated using the following parameter values: Empirical data transformation, maximum of 100 points in each local model, 100 simulated exponential semivariograms, a Standard Circular search neighborhood, and an output cell size of 0.01 degrees (Figure [A-28](#)).

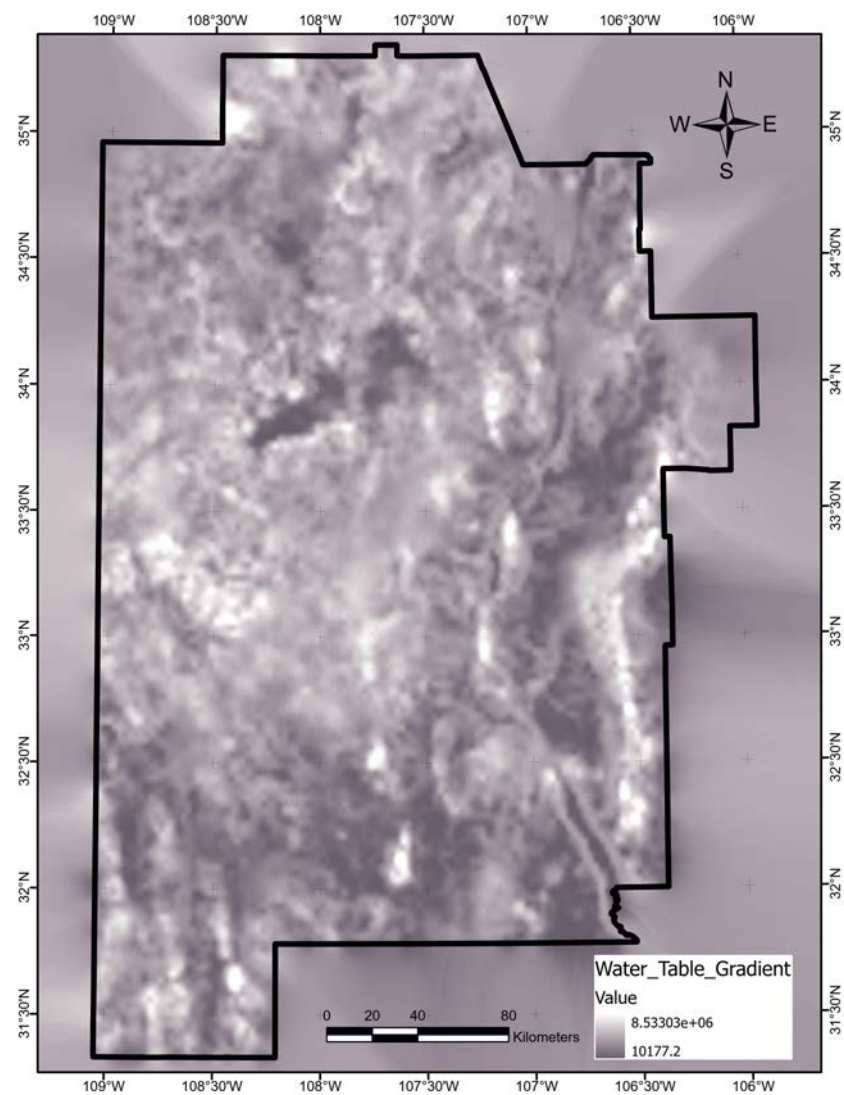


Figure A-28: Water-table gradient data layer. Units are in feet/degree. Based on the raster from [Bielicki et al. \(2015\)](#).

A.26 Geothermal Gradient Points

The response variable for this analysis comes from observations stored in the SMU *Heat Flow Database from Bottom Hole Temperature Data*, accessed via the SMU node of the National Geothermal Data System (SMU, 2021). This database focuses specifically on heat-flow values derived using geothermal-gradient and conductivity measurements from well data in journal articles, books, reports, and from other sources (Blackwell et al., 2014). Geothermal-gradient values are provided in two forms: reported gradient and corrected gradient. The latter incorporates both temperature and terrain corrections based on the well depth interval that a measurement was taken. Corrected geothermal gradient values were used when available; otherwise, the uncorrected geothermal gradient value was selected. The well data set was clipped to the bounds of the Regional Polygon, then loaded into Python for conditioning. Well records missing geographic coordinates or a geothermal-gradient value of either type were dropped. The list was sorted on coordinates and gradient, and only the first record for each location was kept. The sort was in descending order, so this method preserved the highest gradient value captured per well. The final set of 596 values shown in Figure A-29 comprise the raw input data used for predictive modeling of geothermal gradient across the study AOI.

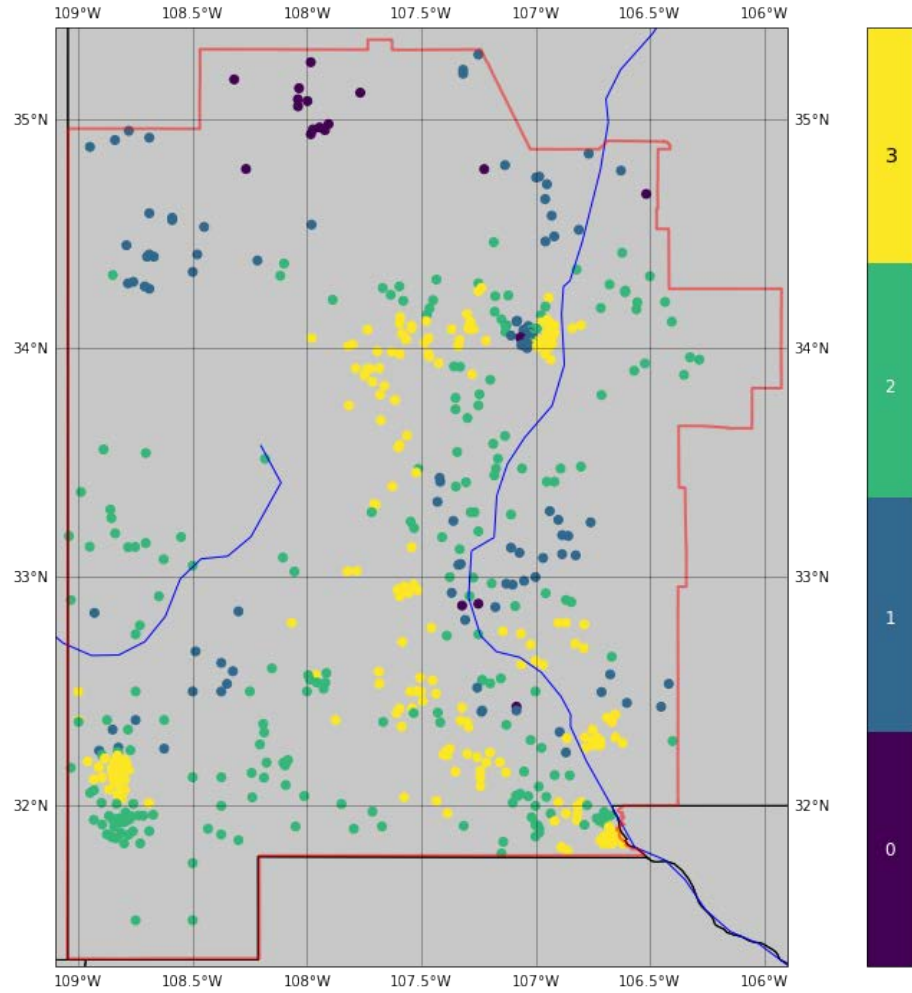


Figure A-29: Geothermal-gradient observations from well data. Markers are colored by geothermal gradient class as defined in Section 3.2.2. Data were retrieved from the SMU NGDS portal (SMU, 2021).

A.27 Geothermal Gradient Layer

A geothermal gradient GIS layer also exists within the [Bielicki et al.](#) OpenEI submission ([Kelley, 2015](#)), originally constructed by interpolating a prior version of the SMU well data set paired with additional oil & gas well measurements. The layer nearly covers the entire AOI, except for a missing section in the southwestern “boot-heel” of the state (Figure A-30A). To fill this gap, extrapolation was performed with the ArcGIS *Empirical Bayes Kriging* function using the following parameters: Empirical data transformation, a maximum of 100 points in each local model, 100 simulated semivariograms with Exponential model type, a Standard Circular search neighborhood, and an output cell size of 0.01 degrees. This layer was saved as a reference map for later comparison with predictive model results (Figure A-30B).

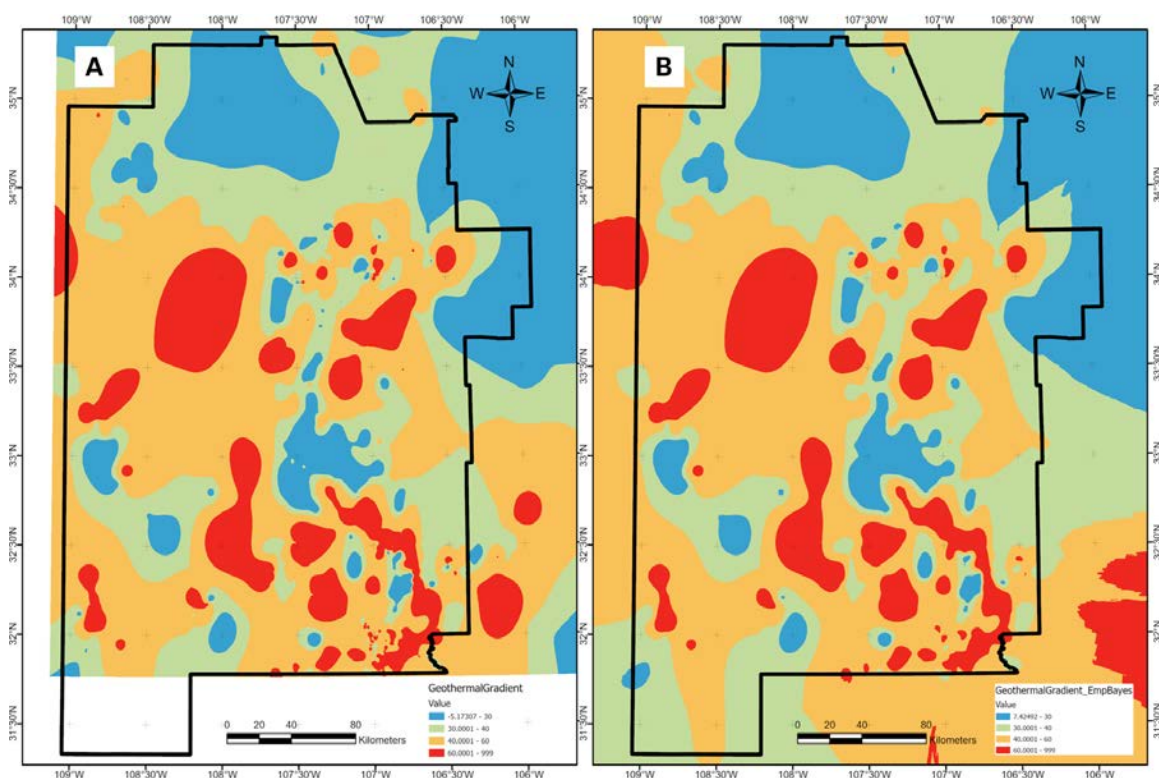


Figure A-30: Geothermal gradient data layer. Values are binned into four classes using the ranges in Section 3.2.2. Units for the class ranges are K/km. A. Raster originally created by [Bielicki et al. \(2015\)](#). B. Extrapolation performed using the ArcGIS *Empirical Bayes Kriging* to fill in the southwestern data gap.

Appendix B

Cost Model Spreadsheets

B.1 Static NPV Model

The static model described in Section 4.1 was implemented in Microsoft Excel as a single worksheet for cost analysis. Figures B-1 and B-2 show the model when flow rate is pre-defined and capacity depends on the input temperature of the produced brine. Not shown is the supporting look-up table for the EIA STEO-based electricity price forecast (Figure 4-3).

RESOURCE	VALUE	UNITS	REFERENCE
Surface Temperature (2020)	15.8	degrees	Dahal, 2012
Average Geothermal Gradient	100	K/km	Crowell, 2014
Initial Average Reservoir Temperature	149	degrees	Schochet, 2001
Average Well Depth (near vertical, MD)	1.332	km	
Production Well Temperature Loss	11	degrees	Lowry, 2017
Production Temperature (@WH)	138	degrees	
Water Loss Rate	2%	% of injected water	Freeman, 2018 (SAM)
Production Flow Rate per Module (7 units)	40	kg/s	GETEM
CAPEX (per module)	VALUE	UNITS	REFERENCE
Drilling & Completions Costs	\$ 1,797,145	USD (1 well)	Beckers, 2013
Wells per Module	2	well count per unit	
Surface Plant Costs	2000	\$/kW (e)	Beckers 2013
Reservoir Stimulation per injection well	\$ 1,250,000	USD	Lowry, 2017
Fluid Distribution Costs	\$ 425,173	USD	Beckers 2013
Redevelopment Factor	0.85		pers. comm., Prestidge
Thermal Drawdown Threshold	16.7	degrees	GETEM
Thermal Drawdown Rate	0.5%		GETEM
Redevelop Every	27	years	
Exploration Success Rate	100.0%		
Total Capital Costs (exploration)	\$ 2,486,084	USD	Beckers, 2013
Total Capital Costs (drilling)	\$ 3,594,290	USD	Beckers, 2013
Total Capital Costs (non-drilling)	\$ 5,942,834	USD	Beckers, 2013
POWER PLANT (Binary ORC modules)	VALUE	UNITS	REFERENCE
Plant Type	Binary ORC		
Plant Useful Life	30	years	Augustine, 2009
Units per module	7		
Output per Unit	150	kW	
Expected Net Production	1050	kW	
Inlet Temperature (brine)	138	degrees	
Net Brine Effectiveness (be)	7.45	59153	w-hr/lb, w-s/kg; GETEM
Parasitic Pumping	120	kW	GETEM, 2006
Required Flow Per Module	40	kg/s	
Capacity Factor	95%		Glassly 2015
Degradation Factor	0.5%		NREL, 2002
Avg Net Power Output per Unit	2.134	MW (e)	

Figure B-1: First part of static NPV cost model spreadsheet for the geothermal expansion project. Values in gray cells with orange font are calculated using inputs from the rest of the sheet.

OPEX	VALUE	UNITS	REFERENCE
Labor (per module)	\$ 325,813	USD	Beckers 2013
Power Plant Ops & Maintenance (per module)	\$ 308,375	USD	Beckers 2013
Field Ops & Maintenance (per module)	\$ 117,396	USD	Beckers 2013
Water Ops & Maintenance	\$ 6,174	USD	GETEM model
Total Annual O&M costs (per module)	\$ 431,945	USD	
FACTORS/INDICES	VALUE	UNITS	REFERENCE
Price Index from for Q4 2004 to 2020 USD	143%		PPI electric sector
Price Index from for Q4 2009 to 2020 USD	104%		PPI electric sector
Price Index from for Q4 2012 to 2020 USD	107%		PPI electric sector
Employment Cost Index (Utilities) compared to 200	122%		BLS
Discount rate	7%		GETEM, 2012
Learning rate	9%		Lukawski 2014
Contract rate above wholesale	50%		PNM, 2014
Learning rate exponent	-0.136		
Calendar Year	2020	2021	2022
Nominal Year	0	1	2
Calculated Price	0.086	0.090	0.000
Price (\$/kWh)	0.086	0.090	0.090
Capacity Level Increase [Input]	2	1	0
Unit Count	2	3	3
Wells Drilled or Redrilled	4	6	6
Total Capacity Increase (kWh)	37384709	18692354	0
Total Undegraded Capacity (kWh)	0	37384709	56077063
Declined Wellhead Temperature (degrees)	138	137	136
Updated Brine Effectiveness (w-s/kg)	59153	58009	56881
Degraded Capacity (kWh)	0	36622759	53807472
Power Plant Degradation Factor	1.000	1.000	0.995
Total Overall Capacity (kWh)	0	36622759	53538435
Revenue	\$0.00	\$3,285,061.49	\$4,802,397.62
CAPEX (Drilling)	\$7,188,579.83	\$2,976,431.48	\$0.00
CAPEX (Other)	\$14,371,751.75	\$5,942,833.98	\$0.00
OPEX	\$0.00	\$863,890.79	\$1,295,836.18
Cashflow	\$0.00	(\$6,498,094.76)	\$3,506,561.44
DCF	\$0.00	(\$6,072,985.75)	\$3,062,766.56
Present value of cashflow	\$22,594,047		
Up-front investment	\$21,560,332		
Net present value	\$1,033,715		

Figure B-2: Second part of static NPV cost model spreadsheet for the geothermal expansion project. Parameters in gray cells with orange font are calculated using inputs from the rest of the sheet. The orange cells in the annual cash flow analysis are manual entry fields for constructing power-plant modules. The analysis only extends out to year 2 for visualization purposes, but continues to year 30 in the complete spreadsheet.

B.2 Probabilistic NPV Model

The probabilistic NPV model was implemented as an extension of the static NPV model in Excel, with variable look-ups using the PDFs described in Section 4.3.3.

Flexible design options described in Section 4.4 were implemented as decision rules in the cash flow analysis. Figures B-3 and B-4 illustrate the spreadsheet for the Full Flexibility case (see Section 4.4.3). The results histogram, target curve, and summary statistics were generated using a 2000-row data table tied to the NPV calculation cell (not shown).

RESOURCE	VALUE	UNITS	REFERENCE
Surface Temperature (2020)	15.8	degrees	Dahal, 2012
Average Geothermal Gradient	92	K/km	Crowell, 2014
Initial Average Reservoir Temperature	141	degrees	Schochet, 2001
Average Well Depth (near vertical, MD)	1.357	km	
Production Well Temperature Loss	11	degrees	Lowry, 2017
Production Temperature (@WH)	131	degrees	
Water Loss Rate	2%	% of injected water	Freeman, 2018 (SAM)
Production Flow Rate per Module (7 units)	40	kg/s	GETEM

CAPEX (per module)	VALUE	UNITS	REFERENCE
Drilling & Completions Costs	\$ 2,276,729	USD (1 well)	
Wells per Module	2	well count per unit	
Surface Plant Costs	2000	\$/kW (e)	Beckers 2013
Reservoir Stimulation per injection well	\$ 1,250,000	USD	Lowry, 2017
Fluid Distribution Costs	\$ 364,701	USD	Beckers 2013
Redevelopment Factor	0.85		pers. comm., Prestidge
Thermal Drawdown Threshold	15.3	degrees	GETEM
Thermal Drawdown Rate	0.84%		
Redevelop Every	15	years	
Exploration Success Rate	100.0%		
Total Capital Costs (exploration)	\$ 2,830,296	USD	Beckers, 2013
Total Capital Costs (drilling)	\$ 4,553,458	USD	Beckers, 2013
Total Capital Costs (non-drilling)	\$ 5,025,234	USD	Beckers, 2013

POWER PLANT (modules)	VALUE	UNITS	REFERENCE
Plant Type	Binary ORC		
Plant Useful Life	30	years	Augustine, 2009
Units per module	7		
Output per Unit	150	kW	
Expected Net Production	1050	kW	
Inlet Temperature (brine)	131	degrees	
Net Brine Effectiveness (be)	6.03	47875	GETEM, 2006
Parasitic Pumping	120	kW	GETEM, 2006
Required Flow Per Module	40	kg/s	
Capacity Factor	95%		Glassly 2015
Degradation Factor	0.5%		NREL, 2002
Avg Net Power Output per Unit	1.71	MW (e)	

OPEX	VALUE	UNITS	REFERENCE
Labor (per module)	\$ 325,813	USD	GETEM
Power Plant Ops & Maintenance (per module)	\$ 295,518	USD	Beckers 2013
Field Ops & Maintenance (per module)	\$ 126,988	USD	Beckers 2013
Water Ops & Maintenance	\$ 6,174	USD	GETEM
Total Annual O&M costs (per module)	\$ 428,680	USD	

Figure B-3: First part of probabilistic NPV cost model spreadsheet for the geothermal expansion project. Values in gray cells with orange font are calculated using inputs from the sheet or distributions in other worksheets. PDF look-ups are implemented for Average Geothermal Gradient, Initial Average Reservoir Temperature, Drilling & Completions Costs, Thermal Drawdown Rate, and Price Forecast.

FACTORS/INDICES	VALUE	UNITS	REFERENCE
Price Index from for Q4 2004 to 2020 USD	143%		PPI electric sector
Price Index from for Q4 2009 to 2020 USD	104%		PPI electric sector
Price Index from for Q4 2012 to 2020 USD	107%		PPI electric sector
Employment Cost Index (Utilities) compared to 2004	122%		BLS
Discount rate	7%		GETEM, 2012
Learning rate	9%		Lukawski 2014
Contract rate above wholesale	50%		PNM, 2014
Learning rate exponent	-0.136		
Calendar Year	2020	2021	2022
Nominal Year	0	1	2
Price Forecast (\$/kWh)	0.057	0.060	0.060
Contract Price (\$/kWh)	0.086	0.090	0.090
Price Deviance (Wholesale)	0.0%	0.0%	0.3%
Capacity Level Increase [Input]	2	1	0
Capacity Level Change [Auto]	0	0	0
Auto Renegotiated Price (\$/kWh)	0.000	0.000	0.000
Unit Count	2	3	3
Decommission Flag	0	0	0
Potential Redevelopment	0	0	0
Redevelopment Balance	0	0	0
Units Redeveloped	0	0	0
Wells Drilled or Redrilled	4	6	6
Total Capacity Change (kWh)	29876266	14938133	0
Total Undegraded Capacity (kWh)	0	29876266	44814400
Years Since Development	0	1	2
Declined Wellhead Temperature (degrees)	131	130	129
Updated Brine Effectiveness (w-s/kg)	47875	46247	44662
Degraded Capacity (kWh)	0	28792385	41605101
Power Plant Degradation Factor	1.000	0.995	0.990
Total Overall Capacity (kWh)	0	28648423	41190090
Revenue	\$0.00	\$2,569,763.51	\$3,694,751.04
CAPEX (Drilling)	\$9,106,916.22	\$3,770,718.66	\$0.00
CAPEX (Redevelopment)	\$0.00	\$0.00	\$0.00
CAPEX (Other)	\$12,880,764.03	\$5,025,234.01	\$0.00
OPEX	\$0.00	\$857,360.31	\$1,286,040.46
Cashflow	\$0.00	(\$7,083,549.47)	\$2,408,710.58
DCF	\$0.00	(\$6,620,139.69)	\$2,103,861.10
Present value of cashflow	\$14,898,319		
Up-front investment	\$21,987,680		
Net present value	-\$7,089,362		

Figure B-4: Second part of probabilistic NPV cost model spreadsheet for the geothermal expansion project. Parameters in gray cells with orange font are calculated using inputs from the rest of the sheet. PDF look-ups are implemented for Average Geothermal Gradient, Initial Average Reservoir Temperature, Drilling & Completions Costs, Thermal Drawdown Rate, and Price Forecast. The orange cells in the annual cash flow analysis are manual entry fields for constructing power-plant modules. Decision rules are implemented in the annual cash flow section. The yearly breakdown of cost and revenue only extends out to year 2 for visualization purposes, but continues to year 30 in the full spreadsheet.

Appendix C

Risk Mitigation Log

This appendix provides additional context behind the scores assigned to likelihood and consequence for geothermal power-generation project risks before and after mitigation as presented in Tables 7.3 and 7.4 and plotted in Figure 7-2 in Chapter 7. Each of the risks described below have proposed mitigations using the methods outlined in this thesis. Specifically, machine-learning techniques paired with uncertainty analysis can be applied to several risks in the exploration and appraisal phases of the geothermal field lifecycle (Figure 7.1). And multiple risks in the development and production phases of the geothermal lifecycle can be reduced using probabilistic cost models with decision rules for evaluating dynamic operational strategies over the lifetime of a facility.

Table C.1 describes the likelihood scores for the six project risks shown in Figure 7-2. The corresponding consequence scores are reviewed in Table C.2. Risk-mitigation actions reduce risk likelihood, consequence, or both. Table C.3 explains the impact the proposed mitigation strategies have on risk-likelihood scores, while Table C.4 steps through changes to individual consequence scores. Likelihood-score definitions follow guidelines adapted from NASA as listed in Table 7.1. Similarly, consequence scores follow the descriptions in Table 7.2.

ID	Description	Likelihood	Score	Explanation
EXP1	Insufficient exploration budget	Highly Likely	4	Primarily due to high cost of gathering sufficient data sets covering risk elements and drilling of exploration wells. Exploration has 31% success rate (Doughty et al., 2018).
EXP2	Poor subsurface characterization	Highly Likely	4	Comprehensive data availability tends to be poor. Seismic is expensive and not as diagnostic for structural elements in geothermal. Well data available, but provides single-borehole views of complex 3D systems.
DRL2	Drilling cost overruns	Likely	3	Drill bits wearing out on hard rock and heat failure of equipment can require tripping and delays, equating to additional rig and equipment costs.
PRD1	Insufficient production budget	Likely	3	Production costs depend on flow rate/pumps, thermal drawdown and well recompletion, over-estimates of resource temperature, etc., leading to lower efficiency and/or additional costs.
PRD5	Demand variability	Low Likelihood	2	Large fraction of demand increases due to greater electrification likely be absorbed by other competitive generating technologies (e.g., natural gas, solar, wind). Decrease in electricity demand is unlikely.
PRD6	Wrong-sized infrastructure	Likely	3	Over-estimation of accessible resource can and has led to over-construction of surface facilities. For EGS, this extends to unsustainable rates of heat extraction and enhanced thermal drawdown.

Table C.1: Risk likelihood scores and score explanations for a subset of potential risks in a geothermal power-generation project, primarily in the exploration and production phases of a field lifecycle. Likelihood scores correspond with the score rubric listed in Table 7.1.

ID	Description	Consequence	Score	Explanation
EXP1	Insufficient exploration budget	Medium	3	Will impact comprehensive exploration activities, so project may be sub-optimal or opportunities missed. Impacts are cost and performance, but project may stay on schedule.
EXP2	Poor subsurface characterization	Medium-High	4	Equates to poor understanding of the risk elements (heat, permeability, seal, fluids), each of which could completely derail the project in performance, or in cost or schedule from addressing unexpected conditions.
DRL2	Drilling cost overruns	Medium	3	Multi-well drilling can be split across multiple years to manage FY budget. Depending on reservoir enthalpy, shallower wells could be drilled. Overall, impact on cost, but potentially on performance and schedule.
PRD1	Insufficient production budget	Medium-High	4	Underfunded production costs impacts both cost and performance. Reduced power production. Cuts into project revenue, possibly making project uneconomic. Could also require renegotiation of PPA.
PRD5	Demand variability	Medium-High	4	Missed opportunities for higher demand, but not much risk to the project. Lower demand due to competitive pressures or removal of dedicated carve-outs from state RPS policies could sink a geothermal plant, particularly if subsidies remain lopsided toward solar and wind and natural gas doesn't face abatements like carbon taxes.
PRD6	Wrong-sized infrastructure	Medium-High	4	Under-sized facilities miss an opportunity to produce more power. Oversized facilities may be unprofitable and require hybrid energy retrofitting, e.g. Stillwater, Nevada. Biggest impacts are on both performance and cost.

Table C.2: Risk-consequence scores and score explanations for a subset of potential risks in a geothermal power-generation project, primarily in the exploration and production phases of a field lifecycle. Consequence scores correspond with the score rubric listed in Table 7.2.

ID	Mitigation	Updated Likelihood	Score	Explanation
EXP1	Use ML model predictions to reduce costs in exploration	Low Likelihood	2	ML modeling reduces risk of exploration-well failures and focuses data-acquisition expenditures on the most important data sets to acquire. Costs are lower and more predictable.
EXP2	Use ML for baseline model, acquire data based on importances	Low Likelihood	2	ML models integrate many different data sets for a combined assessment of the heat risk element and potentially including permeability, fluids, and seal, for a more complete assessment of reservoir potential.
DRL2	Use ML to identify high-gradient areas faster, shallower drilling	Low Likelihood	2	Focusing on high-gradient areas can reduce overall drill depths. Modeling could also target specific rock types to create a map of lithologic complexity or bedrock hardness to better inform the drillers.
PRD1	Cost modeling for optimized spending and return	Not Likely	1	Economic models that incorporate uncertainty and validated inputs will set realistic bounds on costs both up-front and in the future. Building and revisiting these models will inform budgetary planning and reduce the risk of overruns.
PRD5	Flexibility in power generation based on market	Low Likelihood	2	Cost models that include decision rules enable testing of field operational strategies, including growth and reduction using plant modularity. With an appropriate range of demand models, project managers can optimize the power plant size to meet variability without large losses.
PRD6	Flexibility in design for demand-triggered capacity changes	Not Likely	1	Cost models that include flexibility and decision rules allow project managers to examine different build-out schedules that could be sensitive to resource viability. Machine learning can also help in predicting the resource grade, further constraining the surface-facility needs.

Table C.3: Updated risk-likelihood scores and score explanations for a subset of potential risks in a geothermal power-generation project. Score updates reflect the impact of proposed mitigation actions. Likelihood scores correspond with the score rubric listed in Table 7.1.

ID	Mitigation	Updated Consequence	Score	Explanation
EXP1	Use ML model predictions to reduce costs in exploration	Medium-Low	2	Use of ML models can better constrain data-acquisition needs and derisk further drilling activities, leading to minor cost consequences on project if the budget is exceeded.
EXP2	Use ML for baseline model, acquire data based on importances	Medium	3	Poor subsurface characterization still impacts performance, cost, and schedule post-mitigation, but to a lesser extent if ML-based uncertainty measures are used to screen out the most at-risk areas for failure.
DRL2	Use ML to identify high-gradient areas faster, shallower drilling	Medium	3	If drilling cost overruns still occur, the same consequences will apply —impact felt in cost itself, but also in project performance from holes that TD too shallow, or schedule due to required approvals and other delays.
PRD1	Cost modeling for optimized spending and return	Medium	3	Cost models will not always be correct, and decisions may be made on the high side while reality follows the low side. Nevertheless, early awareness of the scenario ranges enables mitigation behaviors earlier than otherwise would be true.
PRD5	Flexibility in power generation based on market	Medium	3	Sudden threshold behavior in demand may occur without warning. Models can pick this up as unlikely scenarios, but if they occur in reality, the impact will be felt as project losses. With appropriate modeling, these losses can be mitigated more than for blind project execution.
PRD6	Flexibility in design for demand-triggered capacity changes	Medium	3	Economic models will steer project managers toward less risky initial investments, protecting the operator from heavier losses if the accessible resource is below expectations. But the impact of overly-large infrastructure will still hit project performance and cost, just with lesser losses because of a mitigation mentality.

Table C.4: Updated risk consequence scores and score explanations for a subset of potential risks in a geothermal power-generation project. Score updates reflect the impact of proposed mitigation actions. Consequence scores correspond with the score rubric listed in Table 7.2.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.
- Allegre, C. J., Manhès, G., & Göpel, C. (1995). The age of the Earth. *Geochimica et Cosmochimica Acta*, 59(8), 1445–1456. (Publisher: Elsevier)
- Anderson, A., Blackwell, D., Chikering, C., Boyd, T., Horne, R., MacKenzie, M., ... Shevenell, L. A. (2013). *National Geothermal Data System (Ngds) Geothermal Data Domain: Assessment Of Geothermal Community Data Needs* (Tech. Rep.). Boise State Univ., ID (United States). Retrieved 2021-06-04, from <https://www.osti.gov/biblio/1130333>
- Armstead, H. C. H., & Tester, J. W. (1987). *Heat mining: a new source of energy*. Spon Press.
- Augustine, C. (2009). *Hydrothermal spallation drilling and advanced energy conversion technologies for engineered geothermal systems* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Augustine, C., Ho, J., & Blair, N. (2019, May). *GeoVision Analysis Supporting Task Force Report: Electric Sector Potential to Penetration* (Tech. Rep. No. NREL/TP-6A20-71833). National Renewable Energy Lab. (NREL), Golden, CO (United States). Retrieved 2021-04-10, from <https://www.osti.gov/biblio/1524768/> doi: 10.2172/1524768
- Augustine, C., Tester, J. W., Anderson, B., Petty, S., & Livesay, B. (2006). A comparison of geothermal with oil and gas well drilling costs. In (pp. 5–19). Curran Associates Inc New York, New York.
- Baird, D. C. (1962). *Experimentation: an introduction to measurement theory and experiment design*. Prentice Hall.
- Bankey, V., Cuevas, A., Daniels, D., Finn, C. A., Hernandez, I., Hill, P., ... Roberts, C. (2002). *Digital data grids for the magnetic anomaly map of North America* (Open-File Report No. 2331-1258). Reston, VA: USGS. Retrieved from <https://pubs.usgs.gov/of/2002/ofr-02-414/>

- Beaudette, D. (2020, February). *Accuracy and Uncertainty for Categorical Predictions*. Retrieved from http://ncss-tech.github.io/stats_for_soil_survey/chapters/9_uncertainty/class-accuracy-uncertainty.html
- Beckers, K. (2016). *Low-temperature geothermal energy: systems modeling, reservoir simulation, and economic analysis* (Doctoral dissertation, Cornell University, Ithaca, NY). Retrieved from <https://ecommons.cornell.edu/handle/1813/44328>
- Beckers, K., Lukawski, M., Reber, T., Anderson, B., Moore, M., & Tester, J. W. (2013). Introducing GEOPHIRES v1.0: Software Package for Estimating Levelized Cost of Electricity and/or Heat from Enhanced. In *Geothermal Systems, Proceedings, Thirty-Eighth Workshop on Geothermal Reservoir Engineering*.
- Beckers, K., & McCabe, K. (2019, February). GEOPHIRES v2.0: updated geothermal techno-economic simulation tool. *Geothermal Energy*, 7(1), 5. Retrieved 2021-04-07, from <https://doi.org/10.1186/s40517-019-0119-6> doi: 10.1186/s40517-019-0119-6
- Bertsimas, D., O'Hair, A., & Pulleyblank, W. (2016). *The Analytics Edge*. Belmont, Massachusetts: Dynamic Ideas LLC.
- Bielicki, J., Blackwell, D., Harp, D., Karra, S., Kelley, R., Kelly, S., ... Sutula, G. (2015). *Hydrogeologic windows: Regional signature detection for blind and traditional geothermal play fairways, Final Reports, Los Alamos National Laboratory* (Tech. Rep.). Los Alamos, NM: LA-UR-15-28360.
- Blackwell, D., Golm, M., Cutright, B., Gosnold, W., Kay, M., Nagihara, S., ... Tester, J. W. (2014, June). *Geothermal Data Aggregation Submission of Information into the National Geothermal Data System* (Final Report No. DOE Project DE-EE0002852). Washington, D.C: United States. Dept. of Energy. Retrieved 2021-06-16, from <http://www.osti.gov/servlets/purl/1135820> (Medium: electronic resource)
- Blackwell, D., Richards, M., Frone, Z., Batir, J., Ruzo, A., Dingwall, R., & Williams, M. (2011). *Temperature-At-Depth Maps For the Conterminous US and Geothermal Resource Estimates* (Tech. Rep. No. GRC1029452). Southern Methodist University Geothermal Laboratory, Dallas, TX (United States). Retrieved 2020-12-07, from <https://www.osti.gov/biblio/1137036>
- Blackwell, D., Steele, J., & Carter, L. (1990). *Heat flow patterns of the North American continent: A discussion of the DNAG Geothermal Map of North America* (Tech. Rep.). DOE/EEGTP (USDOE Office of Energy Efficiency and Renewable Energy Geothermal Tech Pgm). Retrieved 2021-06-04, from <https://www.osti.gov/biblio/896322> doi: 10.2172/896322
- Blackwell, D., Wisian, K., & Steele, J. (1995). Geothermal regime in the central and eastern United States east of the Rocky Mountains. *Proc. World Geothermal Cong*, 649–653.

- Blair, N., DiOrio, N., Freeman, J., Gilman, P., Neises, T., & Wagner, M. (2018). *System Advisor Model (SAM) general description (version 2017.9. 5)* (Technical Report No. NREL/TP-6A20-70414). Golden, CO, USA: National Renewable Energy Lab.(NREL). Retrieved from [doi:10.2172/1440404](https://doi.org/10.2172/1440404)
- Blessing, E. (2021, January). *What Happened to Oil Prices in 2020*. Retrieved 2021-05-26, from <https://www.investopedia.com/articles/investing/100615/will-oil-prices-go-2017.asp>
- Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural network. In (pp. 1613–1622). PMLR.
- Bonafin, J., & Dickey, H. K. (2019). The Repowering of the Lightning Dock Geothermal Plant in New Mexico. *Geothermal Resources Council Bulletin*, 48(6), 34–40.
- Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces* (Vol. 424). Wiley New York.
- BP. (2020, August). *From International Oil Company to Integrated Energy Company: bp sets out strategy for decade of delivery towards net zero ambition / News and insights / Home* [Press Release]. Retrieved 2021-03-06, from <https://www.bp.com/en/global/corporate/news-and-insights/press-releases/from-international-oil-company-to-integrated-energy-company-bp-sets-out-strategy-for-decade-of-delivery-towards-net-zero-ambition.html>
- Breede, K., Dzebisashvili, K., Liu, X., & Falcone, G. (2013, November). A systematic review of enhanced (or engineered) geothermal systems: past, present and future. *Geothermal Energy*, 1(1), 4. Retrieved 2021-03-20, from <https://doi.org/10.1186/2195-9706-1-4> doi: 10.1186/2195-9706-1-4
- Bremmer, I. (2010, May). The Long Shadow of the Visible Hand. *Wall Street Journal*. Retrieved 2021-07-26, from <https://www.wsj.com/articles/SB10001424052748704852004575258541875590852>
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1), 1–3.
- Brown, S. C. (2021). *personal communication*. (Great Basin Machine Learning Project)
- Brown, S. C., Coolbaugh, M. F., DeAngelo, J., Faulds, J. E., Fehler, M., Gu, C., ... Mlawsky, E. (2020). Machine learning for natural resource assessment: An application to the blind geothermal systems of Nevada. In *Geothermal Resources Council Transactions* (Vol. 44, p. 13). Nevada.
- Brownlee, J. (2017, July). *Gentle Introduction to the Adam Optimization Algorithm for Deep Learning*. Retrieved 2021-06-30, from <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>

- Brownlee, J. (2019a, January). *A Gentle Introduction to the Rectified Linear Unit (ReLU)*. Retrieved 2021-06-29, from <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/>
- Brownlee, J. (2019b, January). *How to Control the Stability of Training Neural Networks With the Batch Size*. Retrieved 2021-06-30, from <https://machinelearningmastery.com/how-to-control-the-speed-and-stability-of-training-neural-networks-with-gradient-descent-batch-size/>
- Brownlee, J. (2020a, January). *How to Fix k-Fold Cross-Validation for Imbalanced Classification*. Retrieved 2021-06-23, from <https://machinelearningmastery.com/cross-validation-for-imbalanced-classification/>
- Brownlee, J. (2020b, April). *One-vs-Rest and One-vs-One for Multi-Class Classification*. Retrieved 2021-06-23, from <https://machinelearningmastery.com/one-vs-rest-and-one-vs-one-for-multi-class-classification/>
- Cappetti, G., Fiordelisi, A., Casini, M., Ciuffi, S., & Mazzotti, A. (2005). A new deep exploration program and preliminary results of a 3D seismic survey in the Larderello-Travale geothermal field (Italy). In (pp. 24–29).
- Cavur, M., Moraga, J., Duzgun, H. S., Soydan, H., & Jin, G. (2021, February). The DInSAR Analysis with Machine Learning for Delineating Geothermal Sites at the Brady Geothermal Field. In *46th Workshop on Geothermal Reservoir Engineering*. Stanford, California.
- Chen, T. (2021, June). *dmlc/xgboost*. Distributed (Deep) Machine Learning Community. Retrieved 2021-06-26, from <https://github.com/dmlc/xgboost> (original-date: 2014-02-06T17:28:03Z)
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In (pp. 785–794).
- Chevron. (2021, March). *Chevron Reinforces Plan to Deliver Higher Returns, Lower Carbon*. Retrieved 2021-05-27, from <https://www.businesswire.com/news/home/20210309005302/en/Chevron-Reinforces-Plan-to-Deliver-Higher-Returns-Lower-Carbon>
- Climeon. (2018). *Why low temperature geothermal heat power will become bigger than high temperature geothermal heat power?* Reykjavik. Retrieved 2021-05-01, from <https://www.slideshare.net/IcelandGeothermal/a3-why-low-temperature-geothermal-heat-power-will-become-bigger-than-high-temperature-geothermal-heat-power>
- Climeon. (2021). *The Climeon Heat Power System - How does it work?* Retrieved 2021-05-01, from <https://climeon.com/how-it-works/>

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46. (Publisher: Sage Publications Sage CA: Thousand Oaks, CA)
- ConocoPhillips. (2020, October). *ConocoPhillips Adopts Paris-Aligned Climate Risk Framework to Meet Net-Zero Operational Emissions Ambition by 2050* [Press Release]. Retrieved 2021-03-07, from <https://www.conocophillips.com/sustainability/sustainability-news/story/conocophillips-adopts-paris-aligned-climate-risk-framework-to-meet-net-zero-operational-emissions-ambition-by-2050/>
- Crawley, E., Cameron, B., & Selva, D. (2015). *System architecture: strategy and product development for complex systems*. Prentice Hall Press.
- Crowell, J., & Crowell, A. (2014). The History of Lightning Dock KGRA: Identifying a Blind Geothermal Resource. *GRC Transactions*, 38, 8.
- Cunniff, R. A., & Bowers, R. L. (2003, December). *FINAL REPORT ENHANCED GEOTHERMAL SYSTEMS TECHNOLOGY PHASE II ANIMAS VALLEY, NEW MEXICO* (Tech. Rep. No. DOE/ID/14203). LIGHTNING DOCK GEOTHERMAL INC. (US). Retrieved 2021-04-15, from <https://www.osti.gov/biblio/820539> doi: 10.2172/820539
- Cunniff, R. A., & Bowers, R. L. (2005, August). *Final Technical Report; Geothermal Resource Evaluation and Definition (GRED) Program-Phases I, II, and III for the Animas Valley, NM Geothermal Resource* (Tech. Rep. No. DOE-AL66977). Lightning Dock Geothermal, Inc., Las Cruces, NM. Retrieved 2021-04-10, from <https://www.osti.gov/biblio/850199> doi: 10.2172/850199
- Dahal, S., McDonald, M. R., Bubach, B., & Crowell, A. M. (2012). Evaluation of geothermal potential of lightning dock KGRA, New Mexico. *Transactions of the Geothermal Resources Council*, 36, 637–640.
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., ... Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology: a Journal of the Royal Meteorological Society*, 28(15), 2031–2064. (Publisher: Wiley Online Library)
- Daso, F. (2020, January). *Eden GeoPower, An MIT Green Tech Startup, Improves America's Oil Production*. Retrieved 2021-07-28, from <https://www.forbes.com/sites/frederickdaso/2020/01/07/eden-geopower-an-mit-green-tech-startup-improves-americas-oil-production/> (Section: Enterprise Tech)
- Deloitte. (2020). 2021 Oil and Gas Industry Outlook. *Deloitte*. Retrieved from <https://www2.deloitte.com/us/en/pages/energy-and-resources/articles/oil-and-gas-industry-outlook.html>

- de Neufville, R., & Scholtes, S. (2011). *Flexibility in engineering design*. MIT Press.
- Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., ... Saurous, R. A. (2017). Tensorflow distributions. *arXiv preprint arXiv:1711.10604*.
- DiPippo, R. (2012). *Geothermal power plants: principles, applications, case studies and environmental impact*. Butterworth-Heinemann.
- DOE. (2021a, May). *A History of Geothermal Energy in America*. Retrieved 2021-05-29, from <https://www.energy.gov/eere/geothermal/history-geothermal-energy-america> (United States Department of Energy)
- DOE. (2021b, May). *Low Temperature Deep Direct-Use Program Draft White Paper*. Retrieved 2021-05-31, from <https://www.energy.gov/eere/geothermal/low-temperature-deep-direct-use-program-draft-white-paper>
- Doughty, C., Dobson, P. F., Wall, A., McLing, T., & Weiss, C. (2018). *GeoVision Analysis: Exploration Task Force Report* (Tech. Rep.). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States). Retrieved from <https://doi.org/10.2172/1457012>
- Doust, H. (2010, November). The exploration play: What do we mean by it? *AAPG Bulletin*, 94(11), 1657–1672. Retrieved 2021-02-26, from <https://pubs.geoscienceworld.org/aapgbull/article-abstract/94/11/1657/133030/The-exploration-play-What-do-we-mean-by-it> (Publisher: GeoScienceWorld) doi: 10.1306/06301009168
- DSIRE. (2021). *DSIRE*. Retrieved 2021-05-02, from <https://programs.dsireusa.org/system/program/detail/720>
- Durst, P. (2013, August). Overview of the Soultz geothermal project. In (p. 35). Pisa, Italy. Retrieved from http://www.geoelec.eu/wp-content/uploads/2013/07/GEOELEC_Soultz_Pise.pdf
- Duval, J., Carson, J., Holman, P., & Darnley, A. (2005). *Terrestrial Radioactivity and Gamma-ray Exposure in the United States and Canada* (Open-File Report No. 2005-1413). USGS. Retrieved 2021-06-14, from <https://pubs.usgs.gov/of/2005/1413/index.htm>
- EERE. (2012). *GETEM -Geothermal Electricity Technology Evaluation Model*. U.S. Department of Energy. Retrieved from <https://www.energy.gov/eere/geothermal/getem-manuals-and-revision-notes> (Office of Energy Efficiency & Renewable Energy)
- EERE. (2021). *Pumped-Storage Hydropower*. Retrieved 2021-05-27, from <https://www.energy.gov/eere/water/pumped-storage-hydropower>
- EERI. (2014, August). *Play Fairway Analysis*. Retrieved 2021-06-09, from <https://www.energy.gov/eere/geothermal/play-fairway-analysis>

- EIA. (2020a, October). *Electric Power Annual 2019*. Retrieved 2021-06-03, from <https://www.eia.gov/electricity/annual/>
- EIA. (2020b). *International Energy Outlook 2020 (IEO2020)* (Tech. Rep.). Government Printing Office. Retrieved from <https://www.eia.gov/outlooks/ieo/> (US Energy Information Administration Publication Title: International Energy Outlook Volume: 2020)
- EIA. (2021a, February). *Annual Energy Outlook 2021 (AEO2021)* (Technical Report). Government Printing Office. Retrieved from <https://www.eia.gov/outlooks/aeo/> (U.S. Energy Information Administration Publication Title: Annual Energy Outlook Volume: 2021)
- EIA. (2021b, July). *New Mexico - State Energy Profile Overview*. Retrieved 2021-07-08, from <https://www.eia.gov/state/index.php?sid=NM> (U.S. Energy Information Administration (EIA))
- EIA. (2021c). *Short-Term Energy Outlook (STEO)* (Tech. Rep.). Retrieved from <https://www.eia.gov/outlooks/steo/> (ISSN: 0140-6736 Issue: October 6 Publication Title: US EIA - Short-Term Energy Outlook)
- Elston, W. E., Deal, E. G., & Logsdon, M. J. (1983). *Geology and geothermal waters of Lightning Dock region, Animas Valley and Pyramid Mountains, Hidalgo County, New Mexico* (Tech. Rep.). Socorro, NM, USA: New Mexico Bureau of Mines and Mineral Resources.
- Entingh, D., Mines, G., Nix, G., Mansure, A., Bauer, S., Petty, S., & Livesay, B. (2006). Volume I—technical reference manual. *Geothermal electricity technology evaluation model (GETEM)*. Washington DC: US DoE and NREL. (Publisher: Citeseer)
- Equinor. (2020, February). *Equinor sets ambition to reduce net carbon intensity by at least 50% by 2050 - equinor.com* [Press Release]. Retrieved 2021-03-07, from <https://www.equinor.com/en/news/2020-02-06-climate-roadmap.html>
- ESRI. (2021a). *Empirical Bayes Kriging*. Retrieved 2021-06-17, from <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm>
- ESRI. (2021b). *Kernel Density*. Retrieved 2021-06-17, from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-kernel-density-works.htm>
- ESRI. (2021c). *Kriging*. Retrieved 2021-06-17, from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-kriging-works.htm>
- ESRI. (2021d). *Spline*. Retrieved 2021-06-17, from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-spline-works.htm>

- ESRI. (2021e). *Topo to Raster*. Retrieved 2021-06-17, from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-topo-to-raster-works.htm>
- ExxonMobil. (2021, April). *ExxonMobil Presentation Details Strategy to Grow Shareholder Value, Protect Dividend and Transition to Lower-Carbon Future*. Retrieved 2021-05-27, from <https://www.businesswire.com/news/home/20210427005455/en/ExxonMobil-Presentation-Details-Strategy-to-Grow-Shareholder-Value-Protect-Dividend-and-Transition-to-Lower-Carbon-Future>
- Faulds, J., & Hinz, N. (2015). Favorable tectonic and structural settings of geothermal systems in the Great Basin region, western USA: Proxies for discovering blind geothermal systems. Nevada Bureau of Mines and Geology, University of Nevada, Reno. (Issue: DOE-UNR-06731-02)
- Faulds, J., Hinz, N., Ayling, B., Coolbaugh, M., Glen, J., Siler, D., ... Hardwick, C. (2019). *Discovering Blind Geothermal Systems in the Great Basin Region: An Integrated Geologic and Geophysical Approach for Establishing Geothermal Play Fairways: All Budget Periods* (Tech. Rep.). Reno, NV: Nevada Bureau of Mines and Geology, University of Nevada, Reno.
- Faulds, J., Hinz, N., Coolbaugh, M., Sadowski, A., Shevenell, L., McConville, E., ... Siler, D. (2017). Progress report on the Nevada play fairway project: Integrated geological, geochemical, and geophysical analyses of possible new geothermal systems in the Great Basin region. Nevada Bureau of Mines and Geology, University of Nevada, Reno. (Issue: NoneDOE-UNR-06731-04)
- Fawcett, T. (2006, June). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. Retrieved 2021-06-22, from <https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X> doi: 10.1016/j.patrec.2005.10.010
- Fournier, R. (1977). Chemical geothermometers and mixing models for geothermal systems. *Geothermics*, 5(1-4), 41–50. (Publisher: Elsevier)
- Fowler, C. (2005). The Solid Earth. *The Solid Earth*, 500.
- Frankson, R., Kunkel, K., Stevens, L., & Easterling, D. (2019). New Mexico State Climate Summary. *NOAA Technical Report NESDIS*.
- Fraser, A. (2010). A regional overview of the exploration potential of the Middle East: a case study in the application of play fairway risk mapping techniques. In (Vol. 7, pp. 791–800). Geological Society of London. (Issue: 1)
- Frisch, W., Meschede, M., & Blakey, R. (2011). Continental graben structures. In W. Frisch, M. Meschede, & R. C. Blakey (Eds.), *Plate Tectonics: Continental Drift and Mountain Building* (pp. 27–41). Berlin, Heidelberg: Springer. Retrieved 2021-06-11, from https://doi.org/10.1007/978-3-540-76504-2_3 doi: 10.1007/978-3-540-76504-2_3

- Gao, K., Huang, L., Lin, R., Hu, H., Zheng, Y., & Cladohous, T. (2021, February). Delineating Faults at the Soda Lake Geothermal Field Using Machine Learning. In *PROCEEDINGS, 46th Workshop on Geothermal Reservoir Engineering*. Stanford, California.
- Garchar, L., Badgett, A., Nieto, A., Young, K., Hass, E., & Weathers, M. (2016). Geothermal play fairway analysis: phase I summary..
- GEM. (2014). *Strain Rate Model / GEM - Global Earthquake Model*. Retrieved 2021-06-14, from <https://storage.globalquakemodel.org/what/seismic-hazard/strain-rate-model/>
- GeoEnergy, T. (2021, March). *Saskatchewan geothermal project ready for final engineering*. Retrieved 2021-06-03, from <https://www.thinkgeoenergy.com/saskatchewan-geothermal-project-ready-for-final-engineering/>
- Gifford, J. S., & Grace, R. C. (2013, July). *CREST Cost of Renewable Energy Spreadsheet Tool: A Model for Developing Cost-Based Incentives in the United States; User Manual Version 4, August 2009 - March 2011 (Updated July 2013)* (Tech. Rep. Nos. NREL/SR-6A20-50374, 1010865). Golden, CO, USA: National Renewable Energy Lab. Retrieved 2021-07-22, from <http://www.osti.gov/servlets/purl/1010865/> doi: 10.2172/1010865
- Gifford, J. S., Grace, R. C., & Rickerson, W. H. (2011, May). *Renewable Energy Cost Modeling. A Toolkit for Establishing Cost-Based Incentives in the United States* (Tech. Rep. Nos. NREL/SR-6A20-51093, 1219218). Golden, CO, USA: National Renewable Energy Lab. Retrieved 2021-07-22, from <http://www.osti.gov/servlets/purl/1219218/> doi: 10.2172/1219218
- Gimond, M. (2021). *Intro to GIS and Spatial Analysis*. Retrieved 2021-06-18, from <https://mgimond.github.io/Spatial/>
- Glassley, W. E. (2015). *Geothermal energy: renewable energy and the environment*. CRC press.
- Goodman, N., & Smiley, C. (2013, September). *Lightning Dock Geothermal* [Legislative Committee]. Deming, NM, USA. Retrieved from <https://www.nmlegis.gov/handouts/STTC%20091312%20Item%200%20Cryq%20Energy%20Presenation.pdf>
- Gunderson, K. L., Holmes, R. C., & Loisel, J. (2020). Recent digital technology trends in geoscience teaching and practice. *GSA Today*, 30(1). doi: 10.1130/GSATG404GW.1
- Hadi, J., Quinlivan, P., Ussher, G., Alamsyah, O., Pramono, B., & Masri, A. (2010). Resource risk assessment in geothermal greenfield development; an economic implications. In *Proceedings World Geothermal Congress 2010*. Bali, Indonesia.
- Hallett, K. (2010). *Open Energy Info (OpenEI)*. NREL. Retrieved 2021-07-25, from <https://www.nrel.gov/docs/fy11osti/48798.pdf> (openei.org)

- Hamm, S. G., Anderson, A., Blankenship, D., Boyd, L. W., Brown, E. A., Frone, Z., ... Winick, J. (2021, February). Geothermal Energy R&D: An Overview of the U.S. Department of Energy's Geothermal Technologies Office. *Journal of Energy Resources Technology*, 143(100903). Retrieved 2021-05-31, from <https://doi.org/10.1115/1.4049581> doi: 10.1115/1.4049581
- Hamm, S. G., Augustine, C. R., Tasca, C., & Winick, J. (2019). *An Overview of the US Department of Energy's GeoVision Report* (Tech. Rep.). National Renewable Energy Lab. Retrieved from <https://www.energy.gov/eere/geothermal/geovision> (Publisher: National Renewable Energy Lab.(NREL), Golden, CO (United States))
- HARC. (2021). *Winter Storm Uri's Impacts & Pathways to Resilience in Texas*. Retrieved 2021-05-26, from https://experience.arcgis.com/experience/cc48fcfebfae414b99b3d18f86c72c27/page/page_95/?views=view_10
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Henry, C. D., & Aranda-Gomez, J. J. (1992). The real southern Basin and Range: Mid-to late Cenozoic extension in Mexico. *Geology*, 20(8), 701–704. (Publisher: Geological Society of America)
- Herzog, H. J., Tester, J. W., & Frank, M. G. (1997). Economic analysis of heat mining. *Energy sources*, 19(1), 19–33. (Publisher: Taylor & Francis)
- Hohmeyer, O. (2008). IPCC scoping meeting on renewable energy sources proceedings. *IPCC*. Retrieved from <https://www.ipcc.ch/publication/ipcc-scoping-meeting-on-renewable-energy-sources/>
- Holl, H.-G. (2015). *What did we learn about EGS in the Cooper Basin?* doi: 10.13140/RG.2.2.33547.49443
- Holmes, R. C. (2009). *Oceanic crustal structure from two-dimensional active source traveltimes tomography* (Vol. 70). (Issue: 08)
- IHS. (2021, August). *IHS Markit - Leading Source of Critical Information*. Retrieved 2021-08-17, from <https://ihsmarkit.com/index.html>
- IPCC. (2018). Global warming of 1.5°C. *An IPCC Special Report on the impacts of global warming of 1.5 °C*, 1, 1–9. Retrieved from <https://www.ipcc.ch/sr15/>
- IRENA. (2021, May). *Country Rankings*. Retrieved 2021-06-01, from <https://www.irena.org/Statistics/View-Data-by-Topic/Capacity-and-Generation/Country-Rankings>
- Jain, A. (2016, March). *XGBoost Parameters / XGBoost Parameter Tuning*. Retrieved 2021-06-26, from <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.
- Jelacic, A., Fortuna, R., LaSala, R., Nathwani, J., Nix, G., Visser, C., ... Kennedy, M. (2008, April). *An Evaluation of Enhanced Geothermal Systems Technology* (Tech. Rep. No. 1219317). Washington, D.C.: Office of Energy Efficiency and Renewable Energy (EERE). Retrieved from <https://www.osti.gov/biblio/1219317>
- Jordan, T. E., Richards, M. C., Horowitz, F. G., Camp, E., Smith, J. D., Wheaton, C. A., ... Tester, J. W. (2016). *Low Temperature Geothermal Play Fairway Analysis For The Appalachian Basin: Phase 1 Revised Report November 18, 2016* (Tech. Rep.). Ithaca, NY: Cornell University.
- Jousset, P., Haberland, C., Bauer, K., & Arnason, K. (2011, March). Hengill geothermal volcanic complex (Iceland) characterized by integrated geophysical observations. *Geothermics*, 40(1), 1–24. Retrieved 2020-12-17, from <http://www.sciencedirect.com/science/article/pii/S0375650511000034> doi: 10.1016/j.geothermics.2010.12.008
- Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(4), 1–21. (Publisher: ACM New York, NY, USA)
- Kehle, R. O., Schoepel, R. J., & Deford, R. K. (1970, January). The AAPG geothermal survey of North America. *Geothermics*, 2, 358–367. Retrieved 2021-06-04, from <https://www.sciencedirect.com/science/article/pii/0375650570900349> doi: 10.1016/0375-6505(70)90034-9
- Keller, G. R., & Baldrige, W. S. (1999, January). The Rio Grande rift : A geological and geophysical overview. *Rocky Mountain Geology*, 34(1), 121–130. Retrieved 2021-06-11, from <https://doi.org/10.2113/34.1.121> doi: 10.2113/34.1.121
- Keller, G. R., Khan, A., Morganc, P., Wendlandt, R., Baldrige, S., Olsen, K., ... Braile, L. W. (1991, October). A comparative study of the Rio Grande and Kenya rifts. *Tectonophysics*, 197(2), 355–371. Retrieved 2021-02-11, from <https://www.sciencedirect.com/science/article/pii/0040195191900503> doi: 10.1016/0040-1951(91)90050-3
- Kelley, R. (2015). *Geothermal Data Repository (GDR)*. Retrieved 2021-06-14, from <https://gdr.openei.org/>
- Kingma, D. P., & Ba, J. (2017, January). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. Retrieved 2021-06-30, from <http://arxiv.org/abs/1412.6980> (arXiv: 1412.6980)
- Kreemer, C. (2020, April). *Global Strain Rate Model v.2.1*. Retrieved 2021-06-14, from <http://geodesy.unr.edu/GSRM/>

- Kreemer, C., Blewitt, G., & Klein, E. C. (2014). A geodetic plate motion and Global Strain Rate Model. *Geochemistry, Geophysics, Geosystems*, 15(10), 3849–3889. Retrieved 2021-02-11, from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GC005407> (_eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014GC005407>) doi: <https://doi.org/10.1002/2014GC005407>
- LANL, Kron, A., Wohletz, K., & Tubb, J. (1991). *Geothermal gradient contour map of the United States*. US Department of Energy, Los Alamos National Laboratory.
- Larson, E. (2009, September). *AltaRock suspends Geysers drilling project*. Retrieved 2021-07-21, from <https://www.lakeconews.com/component/content/article/142-local-government/10159-altarock-suspends-geysers-drilling-project->
- Larson, J., Mohan, S., Marsters, P., & Herndon, W. (2018). *Energy and Environmental Implications of a Carbon Tax in the United States* (Tech. Rep.). New York, NY: Columbia University. Retrieved from https://energypolicy.columbia.edu/sites/default/files/pictures/CGEP_Energy_Environmental_Impacts_CarbonTax_FINAL.pdf (An independent report prepared by Rhodium Group for Columbia SIPA Center on Global Energy Policy)
- Lazard. (2020, October). Lazard’s Levelized Cost of Energy Analysis Version 14.0. *Lazard Freres & Co.*. Retrieved from <https://www.lazard.com/perspective/levelized-cost-of-energy-and-levelized-cost-of-storage-2020/>
- Leighty, R. S. (1997). *Neogene tectonism and magmatism across the Basin and Range-Colorado Plateau boundary, central Arizona*. Arizona State University.
- Lerch, D. W., Klemperer, S. L., Glen, J. M. G., Ponce, D. A., Miller, E. L., & Colgan, J. P. (2007). Crustal structure of the northwestern Basin and Range Province and its transition to unextended volcanic plateaus. *Geochemistry, Geophysics, Geosystems*, 8(2). Retrieved 2021-06-11, from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GC001429> (_eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006GC001429>) doi: 10.1029/2006GC001429
- Levander, A., Schmandt, B., Miller, M., Liu, K., Karlstrom, K., Crow, R., ... Humphreys, E. (2011). Continuing Colorado plateau uplift by delamination-style convective lithospheric downwelling. *Nature*, 472(7344), 461–465. (Publisher: Nature Publishing Group)
- Lichoro, C. M., Arnason, K., & Cumming, W. (2019). Joint interpretation of gravity and resistivity data from the Northern Kenya volcanic rift zone: Structural and geothermal significance. *Geothermics*, 77, 139–150. (Publisher: Elsevier)

- Lillian, B. (2019, March). *New Mexico Enacts Bill For 100% Carbon-Free Power*. Retrieved 2021-05-02, from <https://nawindpower.com/new-mexico-enacts-bill-100-carbon-free-power>
- Lioudis, N. (2021, January). *What Causes Oil Prices to Fluctuate?* Retrieved 2021-05-26, from <https://www.investopedia.com/ask/answers/012715/what-causes-oil-prices-fluctuate.asp>
- Liu, X., Gluesenkamp, K., & Momen, A. M. (2015, May). *Overview of Available Low-Temperature/ Coproduced Geothermal Resources in the United States and the State of the Art in Utilizing Geothermal Resources for Space Conditioning in Commercial Buildings* (Tech. Rep. No. ORNL/TM-2015/131). Oak Ridge, Tennessee: Oak Ridge National Laboratory.
- Lottaroli, F., Craig, J., & Cozzi, A. (2018, May). Evaluating a vintage play fairway exercise using subsequent exploration results: did it work? *Petroleum Geoscience*, 24(2), 159–171. Retrieved 2021-06-09, from <https://pubs.geoscienceworld.org/pg/article-abstract/24/2/159/519277/Evaluating-a-vintage-play-fairway-exercise-using> (Publisher: GeoScienceWorld) doi: 10.1144/petgeo2016-150
- Lowrie, W. (2007). *Fundamentals of Geophysics*. Cambridge University Press.
- Lowry, T. S., Finger, J. T., Carrigan, C. R., Foris, A., Kennedy, M. B., Corbet, T. F., ... Sonnenthal, E. L. (2017). *GeoVision Analysis Supporting Task Force Report: Reservoir Maintenance and Development* (SAND2017-9977). Albuquerque, NM: Sandia National Lab.
- Lowry, T. S., Foris, A., Finger, J. T., Pye, S., & Blankenship, D. A. (2017). Implications of drilling technology improvements on the availability of exploitable EGS resources..
- Lucazeau, F. (2019). Analysis and Mapping of an Updated Terrestrial Heat Flow Data Set. *Geochemistry, Geophysics, Geosystems*, 20(8), 4001–4024. Retrieved 2021-02-15, from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GC008389> (_eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019GC008389>) doi: <https://doi.org/10.1029/2019GC008389>
- Lukawski, M. Z., Anderson, B. J., Augustine, C., Capuano, L. E., Beckers, K., Livesay, B., & Tester, J. W. (2014, June). Cost analysis of oil, gas, and geothermal well drilling. *Journal of Petroleum Science and Engineering*, 118, 1–14. Retrieved 2021-04-07, from <https://www.sciencedirect.com/science/article/pii/S0920410514000813> doi: 10.1016/j.petrol.2014.03.012
- Lukawski, M. Z., Silverman, R. L., & Tester, J. W. (2016, November). Uncertainty analysis of geothermal well drilling and completion costs. *Geothermics*, 64,

- 382–391. Retrieved 2021-03-20, from <https://www.sciencedirect.com/science/article/pii/S0375650516300736> doi: 10.1016/j.geothermics.2016.06.017
- Lund, J. W. (2007). Characteristics, development and utilization of geothermal resources. *GHC Bulletin, June*. (Publisher: Citeseer)
- Lundberg, S. M. (2020, October). *Interpretable Machine Learning with XGBoost*. Retrieved 2021-06-28, from <https://towardsdatascience.com/interpretable-machine-learning-with-xgboost-9ec80d148d27>
- Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2019, March). Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv:1802.03888 [cs, stat]*. Retrieved 2021-06-28, from <http://arxiv.org/abs/1802.03888> (arXiv: 1802.03888)
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30. Retrieved 2021-06-28, from <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- Ma, Z., & Redmond, R. L. (1995). Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, 61(4), 435–439. (Publisher: [Falls Church, Va.] American Society of Photogrammetry.)
- Mack, G., Witcher, J., & Lueth, V. W. (2008). Geology of the Gila Wilderness-Silver City area..
- Malone, R. W., & Moses, K. (2004, January). *Development of Risk Assessment Matrix for NASA Engineering and Safety Center*. Retrieved 2021-07-07, from <https://ntrs.nasa.gov/citations/20050123548> (NTRS Author Affiliations: NASA Marshall Space Flight Center, Futron Corp. NTRS Meeting Information: Risk Analysis: The Profession and the Future; 2004-12-05 to 2004-12-08; undefined NTRS Document ID: 20050123548 NTRS Research Center: Marshall Space Flight Center (MSFC))
- Manente, G., Field, R., DiPippo, R., Tester, J. W., Paci, M., & Rossi, N. (2011). Hybrid solar-geothermal power generation to increase the energy production from a binary geothermal plant. In (Vol. 54907, pp. 109–119).
- Mariner, R. H., Brook, C., Reed, M., Bliss, J., Rapport, A., & Lieb, R. (1983). Low-temperature geothermal resources in the western United States. *Assessment of Low-Temperature Geothermal Resources of the United States-1982, US Circular, 892*, 31–50.
- McIntosh, A. (2016). The Jackknife estimation method. *arXiv preprint arXiv:1606.00497*.

- McIntosh, W. C., Chapin, C. E., Ratté, J. C., & Sutter, J. F. (1992). Time-stratigraphic framework for the Eocene-Oligocene Mogollon-Datil volcanic field, southwest New Mexico. *Geological Society of America Bulletin*, 104(7), 851–871. (Publisher: Geological Society of America)
- McWilliams, G. (2021, May). Investors, court deliver 'stark warning for Big Oil' on climate. *Reuters*. Retrieved 2021-05-27, from <https://www.reuters.com/article/us-global-oil-exxon-shell-wrap-idUSKCN2D72V2>
- Melosh, G. (2017). Geothermal probability of success and conceptual well design. In (pp. 13–15).
- Microsoft. (2021). *FORECAST.ETS function*. Retrieved 2021-07-13, from <https://support.microsoft.com/en-us/office/forecast-ets-function-15389b8b-677e-4fbd-bd95-21d464333f41>
- Millot, R., & Négrel, P. (2007, October). Multi-isotopic tracing (d7Li, d11B, 87Sr/86Sr) and chemical geothermometry: evidence from hydro-geothermal systems in France. *Chemical Geology*, 244(3), 664–678. Retrieved 2021-06-07, from <https://www.sciencedirect.com/science/article/pii/S0009254107003427> doi: 10.1016/j.chemgeo.2007.07.015
- Mines, G. (2008, January). Geothermal electricity technologies evaluation model DOE tool for assessing impact of research on cost of power. In *PROCEEDINGS, Thirty-Third Workshop on Geothermal Reservoir Engineering* (Vol. SGP-TR-185, p. 6). Stanford, California. Retrieved from <https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2008/mines.pdf>
- Mines, G. L. (2016). GETEM User Manual. *Idaho National Laboratory, Idaho Falls, ID*.
- Montgomery, D. C. (2012). *Statistical quality control*. Wiley Global Education.
- Moore, J., McLennan, J., Allis, R., Pankow, K., Simmons, S., Podgorney, R., ... Rickard, W. (2019, February). The Utah Frontier Observatory for Research in Geothermal Energy (FORGE): An International Laboratory for Enhanced Geothermal System Technology Development. In *PROCEEDINGS, 44th Workshop on Geothermal Reservoir Engineering*. Stanford, California: Stanford University. Retrieved from <https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2019/Moore.pdf>
- Moore, J. N., & Simmons, S. F. (2013). More power from below. *Science*, 340(6135), 933–934. (Publisher: American Association for the Advancement of Science)
- Moorkamp, M., Heincke, B., Jegen, M., Roberts, A. W., & Hobbs, R. W. (2011, January). A framework for 3-D joint inversion of MT, gravity and seismic refraction data. *Geophysical Journal International*, 184(1), 477–493. Retrieved 2020-12-24, from <https://doi.org/10.1111/j.1365-246X.2010.04856.x> doi: 10.1111/j.1365-246X.2010.04856.x

- Morgan, M. G. (2009). *Best Practice Approaches for Characterizing, Communicating and Incorporating Scientific Uncertainty in Climate Decision Making: Synthesis and Assessment Product 5.2 Report* (Vol. 5). US Climate Change Science Program.
- Moss, S. (2021, May). *Google and Fervo to develop 5MW geothermal project to power Nevada data centers*. Retrieved 2021-06-03, from <https://www.datacenterdynamics.com/en/news/google-and-fervo-to-develop-5mw-geothermal-project-to-power-nevada-data-centers/>
- Moucha, R., Forte, A. M., Rowley, D. B., Mitrovica, J. X., Simmons, N. A., & Grand, S. P. (2009). Deep mantle forces and the uplift of the Colorado Plateau. *Geophysical Research Letters*, 36(19). (Publisher: Wiley Online Library)
- Muffler, L. J. P. (1979, January). *Assessment of geothermal resources of the United States, 1978* (Tech. Rep. No. USGS-CIRC-790). Geological Survey, Reston, VA (USA). Geologic Div. Retrieved 2021-06-04, from <https://www.osti.gov/biblio/6870401> doi: 10.2172/6870401
- Murphy, C., Mai, T., Sun, Y., Jadun, P., Muratori, M., Nelson, B., & Jones, R. (2021). *Electrification futures study: Scenarios of power system evolution and infrastructure development for the united states* (Tech. Rep.). Golden, CO: National Renewable Energy Lab.
- Nair, V., & Hinton, G. E. (2010, June). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning* (pp. 807–814). Madison, WI, USA: Omnipress.
- NASA. (2017, October). *S3001: Guidelines for Risk Management Version: G*. Retrieved 2021-07-24, from https://www.nasa.gov/sites/default/files/atoms/files/s3001_guidelines_for_risk_management_-_ver_g_-_10-25-2017.pdf
- Nash, G. D., & Bennett, C. R. (2015). Adaptation of a Petroleum Exploration Tool to Geothermal Exploration: Preliminary Play Fairway Model of Tularosa Basin, New Mexico, and Texas. In *GRC Transactions* (Vol. 39, pp. 743–750). Retrieved from <https://publications.mygeoenergynow.org/grc/1032215.pdf>
- Nash, G. D., Brandt, A., Pfaff, B., Hardwick, C., Gwynn, M., Blake, K., . . . Bennett, C. R. (2017). Phase 2: Updated Geothermal Play Fairway Analysis of the Tularosa Basin, New Mexico. *Transactions-Geothermal Resources Council*, 41. (Publisher: Ruby Mountain Inc., Salt Lake City, UT (United States))
- Ng, A. (2011a). *Logistic Regression / Simplified Cost Function And Gradient Descent*. Retrieved from <https://www.coursera.org/learn/machine-learning>
- Ng, A. (2011b). *Neural Networks Learning / Backpropagation Algorithm*. Retrieved from <https://www.coursera.org/learn/machine-learning>

- Ng, A. (2011c). *Regularization / Regularized Logistic Regression*. Retrieved from <https://www.coursera.org/learn/machine-learning>
- NIST. (2021). *The NIST Reference on Constants, Units, and Uncertainty*. Retrieved 2021-06-10, from <https://physics.nist.gov/cuu/Uncertainty/coverage.html>
- NMBGMR. (2021). *NMBGMR Interactive Map*. Retrieved 2021-06-16, from <https://maps.nmt.edu/>
- NREL. (2020). *2020 Annual Technology Baseline* (Tech. Rep.). National Renewable Energy Laboratory. Retrieved from <https://atb.nrel.gov/electricity/2021/geothermal> (ISSN: 20958099 Publication Title: Annual Technology Baseline)
- NREL. (2021a). *Geothermal Prospector*. Retrieved 2021-07-22, from <https://maps.nrel.gov/geothermal-prospector/?aL=tJ4wAm%255Bv%255D%3Dt&bL=clight&cE=0&lR=0&mC=37.03763967977139%2C-102.9638671875&zL=5>
- NREL. (2021b). *System Advisor Model (SAM)*. Retrieved 2021-07-22, from <https://sam.nrel.gov/>
- O’Connell. (2018, June). *In the Matter of Public Service Company of New Mexico’s Application for Approval of its Renewable Energy Act Plan for 2019 and Proposed 2019 Rider Rate Under Rate Rider No. 36*. Santa Fe, NM, USA. Retrieved from <https://www.pnm.com/documents/396023/8638670/2019+Renewable+Plan--+Direct+Testimonies.pdf/e15005a0-809c-16f4-f722-ee9c0214d938>
- Olsen, K. H., Scott Baldrige, W., & Callender, J. F. (1987, November). Rio Grande rift: An overview. *Tectonophysics*, 143(1), 119–139. Retrieved 2021-06-11, from <https://www.sciencedirect.com/science/article/pii/0040195187900837> doi: 10.1016/0040-1951(87)90083-7
- Palma, R. S. G. (2014). *Dynamic capabilities in related diversification: the case of geothermal technology development by oil companies* (Unpublished master’s thesis). Massachusetts Institute of Technology, Cambridge, Mass.. (Publisher: Massachusetts Institute of Technology)
- Pan, S.-Y., Gao, M., Shah, K. J., Zheng, J., Pei, S.-L., & Chiang, P.-C. (2019, March). Establishment of enhanced geothermal energy utilization plans: Barriers and strategies. *Renewable Energy*, 132, 19–32. Retrieved 2021-03-20, from <https://www.sciencedirect.com/science/article/pii/S0960148118309248> doi: 10.1016/j.renene.2018.07.126
- Patterson, C. (1956, October). Age of meteorites and the earth. *Geochimica et Cosmochimica Acta*, 10(4), 230–237. Retrieved 2021-03-17, from <https://www.sciencedirect.com/science/article/pii/0016703756900369> doi: 10.1016/0016-7037(56)90036-9

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825–2830. (Publisher: JMLR. org)
- Pepin, J. (2019). *New Approaches to Geothermal Resource Exploration and Characterization* (Unpublished doctoral dissertation). New Mexico Institute of Mining and Technology. (ISBN: 0438764366)
- Pepin, J., Person, M., Phillips, F., Kelley, S., Timmons, S., Owens, L., ... Gable, C. (2015). Deep fluid circulation within crystalline basement rocks and the role of hydrologic windows in the formation of the Truth or Consequences, New Mexico low-temperature geothermal system. *Geofluids*, 15(1-2), 139–160. (Publisher: Wiley Online Library)
- Petty, S., Bour, D. L., Livesay, B. J., Baria, R., & Adair, R. (2009, March). Synergies and Opportunities Between EGS Development and Oilfield Drilling Operations and Producers. OnePetro. Retrieved 2021-04-18, from <https://onepetro.org/SPEWRM/proceedings/09WRM/All-09WRM/SPE-121165-MS/146156> doi: 10.2118/121165-MS
- Pirog, R. L. (2007). The role of national oil companies in the international oil market. Congressional Research Service, Library of Congress.
- PNM. (2014, June). *Public Service Company of New Mexico's Renewable Energy Portfolio Procurement Plan For 2015*. Retrieved 2021-05-02, from https://www.pnm.com/documents/396023/428013/Renewable_Plan_2015/8e4a68b7-5177-4e38-99c2-d3edf7f598c6
- Prensky, S. (1992). Temperature measurements in boreholes-An overview of engineering and scientific applications. *The Log Analyst*, 33(03), 313–333. (Publisher: OnePetro)
- Press, F., Siever, R., Grotzinger, J., & Jordan, T. H. (2004). *Understanding Earth*. Macmillan. (Google-Books-ID: P8iEVK1yGKwC)
- Prestidge, K. (2021). *personal communication*. (oilfield employee)
- PRISM. (2021). *PRISM Climate Group, Oregon State U*. Retrieved 2021-06-13, from <https://prism.oregonstate.edu/normals/>
- Pursley, J. (2013). Earthquake catalogs for New Mexico and bordering areas: 2005–2009. *New Mexico Geology*, 35(1), 3–12.
- Rosenkjaer, G. K., Gasperikova, E., Newman, G. A., Arnason, K., & Lindsey, N. J. (2015). Comparison of 3D MT inversions for geothermal exploration: Case studies for Krafla and Hengill geothermal systems in Iceland. *Geothermics*, 57, 258–274. (ISBN: 0375-6505 Publisher: Elsevier)

- Ross, M. (2020, October). *The Energy of the Future is Right Beneath Your Feet*. Retrieved 2021-06-03, from <https://www.rollingstone.com/culture/culture-news/eavor-energy-company-review-1068287/>
- Sanford, A. (2006). Earthquake catalogs for New Mexico and bordering areas II: 1999–2004. *New Mexico Geology*, 28(4), 99–109.
- Sanford, A., Lin, K.-w., Tsai, I.-c., & Jaksha, L. H. (2002). Earthquake catalogs for New Mexico and bordering areas: 1869–1998. *New Mexico Bureau of Geology and Mineral Resources Circular*, 210, 1–9.
- Sanger, D. E., & Perlroth, N. (2021, May). Pipeline Attack Yields Urgent Lessons About U.S. Cybersecurity. *The New York Times*. Retrieved 2021-05-26, from <https://advance-lexis-com.libproxy.mit.edu/api/document?collection=news&id=urn:contentItem:62NS-5DW1-DXY4-X201-00000-00&context=1516831>.
- Sani, H. M., Lei, C., & Neagu, D. (2018). Computational complexity analysis of decision tree algorithms. In (pp. 191–197). Springer.
- Schneider, R. V., Keller, G. R., & Cather, S. (1994). Crustal structure of the western margin of the Rio Grande rift and Mogollon-Datil volcanic field, southwestern New Mexico and southeastern Arizona. *SPECIAL PAPERS-GEOLOGICAL SOCIETY OF AMERICA*, 207–207. (Publisher: GEOLOGICAL SOCIETY OF AMERICA, INC)
- Schochet, D. N., & Cunniff, R. A. (2001, February). *Development of a Plan to Implement Enhanced Geothermal Systems (EGS) in the Animas Valley, New Mexico - Final Report - 07/26/2000 - 02/01/2001* (Tech. Rep. No. DOE/ID/13989). ORMAT International, Inc., Sparks, NV (US); Lightning Dock Geothermal, Inc., Las Cruces, NM (US). Retrieved 2021-04-13, from <https://www.osti.gov/biblio/790039> doi: 10.2172/790039
- scikit-learn. (2021a). *Logistic regression*. Retrieved 2021-06-23, from https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
- scikit-learn. (2021b). *Multiclass and multioutput algorithms*. Retrieved 2021-06-23, from <https://scikit-learn.org/stable/modules/multiclass.html#multiclass-classification>
- scikit-learn. (2021c). *Pipelines and composite estimators*. Retrieved 2021-08-05, from <https://scikit-learn.org/stable/modules/compose.html#pipeline>
- scikit-learn. (2021d). *sklearn.metrics.roc_auc_score*. Retrieved 2021-06-22, from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html#sklearn.metrics.roc_auc_score

- scikit-learn. (2021e). *sklearn.model_selection.StratifiedShuffleSplit*. Retrieved 2021-06-20, from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedShuffleSplit.html
- scikit-learn. (2021f). *sklearn.preprocessing.PowerTransformer*. Retrieved 2021-06-18, from <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PowerTransformer.html#rf3e1504535de-1>
- scikit learn. (2021). *sklearn.preprocessing.StandardScaler*. Retrieved 2021-06-18, from <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>
- Seager, W. R., Shafiqullah, M., Hawley, J. W., & Marvin, R. F. (1984, January). New K-Ar dates from basalts and the evolution of the southern Rio Grande rift. *GSA Bulletin*, 95(1), 87–99. Retrieved 2021-06-11, from <https://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/95/1/87/202952/New-K-Ar-dates-from-basalts-and-the-evolution-of> (Publisher: GeoScienceWorld) doi: 10.1130/0016-7606(1984)95<87:NKDFBA>2.0.CO;2
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379–423. (Publisher: Nokia Bell Labs)
- Shapley, L. S. (1997). A value for n-person games. *Classics in game theory*, 69. (Publisher: Princeton Univ Pr)
- Shell. (2020, April). *Responsible Investment Annual Briefing updates* [Press Release]. Retrieved 2021-03-06, from <https://www.shell.com/media/news-and-media-releases/2020/responsible-investment-annual-briefing-updates.html>
- Shell. (2021, February). *Shell accelerates drive for net-zero emissions with customer-first strategy* [Press Release]. Retrieved 2021-03-07, from <https://www.shell.com/media/news-and-media-releases/2021/shell-accelerates-drive-for-net-zero-emissions-with-customer-first-strategy.html>
- Shieber, J. (2019, March). *Bill Gates and Jeff Bezos-backed fund invests in a global geothermal energy project developer* | *TechCrunch*. Retrieved 2021-05-01, from <https://techcrunch.com/2019/03/03/bill-gates-and-jeff-bezos-backed-fund-invests-in-a-global-geothermal-energy-project-developer/?guccounter=1>
- Shieber, J. (2021, April). *Geothermal technology has enormous potential to power the planet and Fervo wants to tap it*. Retrieved 2021-06-03, from <https://social.techcrunch.com/2021/04/30/geothermal-technology-has-enormous-potential-to-power-the-planet-and-fervo-wants-to-tap-it/>
- Shuen, A., Feiler, P. F., & Teece, D. J. (2014, September). Dynamic capabilities in the upstream oil and gas sector: Managing next generation competition. *Energy Strategy Reviews*, 3, 5–13. Retrieved 2021-03-08, from <https://www>

- [.sciencedirect.com/science/article/pii/S2211467X14000194](https://www.sciencedirect.com/science/article/pii/S2211467X14000194) doi: 10.1016/j.esr.2014.05.002
- Silverman, B. (1986). *Density estimation for statistics and data analysis*. (Publisher: Chapman & Hall)
- Singh, M. (2021, January). *The 2007-2008 Financial Crisis in Review*. Retrieved 2021-05-26, from <https://www.investopedia.com/articles/economics/09/financial-crisis-review.asp>
- Small, S. (2019, April). *Theory meets application: Machine learning techniques for geothermal exploration / Penn State University*. Retrieved 2021-06-09, from <https://news.psu.edu/story/570047/2019/04/18/research/theory-meets-application-machine-learning-techniques-geothermal>
- Smith, C. M., Faulds, J. E., Brown, S., Coolbaugh, M., Lindsey, C. R., Treitel, S., ... Mlawsky, E. (2021, February). Characterizing Signatures of Geothermal Exploration Data with Machine Learning Techniques: An Application to the Nevada Play Fairway Analysis. In *PROCEEDINGS, 46th Workshop on Geothermal Reservoir Engineering*. Stanford, California.
- SMU. (2021). *Geothermal Data Aggregation*. Retrieved 2021-06-16, from <http://geothermal.smu.edu/gtda/>
- Sorey, M. L., Reed, M. J., Foley, D., & Renner, J. (1983). Low-temperature geothermal resources in the central and eastern United States. *Assessment of low-temperature geothermal resources of the United States-1982: US Geological Survey Circular*, 892, 51–65.
- Stanfield, J. (2017, November). *New Mexico approves solar project, geothermal contract, wind upgrade for PNM*. Retrieved 2021-05-02, from <https://www.spglobal.com/marketintelligence/en/news-insights/trending/ZQmWEcw4eDx8l4y8P5we5g2>
- Stein, C. A. (1995). Heat flow of the Earth. *Global earth physics: a handbook of physical constants*, 1, 144–158. (Publisher: Wiley Online Library)
- Stephens, J. C., & Jiusto, S. (2010, April). Assessing innovation in emerging energy technologies: Socio-technical dynamics of carbon capture and storage (CCS) and enhanced geothermal systems (EGS) in the USA. *Energy Policy*, 38(4), 2020–2031. Retrieved 2021-06-02, from <https://www.sciencedirect.com/science/article/pii/S0301421509009392> doi: 10.1016/j.enpol.2009.12.003
- Stevenson, D. J., & Halliday, A. N. (2014, September). The origin of the Moon. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 372(2024). Retrieved 2021-03-17, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4128275/> doi: 10.1098/rsta.2014.0289

- Stoesser, D., Green, G., Morath, L., Heran, W., Wilson, A., Moore, D., & Van Gosen, B. (2005). *USGS Geologic Map Database, New Mexico* (Open-File Report No. 2005-1351). Reston, VA: USGS. Retrieved 2021-06-16, from <http://pubs.usgs.gov/of/2005/1351/>
- Tester, J. W., Anderson, B. J., Batchelor, A., Blackwell, D., DiPippo, R., Drake, E., ... Nichols, K. (2006). The future of geothermal energy. *Massachusetts Institute of Technology*, 358.
- Tester, J. W., & Herzog, H. J. (1990). *Economic predictions for heat mining : a review and analysis of hot dry rock (HDR) geothermal energy technology* (Tech. Rep. No. MIT-EL 90-001). Cambridge, Mass.: Massachusetts Institute of Technology. Energy Laboratory. Retrieved 2021-02-26, from <https://dspace.mit.edu/handle/1721.1/60650> (Accepted: 2011-01-14T23:32:31Z Publisher: [Cambridge, Mass.] : Energy Laboratory, Massachusetts Institute of Technology, 1990)
- Tester, J. W., Reber, T., Beckers, K., Lukawski, M., Camp, E., Aguirre, G. A., ... Horowitz, F. (2015). Integrating geothermal energy use into re-building American infrastructure..
- Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*. (Publisher: Emerald Publishing Limited)
- Think GeoEnergy. (2020, January). *Turboden on its repowering of the Lightning Dock geothermal plant in New Mexico | ThinkGeoEnergy - Geothermal Energy News*. Retrieved 2021-05-02, from <https://www.thinkgeoenergy.com/turboden-on-its-repowering-of-the-lightning-dock-geothermal-plant-in-new-mexico/>
- Thompson, G. A., & Zoback, M. L. (1979). Regional geophysics of the Colorado Plateau. *Tectonophysics*, 61(1-3), 149–181. (Publisher: Elsevier)
- Total. (2020, May). *Total adopts a new Climate Ambition to Get to Net Zero by 2050* [Press Release]. Retrieved 2021-03-07, from <https://www.total.com/media/news/total-adopts-new-climate-ambition-get-net-zero-2050>
- Total. (2021, February). *2020 Results & Outlook* [Press Release]. Retrieved 2021-03-06, from <https://www.total.com/investors/results-investor-presentations/results>
- Turcotte, D. L., & Schubert, G. (2002). *Geodynamics* (2nd edition ed.). Cambridge university press.
- UNFCCC. (2015). *PARIS AGREEMENT - Conference of the Parties COP 21* (Vol. 21932) (No. December). (ISSN: 1098-6596 Publication Title: Adoption of the Paris Agreement. Proposal by the President.)
- U.S. BLS. (2021a). *ECI total compensation for private industry workers in utilities*. Retrieved 2021-07-10, from <https://data.bls.gov/timeseries/>

- CIS2014400000000I (total compensation United States Bureau of Labor Statistics Data)
- U.S. BLS. (2021b). *PPI industry data for electric power generation-utilities*. Retrieved 2021-07-09, from <https://data.bls.gov/timeseries/PCU2211102211104> (not seasonally adjusted, Bureau of Labor Statistics Data)
- USGS. (2021a). *Earthquake Hazards Program*. Retrieved 2021-06-14, from <https://earthquake.usgs.gov/earthquakes/search/>
- USGS. (2021b). *EERMA*. Retrieved 2021-06-16, from <https://my.usgs.gov/eerma/>
- USGS. (2021c). *National Water Information System*. Retrieved 2021-06-15, from <https://waterdata.usgs.gov/nwis>
- USGS. (2021d). *TNM Download v2*. Retrieved 2021-06-16, from <https://apps.nationalmap.gov/downloader/#/>
- UTEP. (2011). *Gravity Dataset for the United States Lower 48 states*. Retrieved 2021-06-14, from <http://cybershare.utep.edu/dataset/gravity-dataset-united-states-lower-48-states>
- Vesselinov, V. V., Mudunuru, M. K., Ahmmed, B., Karra, S., & Middleton, R. S. (2020). *Discovering signatures of hidden geothermal resources based on unsupervised learning* (Tech. Rep.). Los Alamos, NM: Los Alamos National Lab.
- Vose, R., Easterling, D., Kunkel, K., LeGrande, A., & Wehner, M. (2017). *Temperature changes in the United States* (Tech. Rep.). Washington, D.C: U.S. Global Change Research Program.
- Vozoff, K., & Jupp, D. (1975). Joint inversion of geophysical data. *Geophysical Journal International*, 42(3), 977–991. (Publisher: Blackwell Publishing Ltd Oxford, UK)
- Wannamaker, P. E. (2016). *Structurally controlled geothermal systems in the central Cascades arc-backarc regime, Oregon* (Tech. Rep.). Salt Lake City, UT: University of Utah.
- Webster, K. (2021). *Probabilistic layers and Bayesian neural networks* [e-learning]. Retrieved from <https://www.coursera.org/learn/probabilistic-deep-learning-with-tensorflow2/>
- Williams, C. F., Reed, M. J., Mariner, R. H., DeAngelo, J., & Galanis, S. P. (2008). *Assessment of moderate-and high-temperature geothermal resources of the United States* (Tech. Rep. No. 2327-6932). Menlo Park, California: U.S. Geological Survey.

- Wilson, D., Aster, R., Ni, J., Grand, S., West, M., Gao, W., ... Semken, S. (2005). Imaging the seismic structure of the crust and upper mantle beneath the Great Plains, Rio Grande Rift, and Colorado Plateau using receiver functions. *Journal of Geophysical Research: Solid Earth*, 110(B5). Retrieved 2021-06-12, from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JB003492> (__eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004JB003492>) doi: 10.1029/2004JB003492
- Wilson, O. (2020). *Machine Learning for Well Rate Estimation: Integrated Imputation and Stacked Ensemble Modeling* (Unpublished master's thesis). Massachusetts Institute of Technology, Cambridge, Mass..
- Winther, J. (2018). *Power production from excess heat in a district heating system* (Unpublished master's thesis). Chalmers University of Technology, Gothenburg, Sweden.
- Wisian, K., Blackwell, D., & Richards, M. (1999). Heat flow in the western United States and extensional geothermal systems. In (Vol. 219). Citeseer.
- Witcher, J. C. (2002, January). *Deep Production Well for Geothermal Direct-Use Heating of A Large Commercial Greenhouse, Radium Springs, Rio Grande Rift, New Mexico* (Tech. Rep. No. DOE/ID13747). Alex R. Masson, Inc. (US). Retrieved 2021-08-02, from <https://www.osti.gov/biblio/791568> doi: 10.2172/791568
- Wood Mackenzie. (2019, February). *Energy Research & Consultancy*. Retrieved 2021-08-17, from <https://www.woodmac.com/>
- Woskov, P. (2017). Millimeter-Wave Directed Energy Deep Boreholes Melting and Vaporizing Rock Thermodynamics of Rock Replacing Drilling Mud..
- xgboost developers. (2020). *XGBoost Parameters — xgboost 1.5.0-dev documentation*. Retrieved 2021-06-26, from <https://xgboost.readthedocs.io/en/latest/parameter.html>
- Yeo, I., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954–959. (Publisher: Oxford University Press)