

EXPLORATORY SPATIAL DATA ANALYSIS AND UNCERTAINTY PROPAGATION  
FOR GEOTHERMAL RESOURCE ASSESSMENT AND RESERVOIR MODELS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jared David Smith

May 2019

© 2019 Jared David Smith

# EXPLORATORY SPATIAL DATA ANALYSIS AND UNCERTAINTY PROPAGATION FOR GEOTHERMAL RESOURCE ASSESSMENT AND RESERVOIR MODELS

Jared David Smith, Ph. D.

Cornell University 2019

In the exploration and planning phases of geothermal energy projects, data are often limited and imprecise. As a result, the uncertainty in derived geothermal variables can be large. To address those problems, this thesis develops exploratory spatial data analysis procedures to identify discordant observations in datasets of derived geothermal variables, and a stochastic framework to propagate uncertainty through geothermal resource assessment and geothermal reservoir models. The resulting uncertainty distributions should inform project decisions and guide prioritization of additional data collection and modeling. This research is applied to Appalachian Basin geothermal resource characterization, and to assess the feasibility of using geothermal reservoirs to meet direct-use heating objectives for the Cornell University campus in Ithaca, NY.

Exploratory spatial data analysis (ESDA) can identify discordant observations in datasets derived from low-quality bottom-hole temperature (BHT) measurements taken during well drilling. Spatial regressions that characterize geothermal resources commonly employ such datasets. A local median deviation procedure informs a minimum depth for BHT measurements to represent conduction-dominated heat transfer for the Appalachian Basin. A subsequent local spatial outlier identification procedure detects discordant observations using asymmetric boxplots. Using ESDA procedures and removing discordant observations significantly reduced

and stabilized semi-variance estimates of spatial autocorrelations. Striking differences are found in spatial autocorrelation functions employed for regressions in separate geologic regions.

A stochastic framework supports uncertainty and sensitivity analysis of geothermal resource assessment models. Assessments require transforming BHT measurements into heat flows at well sites, and employing spatial regression to estimate heat flow across the region. Uncertainties in geologic properties and predicted surface heat flows are propagated through models to obtain uncertainty distributions of temperature-depth profiles across the Appalachian Basin, and distributions of thermal energy within potential geothermal reservoirs. Sensitivity analyses reveal variables whose uncertainties contribute most to thermal energy uncertainty; the results are a function of a reservoir's spatial location and the depth of available temperature data. Many geologic variables may be assigned regional values with large uncertainties with little impact on the thermal energy uncertainty.

Stochastic evaluation of geothermal reservoir production models transforms geothermal resource assessment uncertainties into decision-relevant information for project planning. Several reservoir flow geometries and production scenarios are compared using analytical and numerical reservoir models. The analysis generates the joint probability over time that a production scenario achieves specified heating objectives. Uncertainties in reservoir flow geometry provided a wide range of possible useful lifetimes for target geothermal reservoirs.

## BIOGRAPHICAL SKETCH

Jared David Smith is from Bay Shore, New York. Growing up on Long Island, Jared had varied interests in environmental sciences and conservation, which matured through a journey to the rank of Eagle Scout in the Boy Scouts of America. Jared attended Clarkson University for undergraduate studies and received a B.S. in Environmental Engineering with great distinction in May of 2013. Jared's graduation awards included the Phalanx Commendable Leadership Award and the Levinus Clarkson Award. Jared was recognized nationally for his undergraduate scholastic achievements with fellowships for graduate study from the Tau Beta Pi Engineering Honor Society and The Honor Society of Phi Kappa Phi.

Jared began graduate school at Cornell University in August of 2013 as part of Jery Stedinger's research group in Environmental and Water Resources Systems Engineering, and the interdisciplinary National Science Foundation-funded Earth-Energy Systems Integrated Graduate Education and Research Traineeship (IGERT) research group. Numerous IGERT-related field trips, conferences, and workshops broadened Jared's research interests to include geological sciences in addition to water and energy systems science and engineering.

Over 5.5 years of graduate study, Jared contributed to several geothermal energy research projects funded by the United States Department of Energy, including a Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin, and two Deep Direct Use geothermal feasibility studies for the campuses of Cornell University in Ithaca, NY, and West Virginia University in Morgantown, WV. In May of 2016, Jared conducted research at Reykjavik University's Iceland School of Energy and the Iceland Geosurvey, ÍSOR. In August of 2016, Jared completed an M.S. that developed analytical and geostatistical methods for geothermal resource reconnaissance in the Appalachian Basin. In recognition of Jared's contributions to the geothermal community, he received a Geothermal Resources Council Graduate Scholarship for the 2018-2019 academic year.

While at Cornell, Jared had the pleasure of being a teaching assistant for three courses, and in the Fall semester of 2017 was the instructor of Uncertainty Analysis in Engineering offered by the School of Civil and Environmental Engineering. Jared was an active graduate member of the New York Delta chapter of Tau Beta Pi, for which he served as an advisor for four years. The chapter developed into one of the more active chapters in the honor society over that time, and received a Chapter Excellence Award for the 2017-2018 academic year.

## ACKNOWLEDGEMENTS

I would like to thank my thesis committee, Jery Stedinger, Teresa Jordan, and David Rossiter, for their support and guidance throughout my graduate education. Each of you have challenged me to deepen my knowledge of your respective fields of study, and have enabled me to grow as a researcher.

Several chapters in this thesis contain sentences with “we” or “our” in reference to myself and coauthors of planned journal manuscripts. Coauthors are acknowledged within relevant chapters of this thesis. I appreciate all of the feedback that they provided to me, which improved my work.

I am grateful to have completed research with several project leaders who encouraged field trips and experiential learning to be critical components of my development as an earth-energy systems researcher. The organizations that supported these trips and those who arranged these trips included:

- The National Science Foundation-funded Earth-Energy Systems Integrated Graduate Education and Research Traineeship (IGERT). Cornell University trip leaders: Jefferson Tester, Teresa Jordan, and John Thompson.
- Iceland School of Energy at Reykjavík University, and the Iceland Geosurvey, ÍSOR. Trip leaders and organizers: Jefferson Tester, Sverrir Thorhallsson, R. Morgan Greene, and Guðni Axelsson. Jeff: Thank you for giving me the opportunity to conduct research abroad and to learn about the geothermal energy culture of Iceland.
- The National Science Foundation-funded SEDHeat program. Trip and workshop leaders: Jeff Bielicki, The Ohio State University; Stewart Simmons, University of Utah; John Holbrook, Texas Christian University; and John Muntean, University of Nevada, Reno. Thank you for welcoming me on a research trip and a workshop that broadened my knowledge of geothermal systems and their energy utilization potential.

I enjoyed research collaborations with three geothermal project teams throughout my graduate studies. These teams and individuals whom I would like to particularly thank are:

- Cornell University Deep Direct Use Project Team: J. Olaf Gustafson, Steve Beyers, Jood Al Aswad, Tasnuva Ming Khan, Jefferson Tester, and Teresa Jordan. I feel fortunate to have contributed to a real engineering project for part of my Ph. D. research. The integration of academic research with facilities engineering is among the most valuable components of my graduate education. Jood and Ming: thank you for processing and providing data, and for discussions that were critical for my thesis work. This research was supported by the U.S. Department of Energy (Award Number: DE-EE0008103).
- West Virginia University Deep Direct Use Project Team: Nagasree Garapati, Brian Anderson, R. Scott McCleery, Ronald McDowell, Jessica Moore, Christine Doughty, and Yingqi Zhang. Thank you for sharing data and for discussions that informed my thesis work. This research was supported by the U.S. Department of Energy (Award Number: DE-EE0008105).
- Appalachian Basin Geothermal Play Fairway Analysis Team: Calvin Whealton, Erin Camp, Franklin Horowitz, Teresa Jordan, Jefferson Tester, and Jerry Stedinger of the Cornell University team; Zachary Frone, David Blackwell, and Maria Richards of the Southern Methodist University team; and Brian Anderson and Kelydra Welcker of the West Virginia University team. This research was supported by the U.S. Department of Energy (Award Number: DE-EE0006726).

In addition to the aforementioned U.S. Department of Energy awards, portions of the research presented in this thesis were completed while I was supported by the National Science Foundation IGERT program (Award Number: DGE-0966045), and a Geothermal Resources Council Graduate Scholarship.

I would like to thank several individuals who provided me access to datasets and software that were used in my thesis work:

- Jessica Moore and Ronald McDowell at the West Virginia Geologic and Economic Survey provided access to confidential data contained within their state databases.
- Kris Carter at the Pennsylvania Department of Conservation and Natural Resources provided access to the EDWIN database of well logs for Pennsylvania.
- Brian Slater at the New York State Geological Survey provided access to the ESOGIS database of well logs for New York.
- Tom Pasquini of Gulf Plains Prospecting Company, LLC provided field trip guidance and data on fractures in the Adirondack Mountains in New York.
- Richard Allmendinger at Cornell University provided software for fracture orientation and spacing analysis.

I would like to thank several faculty members at Cornell for their contributions to my graduate and professional education:

- Jerry Stedinger and Linda Nozick: Thank you for giving me the opportunity to be a teaching assistant for two semesters and the instructor for Uncertainty Analysis in Engineering in Fall, 2017. I appreciated the assistance of the School of Civil and Environmental Engineering administrative staff during this time, particularly Nadine Porter, Tania Sharpsteen, Charissa King-O'Brien, Megan Keene, and Jeanette Little.
- R. Bruce van Dover: Thank you for giving me the opportunity to be a teaching assistant for your summer course. I enjoyed exploring engineering with you and the students.
- Patrick Reed: Thank you for including me in your group's weekly meetings and discussions with visiting researchers and individuals in water systems engineering.
- D. Pete Loucks: Thank you for several dinner conversations about life in civil engineering, and for allowing me to impress students with the portion of your library that was in my office.

Finally, thanks to many friends and fellow graduate students and postdocs for academic discussions and personal support throughout my graduate studies: Lena Abu-Ali, G. Andrea Aguirre, Jood Al Aswad, Koenraad Beckers, Stephen Borregine, Erin Camp, Corey Carpenter, Esther Chiew, Casey Ching, Kwangbae Choi, Kenji Doering, Kate Donahue, David Gold, Amandeep Gupta, Rohini Gupta, Antonia Hadjimichael, Deniz Harland, Katherine Herleman, Jonathan Herman, Eric Huber, Paige Jacob, Courtney Karszes, Sarah Kayali El-Alem, David Kemmenoe, Tasnuva Ming Khan, Justyna Kosianka, Jonathan Lamontagne, Yolanda Lin, Scott Maguffin, John Mason, Joseph Mattson, Philip McGuire, Marika Nell, Arna Pálsdóttir, Dana Peterson, Julianne Quinn, Charles Rogué, Andrew Shah, Laura Sinclair, Jane Steele, Peter Storm, Sue Nee Tan, Bernardo Trindade, Kezhen Jenny Wang, Victoria Ward, Calvin Whealton, Sarah Wilkinson, Xin Candice Yu, Jazmín Zatarain-Salazar, and Michael Zumstein.

## TABLE OF CONTENTS

Biographical Sketch .....	iii
Acknowledgements .....	v
Table of Contents .....	ix
List of Figures .....	xii
List of Tables .....	xvii
CHAPTER 1 Introduction.....	1
1.0 Background and Research Motivation .....	1
1.1 Chapter Organization .....	2
1.1.1 Chapter 2: Exploratory Spatial Data Analysis for Geothermal Resource Assessments.....	4
1.1.2 Chapter 3: Stochastic Uncertainty Analysis and Sensitivity Analysis of Geothermal Resource Assessments for Direct-Use Heating .....	5
1.1.3 Chapter 4: Uncertainty Analysis of Geothermal Reservoir Simulations for Geothermal Greenfield Feasibility Assessments .....	6
1.1.4 Chapter 5: Conclusions and Future Research Opportunities .....	7
CHAPTER 2 Exploratory Spatial Data Analysis for Geothermal Resource Assessments: An Appalachian Basin Case Study .....	9
2.0 Abstract .....	9
2.1 Introduction .....	10
2.1.1 Definition and Causes of Rogue Observations in Geothermal Datasets .....	11
2.1.2 Paper Organization.....	15
2.2 Description of the Dataset .....	15
2.3 Initial Data Processing .....	18
2.3.1 Physically Implausible Values .....	18
2.3.2 Records with the Same Spatial Coordinates .....	20
2.4 Exploratory Spatial Data Analysis .....	23
2.4.1 Local Median Deviation Analysis for Selecting a Minimum Depth for BHTs.....	23
2.4.2 Spatial Outlier Detection and Analysis .....	30
2.4.3 Geostatistical Evaluation of ESDA Procedures for Geothermal Resource Assessments....	40
2.5 Discussion .....	44
2.6 Conclusions .....	48
2.7 Acknowledgements .....	49
2.8 Disclaimer .....	50
2.9 Code and Data Availability, and Software Credits .....	50
CHAPTER 3 Stochastic Uncertainty Analysis and Sensitivity Analysis of Geothermal Resource Assessments for Direct-Use Heating .....	57
3.0 Abstract .....	57
3.1 Introduction .....	58

3.1.1	Case Study Site Description.....	61
3.1.2	Organization.....	65
3.2	One-Dimensional Heat Conduction Model.....	65
3.2.1	Heat Conduction Model Variables.....	67
3.2.2	Uncertainty Analysis Methods for Temperature-Depth Profiles across the Appalachian Basin	74
3.2.3	Sensitivity Analysis Methods for the Surface Heat Flow and Temperature-Depth Profiles at Well Locations .....	75
3.3	Heat-in-Place Model .....	76
3.3.1	Heat-in-Place Model Equations .....	77
3.3.2	Heat-in-Place Model Variables .....	79
3.3.3	Uncertainty and Sensitivity Analysis Methods for the Heat-In-Place Model .....	86
3.4	Sensitivity Analysis Methods.....	87
3.4.1	Sensitivity Analysis Implementation .....	88
3.5	Results .....	90
3.5.1	Distributions of Temperatures at the Basement Depth in the Appalachian Basin .....	91
3.5.2	Sensitivity Analysis for the Surface Heat Flow and Temperatures at 1.5 km Depth at Well Locations	94
3.5.3	Distributions of Temperatures at Reservoir Depths .....	100
3.5.4	Distributions of Stored Thermal Energy for Reservoirs .....	104
3.5.5	Sensitivity Analysis of Stored Thermal Energy for Reservoirs .....	110
3.6	Discussion .....	120
3.7	Conclusions .....	123
3.8	Acknowledgements .....	125
3.9	Code and Data Availability, and Software Credits .....	125
CHAPTER 4 Uncertainty Analysis of Geothermal Reservoir Simulations for Geothermal Greenfield Feasibility Assessments: A Case Study of Direct-Use Heating for Cornell University.....		
		134
4.0	Abstract .....	134
4.1	Introduction .....	135
4.1.1	Case Study Site Description.....	137
4.1.2	Paper Organization.....	140
4.2	Geothermal Resource Below the Cornell Campus.....	140
4.3	Geologic Column Below Cornell.....	145
4.3.1	Caprocks: Lorraine / Utica Shale .....	148
4.3.2	Reservoir Target: Trenton-Black River .....	148
4.3.3	Reservoir Target: Precambrian Crystalline Basement .....	154
4.4	Geothermal Reservoir Models .....	156
4.4.1	Analytical Reservoir Models .....	157
4.4.2	Numerical Reservoir Model TOUGH2.....	161

4.4.3	Extraction Well Borehole Heat Transfer.....	164
4.4.4	Model Implementation and Uncertainty Analysis .....	165
4.5	Production Scenarios.....	167
4.5.1	Well Flow Rates and Diameters.....	167
4.5.2	Injection Temperatures .....	168
4.5.3	Well Separation Distance.....	168
4.5.4	Capacity Factor .....	169
4.6	Results .....	169
4.6.1	Trenton-Black River Analytical Model Results.....	169
4.6.2	Trenton-Black River Numerical Model .....	173
4.6.3	Basement.....	177
4.7	Discussion .....	179
4.8	Conclusions .....	183
4.9	Acknowledgements .....	184
4.10	Disclaimer .....	184
4.11	Code and Data Availability, and Software Credits .....	185
CHAPTER 5 Conclusions and Future Research Opportunities.....		192
5.0	Summary of Thesis Research Progression.....	192
5.1	Summary of Major Conclusions .....	193
5.2	Future Research Opportunities.....	196
5.2.1	Geothermal Resource Exploration Phase.....	196
5.2.2	Geothermal Project Planning Phase.....	199
Appendix A Performance Tests for the Asymmetric Boxplot Outlier Criterion Presented in Chapter 2.....		200
Appendix B Supplementary Material for Chapter 2.....		205
Appendix C Supplementary Material for Chapter 3.....		219
Appendix D Supplementary Material for Chapter 4.....		223

## LIST OF FIGURES

### Chapter 1

Figure 1.1: Graphical abstract of the research progression and methods developed in this thesis.....	4
--	---

### Chapter 2

Figure 2.1: Locations of 20,750 wells used in this study, colored by the BHT measurement depth.....	17
Figure 2.2: BHTs for records in the same spatial locations.....	21
Figure 2.3: Exploratory spatial data analysis of the surface heat flow ( $Q_s$ ).....	27
Figure 2.4: BHTs deeper than 1000 m (black) and data that were removed based on the 600 m (light colors) and 1000 m (dark colors) minimum depth thresholds. ....	29
Figure 2.5: Sensitivity analysis of the local spatial outlier algorithm for the surface heat flow within a neighborhood defined by the maximum radius (vertical axis) to achieve the number of data points required to run the outlier test (horizontal axis). ....	33
Figure 2.6: Spatial distribution of surface heat flow ( $Q_s$ ) local outliers identified using the spatial outlier detection algorithm.....	35
Figure 2.7: Histograms and maps of the depth ranks of low and high outliers. ....	37
Figure 2.8: Normal q-q plots of the surface heat flow within each of the geologic regions shown on the map. ....	39
Figure 2.9: Average semi-variance of the surface heat flow as a function of separation distance within the geologic regions provided in Figure 2.8. ....	43

### Chapter 3

Figure 3.1: Schematic of the one-dimensional heat conduction model.....	68
Figure 3.2: Schematic of the heat-in-place model used to calculate stored thermal energy for a reservoir volume consisting of a $1 \text{ km}^2$ grid cell area and thickness $Z_b - Z_t$ . ....	80
Figure 3.3: Estimated quantiles of temperatures at the basement depth for each of 138,400 $1 \text{ km}^2$ grid cells, and the depth to basement rocks in the Appalachian Basin. ....	93
Figure 3.4: Total order sensitivity effects for the surface heat flow (left panel) and the temperature at 1.5 km depth (right panel) for each of 12,155 wells (points on plots) for five individual input variables and two input variable groups. ....	96

Figure 3.5: Maps of the total order sensitivity effects for the surface heat flow at 12,155 well locations (points) for five individual input variables and two input variable groups.....	98
Figure 3.6: Maps of the total order sensitivity effects for the temperature at 1.5 km depth at 12,155 well locations (points) for five individual input variables and two input variable groups.....	99
Figure 3.7: Estimated temperatures at the center depth of each of 16,700 1 km <sup>2</sup> areas of potential geothermal reservoirs in the Appalachian Basin.....	102
Figure 3.8: Left: 5-95% quantile range of temperatures at the center depth of each of 16,700 1 km <sup>2</sup> areas of the reservoirs evaluated. Right: Geologic regions used for interpolation by stratified kriging of the surface heat flow by Smith (2016). .....	104
Figure 3.9: Energy density (stored thermal energy per unit volume), specified for each of 16,700 1 km <sup>2</sup> spatial areas times the corresponding reservoir thickness (meters) at that location.....	106
Figure 3.10: Standard deviation of the stored thermal energy density vs. the average stored thermal energy density for each of 16,700 1 km <sup>2</sup> areas of reservoirs evaluated. ....	107
Figure 3.11: Empirical cumulative distribution functions of the useful life for each 1 km <sup>2</sup> area the Corning Black River Dolomite (left, n = 5 km <sup>2</sup> ), and the Morgantown Tuscarora Sandstone (right, n = 241 km <sup>2</sup> ).....	110
Figure 3.12: Total order sensitivity of the stored thermal energy to each input variable for each of 16,700 1 km <sup>2</sup> areas of reservoirs evaluated.....	113
Figure 3.13: Maps of total order sensitivity of the stored thermal energy for each of 16,700 1 km <sup>2</sup> areas of reservoirs for each input variable (A-F), and plots of the temperature and thickness uncertainty for each 1 km <sup>2</sup> area (G, H).....	116

## Chapter 4

Figure 4.1: Predicted temperature-depth profile below Cornell University in Ithaca, NY.....	142
Figure 4.2: Left: Location of Cornell (purple diamond) and wells with digital logs that intersect the Black River formation. Right: Violin plots of effective porosity (well log neutron porosity observations corrected for gas and shale) in the Black River formation for wells on the map. ....	150
Figure 4.3: Tobit censored regression for maximum (red) and 90°-to-maximum (blue) horizontal permeability from the Whiteman #1 core (New York State Museum, 2014) of a Trenton-Black River reservoir.....	152
Figure 4.4: Conceptual diagrams for analytical reservoir models after a lengthy time has elapsed.....	158

Figure 4.5: Modeled geologic strata and simulation grid in the TOUGH2 numerical model.....	162
Figure 4.6: Top Row: Production temperature over time for the Trenton-Black River reservoir using the parallel fractures model (left) and the heat sweep model (right). The injection temperature is 30 °C and the flow rate is 50 kg/s. Each blue line provides the results of a single Monte Carlo replicate. Selected quantiles over time are provided in red. Bottom row: The probability of meeting the listed objectives over time. The horizontal dashed gray line is located at 95% probability.....	171
Figure 4.7: The time after geothermal production began at which the joint probability of a Trenton-Black River production scenario meeting a temperature of 60 °C and heat production of 5 MW <sub>th</sub> declines to less than 95%.....	172
Figure 4.8: Rock temperature over 40 years of production from the Trenton-Black River (TBR) reservoir for the condition of flow through matrix porosity for the warmest 5 <sup>th</sup> percentile initial rock temperature conditions. ....	174
Figure 4.9: Pressure over 40 years of production from the Trenton-Black River (TBR) reservoir for the warmest 5 <sup>th</sup> percentile initial rock temperature conditions. ....	175
Figure 4.10: Estimated heat production and temperature over time for the Trenton-Black River for injection of 23.5 °C fluid. ....	177
Figure 4.11: Multiple parallel fractures temperature production results for basement rock at 3 – 3.5 km depth for 20 °C injection at 50 kg/s. ....	178
Figure 4.12: The time at which the joint probability of a basement reservoir production scenario meeting a temperature of 60 °C and heat production of 5 MW <sub>th</sub> becomes less than 95%.....	179

## Appendix A

Figure A.1: Type I error (%) as a function of $k$ for the asymmetric boxplot outlier criterion based on 100,000 replicates of the symbolized number of points.....	203
Figure A.2: Real-space mean squared error (MSE) of the mean as a function of $k$ for the asymmetric boxplot outlier criterion based on 10,000 replicates of size 50 from a lognormal distribution with the colored real-space coefficient of variation (CV) and log-space mean of 0. The MSE is provided before and after outlier censoring using the asymmetric boxplot. ....	204

## Appendix B

Figure B.1: Locations of records with negative geothermal gradients (red), and records for which the original source was inspected (Table 2.1, Table B.1).....	205
--	-----

Figure B.2: Histogram showing by state the number of records that are in locations (on the map) with multiple BHTs.....	207
Figure B.3: Surface heat flow vs. BHT depth of measurement for locations with multiple BHTs. ....	208
Figure B.4: Boxplots of the average sample nugget semi-variance for records in the same spatial location within each geologic (interpolation) region.....	210
Figure B.5: Local median deviation for the surface heat flow values for wells in Figure 2.4 in northwestern Pennsylvania. ....	212
Figure B.6: Points that were not tested in the local median deviation analysis for the surface heat flow (red). All other points were tested. ....	213
Figure B.7: Average semi-variance of the surface heat flow (solid points) with jackknife 95% confidence intervals (open points connected with dashed lines) within the geologic regions provided in Figure 2.8.....	216
Figure B.8: Average semi-variance of the surface heat flow (solid points) with jackknife 95% confidence intervals (open points) within the geologic regions provided in Figure 2.8. ....	217
Figure B.9: Average semi-variance of the surface heat flow within the geologic regions provided in Figure 2.8. This figure illustrates the sensitivity of the spatial outlier detection results to the value of the asymmetric boxplot outlier criterion parameter, $k$ , in Equation 2.1.....	218

## Appendix C

Figure C.1: Radiogenic heat generation rate with depth estimated from a spectral gamma ray log for Wetzel-2973 well in West Virginia (API: 4710302973; WVGES, 2019). ....	220
---	-----

## Appendix D

Figure D.1: Locations of reference wells for lithologic properties and well logs near the Cornell project site. ....	230
Figure D.2: Temperature digital log from the Lansing T1 well located about 20 km north of Cornell. ....	233
Figure D.3: Vertical one-dimensional variograms for Trenton density for wells in Table D.5.....	239
Figure D.4: Vertical one-dimensional variograms for Trenton porosity for wells in Table D.5.....	240
Figure D.5: Vertical one-dimensional variograms of Black River density well logs for wells in Table D.8. ....	244

Figure D.6: Vertical one-dimensional variograms for Black River porosity for wells in Table D.8.....	245
Figure D.7: Scanlines (red) used to sample fractures with orientations of $45^\circ \pm 15^\circ$ from North (black). Fractures with other orientations are shown in gray. ....	252
Figure D.8: Scanlines (red) used to sample fractures with orientations of $95^\circ \pm 15^\circ$ from North (black). Fractures with other orientations are shown in gray. ....	253
Figure D.9: Scanlines (red) used to sample fractures with orientations of $135^\circ \pm 15^\circ$ from North (black). Fractures with other orientations are shown in gray. ....	254

## LIST OF TABLES

### Chapter 2

Table 2.1: Types of data recording errors for fourteen inspected records with negative geothermal gradients. ....	19
Table 2.2: Causes for records in the same spatial location, data processing methods for those records, and number of locations affected and retained for further analysis. ....	22

### Chapter 3

Table 3.1: Probability distributions for input variables in the uncertainty and sensitivity analyses of temperature-depth profiles (first seven variables) and surface heat flow (last seven variables) using the heat conduction model. ....	73
Table 3.2: Uncertainty levels and probability distributions used for reservoir thickness and porosity. The bounds of the symmetric triangular distribution are provided as a percentage of the mean value. ....	83
Table 3.3: Density and heat capacity for rocks by lithology. ....	85
Table 3.4: Probability distributions selected for variables in the uncertainty analysis and the sensitivity analysis of the stored thermal energy for each 1 km <sup>2</sup> area of each reservoir. ....	86
Table 3.5: Regression models and parameter estimates for the stored thermal energy sensitivity effects. ....	119

### Chapter 4

Table 4.1: Estimates of the 5 <sup>th</sup> , 50 <sup>th</sup> , and 95 <sup>th</sup> percentiles of temperatures at selected depths below Cornell, and corresponding geothermal gradients. ....	143
Table 4.2: Generalized geologic column below Cornell with estimated formation depths, geologic properties, and grid cell sizes used in numerical geothermal reservoir simulations. ....	147
Table 4.3: Summary of parameters used in GEOPHIRES analytical reservoir models. ....	166
Table 4.4: Summary of production scenarios for geothermal reservoir simulations. ....	167

## Appendix A

Table A.1: Type I error (%) for the asymmetric boxplot outlier criterion based on 100,000 replicates of sample size 25.....	202
Table A.2: Type I error (%) for the asymmetric boxplot outlier criterion based on perfect knowledge of the parameters (large sample). .....	202

## Appendix B

Table B.1: State and API numbers for inspected records (blue points in Figure B.1).....	206
Table B.2: Estimates of mean semi-variance for selected separation distance lags after each major ESDA procedure: Dataset 1) pre-ESDA, Dataset 2) data shallower than 1 km removed, and Dataset 3) post-ESDA with outliers removed.....	214

## Appendix D

Table D.1: Deep wells with density (+) and porosity (*) logs near the Cornell project site. Some of these wells were also used to estimate formation tops at the Cornell project site ( <sup>T</sup> ).....	231
Table D.2: Wells with temperature logs near Cornell. ....	232
Table D.3: Wells used to estimate Utica density. ....	237
Table D.4: Quantile estimates of density (g/cm <sup>3</sup> ) for the Utica formation from wells in Table D.3.....	237
Table D.5: Wells used to estimate Trenton density and porosity. ....	238
Table D.6: Quantile estimates of density (g/cm <sup>3</sup> ) for the Trenton formation from wells in Table D.5.....	238
Table D.7: Quantile estimates of (decimal) porosity for the Trenton formation from wells in Table D.5.....	238
Table D.8: Wells used to estimate Black River density and porosity.....	241
Table D.9: Quantile estimates of density (g/cm <sup>3</sup> ) for the Black River dolomite from wells in Table D.8.....	241
Table D.10: Quantile estimates of (decimal) porosity for Black River dolomite from wells in Table D.8.....	242
Table D.11: Quantile estimates of density (g/cm <sup>3</sup> ) for the Black River limestone from wells in Table D.8. ....	242
Table D.12: Quantile estimates of (decimal) porosity for Black River limestone from wells in Table D.8. ....	243

Table D.13: Wells used to estimate Beekmantown Group density and porosity.....	246
Table D.14: Quantile estimates of density (g/cm <sup>3</sup> ) for the Beekmantown Group from wells in Table D.13. ....	246
Table D.15: Quantile estimates of (decimal) porosity for the Beekmantown Group from wells in Table D.13. ....	246
Table D.16: Wells used to estimate Galway/Theresa density and porosity.....	247
Table D.17: Quantile estimates of density (g/cm <sup>3</sup> ) for the Galway/Theresa from wells in Table D.16.....	247
Table D.18: Quantile estimates of (decimal) porosity for the Galway/Theresa from wells in Table D.16.....	248
Table D.19: Wells used to estimate Potsdam density and porosity. ....	249
Table D.20: Quantile estimates of density (g/cm <sup>3</sup> ) for the Potsdam from wells in Table D.19.....	249
Table D.21: Quantile estimates of (decimal) porosity for the Potsdam from wells in Table D.19.....	250
Table D.22: Scanline fracture spacings and fracture lengths estimated from the scanlines in Figure D.7. ....	252
Table D.23: Scanline fracture spacings and fracture lengths estimated from the scanlines in Figure D.8. ....	253
Table D.24: Scanline fracture spacings and fracture lengths estimated from the scanlines in Figure D.9. ....	254
Table D.25: Parameter estimates, standard errors, and significance levels for the censored regression on maximum permeability presented in Equation 4.2.....	255
Table D.26: Parameter estimates, standard errors, and significance levels for the censored regression on 90° permeability presented in Equation 4.2.....	255

# CHAPTER 1

## INTRODUCTION

### **1.0 Background and Research Motivation**

Cold-climate regions in the Northern Tier of the United States currently rely primarily on combustion of fossil fuels to meet space and water heating demands. In Northeast and Midwest states, some 65% - 70% of residential heating demands are met by non-electric fossil fuel sources (EIA, 2018a; 2018b). Thus, current widespread efforts to transition to renewable electricity will not reduce a large source of carbon emissions that are attributable to heating buildings in the Northern Tier. Additionally, combustion of fossil fuels is generally at higher temperatures than are required by the end uses (e.g. Fox et al., 2011). Meeting the heating needs of such low-temperature utilization processes with a lower carbon emission alternative would be desirable. Direct-use of heat from deep (> 1 km) geothermal reservoirs could provide baseload heating as part of a long-term solution to renewable energy portfolios for population centers and industrial users of heat in this region of the country (e.g. Jordan et al., 2017; Tester et al., 2006).

In the Northern Tier of the U.S., several sedimentary basins have been explored for oil and gas resources. Exploration has provided subsurface data with which to inform geothermal resource assessments. However, uncertainties in geothermal resource variables (e.g. the surface heat flow) that are estimated from available temperature and geologic data are commonly large in the early phases of geothermal project feasibility studies, especially since oil and gas datasets were not collected for the purpose of geothermal energy development (e.g. Blackwell et al., 1999). Despite the limitations of the available data, quantification of geothermal resource uncertainty in these early phases of geothermal projects is often ignored (Witter et al., 2019). A thorough evaluation of uncertainty in geothermal resources derived from available datasets, and

propagation of uncertainty from the resource exploration phase through to conceptual reservoir modeling and utilization system planning phases of a project is particularly important to inform project decisions because uncertainties can be so large.

This thesis presents methods of exploratory spatial data analysis that aim to identify discordant observations and to reduce uncertainty in subsurface temperature estimates, and stochastic methods to propagate uncertainty associated with available datasets through geothermal resource assessment and reservoir models. These methods allow for probabilistic interpretations of the likelihood of meeting community heating objectives that may be used to inform risk and financial analyses for project decision making – all before a well is drilled at a site. Such assessments are critically important to evaluate the potential success of any geothermal project, as the largest financial risk for new geothermal projects, or “greenfield” projects (e.g. Hadi et al., 2010), is the cost associated with drilling the first well (e.g. Lowry et al., 2017).

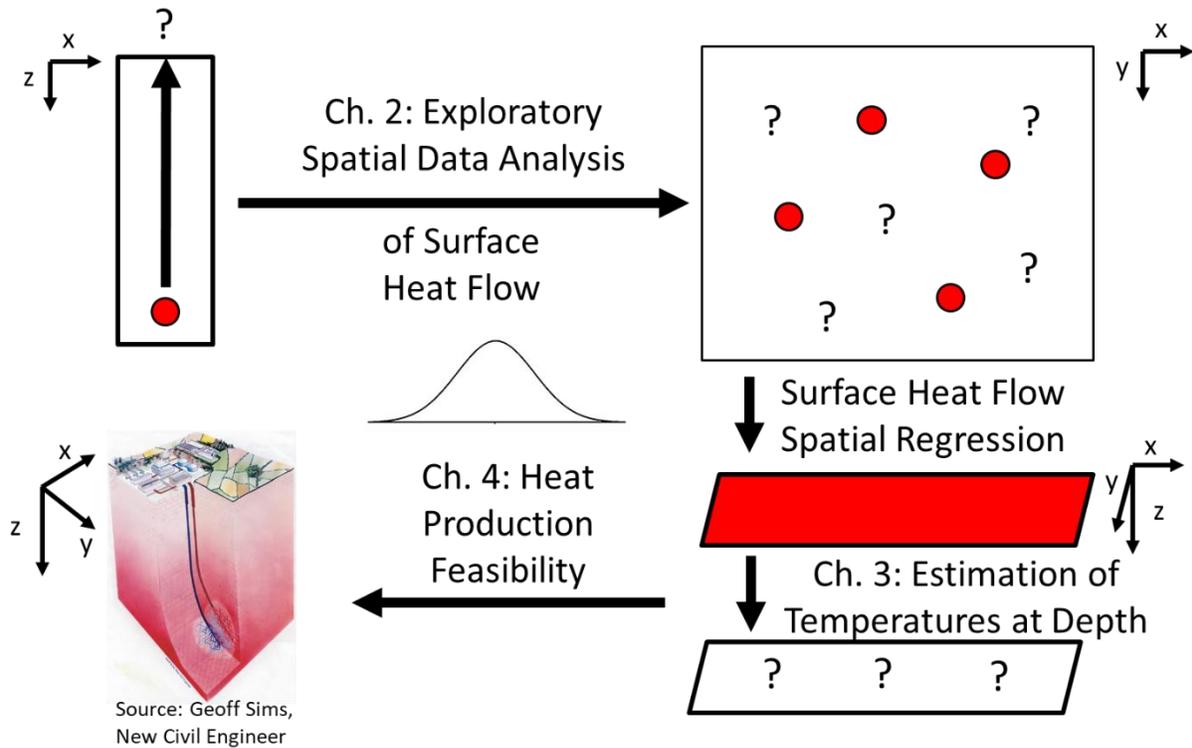
The methods presented in this thesis are applied to heat conduction-dominated geothermal systems in the north-central Appalachian Basin states of New York, Pennsylvania, and West Virginia. The following paragraphs discuss the motivation for each of the chapters and summarize the primary methodologies developed therein.

## **1.1 Chapter Organization**

This thesis develops exploratory spatial data analysis methods and stochastic methods for uncertainty propagation that are especially useful before site-specific knowledge of a geothermal resource is available. Figure 1.1 provides a graphical abstract of the research progression and the methods developed in this thesis. These methods are applied within two phases of geothermal projects.

Chapter 2 and Chapter 3 are applied to the geothermal exploration phase at regional or basin scales. In this phase, bottom-hole temperature data from oil and gas wells are used to estimate the surface heat flow at well locations (Figure 1.1, top left). The surface heat flow estimates at well locations are then used in spatial regressions to estimate the surface heat flow at specified locations across a region of interest (Figure 1.1, top right). Finally, temperatures at depth and stored thermal energy within a rock volume are estimated using the surface heat flow as the input thermal information (Figure 1.1, bottom right).

Chapter 4 uses the results of Chapter 2 and Chapter 3 to evaluate the feasibility of potential geothermal reservoirs in the Appalachian Basin to meet heating objectives (Figure 1.1, bottom left). The stochastic methods for geothermal reservoir modeling presented in Chapter 4 can inform decisions in the planning phase of a geothermal project. In aggregate, this thesis provides a framework for uncertainty analysis of geothermal greenfields from data acquisition to uncertainty characterization and translation to project risk analysis.



**Figure 1.1:** Graphical abstract of the research progression and methods developed in this thesis. Large black arrows indicate the progression of research phases, which begin in the top left. Red items are known or estimated quantities, and question marks indicate where geothermal resource information is estimated in each phase. The uncertainty distribution in the middle indicates that uncertainty is considered in all research phases.

### 1.1.1 Chapter 2: Exploratory Spatial Data Analysis for Geothermal Resource Assessments

Chapter 2 presents exploratory spatial data analysis (ESDA) procedures for geothermal datasets that are derived from oil and gas well bottom-hole temperature (BHT) measurements. BHT datasets are often abundant in sedimentary basins; however, BHT measurements are affected by drilling related thermal disturbances and data recording errors such that the measurements may not represent thermal equilibrium conditions (e.g. Deming, 1989; Whealton, 2016). For accurate geothermal resource assessments, the temperature data must be representative of the physical heat transfer processes that are expected or assumed in the models of a region. A local median deviation procedure is used to inform the selection of data that are

more likely representative of conduction-dominated heat transfer conditions, which is an assumption made in this analysis of the Appalachian Basin.

Spatial regressions are commonly used to estimate geothermal resources, such as the surface heat flow, across a region of interest using point values of surface heat flow derived from BHTs at those well sites. For least-squares regression analysis, spatially discordant observations in derived datasets are of concern because they may cause large residual errors that are unrepresentative of actual geothermal resource variability. Spatially discordant observations can also affect the precision of spatial correlation estimates, which are used in geostatistical spatial regression analyses. This chapter presents a local spatial outlier detection algorithm to identify spatially discordant observations in datasets derived from BHTs. Statistical diagnostics are used to evaluate the impact of removing identified discordant observations on the estimation of spatial correlation in an Appalachian Basin geothermal dataset.

### **1.1.2 Chapter 3: Stochastic Uncertainty Analysis and Sensitivity Analysis of Geothermal Resource Assessments for Direct-Use Heating**

After ESDA methods are applied to the geothermal variable in the dataset derived from BHTs, spatial regression is used to estimate values of the geothermal variable across a region of interest. Chapter 3 uses the results of a geostatistical estimation that provided the mean and standard error of the surface heat flow in the Appalachian Basin. To estimate uncertainty distributions of computed temperatures at depth within the Appalachian Basin, a stochastic methodology is developed to propagate uncertainty in geologic properties, and uncertainty in the surface heat flow that results from spatial interpolation uncertainty. These uncertainty distributions for temperature-depth profiles are used to estimate uncertainty distributions of

stored thermal energy within potential geothermal reservoirs. The results from this chapter update regional maps of the Appalachian Basin geothermal resource, with a particular advance in providing prediction uncertainties that reflect spatial correlation uncertainty and uncertainties in geologic properties.

Geothermal resource assessments often rely on low quality or limited temperature and geologic property data to estimate temperatures at depths of up to several kilometers and the stored thermal energy in rock volumes. With many uncertain variables, it is not initially clear which variables will contribute most to temperature uncertainty and stored thermal energy uncertainty for a site or reservoir of interest. Sensitivity analyses are presented for the models used to compute temperatures at depth and stored thermal energy in order to identify spatially which variables contribute most to the uncertainty in these variables. This exercise provides guidance for how to allocate time and financial resources to reduce uncertainty in resource estimates for sites of interest.

### **1.1.3 Chapter 4: Uncertainty Analysis of Geothermal Reservoir Simulations for Geothermal Greenfield Feasibility Assessments**

Transformation of geothermal resource uncertainty characterizations to decision-relevant information is critical for informed project planning (Witter et al., 2019). Using the estimated uncertainty distributions of temperatures at depth from Chapter 3, Chapter 4 develops uncertainty propagation methods using stochastic modeling of geothermal reservoirs. Geothermal reservoir production results are translated into probabilistic interpretations of the feasibility of produced fluids to meet target heating objectives over time. The case study for this chapter evaluates the feasibility of using geothermal production from either of two target

reservoirs to meet the heating objectives of Cornell University's main campus in Ithaca, New York.

#### **1.1.4 Chapter 5: Conclusions and Future Research Opportunities**

Chapter 5 summarizes the major findings and conclusions of this thesis. It includes a discussion of areas of future research that would build upon the results obtained in each chapter. Future research ideas for the resource exploration phase of geothermal projects address: 1) outlier detection that considers uncertainty in the derived geothermal variable at well locations, 2) the use of these uncertainties at well locations and additional basin characteristics to update geostatistical regressions of the Appalachian Basin surface heat flow, and 3) using additional local well log data to refine geothermal resource estimates for specific sites. Future research ideas for the geothermal project planning phase address using the probabilistic reservoir production results from Chapter 4 to inform probabilistic assessments of surface utilization system performance.

## REFERENCES

- Blackwell, D.D., G.R. Beardsmore, R.K. Nishimori, and R.J. McMullen Jr.. (1999). High-resolution temperature logs in a petroleum setting: Examples and applications. In A. Förster and D.F. Merriam (Eds.), *Geothermics in Basin Analysis*. Kluwer Academic / Plenum Publishers. pp. 1–34. <https://doi.org/10.1007/978-1-4615-4751-8>
- Deming, D.. (1989). Application of bottom-hole temperature corrections in geothermal studies. *Geothermics*, 18. pp. 775 - 786. [https://doi.org/10.1016/0375-6505\(89\)90106-5](https://doi.org/10.1016/0375-6505(89)90106-5)
- EIA. (2018a). TableCE2.3 Annual household site fuel consumption in the Midwest -totals and averages, 2015. Available online: <https://www.eia.gov/consumption/residential/data/2015/c&e/pdf/ce2.3.pdf>
- EIA. (2018b). TableCE2.2 Annual household site fuel consumption in the Northeast -totals and averages, 2015. Available online: <https://www.eia.gov/consumption/residential/data/2015/c&e/pdf/ce2.2.pdf>
- Fox, D.B., D. Sutter, and J.W. Tester. (2011). The thermal spectrum of low-temperature energy use in the United States. *Proceedings Thirty-Sixth Workshop on Geothermal Reservoir Engineering*. Stanford, CA. SGP-TR-191. 15 p.
- Hadi, J., P. Quinlivan, G. Ussher, O. Alamsyah, B. Promono, and A. Masri. (2010). Resource risk assessment in geothermal greenfield development: An economic implications. *Proceedings World Geothermal Congresss 2010*. Bali, Indonesia.
- Jordan, T.E., F.G. Horowitz, J.R. Stedinger, J.W. Tester, E.R. Camp, C.A. Whealton, J.D. Smith, B.J. Anderson, K. Welcker, X. He, M.C. Richards, C. Chickering-Pace, M. Hornbach, Z.S. Frone, C. Ferguson, R. Bolat, and M.B. Magnani. (2017). Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. U.S. Dept. of Energy Award No. DE-EE0006726. Retrieved from <http://gdr.openei.org/submissions/899>.
- Lowry, T.S., J.T. Finger, C.R. Carrigan, A. Foris, M.B. Kennedy, T.F. Corbett, C.A. Doughty, S. Pye, and E.L. Sonnenthal. (2017). Reservoir maintenance and development task report for the DOE geothermal technologies office GeoVision study. Sandia Report SAND2017-9977. 81 p.
- Tester, J.W., B.J. Anderson, A.S. Batchelor, D.D. Blackwell, R. DiPippo, E.M. Drake, J. Garnish, B. Livesay, M.C. Moore, K. Nichols, S. Petty, M.N. Toksoz, R.W.J. Veatch, R. Baria, C. Augustine, E. Murphy, P. Negraru, and M.C. Richards. (2006). *The Future of Geothermal Energy*. The Future of Geothermal Energy.
- Whealton, C.A.. (2016). Statistical data analysis, global sensitivity analysis, and uncertainty propagation applied to evaluating geothermal energy in the Appalachian Basin. Cornell University, Ph.D. Thesis. 267 p.
- Witter, J.B., W.J. Trainor-Guitton, and D.L. Siler. (2019). Uncertainty and risk evaluation during the exploration stage of geothermal development: A review. *Geothermics*, 78. Pp. 233 - 242.

## CHAPTER 2

### EXPLORATORY SPATIAL DATA ANALYSIS FOR

### GEOHERMAL RESOURCE ASSESSMENTS: AN APPALACHIAN BASIN CASE STUDY

#### **2.0 Abstract**

The most abundant temperature-depth data for geothermal studies in sedimentary basins are low-quality bottom-hole temperature (BHT) measurements. BHT measurement errors result from drilling-related thermal disturbances and data recording problems. The precision of variables derived from BHTs, e.g. the surface heat flow, which is independent of the BHT depth, is limited by those errors. Large errors may result in spatially discordant observations. This paper considers detection of discordancy in derived datasets, with the goal of improving geothermal resource assessments. This paper summarizes possible causes of discordancy, and formalizes detection methods using exploratory spatial data analysis (ESDA) procedures. ESDA procedures include (i) a local median deviation analysis used to select a minimum depth for BHT measurements to avoid near-surface thermal disturbances, and (ii) local spatial outlier detection using asymmetric boxplots. Identified outliers are examined spatially and with depth to understand possible causes of the discordancy. Applying our ESDA procedures to an Appalachian Basin dataset classified 37% of the records as being too shallow for our resource assessment, and identified 7.6% of the remaining records as local spatial outliers. Removing these records sometimes substantially reduced semi-variance estimates of spatial autocorrelation and stabilized the autocorrelation trend for critical cases, illustrating the value of employing ESDA procedures with BHT datasets as part of geothermal resource assessments. Evaluating spatial autocorrelation within spatially distinct geologic regions of the basin revealed striking differences between regions.

## 2.1 Introduction

Geothermal resource assessments aim to predict the surface heat flow or subsurface temperatures (e.g., Stutz et al., 2015), or to evaluate the thermal energy stored within rock volumes (e.g., Garg and Combs, 2015; Tester et al., 2006). These resource assessments typically have limited high-quality temperature-depth information (e.g., equilibrium temperature logs and drill stem tests; Blackwell et al., 1999). To supplement high-quality information and to fill in spatial data gaps, low-quality well bottom-hole temperature (BHT) datasets are commonly employed; these are often affected by drilling-related thermal disturbances so they would not represent thermal equilibrium conditions. BHTs may be adjusted to approximate thermal equilibrium with correction equations that provide a relationship between the thermal disturbance and factors such as measurement depth, rock properties, time since drilling ceased, or the type of drilling fluids (e.g., Deming, 1989; Whealton et al., 2015). While such corrections address systematic deviations that can be estimated using the factors included in the equations, discordant observations can remain because of a host of related and unrelated random factors (Aguirre et al., 2013).

Exploratory data analysis (EDA) is a recommended practice to detect discordant (Barnett and Lewis, 1994) or “rogue” (Section 2.1.1) observations in feature-space statistical analyses; exploratory spatial data analysis (ESDA) is an extension of EDA to include geographic- and feature-space domain information (e.g., Anselin, 1999). EDA, and by extension ESDA, are essential precursors for reliable empirical statistical studies. In this paper, ESDA procedures are developed for use with geothermal datasets to identify potentially rogue observations so that they may be analyzed and corrected or removed before spatial regressions are employed for geothermal resource assessments. Analyzing potentially rogue observations may lead to

discoveries of real thermal anomalies that are of interest in geothermal resource assessments. Potentially rogue observations that cannot be explained by the available information may be indicative of bad data; correcting or removing these observations may improve regression parameter estimates and estimates of spatial autocorrelation, as we demonstrate.

### **2.1.1 Definition and Causes of Rogue Observations in Geothermal Datasets**

A rogue observation is defined in this paper as a point that is unrepresentative of the geothermal processes that are being considered in a study. We consider a sedimentary basin that is heat conduction-dominated at depths greater than a certain position; the appropriate minimum depth for a BHT observation is to be discovered through analysis of the dataset (Section 2.4.1). Having made this heat conduction assumption, we transform our BHT measurements into a dataset of a derived variable, the (ground) surface heat flow, computed using an analytical one-dimensional heat conduction model (Smith and Horowitz, 2017; code available in Horowitz et al., 2015). Transforming the BHT data to surface heat flow data removes the vertical trend in expected temperature with depth that results from the geothermal gradient, and refers all data to a common surface to which ESDA methods may be applied. We presume that the heat flow dataset will be used for geostatistical interpolations or other spatial regressions in a geothermal resource assessment (e.g., Stutz et al., 2015). For this study, rogue surface heat flow observations may be caused by erroneous BHT-depth data, or by inappropriate modeling assumptions for particular BHT spatial locations (Sections 2.1.1.2 – 2.1.1.4). The provided definitions and causes of rogue observations are broadly applicable to datasets used for geothermal resource assessments; our examples focus on conduction-dominated geothermal systems.

### **2.1.1.1 Influence of Rogue Observations on Geothermal Resource Predictions**

A critical component of geographic-space analysis is the estimation of spatial autocorrelation of the spatial process to be modeled. When spatially discordant rogue observations exist in a dataset, noise is introduced that affects the estimation of the spatial autocorrelation signal (e.g., the semi-variogram) at potentially many length scales (Kerry and Oliver, 2007; Section 2.4.3, this paper), although robust semi-variogram parameter estimation may mitigate these issues (Cressie and Hawkins, 1980; Marchant and Lark, 2007). In least-squares regression analysis of geothermal resources, rogue observations are of concern because they may cause high squared residual errors that are unrepresentative of actual geothermal variability. Rogue observations limit the estimated accuracy of the regression parameters, and may be influential in prediction if the selected parameter estimation method is not robust to extreme values (standard linear regression is not robust, Chatterjee and Hadi [1986]). This can be avoided to some degree if robust regression methods are used (e.g., Aguirre et al., 2013; Andersen, 2008); however, robust methods become less effective as the proportion of rogue observations in a dataset increases and approaches the breakdown point (Davies, 1993). Thus, ESDA procedures are also valuable precursors to robust regressions to improve parameter estimation.

### **2.1.1.2 Measurement Errors and Data Recording Errors in BHT Datasets**

Probably the most common interpretation of the cause of a rogue observation is a measurement error or a data recording error in the BHT dataset (e.g., spatial positioning errors or bad BHT-depth records). BHT measurements taken at shallow depths using maximum recording thermometers may provide erroneously high temperatures if the ground temperature doesn't

exceed the surface air temperature (Gray et al., 2012). Other measurement errors caused by the inherent precision of equipment (e.g., thermometers, positional surveying devices) are unavoidable, but are quantifiable by repeat measurements. Equipment measurement errors are generally not of concern as a cause of rogue observations; the maximum magnitude of equipment measurement error is likely small relative to the errors that can result from other causes.

If one suspects a data recording error generated during transfer from the original BHT data source to a digital format, the original source should be reviewed for agreement. Even if the records agree, the original source may have an error that is impossible to diagnose or correct. We examine the original BHT data source of some discordant observations and discuss this idea further in Section 2.3.1.

### **2.1.1.3 Neglected Physical Processes**

Datasets should be processed in a manner that considers the models in which they will be or have been used. Physical processes that are neglected in the models, or are not modeled appropriately within a specific region, can be causes of rogue data. These rogue data could reveal spatial regions for which an alternative model should be applied. For example, rogue surface heat flow data could result from thermal disturbances that are not properly addressed with BHT correction equations, like localized groundwater advection. Groundwater advection is also not addressed in heat conduction models, which may result in a spatially discordant surface heat flow estimate. We discuss possible causes of thermal disturbances further in Section 2.4.1, associated with the selection of a minimum depth for BHTs.

A rogue observation may also result from an inappropriate transformation from a measured quantity to a modeled quantity. For example, if the selected values of geological

properties in a heat conduction model do not reflect local geologic conditions, the calculated surface heat flow may be spatially discordant even if the BHT is accurate.

#### **2.1.1.4 Microscale Spatial Variability**

In geostatistics, a rogue observation could be a result of “microscale” spatial variation; see for example the discussion of the so-called “nugget effect” in Cressie (1988, p. 410).

Microscale is defined as spatial (geological) phenomena with length scales that are either 1) smaller than the spacing of observations, or 2) smaller than the length scales that may be used to reliably assess spatial autocorrelation (e.g. semi-variogram spatial lags). For example, an intrusive dike or roll-front uranium with anomalous concentrations of radiogenic heat producing elements may result in surface heat flow estimates that are discordant with the local or regional heat flow. Such anomalies could provide rogue heat flow data, which would interfere with resource assessments that aim to capture trends and stochastic behavior over larger spatial scales.

Both microscale variations and measurement errors are included in the nugget semi-variance, and there is no theoretical way to distinguish between them. The nugget semi-variance represents noise in a dataset at relatively small spacing (Cressie, 1988). It is possible to directly estimate the nugget for a site if multiple measurements are available, and to estimate for a region if many sites in the region have multiple measurements (e.g. Laslett and McBratney, 1990). A detailed estimation and analysis of the nugget semi-variance is beyond the scope of this paper; we provide an example estimation in supplementary material (Appendix B section 3) and employ the results in Section 2.3.2.

### **2.1.2 Paper Organization**

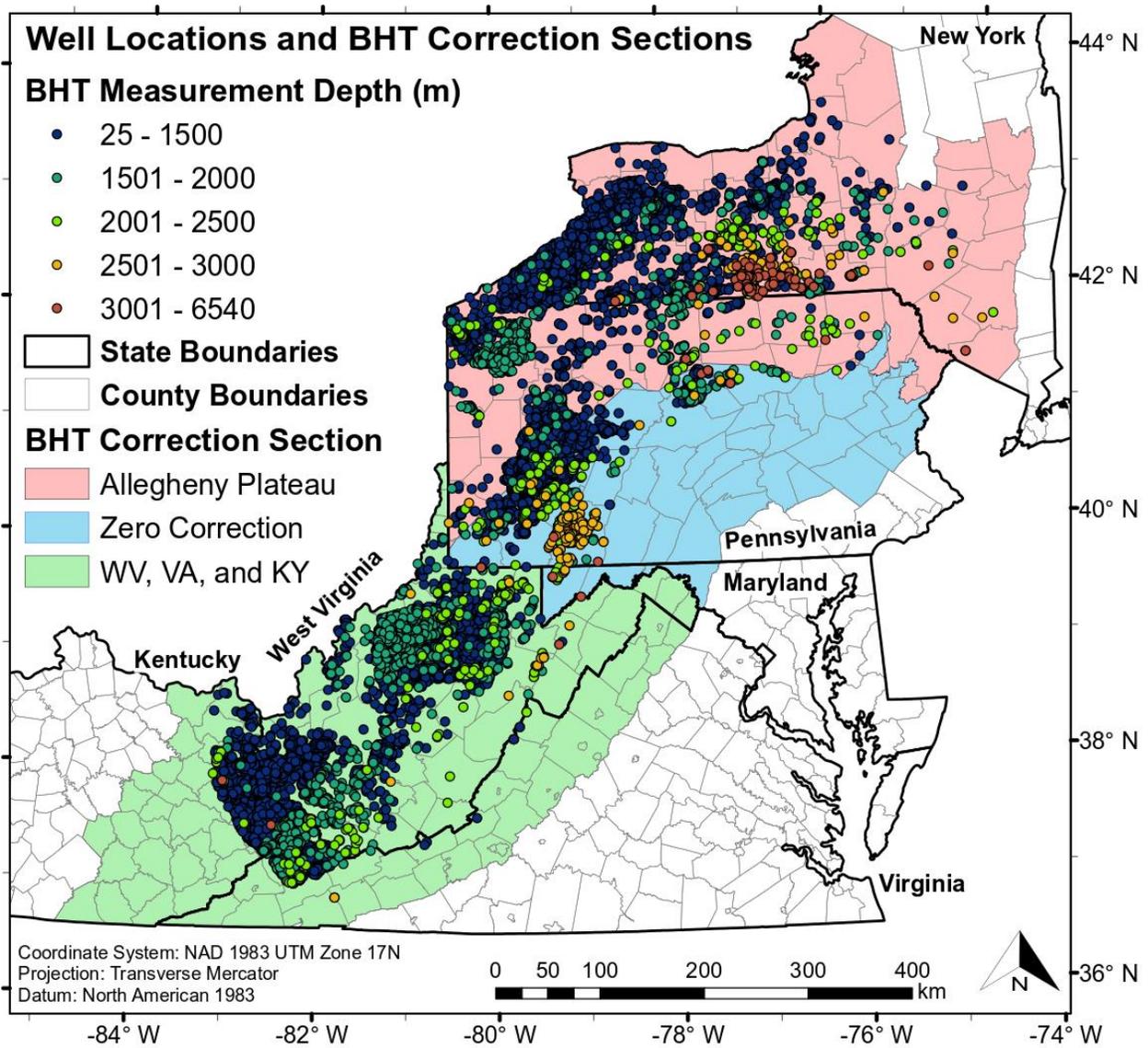
The following sections of this paper describe ESDA procedures to identify and analyze potentially rogue observations in geothermal datasets. We apply our ESDA procedures to an Appalachian Basin dataset, which contains BHTs that are used to compute surface heat flow estimates. Initial data processing is used to detect physically implausible rogue surface heat flow values, and to handle special cases. ESDA procedures are used to identify potentially rogue observations by informing the selection of a minimum depth for BHTs to be representative of heat conduction, and detecting local spatial outliers. We evaluate the performance of our ESDA procedures by computing and comparing spatial autocorrelation as estimated from the datasets that result after each procedure is applied. We conclude with a discussion of limitations of our analyses, and suggest analyses that could advance the methods presented in this paper.

### **2.2 Description of the Dataset**

The dataset used in this study was assembled for the Department of Energy funded Geothermal Play Fairway Analysis of the Appalachian Basin (Cornell University, 2017). The full dataset contains 32,385 bottom-hole temperature (BHT) measurements gathered from the American Association of State Geologists (AASG) Geothermal Data Repository for the states of New York, Pennsylvania, West Virginia and a 50 km radial buffer into surrounding Appalachian Basin states (Brezinski, 2011; Curl, 2011; Leftwich, 2011; Shank et al., 2012; Slater, 2012; Virginia Division of Geology and Mineral Resources (VDGMR), 2011; West Virginia Geological & Economic Survey (WVGES), 2011). A buffer is used to limit edge effects that would occur from using ESDA procedures near state boundary lines.

Only those records with a reported BHT and a specific BHT depth of measurement are retained for analysis in this paper. The depth filter eliminates those records for which the only reported depths are the vertical depth below the surface or a total borehole length, and there is no clarification of the position of the BHT within the borehole. This filter is necessary because BHT measurements are sometimes collected at intermediate positions in a borehole, reflective of the progressive stages of drilling or targets for logging. Even for the retained BHT records, the AASG source database lacks specificity of whether a BHT depth of measurement is a vertical depth below the surface or a total borehole length; this can be an undetectable cause of rogue data (Section 2.1.1.2).

The BHT measurements were adjusted to approximate thermal equilibrium using the Whealton et al. (2015) correction equations (code available in Whealton and Smith [2015]), which are defined for the spatial regions of the Appalachian Basin provided in Figure 2.1. The BHTs are spatially non-uniformly distributed within the basin volume, and generally cluster where oil and gas resources were targeted (Figure 2.1). The corrected BHTs were transformed into a surface heat flow dataset (Section 2.1.1), to which further analyses are applied. To compute the surface heat flow, a heat conduction model requires not only a BHT but also spatial information about the depth to basement rock (West Virginia Geological and Economic Survey [WVGES], 2006; processed dataset available in Smith and Horowitz, 2017) and stratigraphic information (Patchen et al., 1985a, 1985b). Spatial coverage of these geologic data reduced the analyzed dataset to 20,750 records; the dataset is available on the Department of Energy Geothermal Data Repository (Cornell University, 2016).



**Figure 2.1:** Locations of 20,750 wells used in this study, colored by the BHT measurement depth. Where points overlap in space, wells with deeper BHTs are plotted on top of wells with shallower BHTs. The BHT correction sections of Whealton et al. (2015) are provided for reference.

## **2.3 Initial Data Processing**

Initial data processing is used to identify physically implausible surface heat flow estimates, and to correct or remove them prior to using exploratory spatial data analysis (ESDA) (Section 2.4). We also handle observations that share the same spatial coordinates. Points in the same location usually differ in their surface heat flow value (Section 2.1.1), but a location should only be counted once in ESDA procedures. We describe a simple method to extract higher quality information from locations with multiple BHT observations.

### **2.3.1 Physically Implausible Values**

The calculated surface heat flow values were negative (downward) for 38 records, and the straight-line surface-to-BHT-depth geothermal gradient was negative for one additional record. For heat conduction-dominated systems, a negative surface heat flow can be calculated when the assumed annual average surface temperature for a site (Gass, 1982; Southern Methodist University [SMU] Geothermal Lab, 2016) is greater than the BHT measurement and the radiogenic heat generation term doesn't offset the negative gradient term in the heat conduction equation (e.g., Jaeger, 1965, sec. 4.1). A negative surface heat flow can also be caused by a rogue observation (Section 2.1.1).

To diagnose possible causes of these implausible values, we examined fourteen of the deeper records in New York, Pennsylvania, and West Virginia. We compared the reported uncorrected BHT and measurement depth in the AASG database with the original sources of well log data (available in the respective state databases [API unique identifiers provided in Appendix B section 1]: Bureau of Topographic & Geologic Survey, 2018; New York State Museum Hydrocarbon Reservoir Characterization Group, 2018; West Virginia Geological &

Economic Survey (WVGES), 2018). We identified several data recording errors in the original well logs, and also in data transfer from the well logs to the AASG database (Table 2.1). We corrected two typographical errors, which resulted in positive geothermal gradients and heat flows for those records, but we were unable to correct other data recording errors. Considering the time required to read each well log, and the amount of non-negative surface heat flow data available to use in further analyses, we classified all other records that provided negative gradients as rogue observations and removed them from the dataset.

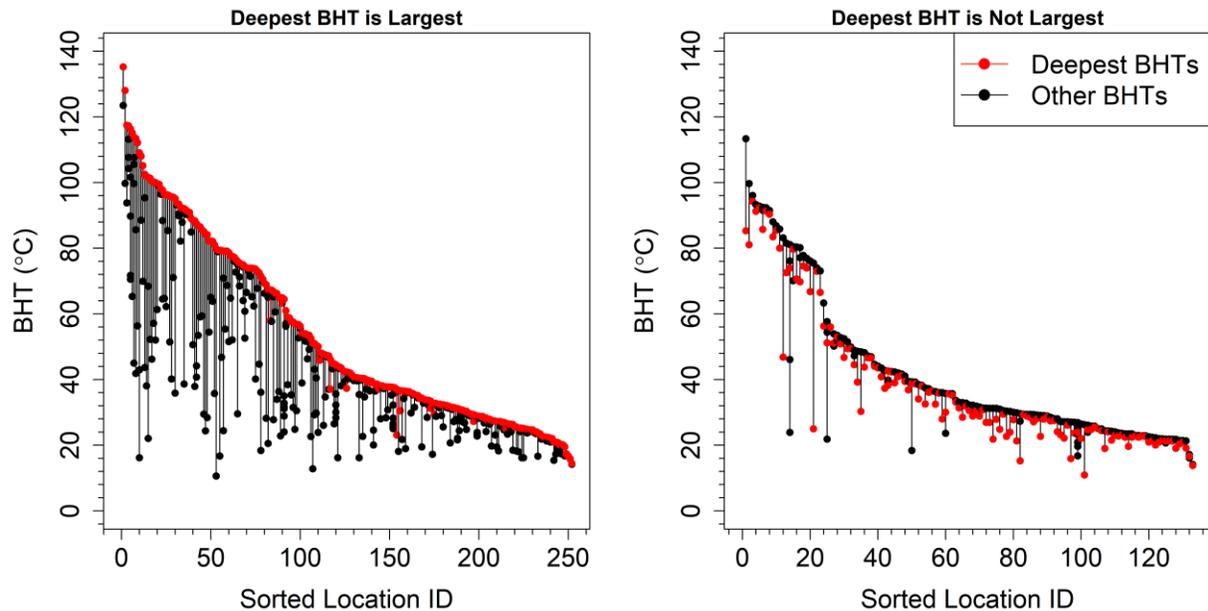
**Table 2.1:** Types of data recording errors for fourteen inspected records with negative geothermal gradients. Errors are either in the original well log sources or in the AASG state databases.

<b>Type of Error</b>	<b>Description</b>	<b>Number of Records</b>
Data Recording	The temperature scale used on the well log is likely incorrect or misinterpreted.	6
	The AASG BHT is less than the well log's temperature scale lower limit.	2
Or	The surface temperature was listed as the BHT in the AASG database. A BHT was not provided on the well log.	2
Data Transcription	Typographical error for the BHT in the AASG database.	2
	The well unique identifier (API number) in the AASG database was not listed in the state database. This record was not checked for agreement.	1
Physical Process	The BHT was taken in a gas producing interval. A maximum recording thermometer was not used.	1

### 2.3.2 Records with the Same Spatial Coordinates

The processed AASG database has 1190 records located in 478 unique latitude-longitude coordinates. Records with the same coordinates may occur because of: 1) a well with multiple BHTs, potentially for different depths or segments (e.g., after a failed “fishing” attempt for a drill string, an offshoot may be created [Friðleifsson et al., 2005]), 2) wells located on the same drill pad that have been assigned the pad coordinates, 3) a data recording error, or 4) duplicate records.

For records with the same spatial location, we retained only the surface heat flow that corresponded to the deepest BHT to limit the effects of groundwater and concerns about measurement devices at shallow depths (Frone et al., 2015; Gray et al., 2012; discussed further in Section 2.4.1). Deeper BHTs are also preferred because they generally provide more accurate temperature estimates. Plots of the BHTs for records in the same spatial locations are provided in Figure 2.2. About 65% of the locations with multiple BHTs had the deepest BHT as the largest. For the remaining 133 locations, about 50% of the locations have a temperature difference  $< 2$  °C, and only 33 locations have a temperature difference  $> 5$  °C.



**Figure 2.2:** BHTs for records in the same spatial locations. Locations are plotted on the x-axis and are sorted by the maximum BHT value. Vertical lines connect records in the same spatial location. The deepest BHT in a location is colored red; some locations had more than one BHT at the deepest position.

Multiple measurements at different depths could have been used to compute the surface heat flow, but there were complications that challenged such a computation. Some locations had several different BHT measurements at the same depth. The measurement date and time or shut-in-time were not available for these records, so it was impossible to distinguish if differences were indicative of progress through time toward thermal equilibration, or if one of the depths was incorrect. Blackwell et al. (2010) encounter a similar problem, and we infer that they retain the BHT value that is most similar to BHTs in nearby wells, and are more likely to use the data in their analysis if the BHTs are similar.

An adaptation of the Blackwell et al. (2010) reasoning was used for quality control on the locations with different BHTs at the same depth. A threshold of  $\pm 2$  °C was used to diagnose whether or not these BHTs were reliable. This threshold was selected based on the smallest estimated nugget semi-variance for the geologic regions within our study area (geologic regions

are displayed in Figure 2.8; see Appendix B section 3 for analysis). Using this method ensures that differences in the retained observations are within the estimated noise.

A summary of the data processing methods used to handle records in the same spatial locations is provided in Table 2.2. Only ten locations were dropped from the dataset as a result of these methods. After applying these methods, the surface heat flow dataset available for ESDA procedures consisted of 19,991 unique locations.

**Table 2.2:** Causes for records in the same spatial location, data processing methods for those records, and number of locations affected and retained for further analysis. Processing methods proceed sequentially down the table.

<b>Cause for Records in the Same Spatial Location</b>	<b>Data Processing Method</b>	<b>Locations Affected</b>	<b>Locations Retained</b>
Duplicate records	Retain only one record	93	93
Several BHTs at different depths	Retain the deepest BHT	385	385
Several BHTs at the deepest depth	Retain BHTs within $\pm 2$ °C	27	17
For some BHTs within $\pm 2$ °C, measurement depth may be mislabeled because other depth information is provided.	Retain BHT that more likely matches the reported depth of measurement, assuming temperature increases with increasing depth.	12	12
For other BHTs within $\pm 2$ °C, the cause is unknown.	Average BHTs and recompute the surface heat flow.	5	5

## **2.4 Exploratory Spatial Data Analysis**

The reasonableness of an observation's value depends on its spatial context. This is because we expect surface heat flow values within a small area to be similar (Anselin, 1999), unless there are abrupt geologic boundaries that affect heat transfer. Relying upon this hypothesis of physical continuity, each surface heat flow estimate was subject to exploratory spatial data analysis (ESDA) procedures that rely on local neighborhoods to identify spatial discordancy. The three ESDA procedures employed are 1) local median deviation, considering each datum's surface heat flow relative to the median of its local neighborhood (Section 2.4.1), 2) local spatial outlier detection using asymmetric boxplots (Section 2.4.2), and 3) quantile-quantile plots to scan for outliers in data-sparse areas that were unable to be tested using other procedures (Section 2.4.2.3). Each ESDA procedure identifies potentially rogue observations. We evaluate the impact of removing potentially rogue observations on estimates of surface heat flow spatial autocorrelation (Section 2.4.3).

### **2.4.1 Local Median Deviation Analysis for Selecting a Minimum Depth for BHTs**

If there are no rogue observations in a local area and the heat conduction model and associated geologic assumptions are reasonable, values of the surface heat flow should be similar locally regardless of the BHT measurement depth. We assessed our surface heat flow values by their BHT measurement depth to search for trends that would necessitate introducing a minimum depth threshold. Implementing a minimum depth threshold may be justified by potential near-surface thermal disturbances that are not considered in heat conduction models, and concerns about measurement device accuracy at shallow depths. Examples include: 1) the historical use of maximum recording thermometers to measure BHTs (Gray et al. [2012] use a 1200 m minimum

depth, which corresponds to the likely depth to the maximum summer surface air temperature); 2) shallow groundwater systems (500 m minimum depth in Hodge et al. [1981] in New York and Hendry et al. [1982] in West Virginia; 600 m minimum depth in several Southern Methodist Geothermal Laboratory studies [Frone et al., 2015] and [Stutz et al., 2015] for New York and Pennsylvania); and 3) surface temperature fluctuations on millennial (climatic) time scales, for which residual effects may remain at depths down to 600 m (Beltrami et al., 2015). In the Appalachian Basin, the maximum summer surface air temperature is about 35 °C, which would result in an air-temperature-based minimum depth of about 1 km on average across the basin (informed by our BHT dataset). The depth to 35 °C varies spatially, primarily as a result of spatial variability in the geothermal resource.

Deep groundwater systems may also affect BHT measurements (e.g. upwelling in Davis, 2012; downwelling in Rao and Rao, 1980; groundwater recharge in Majorowicz and Jessop, 1981; and lateral groundwater flow in Lewis and Beck, 1977). We do not have sufficient information to evaluate groundwater hydrology in the Appalachian Basin. Groundwater could be the cause of some rogue data in our analysis, and it would be useful to assess in greater detail if hydrogeological data become available.

#### **2.4.1.1 Local Median Deviation Analysis**

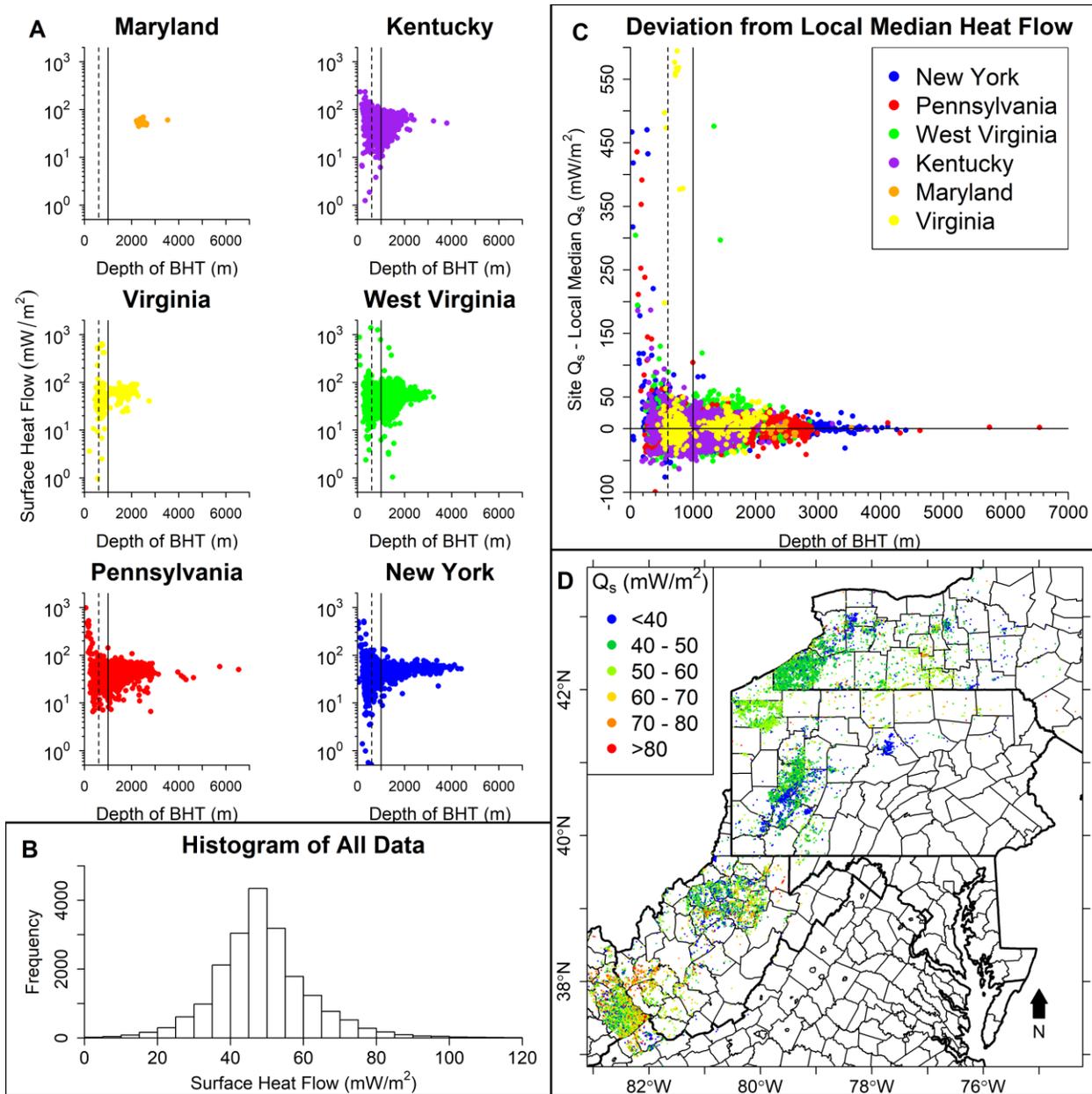
To inform the selection of a minimum BHT depth threshold, we use a local median deviation ESDA procedure for which each surface heat flow datum is compared to the median surface heat flow of its local neighbors. For our algorithm, there are three parameters to define the local neighborhood: 1) minimum number of neighbors, 2), maximum number of neighbors, and 3) maximum size, defined by the Euclidian radius. We choose a minimum of three neighbors

because with at least three points, for an estimate of central tendency, the median is more robust to rogue data than the mean (i.e. with only two points the median equals the mean). The median is robust to rogue data when fewer than 50% of the neighbors are rogue (Rousseeuw and Croux, 1993). We select a maximum number of 25 neighbors, for which the median will be robust even with 12 locally rogue observations. Because in our study the minimum depth threshold will be applied to all spatial locations, we selected a neighborhood radius of 10 km based on the spatial coverage of the points tested by the algorithm (a map is provided in Appendix B section 5). If there were fewer than 3 neighbors within 10 km, then a datum was considered to be without a neighborhood.

The surface heat flow data and the results of the local median deviation analysis are provided in Figure 2.3. Plots of the surface heat flow versus the BHT depth (Figure 2.3A) show that the calculated surface heat flow spans three orders of magnitude for BHTs that were measured shallower than 1 km, but generally spans only one order of magnitude for BHTs that were measured deeper than 1 km. For BHTs deeper than 1 km, most surface heat flow estimates are less than  $100 \text{ mW/m}^2$ . Based on previously published heat flow maps for the Appalachian Basin (Frone and Blackwell, 2010; Stutz et al., 2015), values of the surface heat flow greater than  $100 \text{ mW/m}^2$  are regionally extreme, but may be realistic if the observations are spatially clustered, potentially indicative of hot spots. Using local median deviation captures such spatial information that Figure 2.3A does not.

The local median deviation results are plotted against the BHT measurement depth in Figure 2.3C. Shallower than about 600 m the deviation from the local median tends to be larger, possibly attributable to effects mentioned in Section 2.4.1. Generally, the local median deviations are  $\pm 50 \text{ mW/m}^2$  for BHTs taken at depths between 600 m and 1000 m. The deviations become

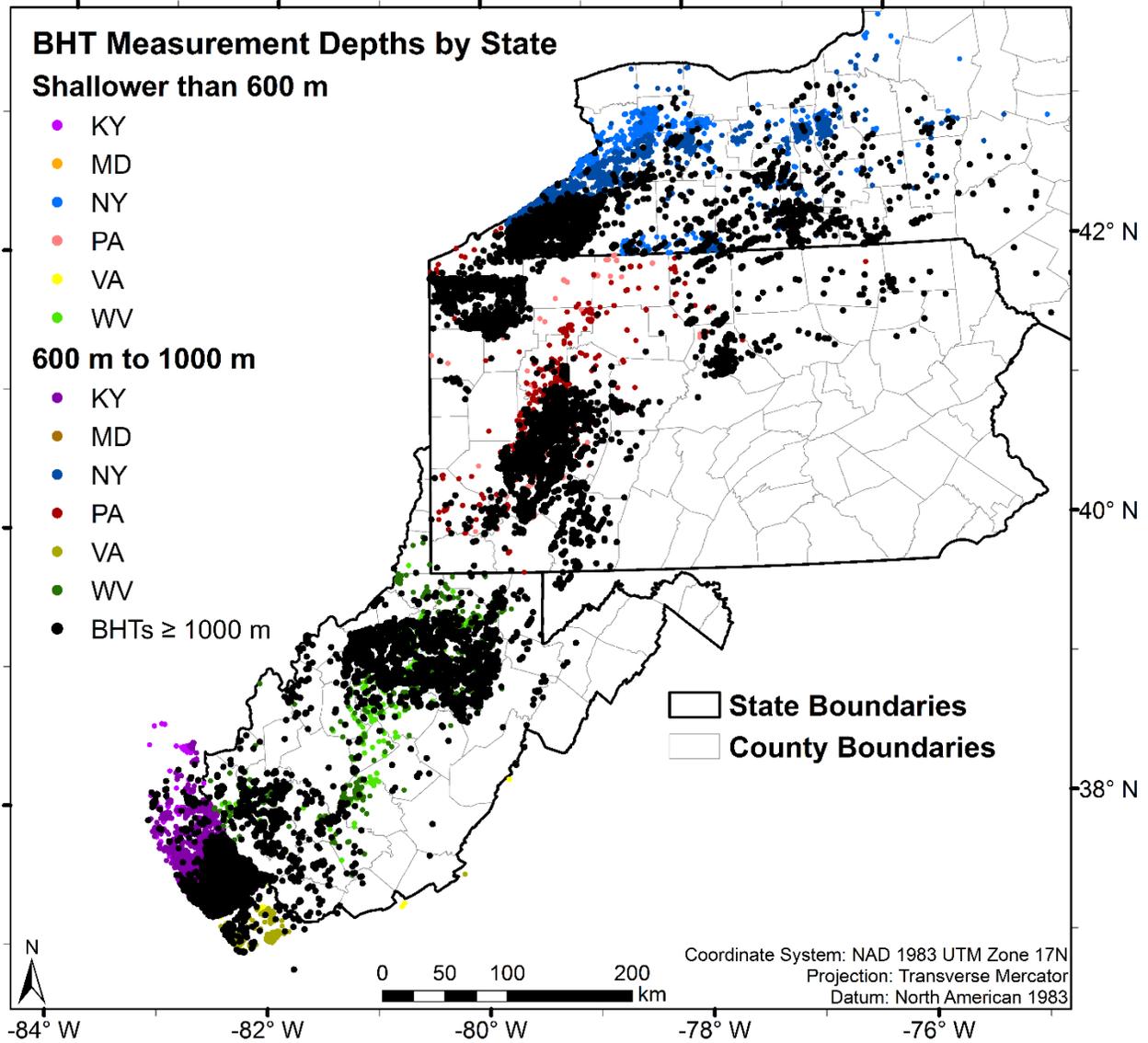
smaller with increasing depth to about  $\pm 20 \text{ mW/m}^2$  at 2000 m, and  $\pm 10 \text{ mW/m}^2$  deeper than 3000 m. Spatial clustering of the wells deeper than 3000 m (86% in the eastern Southern Tier of New York, Figure 2.1) may explain the consistency at these depths. Points that have large deviations from their local median may be rogue observations. Motivation for a local spatial outlier analysis (Section 2.4.2) is provided by the map of the surface heat flow values in Figure 2.3D, which reveals that the estimated surface heat flow values are spatially variable at the scale of the Appalachian Basin, thereby rendering global outlier tests inappropriate at the spatial scale of the basin.



**Figure 2.3:** Exploratory spatial data analysis of the surface heat flow ( $Q_s$ ). **A:**  $Q_s$  versus the depth of the BHT measurement by state. Vertical lines indicate the 600 m (dashed) and 1000 m (solid) minimum depth thresholds. **B:** Histogram of  $Q_s$ . **C:** Deviation of site  $Q_s$  from the local median  $Q_s$ . The y-axis has been trimmed on the positive side to show detail where there are smaller deviations. Vertical lines indicate the 600 m (dashed) and 1000 m (solid) minimum depth thresholds. **D:** Map of  $Q_s$ . Where points overlap, deeper data are plotted on top of shallower data.

### **2.4.1.2 Spatial Data Coverage from Possible Minimum Depth Thresholds**

The local median deviation analysis supports that the surface heat flow estimates become more consistent with increasing depth. We evaluate spatial data coverage to select a minimum depth threshold. Figure 2.4 provides the spatial distribution of the dataset based on the 600 m and 1000 m minimum depth thresholds. A 600 m threshold would remove approximately 10% of the dataset (1996 records). Removing these records would primarily limit spatial data coverage in northern New York. A 1000 m threshold would remove approximately 33% of the dataset (6676 records), but the remaining dataset includes most of the study area. Notable data gaps would occur in northern New York, northeastern Kentucky, and northwestern Pennsylvania (purple circle in Figure 2.4). Smith (2016, chap. 4) examined the surface heat flow data between 600 m and 1000 m in northwestern Pennsylvania using local median deviation (Section 2.4.1.1) and found that a 750 m minimum depth threshold was reasonable (Appendix B section 4). This finding suggests that the minimum depth threshold for BHTs may be better applied as spatially variable or site specific, but we do not investigate this further in this paper.



**Figure 2.4:** BHTs deeper than 1000 m (black) and data that were removed based on the 600 m (light colors) and 1000 m (dark colors) minimum depth thresholds. Where points overlap, deeper BHT data are plotted on top of shallower data. The color scheme for each state matches the Figure 2.3 scatterplots.

Based on these results, we opted to use a 1000 m minimum depth threshold, except in northwestern Pennsylvania where a 750 m threshold was used. This reduced the dataset to 13,315 records that are more likely representative of heat conduction. The remaining data

included some seemingly high and low heat flow values, which were tested for discordancy using a spatial outlier detection algorithm.

## **2.4.2 Spatial Outlier Detection and Analysis**

A spatial outlier detection procedure was used to scan our heat flow dataset for locally spatially discordant observations. An outlier can be defined as “an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data” (Barnett and Lewis, 1994, p. 7). In a spatial context, outliers may be “global” or regional outliers, relating to the whole dataset within the region of interest, or local, relating to a subset of the data defined by a local neighborhood (e.g., 25 closest points, or all points within 32 km). Schubert et al. (2014) provides a review of spatial outlier detection algorithms; most algorithms rely on a local neighborhood. In general, global outlier tests should be used with caution – if a large proportion of the dataset is clustered in a small area, the small variability within that area can bias the bounds (e.g. the “fences” of Carling, 2000) of a global outlier test. For geothermal applications, removing global outliers can bias cold regions to appear warmer, and warm regions to appear cooler; if the extremes of the global dataset are consistent with the heat flow values within their local neighborhoods, they should be retained.

### **2.4.2.1 Spatial Outlier Detection Algorithm and Sensitivity Analysis Results**

Our spatial outlier detection algorithm used the asymmetric boxplot outlier identification test proposed in Aguirre et al. (2013), and further developed by Whealton and Stedinger (2015). A similar asymmetric boxplot was used in Carter et al. (2009) based on the Carling (2000) bounds, but their methods were not applied to spatial datasets. This outlier test is relatively

robust to asymmetry and non-normality because it relies on sample order-statistics to estimate quantiles; thus, for sufficiently large neighborhoods, the test will be insensitive to the values of the largest and smallest observations (e.g., median outlier test in Lu et al. [2003]). The upper and lower bounds of the asymmetric boxplot outlier test are provided in Equation 2.1

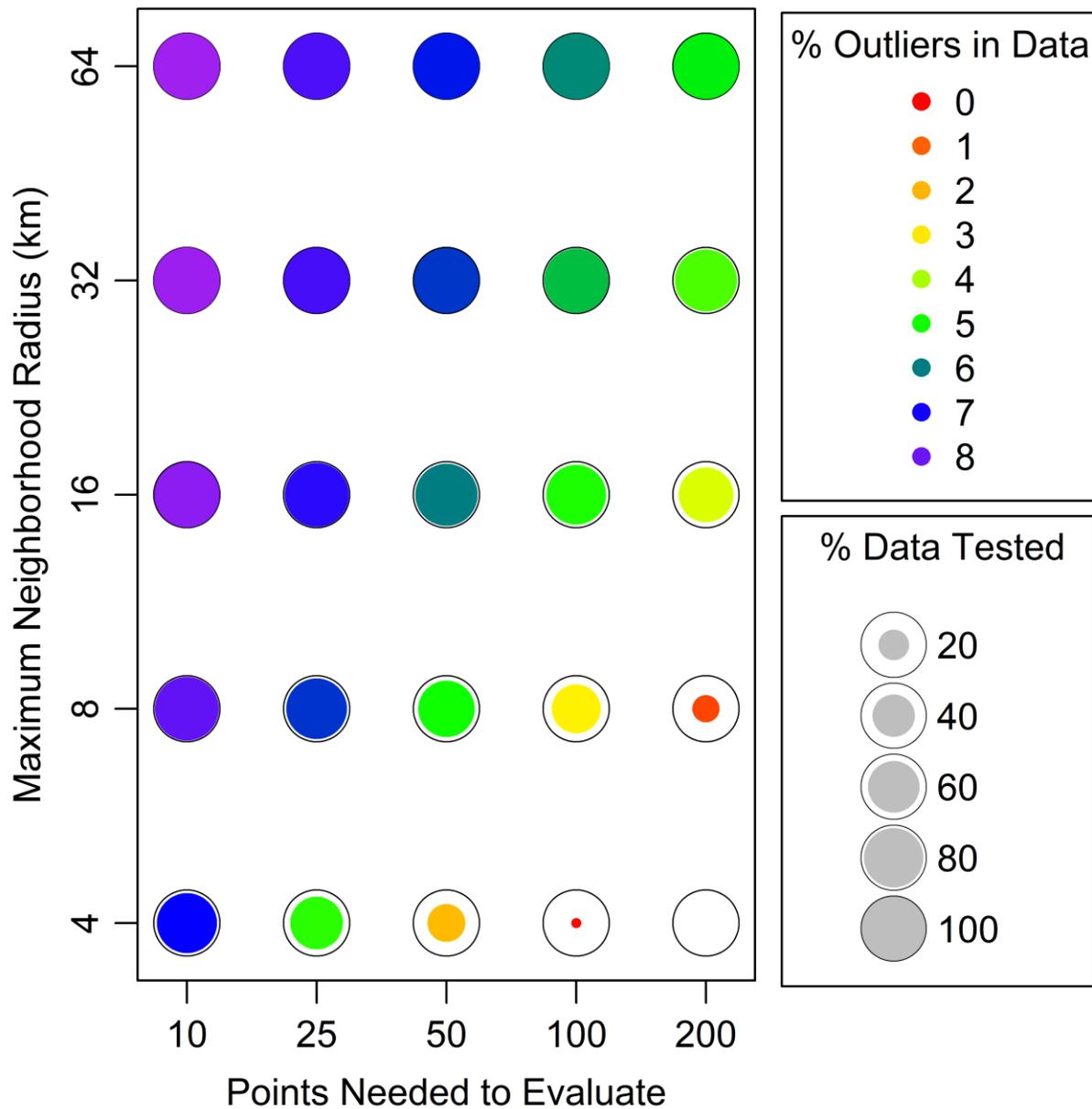
$$[B_{\text{low}}, B_{\text{up}}] = [q_{0.25} - k(q_{0.5} - q_{0.25}), q_{0.75} + k(q_{0.75} - q_{0.5})] \quad [2.1]$$

where  $B_{\text{low}}$  is the lower bound,  $B_{\text{up}}$  is the upper bound,  $q_p$  is the  $p^{\text{th}}$  quantile, and  $k$  is a parameter that determines the placement of the bounds. If the value of the surface heat flow for the point being tested was greater than  $B_{\text{up}}$  or less than  $B_{\text{low}}$ , the point was considered to be a local outlier and a potentially rogue observation. The value of  $k$  is typically selected according to type I error rates (identifying a point as an outlier when it is not an outlier), which we provide in Appendix A for several null hypothesis distributions. We selected  $k = 3$ , which is consistent with the numerical range that could be spanned by the whiskers on a standard boxplot ( $1.5 * [q_{0.75} - q_{0.25}]$ ). Larger values of  $k$  would result in fewer outliers using Equation 2.1 (we present results for several values of  $k$  in Appendix B section 7).

There are two parameters to define a local neighborhood with our algorithm: the maximum size, defined by the Euclidian radius, and the number of points needed to test for outliers. We performed a sensitivity analysis to inform the selection of these two parameters for our heat flow dataset. The point being tested is considered one of the neighborhood points because we desire that neighborhoods within a local area use similar outlier bounds (Equation 2.1). Once a point was identified as an outlier, it was still available as a neighbor for other points. Thus, the local neighborhoods may contain points that are deemed to be outliers relative to other points. The breakdown point for quartile estimation in Equation 2.1 is when 25% of the local

neighbors are rogue observations, so including some rogue observations as neighbors will not greatly limit the use of the equation.

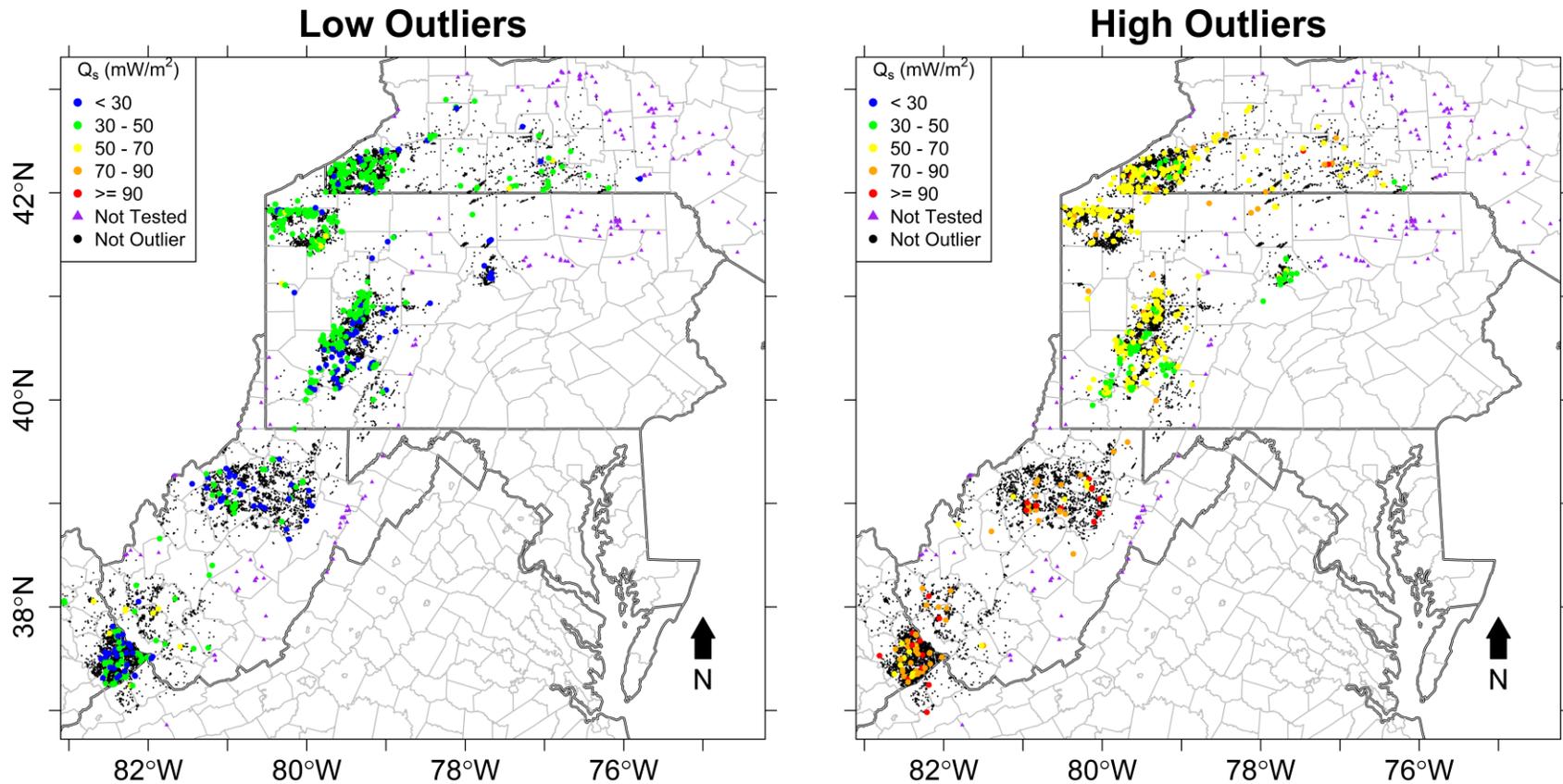
Figure 2.5 displays the results of the sensitivity analysis on our dataset. When a neighborhood is defined by a large number of points and a small maximum radius, a small fraction of the dataset can be evaluated. As the maximum neighborhood radius increases or the points criterion decreases, a greater fraction of the dataset can be tested. The local area and points criterion can be so large that the test converges to a global test; however, the tested cases in Figure 2.5 will never be global tests because of the points criterion restriction. In the upper left portion of Figure 2.5 where most of the dataset is tested, the fraction of the dataset identified as outliers is approximately twice what one would expect for normal data (6% - 8% versus 3% for normal, see Appendix A, Table A.1), and is closer to what one would expect from a fairly heavy-tailed kurtotic Student t distribution. This result seems reasonable based on the histogram of the surface heat flow provided in Figure 2.3B.



**Figure 2.5:** Sensitivity analysis of the local spatial outlier algorithm for the surface heat flow within a neighborhood defined by the maximum radius (vertical axis) to achieve the number of data points required to run the outlier test (horizontal axis). Colors represent the percentage of outliers identified, and point size represents the percentage of the dataset tested. The black circle represents 100% of the dataset being tested in the outlier algorithm.

Based on the results of the sensitivity analysis, we define a local neighborhood as the nearest 25 points within a maximum distance of 32 km. This resulted in 7.6% of the dataset (1015 records) being identified as spatial outliers. Only 1.4% of the dataset (191 records) was not tested because there were less than 25 points within 32 km of these points.

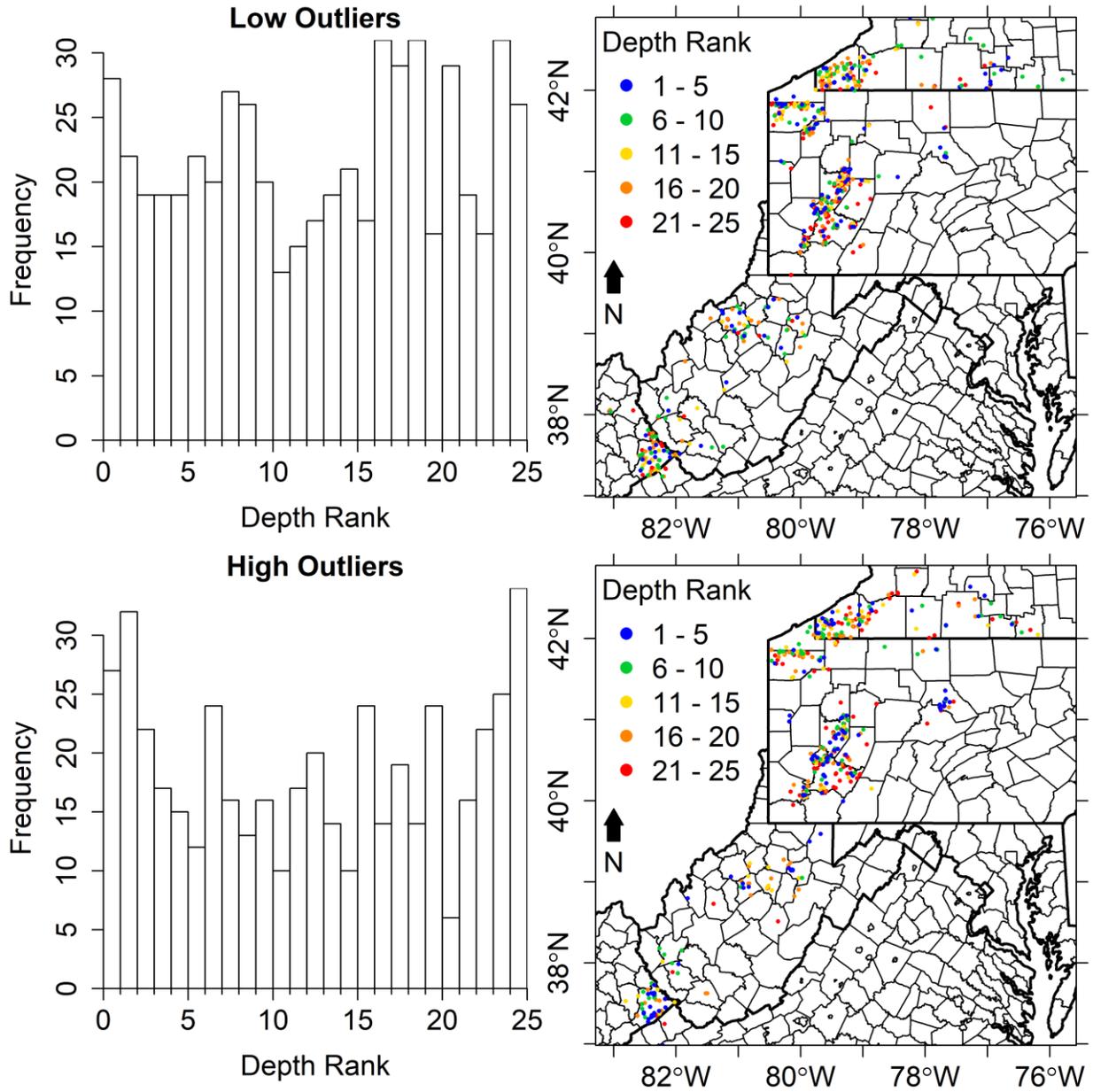
The spatial distribution of the local outliers is provided in Figure 2.6, colored by their surface heat flow value. There are 89 more low outliers than high outliers. Generally, the high outliers  $\geq 90 \text{ mW/m}^2$  are located in West Virginia, and high outliers between  $50 \text{ mW/m}^2$  and  $70 \text{ mW/m}^2$  are located in Pennsylvania and New York. Low outliers less than  $50 \text{ mW/m}^2$  are found throughout the basin.



**Figure 2.6:** Spatial distribution of surface heat flow ( $Q_s$ ) local outliers identified using the spatial outlier detection algorithm. Low outliers (left) and high outliers (right) share the same color scheme. Points that were not identified as outliers are shown as black circles. Points that were not tested for outliers as a result of too few neighborhood points are shown as purple triangles.

### 2.4.2.2 Spatial Outlier Analysis

In the outlier algorithm, we recorded the depth of BHT measurement of the outlier relative to the depth of BHT measurement for the points in its neighborhood. The measurement depths of the outlier and all neighborhood points were sorted and assigned ranks sequentially ranging from 1 (shallowest) to 25 (deepest). Ties in depth were assigned the average of the original ranks assigned to those depths. If throughout the basin outliers within a local neighborhood were not depth dependent, their ranks would follow a discrete uniform distribution. Histograms of the outlier depth ranks are provided in Figure 2.7, along with the spatial distribution of the ranks. We used a Kolmogorov–Smirnov test on the rank distributions of the low outliers and the high outliers. Both distributions are significantly different from discrete uniform at the 3% level; the high outliers are significantly different at the 1% level. Based on the histograms in Figure 2.7, the high outliers tend to be from surface heat flows that are calculated from either the shallowest or the deepest BHT within their local neighborhood. The low outliers do not have a clear trend with depth rank.

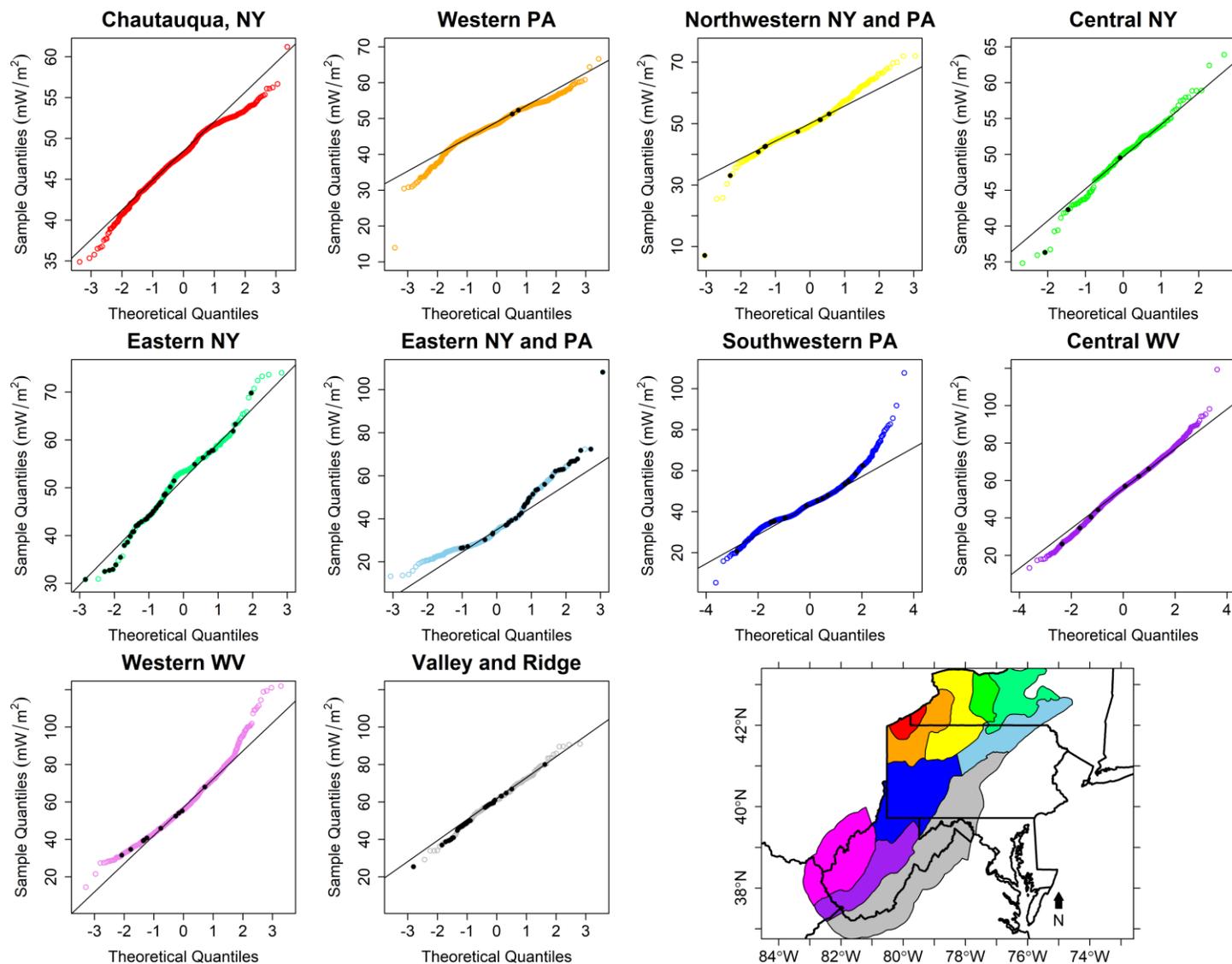


**Figure 2.7:** Histograms and maps of the depth ranks of low and high outliers. Lower ranks indicate that an outlier’s surface heat flow was calculated from a shallower BHT measurement within its local 25-point neighborhood.

### 2.4.2.3 Quantile-Quantile Plots to Evaluate Points in Data-Sparse Areas

Our local outlier identification algorithm was able to test most points on the interior of the Appalachian Basin spatial data domain, but about 200 points along the exterior of the spatial data domain and in other data-sparse areas were not tested (purple triangles, Figure 2.6). To evaluate these points, all surface heat flow values in our dataset were compared using normal quantile-quantile (q-q) plots within their respective geologic regions. Normality is not necessary for this ESDA procedure, and other distribution theoretical quantiles could be selected. Geologic regions were defined by strong lateral contrasts in gravity or magnetic pseudogravity resulting from Horowitz (2015) Poisson wavelet multiscale potential field analyses (Hornby et al., 1999; Moreau et al., 1997) in the Appalachian Basin (geologic region data are available on the Department of Energy Geothermal Data Repository, Cornell University, 2016).

Figure 2.8 provides the q-q plots. Generally, the untested points appear reasonable within their geologic regions. Two notable exceptions are a low point in the geologic region designated as “Northwestern NY and PA”, and a high point in the geologic region designated as “Eastern NY and PA” (Figure 2.8). These points were inconsistent with spatially nearby points, so they were considered spatial outliers. Used in this manner, the q-q plots provide a global spatial outlier detection method within each geologic region, and outlying points are examined among their local neighbors.



**Figure 2.8:** Normal q-q plots of the surface heat flow within each of the geologic regions shown on the map. Colors of points on the q-q plots match the colors of geologic regions. Observations that were not evaluated in the spatial outlier algorithm are shown as solid black points on q-q plots.

In addition to q-q plots, surface heat flow data near all untested points were manually scanned for spatial similarity with the untested point. One untested point in the geologic region designated as “Eastern NY” had a surface heat flow that was  $>20 \text{ mW/m}^2$  higher than nearby points, so we labeled it as a spatial outlier.

In summary, three points that were untested in the spatial outlier algorithm were labeled as spatial outliers, which has a minimal effect on the dataset compared to the 1015 points identified as outliers using the algorithm. All identified outliers are considered potentially rogue observations, and they should be evaluated in further detail to understand causes of their spatial discordancy. For a simple quality control, all outliers could be removed before a spatial regression, which would result in a dataset of 12,365 records. This would remove some valid observations (Appendix A Type I Error) in addition to rogue observations. In the following section, we evaluate the effect of removing these observations on estimation of spatial autocorrelation.

### **2.4.3 Geostatistical Evaluation of ESDA Procedures for Geothermal Resource Assessments**

The purpose of the ESDA procedures adopted in this paper was to identify rogue observations prior to using geostatistical spatial regressions for geothermal resource assessments. A critical requirement for spatial regressions is computing a measure of spatial autocorrelation. To assess the performance of our ESDA procedures, we compare estimated spatial autocorrelation across the datasets that result after each ESDA procedure is applied.

In expectation, if there is spatial autocorrelation, points closer to one another tend to be more similar in their feature-space values than points farther away. A metric that describes this relationship is the semi-variogram, which models the average semi-variance of feature-space

values as a function of the separation distance between two locations. The definition of the semi-variance and an unbiased method of moments sample estimator are provided in Equation 2.2 and Equation 2.3, respectively (e.g. Cressie, 1988; Isaaks and Srivastava, 1989)

$$\gamma(\|\mathbf{h}\|) = \frac{1}{2} \mathbb{E}([Q_s(\mathbf{x}) - Q_s(\mathbf{x} + \mathbf{h})]^2) \quad [2.2]$$

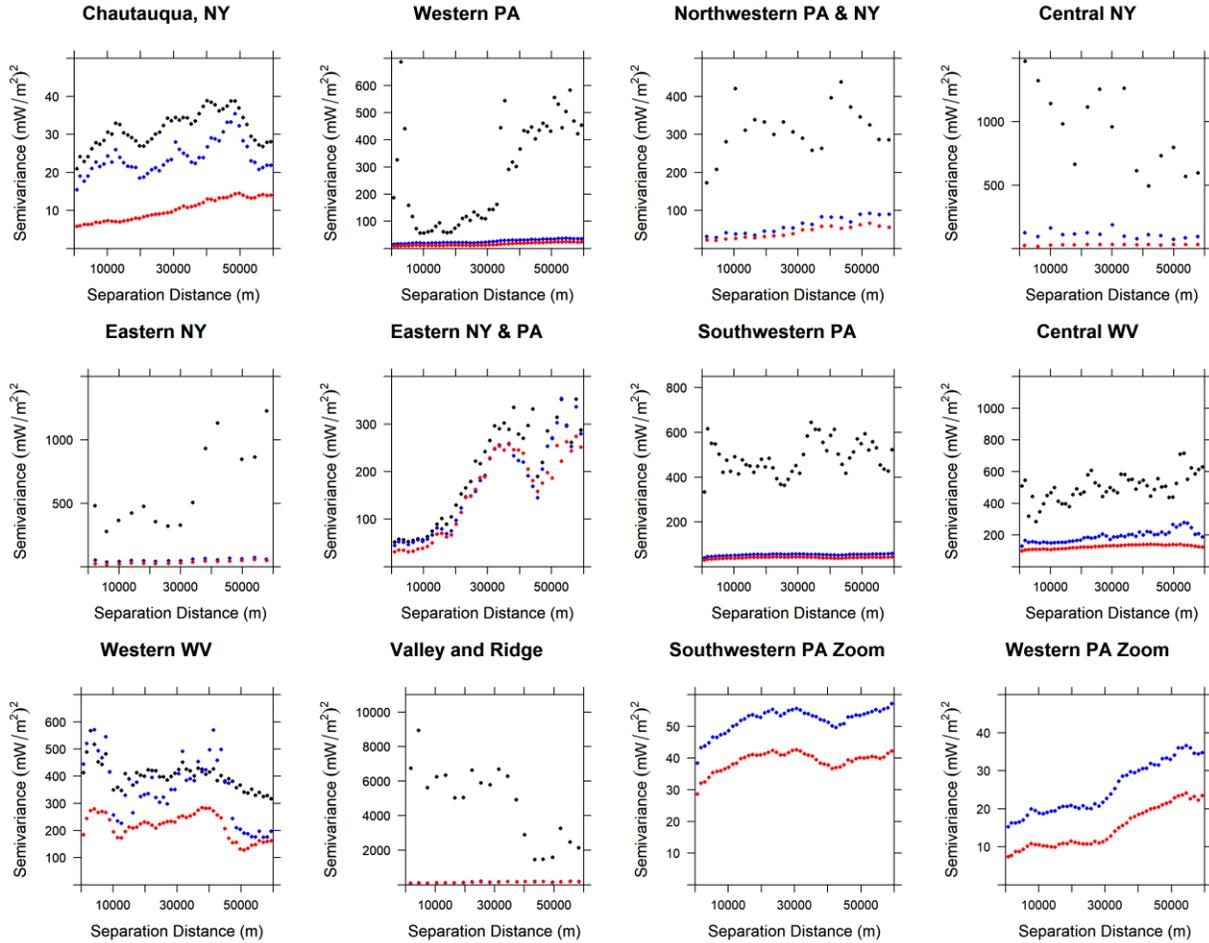
$$\hat{\gamma}(\|\mathbf{h}\|) = \frac{1}{2N(\|\mathbf{h}\|)} \sum_{i=1}^{N(\|\mathbf{h}\|)} [Q_s(\mathbf{x}_i) - Q_s(\mathbf{x}_i + \mathbf{h})]^2 \quad [2.3]$$

where  $\gamma$  is the population semi-variance,  $\hat{\gamma}$  estimates  $\gamma$ ,  $Q_s$  is the surface heat flow,  $\mathbf{x}$  (and  $\mathbf{x}_i$ ) are vectors of spatial position (easting, northing) of the surface heat flow values (for the  $i^{\text{th}}$  point),  $\mathbf{h}$  is a vector that describes the displacement from  $\mathbf{x}$ ,  $\|\cdot\|$  is the Euclidian norm,  $N(\|\mathbf{h}\|)$  is the number of points located at spatial lag distance  $\|\mathbf{h}\|$  from  $\mathbf{x}$ , and  $\mathbb{E}$  is the mathematical expectation. Equation 2.2 describes the expected difference between the values of  $Q_s$  at two locations as a function of the separation distance. The average semi-variance for each lag,  $\|\mathbf{h}\|$ , may be estimated from a spatial dataset using Equation 2.3; the result for all lags is called the sample semi-variogram. A tolerance window of half the lag separation distance is used to gather points that are close to  $\|\mathbf{h}\|$  in separation. The choice of number of lags to compute generally does not affect the resulting sample semi-variogram if the spatial process is well-behaved over the length scales evaluated by the lags, and the data are representative of the spatial process.

To evaluate the performance of our ESDA procedures, we used Equation 2.3 to compute sample semi-variograms for surface heat flow data within each of the geologic regions. Semi-variogram lag estimates are compared across the datasets that resulted after each of the major ESDA procedures: 1) post initial data processing (Section 2.3) and pre-ESDA; 2) post-application of a minimum depth threshold (Section 2.4.1) and pre-spatial outlier analysis

(Section 2.4.2); and 3) post-ESDA with spatial outliers removed from the dataset. The Equation 2.3 method of moments estimator is sensitive to spatial outliers (Cressie and Hawkins, 1980), so we expect a reduction in semi-variance to occur when spatial outliers are removed. The results of this analysis are provided in Figure 2.9.

For all geologic regions, ESDA procedures provided a reduction in semi-variance at shorter separation distances. Order of magnitude reductions in lag estimates of semi-variance resulted in several geologic regions (Figure 2.9; a table is provided in Appendix B section 6). The semi-variogram trend also stabilized for several geologic regions as a result of applying the ESDA procedures. For example, in the geologic regions designated as “Central WV” and “Southwestern PA,” applying ESDA procedures filtered essentially pure noise into signals with clear spatial autocorrelation. For most of the geologic regions, the application of a minimum depth threshold provided the most significant reduction in semi-variance, and removing spatial outliers further stabilized the semi-variogram trend.



**Figure 2.9:** Average semi-variance of the surface heat flow as a function of separation distance within the geologic regions provided in Figure 2.8. Point colors represent: 1) dataset pre-ESDA procedures (black), 2) dataset post-minimum depth threshold and pre-spatial outlier detection (blue), and 3) dataset post-ESDA with spatial outliers removed (red). Two plots in the bottom right corner zoom into two geologic regions to show detail for smaller semi-variances. For the Valley and Ridge region, blue points are covered by red points for the selected y-axis range; all of these points are less than  $200 \text{ (mW/m}^2\text{)}^2$ , and red points are lower than blue points by about  $12 \text{ (mW/m}^2\text{)}^2$  on average. For all geologic regions, Appendix B section 6 provides tables of the average semi-variance, and figures with jackknife approximate 95% confidence intervals using the methods presented in (Shafer and Varljén, 1990).

The post-ESDA sample semi-variograms are different in both magnitude and shape between geologic regions (confidence intervals in Appendix B section 6 further support this claim). Sample semi-variance estimates at short separation distances range from less than  $10 \text{ (mW/m}^2\text{)}^2$  to more than  $100 \text{ (mW/m}^2\text{)}^2$ . The separation distances to essentially uncorrelated data

range from 10 km to 50 km between geologic regions. Further research is needed to evaluate if these observed differences may be explained by physical causes, or if they are a result of data quality issues that were unaddressed by our ESDA procedures.

## **2.5 Discussion**

This paper illustrates the use of initial data processing and exploratory spatial data analysis (ESDA) procedures that identify potentially rogue data. In our application, rogue data may arise because of poor bottom-hole temperature (BHT) measurements, errors in the reported depth or location of the measurement, or from physical processes that are neglected or misrepresented in models used to arrive at surface heat flow estimates. We use a heat conduction model to compute surface heat flows using our BHT dataset. Other geothermal variables estimated from the heat conduction model (e.g., the temperature at 3 km depth) may also be checked using our ESDA procedures. If analyses include multiple geothermal variables that are derived from BHT measurements, note that these variables are related because they rely on the same BHT measurements and heat conduction model, and estimation of temperatures at depth requires the estimated surface heat flows. Thus, if an observation is rogue for one variable, it is more likely to be rogue for other variables.

Initial data processing was used to provide a surface heat flow dataset that contained plausible values with unique spatial coordinates. ESDA procedures applied to our 20,000-point dataset informed the implementation of a minimum BHT depth that resulted in a significant portion of the dataset being removed due to concerns for shallow thermal disturbances, like groundwater advection, that could cause the observed wide variability of calculated heat flow values for shallow BHT measurements (7435 observations removed, or about 37% of the

dataset). 7.6% of the remaining dataset was identified as locally discordant spatial outliers. We consider these shallow observations and spatial outliers to be potentially rogue observations. It would be desirable if all potentially rogue observations are truly rogue, but inevitably some observations are physically correct and unusually extreme. The time required to check the original sources of information and to correct or diagnose causes of discordancy would be intractable for a dataset of this size, with the origin in raster or paper archival records. The spatial information loss that results from removing points that are not rogue and may represent physical reality should be considered. Ideally, the spatial information lost from removing valid observations would be assessed relative to the information gained by removing rogue observations. Our geostatistical evaluation of spatial autocorrelation provides justification for removing potentially rogue observations from our dataset.

The use of a minimum depth threshold for BHTs caused a reduction of spatial data coverage. The removed points were primarily in exterior areas of the Appalachian Basin spatial data domain. If one were interested in reducing spatial extrapolation for geothermal resource assessments in these exterior areas, including shallower data could be beneficial. The consistency of shallower data in exterior areas may be evaluated using local median deviation ESDA procedures as applied in the shallow data area of Pennsylvania (Figure 2.4), using q-q plots (Figure 2.8), or with other methods. In our analysis without these shallower data, the local neighborhoods in exterior areas primarily included data from the interior of the Appalachian Basin spatial data domain. Thus, some information about spatial similarity was lost. We decided that at the scale of the Appalachian Basin, variability in surface heat flow estimates from shallow BHT measurements indicated a higher likelihood of rogue observations (Figure 2.3); allowing for a spatially variable minimum depth threshold may improve upon this analysis.

A critical ESDA procedure was the identification of local spatial outliers. Our asymmetric boxplot outlier criterion (Equation 2.1) depended on the parameter,  $k$ , that determined how extreme a heat flow value needed to be before it was considered a spatial outlier within its local neighborhood. The value of  $k$  for our study was selected based on Type I Error rates for the asymmetric boxplot (Appendix A), and we evaluated our results for several values of  $k$  (Appendix B section 7). We found that our semi-variance results (Figure 2.9) and associated conclusions were not very sensitive to the value of  $k$ . However, theoretical Type I Error rates increase as  $k$  decreases (Appendix A), and mean squared error tends to increase, especially if the true distribution is highly skewed. Future research could evaluate outlier criteria that are better suited for highly skewed distributions (e.g. Hubert and Vandervieren, 2008).

Within any outlier detection algorithm, there are several factors that may contribute to ineffective identification of rogue observations. Masking (Tietjen and Moore, 1972) of rogue observations may occur when a local neighborhood contains several points that are rogue, which “masks,” or prevents their detection in the neighborhood. Carter et al. (2009) found that the Carling (2000) based asymmetric boxplots, which are similar to our Equation 2.1, performed well to limit masking effects; however, these tests were susceptible to swamping. Swamping (Woolley, 2013) occurs where there are so many rogue observations within a local area that the good observations are identified as outliers. For our dataset, and most BHT datasets, swamping is of concern because shallow data outnumber deep data. If a good deep observation has as neighbors many rogue shallow observations, then the outlier algorithm may identify the good observation as the outlier. Future research can address spatial neighborhoods that consider the depth dimension within the outlier detection algorithm.

In areas that had data identified as outliers, an evaluation of those data points in the context of their geologic regions could be informative to diagnose causes of discordancy, or to correct the data. Our nonparametric outlier analysis by depth of the outlier within its local neighborhood allowed for preliminary analysis of the spatial outliers within our dataset, but further evaluation is required. Potential geostatistical evaluations include graphical or quantitative analysis of variogram point-pair clouds (e.g. Haslett et al., 1991; O’Leary et al., 2016; Ploner, 1999), and the local Moran’s I and associated statistics (Anselin, 1999).

We evaluated the performance of our ESDA procedures within geologic regions defined by geophysical potential fields (Horowitz, 2015). Other methods to define geologic regions may be used, as appropriate for the geothermal area being studied. Geologic region domains that are defined by physical properties may be independent of the BHT spatial data domain, as in this study; geologic regions may also be defined in 3D space (e.g., to handle aquifers or dipping structures). If there are multiple geologic regions in a study area, ESDA procedures and results should be evaluated for each region separately, as in this paper.

Evaluating our dataset within geologic regions revealed that surface heat flow spatial autocorrelation is not stationary at the scale of the Appalachian Basin. Spatial regressions for geothermal resources in the Appalachian Basin that use this dataset should be designed to handle the observed spatial autocorrelation in each geologic region. Further research is needed to understand if the observed variations in spatial autocorrelation between geologic regions are a result of physical processes, or due to data quality issues that were not addressed by applying our ESDA procedures. For example, three geologic regions (Figure 2.8) overlap multiple BHT correction sections (Figure 2.1) (e.g. the Southwestern PA geologic region overlaps all three BHT sections). Spatial regressions could consider the BHT correction section as a categorical

independent variable, and compute measures of spatial autocorrelation for the residuals of the regression, as in universal kriging or regression kriging (e.g. Hengl et al., 2007; Journel and Rossi, 1989). For such regression models, rogue observations would impact the estimation of parameters, as in any least-squares approach. Removing rogue data should provide more reliable parameter estimates and spatial autocorrelation estimates (e.g. Araki et al., 2017).

## **2.6 Conclusions**

Bottom-hole temperature (BHT) measurements are often the most abundantly available thermal information to derive variables of interest for earth-energy characterizations and to build regression models for prediction of geothermal resources. The known low quality of BHT datasets results in a high likelihood of discordant or rogue observations in the original temperature dataset and in the corresponding derived datasets. This paper provided a summary of possible causes of rogue observations in geothermal datasets, and demonstrated exploratory spatial data analysis (ESDA) procedures to identify rogue observations.

The local median deviation ESDA procedure was useful to inform the selection of a minimum depth for BHTs to limit concerns about shallow physical processes, like groundwater advection. Our local spatial outlier detection algorithm and ESDA procedure may be used to identify locally rogue observations in geothermal datasets; nonparametric spatial outlier analysis may help to diagnose the causes for discordancy. The quantile-quantile plot ESDA procedure applied within geologic regions may be useful to evaluate points in data-sparse areas in which “local” neighbors are relatively far away.

Processing our Appalachian Basin dataset using our ESDA procedures resulted in a significant portion of the dataset being considered as 1) too shallow for heat conduction

modeling (37%) or 2) potentially rogue spatial outliers (7.6% of the remaining dataset). In our study, removing potentially rogue observations resulted in stabilization of estimated spatial autocorrelation and reduction in estimated semi-variance compared with the initial dataset. Evaluating spatial autocorrelation within distinct geologic regions of the basin revealed non-stationary autocorrelation on the scale of the Appalachian Basin, which should be considered in geothermal resource assessments. These results provide confidence that parameter estimation precision for geothermal resource assessments are likely to improve as a result of applying the ESDA procedures to datasets with rogue observations. Therefore, we recommend that ESDA procedures be applied to BHT datasets or to variables derived from such datasets before building models for geothermal resource assessments. Future work should address the information that would be lost by removing valid observations, relative to the information gained by removing rogue observations.

## **2.7 Acknowledgements**

The author thanks coauthors Calvin A. Wheaton, Jerry R. Stedinger, David G. Rossiter, and Teresa E. Jordan for their contributions to this work. The authors would like to acknowledge Andrea Aguirre and Jefferson Tester for their involvement in developing and presenting an earlier version of the spatial outlier identification algorithm presented in this paper (Aguirre, 2014; Aguirre et al., 2013). We also thank the members of the Geothermal Play Fairway Analysis – Appalachian Basin project.

This research was supported by the National Science Foundation [grant number DGE-0966045] and the Department of Energy [grant number DE-EE0006726].

## **2.8 Disclaimer**

The information, data, or work presented herein was funded in part by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## **2.9 Code and Data Availability, and Software Credits**

The code and data used to run the analyses in this paper are available in a GitHub repository (Smith, 2018a). This repository relies on code contained within several other repositories for data processing and spatial outlier analysis (Smith, 2018b; Whealton and Smith, 2015). The heat conduction model is available in a Bitbucket repository (Horowitz et al., 2015). The processed BHT dataset is available on the Department of Energy Geothermal Data Repository (Cornell University, 2016).

Some maps presented in this paper were created in ArcGIS 10.4.1 (Environmental Systems Resource Institute [ESRI], 2016). Other graphics and statistical analyses were carried out using R version 3.5.0 (R Core Team, 2018) and the packages dgof (Arnold and Emerson,

2011), gstat (Pebesma, 2004), Hmisc (Harrell Jr, 2018), rgdal (Bivand et al., 2016), sp (Pebesma and Bivand, 2005), GISTools (Brunsdon and Chen, 2014), and parallel (R Core Team, 2018).

## REFERENCES

- Aguirre, G.A., 2014. Geothermal resource assessment: A case study of spatial variability and uncertainty analysis for the states of New York and Pennsylvania. Cornell University.
- Aguirre, G.A., Stedinger, J.R., Tester, J.W., 2013. Geothermal Resource Assessment : a Case Study of Spatial Variability and Uncertainty Analysis for the State of New York and Pennsylvania, in: Thirty-Eighth Workshop on Geothermal Reservoir Engineering. Stanford, California, p. 12.
- Andersen, R., 2008. Modern Methods for Robust Regression. SAGE Publications, Inc., Thousand Oaks, California. <https://doi.org/10.4135/9781412985109>
- Anselin, L., 1999. Interactive techniques and exploratory spatial data analysis, in: Longley, P., Goodchild, M., Maguire, D., Rhind, D. (Eds.), Geographical Information Systems. Wiley, New York, NY, pp. 253–266.
- Araki, S., Shimadera, H., Yamamoto, K., Kondo, A., 2017. Effect of spatial outliers on the regression modelling of air pollutant concentrations: A case study in Japan. Atmos. Environ. 153, 83–93. <https://doi.org/10.1016/j.atmosenv.2016.12.057>
- Arnold, T.A., Emerson, J.W., 2011. Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions [WWW Document]. R J. URL [http://journal.r-project.org/archive/2011-2/RJournal\\_2011-2\\_Arnold+Emerson.pdf](http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Arnold+Emerson.pdf)
- Barnett, V., Lewis, T., 1994. Outliers in Statistical Data, 3rd ed. John Wiley and Sons, New York, NY.
- Beltrami, H., Matharoo, G.S., Smerdon, J.E., 2015. Impact of borehole depths on reconstructed estimates of ground surface temperature histories and energy storage. J. Geophys. Res. Earth Surf. 120, 763–778. <https://doi.org/10.1002/2014JF003382>
- Bivand, R., Keitt, T., Rowlingson, B., 2016. rgdal: Bindings for the Geospatial Data Abstraction Library.
- Blackwell, D., Richards, M., Stepp, P., 2010. Final Report Texas Geothermal Assessment for the I35 Corridor East, Texas State Energy Conservation Office Contract CM709.
- Blackwell, D.D., Beardsmore, G.R., Nishimori, R.K., McMullen Jr., R.J., 1999. High-resolution Temperature Logs in a Petroleum Setting: Examples and Applications, in: Förster, A., Merriam, D.F. (Eds.), Geothermics in Basin Analysis. Kluwer Academic / Plenum Publishers, pp. 1–34. <https://doi.org/10.1007/978-1-4615-4751-8>
- Brezinski, D.K., 2011. Maryland borehole temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/cc54f15894222c91e71e4530dc088fec> (accessed 6.22.16).
- Brunsdon, C., Chen, H., 2014. GISTools: Some further GIS capabilities for R.
- Bureau of Topographic & Geologic Survey, 2018. Exploration and Development Well Information Network (EDWIN).
- Carling, K., 2000. Resistant outlier rules and the non-Gaussian case. Comput. Stat. Data Anal. 33, 249–258. [https://doi.org/10.1016/S0167-9473\(99\)00057-2](https://doi.org/10.1016/S0167-9473(99)00057-2)
- Carter, N.J., Schwertman, N.C., Kiser, T.L., 2009. A comparison of two boxplot methods for detecting univariate outliers which adjust for sample size and asymmetry. Stat. Methodol. 6, 604–621. <https://doi.org/10.1016/j.stamet.2009.07.001>
- Chatterjee, S., Hadi, A.S., 1986. Influential Observations, High Leverage Points, and Outliers in Linear Regression. Stat. Sci. 1, 415–416. <https://doi.org/10.1214/ss/1177013630>

- Cornell University, 2017. Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin [WWW Document]. URL <https://gdr.openei.org/submissions/899>
- Cornell University, 2016. Appalachian Basin Play Fairway Analysis: Improvements in 2016 to Thermal Quality Analysis in Low-Temperature Geothermal Play Fairway Analysis (GPFA-AB) [data set] [WWW Document]. URL <https://gdr.openei.org/submissions/879>
- Cressie, N., 1988. Spatial prediction and ordinary kriging. *Math. Geol.* 20, 405–421. <https://doi.org/10.1007/BF00892986>
- Cressie, N., Hawkins, D.M., 1980. Robust estimation of the variogram: I. *Math. Geol.* 12, 115–125. <https://doi.org/10.1007/BF01035243>
- Curl, D., 2011. Kentucky borehole temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/168566464e3d5f8f3cde3b9fc004bd38> (accessed 6.22.16).
- Davies, P.L., 1993. Aspects of Robust Linear Regression. *Ann. Stat.* 21, 1843–1899. <https://doi.org/10.1214/aos/1176349401>
- Davis, R.W., 2012. Deriving geothermal parameters from bottom-hole temperatures in Wyoming. *Am. Assoc. Pet. Geol. Bull.* 96, 1579–1592. <https://doi.org/10.1306/11081110167>
- Deming, D., 1989. Application of bottom-hole temperature corrections in geothermal studies. *Geothermics* 18, 775–786. [https://doi.org/10.1016/0375-6505\(89\)90106-5](https://doi.org/10.1016/0375-6505(89)90106-5)
- Environmental Systems Resource Institute (ESRI), 2016. ArcMap 10.4.
- Friðleifsson, G.O., Blischke, A., Kristjánsson, B.R., Richter, B., Einarsson, G.M., Jónasson, H., Franzson, H., Sigurðsson, O., Danielsen, P.E., Jónsson, S.S., Thordarson, S., Þórhallsson, S., Harðardóttir, V., Egilson, Þ., 2005. Reykjanes Well Report RN-17 & RN-17ST.
- Frone, Z., Blackwell, D.D., 2010. Geothermal Map of the Northeastern United States and the West Virginia Thermal Anomaly. *GRC Trans.* 34, 339–344.
- Frone, Z.S., Blackwell, D.D., Richards, M.C., Hornbach, M.J., 2015. Heat flow and thermal modeling of the Appalachian Basin, West Virginia. *Geosphere* 11, 1279–1290. <https://doi.org/10.1130/GES01155.1>
- Garg, S.K., Combs, J., 2015. A reformulation of USGS volumetric “heat in place” resource estimation method. *Geothermics* 55, 150–158. <https://doi.org/10.1016/j.geothermics.2015.02.004>
- Gass, T.E., 1982. The Geothermal Heat Pump. *Geotherm. Resour. Coun. Bull.* 11, 3–8.
- Gray, A.D., Majorowicz, J., Unsworth, M., 2012. Investigation of the geothermal state of sedimentary basins using oil industry thermal data: case study from Northern Alberta exhibiting the need to systematically remove biased data. *J. Geophys. Eng.* 9, 534–548. <https://doi.org/10.1088/1742-2132/9/5/534>
- Harrell Jr, F.E., 2018. Hmisc: Harrell Miscellaneous [WWW Document]. R Packag. version 4.1-1. URL <https://cran.r-project.org/package=Hmisc>
- Haslett, J., Bradley, R., Craig, P., Unwin, A., Wills, G., 1991. Dynamic Graphics for Exploring Spatial Data with Application to Locating Global and Local Anomalies. *Am. Stat.* 45, 234. <https://doi.org/10.2307/2684298>
- Hendry, R., Hilfiker, K., Hodge, D., Morgan, P., Swanberg, C., Shannon, S.S., 1982. Geothermal Investigations in West Virginia. Los Alamos, NM.
- Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: From equations to case studies. *Comput. Geosci.* 33, 1301–1315.

- <https://doi.org/10.1016/j.cageo.2007.05.001>
- Hodge, D.S., De Rito, R., Hifiker, K., Morgan, P., Swanberg, C.A., 1981. Investigations of Low-Temperature Geothermal Potential in New York State. Los Alamos, NM.
- Hornby, P., Boschetti, F., Horowitz, F.G., 1999. Analysis of potential field data in the wavelet domain. *Geophys. J. Int.* 137, 175–196. <https://doi.org/10.1046/j.1365-246x.1999.00788.x>
- Horowitz, F.G., Smith, J.D., Whealton, C.A., 2015. One dimensional conductive geothermal Python code [WWW Document]. URL <https://bitbucket.org/geothermalcode/onedimensionalgeothermalheatconductionmodel.git> (accessed 8.1.17).
- Horowitz, F.G., 2015. Identifying Potentially Activatable Faults in GPFA-AB, in: Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. pp. 298–316.
- Hubert, M., Vandervieren, E., 2008. An adjusted boxplot for skewed distributions. *Comput. Stat. Data Anal.* 52, 5186–5201. <https://doi.org/10.1016/j.csda.2007.11.008>
- Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York, NY.
- Jaeger, J.C., 1965. Application of the Theory Of Heat Conduction to Geothermal Measurements, in: Lee, W.H.K. (Ed.), *Terrestrial Heat Flow*. American Geophysical Union, Washington, D.C., pp. 7–23. <https://doi.org/10.1029/GM008p0007>
- Journel, A.G., Rossi, M.E., 1989. When do we need a trend model in kriging? *Math. Geol.* 21, 715–739. <https://doi.org/10.1007/BF00893318>
- Kerry, R., Oliver, M.A., 2007. Determining the effect of asymmetric data on the variogram. I. Underlying asymmetry. *Comput. Geosci.* 33, 1212–1232. <https://doi.org/10.1016/j.cageo.2007.05.008>
- Laslett, G.M., McBrantney, A.B., 1990. Estimation and implications of instrumental drift, random measurement error and nugget variance of soil attributes—a case study for soil pH. *J. Soil Sci.* 41, 451–471. <https://doi.org/10.1111/j.1365-2389.1990.tb00079.x>
- Leftwich, T., 2011. Ohio borehole temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/ba2f0b9d21f71acfe10609f76e2699e6> (accessed 6.22.16).
- Lewis, T.J., Beck, A.E., 1977. Analysis of heat-flow data — detailed observations in many holes in a small area. *Tectonophysics* 41, 41–59. [https://doi.org/10.1016/0040-1951\(77\)90179-2](https://doi.org/10.1016/0040-1951(77)90179-2)
- Lu, C.-T., Chen, D., Kou, Y., 2003. Algorithms for spatial outlier detection, in: Third IEEE International Conference on Data Mining. IEEE Comput. Soc, pp. 597–600. <https://doi.org/10.1109/ICDM.2003.1250986>
- Majorowicz, J.A., Jessop, A.M., 1981. Regional heat flow patterns in the Western Canadian Sedimentary Basin. *Tectonophysics* 74, 209–238. [https://doi.org/10.1016/0040-1951\(81\)90191-8](https://doi.org/10.1016/0040-1951(81)90191-8)
- Marchant, B.P., Lark, R.M., 2007. Robust estimation of the variogram by residual maximum likelihood. *Geoderma* 140, 62–72. <https://doi.org/10.1016/j.geoderma.2007.03.005>
- Moreau, F., Gibert, D., Holschneider, M., Saracco, G., 1997. Wavelet analysis of potential fields. *Inverse Probl.* 13, 165–178. <https://doi.org/10.1088/0266-5611/13/1/013>
- New York State Museum Hydrocarbon Reservoir Characterization Group, 2018. Empire State

- Organized Geologic Information System.
- O’Leary, B., Reiners, J.J., Xu, X., Lemke, L.D., 2016. Identification and influence of spatio-temporal outliers in urban air quality measurements. *Sci. Total Environ.* 573, 55–65. <https://doi.org/10.1016/j.scitotenv.2016.08.031>
- Patchen, D.G., Avary, K.L., Erwin, R.B., 1985a. Northern Appalachian Region correlation chart. AAPG Correl. Stratigr. Units North Am.
- Patchen, D.G., Avary, K.L., Erwin, R.B., 1985b. Southern Appalachian Basin correlation chart. AAPG Correl. Stratigr. Units North Am.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30, 683–691. <https://doi.org/10.1016/j.cageo.2004.03.012>
- Pebesma, E.J., Bivand, R.S., 2005. Classes and methods for spatial data in R. *R News* 5.
- Ploner, A., 1999. The use of the variogram cloud in geostatistical modelling. *Environmetrics* 10, 413–437. [https://doi.org/10.1002/\(SICI\)1099-095X\(199907/08\)10:4<413::AID-ENV365>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1099-095X(199907/08)10:4<413::AID-ENV365>3.0.CO;2-U)
- R Core Team, 2018. R: A language and environment for statistical computing.
- Rao, G.V., Rao, R.U.M., 1980. A geothermal study of the Jharia Gondwana basin (India): Heat flow results from several holes and heat production of basement rocks. *Earth Planet. Sci. Lett.* 48, 397–405. [https://doi.org/10.1016/0012-821X\(80\)90204-6](https://doi.org/10.1016/0012-821X(80)90204-6)
- Rousseeuw, P.J., Croux, C., 1993. Alternatives to the Median Absolute Deviation. *J. Am. Stat. Assoc.* 88, 1273. <https://doi.org/10.2307/2291267>
- Schubert, E., Zimek, A., Kriegel, H.-P., 2014. Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection. *Data Min. Knowl. Discov.* 28, 190–237. <https://doi.org/10.1007/s10618-012-0300-z>
- Shafer, J.M., Varljen, M.D., 1990. Approximation of confidence limits on sample semivariograms from single realizations of spatially correlated random fields. *Water Resour. Res.* 26, 1787–1802. <https://doi.org/10.1029/WR026i008p01787>
- Shank, S., Kaiser, R., Sullivan, M., Deemer, M., Stuckert, K., Macklin, E., Suskie, T., 2012. Pennsylvania Borehole Temperatures [WWW Document]. URL <http://repository.stategeothermaldata.org/repository/resource/9e15e1a59b768b330d029e86dc0d6512> (accessed 6.22.16).
- Slater, B., 2012. New York Borehole Temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/9e15e1a59b768b330d029e86dc00481f/> (accessed 6.22.16).
- Smith, J.D., 2018a. Geothermal\_ESDA [WWW Document]. URL [https://github.com/jds485/Geothermal\\_ESDA/new/master](https://github.com/jds485/Geothermal_ESDA/new/master)
- Smith, J.D., 2018b. Geothermal\_DataAnalysis\_CrossSections [WWW Document]. URL [https://github.com/jds485/Geothermal\\_DataAnalysis\\_CrossSections](https://github.com/jds485/Geothermal_DataAnalysis_CrossSections)
- Smith, J.D., 2016. Analytical and Geostatistical Heat Flow Modeling for Geothermal Resource Reconnaissance applied in the Appalachian Basin. Cornell University.
- Smith, J.D., Horowitz, F.G., 2017. Thermal Model Methods and Well Database Organization in GPFA-AB, in: Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. pp. 202–234.
- Southern Methodist University (SMU) Geothermal Lab, 2016. Near ground surface temperature map [WWW Document]. Data Maps / Temp. Maps. URL <http://www.smu.edu/Dedman/Academics/Programs/GeothermalLab/DataMaps/Temperat>

- ureMaps (accessed 6.22.16).
- Stutz, G.R., Shope, E., Aguirre, G.A., Batir, J., Frone, Z., Williams, M., Reber, T.J., Whealton, C.A., Smith, J.D., Richards, M.C., Blackwell, D.D., Tester, J.W., Stedinger, J.R., Jordan, T.E., 2015. Geothermal energy characterization in the Appalachian Basin of New York and Pennsylvania. *Geosphere* 11. <https://doi.org/10.1130/GES00499.1>
- Tester, J.W., Anderson, B.J., Batchelor, A.S., Blackwell, D.D., DiPippo, R., Drake, E.M., Garnish, J., Livesay, B., Moore, M.C., Nichols, K., Petty, S., Toksoz, M.N., Veatch, R.W.J., Baria, R., Augustine, C., Murphy, E., Negraru, P., Richards, M.C., 2006. The Future of Geothermal Energy: Impact of enhanced geothermal systems (EGS) on the United States in the 21st century, *The Future of Geothermal Energy*. Idaho Falls, Idaho.
- Tietjen, G.L., Moore, R.H., 1972. Some Grubbs-Type Statistics for the Detection of Several Outliers. *Technometrics* 14, 583–597. <https://doi.org/10.2307/1267287>
- Virginia Division of Geology and Mineral Resources (VDGMR), 2011. Virginia borehole temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/b99f8f8e3a7d798d77d4c343bd16b987> (accessed 6.22.16).
- West Virginia Geological & Economic Survey (WVGES), 2018. Pipeline Plus.
- West Virginia Geological & Economic Survey (WVGES), 2011. West Virginia Borehole Temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/168566464e3d5f8f3cde3b9fc0052329> (accessed 6.22.16).
- West Virginia Geological & Economic Survey (WVGES), 2006. Appalachian Basin Precambrian Basement Contours [WWW Document]. Trent. Black River Proj. PCMB\_Contours.zip. URL <http://www.wvgs.wvnet.edu/www/tbr/resources.asp> (accessed 6.22.16).
- Whealton, C.A., Smith, J.D., 2015. geothermal\_pfa [WWW Document]. URL [https://github.com/calvinwhealton/geothermal\\_pfa](https://github.com/calvinwhealton/geothermal_pfa)
- Whealton, C.A., Stedinger, J.R., 2015. Thermal Outlier Assessment in GPFA-AB, in: Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. pp. 180–187.
- Whealton, C.A., Stedinger, J.R., Horowitz, F.G., 2015. Application of Generalized Least Squares Regression in Bottom-Hole Temperature Corrections, in: Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. pp. 130–144.
- Woolley, T.W., 2013. An Investigation of the Effect of the Swamping Phenomenon on Several Block Procedures for Multiple Outliers in Univariate Samples. *Open J. Stat.* 03, 229–304. <https://doi.org/10.4236/ojs.2013.35035>

## CHAPTER 3

### STOCHASTIC UNCERTAINTY ANALYSIS AND SENSITIVITY ANALYSIS OF GEOHERMAL RESOURCE ASSESSMENTS FOR DIRECT-USE HEATING

#### **3.0 Abstract**

A stochastic framework is presented for geothermal resource assessments. It computes probability distributions to describe the precision of estimated 1) temperature-depth profiles at sites in a region of interest, and 2) stored thermal energy within potential geothermal reservoirs. Those probability distributions support an uncertainty analysis that identifies favorable areas where the stored thermal energy is defined relatively precisely by the available temperature and geologic data. The distributions of stored thermal energy are estimated using the estimated distributions of temperature-depth profiles, which reflect the spatial correlation employed in a laterally stratified kriging analysis of the surface heat flow derived from available well temperature datasets. Differences in surface heat flow spatial correlation structure across a region where a project may be proposed were important for comparing stored thermal energy site favorability. Sensitivity analysis identifies which temperature and geologic variable uncertainties contribute most to the resource uncertainty for different locations: this can inform prioritization of data collection efforts near potential project sites.

We apply our methods to estimate the stored thermal energy within potential sedimentary geothermal reservoirs in the Appalachian Basin. The intended application is geothermal direct-use heating. For two cities near more favorable reservoirs, we transform the estimated stored thermal energy distributions into distributions of useful life of a hypothetical geothermal heating system in each city. With our dataset, the variables that contributed most to stored thermal energy uncertainty were a function of the geothermal reservoir depth, and among reservoirs

changed spatially as a result of differing geologic conditions. The sensitivity analysis suggests that many geologic variables may be assigned regional values with associated large uncertainties with little impact on the resulting overall uncertainty in temperature-depth profiles and stored thermal energy calculations.

### **3.1 Introduction**

Geothermal resource assessments provide estimates of the surface heat flow, temperatures as a function of depth, and the thermal energy stored within rock volumes that are envisioned as geothermal reservoirs (e.g. Lee et al., 2010; Tester et al., 2006; Zafar and Cutright, 2014). Often, the temperatures of interest for utilization of a geothermal resource are located at greater depths than available thermal and geological data, or a project is proposed for a location at which no local data exist at relevant depths. Locations with insufficient data due to lack of prior geothermal well field exploration are referred to as geothermal greenfields (Hadi et al., 2010). This paper develops stochastic methods for uncertainty and sensitivity analyses for resource assessment of geothermal greenfields with conduction-dominated heat transfer. These methods are intended to be applied during the exploration phase in such geothermal regions, to inform decision making under unavoidable uncertainty.

The common practice in resource assessment of greenfields proceeds as follows. In the exploration phase, heat transfer models are employed to extrapolate the available thermal information to depths of interest while accounting for the expected trend of increasing temperature with increasing depth (e.g. transformation from a well bottom-hole temperature measurement to a surface heat flow in Smith [2019, Ch. 2] and Stutz et al. [2015]). With the derived geothermal variable at point locations on a common surface, spatial interpolation is used

to estimate values of the geothermal variable across a region of interest (e.g. Williams and DeAngelo, 2011; Smith, 2016). Then, heat transfer models are again employed to compute temperature-depth profiles throughout the region's rock volume (e.g. Lee et al., 2010). Extrapolations can be over several kilometers depth, and the locations of interest for spatial predictions can be tens of kilometers away from the nearest data points. Nevertheless, for stochastic simulations used to estimate stored thermal energy, uncertainty in geothermal reservoir rock temperature-depth profiles is commonly assigned without a clear relation to these underlying model uncertainties (e.g. Palmer-Wilson et al., 2018; Limberger et al., 2018; Banks and Harris, 2018; Garg and Combs, 2015; Siégel et al., 2014).

Moreover, spatial correlation among available thermal data and the prediction uncertainty in spatial interpolations are commonly ignored in stored thermal energy calculations. For example, several frequently employed geothermal resource assessment methods, including spatial interpolation by inverse distance weighting or contouring (e.g. Banks and Harris, 2018; Siégel et al., 2014) and the use of simple average values of measured temperatures within a region (e.g. Palmer-Wilson et al., 2018; Banks and Harris, 2018), ignore any spatial correlation among available temperature observations in datasets. Such approaches also neglect the geostatistical “nugget effect” (e.g. Cressie, 1988), which reflects noise in a dataset at very small spatial scales that may result from model errors, measurement and data collection errors, or very short-range spatial variability. Such effects are often found in geothermal datasets, as discussed in Smith (2019, Ch. 2). The nugget effect may be large and spatially variable for geothermal datasets that are derived from oil and gas well bottom-hole temperature measurements, and spatial correlation structures may be statistically significantly different across a region, as observed in the Appalachian Basin (Smith, 2019, Ch. 2; Smith, 2016). For a basin-scale

assessment of geothermal resources that informs potential geothermal reservoir favorability, the impact of these spatial effects on the uncertainty in stored thermal energy could be important and thus ought to be evaluated.

This paper presents modified stochastic simulation methods used for stored thermal energy calculations to also account for the uncertainty in spatial predictions and temperature-depth estimations that characterize the geothermal resource in a rock volume. The goals of this paper are thus 1) to employ stochastic methods of uncertainty propagation to the characterization of geothermal resources over spatial scales of hundreds of kilometers, and 2) to use sensitivity analysis to explore which geologic and thermal variable uncertainties contribute most to geothermal resource uncertainty in any locations of interest. In particular, this paper considers as measures of the geothermal resource the surface heat flow, temperature at 1.5 km depth, and stored thermal energy.

Our stochastic method uses Monte Carlo analyses to propagate resource characterization uncertainties that result from two data processing steps. The first step propagates uncertainty in the surface heat flow, which results from kriging geostatistical spatial interpolation prediction uncertainties, to uncertainty in predictions of temperatures at depth using a one-dimensional heat conduction model (Smith and Horowitz, 2017). The empirical distributions of estimated temperature-depth profiles reflect the spatial correlation of the heat flow map and associated prediction uncertainty. The second step propagates uncertainty in the estimated temperature-depth profiles to uncertainty in stored thermal energy, which is calculated using a volumetric heat-in-place model (e.g. Garg and Combs, 2015; Muffler and Cataldi, 1978). Uncertainty in geologic reservoir properties are considered in both models.

Uncertainty analysis is useful for understanding the precision of estimated geothermal resources and which areas of a geothermal region have more precise estimates, given the available data. We show that spatial heterogeneity in the magnitude of uncertainty is in part a result of differences in spatial correlation models among different geologic regions across our region of interest. Williams and DeAngelo (2011) note that uncertainties are also a function of the number of observations, the proximity of observations to the prediction location and to each other, and the variation in values among neighboring points (e.g. kriging covariance and distance matrices in Cressie, 1988). These specific effects are not evaluated explicitly in this paper.

For the heat conduction model and the heat-in-place (stored thermal energy) model, we provide global, variance-based sensitivity analyses (Pianosi et al., 2016). These analyses can inform how time and financial resources may be directed towards reducing uncertainty in those variables that contribute most to uncertainty in the modeled geothermal resource near a potential project site.

### **3.1.1 Case Study Site Description**

A case study applies our uncertainty analysis methods to the estimation of surface heat flow and temperatures at depth at well locations in the Appalachian Basin, temperature-depth profiles across the basin, and the stored thermal energy contained within potential sedimentary rock geothermal reservoirs. The Appalachian Basin covers an area of about 480,000 km<sup>2</sup> along the western margin of the Appalachian Mountains from northern Alabama, USA to southern Ontario, Canada (Ryder, 1995). The sedimentary rocks contain siliciclastics and carbonates with some evaporites (Roen and Walker, 1996). Given this focus on sedimentary rocks, our geothermal resource assessment is notably shallower than previous studies in the Appalachian

Basin that assessed enhanced geothermal systems (EGS) potential and electricity generation potential at depths greater than 4 km (Blackwell et al., 2010; Blackwell et al., 2011). With EGS, rock is stimulated to obtain suitable flow rates in rocks whose natural (unstimulated) permeability is low. Creating an EGS can be a costly investment (e.g. Lowry et al., 2017) and generation of desirable flow rates is not guaranteed, which increases the technical risk of a proposed project. Developing a geothermal system in naturally permeable rocks offers an opportunity for relatively less risk.

Camp et al. (2018) evaluated the potential flow productivity for geothermal projects within sedimentary rock reservoirs, which were explored for hydrocarbons in the Appalachian Basin states of New York, Pennsylvania, and West Virginia. Camp et al. (2018) provided estimates of reservoir rock geologic properties, such as porosity and thickness, along with their uncertainties (the data are available on the Geothermal Data Repository [Cornell University, 2016b]). In this paper, we estimate the thermal energy contained within what Camp et al. (2018) designated as reservoirs, for which the uncertainty in geologic properties is relatively well-characterized compared to other sedimentary rocks within the basin and the basement rocks. The primary datasets that informed the Camp et al. (2018) dataset are biased toward hydrocarbon-bearing reservoirs, hence they likely omit other naturally suitable geothermal reservoirs in the Appalachian Basin. Given this focus, this paper does not fully characterize the geothermal resource opportunities in naturally permeable rocks throughout the Appalachian Basin.

Geothermal resources across the Appalachian Basin were most recently evaluated by Smith (2016) as part of a U.S. Department of Energy funded Geothermal Play Fairway Analysis of the Appalachian Basin (Cornell University, 2017). A brief summary of the data processing in that study is provided here. Well bottom-hole temperatures (BHTs) were gathered from state

databases (Brezinski, 2011; Curl, 2011; Leftwich, 2011; Shank et al., 2012; Slater, 2012; Virginia Division of Geology and Mineral Resources [VDGMR], 2011; West Virginia Geological & Economic Survey [WVGES], 2011) and corrected to approximate thermal equilibrium using a set of Appalachian Basin-specific correction equations (Wheaton, 2016; Wheaton et al., 2015; code is available in a Github repository [Wheaton and Smith, 2015]). These correction equations were developed using generalized least squares regression, for which temperature data that likely reflected thermal equilibrium conditions (e.g. drill stem tests and equilibrium temperature-depth logs [Deming, 1989]) were used to define spatial neighborhoods that served as statistical clusters in the regression. The corrected BHTs were used to compute the surface heat flow using an analytical solution to the one-dimensional vertical heat conduction model (Smith and Horowitz, 2017) described in Section 3.2 of this paper. Smith (2019, Ch. 2; 2016) used an exploratory spatial data analysis (ESDA) (e.g. Anselin, 1999) of the computed surface heat flows at well locations to define a minimum depth for BHT measurements so that they likely reflected conduction-dominated heat transfer. A local spatial outlier analysis further refined the dataset to a total of 12,155 wells by removing heat flow values that were locally spatially discordant. This dataset contains wells from New York, Pennsylvania, West Virginia, and a 50 km buffer into Kentucky and Virginia.

Using the estimated point surface heat flows at those well locations, Smith (2016) used a laterally-stratified ordinary kriging geostatistical interpolation (e.g. Cressie, 1988; Stein et al., 1988) to estimate the surface heat flow in the Appalachian Basin for a total of 138,400 1 km<sup>2</sup> spatial areas, corresponding to typical development areas for geothermal projects. The lateral stratification boundaries (Figure 8 of this paper) were defined by lateral contrasts (i.e. “edges” or “worms” in Hornby et al., 1999 and Moreau et al., 1997) in gravity and magnetic potential fields

(Horowitz, 2015). Stratified kriging allowed for modeling the observed non-stationarity in the surface heat flow spatial correlation at well locations (i.e. the variogram and nugget effect) across the Appalachian Basin. In this paper, we use the Smith (2016; data available on the Geothermal Data Repository [Cornell University, 2016a]) predicted mean surface heat flow and standard error maps as input thermal information with which to estimate temperatures at depth for each 1 km<sup>2</sup> grid cell in the basin, and, subsequently, to calculate stored thermal energy within potential geothermal reservoirs at those grid cells (~16,700 total cells). An analysis of variance within the Smith (2016) geologic regions is used in this present paper to examine the effect of spatial correlation structure on geothermal resource prediction uncertainties.

In the present study, we assess stored thermal energy and its uncertainty to provide insight into those reservoirs that have 1) relatively large stored thermal energy, coupled to 2) a relatively small uncertainty. Such reservoirs may be favorable targets for geothermal energy extraction projects. Following Camp et al. (2018), given the shallow depths of many reservoirs, we anticipate they could be used for geothermal direct-use heating applications. Common temperature demands for these applications range from > 80 °C for district heating of communities, to as low as 20 °C for greenhouse warming (e.g. Gudmundsson et al., 1985). The stochastic methods presented in this paper to propagate uncertainty allow for site-specific probabilistic interpretations of each reservoir's ability to meet certain targets for heating. To illustrate how the results from these methods may be used for project planning, we provide simple examples that transform the distribution of stored thermal energy into a distribution of reservoir useful lives. Useful life is estimated for favorable geothermal reservoirs near two specific population centers.

### **3.1.2 Organization**

Section 3.2 and Section 3.3 describe the two models used to estimate the surface heat flow, temperatures at depth, and stored thermal energy. Both sections describe the model variables, and the selection of their values and probability distributions that are used in a Monte Carlo uncertainty analysis and sensitivity analysis of each model. Section 3.4 develops the sensitivity analysis methods used in this study. Section 3.5 presents the results of the geothermal resource assessments and the analyses of sensitivities. We conclude with a discussion of the implications of our results for probabilistic modeling of geothermal resources, and discuss limitations of our analyses.

### **3.2 One-Dimensional Heat Conduction Model**

A one-dimensional vertical heat conduction model was used to compute temperature-depth profiles. Details of the model can be found in Smith and Horowitz (2017) (the code is available in Horowitz et al., 2015). This model assumes two layers of radiogenic heat generation: 1) a constant value in sedimentary rocks, and 2) an exponential decrease in radiogenic heat producing elements (potassium, uranium, thorium) with increasing depth in basement rocks (e.g. Lachenbruch, 1968). Using surface heat flow data as an input thermal variable, temperatures at depth may be computed using Equation 3.1 (e.g. Eq. 10 in Jaeger, 1965)

$$T_{Z_{calc}} = \begin{cases} T_s + \frac{Q_s Z_{calc}}{\bar{k}_{Z_{calc}-0}} - \frac{A_s Z_{calc}^2}{2\bar{k}_{Z_{calc}-0}}, & Z_{calc} \leq Z_s \\ T_{Z_s} + \frac{(Q_m - A_B B e^{-3})(Z_{calc} - Z_s)}{k_B} + \frac{A_B B^2 \left(1 - e^{-\left(\frac{Z_{calc} - Z_s}{B}\right)}\right)}{k_B}, & Z_s < Z_{calc} \leq Z_s + 3B \end{cases} \quad [3.1]$$

where  $T_{Z_{calc}}$  ( $^{\circ}\text{C}$ ) is the temperature at the depth of interest,  $Z_{calc}$  (m).  $T_s$  is the average annual ground surface temperature ( $^{\circ}\text{C}$ ),  $Q_s$  is the surface heat flow ( $\text{W}/\text{m}^2$ ),  $A_s$  is the heat generation rate within sedimentary rocks ( $\text{W}/\text{m}^3$ ),  $\bar{k}_{Z_{calc}-0}$  is the harmonic average thermal conductivity of rocks within the interval from the surface to  $Z_{calc}$  ( $\text{W}/\text{m}\cdot^{\circ}\text{C}$ ),  $Z_s$  is the total sedimentary rock thickness (i.e. depth to basement) (m),  $T_{Z_s}$  is the temperature at the top of the basement ( $^{\circ}\text{C}$ ),  $Q_m$  is the mantle heat flow ( $\text{W}/\text{m}^2$ ),  $A_B$  is the heat generation in rocks at the top of the basement ( $\text{W}/\text{m}^3$ ),  $B$  is the distance corresponding to a one-log decrease in radiogenic heat production with depth in basement rocks (m), and  $k_B$  is the basement rock thermal conductivity ( $\text{W}/\text{m}\cdot^{\circ}\text{C}$ ).

In Equation 3.1, there are two conditions for calculating temperatures at depth, corresponding to the two layers of the heat generation model. To calculate temperatures at depth in sedimentary reservoirs, the first condition in Equation 3.1 is used. The second condition in Equation 3.1 is used to calculate temperatures at depth in basement rocks.

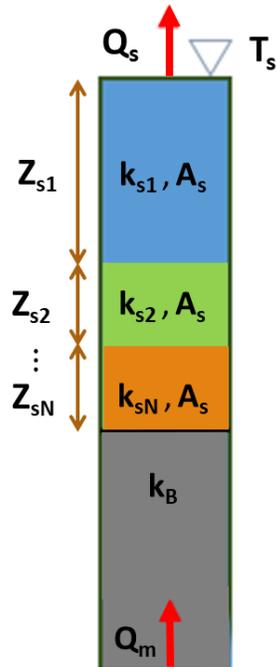
Equation 3.1 may be inverted to solve for the surface heat flow given a well bottom-hole temperature (BHT) measurement, which is how Smith (2016) estimated surface heat flow at well locations in the Appalachian Basin. Equation 3.2 presents the equations used to calculate the surface heat flow at a well location

$$Q_s = \begin{cases} G(Z_w) * \bar{k}_w + \frac{A_s Z_w}{2} & , \quad Z_w \leq Z_s \\ \frac{\left[ (BHT - T_s) + \frac{A_s Z_s^2}{2\bar{k}_s} + \frac{A_s Z_s (Z_w - Z_s)}{k_B} + \frac{(Q_m + A_s Z_s)}{k_B (1 - e^{-3})} \left[ B \left( 1 - e^{-\frac{Z_w - Z_s}{B}} \right) - (Z_w - Z_s) \right] \right]}{\frac{Z_w}{\bar{k}_w} + \frac{B \left( 1 - e^{-\frac{Z_w - Z_s}{B}} \right) - (Z_w - Z_s)}{k_B (1 - e^{-3})}} & , \quad Z_s < Z_w \leq Z_s + 3B \end{cases} \quad [3.2]$$

where  $G(Z_w)$  is the straight-line geothermal gradient computed from the surface to the well depth ( $Z_w$ ) ( $^{\circ}\text{C}/\text{m}$ ),  $\bar{k}_w$  is the harmonic average thermal conductivity of rocks within the interval from the surface to  $Z_w$  ( $\text{W}/\text{m}\cdot^{\circ}\text{C}$ ),  $\bar{k}_s$  is the harmonic average thermal conductivity of the entire sedimentary rock column, and other terms are as previously defined. Equation 3.2 is an analytical solution for the surface heat flow, given the assumptions of two-layer heat generation.

### 3.2.1 Heat Conduction Model Variables

There are seven variables whose values need to be resolved to use the heat conduction model (Equations 3.1 and 3.2). Figure 3.1 illustrates these variables within a geologic column with  $N$  sedimentary rock layers, followed by a basement rock layer. A BHT could be measured at any depth within such a column.



**Figure 3.1:** Schematic of the one-dimensional heat conduction model.  $Q_s$  – surface heat flow,  $T_s$  – surface temperature,  $A_s$  – sedimentary rock heat generation,  $Z_s$  – thickness of a sedimentary rock formation,  $k_s$  – thermal conductivity for a sedimentary rock formation,  $k_B$  – basement rock thermal conductivity,  $Q_m$  – mantle heat flow. Red arrows indicate the direction of heat flow.

Values and probability distributions assigned for Monte Carlo uncertainty analysis of temperature-depth profiles across the basin (138,400 grid cells), and for sensitivity analysis of the surface heat flow and temperatures at 1.5 km depth at 12,155 well locations are described in the following subsections. A summary of selected values and probability distributions follows in Table 3.1. All variables are assumed to have triangular distributions, which allow for a minimum, most likely and maximum value to be specified. Distributions are specified uniquely for each location in the basin. For geothermal greenfield studies, a range of likely values for geologic properties is often the best available information. Given the expected ranges, triangular distributions avoid the random generation of physically unreasonable or impossible values. The triangular distribution is commonly employed for stochastic heat-in-place evaluations (Palmer-

Wilson et al., 2018; Garg and Combs, 2015) and is standard practice for the United States Geological Survey (Williams et al., 2008). Several authors note the need to evaluate the effect of the epistemic uncertainty in the assumed shape of probability distributions (Garg and Combs, 2015; Williams, 2014), but that is beyond the scope of this paper.

### **3.2.1.1 Surface Heat Flow**

Maps of the mean and prediction standard error of surface heat flow from Smith (2016) are used in this study (data are available on the Geothermal Data Repository, Cornell University [2016a]). For each of the 138,400 1 km<sup>2</sup> grid cells on those maps, the bounds of the triangular distribution were specified as  $\pm 2$  standard errors from the mean at that cell. If the true error distribution were Normal, using a triangular distribution with bounds of  $\pm 2.45$  standard errors would be optimal to minimize the mean squared error of the distribution's moments (Scherer et al., 2003). Using  $\pm 2$  standard errors reduces tail variability relative to a Normal distribution.

### **3.2.1.2 Surface Temperature**

A map of the ground surface temperature for the United States is provided by Gass (1982; data are available from SMU Geothermal Lab, 2016). This map was estimated from temperature measurements in 15 m to 46 m deep groundwater wells. To estimate the error in the temperature measured at these depths that results from solar energy thermal propagation, Smith and Horowitz (2017) assumed a relatively high thermal diffusivity of rocks and found a range of about  $\pm 0.5$  °C for 15 m depth over a calendar year. We assume that map interpolation prediction uncertainties, which are not quantified in Gass (1982), contribute an additional  $\pm 0.5$  °C,

comparable to the thermal diffusivity range just mentioned. Thus, the total assumed error from physical measurement errors and statistical methods is  $\pm 1$  °C.

### **3.2.1.3 Sedimentary Rock Radiogenic Heat Generation**

The range of radiogenic heat generation in sedimentary rocks is small, with typical values of  $0.5 \mu\text{W}/\text{m}^3$  for non-siliciclastic rocks and  $2.0 \mu\text{W}/\text{m}^3$  for radiogenic siliciclastic rocks (Rybach, 1986; Waples, 2002). Following Smith (2016), heat generation is assigned the same constant value for all sedimentary rocks, regardless of lithology. We selected a triangular distribution with bounds of  $1 \pm 0.25 \mu\text{W}/\text{m}^3$ . The  $0.5 \mu\text{W}/\text{m}^3$  range is similar to the range of variability observed for individual formations in a recently acquired spectral gamma digital well log taken in West Virginia (WVGES, 2019; Appendix C Section 1). However, the assumed mean tends to be smaller on average compared to values observed in siliciclastic rocks, and larger on average for carbonate rocks. For improved resource assessment accuracy, future analyses can evaluate using heat generation values that adhere to trends observed in this log.

### **3.2.1.4 Sedimentary Rock Formation Thicknesses**

Formation tops and thicknesses for the Appalachian Basin were gathered from the Correlation of Stratigraphic Units of North America (COSUNA) generalized stratigraphic columns (Patchen et al., 1985a; 1985b). These columns are defined for broad spatial regions of the basin (Figure 3.3, this paper). Smith (2016) used these columns to estimate surface heat flow for the basin, so we employ these columns to match the stratigraphic assumptions of that analysis. The triangular distribution for each formation's thickness was assigned using the ranges and most likely values reported on the COSUNA columns, as interpreted and digitized in Cornell

University (2016a). Following Smith (2016), to estimate the depth at which a formation top occurs for a particular location within the COSUNA column region, the total column thickness was linearly scaled to match the depth of basement rocks in each location (WVGES, 2006).

### **3.2.1.5 Sedimentary Rock Thermal Conductivity**

Estimates of thermal conductivity for Appalachian Basin formations were made as part of a Geothermal Play Fairway Analysis project (the data are available on the Geothermal Data Repository, Cornell University [2016a]). Few thermal conductivity measurements are available for Appalachian Basin formations, so values were primarily from Carter et al. (1998) for rock samples in the Anadarko Basin, which has a similarly deep burial history as the Appalachian Basin (e.g. Rowan, 2006; Johnsson, 1986; Carter et al., 1998). Formation thermal conductivity values were estimated using a Monte Carlo analysis of the thermal conductivity values, which were assigned to each formation by lithology. In this study, we use the estimated means as the most likely values, and  $\pm 40\%$  of the means to define the bounds of the triangular distributions. For the smallest values of thermal conductivity, around  $1 \text{ W/m-}^\circ\text{C}$ , these uncertainty bounds adhere to physical expectations of lower bounds of thermal conductivity for rocks with water-filled pores (e.g. Robertson, 1988). As a result of this physical lower bound, measurement errors and errors resulting from using rock samples from a different sedimentary basin are expected to be larger for larger values of thermal conductivity. Using a percentage of the mean to define bounds of the triangular distribution reflects these expectations.

### **3.2.1.6 Basement Thermal Conductivity**

Few measurements of basement rock thermal conductivity are available for the Appalachian Basin. We based the mean and the range of thermal conductivity uncertainty on the distribution of basement rock thermal conductivity in the heat flow database for the United States (Blackwell et al., 2007). This dataset informed the selection of possible thermal conductivities for Appalachian Basin basement rocks in the Monte Carlo analysis described in Section 3.2.1.5. The basement rock thermal conductivity in this paper was modeled as a symmetric triangular distribution with bounds of the mean  $\pm$  two standard deviations, as obtained from the Monte Carlo analysis. We do not expect basement conductivity to be a critical variable in this study because the value will only affect calculations of temperatures at depth in basement rocks, and estimations of surface heat flow from wells drilled into basement rocks (Equation 3.1, Equation 3.2).

### **3.2.1.7 Mantle Heat Flow**

A mantle heat flow of  $30 \text{ mW/m}^2$  is assumed for the Appalachian Basin region of interest based on the ranges of values for stable continents and estimates near the Appalachian Basin (Roy, Blackwell, and Birch, 1968; Jaupart and Mareschal, 1999; Sclater, Jaupart, and Galson, 1980; Artemieva and Mooney, 2001). The value of the mantle heat flow provides a boundary condition in the heat conduction model that defines what value the heat flow achieves after the exponential decrease in heat producing elements in basement rocks has ceased. We do not expect this variable to be critical in this study for the same reasons as presented in Section 3.2.1.6.

### 3.2.1.8 Bottom-hole Temperature

For each of the 12,155 wells, BHT uncertainty was assigned as the greater of 5 °C or 10% of the BHT measurement. This was based on knowledge of the BHT data collection procedures, observed spatial variability in BHTs in areas less than 5 km in diameter (e.g. Smith, 2019, Ch. 2), and uncertainty in the BHT correction equations provided by Whealton (2016; Whealton et al., 2015; code available in Whealton and Smith, 2015). The assignment of greater uncertainty to larger BHTs is because the uncertainty in the BHT correction equation tends to increase with increasing depth (Whealton, 2016).

**Table 3.1:** Probability distributions for input variables in the uncertainty and sensitivity analyses of temperature-depth profiles (first seven variables) and surface heat flow (last seven variables) using the heat conduction model.

Variable	Triangular Distribution Specifications	Sources
Surface Heat Flow (mW/m <sup>2</sup> )	Mean ± 2 std. dev.	Smith (2016), Cornell University (2016a)
Surface Temperature (°C)	Mean ± 1	Gass (1982), Smith (2016)
Sedimentary Rock Heat Generation (μW/m <sup>3</sup> )	1 ± 0.25	Rybach (1986), Waples (2002)
Sedimentary Rock Thicknesses (m)	Minimum, Most Likely, Maximum	Patchen et al. (1985a; 1985b)
Sedimentary Rock Thermal Conductivities (W/m-°C)	Mean ± 40% *Mean	Cornell University (2016a)
Basement Rock Thermal Conductivity (W/m-°C)	2.83 ± 0.72	Blackwell et al. (2007)
Mantle Heat Flow (mW/m <sup>2</sup> )	30 ± 6	Roy, Blackwell, and Birch (1968); Jaupart and Mareschal (1999); Sclater, Jaupart, and Galson (1980); Artemieva and Mooney (2001)
Bottom-hole Temperature (BHT) (°C)	BHT ± max(5, 0.1*BHT)	Whealton et al. (2015), Whealton (2016)

### **3.2.2 Uncertainty Analysis Methods for Temperature-Depth Profiles across the Appalachian Basin**

Monte Carlo analysis of the heat conduction model provided estimates of mean temperatures and the uncertainty distribution as a function of depth across the Appalachian Basin. Computations were completed at all centers of the 138,400 1 km<sup>2</sup> grid cells of the Smith (2016) heat flow maps. Equation 3.1 was used for these calculations because heat flow was the input thermal variable in the heat conduction model.

The uncertainty analysis considered 10,000 Monte Carlo random samples of all variables in Table 3.1 except the BHT, which is reflected in the surface heat flow maps by Smith (2016). Therefore, this analysis considers the spatial correlation of the BHT-derived heat flows and the associated spatial prediction uncertainties.

The spatial correlation of the geologic variables in Table 3.1 is unknown because of the limited data for the Appalachian Basin, so we sampled values for all variables independently of the values sampled for neighboring grid cells. This approach is sufficient for our analysis because the intent is to summarize results at the grid cell scale. Variances and quantiles computed by aggregating results over larger areas would require assumptions about the spatial correlation of these geologic variables. The implication of this approach for geothermal resource exploration is that each 1 km<sup>2</sup> area would be developed independently, which is reasonable for initial evaluation of a geothermal greenfield (e.g. licensed areas for geothermal development in the Dogger aquifer, Paris, France [BRGM, 2015]).

Our uncertainty analysis provides empirical distributions of temperature-depth profiles for each grid cell of the Smith (2016) heat flow map. By using the spatially correlated heat flow values and prediction uncertainties as input to the heat conduction model, these temperature-

depth profiles reflect the spatial correlation of the surface heat flow map and associated prediction uncertainties, and the trends in surface heat flow across the basin. Similarly, because the ground surface temperature map is an input variable, the trend and assumed error in the ground surface temperature map across the basin is also reflected in the resulting temperature-depth profiles.

### **3.2.3 Sensitivity Analysis Methods for the Surface Heat Flow and Temperature-Depth Profiles at Well Locations**

The surface heat flow maps provided by Smith (2016) do not consider uncertainty that results from uncertain input variables to the calculation of the surface heat flow at the 12,155 well locations that were used to make the maps. These uncertainties should be considered in future geothermal resource assessments that map the surface heat flow. The intent of our sensitivity analysis is to determine which heat conduction model variables contribute most to the uncertainty in surface heat flow computations because the estimation of temperature-depth profiles across the basin should depend on the uncertainty in the surface heat flow calculations at well locations. The results of this sensitivity analysis have most value when placed into the context of how each well was weighted in the Smith (2016) kriging spatial regression to estimate surface heat flow. We present here results for individual wells, and leave to future work evaluation of uncertainty in values of individual wells aggregated into an updated regional spatial regression analysis.

A sensitivity analysis is provided for the surface heat flow at each well location used by Smith (2016). The BHT is the input thermal variable for this analysis. BHT values are computed exactly at their depth of measurement when using Equation 3.1 and 3.2 to calculate the surface

heat flow and subsequent temperature-depth profiles. Using the BHT-derived surface heat flows, we also completed a sensitivity analysis for the estimated temperature at 1.5 km depth for each well location. This is approximately the median depth of the potential geothermal reservoirs in the Camp et al. (2018) reservoir dataset. We compare the results of the sensitivity analysis for the surface heat flow and temperature at 1.5 km depth for each well location.

These sensitivity analyses used 1000 random and independent samples of all variables in Table 3.1 except the surface heat flow, whose variability results from variability in the BHT. Whealton (2016) provided a similar sensitivity analysis of this heat conduction model for a set of relatively deep wells near or within basement rocks in the Appalachian Basin; our analysis includes wells from a wider variety of well depths, and provides a spatial analysis of the sensitivity results. The sensitivity analysis methods and equations are provided in Section 3.4, and the results are provided in Section 3.5.2.

### **3.3 Heat-in-Place Model**

Stored thermal energy is computed for reservoirs in the dataset provided by Camp et al. (2018; the data are available on the Geothermal Data Repository [Cornell University, 2016b]). This dataset was compiled from regional studies, including the Midwest Regional Carbon Sequestration Partnership (MRCSP, 2015) and the New York state geological database (ESOGIS, 2015). In addition to these reservoirs, as part of an ongoing U.S. Department of Energy-funded deep direct-use geothermal feasibility study at West Virginia University, one additional reservoir near Morgantown, WV was added to the earlier Camp et al. (2018) dataset. There are currently no wells in Morgantown for this potential reservoir, so data from wells near Morgantown were used to inform the selection of geologic properties (Appendix C Section 2).

We only consider reservoirs deeper than 1 km in this study because shallower BHT data were generally not used in the estimation by Smith (2016). Additionally, temperatures of reservoirs at shallower depths tended to be cooler than would be required by typical direct-use heat utilization processes. Reservoirs whose uncertainty in the depth to the formation top could result in a reservoir shallower than 1 km were not evaluated. We estimated stored thermal energy only for areas of reservoirs that had surface heat flow predictions in the map by Smith (2016). We filtered the reservoir dataset to contain reservoirs with at least 4 km<sup>2</sup> area, defined by intersecting 2 raster grid cell centers on the heat flow map by Smith (2016). These reservoirs tended to have more than one oil or gas wells whose locations support the estimated spatial extent. These processing steps resulted in a database of 510 potential geothermal reservoirs. Stored thermal energy was computed for each 1 km<sup>2</sup> spatial area of each reservoir over the corresponding reservoir thickness, for a total of 16,700 grid cells evaluated.

### **3.3.1 Heat-in-Place Model Equations**

The volumetric heat-in-place method (e.g. Garg and Combs, 2015; Muffler and Cataldi, 1978) is used to estimate the stored thermal energy contained within a volume of rock and its water-filled pore space. The thermal energy contained within the  $i^{\text{th}}$  1 km<sup>2</sup> surface area ( $A_i$ , in m) of a reservoir over its thickness is provided in Equation 3.3 (e.g. Tester et al., 2012, Ch. 11)

$$E_i = A_i(\phi\rho_w C_{p_w} + [1 - \phi]\rho_r C_{p_r}) \int_{z_t}^{z_b} (T(z_t) + G[Z - z_t] - T_{ref}) dZ$$

$$E_i = A_i(\phi\rho_w C_{p_w} + [1 - \phi]\rho_r C_{p_r}) \left[ G \left( \frac{z_b^2}{2} + \frac{z_t^2}{2} - z_b z_t \right) + (T(z_t) - T_{ref})(z_b - z_t) \right]$$

[3.3]

where  $E_i$  is the stored thermal energy (J),  $\phi$  is the porosity of the rock,  $\rho_w$  is the density of the water ( $\text{kg/m}^3$ ),  $\rho_r$  is the density of the rock ( $\text{kg/m}^3$ ),  $C_{p_w}$  is the heat capacity of the water ( $\text{J/kg-}^\circ\text{C}$ ),  $C_{p_r}$  is the heat capacity of the rock ( $\text{J/kg-}^\circ\text{C}$ ),  $Z$  is the depth below the ground surface (m),  $z_t$  is the depth to the top of the reservoir (m),  $z_b$  is the depth to the bottom of the reservoir (m),  $T(z_t)$  is the temperature at the top of the reservoir ( $^\circ\text{C}$ ),  $G$  is the geothermal gradient in the reservoir ( $^\circ\text{C/m}$ ), and  $T_{ref}$  is the reference temperature ( $^\circ\text{C}$ ) that is determined by the engineering design. The effects of heat generation (Equation 3.1) could be added to this integration, but those effects are negligible over the thicknesses of reservoirs considered in this paper (thinnest 5<sup>th</sup> percentile: 1.5 m, median thickness: 18 m, thickest 5<sup>th</sup> percentile: 47.2 m).

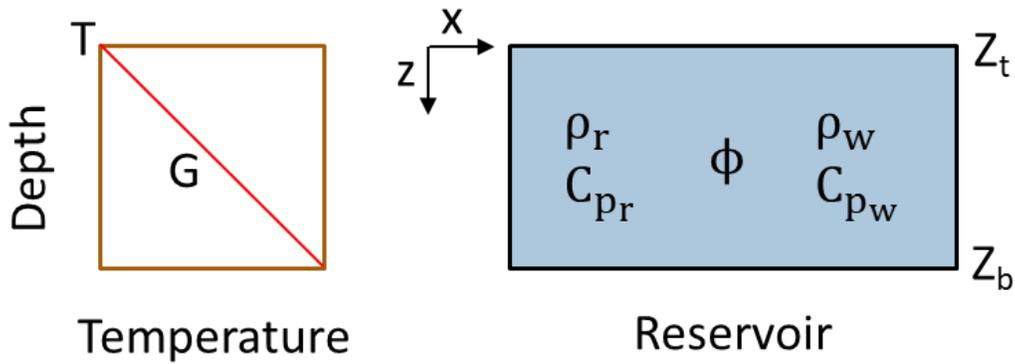
In a practical geothermal energy project, the useful thermal energy extracted from a discrete rock volume will be smaller than the value computed using Equation 3.3 because heat cannot be uniformly extracted, and because of the utilization efficiency of surface equipment. Recovery factors are commonly included in Equation 3.3 to assess the resource potential according to the expected flow geometry of the reservoir (e.g. Garg and Combs, 2015; Williams et al., 2008), and utilization efficiency factors are used to account for surface processes (e.g. Garg and Combs, 2015). These factors are both site-specific (e.g. recovery factors for the Trenton-Black River sedimentary reservoir in Camp and Jordan [2017]), and related to actual production from the geothermal reservoir. Therefore, it is beyond the scope of this paper to

consider them. Additionally, because we do not evaluate production from the reservoirs, we also do not evaluate heat transfer from caprocks or base rocks, as in Smith et al. (2016), who showed that these effects may be important for reservoirs less than about 10 m thick, depending on the properties of the rock, flow rates, and the length of time the reservoir is in operation.

The primary engineering-controlled parameter in Equation 3.3 is the reference temperature, also known as the reinjection temperature for a geothermal system. In practice, the reference temperature is site-specific as a function of the surface utilization processes (e.g. Takahashi and Yoshida, 2016; Garg and Combs, 2015). In this paper, we aim to provide an analysis that allows for screening of more favorable reservoirs independent of the utilization process, so we choose the same reference temperature for all reservoirs. We selected a reference temperature of 20 °C because our reservoirs all have estimated temperatures greater than this value, as presented in Section 3.5.3 (Figure 3.8).

### **3.3.2 Heat-in-Place Model Variables**

From Equation 3.3, we defined six variables whose values need to be specified for heat-in-place calculations: rock temperature-depth profile in the reservoir, reservoir depth ( $z_t$ ), reservoir thickness ( $z_b - z_t$ ), rock porosity, rock density, and rock heat capacity. Using the specified rock temperature, the density and heat capacity of water are computed as described in Section 3.3.2.2. Figure 3.2 provides a schematic of a reservoir that illustrates these variables.



**Figure 3.2:** Schematic of the heat-in-place model used to calculate stored thermal energy for a reservoir volume consisting of a 1 km<sup>2</sup> grid cell area and thickness  $Z_b - Z_t$ .  $T$  – temperature at the top of the reservoir,  $G$  – geothermal gradient,  $Z_t$  – depth to top of the reservoir,  $Z_b$  – depth to bottom of the reservoir,  $\phi$  – rock porosity,  $\rho_r$  – rock density,  $C_{p_r}$  – rock heat capacity,  $\rho_w$  – water density,  $C_{p_w}$  – water heat capacity.

The following subsections describe the selection of values and probability distributions for the temperature-depth profiles and reservoir geologic variables. These are used in uncertainty analysis and sensitivity analysis of stored thermal energy. A summary of the selected probability distributions is provided in Table 3.4 at the end of this section.

### 3.3.2.1 Reservoir Temperature-Depth Profiles

The temperatures at depth within each reservoir were specified exactly by the empirical distribution of temperature-depth profiles that resulted from the Monte Carlo uncertainty analysis of the heat conduction model for each 1 km<sup>2</sup> grid cell in the Appalachian Basin. In that analysis, temperatures were reported in 0.5 km depth increments. To compute temperatures in each reservoir, the temperature at the top and bottom of the reservoir were assigned by assuming a linear gradient between each 0.5 km depth increment. Using the temperature at the reservoir top and bottom, the geothermal gradient was computed for the reservoir thickness. A linear gradient

is used for the reservoir because we expect the average thermal conductivity of the reservoir to be the same over its thickness, and heat generation is negligible over the thicknesses of reservoirs in this study.

### **3.3.2.2 Water Thermodynamic Properties**

The fluid filling the pore spaces of reservoir rock is likely brine. Observations of oilfield brine composition, produced waters, and disposal well data in the Appalachian Basin reveal high variability in the salt concentration of brines across the basin (Dresel, 1985; Lynch and Castor, 1983; Matsumoto et al., 1996; Siegel et al., 1990; Skeen, 2010; Waller et al., 1978). Appreciable variability in salt concentration is likely from site to site. Lacking data for each reservoir in this study, for a fair comparison across reservoirs, we assume pure water instead of brine. The density of brine is higher than the density of pure water and the heat capacity is smaller (e.g. Ramalingam and Arumugam, 2012). These effects tend to not cancel out for the temperatures evaluated in this study, so our use of pure water will likely underestimate the total stored thermal energy.

The density and heat capacity of pure water were determined using the temperature correlations provided in the GEOPHIRES software, which was developed for geothermal reservoir simulations and economic analyses (Beckers and McCabe, 2018). These density and heat capacity relationships are provided in Equation 3.4 and Equation 3.5, respectively

$$\rho_w(T_w) = 0.7983223 + (1.50896E^{-3} - 2.9104E^{-6} * T_w) * T_w * 1000 \left[ \frac{\text{kg}}{\text{m}^3} \right] \quad [3.4]$$

$$C_{p_w}(T_w) = \frac{-203.606 + 1523.29T_w - 3196.413T_w^2 + 2474.455T_w^3 + \frac{3.855326}{T_w^2}}{18.02} * 1000 \left[ \frac{\text{J}}{\text{kg} - \text{K}} \right] \quad [3.5]$$

where  $\rho_w(T_w)$  is the density of water as a function of the water temperature ( $T_w$  in K), and  $C_{p_w}(T_w)$  is the heat capacity of water as a function of the water temperature ( $T_w$  in mK). We assume that the water in the pore spaces is initially in thermal equilibrium with the rock, such that the initial rock temperature determines the water temperature. For simplicity, we use Equations 3.4 and 3.5 to compute the heat capacity and density at the center of each reservoir's vertical thickness in each simulation.

### 3.3.2.3 Reservoir Depth, Thickness, and Porosity

Camp et al. (2018) provide average values for porosity, reservoir thickness, and depth, and uncertainty levels for the former two variables. Average values of each variable were gathered from well logs for wells within each reservoir (Camp et al., 2018). The uncertainty levels adopted from Camp et al. (2018) are summarized in Table 3.2. The reservoir thickness uncertainty level was based on the standard deviation of the net pay thickness for oil and gas reservoirs. Using the hydrocarbon pay thickness provides a conservatively low thickness estimate for geothermal reservoirs (i.e. a reservoir can be permeable, and yet not contain hydrocarbons). For porosity, we assumed that the uncertainty level was the same as provided in the Camp et al. (2018) database for permeability. The highest uncertainty level for permeability

was 200% of the mean value. Using 200% uncertainty in a triangular distribution could generate negative values, so we used 100% of the mean value as the highest porosity uncertainty level.

**Table 3.2:** Uncertainty levels and probability distributions used for reservoir thickness and porosity. The bounds of the symmetric triangular distribution are provided as a percentage of the mean value.

Uncertainty Level	Thickness Uncertainty (Bounds as a % of the Mean)	Porosity Uncertainty (Bounds as a % of the Mean)
Distribution	Symmetric Triangular	Symmetric Triangular
1	20	12.5
2	40	25
3	60	50
4	80	100
5	100	100

Uncertainty in the depth to the top of the reservoir is not provided in Camp et al. (2018). The depths of reservoirs were based on well log interpretations, so we assigned uncertainty based on the authors’ observations of variability in the depths to formation tops that were selected by different geologists within the New York Empire State Organized Geologic Information System (ESOGIS) database (ESOGIS, 2015). The depths for formation tops tended to vary by about +/- 20 m, depending on the formation and the available well log data. For depth uncertainty, we assigned a normal distribution with a mean as listed in the database and a standard deviation of 10 m, so that two standard deviations corresponds closely to the observed variability.

#### 3.3.2.4 Reservoir Rock Heat Capacity and Density

The Camp et al. (2018) database specifies a lithology for each reservoir. Values of density and heat capacity in this paper were selected by lithology, as summarized in Table 3.3. Many reservoirs in West Virginia had the lithology specified as “Unknown,” to which we assigned the range of likely density and heat capacity values for sedimentary rocks.

Density ranges and most likely values were gathered from Manger (1963), who provides tables by lithology. Many of the sampled formations listed in Manger (1963) are in the Appalachian Basin. The ranges of density and the most likely values were selected based on these tables, rounded to the nearest 50 kg/m<sup>3</sup>.

Waples and Waples (2004) provide a normalized heat capacity relationship as a function of temperature for a wide variety of rock types. These authors show that the effect of pressure on heat capacity of rocks from temperatures ranging from 0 – 1200 °C is negligible, so it is ignored in this paper. The equation for normalized rock heat capacity and the correction from an initial temperature (e.g. lab measured) to a final temperature (e.g. at the depth of a reservoir) are provided in Equation 3.6 and Equation 3.7, respectively

$$C_{pn}(T) = (8.95E^{-10})T^3 - (2.13E^{-6})T^2 + 0.00172T + 0.716 \quad [3.6]$$

$$C_{pr}(T_2) = C_{pr}(T_1) * \frac{C_{pn}(T_2)}{C_{pn}(T_1)} \left[ \frac{\text{J}}{\text{kg} - ^\circ\text{C}} \right] \quad [3.7]$$

where  $C_{pn}(T)$  is the normalized rock heat capacity as a function of temperature ( $T$  in °C),  $T_1$  is the initial temperature,  $T_2$  is the final temperature,  $C_{pr}(T_1)$  is the heat capacity of rock at temperature  $T_1$ , and  $C_{pr}(T_2)$  is the heat capacity of rock at temperature  $T_2$ . We use this relationship to compute the heat capacity corresponding to the temperature at the center depth of each of our reservoirs.

**Table 3.3:** Density and heat capacity for rocks by lithology.

<b>Lithology</b>	<b>Density (kg/m<sup>3</sup>)</b> Triangular Minimum, Mode, Maximum	<b>Heat Capacity at 20 °C (J/kg-°C)</b> Symmetric Triangular Mean, Bounds as % of Mean	<b>Sources</b>
Limestone	2600, 2700, 2750	780, 12%	Density: Range of limestone values in Manger (1963) Heat Capacity: Range of 2 measurements in Waples and Waples (2004)
Dolomite	2700, 2850, 2900	870, 5%	Density: Range of dolomite values in Manger (1963) Heat Capacity: Robertson (1988), Waples and Waples (2004)
Sandstone	2300, 2650, 2700	790, 7%	Density: Range of sandstone values in Manger (1963) Heat Capacity: Range of values in Abdulagatov et al. (2015)
Mudstone	2400, 2600, 2800	810, 10%	Density: Range of shale/siltstone values in Manger (1963) Heat Capacity: Value in Robertson (1988) at 0 °C, corrected to 20 °C using the Waples and Waples (2004) correction.
Chert	2500, 2600, 2700	810, 15%	Density: Chert formations in Manger (1963) Heat Capacity: quartzite in Waples and Waples (2004)
Unknown	2300, 2500, 2900	780, 20%	Range of values observed for other rocks.

### 3.3.2.5 Summary of Probability Distributions for Heat-In-Place Model Variables

A summary of the probability distributions for each of the variables used in the stored thermal energy calculations is provided in Table 3.4. For the temperature-depth profiles, probability distributions are specified uniquely for each 1 km<sup>2</sup> area of the reservoir. For all other variables, probability distributions have the same shape for all spatial areas of a specific reservoir.

**Table 3.4:** Probability distributions selected for variables in the uncertainty analysis and the sensitivity analysis of the stored thermal energy for each 1 km<sup>2</sup> area of each reservoir.

<b>Variable</b>	<b>Distribution Specifications</b>	<b>Sources</b>
Reservoir Temperature-Depth Profile (°C)	Empirical distribution from uncertainty analysis using the heat conduction model.	This paper
Reservoir Formation Top Depth (m)	Normal Mean, Standard Deviation = 10 m	Camp et al. (2018)
Reservoir Thickness (m)	Symmetric Triangular Bounds: Mean ± Uncertainty Level	Camp et al. (2018); Table 3.2, this paper
Reservoir Porosity (-)	Symmetric Triangular Bounds: Mean ± Uncertainty Level	Camp et al. (2018); Table 3.2, this paper
Rock Density (kg/m <sup>3</sup> )	Triangular Min, Mode, Max	Table 3.3, this paper
Rock Heat Capacity (J/kg-°C)	Symmetric Triangular Bounds: Mean ± Uncertainty	Table 3.3, this paper

### 3.3.3 Uncertainty and Sensitivity Analysis Methods for the Heat-In-Place Model

Stored thermal energy was estimated using a Monte Carlo uncertainty analysis for each of the 16,700 1 km<sup>2</sup> areas corresponding to potential geothermal reservoirs. The uncertainty analysis used the empirical distribution of 10,000 temperature-depth profiles, each one coupled to a single random sample of the other five variables in Table 3.4. For reasons stated in Section 3.2.2, we sampled values for all variables independently of the values sampled for neighboring grid cells. The result of our uncertainty analysis is empirical distributions of stored thermal energy for each grid cell of each reservoir in the basin. The stored thermal energy distributions are spatially correlated due to the propagation of a spatially correlated surface heat flow map and associated prediction uncertainties through the heat conduction model and the heat-in-place model.

Sensitivity analysis of stored thermal energy was completed for each 1 km<sup>2</sup> area of each reservoir. As with the uncertainty analysis, sensitivity analysis used the 10,000 temperature-

depth profiles, each one coupled to a random sample of the other five variables in Table 3.4. The sensitivity analysis methods and equations are provided in Section 3.4, and the results are provided in Section 3.5.5.

### 3.4 Sensitivity Analysis Methods

We use sensitivity analyses to explore how much of the uncertainty in model output variables can be explained by the uncertainties in the various input variables. Sensitivity analyses also provide insight into which input variables may be fixed at constant regional average values due to negligible impacts on the uncertainty in model output variables. Sensitivity analysis may be local (often derivative-based), referring to the sensitivity at a fixed point in the domain of the input variables, or it may be global (generally variance-based), referring to sensitivity of a specified output variable to variations of an input variable over its entire uncertainty range. A review of sensitivity analysis methods is provided in Pianosi et al. (2016). A global Monte Carlo approach is appropriate for our study rather than a local approach because we are interested in the impact of the entire range of possible values of input variables, and their interactions, on the output variables.

A simple form of global variance-based sensitivity analysis computed the total order effect of each input variable on the output variance (e.g. Pianosi et al., 2016). The total order effect is computed using Equation 3.8

$$\hat{S}_T(X_i) = 1 - \frac{\widehat{V}_Y[Y|\mathbf{X}_{\sim i}]}{\widehat{V}_Y[Y|\mathbf{X}]}$$
[3.8]

where  $\hat{S}_T(X_i)$  is the estimated total order effect for variable  $X_i$ ,  $\widehat{V}_Y[Y|\mathbf{X}]$  is the estimated variance of output variable Y that results when all of the input variables are allowed to vary, and

$\widehat{V}_Y[Y|\mathbf{X}_{\sim i}]$  is the estimated variance that is computed when the  $X_i$  variable is fixed and the rest are allowed to vary. The total order effect measures the proportion of variance attributable to variable  $X_i$  and that variables' interactions with other variables. Thus, in expectation, the total order effect represents the fraction of the total variance in  $Y$  that could be eliminated if the true value of the variable  $X_i$  were known (e.g. Saltelli et al., 2008).

As described in Section 3.2.3 and 3.3.3, Monte Carlo random samples from the probability distribution of each input variable were used to estimate  $\widehat{V}_Y[Y|\mathbf{X}]$  for each output variable. To compute  $\widehat{V}_Y[Y|\mathbf{X}_{\sim i}]$ , the same random sample as was used to compute  $\widehat{V}_Y[Y|\mathbf{X}]$  was used for every variable, except the  $i^{\text{th}}$  variable, which was generally fixed at the mode of its probability distribution. The only variable specified differently was the empirically-derived distribution of temperature-depth profiles, which was fixed at the mean value.

### 3.4.1 Sensitivity Analysis Implementation

Total order effects were computed for both the surface heat flow and the temperature at 1.5 km depth (output variables from the heat conduction model) at all 12,155 well locations. They were also computed for the stored thermal energy (output variable from the heat-in-place model) for each of the 16,700 1 km<sup>2</sup> grid cells of reservoirs. Sensitivity analyses considered all possible input variables that contributed uncertainty at a well location or a grid cell.

A challenge with the sensitivity analysis for the surface heat flow and temperature at 1.5 km depth at well locations is the number of possible input variables that might be considered. The majority of input variables result from formation-specific values that are related to the same physical issue. For example, for a well location with 45 different sedimentary rock formations (layers), there are 45 thermal conductivity variables. To reduce the computational complexity,

the question we considered in this sensitivity analysis was if all the thermal conductivities at a well location were collectively important. An analogous approach was used for sedimentary formation thicknesses. Thus, for each well location we computed total order effects for the five individual input variables listed in Table 3.1, and two groups of variables: sedimentary rock thermal conductivities and formation thicknesses. Saltelli and Tarantola (2002) suggest creating groups of variables to allow evaluation of their total impact on model output variables when input variables are considered together.

Using variable groups changes the interpretation of terms in Equation 3.8; each  $X_i$  becomes a variable group instead of a single input variable. For variable groups, interactions in the model are interpreted as “closed” on the group (e.g. Saltelli and Tarantola, 2002), meaning that when we fix all of the values of variables in a group, we remove each variable’s first-order contribution to the total variance in the output variable and all of the possible higher-order contributions that result from interactions between those variables.

For sensitivity analysis of the surface heat flow and temperature at 1.5 km depth at well locations, we consider seven input variables listed in Table 3.1. For each well location, individual input variables of interest are: 1) the bottom-hole temperature (BHT), 2) the surface temperature at the well location, 3) the heat generation in all sedimentary rocks, 4) the mantle heat flow, and 5) the basement rock thermal conductivity. The final two variables are groups for each well: 6) sedimentary rock thermal conductivities (group size equal to the number of formations), and 7) sedimentary rock formation thicknesses (group size equal to the number of formations). Grouping the sedimentary formation thicknesses and conductivities reduced the number of input variables from some 95 variables (45 formation thicknesses + 45 formation conductivities + 5 other variables) to 7 input variables and variable groups for each well location.

For sensitivity analysis of the stored thermal energy computed for reservoir grid cells over the simulated thickness of the reservoir, each input variable in Table 3.4 was considered individually. For each grid cell, the six individual variables of interest are: 1) the reservoir temperature-depth profile, 2) the reservoir top depth, 3) the reservoir thickness, 4) the reservoir porosity, 5) the reservoir rock density, and 6) the reservoir rock heat capacity. Other variables that contribute uncertainty to the temperature of the reservoir, like rock thermal conductivity, are not considered explicitly in this analysis because they have already been considered in the temperature-depth profile uncertainty analysis.

Our global sensitivity analysis is based on a Monte Carlo analysis of the three system output variables' values, where each was computed with randomly drawn values of input variables. All of the input variables are sampled independently and randomly according to their specified distributions in Table 3.1 and Table 3.4. For the variables and variable groups considered in our two models, only density and porosity at a grid cell would likely be strongly correlated with each other within a reservoir, and could thus require the use of a joint distribution. Few measurements are available to derive density-porosity correlations for the reservoirs considered in this analysis. Additionally, the results of our stored thermal energy sensitivity analysis considering density and porosity as independent variables found that both variables had a relatively minor impact on uncertainty in the stored thermal energy. So, we did not consider correlated sampling of density and porosity.

### **3.5 Results**

The following discussion considers the results for each of the uncertainty and sensitivity analyses for the Appalachian Basin geothermal resource assessment considered in this paper.

Section 3.5.1 presents maps of the estimated quantiles of temperatures at the basement depth across the basin. Section 3.5.2 provides results of the sensitivity analysis for the surface heat flow and temperatures at 1.5 km depth for all well locations. Section 3.5.3 and 3.5.4 provide results for the temperatures at the depth of potential geothermal reservoirs at each grid cell, and the stored thermal energy within the reservoirs. Section 3.5.5 provides sensitivity analysis results for the stored thermal energy calculations.

### **3.5.1 Distributions of Temperatures at the Basement Depth in the Appalachian Basin**

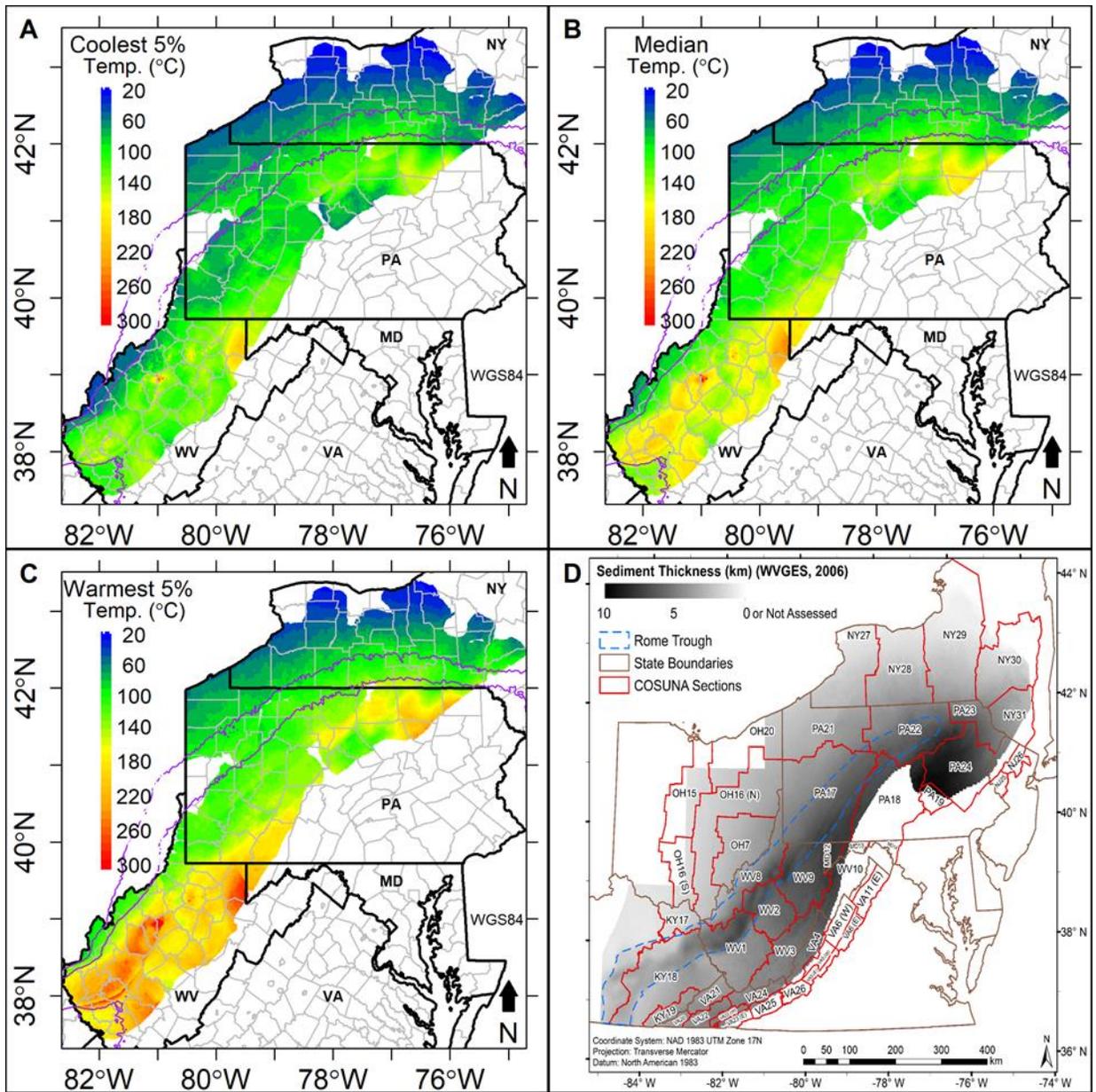
The first step in the uncertainty analysis was to estimate temperature-depth profiles at all grid cell sites in the Appalachian Basin study area. Figure 3.3A-C shows the results of that Monte Carlo uncertainty analysis displayed as quantiles of temperature at the basement depth. If adequate temperatures for a heat utilization process are not found within sedimentary rocks below a site, the site would be infeasible for developing a geothermal sedimentary reservoir.

Quantile maps are used in Figure 3.3A-C instead of mean and standard deviation maps because the Monte Carlo distributions of temperature tended to be positively skewed. Furthermore, the 5<sup>th</sup> percentile map provides estimates of design temperatures at the basement depth which are 95% reliable. Figure 3.3D provides a map of the depth to basement rocks for reference (WVGES, 2006; the raster file is available on the Geothermal Data Repository [Cornell University, 2016a]). The Rome Trough failed rift structural feature (Harris et al., 2002) provides relatively thick sedimentary rock cover in western West Virginia and southwestern Pennsylvania that correlate with the relatively warmer temperatures estimated for that region of the basin.

We would like to have temperatures greater than 60 °C in sedimentary reservoirs for direct-use heating projects. The coolest 5<sup>th</sup> percentile map in Figure 3.3A shows that

temperatures greater than 60 °C are likely to be found in sedimentary rocks across most of the basin area (81% of the area evaluated) with 95% reliability. Exceptions occur along the northern rim of the basin in New York, where the basement is located at shallower depths. Such temperatures could be adequate for direct-use heating needs if a permeable reservoir were available from which to extract heat. This observation further motivates the analysis of potential geothermal reservoirs in this paper. Figure 3.3B reports median temperature values, which provides a less risk-averse view of the geothermal resource.

The temperature quantile maps (Figure 3.3A-C) illustrate that the uncertainty in temperatures at the basement depth, expressed as the difference between temperatures in Figure 3.3A compared to Figure 3.3C, is different across the basin. For any location, the uncertainty in the predicted temperature increases as the prediction depth increases, so regions with deeper depths to basement rock generally have more uncertainty in map temperature predictions. Variations in uncertainty as a function of spatial location also result from differences in surface heat flow spatial interpolation precision. For example, consider regions with depths to basement around 3 - 4 km (between the purple contours on Figure 3.3A-C). In New York and northwestern Pennsylvania, these regions tend to have a smaller 5<sup>th</sup> – 95<sup>th</sup> quantile range compared to southern West Virginia. By using a surface heat flow map with spatially correlated values and prediction uncertainties as input information to the heat conduction model, the estimated temperature-depth profiles also reflect the spatial correlation of the heat flow values and prediction uncertainties.



**Figure 3.3:** Estimated quantiles of temperatures at the basement depth for each of 138,400 1 km<sup>2</sup> grid cells, and the depth to basement rocks in the Appalachian Basin. A: coolest 5<sup>th</sup> quantile estimates, B: median estimates, C: warmest 5<sup>th</sup> quantile estimates, D: depths to basement rock. On A-C, purple lines indicate 3 km and 4 km depth contours. On D, the COSUNA geologic regions are provided for reference. These are labeled by state, followed by the number of the column and, where needed, the cardinal direction in parenthesis. The estimated location of the Rome Trough (Repetski et al., 2008) coincides with much of the area of maximum depths. State abbreviations: KY: Kentucky, MD: Maryland, NY: New York, OH: Ohio, PA: Pennsylvania, VA: Virginia, WV: West Virginia.

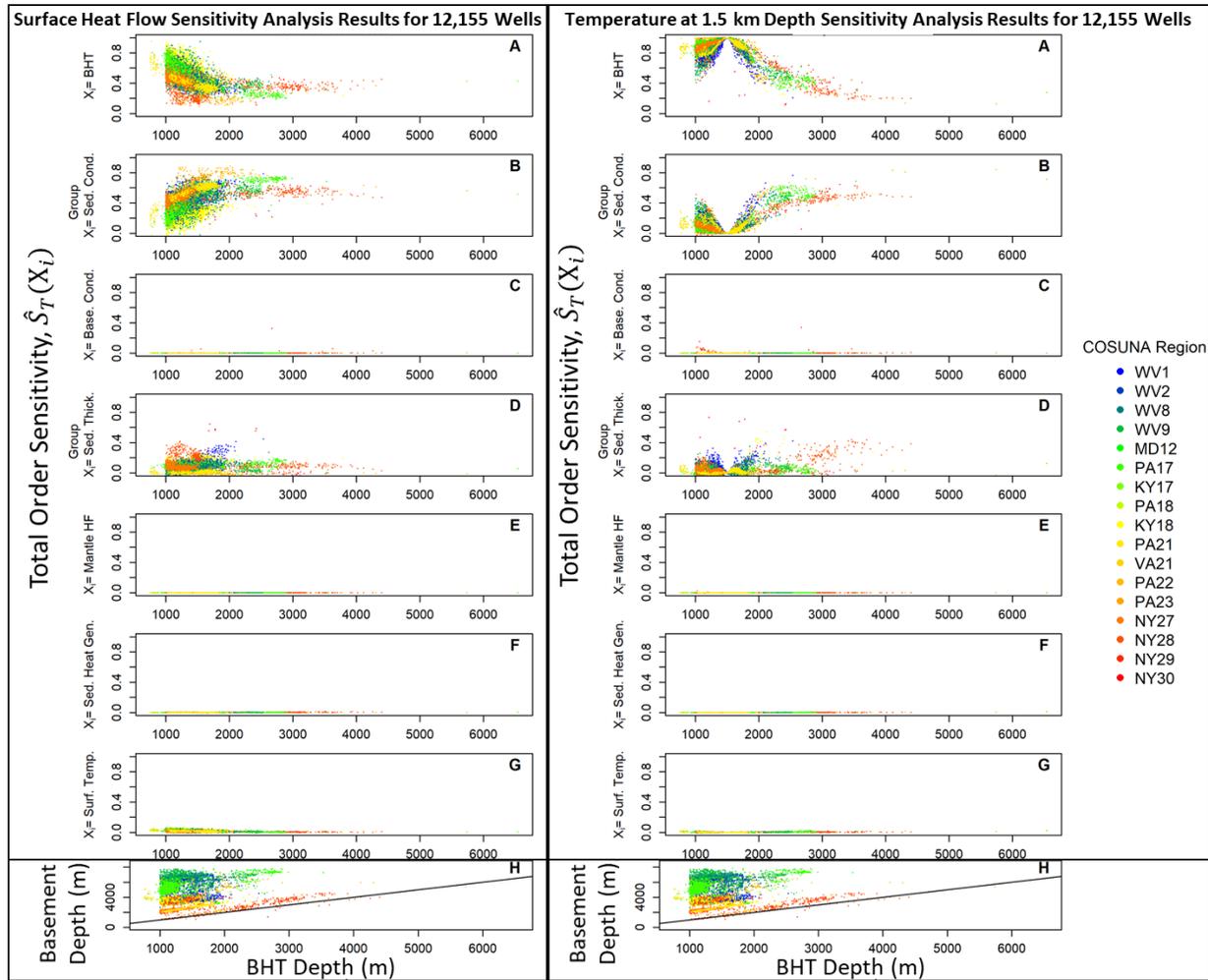
### **3.5.2 Sensitivity Analysis for the Surface Heat Flow and Temperatures at 1.5 km Depth at Well Locations**

Figure 3.4A-G present the computed total order effects from the sensitivity analysis of the surface heat flow and the temperature at 1.5 km depth at 12,155 well locations, given all of the uncertainties in model input variables. Each point in Figure 3.4 represents a single well. Each panel A-G represents a different input variable or variable group used in the heat conduction model (Table 3.1). Figure 3.4H provides a plot of the basement depth vs. bottom-hole temperature (BHT) measurement depth because each of these factors impacts the sensitivity results. In particular, the uncertainty in basement conductivity is important for the few wells with BHT measurements in basement rocks (points below the black line in Figure 3.4H). Points are colored by the COSUNA geologic region (Figure 3.3D) that they are located in, because sedimentary rock thermal conductivities and thicknesses are different across the geologic regions of the basin. Sensitivity analysis results appear to cluster by COSUNA geologic region (Figure 3.4A, B, D). This result suggests that future spatial regressions of surface heat flow for geothermal resource assessment of the Appalachian Basin should consider these or an alternative definition of geologic regions as categorical variables to account for potentially different error structures between regions.

The sensitivity analysis results reveal that uncertainties in three input variables or groups tend to have the largest effect on the uncertainty variance of the computed heat flow and temperature at 1.5 km depth at well locations. These three inputs are the BHT, the sedimentary rock thermal conductivities group, and the sedimentary rock thicknesses group. For the surface heat flow, the BHT tends to be a more critical variable when measured at shallower depths compared to deeper depths (Figure 3.4A). BHTs measured at deeper depths require propagation

of heat flow through greater rock cover, so thermal conductivities become more critical variables for wells with deeper BHT measurements (Figure 3.4B). The sensitivity of temperature at 1.5 km depth to thermal conductivity shows a similar trend with BHT depth (Figure 3.4B). When 1.5 km depth is in basement rocks, the basement thermal conductivity becomes a more critical variable, as expected (Figure 3.4C). When the BHT is located at 1.5 km depth, it is the most critical variable for calculating the temperature at 1.5 km depth, which demonstrates that the model is working as expected (Figure 3.4A).

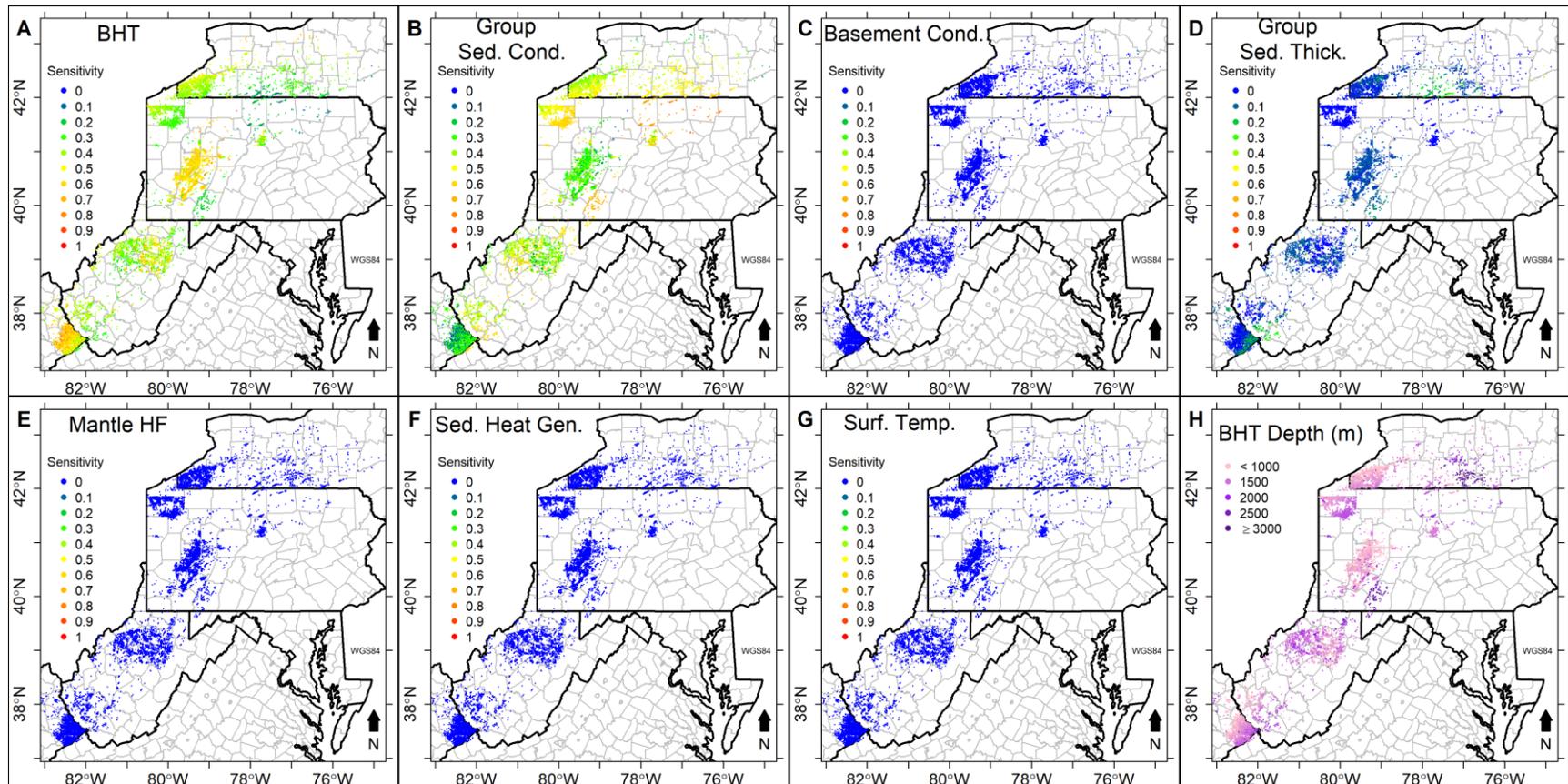
Several variables have relatively minor contributions to the uncertainty variance of surface heat flow and temperature at 1.5 km depth, given our probability distribution assumptions. These sensitivity analysis results suggest that values assigned to the mantle heat flow (Figure 3.4E) and the sedimentary rock heat generation (Figure 3.4F) may be fixed at their most likely values in probabilistic assessments of heat conduction models; each of these variables explain less than 1% of the variance in the output variables for our dataset. For BHTs measured deeper than about 2 km, the surface temperature (Figure 3.4G) may also be fixed.



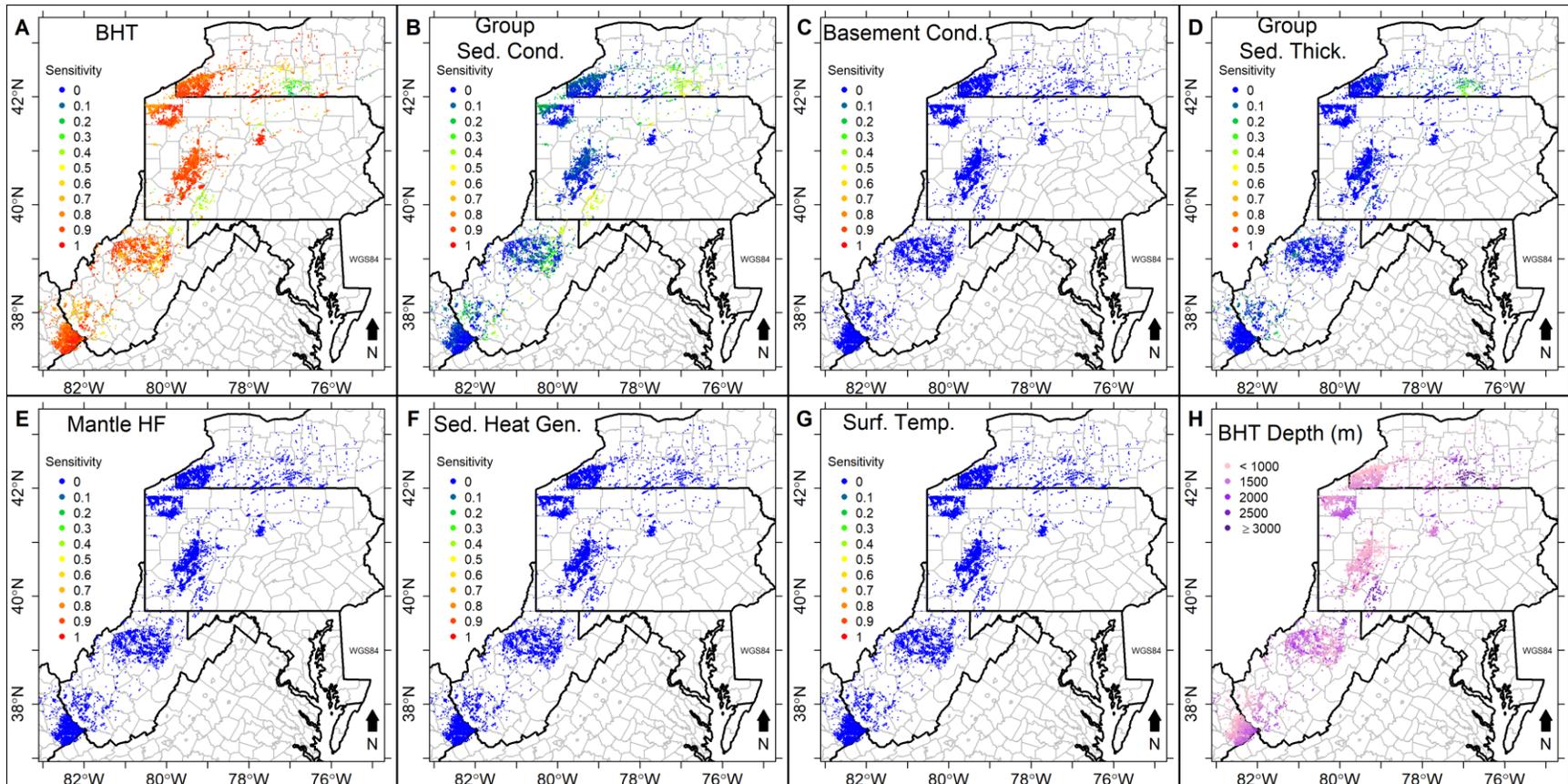
**Figure 3.4:** Total order sensitivity effects for the surface heat flow (left panel) and the temperature at 1.5 km depth (right panel) for each of 12,155 wells (points on plots) for five individual input variables and two input variable groups. A-G: total order effect vs. bottom-hole temperature (BHT) measurement depth. A: BHT, B: sedimentary rock thermal conductivity group, C: basement rock thermal conductivity, D: sedimentary rock thickness group, E: mantle heat flow, F: sedimentary rock heat generation, G: surface temperature. H: Basement depth vs. BHT depth for wells used in the sensitivity analysis. Point colors correspond to each of seventeen COSUNA geologic regions (Figure 3.3D).

Maps of the total order sensitivity of the surface heat flow (Figure 3.5A-G) and temperature at 1.5 km depth (Figure 3.6A-G) at the 12,155 well locations are provided for each input variable. For reference, Figure 3.5H and Figure 3.6H provide maps of the BHT measurement depth for each well. One can observe a relationship between the sensitivity and the BHT depth for several variables (Figure 3.4A, B, D).

Viewing spatially the sensitivity analysis results in Figure 3.5 and Figure 3.6 reveals which data to acquire in support of geothermal projects at specific sites: the data that would be most valuable to acquire to reduce uncertainty in the surface heat flow and temperatures at depth depends on location within the basin. This is mainly a result of differences in the depths of BHT data across the basin. To reduce uncertainty in the surface heat flow in regions that have primarily deep wells, such as south-central New York, more precise estimates of rock thermal conductivities (Figure 3.5B) and thicknesses (Figure 3.5D) would be preferred over more precise BHTs. Most of the Appalachian Basin region contains wells at depths near 1.5 km, so the temperature at 1.5 km depth is most sensitive to the BHT over much of the region. In contrast, in areas such as south-central New York and southern Pennsylvania, wells tend to be more than twice as deep as the 1.5 km depth of interest. In those regions, calculations of temperature at 1.5 km depth tend to be similarly sensitive to sedimentary rock thermal conductivity and the BHT. This result implies that for existing Appalachian Basin data, sensitivity results will be spatially variable as a function of the target depth of a geothermal reservoir. Therefore, unsurprisingly, improving estimates of the depths to target reservoirs at potential geothermal project sites should inform data collection efforts that aim to reduce uncertainty in the predicted temperatures at the target depths.



**Figure 3.5:** Maps of the total order sensitivity effects for the surface heat flow at 12,155 well locations (points) for five individual input variables and two input variable groups. Each input variable has its own map of total order sensitivity: A: bottom-hole temperature (BHT), B: sedimentary rock thermal conductivity group, C: basement rock thermal conductivity, D: sedimentary rock thickness group, E: mantle heat flow, F: sedimentary rock heat generation, G: surface temperature. H: map of the BHT measurement depths at each well location.



**Figure 3.6:** Maps of the total order sensitivity effects for the temperature at 1.5 km depth at 12,155 well locations (points) for five individual input variables and two input variable groups. Each input variable has its own map of total order sensitivity: A: bottom-hole temperature (BHT), B: sedimentary rock thermal conductivity group, C: basement rock thermal conductivity, D: sedimentary rock thickness group, E: mantle heat flow, F: sedimentary rock heat generation, G: surface temperature. H: map of the BHT measurement depths at each well location.

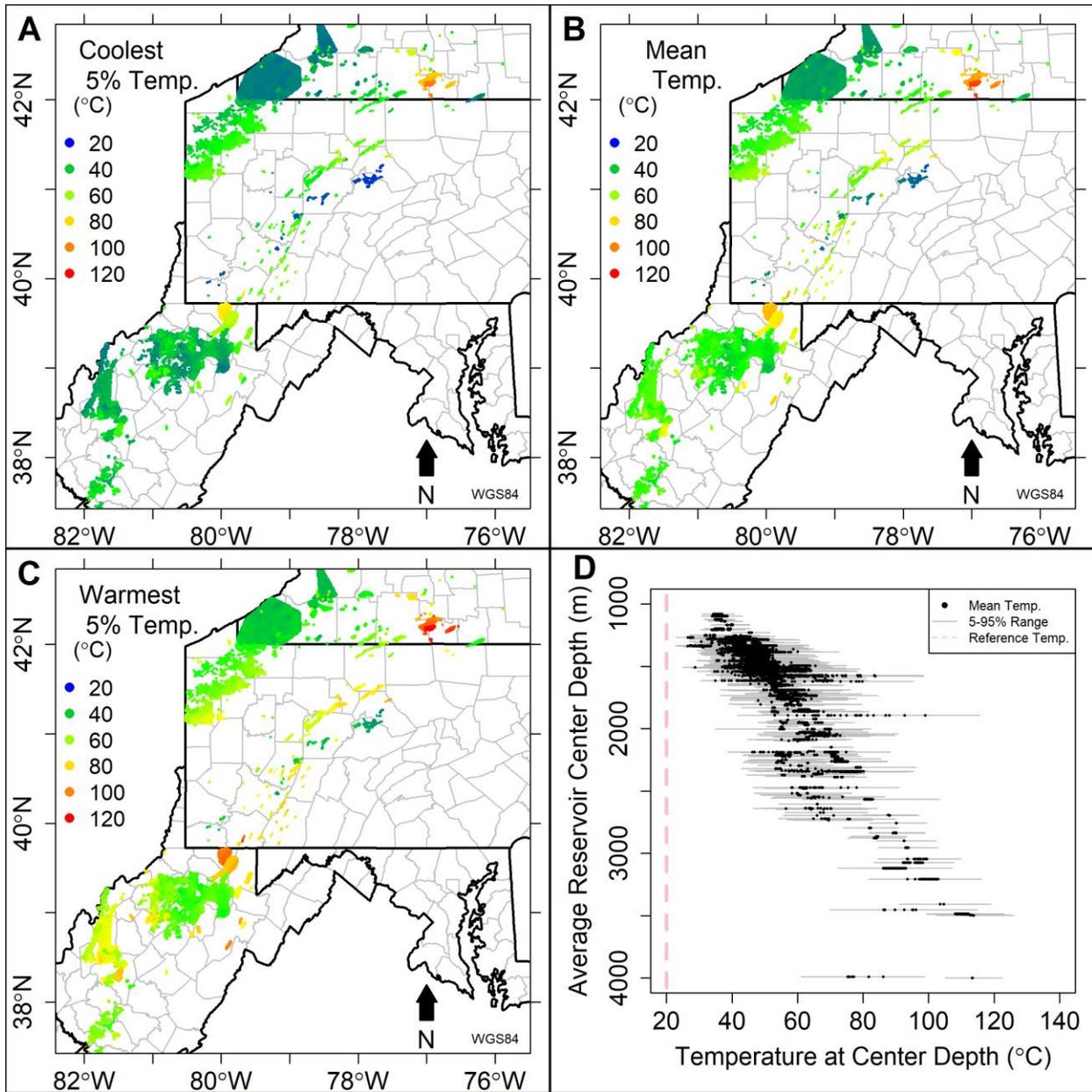
Figure 3.5 and Figure 3.6 are particularly useful to display and interpret the sensitivity analysis results for all 12,155 wells spread across the basin. In locations with a high density of wells, it is not necessary to conduct a sensitivity analysis at every well if BHT values and their depths of measurement are similar, and the locations have similar geologic characteristics. To reduce the number of points to evaluate in a sensitivity analysis, one could use a random subsample of the wells in these high-density locations to understand the trends in sensitivity. Similarly, to construct Figure 3.4, a sensitivity analysis for all wells was not necessary. Instead, one might divide the wells into BHT depth categories of specified width (e.g. 500 m), and then select at random N number of wells from each BHT depth category. To preserve the observed effect of geologic region on the sensitivity analysis results, the random sample could be specified for each geologic region. Using such a random sample would generate in Figure 3.4 a more uniform density of points, and could make the graphs easier to interpret.

### **3.5.3 Distributions of Temperatures at Reservoir Depths**

The second step in the uncertainty analysis focused on the bodies of rock with properties suitable to be geothermal reservoirs, as defined by a reservoir dataset that is biased toward known oil and gas occurrences (Camp et al., 2018). This step used the estimated distributions of temperature-depth profiles at each 1 km<sup>2</sup> grid cell in the Appalachian Basin to extract distributions of temperatures at the depths of potential geothermal reservoirs. Reservoirs were located in a total of 16,700 grid cells. These temperature distributions were used for stored thermal energy calculations over the thickness of each potential reservoir for each 1 km<sup>2</sup> grid cell of the reservoir.

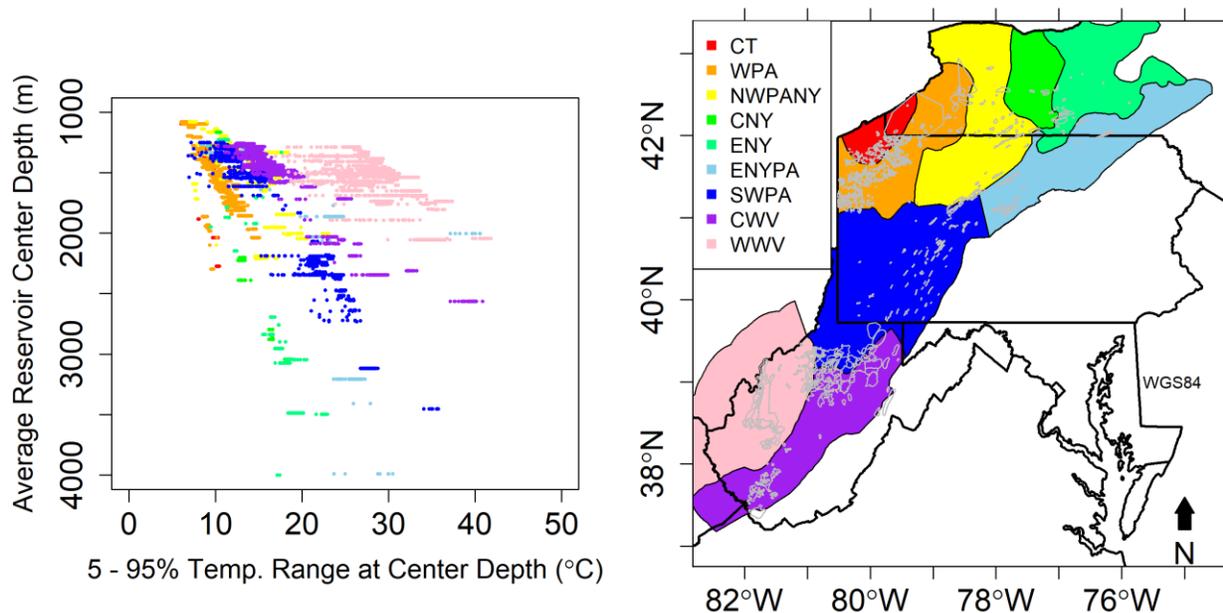
Figure 3.7A-C provides quantile and mean estimates for the temperatures at the center depth of each reservoir grid cell. In Figure 3.7A-C, multiple reservoirs may overlap in space when they are located at different depths. As a result, abrupt changes in temperature among neighboring map grid cells is generally a result of different reservoirs, which may be located at substantially different depths. Figure 3.7D provides the estimated mean reservoir temperatures, with uncertainty, plotted at the average center depth of each reservoir.

The mean temperatures at the depth of these reservoirs range from about 25 °C to 110 °C (Figure 3.7B, D), which allows for a variety of options for direct-use of heat across the basin. For reservoirs that span large spatial areas, including a reservoir in western New York, there is spatial variability in the predicted temperatures at the depth of the reservoir. As stated in Section 3.3.2.4, the probability distributions assigned to the depth and thickness of a reservoir are the same across the reservoir, so the spatial variability in predicted temperatures primarily results from the spatial variability in the surface heat flow map. This result highlights a need to evaluate the stored thermal energy for a reservoir using a spatial analysis grid that is comparable in size to the spatial area of rock that would be used to extract geothermal energy. The 1 km<sup>2</sup> grid resolution used in this study is comparable in size to permitted exploration areas in basins that have been developed for direct-use of geothermal energy.



**Figure 3.7:** Estimated temperatures at the center depth of each of 16,700 1 km<sup>2</sup> areas of potential geothermal reservoirs in the Appalachian Basin. A: coolest 5<sup>th</sup> percentile, B: mean, C: warmest 5<sup>th</sup> percentile. D: the estimated mean temperature (black dot) and 5-95% temperature range (gray) for each 1 km<sup>2</sup> area of the reservoirs evaluated. Points are plotted at the average depth to the center of the reservoir. The vertical dashed pink line indicates the reference temperature for stored thermal energy calculations.

Figure 3.8 provides a plot of the 5-95% temperature range at the center depth of reservoirs as a function of reservoir center depth for each 1 km<sup>2</sup> area of each reservoir. The 5-95% temperatures ranges vary from less than 10 °C for the shallowest reservoirs to as much as 45 °C for deeper reservoirs. The uncertainty in the temperatures at depth is well-explained by the reservoir's location within the geologic regions that defined interpolation sections for stratified kriging of the surface heat flow (Smith, 2016). An important discovery of Smith (2019, Ch.2; 2016) was the need for a spatially varying nugget semi-variance for the surface heat flow between geologic regions. In Smith (2019, Ch. 2), the geologic regions with the largest nugget semi-variances for the surface heat flow were Western West Virginia, Central West Virginia, and Eastern New York and Pennsylvania (Figure 3.8). Reservoirs in these three regions tend to have the largest temperature uncertainty ranges at depth (Figure 3.8). Additional variability in the temperature at depth uncertainty within an interpolation region results from differences in well clustering, and differences in geologic rock column properties (e.g. COSUNA regions and the basement depth shown in Figure 3.3).

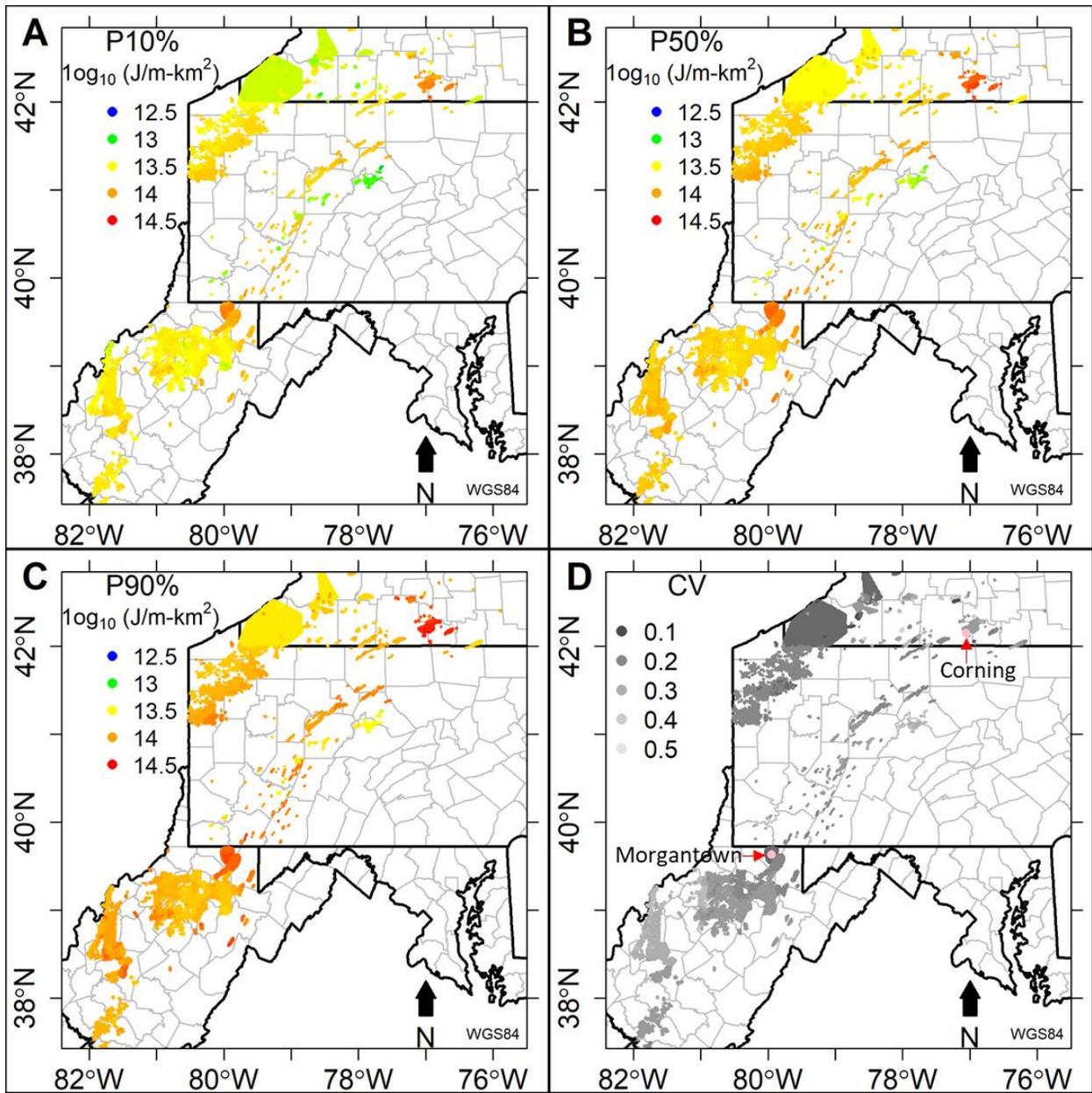


**Figure 3.8:** Left: 5-95% quantile range of temperatures at the center depth of each of 16,700 1 km<sup>2</sup> areas of the reservoirs evaluated. Right: Geologic regions used for interpolation by stratified kriging of the surface heat flow by Smith (2016). Reservoir locations are shown in gray outlines. Colors of the geologic regions match the scatterplot point colors (left). Geologic region abbreviations: CT-Chautauqua, NY, WPA-Western Pennsylvania, NWPANY-Northwestern New York and Pennsylvania, CNY-Central New York, ENY-Eastern New York, ENYPA-Eastern New York and Pennsylvania, SWPA-Southwestern Pennsylvania, CWV-Central West Virginia, WWV-Western West Virginia.

### 3.5.4 Distributions of Stored Thermal Energy for Reservoirs

For each of the 16,700 1 km<sup>2</sup> grid cells, the estimated distributions of temperatures at reservoir depths were used to compute the distribution of stored thermal energy using Monte Carlo uncertainty analysis of the heat-in-place model variables (Table 3.4). Figure 3.9A-C shows quantiles of the resulting stored thermal energy distributions for each 1 km<sup>2</sup> area of each reservoir. Figure 3.9D provides the coefficient of variation for the stored thermal energy. To follow a literature convention of reporting energy recoverable from geothermal reservoirs (e.g. Sanyal and Sarmiento, 2005), we present stored thermal energy for the smallest 10<sup>th</sup> percentile (P10% “proven reserves”), median (P50% “probable reserves”), and largest 10<sup>th</sup> percentile

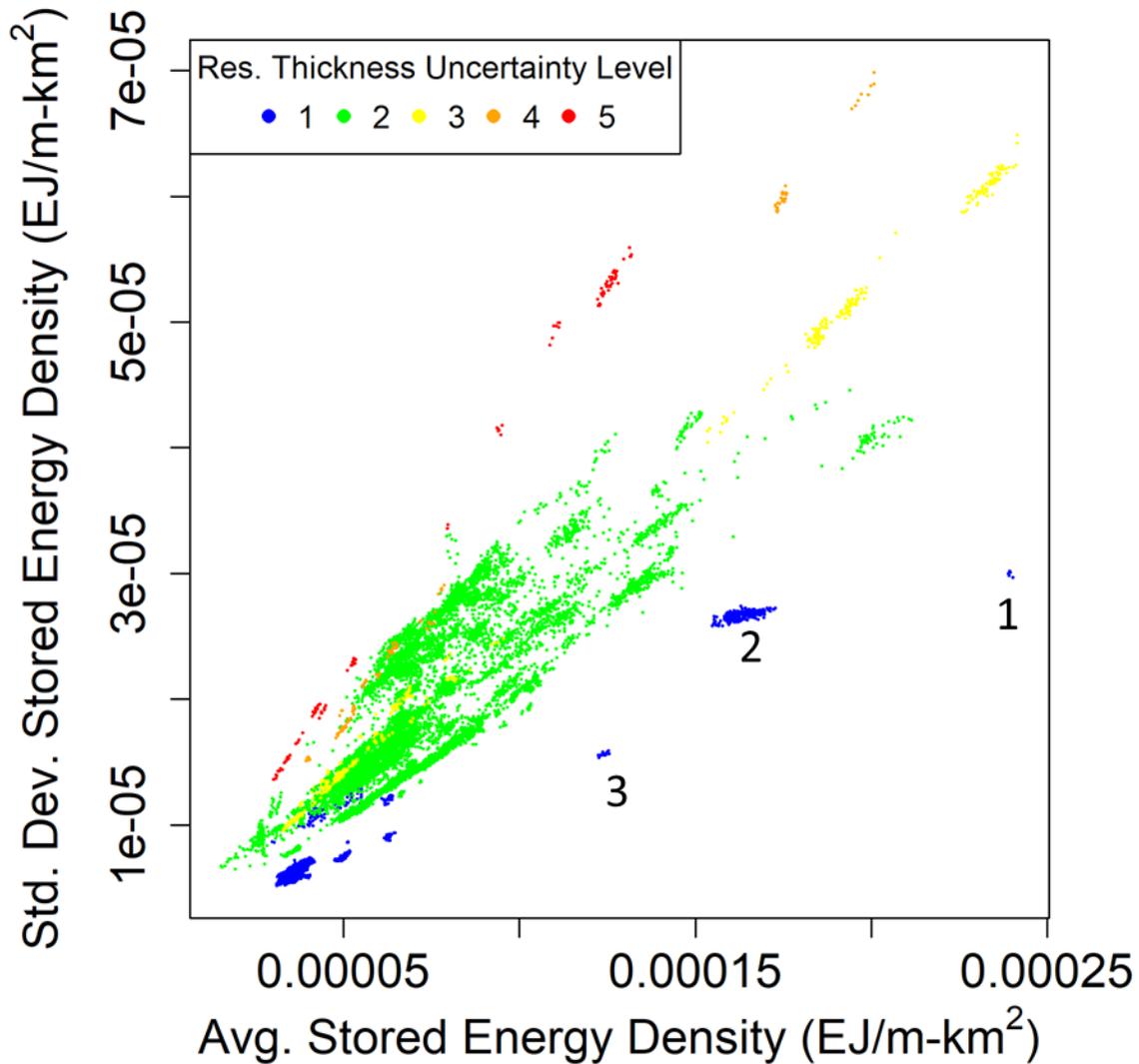
(P90% “possible reserves”). Quantiles in Figure 3.9A-C provide the energy density (energy per unit volume of reservoir rock, specified as the 1 km<sup>2</sup> spatial area multiplied by reservoir thickness in m). The most favorable reservoirs by energy density are located in south-central New York and in West Virginia near the border with Pennsylvania. Reservoirs in northwestern Pennsylvania and much of West Virginia also have moderate to high energy densities. Coefficients of variation (Figure 3.9D) tend to be smaller for the reservoirs in northwestern Pennsylvania, and therefore may be preferable to the West Virginia reservoirs with comparable amounts of stored heat in terms of relative uncertainty, i.e., risk.



**Figure 3.9:** Energy density (stored thermal energy per unit volume), specified for each of 16,700  $1 \text{ km}^2$  spatial areas times the corresponding reservoir thickness (meters) at that location. A: estimated 10<sup>th</sup> percentile, B: estimated median, C: estimated 90<sup>th</sup> percentile, D: coefficient of variation in real space. On D, pink circles represent the locations of the two labeled cities.

We considered favorable reservoirs to have high stored thermal energy density and low uncertainty. Figure 3.10 provides a plot of the standard deviation in the stored thermal energy density as a function of the average stored thermal energy density. Each of the  $1 \text{ km}^2$  reservoir

areas (points) are colored by the uncertainty in the reservoir thickness, which has an appreciable impact on the uncertainty in the stored thermal energy. Generally, the uncertainty in the stored thermal energy increases as the average value increases. This results from larger uncertainties in temperature at greater depths (Figure 3.8).



**Figure 3.10:** Standard deviation of the stored thermal energy density vs. the average stored thermal energy density for each of 16,700 1 km<sup>2</sup> areas of reservoirs evaluated. Energy density is the 1 km<sup>2</sup> area multiplied by the reservoir thickness (meters). Points are colored by the uncertainty level of the reservoir thickness. Numbers in the plot correspond to favorable reservoirs evaluated in further detail: 1 and 3) Corning, NY Black River dolomite, 2) Morgantown, WV Tuscarora sandstone.

Of the reservoirs in Figure 3.10 with high average stored thermal energy density, the three most favorable reservoirs in terms of low uncertainty are 1) a Black River dolomite reservoir near Corning, NY (location in Figure 3.9D), 2) a Tuscarora Sandstone reservoir near Morgantown, WV (location in Figure 3.9D), and 3) a second Black River dolomite reservoir near Corning, NY. The more favorable Black River reservoir is estimated to be about 45 m thick over a 5 km<sup>2</sup> area, with mean total energy of 0.054 EJ. The other Black River reservoir is estimated to be only 8 m thick over an 18 km<sup>2</sup> area, with mean total energy of about 0.019 EJ. The Tuscarora reservoir is estimated to have the largest volume of about 120 m over a 241 km<sup>2</sup> area, with mean total energy of about 4.80 EJ.

The average stored thermal energy for the Corning reservoirs could be sufficient to meet the heating needs for the Corning community for decades. Corning has a population of about 10,000 people who live in about 3,000 households (Data USA, 2018a). The average household energy use in New York State in 2009 was about 103 million BTU/year, of which about 70% was used for space heating and water heating (EIA, 2009). Using this information, for the estimated mean stored thermal energy, the maximum useful life for each of the two reservoirs would be:

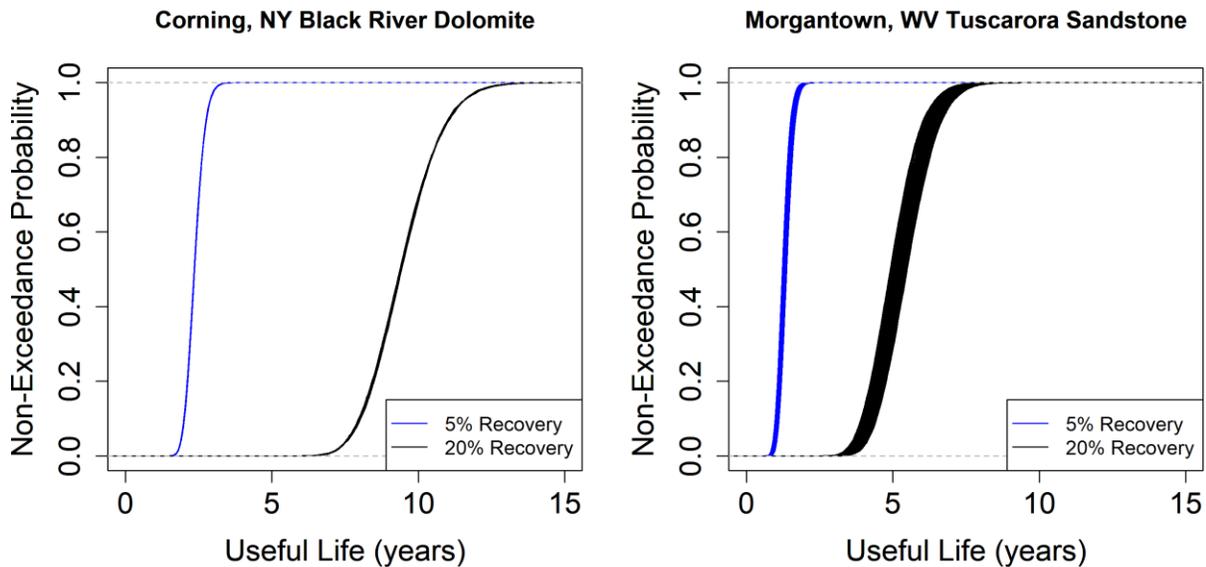
$$\frac{0.054E^{18} J}{\left(3000 \text{ households} * \frac{103E^6 \text{ BTU}}{\text{yr} \cdot \text{household}} * \frac{1055.06 J}{\text{BTU}} * 0.7\right)} \approx 240 \text{ years.}$$

$$\frac{0.019E^{18} J}{\left(3000 \text{ households} * \frac{103E^6 \text{ BTU}}{\text{yr} \cdot \text{household}} * \frac{1055.06 J}{\text{BTU}} * 0.7\right)} \approx 83 \text{ years}$$

These estimates do not consider a thermal energy recovery factor that describes how much of the total energy could be extracted from the reservoir. Recovery factors typically range from 5 – 20% for geothermal projects (Williams et al., 2008; Nathanson, 1975). A low-end

recovery factor of 5% would result in about 4-5 years of useful life for the smaller reservoir, and a high-end recovery factor of 20% would result in about 50 years of useful life for the larger reservoir. Using the same energy consumption assumptions for Morgantown and adjusting for 10,000 households (Data USA, 2018b), the mean total stored thermal energy of the Tuscarora reservoir could provide a low-end estimate of about 300 years of heating if all 241 km<sup>2</sup> were developed.

Geothermal projects tend to develop spatial areas on the order of 1 km<sup>2</sup>. To inform economic evaluations, the useful life of the developed portion of a reservoir for a project should be considered. Figure 3.11 provides empirical distributions of the useful life for each 1 km<sup>2</sup> area of the more favorable Corning reservoir, and the Morgantown Tuscarora reservoir. Recovery factors of 5% and 20% are provided as bounds on the production performance. For even a high-end recovery factor of 20%, the useful lifetime of each 1 km<sup>2</sup> area is estimated to be less than 15 years. Within the Morgantown Tuscarora spatial area there are more and less favorable areas to develop, given the currently available data (spread of curves in Figure 3.11). For example, a useful life of 5 years in some locations corresponds to about 30% non-exceedance probability, and in other locations corresponds to about 60% non-exceedance probability. Areas of the Corning reservoir exhibit less variability as a result of the reservoir having about 48 times smaller spatial coverage.



**Figure 3.11:** Empirical cumulative distribution functions of the useful life for each 1 km<sup>2</sup> area the Corning Black River Dolomite (left, n = 5 km<sup>2</sup>), and the Morgantown Tuscarora Sandstone (right, n = 241 km<sup>2</sup>). Each black line represents a unique 1 km<sup>2</sup> area of the reservoir. Results are shown for recovery factors of 5% (blue) and 20% (black).

### 3.5.5 Sensitivity Analysis of Stored Thermal Energy for Reservoirs

Sensitivity analysis of the heat-in-place model results was used to explore which of the temperature and geologic variable uncertainties contributed the most to the uncertainty in the stored thermal energy for reservoirs in the Appalachian Basin. Figure 3.12A-F provides the results of the sensitivity analysis for each input variable for each of 16,700 1 km<sup>2</sup> grid cell areas of reservoirs. Each point represents a 1 km<sup>2</sup> area of a reservoir. To linearize the scatterplots, total order sensitivity is plotted as a function of the temperature variability (5-95% range) relative to the mean temperature at the center depth of the reservoir. Points are colored according to which factors most influenced the observed trends in sensitivity.

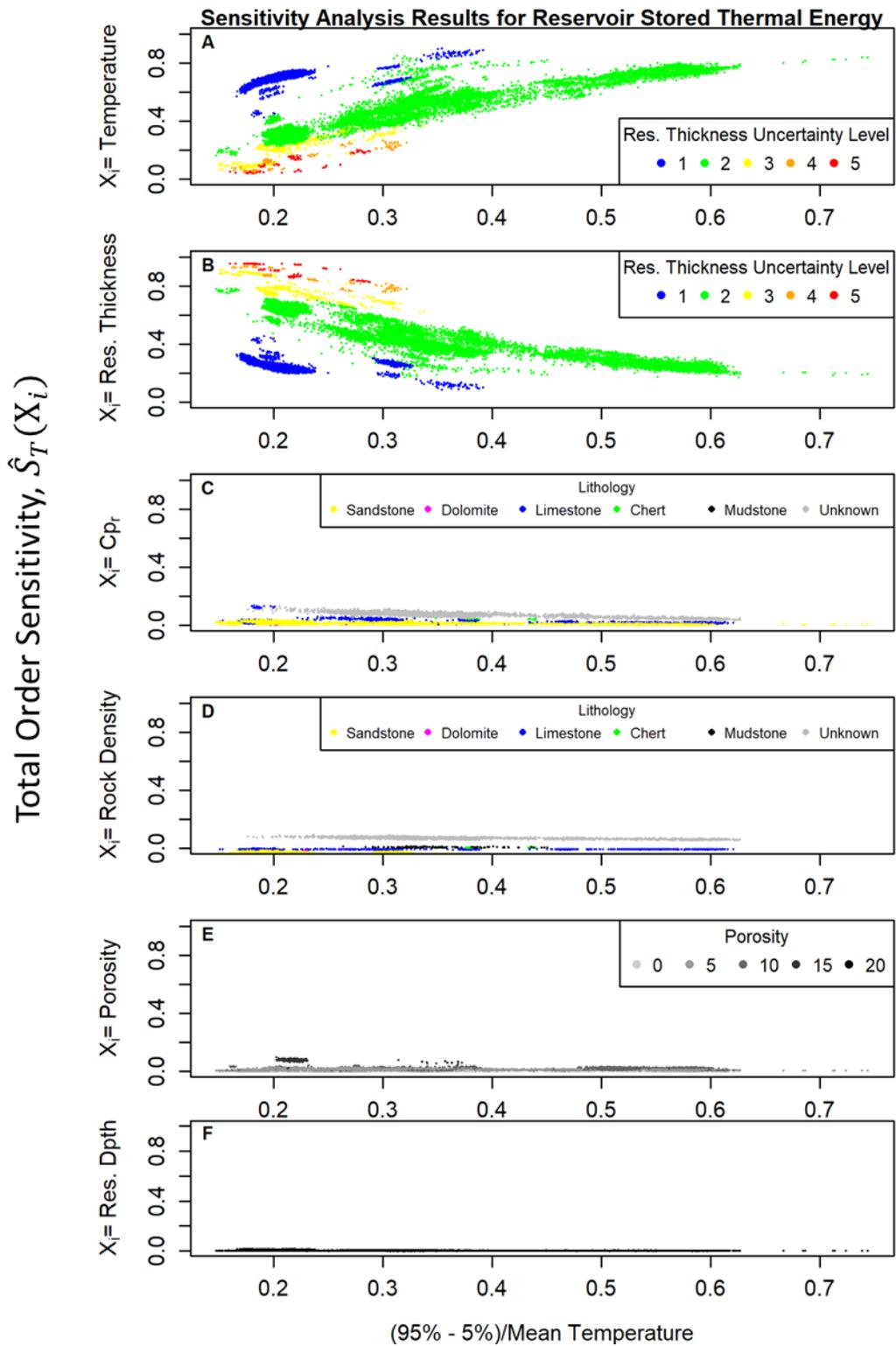
Temperature and reservoir thickness are colored by the uncertainty in reservoir thickness (Figure 3.12A, B). Formation thickness and temperature tend to trade off as the variables that

contribute most to stored thermal energy uncertainty for a site. This result suggests that if local well data were to become available with which to further constrain formation tops and reduce uncertainty in formation thickness, the sensitivity of the stored thermal energy estimate to the temperature estimate would tend to proportionally increase. Palmer-Wilson et al. (2018) also found that temperature and reservoir thickness were the most critical variables for electricity generation estimation in low-temperature geothermal settings.

The rock heat capacity and density are colored by lithology of the reservoir (Figure 3.12C, D). For reservoirs with “Unknown” lithologies, which spanned the range of physically plausible rock density and heat capacity, the sensitivity results indicate that about 10-15% of the uncertainty variance in the stored thermal energy is explained by the uncertainty in these variables. For reservoirs with known lithologies, the explained variance tended to be less than 5%. These results are remarkable because they suggest that heat capacity and density, whose values are rarely available for specific sites, may be assigned generic ranges by lithology and have little impact on the resulting estimates of uncertainty in stored thermal energy.

Porosity is colored by the average porosity of the reservoir (Figure 3.12E). The sensitivity analysis results suggest that in our analysis porosity values across the Appalachian Basin reservoirs may be specified as constants in probabilistic modeling when a reservoir is expected to have small to moderate porosity; the total order sensitivity values are less than 5% for reservoirs with less than 15% average porosity. Porosity became more critical with increasing porosity as a result of relatively more water filling the analyzed volume. The heat capacity of water is about 3-4 times the heat capacity of rocks, so having more water in the volume increases the stored energy estimate.

There were no meaningful trends in the total order sensitivity of stored thermal energy to the reservoir depth, so a color scheme for this variable is not assigned by relation to another variable (Figure 3.12F). The depth to the reservoir had little to no impact on the stored thermal energy uncertainty, given the probability distribution that we used to describe depth uncertainty for each reservoir.

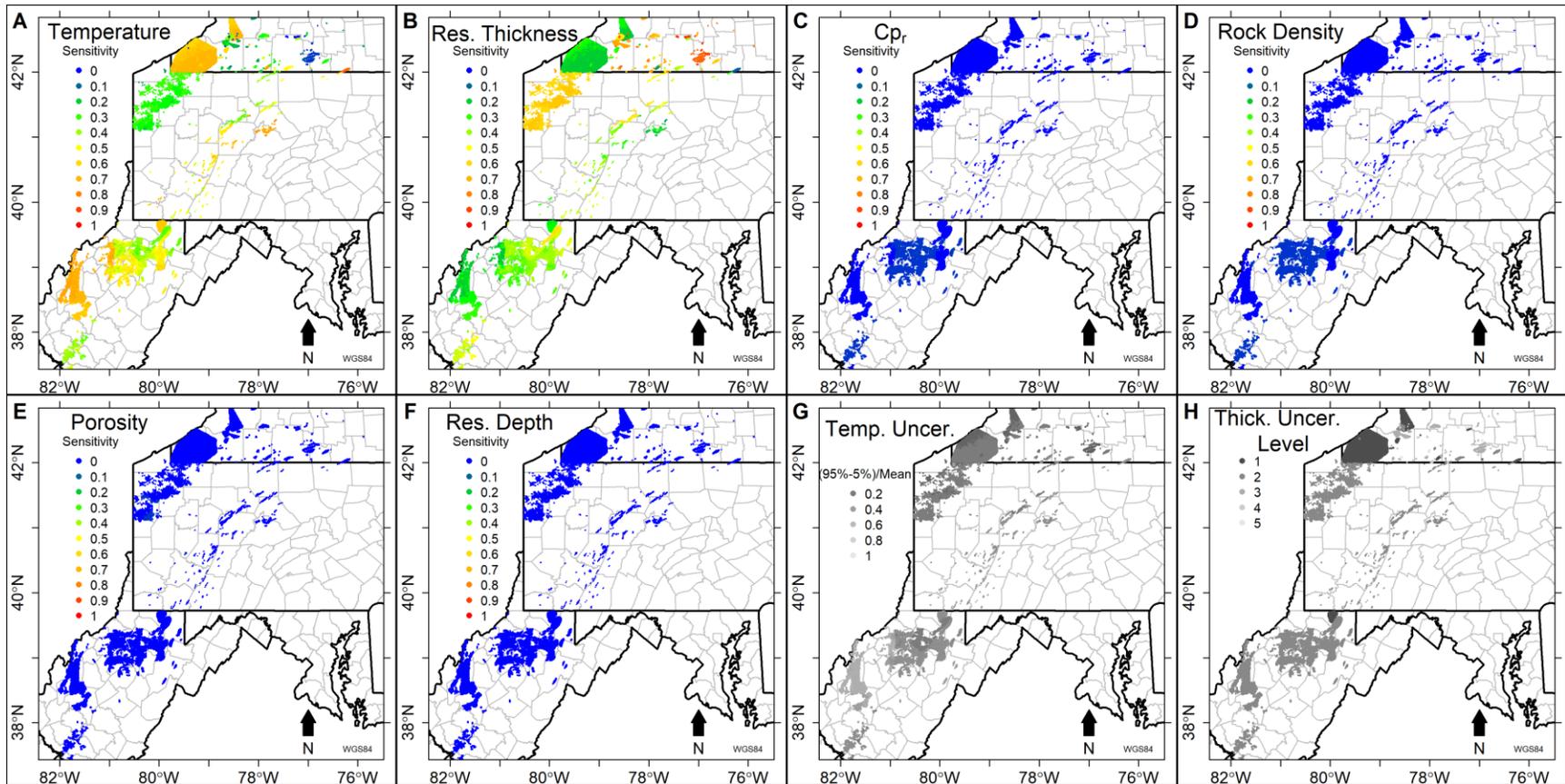


**Figure 3.12:** Total order sensitivity of the stored thermal energy to each input variable for each of 16,700 1 km<sup>2</sup> areas of reservoirs evaluated. Each point corresponds to a 1 km<sup>2</sup> reservoir area. A: reservoir temperature, B: reservoir thickness, C: rock heat capacity, D: rock density, E: rock porosity, F: reservoir depth.

Figure 3.13A-F provide maps of the total order sensitivity of the stored thermal energy to each input variable in the heat-in-place model. Each point in Figure 3.13 corresponds to one of 16,700 1 km<sup>2</sup> areas of reservoirs evaluated. Figure 3.13A, B, G, and H illustrate spatially the observed Figure 3.12A and B trends in sensitivity as a function of reservoir thickness uncertainty and temperature uncertainty. Where the temperature uncertainty is relatively large and the reservoir thickness uncertainty is relatively small, the stored thermal energy tends to be most sensitive to rock temperature. Reservoirs in western New York exhibit this trend. The complement of these conditions occurs in some northwestern Pennsylvania reservoirs, wherein reservoir rock thickness uncertainty contributes more to the stored thermal energy uncertainty than the rock temperature (Figure 3.13A, B, G, H).

If the uncertainty in stored thermal energy at a particular site within a reservoir is too large to be acceptable for a decision maker to proceed with drilling the first well in the greenfield, the spatial presentation of the sensitivity analysis results in Figure 3.13 may be used to inform which variables would be most valuable to collect to reduce uncertainty in the stored thermal energy for the site. With the currently available data, for the two favorable reservoirs near Corning, NY, reducing uncertainty in estimates of the reservoir rock thickness would be more useful to collect than reducing temperature uncertainty in the reservoir. However, the uncertainty in the reservoir rock thickness for Corning, NY reservoirs is already the lowest level implemented on the Camp et al. (2018) uncertainty scale (Figure 3.10, Table 3.2) as a result of the well log data available. If further reduction of uncertainty were required for the risk of developing a reservoir to be acceptable, then additional well logs from the region may be acquired.

For the Morgantown, WV target reservoir, to reduce uncertainty in stored thermal energy, reducing rock temperature uncertainty would be most useful. Reducing uncertainty in the predicted temperature-depth profiles could involve: 1) re-measuring temperatures in wells that have likely achieved thermal equilibrium with the rock since the wells were drilled, 2) reducing uncertainty in the heat conduction model variables that were identified to contribute most to the uncertainty in the surface heat flow for wells near Morgantown (Figure 3.5, Figure 3.6), and 3) improving the regression methods used to estimate surface heat flow and updating the regressions with newly acquired datasets from options 1 and 2 (e.g. using regression kriging [Hengl et al., 2007; Journel and Rossi, 1989] within the stratified regions used by Smith [2016]). Based on the heat conduction model sensitivity analysis results in Figure 3.5, to reduce uncertainty in the surface heat flow near Morgantown, it would be most valuable to reduce uncertainty in temperature measurements and to improve estimates of sedimentary rock thermal conductivities. With more precise surface heat flow predictions, the subsequent distributions of temperature-depth profiles and stored thermal energy would also become more precise.



**Figure 3.13:** Maps of total order sensitivity of the stored thermal energy for each of 16,700 1 km<sup>2</sup> areas of reservoirs for each input variable (A-F), and plots of the temperature and thickness uncertainty for each 1 km<sup>2</sup> area (G, H). A: reservoir temperature, B: reservoir thickness, C: rock heat capacity, D: rock density, E: rock porosity, F: reservoir depth, G: temperature uncertainty, H: reservoir thickness uncertainty level.

For the four input variables whose uncertainties contributed most to uncertainty in stored thermal energy, we regressed their total order sensitivity effects on selected subsets of the input variables to the heat-in-place model. This analysis revealed which of the input variables contributed the most to the observed variability in the sensitivity effects for these four input variables. The regression models used in this analysis included some categorical variables (e.g. lithology), for which we defined a base category, and one binary variable for each other category whose effect is estimated relative to the base category.

Regression models adopted here are all linear as provided in Equation 3.9

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3.9}$$

where  $\mathbf{Y}$  is the vector of sensitivity effects,  $\mathbf{X}$  is a matrix containing  $N$  rows corresponding to the number of grid cells and  $k+1$  columns corresponding to the number of regression parameters plus one column of ones for the intercept,  $\boldsymbol{\beta}$  is the vector of regression parameters, and  $\boldsymbol{\epsilon}$  is the vector of residuals. This analysis assumes that residuals are spatially independent, which is unlikely given the presentation of spatial sensitivity analysis results (Figure 3.13); however, the simple regression analysis is informative to explain variability across the input space.

To parameterize the regression equations, initially all input variables were included individually in regression models. Input variables included mean values and measures of variability for each heat-in-place model variable in Table 3.4. For the temperature input variable, the metric used on the x-axis in Figure 3.12 was used to linearize the effect of temperature on the sensitivity effects. Categorical variables evaluated included reservoir rock lithology, COSUNA geologic region, and the uncertainty levels for variables. For all sensitivity effects, a single input variable determined at least 80% of the observed variability. The input variables with the most

statistically significant parameter effects were additively included into multiple regression models. Interaction terms were considered, but were generally not statistically significant or practically significant, given the magnitude of the effect.

For each of the four input variable total order sensitivity effects evaluated, the regression model parameters and their estimated values and significance levels are provided in Table 3.5. For reservoir temperature and thickness sensitivity effects, relative temperature uncertainty and the uncertainty level for reservoir thickness explained about 90% of the variability. Adding the COSUNA column region as a predictor increased the adjusted  $R^2$  to about 94% for these regressions, but the regression parameters were generally an order of magnitude or more smaller than the thickness uncertainty levels.

For rock heat capacity and density, the lithology of the rock was the most important variable to describe variability in these sensitivity effects. This meets expectations because we defined probability distributions for these variables by their lithology. For heat capacity, additional variability is described by the temperature of the rock, which is expected because heat capacity is modeled as a function of temperature in this study.

**Table 3.5:** Regression models and parameter estimates for the stored thermal energy sensitivity effects. Bold and italics indicates significance less than 0.001.  $T_{95}$  is the warmest 5<sup>th</sup> percentile of temperature,  $T_5$  is the coolest 5<sup>th</sup> percentile of temperature, and  $T_{\text{mean}}$  is the mean temperature.

<b>Y: Total Order Effect Variable</b>	<b>X: Regression Model Variables</b>	<b><math>\beta</math>: Parameter Estimates Value, Standard Deviation and Model Adjusted <math>R^2</math></b>
Reservoir Temperature-Depth Profile	<p>Intercept <math>\beta_0</math> is defined for Reservoir Thickness Uncertainty Level 1</p> <p><math>X_1: (T_{95} - T_5)/T_{\text{mean}}</math>  <math>X_2</math>: Thickness Uncertainty Level 2  <math>X_3</math>: Thickness Uncertainty Level 3  <math>X_4</math>: Thickness Uncertainty Level 4  <math>X_5</math>: Thickness Uncertainty Level 5</p>	<p><math>\beta_0</math>: <b><i>0.442</i></b>, 0.0012  <math>\beta_1</math>: <b><i>1.186</i></b>, 0.0038  <math>\beta_2</math>: <b><i>-0.356</i></b>, 0.0011  <math>\beta_3</math>: <b><i>-0.498</i></b>, 0.0024  <math>\beta_4</math>: <b><i>-0.573</i></b>, 0.0049  <math>\beta_5</math>: <b><i>-0.596</i></b>, 0.0049</p> <p>Model Adjusted <math>R^2 = 0.91</math></p>
Reservoir Thickness	<p>Intercept <math>\beta_0</math> is defined for Reservoir Thickness Uncertainty Level 1</p> <p><math>X_1: (T_{95} - T_5)/T_{\text{mean}}</math>  <math>X_2</math>: Thickness Uncertainty Level 2  <math>X_3</math>: Thickness Uncertainty Level 3  <math>X_4</math>: Thickness Uncertainty Level 4  <math>X_5</math>: Thickness Uncertainty Level 5</p>	<p><math>\beta_0</math>: <b><i>0.466</i></b>, 0.0012  <math>\beta_1</math>: <b><i>-1.041</i></b>, 0.0039  <math>\beta_2</math>: <b><i>0.345</i></b>, 0.0011  <math>\beta_3</math>: <b><i>0.547</i></b>, 0.0024  <math>\beta_4</math>: <b><i>0.631</i></b>, 0.0051  <math>\beta_5</math>: <b><i>0.662</i></b>, 0.0051</p> <p>Model Adjusted <math>R^2 = 0.90</math></p>
Rock Heat Capacity	<p>Intercept <math>\beta_0</math> is defined for Chert lithology</p> <p><math>X_1: (T_{95} - T_5)/T_{\text{mean}}</math>  <math>X_2</math>: Dolomite  <math>X_3</math>: Limestone  <math>X_4</math>: Mudstone  <math>X_5</math>: Sandstone  <math>X_6</math>: Unknown</p>	<p><math>\beta_0</math>: <b><i>0.066</i></b>, 0.0015  <math>\beta_1</math>: <b><i>-0.0552</i></b>, 0.00053  <math>\beta_2</math>: <b><i>-0.050</i></b>, 0.0024  <math>\beta_3</math>: <b><i>-0.018</i></b>, 0.0016  <math>\beta_4</math>: <b><i>-0.037</i></b>, 0.0017  <math>\beta_5</math>: <b><i>-0.035</i></b>, 0.0015  <math>\beta_6</math>: <b><i>0.027</i></b>, 0.0015</p> <p>Model Adjusted <math>R^2 = 0.91</math></p>
Rock Density	<p>Intercept <math>\beta_0</math> is defined for Chert lithology</p> <p><math>X_1</math>: Dolomite  <math>X_2</math>: Limestone  <math>X_3</math>: Mudstone  <math>X_4</math>: Sandstone  <math>X_5</math>: Unknown</p>	<p><math>\beta_0</math>: 0.003, 0.0016  <math>\beta_1</math>: <b><i>-0.020</i></b>, 0.0025  <math>\beta_2</math>: <b><i>-0.012</i></b>, 0.0016  <math>\beta_3</math>: 0.003, 0.0018  <math>\beta_4</math>: <b><i>-0.046</i></b>, 0.0016  <math>\beta_5</math>: <b><i>0.066</i></b>, 0.0016</p> <p>Model Adjusted <math>R^2 = 0.97</math></p>

### 3.6 Discussion

This paper provided methods for stochastic modeling to generate distributions of temperature-depth profiles and stored thermal energy. The methodology was applied in the Appalachian Basin using available temperature and geologic datasets. Drilling to basement rock was not evaluated in this study because of the relatively greater risk of suitable permeability relative to the former hydrocarbon reservoirs evaluated. However, basement rocks would have higher temperatures than the sedimentary reservoirs evaluated in this study, and future work could assess the suitability of basement geothermal reservoirs using the temperature-depth profile uncertainty analysis methodology presented in this paper.

The primary interest of this paper was the suitability of using rock volumes whose productivity were well-characterized by available hydrocarbon production and drilling data, relative to other rocks in the basin volume and basement rocks. Given this focus, this paper did not fully characterize the geothermal resource production potential of the Appalachian Basin, which may have other rock volumes with desirable natural permeability. Nevertheless, the uncertainty analysis framework presented in this paper would be essential for propagating uncertainties in the limited information for those other potential greenfield geothermal reservoirs, whose uncertainties would be comparably larger than those reservoirs evaluated in this study. The results of the sensitivity analysis in this paper inform which of the geologic variables are most important to devote time to characterizing.

Analysis of stored thermal energy revealed several reservoirs with relatively high thermal energy density, some of which also had relatively low energy density uncertainty. Such reservoirs located near population centers, like those near Morgantown, WV and Corning, NY, may be suitable for long-term geothermal district heating operations. Other reservoirs located

away from population centers can provide opportunities for new industrial users of low-temperature heat. For either end use application, the results from this resource assessment provide an opportunity for probabilistic evaluations of a reservoir's ability to meet the objectives of surface utilization projects.

For such site-specific analyses, geothermal reservoir recovery factors should be considered based on expected flow geometries of the reservoir, instead of the simpler percentage extraction proxy method used in this paper. Additional considerations include the utilization efficiency of the surface equipment, and whether the temperature of the reservoir matches the required temperature of the utilization processes. The distributions of temperature-depth profiles obtained in this study allow for probabilistic assessment of meeting temperature objectives over the operating lifetime of a geothermal system.

If the uncertainty in meeting site-specific objectives provides too much risk for an institution to proceed with drilling a geothermal well, additional data may be collected that aims to reduce the uncertainty in the resource assessment. The sensitivity analysis results in this paper are useful for such geothermal greenfield evaluations; it would indicate which of the geologic and temperature variables would be most valuable to collect for uncertainty reduction. Selection of which data to acquire should consider both the economics of collecting particular datasets, and which variables contribute most to the resource uncertainty. Neglecting economics, the best variables to collect to reduce the uncertainty of the surface heat flow at well locations are bottom-hole temperature (BHT) measurements, sedimentary rock thermal conductivities, and sedimentary rock thicknesses. Of these three variables, which one is the most critical for specific sites is a function of the BHT measurement depth.

For the calculation of stored thermal energy in reservoirs, uncertainty in the reservoir temperature contributed most to the stored thermal energy uncertainty for many reservoirs. Yet surprisingly, uncertainty in the reservoir temperature was not always the most critical contribution to stored thermal energy uncertainty; in some places, reservoir thickness tended to trade off with temperature as the most critical variable across the basin. If reductions in the stored thermal energy uncertainty are required for the risk of reservoir development to be acceptable, for sites where stored thermal energy is most sensitive to reservoir thickness, existing local well log datasets could be used to refine estimates of the depth to formation tops. For sites where stored thermal energy is most sensitive to temperature, new temperatures could be acquired for existing wells near a site of interest, or alternative methods to predict or to constrain temperatures at depth may be implemented (e.g. Vogt et al., 2010). Reservoir temperature uncertainty may also be reduced by reducing the uncertainty in heat conduction model variables, particularly thermal conductivities.

A goal of this study was to evaluate the effect of spatially correlated datasets on the resulting uncertainty of geothermal resources. The input thermal information for calculating stored thermal energy was a spatially correlated map of the surface heat flow. The uncertainty analysis framework presented in this paper allowed for propagation of these spatially-correlated values and their prediction uncertainties through to the calculations of temperature-depth profiles and stored thermal energy. The magnitude of the uncertainty in temperatures at the depths of reservoirs were found to be well-explained by location within regions of the Appalachian Basin which had different spatial correlation structures. As a result of region-specific uncertainty, considering spatial correlation structures of underlying datasets was useful for comparing site favorability across a large region where a project may be proposed. Within a specific reservoir,

the stored thermal energy was spatially variable as a result of surface heat flow spatial variability. This result highlights a need to evaluate the stored thermal energy for a reservoir using a spatial grid that is comparable in size to the spatial area of rock that would be used to extract geothermal energy.

Garg and Combs (2010) state that stochastic heat-in-place methods rarely use site-specific uncertainties. For our dataset, sensitivity analysis results suggest that it is less fruitful to obtain site-specific values of rock porosity, density, and heat capacity than it is to obtain reservoir temperatures and formation tops that define a reservoir's thickness. For stochastic modeling of stored thermal energy, assigning most likely values to variables whose uncertainties have little impact on the stored thermal energy may provide computationally tractable probabilistic resource assessments over large areas (e.g. global in Limberger et al., 2018). For large-area analyses, the results from our regression analysis of the sensitivity effects may be useful to predict which input variables will contribute most to the sensitivity effects, given the site characteristics and the uncertainties in the available data.

### **3.7 Conclusions**

This paper presented stochastic methods for uncertainty analysis of conduction-dominated geothermal systems. For the Appalachian Basin, uncertainty analysis methods were used to calculate distributions of temperature-depth profiles that were used directly to estimate distributions of stored thermal energy in potential reservoirs. By using a spatially correlated heat flow dataset as the input thermal information to the heat conduction model and the heat-in-place model, the resulting uncertainty distributions of temperature-depth profiles and stored thermal energy are also spatially correlated. We showed that differences in the magnitude of temperature-

depth profile uncertainty across the Appalachian Basin could be important to consider when comparing reservoir favorability across the region.

The methods presented in this paper may be used to inform probabilistic assessments of a reservoir's ability to meet future heating demands, and to rank a reservoir's resource favorability relative to other reservoirs in a region where geothermal projects may be proposed. Using the distributions of stored thermal energy, we demonstrated a simple analysis of the estimated useful life of reservoirs near two population centers. The results of our sensitivity analysis of stored thermal energy were used to suggest prioritization of data collection efforts that aim to reduce uncertainty at each site.

Across the Appalachian Basin, the variables that contributed most to uncertainty in stored thermal energy varied spatially. Generally, the uncertainty in the temperature of the reservoir and the uncertainty in the reservoir thickness, which was estimated from oil and gas well logs, contributed most to the uncertainty in the stored thermal energy. The uncertainty in reservoir temperature results from uncertainty in the temperature-depth profile predictions. For individual well locations, the temperature-depth uncertainty tends to result from uncertainty in the temperature measurement (bottom-hole temperatures in our study), in sedimentary rock thermal conductivities, and in the thicknesses of sedimentary rock formations within the geologic column, as estimated from generic geologic columns defined over broad spatial areas. Other variables that we evaluated for stored thermal energy calculations, including rock heat capacity, rock density, rock porosity, rock heat generation, and mantle heat flow, each explained a relatively small amount of variability. This result implies that geothermal exploration efforts that aim to reduce uncertainty in stored thermal energy should prioritize refining temperature,

thermal conductivity, and depths to formation tops through improvements in spatial predictive modeling, or acquisition or processing of additional data, where possible.

### **3.8 Acknowledgements**

The author would like to thank coauthors Teresa Jordan, Jery Stedinger, Erin Camp and Calvin Whealton for discussions about the Appalachian Basin geologic reservoirs, and uncertainty and sensitivity analysis methods presented in this paper.

This work was completed while some authors were supported by a U.S. Department of Energy-funded project for deep direct-use of geothermal energy at Cornell University (Award Number DE-EE0008103). The analysis of the Morgantown, WV Tuscarora Sandstone reservoir was supported by the U.S. Department of Energy for a deep direct-use project at West Virginia University (Award Number DE-EE0008105). The authors would like to thank members of the West Virginia University project team, particularly Nagasree Garapati, Scott McCleery, Ron McDowell, and Jessica Moore, for sharing data and for discussions about the Tuscarora sandstone properties.

### **3.9 Code and Data Availability, and Software Credits**

The code for the one-dimensional heat conduction model is available on a Github repository (Horowitz et al., 2015). Other code and data are available upon email request to Jared Smith (jds485@cornell.edu).

Statistical analyses were completed using R version 3.5.0 (R Core Team, 2018) and the packages doParallel (Microsoft Corporation and S. Weston, 2017), sp (Pebesma and Bivand,

2005), rgdal (Bivand, Keitt, and Rowlingson, 2017), raster (Hijmans, 2017), GISTools (Brunsdon and Chen, 2014), pracma (Borchers, 2018), and EnvStats (Millard, 2013).

## REFERENCES

- Abdulagatov, I.M., Z. Z. Abdulagatova, S. N. Kallaev, A. G. Bakmaev, and P. G. Ranjith. (2015). Thermal-diffusivity and heat-capacity measurements of sandstone at high temperatures using laser flash and DSC methods. *Int J Thermophys.*, 36(4). Pp. 658 - 691. DOI 10.1007/s10765-014-1829-4
- Anselin, L.. (1999). Interactive techniques and exploratory spatial data analysis. In Longley, P., M. Goodchild, D. Maguire, D. Rhind (Eds.). *Geographical Information Systems*. Wiley. New York, NY. pp. 253 - 266.
- Artemieva, I.M., and W.D. Mooney. (2001). Thermal thickness and evolution of Precambrian lithosphere: A global study. *J. Geophys. Res.*, 106(B8). Pp. 16,387 - 16,414.
- Avary, K.L.. (1996). The Lower Silurian Tuscarora Sandstone fractured anticlinal play. In J.B. Roen, and B.J. Walker (Eds.). *The atlas of major Appalachian gas plays*. West Virginia Geological and Economic Survey, Publication V-25. Pp. 151 - 155.
- Banks, J., and N.B. Harris. (2018). Geothermal potential of foreland basins: A case study from the Western Canada Sedimentary Basin. *Geothermics*, 76. Pp. 74 - 92.
- Beckers, K. and K. McCabe. (2018). GEOPHIRES. Online GitHub Repository. <https://github.com/kfbeckers/GEOPHIRES>.
- Bivand, R., T. Keitt, and B. Rowlingson. (2017). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.2-16. <https://CRAN.R-project.org/package=rgdal>
- Blackwell, D.D., P.T. Negraru, and M.C. Richards. (2007). Assessment of the enhanced geothermal system resource base of the United States. *Nat. Res. Research*, 15(4). Pp. 283 - 308. DOI: 10.1007/s11053-007-9028-7.
- Blackwell, D.D., M. Richards, J. Batir, Z. Frone, and J. Park. (2010). New geothermal resource map of the northeastern US and technique for mapping temperature at depth. *GRC Trans.*, 34. Pp. 313 - 318.
- Blackwell, D., M. Richards, Z. Frone, J. Batir, A. Ruzo, R. Dingwall, and M. Williams. (2011). Temperature at depth maps for the conterminous US and geothermal resource estimates. *GRC Trans.*, 35. Pp. 1545 - 1550.
- Borchers, H.W.. (2018). pracma: Practical Numerical Math Functions. R package version 2.1.4. <https://CRAN.R-project.org/package=pracma>.
- Brezinski, D.K.. (2011). Maryland borehole temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/cc54f15894222c91e71e4530dc088fec> (accessed 6.22.16).
- BRGM. (2015). A "geothermal" database on the Dogger aquifer in the Paris Basin. URL: <https://www.brgm.eu/project/geothermal-database-on-dogger-aquifer-paris-basin> (accessed 3.31.19)
- Brunsdon, C. and H. Chen. (2014). GISTools: Some further GIS capabilities for R. R package version 0.7-4. <https://CRAN.R-project.org/package=GISTools>
- Bücker, C. and L. Rybach. (1996). A simple method to determine heat production from gamma-ray logs. *Marine and Petroleum Geology*, 13(4). Pp. 373 - 375.
- Camp, E.. (2017). Repurposing Petroleum Reservoirs for Geothermal Energy: A Case Study of the Appalachian Basin. Cornell University, Ph.D. Dissertation. pp. 1–219.

- Camp, E.R., T.E. Jordan, M.J. Hornbach, and C.A. Whealton. (2018). A probabilistic application of oil and gas data for exploration stage geothermal reservoir assessment in the Appalachian Basin. *Geothermics*, 71. pp. 187 - 199.
- Carter, L.S., S.A. Kelley, D.D. Blackwell, and N.D. Naeser. (1998). Heat flow and thermal history of the Anadarko Basin, Oklahoma. *AAPG Bulletin*, 82(2). Pp. 291 - 316.
- Cornell University. (2016a). Appalachian Basin Play Fairway Analysis: Improvements in 2016 to thermal quality analysis in Low-Temperature Geothermal Play Fairway Analysis (GPFA-AB) [data set] [WWW Document]. URL <https://gdr.openei.org/submissions/879>
- Cornell University. (2016b). Natural sedimentary reservoirs data geothermal play fairway analysis 2016 revision [data set]. URL <https://gdr.openei.org/submissions/881>
- Cornell University. (2017). Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. URL <https://gdr.openei.org/submissions/899>
- Cressie, N.. (1988). Spatial prediction and ordinary kriging. *Math. Geol.*, 20. Pp. 405–421. doi:10.1007/BF00892986
- Curl, D.. (2011). Kentucky borehole temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/168566464e3d5f8f3cde3b9fc004bd38> (accessed 6.22.16).
- Data USA. (2018a). Data USA - Bar Chart of Property Value in Corning, Ny [dataset]. URL <https://datausa.io/profile/geo/corning-ny/>
- Data USA. (2018b). Data USA - Bar Chart of Household Income in Morgantown, WV [dataset]. URL <https://datausa.io/profile/geo/morgantown-wv/#housing>
- Deming, D.. (1989). Application of bottom-hole temperature corrections in geothermal studies. *Geothermics*, 18. Pp. 775–786. DOI:10.1016/0375-6505(89)90106-5
- Dresel, P.E.. (1985). The geochemistry of oilfield brines from western Pennsylvania: Pennsylvania State University, State College, PA, 237 p.
- EIA. (2009). Household Energy Use in New York. URL [https://www.eia.gov/consumption/residential/reports/2009/state\\_briefs/pdf/NY.pdf](https://www.eia.gov/consumption/residential/reports/2009/state_briefs/pdf/NY.pdf)
- Empire State Oil and Gas Information System (ESOGIS). (2015). ESOGIS [WWW Document]. URL [www.esogis.nysm.nysed.com](http://www.esogis.nysm.nysed.com)
- Frone, Z., and D.D. Blackwell. (2010). Geothermal map of the northeastern United States and the West Virginia thermal anomaly. *GRC Trans.*, 34. Pp. 339 - 344.
- Frone, Z.S., D.D. Blackwell, M.C. Richards, and M.J. Hornbach. (2015). Heat flow and thermal modeling of the Appalachian Basin, West Virginia. *Geosphere*, 11(5). Pp. 1279 - 1290. DOI: 10.1130/GES01155.1.
- Garg, S.K., and J. Combs. (2015). A reformulation of USGS volumetric “heat in place” resource estimation method. *Geothermics*, 55. Pp. 150 - 158.
- Gass, T.E.. (1982). The geothermal heat pump. *GRC Bulletin*, 11. Pp. 3 - 8.
- Gudmundsson, J.S., D.H. Freeston, and P.J. Lienau. (1985). The Lindal diagram. *GRC Trans.*, 9. Pp. 15 - 19. DOI:10.1002/yd.430
- Hadi, J., P. Quinlivan, G. Ussher, O. Alamsyah, B. Promono, and A. Masri. (2010). Resource risk assessment in geothermal greenfield development: An economic implications. *Proceedings World Geothermal Congresss 2010*. Bali, Indonesia.
- Harris, D.C., J.A. Drahovzal, J.B. Hickman, B.C. Nuttall, M.T. Baranoski, and K.L. Avery. (2002). Rome Trough Consortium Final Report and Data Distribution.

- Hengl, T., G.B.M. Heuvelink, and D.G. Rossiter. (2007). About regression-kriging: From equations to case studies. *Comput. Geosci.*, 33. Pp. 1301 - 1315.  
<https://doi.org/10.1016/j.cageo.2007.05.001>
- Hijmans, R.J. (2017). raster: Geographic Data Analysis and Modeling. R package version 2.6-7.  
<https://CRAN.R-project.org/package=raster>
- Hornby, P., F. Boschetti, and F.G. Horowitz. (1999). Analysis of potential field data in the wavelet domain. *Geophys. J. Int.*, 137. Pp. 175–196. <https://doi.org/10.1046/j.1365-246x.1999.00788.x>
- Horowitz, F.G.. (2015). Identifying potentially activatable faults in GPFA-AB. In: Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. Pp. 298 - 316.
- Horowitz, F.G., J.D. Smith, and C.A. Whealton. (2015). One dimensional conductive geothermal Python code [Data repository]. GeothermalCode. URL  
<https://bitbucket.org/geothermalcode/onedimensionalgeothermalheatconductionmodel.git>
- Jaeger, J. C. (1965). Application of the theory of heat conduction to geothermal measurements. In W.H.K. Lee (Ed.). *Terrestrial Heat Flow*. American Geophysical Union. Washington, D.C.. Pp. 7 - 23. DOI: 10.1029/GM008p0007.
- Jaupart, C., and J.C. Mareschal. (1999). The thermal structure and thickness of continental roots. *Lithos*, 48. Pp. 93 - 114.
- Johnsson, M.J.. (1986). Distribution of maximum burial temperatures across northern Appalachian Basin and implications for Carboniferous sedimentation patterns. *Geology*, 14. Pp. 384 - 387.
- Journel, A.G., Rossi, M.E., 1989. When do we need a trend model in kriging? *Math. Geol.*, 21, pp. 715 - 739. <https://doi.org/10.1007/BF00893318>
- Lachenbruch, A.H. (1968). Preliminary geothermal model of the Sierra Nevada. *J. of Geophys. Res.*, 73(2). Pp. 6977 - 6989.
- Lee, Y., S. Park, J. Kim, H.C. Kim, and M. Koo. (2010). Geothermal resource assessment in Korea. *Renewable and Sustainable Energy Reviews*, 14. pp. 2392 - 2400.
- Leftwich, T.. (2011). Ohio borehole temperatures [WWW Document]. AASG Geotherm. Data Repos. URL  
<http://repository.stategeothermaldata.org/repository/resource/ba2f0b9d21f71acfe10609f76e2699e6> (accessed 6.22.16)
- Limberger, J., T. Boxem, M. Pluymaekers, D. Bruhn, A. Manzella, P. Calcagno, F. Beekman, S. Cloetingh, and J.-D. van Wees. (2018). Geothermal energy in deep aquifers: A global assessment of the resource base for direct heat utilization. *Renewable and Sustainable Energy Reviews*, 82. Pp. 961 - 975.
- Lowry, T.S., J.T. Finger, C.R. Carrigan, A. Foris, M.B. Kennedy, T.F. Corbett, C.A. Doughty, S. Pye, and E.L. Sonnenthal. (2017). Reservoir maintenance and development task report for the DOE geothermal technologies office GeoVision study. Sandia Report SAND2017-9977. 81 p.
- Lynch, R.S., and Castor, T.P., 1983, Auburn low-temperature geothermal well final report. NYSERDA report 84-18.
- Manger, G.E. (1963). Porosity and bulk density of sedimentary rocks. Geological Survey Bulletin 1144-E. Washington, D.C. 60 p.

- Matsumoto, M.R., J.F. Atkinson, M.D. Bunn, and D.S. Hodge. (1996). Disposal/recovery options for brine waters from oil and gas production in New York State. NYSERDA Report 96-4, pp. 1 - 168.
- McCleery, R.S., J.P. Moore, R.R. McDowell, N. Garapati, T.R. Carr, and B.J. Anderson. (2018). Development of 3-D geological model of Tuscarora Sandstone for feasibility of deep direct-use geothermal at West Virginia University's main campus. *GRC Transactions*, 42. Pp. 192 - 208.
- McDowell, R., compiler. (2018). Summary of petrographic observations, Silurian Tuscarora Sandstone interval - Clay 513 drill core [dataset]. West Virginia Geological and Economic Survey.
- Microsoft Corporation and S. Weston (2017). doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.11. <https://CRAN.R-project.org/package=doParallel>
- Midwest Regional Carbon Sequestration Partnership (MRCSP). URL [www.mrcsp.org](http://www.mrcsp.org). Accessed January, 2015.
- Millard, S.P.. (2013). EnvStats: An R Package for Environmental Statistics. Springer, New York. ISBN 978-1-4614-8455-4, <http://www.springer.com>.
- Moreau, F., D. Gibert, M. Holschneider, and G. Saracco. (1997). Wavelet analysis of potential fields. *Inverse Probl.*, 13. Pp. 165 - 178. <https://doi.org/10.1088/0266-5611/13/1/013>
- Muffler, P., and R. Cataldi. (1978). Methods for regional assessment of geothermal resources. *Geothermics*, 7(2-4). pp. 53 – 89.
- Nathenson, M. (1975). Physical factors determining the fraction of stored energy recoverable from hydrothermal convection systems and conduction-dominated areas. U.S. Geological Survey, Open File Report 75-525, pp. 35.
- Palmer-Wilson, K., J. Banks, W. Walsh, and B. Robertson. (2018). Sedimentary basin geothermal favourability mapping and power generation assessments. *Renewable Energy*, 127. Pp. 1087 - 1100.
- Patchen, D.G., K.L. Avary, and R.B. Erwin. (1985a). Northern Appalachian Region correlation chart.
- Patchen, D.G., K.L. Avary, and R.B. Erwin. (1985b). Southern Appalachian Region correlation chart.
- Pebesma, E.J. and R.S. Bivand. (2005). Classes and methods for spatial data in R. *R News*, 5(2). <https://cran.r-project.org/doc/Rnews/>.
- Pianosi, F., K. Beven, J. Freer, J.W. Hall, J. Rougier, D.B. Stephenson, and T. Wagener. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79. Pp. 214 - 232.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ramalingam, A., and S. Arumugam. (2012). Experimental study on specific heat of hot brine for salt gradient solar pond application. *International Journal of ChemTech Research*, 4(3). Pp. 956 - 961.
- Repetski, J.E., R.T. Ryder, D.J. Weary, A.G. Harris, and M.H. Trippi. (2008). Thermal maturity patterns (CAI and %Ro) in Upper Ordovician and Devonian rocks of the Appalachian Basin: A major revision of USGS map I-917-E using new subsurface collections. USGS Scientific Investigations Map 3006. Reston, Virginia.

- Robertson, E.C.. (1988). Thermal properties of rocks. USGS open file report 88-441. Reston, VA. 110 p.
- Roen, J.B., and B.J. Walker. (1996). The atlas of major Appalachian gas plays. West Virginia Geological and Economic Survey. Publication V-25.
- Roy, R.F., D.D. Blackwell, and F. Birch. (1968). Heat generation of plutonic rocks and continental heat flow provinces. *Earth and Planetary Science Letters*, 5. Pp. 1 - 12.
- Ryder, R.T.. (1995). Appalachian Basin Province (067), In D.L. Gautier, G.L. Dolton, K.I. Takahashi, and K.L. Varnes (eds.). 1995 National assessment of United States oil and gas resources--Results, methodology, and supporting data. U.S. Geological Survey Digital Data Series DDS-30, Release 2. 144 p.
- Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008). Global Sensitivity Analysis: The Primer. John Wiley & Sons.
- Saltelli, A. and S. Tarantola (2002). On the relative importance of input factors in mathematical models. *Journal of the American Statistical Association*, 97(459). Pp. 702–709
- Sanyal, S.K. and Z. Sarmiento. (2005). Booking geothermal energy reserves. *GRC Transactions*, 29. Pp. 467 - 474.
- Sclater, J.G., C. Jaupart, and D. Galson. (1980). The heat flow through oceanic and continental crust and the heat loss of the Earth. *Reviews of Geophysics and Space Physics*, 18(1). Pp. 269 - 311.
- Scherer, W.T., T.A. Pomroy, and D.N. Fuller. (2003). The triangular density to approximate the normal density: decision rules-of-thumb. *Reliability Engineering and System Safety*, 82. Pp. 331 - 341.
- Shank, S., R. Kaiser, M. Sullivan, M. Deemer, K. Stuckert, E. Macklin, and T. Suskie. (2012). Pennsylvania Borehole Temperatures [WWW Document]. URL <http://repository.stategeothermaldata.org/repository/resource/9e15e1a59b768b330d029e86dc0d6512> (accessed 6.22.16).
- Siegel, D.I., R.J. Szustakowski, and S. Frape. (1990). Regional appraisal of brine chemistry in the Albion Group sandstones (Silurian) of New York: Pennsylvania and Ohio. *Association of Petroleum Geochemical Explorationists Bulletin*, 6. Pp. 66 - 77.
- Siégel, C., C.E. Schrank, S.E. Bryan, G.R. Beardsmore, and D.J. Purdy. (2014). Heat-producing crust regulation of subsurface temperatures: A stochastic model re-evaluation of the geothermal potential in southwestern Queensland, Australia. *Geothermics*, 51. Pp. 182 - 200.
- Skeen, J.C.. (2010). Basin analysis and aqueous chemistry of fluids in the Oriskany Sandstone, Appalachian Basin, USA, M.S. Thesis, West Virginia State University, Morgantown, WV.
- Slater, B.. (2012). New York Borehole Temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/9e15e1a59b768b330d029e86dc00481f/> (accessed 6.22.16).
- Smith, J.D. (2015). [Geothermal\_DataAnalysis\_CrossSections]. Online Git Repository. [https://github.com/jds485/Geothermal\\_DataAnalysis\\_CrossSections](https://github.com/jds485/Geothermal_DataAnalysis_CrossSections).
- Smith, J.D.. (2016). Analytical and geostatistical heat flow modeling for geothermal resource reconnaissance applied in the Appalachian Basin. Cornell University, MS Thesis. 254 p.

- Smith, J.D. and F.G. Horowitz. (2017). Thermal model methods and well database organization in GPFA-AB. In: Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. pp. 202–234.
- Smith, J.D., G. Axelsson, and E.R. Camp. (2016). The importance of caprock heating for geothermal heat in place calculations: An Appalachian Basin case study. *GRC Transactions*, 40. Pp. 105-112.
- SMU Geothermal Lab. (2016). Near ground surface temperature map [dataset]. Data and Maps / Temperature Maps. URL <http://www.smu.edu/Dedman/Academics/Programs/GeothermalLab/DataMaps/TemperatureMaps>. Accessed 22 June 2016.
- Stein, A., M. Hoogerwerf, and J. Bouma. (1988). Use of soil-map delineations to improve (co)kriging of point data. *Geoderma*, 43. pp. 163–177.
- Stutz, G.R., E. Shope, G.A. Aguirre, J. Batir, Z. Frone, M. Williams, T.J. Reber, C.A. Whealton, J.D. Smith, M.C. Richards, D.D. Blackwell, J.W. Tester, J.R. Stedinger, and T.E. Jordan. (2015). Geothermal energy characterization in the Appalachian Basin of New York and Pennsylvania. *Geosphere* 11(5). Pp. 1291 – 1304. DOI 10.1130/GES00499.1.
- Tester, J.W., E.M. Drake, M.J. Driscoll, M.W. Golay, and W.A. Peters. (2012). Geothermal Energy. Ch. 11 in Sustainable Energy Choosing Among Options. MIT Press. Pp. 546 – 618.
- Tester, J.W., B.J. Anderson, A.S. Batchelor, D.D. Blackwell, R. DiPippo, E.M. Drake, J. Garnish, B. Livesay, M.C. Moore, K. Nichols, S. Petty, M.N. Toksoz, R.W.J. Veatch, R. Baria, C. Augustine, E. Murphy, P. Negraru, and M.C. Richards. (2006). The Future of Geothermal Energy. The Future of Geothermal Energy.
- United States Geologic Survey (USGS) National Map. (2015). DEM Product Index 1/3rd arc-second. URL <http://viewer.nationalmap.gov/viewer/>. Accessed 22 June 2016.
- Virginia Division of Geology and Mineral Resources (VDGMR). (2011). Virginia borehole temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/b99f8f8e3a7d798d77d4c343bd16b987> (accessed 6.22.16)
- Waller, R.M., J.T. Turk, and R.J. Dingman. (1978). Potential effects of deep-well waste disposal in Western New York: US Geol. Surv. Professional Paper 1053. Pp. 1 - 39.
- Waples, D.W., and J.S. Waples. (2004). A review and evaluation of specific heat capacities of rocks, minerals, and subsurface fluids. Part 1: Minerals and nonporous rocks. *Natural Resources Research*, 13(2). Pp. 97 - 122.
- West Virginia Geological & Economic Survey (WVGES). (2019). WVGES Pipeline. URL <http://www.wvgs.wvnet.edu/oginfo/pipeline/pipeline2.asp> (accessed 3.20.19).
- West Virginia Geological & Economic Survey (WVGES). (2011). West Virginia Borehole Temperatures [WWW Document]. AASG Geotherm. Data Repos. URL <http://repository.stategeothermaldata.org/repository/resource/168566464e3d5f8f3cde3b9fc0052329> (accessed 6.22.16).
- West Virginia Geological and Economic Survey (WVGES). (2006). Trenton Black River Project. Data [Precambrian structure contours] URL <http://www.wvgs.wvnet.edu/www/tbr/resources.asp>. Map viewer available online [http://ims.wvgs.wvnet.edu/TBR\\_ver3/viewer.htm](http://ims.wvgs.wvnet.edu/TBR_ver3/viewer.htm).

- Whealton, C.A.. (2016). Statistical data analysis, global sensitivity analysis, and uncertainty propagation applied to evaluating geothermal energy in the Appalachian Basin. Cornell University, Ph.D. Thesis. 267 p.
- Whealton, C.A. and J.D. Smith (2015). [geothermal\_pfa] code repository. Available Online [https://github.com/calvinwhealton/geothermal\\_pfa](https://github.com/calvinwhealton/geothermal_pfa). Accessed 22 June 2016.
- Whealton, C.A., J.R. Stedinger, and F.G. Horowitz. (2015). Methodology Memo 2: Application of generalized least squares regression in bottom-hole temperature corrections. In: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin: Final Phase 1 Research Report.
- Williams, C.F.. (2014). Evaluating the volume method in the assessment of identified geothermal resources. *GRC Transactions*, 38. Pp. 967 - 974.
- Williams, C.F., and J. DeAngelo. (2011). Evaluation of approaches and associated uncertainties in the estimation of temperatures in the upper crust of the western United States. *GRC Transactions*, 35. Pp. 1599 - 1606.
- Williams, C.F., M.J. Reed, and R.H. Mariner. (2008). A review of methods applied by the U.S. Geological Survey in the assessment of identified geothermal resources. U.S. Geological Survey Open-File Report 2008-1296. 27 p.
- Zafar, S.D., and B.L. Cutright. (2014). Texas' geothermal resource base: A raster-integration method for estimating in-place geothermal-energy resources using ArcGIS. *Geothermics*, 50. pp. 148 - 154.

## CHAPTER 4

### UNCERTAINTY ANALYSIS OF GEOTHERMAL RESERVOIR SIMULATIONS FOR

### GEOTHERMAL GREENFIELD FEASIBILITY ASSESSMENTS:

### A CASE STUDY OF DIRECT-USE HEATING FOR CORNELL UNIVERSITY

#### **4.0 Abstract**

Geothermal reservoir simulations use analytical or numerical models with specified geologic reservoir properties and temperatures to simulate reservoir heat extraction. Most models are deterministic. Considering uncertainties in geologic reservoir properties and temperatures, and propagating those uncertainties using stochastic reservoir simulations, allows for probabilistic interpretations of a geothermal reservoir's ability to meet heating demands. This is particularly useful in the planning phase of a geothermal greenfield project, when site-specific data are limited and uncertainty of parameters is large because data are sparse. In this paper, uncertainties in geologic properties and temperature are propagated through simple analytical reservoir models using Monte Carlo analysis to estimate the temperature and heat produced for several production scenarios. Analytical geothermal reservoir models approximate possible flow geometries. These include: 1) porous media heat sweep, and 2) parallel fractures for fracture-dominated geothermal systems. The resultant estimated probabilities of meeting direct-use heating objectives are used to select production scenarios to model with numerical thermal-hydraulic simulations using the TOUGH2 software. Numerical models consider reservoir fluid flow complexities that are not included in our analytical models. We apply these analytical and numerical modeling methods to estimate the probability of meeting direct-use heating demands of Cornell University's main campus in New York State. We evaluate a potential injection-extraction well pair drilled into a Trenton-Black River dolomitic reservoir at about 2.3 km depth

below campus, or in crystalline basement rocks beginning at 3 km depth. This case study demonstrates use of probability of meeting target heating objectives to inform if the economics of drilling to the deeper basement target could be worthwhile for Cornell. For Cornell, sensitivity of geothermal reservoir energy extraction to the uncertainty in the initial rock temperature is important for the shallower target reservoir, and uncertainty in fracture spacing is important for the fracture-dominated basement target to determine the time to thermal breakthrough.

#### **4.1 Introduction**

Stochastic analysis of geothermal reservoir models allows for probabilistic feasibility studies that can inform techno-economic evaluations, risk assessments, and decisions regarding energy utilization projects (e.g. Daniilidis et al., 2017; Olasolo et al., 2016; Sener and van Dorp, 2005). Reservoir models may range in complexity from simple conceptual reservoir geometries that are postulated before drilling at a site, to detailed knowledge of producing reservoirs, for which site-specific data are available to calibrate the initial conditions and geologic properties, and to constrain the flow geometries (e.g. Held et al., 2014; Watanabe et al., 2010). This study focuses on uncertainty analysis for reservoir modeling in the planning phase of a geothermal greenfield project (e.g. Hadi et al., 2010), at which time limited site-specific information is available to characterize reservoir properties.

In this phase, uncertainty in geologic properties, hydrogeology, and temperature data can be large. For example, using regionally available bottom-hole temperature (BHT) data and geologic properties in an uncertainty analysis of the temperatures at 3 km depth in the Appalachian Basin, Smith (2019, Ch. 3) found uncertainty ranges of 20 °C in areas with high BHT data density to 60 °C in areas with sparse BHT data. This present paper describes an

uncertainty propagation from geothermal resource characterization completed at the scale of a basin to decision-relevant information (e.g. Witter et al., 2019) on the scale of a project. Using stochastic reservoir simulations, we estimate the probability that a potential geothermal reservoir can meet target temperature and heating objectives for a project site. These analyses are valuable to decide if the large capital investment of the first well-pair at a site is worth the financial risk, or if additional information should be acquired to inform the decision.

The purpose of this paper is to provide uncertainty analysis methodologies that are suitable for pre-drilling decision analysis of geothermal reservoir feasibility for both porous media and fracture-dominated geothermal systems. Uncertainty in reservoir temperature and key geologic properties, such as fracture spacing, are propagated through simple analytical thermal-hydraulic models of geothermal reservoirs using Monte Carlo simulations. Thermal-hydraulic models are used to estimate the produced temperature and heat over time for an injection-extraction well pair. For each reservoir model, we consider several production scenarios consisting of different injection temperatures and fluid flow rates. In many geothermal reservoir modeling studies these parameters have been found to be among the engineering design parameters to which the thermal energy produced from a reservoir is most sensitive (Asai et al., 2018; Aliyu and Chen, 2017; Zeng et al., 2013).

Using the set of fluid flow rates and Monte Carlo-generated extraction well temperatures over time for each production scenario, we compute the probability of meeting direct-use heating objectives over the proposed project lifetime. We consider the objectives to be met when at least 95% of the Monte Carlo replicates for a production scenario achieve the objectives; other decision makers may select different proportions that reflect their risk tolerances. These

probabilities reflect uncertainty in the available site data, and may be used to inform project investment and modeling decisions (e.g. for economic assessment, Goumas et al., 1999).

We use the analytical reservoir modeling results as screening tools to select production scenarios to evaluate further with more computationally expensive numerical model simulations. We do not intend for the initial analytical model results to exactly match the numerical model results, as in surrogate or reduced order modeling (e.g. Asai et al., 2018; Oakley, 2004). Rather, we use the probability of meeting target objectives to refine the selection of production scenarios over the planning period. This is a useful exercise to reduce the need for numerical model simulations. Numerical simulations in this paper consider complexities that are not considered in our analytical models, including anisotropic permeability within formations and permeability contrasts between formations. Due to computational demands, uncertainty in numerical simulations is only evaluated using variability in rock temperatures. In many geothermal reservoir modeling studies, the rock temperature has been found to be among the naturally-occurring geologic model variables to which the thermal energy produced from a reservoir is most sensitive (e.g. Cheng et al., 2016). It is beyond the scope of this paper to present detailed sensitivity analyses for all of the parameters in and boundary conditions for reservoir models that we implement; such analyses are available in other case studies for analogous models (e.g. for low-temperature settings, Bedre and Anderson, 2012), which generally agree on the variables to which thermal energy production is most sensitive.

#### **4.1.1 Case Study Site Description**

We apply the uncertainty analysis methods presented in this paper to a proposed geothermal project for hot water district heating of the Cornell University main campus in Ithaca,

New York. The campus is located in the Appalachian Basin, which is classified as a low-temperature geothermal setting (e.g. Blackwell et al., 2011). Several geothermal resource assessments of the Appalachian Basin have estimated a region of elevated heat flow in south-central New York State (Smith, 2016; Stutz et al., 2015; Frone and Blackwell, 2010), and in 2016, a geothermal play-fairway analysis of the Appalachian Basin placed the Cornell campus proximal to a “high-priority” geothermal play (Cornell University, 2017). The campus facilities are currently heated by a natural gas-supplied steam district heating system. This system has an annual baseload heating demand of about 12 MW<sub>th</sub>, with peak demands in the winter that exceed 80 MW<sub>th</sub> (e.g. Gustafson et al., 2018, Fig. 3). As part of an ongoing U.S. Department of Energy-funded feasibility study, deep (> 1 km) geothermal energy for the campus is being evaluated to initially meet a minimum of 20% of the total annual heating demand using a pilot injection-extraction well pair. This corresponds to about 5 MW<sub>th</sub> continuous supply, considering system efficiency losses (Gustafson et al., 2019). Given the projected demands for the campus and planned renovations to campus buildings and the heating system to improve efficiency, it is expected that the thermal energy representing 20% of the annual demand will not change significantly over the useful lifetime of geothermal reservoirs (Gustafson et al., 2019).

This present paper evaluates the feasibility of using an injection-extraction well pair within potential geothermal reservoir target formations in sedimentary and basement rocks to meet the initial goal of 20% of the annual heating demands for the Cornell campus. The sedimentary target is located at about 2.3 km depth within the Trenton-Black River (TBR) carbonate group, which regionally contains Darcy-scale permeability in a hydrothermally altered dolomite (Camp and Jordan, 2017). For sedimentary rocks, regional studies and oil and gas well data inform the characterization of uncertainty in values selected for key geologic properties. The

deeper target is Precambrian crystalline basement rock starting at 3 km depth, for which limited information about hydrogeologic and thermal properties is available in the region surrounding Cornell. For basement rocks, we evaluate uncertainty in geologic properties and fracture spacings based on studies of crystalline rock in the Adirondack Mountains, which are located nearly 170 km northeast of Cornell. We compare the heat extraction for these two reservoir targets, which have differing degrees of uncertainty in both the temperatures at their depth and their geologic property values. Given that an injection-extraction well pair in deeper and harder basement rocks is expected to be nearly double or triple the cost of drilling compared to wells in the shallower and softer sedimentary rocks considered in this paper (Lowry et al., 2017; Lukawski et al., 2016), an important decision evaluated in this case study is whether or not drilling to basement rocks would be worthwhile to meet the geothermal system feasibility requirements for Cornell district heating.

Feasibility is defined in this paper as meeting two requirements over the operational lifetime of the geothermal direct-use system: 1) producing fluid at a temperature of at least 60 °C, and 2) producing thermal energy of at least 5 MW<sub>th</sub>. A geothermal system that meets these requirements would jointly achieve the minimum building supply temperature for the Cornell campus, and meet the initial goal of 20% of the annual heating demand. Production of higher temperature fluid could be utilized in a thermal cascade (e.g. Gudmundsson et al., 1985) from higher and standard temperature facilities (80 °C and 70 °C minimum supply, respectively) to lower temperature facility needs on the campus, including animal housing, pasteurization, aquaculture, biomass drying, greenhouse warming, and snow melting (Gustafson et al., 2019). Heat extraction greater than 20% of the annual demand could be utilized to store thermal energy in hot water storage tanks and to meet additional space heating demands.

Modeling the utilization of thermal energy within the campus district heating system is outside the scope of this paper. As a result of planned modifications to Cornell’s district heating system to increase flexibility by using heat pumps to meet the campus building supply temperatures and to control the temperature of reinjected geothermal fluid, as described in Gustafson et al. (2019), we assume that all thermal energy produced by a geothermal reservoir will be fully utilized by the campus. As a result of not modeling campus infrastructure, we also do not consider economic assessment of the integrated reservoir-and-surface-use system in the scope of this paper. However, the results from the uncertainty analysis of heat extraction from potential geothermal reservoirs, presented here, may be used to inform economic assessments.

#### **4.1.2 Paper Organization**

Sections 4.2 and 4.3 below detail the geothermal resource estimates, geologic column, and potential reservoir targets below the Cornell campus. Section 4.4 provides details on the analytical and numerical reservoir models used, and their assumptions. Section 4.5 describes the reservoir production scenarios used in geothermal reservoir simulations, and Section 4.6 presents the results of these production scenarios. Conclusions use the analysis to discuss the feasibility of potential reservoirs to meet the heating demands of the Cornell campus, and to describe analysis limitations that should be addressed in future work.

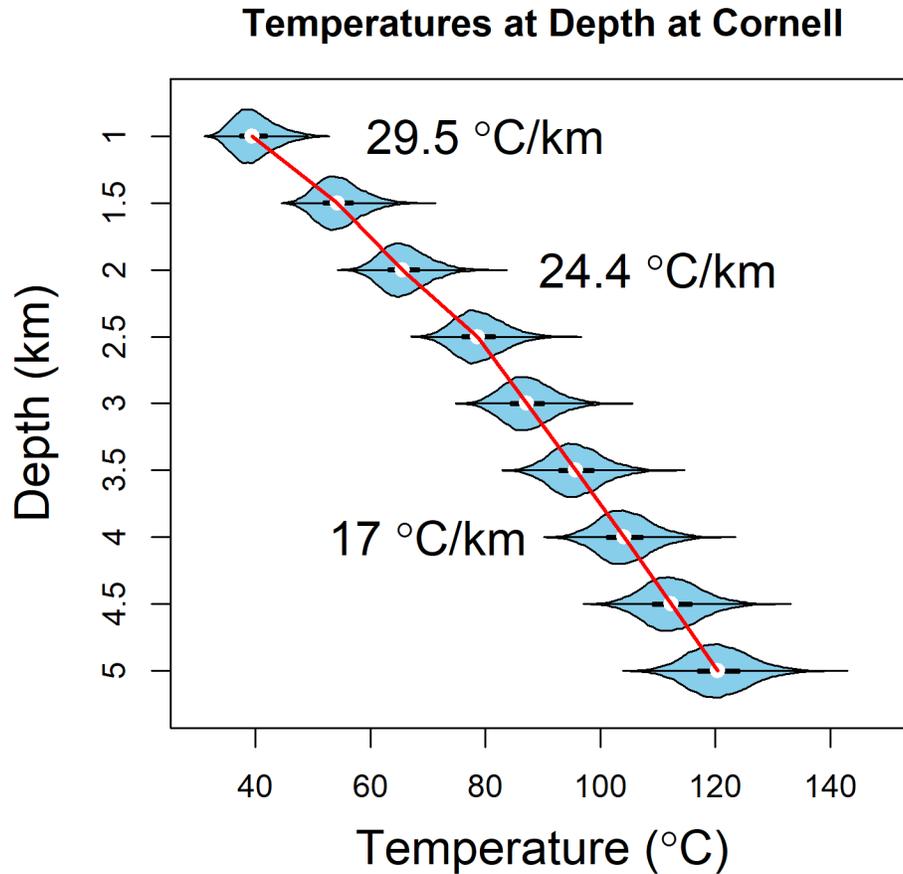
#### **4.2 Geothermal Resource Below the Cornell Campus**

In south-central New York State near Cornell University, temperatures suitable for district heating have been estimated at drilling depths that have been reached by oil and gas wells in the region (Smith, 2019, Ch. 3; Stutz et al., 2015). The Cornell campus does not currently

have a deep well, so the temperatures at target depths for geothermal reservoirs were gathered from estimates made in Smith (2019, Ch. 3). Smith (2019, Ch. 3) estimated temperature-depth profiles across the Appalachian Basin using a Monte Carlo analysis of uncertainty in a spatially correlated surface heat flow map (Smith, 2016; Cornell University, 2017) and uncertainty in geologic properties in a one-dimensional vertical heat conduction model (Horowitz, Smith and Whealton, 2015). The assumption of conduction-dominated heat transfer at depths greater than 1 km in south-central New York is supported by well temperature logs (Section 3 of Appendix D) and evaluations of well bottom-hole temperature data (Smith, 2019, Ch. 2).

The Smith (2019, Ch. 3) temperature-depth profiles estimated for the Cornell campus are presented in Figure 1 for a 1 km<sup>2</sup> area that contains one of the proposed geothermal sites under consideration. Violin plots (Adler, 2005) are used to display uncertainty at each 0.5 km depth increment. A change in geothermal gradient is estimated for basement rocks starting at about 2.8 km depth due to an assumed change in thermal conductivity and heat generation relative to sedimentary rocks. Based on Figure 4.1, the Cornell campus target minimum building supply temperature of 60 °C could be achieved with near certainty within basement rocks and within Trenton-Black River (TBR) rocks around 2.3 km depth; however, the difference in temperatures from the 95<sup>th</sup> quantile to 5<sup>th</sup> quantile is striking at about 13°C to 15 °C for the reservoir depths of interest (Figure 4.1, Table 4.1). Given that rock temperature is among the naturally-occurring geologic variables to which geothermal system performance is most sensitive (e.g. Akar and Young, 2015; Bedre and Anderson, 2012), the uncertainties in potential reservoir temperatures for Cornell motivate the estimation of useful operational lifetimes that could be achieved for each reservoir target. Figure 4.1 also demonstrates that basement rocks are more likely to meet

the minimum building supply temperatures for standard (70 °C) and high-temperature (80 °C) facilities compared to the TBR sedimentary target.



**Figure 4.1:** Predicted temperature-depth profile below Cornell University in Ithaca, NY. The uncertainty distribution of predicted temperatures for each 0.5 km depth interval is shown as a blue violin plot (kernel density plot). White circles are placed at the median predicted temperature at each depth, and a narrow black box in the center spans the 25<sup>th</sup> to the 75<sup>th</sup> percentile estimates. Red line segments connect the predicted median temperature values for each depth. The thermal gradient corresponding to those lines is printed adjacent to the lines.

For the analytical reservoir modeling software used in this project (Section 4.4.1), temperature-depth profiles are specified using a geothermal gradient that is piecewise linear with up to four segments over selected depth intervals (Beckers et al., 2018; Beckers, 2016). We used three geothermal gradient segments, as shown in Figure 4.1 and tabulated in Table 4.1. Table 4.1

provides estimates of the 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles of temperatures at selected depths below Cornell from Smith (2019, Ch. 3). Estimates of the geothermal gradient percentiles in Table 4.1 were computed by assuming that the estimated temperature quantiles at each depth match to those same quantiles at all other depths. These gradients were used to define initial and boundary conditions in reservoir simulations.

**Table 4.1:** Estimates of the 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles of temperatures at selected depths below Cornell, and corresponding geothermal gradients.

Depth (km)	Temperature (°C) 5 <sup>th</sup> Percentile	Temperature (°C) 50 <sup>th</sup> Percentile	Temperature (°C) 95 <sup>th</sup> Percentile	Modeled Gradient (°C/km) 5 <sup>th</sup> Percentile	Modeled Gradient (°C/km) 50 <sup>th</sup> Percentile	Modeled Gradient (°C/km) 95 <sup>th</sup> Percentile
0	8	10	12	26.6	29.5	33.7
1	35	39	46			
1.5	49	54	61			
2	60	66	73	23.7	24.4	24.9
2.5	73	79	86			
2.8 Basement	78	84	92	16.5	17	17.4
3	81	87	95			
3.5	89	96	104			
4	97	104	113			

### *Heat Flow at Cornell*

For numerical modeling of a conduction-dominated geothermal system, a necessary boundary condition is the heat flow upwards into the bottom of the simulation grid. We estimated the heat flow at depth below Cornell by projecting the Smith (2016; available in Cornell University, 2016a) predicted mean surface heat flow of 49 mW/m<sup>2</sup> to the 3 km simulation grid depth using a one-dimensional conduction heat balance. Heat generation is assumed to be constant at an expected value of 1 μW/m<sup>3</sup> in sedimentary rocks, and the concentration of heat producing elements (potassium, uranium, thorium) is modeled as an exponential decrease with increasing depth in basement rocks (e.g. Lachenbruch, 1970), as used in Smith (2016; 2019, Ch. 3) to estimate surface heat flow and temperatures at depth. Under these assumptions, the equation for heat flow as a function of depth is provided in Equation 4.1

$$Q(Z) = Q_s - A_s Z_s - A_b B \left( 1 - e^{-\frac{(Z-Z_s)}{B}} \right) \quad [4.1]$$

where  $Q(Z)$  is the heat flow as a function of depth,  $Z$ ,  $Q_s$  is the surface heat flow,  $A_s$  is the value of heat generation within sedimentary rocks with thickness  $Z_s$ ,  $A_b$  is the value of heat generation within rocks at the top of the basement, and  $B$  is the basement rock thickness corresponding to a one log-decrease in heat generation. The average values of  $Z_s$ ,  $A_b$ , and  $B$  used for the Cornell site were 2.8 km, 1.73 μW/m<sup>3</sup>, and 10 km, respectively. Using these values, the expected heat flow at 3 km depth was estimated to be about 46 mW/m<sup>2</sup>.

### 4.3 Geologic Column Below Cornell

Numerical modeling of a geothermal reservoir requires specification of the geologic formations and their properties for reservoir rocks through which fluid must flow, and the caprocks and base rocks that supply conductive heat recharge to the reservoir. We adopted a simplified geologic column below the Cornell site that was informed by local well logs (Section 2 of Appendix D [Table D.1 and Figure D.1] contains the wells used).

For sedimentary formations, porosity and density data were gathered from six well logs, five of which are within 40 km of Cornell (Figure 4.2; Appendix D Table D.1). These well logs were selected because gamma ray, neutron porosity, and density digital logs were available to at least the depth of the TBR target sedimentary formation; some logs reached basement rocks. This suite of log types was needed to correct neutron-measured porosity for the effects of shale and hydrocarbons (Bassiouni, 1994, Ch. 15 & 16; Appendix D Section 4, Table D.3-Table D.21, Figure D.3-Figure D.6). Changes in the mean and variability of density and porosity in the well logs were used to select groups of vertically adjacent geologic formations with similar properties, for which uncertainty distributions for density and porosity were developed. Summary statistics of density and porosity obtained from each of these wells for each sedimentary formation group are provided in Appendix D.

For reservoir simulations, the most likely values of density and porosity for each formation were selected to be within the interquartile ranges of the well log measurements. The average of the median values obtained from each well log was selected as the most likely value when distributions were highly skewed. The average across all wells was used because sufficient information with which to characterize the observed spatial heterogeneities in the distributions of geologic properties was not readily available on which to base predictions for the Cornell site.

For formations with few well log measurements, values published in studies conducted within New York were used to inform the selection of values in this study (Section 1 of Appendix D). Section 4.3.2.1 presents a detailed discussion of the well log effective porosity distributions for the target TBR dolomite reservoir.

Geologic property values that were not estimated from well log data were gathered from regional studies, when available, or studies of rock of comparable lithology at the thermodynamic conditions expected for each formation. Thermal conductivity values were assigned using the formation mean  $\pm$  two standard deviations from the estimates of Appalachian Basin thermal conductivities made as part of a Geothermal Play Fairway Analysis project (data available on the geothermal data repository, Cornell University [2016]). These conductivity values were primarily from Carter et al. (1998) for rock samples in the Anadarko Basin, which has a similarly deep burial history as the Appalachian Basin (e.g. Rowan, 2006; Johnsson, 1986; Carter et al., 1998). Heat capacity was estimated from Robertson and Hemingway (1995) and Waples and Waples (2004) by lithology at the estimated formation temperatures. Justifications for geologic property values selected for each formation group are provided in Appendix D Section 1.

Table 4.2 provides the estimated stratigraphic column below Cornell and the selected values of geologic properties for each formation group for use in numerical geothermal reservoir simulations. The vertical grid cell discretization for each formation group is also provided in Table 4.2. Finer grid resolution was used near the target TBR formation. Table 4.3 (Section 4.4.4) provides analogous values for analytical models.

**Table 4.2:** Generalized geologic column below Cornell with estimated formation depths, geologic properties, and grid cell sizes used in numerical geothermal reservoir simulations.

Formation Name	Modeled Formation Top Depth (m)	Porosity (-)	Permeability H: horizontal V: vertical (mD)	Density (kg/m <sup>3</sup> )	Thermal Conductivity (W/m-K)	Specific Heat Capacity (J/kg-K)	No. of Vertical Grid Cells: Cell Size (m)	Sources and Notes
Lorraine / Utica Shale	1860	0.04	H: 5E-6 V: 5E-6	2700	0.9	830	Boundary Condition 1: 0.1 m 1: 199.9 m	Permeability: Carter and Soeder (2015) Heat Capacity: Waples and Waples (2004)
Trenton Limestone	2060	0.02	H: 5 V: 0.005	2690	2.11	870	1: 105 m 5: 10.5 m 10: 3.15 m 10: 2.1 m	Permeability anisotropy based on Camp and Jordan (2017)
Black River Dolomite	2270	0.07	H: 250 V: 2.6	2800	2.91	930	15: 2 m	Vertical permeability from Camp and Jordan (2017)
Black River Limestone	2300	0.01	H: 0.5 V: 0.0005	2700	2.11	880	20: 2 m	Permeability anisotropy based on Camp and Jordan (2017)
Upper Beekmantown Group: Tribes Hill / Little Falls Carbonates	2340	0.02	H: 2.6 V: 2.6	2780	3.79	880	5: 11 m 3: 18.3 m 2: 55 m	Permeability: Camp (2017)
Galway / Theresa Carbonates / Rose Run Sandstone	2560	0.01	H: 2.6 V: 2.6	2610	3.34	880	1: 220 m	Porosity and Permeability: Smith et al. (2005; 2010), Camp (2017)
Potsdam Sandstone	2780	0.01	H: 0.002 V: 0.0002	2640	4.27	860	1: 20 m	Porosity and Permeability: Kolkas and Friedman (2007), Waller et al. (1978) Heat Capacity: Abdulagatov et al. (2014)
Precambrian Basement: Granitic Gneiss	2800	0.01	H: 0.001 V: 0.001	2730	2.83	825	1: 199.9 m Boundary Condition 1: 0.1 m	Porosity and Permeability: Selvadurai et al. (2005) Density: Simmons (1964, Table 1) Thermal Conductivity: Southern Methodist University divided bar measurements on 2 samples of Adirondack granite gneiss.

Sections 4.3.1 through 4.3.3 provide further details about the geologic properties used for modeling the expected caprocks at about 1.86 km depth, the Trenton-Black River reservoir target at about 2.27 - 2.3 km depth, and the crystalline basement beginning at about 2.8 km depth. Depths to formations were estimated from interpolations of formation tops selected on local wells (T.E. Jordan, personal communication, 2018; J.A. Al Aswad, personal communication, 2018). The locations of wells used, their selected depths to formation tops, and further details for the selected geologic properties below Cornell are provided in Appendix D Section 1.

#### **4.3.1 Caprocks: Lorraine / Utica Shale**

The Lorraine and Utica shale formations overlie the Trenton-Black River (TBR) reservoir target formation. Based on wells in Ohio (Carter and Soeder, 2015), the Utica shale is expected to have nanodarcy-scale permeability. Based on local well logs (Appendix D), the total thickness of these units below Cornell is expected to be about 200 m. Given these properties, we expect these formations to be caprocks to the TBR reservoir, and we do not model shallower formations in geothermal reservoir simulations.

#### **4.3.2 Reservoir Target: Trenton-Black River**

In south-central New York State near Cornell University, the contact of the Trenton and Black River formations contains dolomite where sub-vertical wrench faults extend upwards from basement rocks through the Black River limestones, and terminate at the upper contact of the Trenton, or within the immediately overlying Utica Shale (Smith, 2006; Camp and Jordan, 2017). A porous dolomite results from hydrothermal alteration of limestone. Camp et al. (2018) estimate that this dolomitized rock is on average orders of magnitude more permeable than the

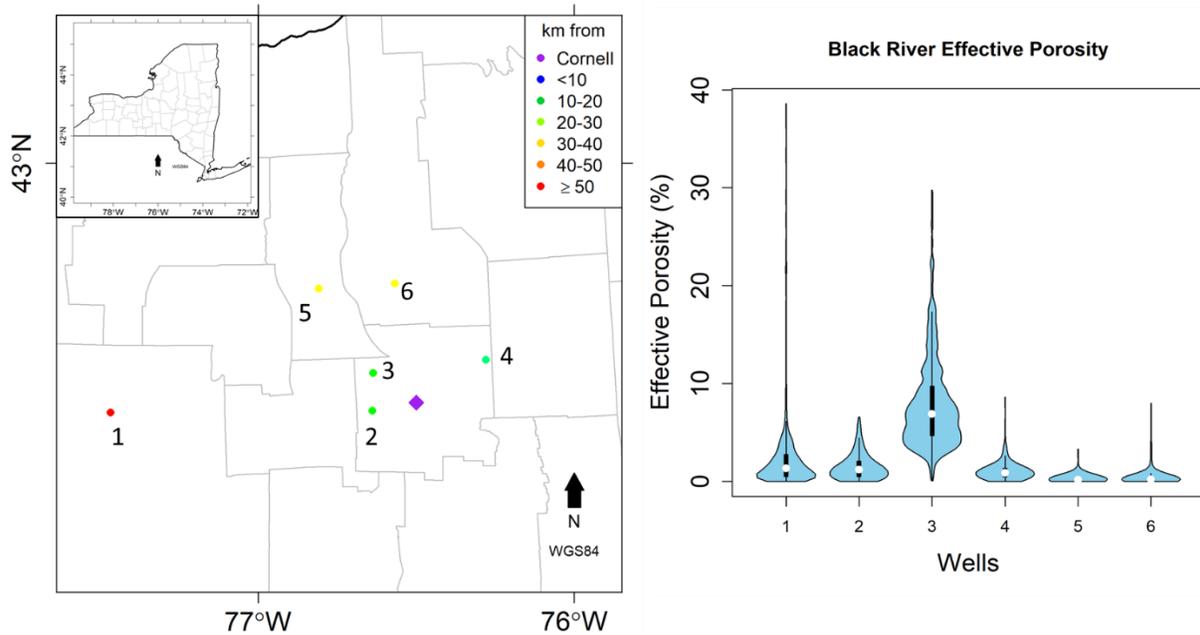
limestone, and the dolomite would be favorable for geothermal reservoir production based on the likely reservoir productivity (inverse of flow impedance).

The existence of Black River dolomite below Cornell depends on the presence of wrench faults, which have no surface expression. Records in the New York Empire State Organized Geologic Information System (ESOGIS) database of locations of permitted TBR exploration boreholes in south-central New York suggest that wrench faults are widely distributed although widely spaced. This implies the possibility that TBR-type reservoirs may exist below the footprint of Cornell's campus. However, the presence of high permeability dolomite in any location is unknown until a well is drilled. Based on regional gas production fields (Camp and Jordan, 2017; Cornell University, 2016b), if high permeability Black River dolomite exists below Cornell, the permeable thickness is expected to be about 30 m, and the spatial coverage area containing permeable zones is expected to have an elliptical shape with a minor axis about 1 - 3 km long. It is not known if the entire rock volume is likely to be permeable, or if permeability would be concentrated in zones, e.g., near the wrench faults (Camp and Jordan, 2017). Our analysis considers two conceptual representations of the geothermal reservoir that reflect these possible flow geometries, as described in Section 4.3.2.3.

#### **4.3.2.1 Trenton-Black River Porosity**

Uncertainty distributions of the corrected well log effective porosity for each of six wells that intersected the Black River formation are provided in Figure 4.2. These distributions include both limestone and dolomite where both rock types are present. Porosity in the matrix limestone is generally less than 2%, whereas dolomite porosity can be greater than 10%. The observed spatial heterogeneity in porosity may result from varying amounts of dolomitization and

categories of dolomite (Smith, 2006). Given the available data, we assumed that the observed distribution from well 3 in Figure 4.2, (Duddleston, Appendix D Table D.10), is representative of a higher productivity dolomite. The median porosity for this well of 6.9% is similar to the average porosity of 7% for productive TBR reservoirs in the region (Camp et al., 2018; data available in Cornell University, 2016b).



**Figure 4.2:** Left: Location of Cornell (purple diamond) and wells with digital logs that intersect the Black River formation. Right: Violin plots of effective porosity (well log neutron porosity observations corrected for gas and shale) in the Black River formation for wells on the map.

#### 4.3.2.2 Trenton-Black River Permeability

We sought a porosity-permeability regression relationship (e.g. Jennings and Lucia, 2003) to estimate permeability for Trenton-Black River rocks using the corrected porosity data presented in Section 4.3.2.1. Camp and Jordan (2017) report 23 porosity and permeability measurements taken on cores from Trenton-Black River rocks from about 50 km southwest of Cornell. The samples contain vugs and fractures. We assume that these samples are representative of a higher productivity dolomite. Two permeability measurements were taken for

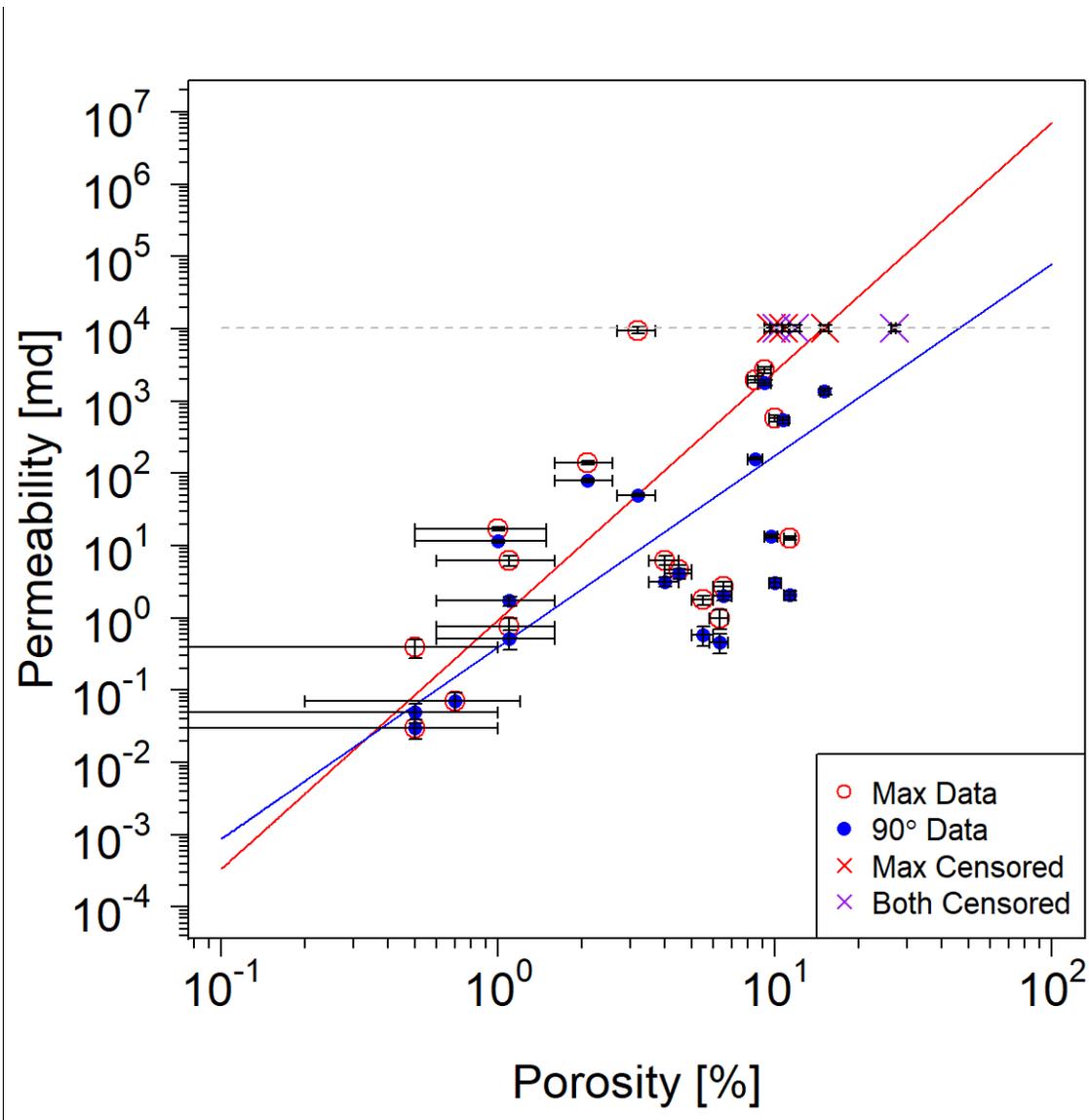
each rock sample: 1) in the direction of maximum permeability, as determined by rotating each sample about the vertical axis while measuring the flow, and 2) 90° from the maximum permeability (e.g. Lucia, 2007). This dataset had several censored observations that resulted from an upper tail detection limit of 10,240 millidarcy (mD) for permeability measurements (i.e. instruments cannot yield larger values). As a result, we used censored regression (e.g. Liu et al., 1997; Tobin, 1958) to estimate the mean permeability for a given porosity. The resulting regression equations are provided in Equation 4.2

$$\ln(\kappa_{max}) = -0.0742 + 3.44 \ln(\phi)$$

$$\ln(\kappa_{90}) = -0.936 + 2.65 \ln(\phi)$$

[4.2]

where  $\kappa_{max}$  is the maximum permeability (mD),  $\kappa_{90}$  is the permeability at 90° to the maximum (mD), and  $\phi$  is the porosity in percent and thus  $0 \leq \phi \leq 100$ . We assume that these regression relationships are applicable for TBR rocks thought to be located below Cornell (parameter significance levels are provided in Appendix D Section 6). The data from Camp and Jordan (2017) and the Equation 4.2 regression lines are provided in Figure 4.3. The expected value of permeability for large porosities is notably larger than the equations presented in Camp and Jordan (2017) as a result of using censored regression to appropriately handle censored data. For a porosity of 7%, the expected permeability ranges from about 70 mD using 90-degree permeability regression to 750 mD using maximum permeability regression. We selected 250 mD for modeling.



**Figure 4.3:** Tobit censored regression for maximum (red) and 90°-to-maximum (blue) horizontal permeability from the Whiteman #1 core (New York State Museum, 2014) of a Trenton-Black River reservoir. The measurement device upper limit of 10,240 millidarcy (mD) is shown as a dashed line. Censored measurements are plotted with x symbols, and uncensored measurements use circles. Error bars are provided for uncensored data.

#### 4.3.2.3 Trenton-Black River Reservoir Production Scenarios

We contrast two conceptual reservoir models that reflect possible flow geometries for a hypothetical Black River dolomite reservoir below Cornell (e.g. Fig. 3 in Camp and Jordan, 2017). The porous media heat sweep model described in Section 4.4.1.1 is used to provide an

upper bound on the reservoir lifetime. The parallel vertical fractures model (Gringarten et al., 1975) described in Section 4.4.1.2 is used to approximate a fracture-dominated enhanced geothermal system (EGS) in which matrix permeability is negligible compared to flow through fractures. If high permeability zones do not exist in rock below Cornell, an EGS could be attempted in TBR rocks using reservoir stimulation techniques (e.g. Breede et al., 2013). Vertical or subvertical fractures are expected from such stimulation of a reservoir because fractures propagate in the direction normal to the plane of least principle stress (Fox et al., 2015). Analytical models assume a 1 km x 1 km extent of the Trenton-Black River permeable zone because wrench fault systems are estimated to be about 500 – 1000 m across (Camp and Jordan, 2017). Fracture spacings of 20 m are assumed in models to limit thermal interaction between the fractures over the simulated lifetime of the geothermal system.

Each of the porous media and fracture-flow analytical models assumes that only the reservoir rocks will be amenable to fluid flow within the simulated rock volume. Numerical modeling of porous media as described in Section 4.4.2 is used to evaluate TBR reservoir production considering heterogeneities in rock permeability between formations, and the possibility of a larger heat sweep area that results from permeable zones extending over a 4 km<sup>2</sup> area. We select 4 km<sup>2</sup> because 2 km is approximately the mean value of the minor axis of the elliptical shape of TBR reservoirs in the region (data available in Cornell University, 2016b). We assume no flow boundary conditions to simulate essentially impermeable rock outside of the 4 km<sup>2</sup> area. The numerical model is expected to provide an upper bound on reservoir heat production.

### 4.3.3 Reservoir Target: Precambrian Crystalline Basement

Knowledge of the values for geologic properties of basement rocks below Cornell is limited. Based on Lidiak and Hinze (1993), basement rocks may reside within the Central Granulite Belt (CGB) of the Grenville basement province (e.g. McLelland, Selleck, and Bickford, 2010), which crops out in the Adirondack Mountains about 170 km northeast of Cornell. The compositions of borehole cuttings for five wells in the Finger Lakes region surrounding Cornell are not inconsistent with this interpretation (Valentino, 2016; the samples are archived by the New York State Museum). As a result, we expect that the composition of basement rocks could consist of a wide variety of petrological groups (metanorthosite and anorthosite gneiss; metasedimentary; granitic, charnockitic, mangeritic, and syenitic gneiss; biotite and/or hornblende granite gneiss; mangerite, pyroxene(hornblende) syenite gneiss; etc.), as observed in the Adirondack Mountains. Based on core samples evaluated by Valentino (2016), and outcrop studies and maps in the Adirondack Mountains (e.g. McLelland et al., 2013), the composition of basement rock changes over horizontal distances on the order of meters to kilometers. This implies that a borehole at Cornell will traverse several basement lithologies, especially if directional drilling is used. Owing to this heterogeneity, we assume geologic properties that are representative of Adirondack Mountain rocks in aggregate from Simmons (1964). These aggregate properties are similar to granitic gneiss, the most common lithology.

Of interest for reservoir modeling is the matrix and fracture permeability of these middle to high-grade metamorphic rocks. Manning and Ingebritsen (1999) compiled permeability measurements for metamorphic rocks with varying composition from several sites, and estimated that conduction-dominated heat transfer occurs for permeability values  $< 100 \mu\text{D}$ . Typical values of matrix permeability for rocks at the depths of interest in this study range between 1 - 100  $\mu\text{D}$ ,

which is similar to measurements taken on granite in Selvadurai et al. (2005). Therefore, aside from fracture flow, these rocks are expected to be practically impermeable for the purpose of heat transfer modeling.

#### **4.3.3.1 Basement Rock Reservoir Modeling: Fracture Properties**

We estimated fracture apertures, spacings, and lengths from Adirondack basement rocks, and treated them as analogues for Cornell basement rocks. Whether or not these properties of the Adirondack rocks prove to be analogous to those at more than 3 km depth below Cornell cannot be tested with existing data.

Outcrop-scale samples of fractures were obtained from a September, 2017 field trip to the Adirondack Mountains for a Cornell University Earth Source Heat geothermal project. Had the examined fractures formed during exhumation of the mountains in the recent geological past, a stage that has no analogy to basement rocks below Cornell, then the fractures would likely not contain mineral seals (Valentino et al., 2016). Among the observed fractures, many were mineral-sealed and therefore are of interest to our analogy.

Outcrop fracture apertures ranged from 0.1 mm to as much as 2 cm. For reservoir modeling, we found this variability in aperture size has a negligible impact on the energy extracted using a parallel fractures model for which aperture size is the same for each fracture in a single simulation.

Large-scale fractures on outcrops tended to be near-vertical and spaced 1 – 10 m apart. Microscale fractures spaced centimeters apart were also observed. In this paper, a range of fracture spacings from 30 m to 200 m was evaluated in reservoir modeling. Shorter spacings reflect outcrop-scale observations, allowing for some fractures to not transmit flow, and longer

spacings reflect observations from an airborne LiDAR dataset (Section 5 of Appendix D) developed in a 20 km x 15 km area in the Adirondack Mountains. If too few fractures that transmit flow exist in basement rocks below Cornell, an enhanced geothermal system (EGS) could be made with the modeled 30 m – 200 m spacings. Creating an EGS could also stimulate the observed microscale fractures, although we currently do not have local information for stress at depth with which to characterize such mechanical processes.

Fracture lengths were assigned based on the faults mapped from LiDAR topography. The mean length of faults in this examined area of the Adirondacks exceeds 1.5 km in length, although we do not know if they remain hydrologically connected over such length scales. For modeling, we assume 500 m connected lengths.

Fractures observed in the Adirondacks exist in nearly all orientations, but are preferentially oriented NNE (e.g. Valentino et al., 2016). For a first-order approximation of heat extraction from a basement reservoir, we use the simple one-dimensional model of vertical parallel fractures described in Section 4.4.1.2.

#### **4.4 Geothermal Reservoir Models**

It is unclear what the reservoir flow geometry will be in the planning phase of a geothermal project. We consider simple analytical reservoir models that are intended to bound the possible heat extraction from porous media reservoir geometries. Evaluating several flow geometries is similar to using an uncertain thermal recovery factor in volumetric heat-in-place calculations (e.g. Garg and Combs, 2015). These analytical models allow ranking of geothermal reservoir production scenarios to evaluate in numerical simulations, which consider additional reservoir geometry complexities. For basement reservoirs, we assume the rocks are impermeable

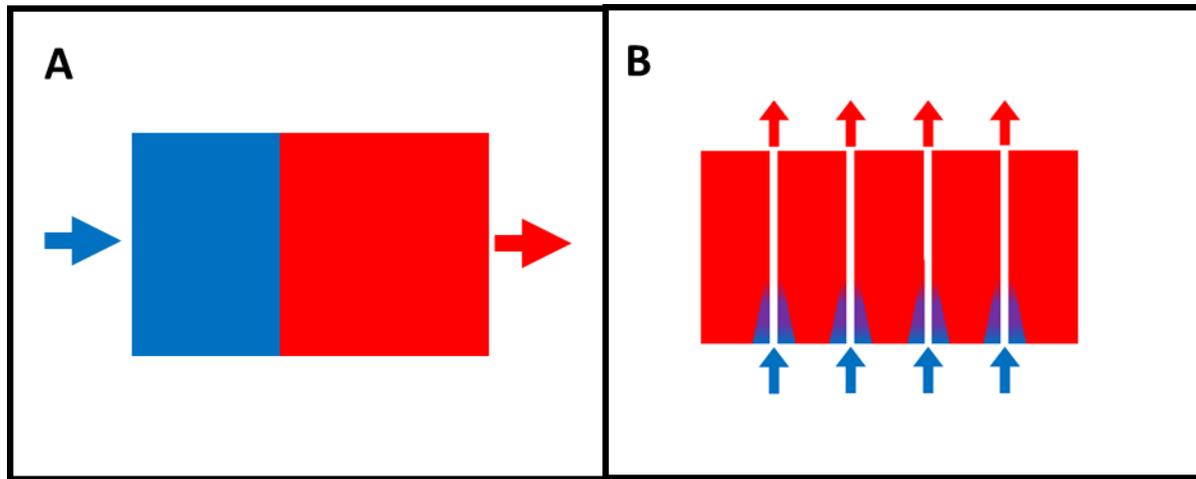
except for fluid flow through fractures. The reservoir models we used are described in Sections 4.4.1 and 4.4.2.

For all models, for simplicity, it's assumed that freshwater will be injected to the geothermal reservoir and be essentially freshwater when it's produced again. Brine that would be initially in the reservoir is used to determine the initial pressure conditions for numerical modeling (Section 4.4.2.1). It is unknown for our geothermal reservoir rocks how much salt would be dissolved into the water over the residence times of fluid from the injection well to the production well. The effect of using brine instead of freshwater may be evaluated numerically in future studies, as in Walsh et al. (2017). Generally, the density of brine is higher than the density of freshwater and the heat capacity is smaller (e.g. Ramalingam and Arumugam, 2012). These effects tend to not cancel out for the temperatures evaluated in this study, so our use of freshwater will likely underestimate the reservoir's total available thermal energy (e.g., Equation 4.3). Additionally, a higher heat capacity of freshwater means that the calculated extraction of heat per unit time will be overestimated (e.g., Equation 4.4; Walsh et al., 2017). These differences between brine and freshwater are acceptable for the purpose of relative ranking of favorable and unfavorable production scenarios in this study, but would be valuable to quantify for probabilistic interpretations of reservoir productivity.

#### **4.4.1 Analytical Reservoir Models**

Two analytical thermal-hydraulic models are used in this paper. The first is a porous media heat sweep model, which considers plug flow of a thermal front from the injection well to the extraction well. The second is a parallel vertical fractures model. Conceptual diagrams of

these reservoir models are provided in Figure 4.4. The equations and assumptions for each model are developed in the following sections.



**Figure 4.4:** Conceptual diagrams for analytical reservoir models after a lengthy time has elapsed. Each large rectangle represents an earth cross-section, with vertical and one horizontal dimension. Warmer temperatures are shown as red and cooler temperatures are shown as blue. Arrows represent well locations, both injection (blue) and extraction (red). The injection well water is cooler than the reservoir rock. Initially the whole block would have been red. A) Porous media heat sweep model. B) Parallel fractures model.

#### 4.4.1.1 Porous Media Heat Sweep Model

A plug flow of the thermal front from the injection well to the extraction well is used to estimate the maximum reservoir lifetime for uniform flow in homogeneous porous media (Figure 4.4A). This model does not consider heat generation, or conduction recharge of the reservoir from caprocks and base rocks. These effects are negligible compared to advective heat transfer over the timescales modeled. A geothermal gradient is not modeled, but the temperature change is negligible over the assumed 30 m thickness of higher permeability for the Trenton-Black River reservoir. The heat sweep model is provided in Equations 4.3 - 4.5.

The thermal energy stored in a rock and its pore fluids is provided in Equation 4.3 (e.g. Garg and Combs, 2015)

$$Q_r = \left( [1 - \phi] \rho_r C_{p_r} + \phi \rho_w C_{p_w} \right) V_r (T_r - T_{inj}) \quad [4.3]$$

where  $Q_r$  is the total thermal energy (J),  $\phi$  is the porosity of the rock,  $\rho_r$  and  $\rho_w$  are the density of the rock and pore fluid, respectively ( $\text{kg/m}^3$ ),  $C_{p_r}$  and  $C_{p_w}$  are the heat capacities of rock and pore fluid, respectively ( $\text{J/kg-K}$ ),  $V_r$  is the volume of the reservoir rock ( $\text{m}^3$ ),  $T_r$  is the temperature of the rock (K), and  $T_{inj}$  is the temperature of the injected fluid (K).

The heat transfer from the rock to the geothermal fluid (freshwater) is provided in Equation 4.4

$$Q_w = \dot{m} C_{p_w} \Delta T, \quad \Delta T = (T_{prod} - T_{inj}) \quad [4.4]$$

where  $Q_w$  is the heat flow rate ( $\text{W}_{th}$ ),  $\dot{m}$  is the fluid mass flow rate ( $\text{kg/s}$ ),  $C_{p_w}$  is the heat capacity of the produced fluid ( $\text{J/kg-K}$ ), and  $\Delta T$  (K) is the difference in temperature from the produced fluid to the reinjected fluid. The mass flow rate and the injection temperatures are selected in production scenarios in this paper.

The residence time (operational lifetime) for the plug flow model is defined to be when the temperature throughout the rock volume reaches the injected fluid temperature and all of the available heat has been extracted by the working fluid, as provided in Equation 4.5.

$$t_{res} = \frac{Q_r}{Q_w} \quad [4.5]$$

Until the residence time elapses, the fluid temperature at the depth of the extraction well is equal to the initial rock temperature.

#### 4.4.1.2 Parallel Vertical Fractures Model

The parallel vertical fractures model (Figure 4.4B) is described in Gringarten et al. (1975). This reservoir model considers a series of parallel, equally spaced vertical fractures with uniform aperture, separated by an impermeable rock matrix. Conduction recharge is modeled from the reservoir rock into the fluid filling the fractures. Vertical conduction recharge is not modeled. Heat generation and buoyancy effects are neglected. The geothermal gradient in the reservoir is modeled, such that the initial water temperature in the fracture depends on depth. Equal mass flow rate is contained within each fracture. Under these assumptions, the temperatures of the water and rock are described by the partial differential equations in Equation 4.6 and Equation 4.7 (Gringarten et al., 1975)

$$b\rho_w C_{p_w} \left[ \frac{\partial T_w(z, t)}{\partial t} + v \frac{\partial T_w(z, t)}{\partial z} \right] = K_r \frac{\partial T_r(x = b, z, t)}{\partial x} \quad [4.6]$$

$$\frac{\partial^2 T_r(x, z, t)}{\partial x^2} = \frac{\rho_r C_{p_r}}{K_r} \frac{\partial T_r(x, z, t)}{\partial t} \quad [4.7]$$

where  $b$  is half of the fracture aperture,  $K_r$  is the thermal conductivity of the rock,  $T_w(z, t)$  is the water temperature at vertical position  $z$  at time  $t$ ,  $v$  is the vertical speed of the water,  $T_r(x, z, t)$  is the temperature of the rock at horizontal position  $x$  and vertical position  $z$  at time  $t$ , and all other terms are as previously defined. Equation 4.6 is the heat balance between the fluid flowing in the fracture and the horizontal conduction recharge from the rock at the fracture interface (location  $b$  along the horizontal  $x$ -axis with origin in the center of a fracture). Equation 4.7 describes the conduction recharge in rock in the  $x$  direction over time. These equations assume that thermal drawdown occurs equally along the  $y$  dimension (into the page in Figure 4.4). The reader is

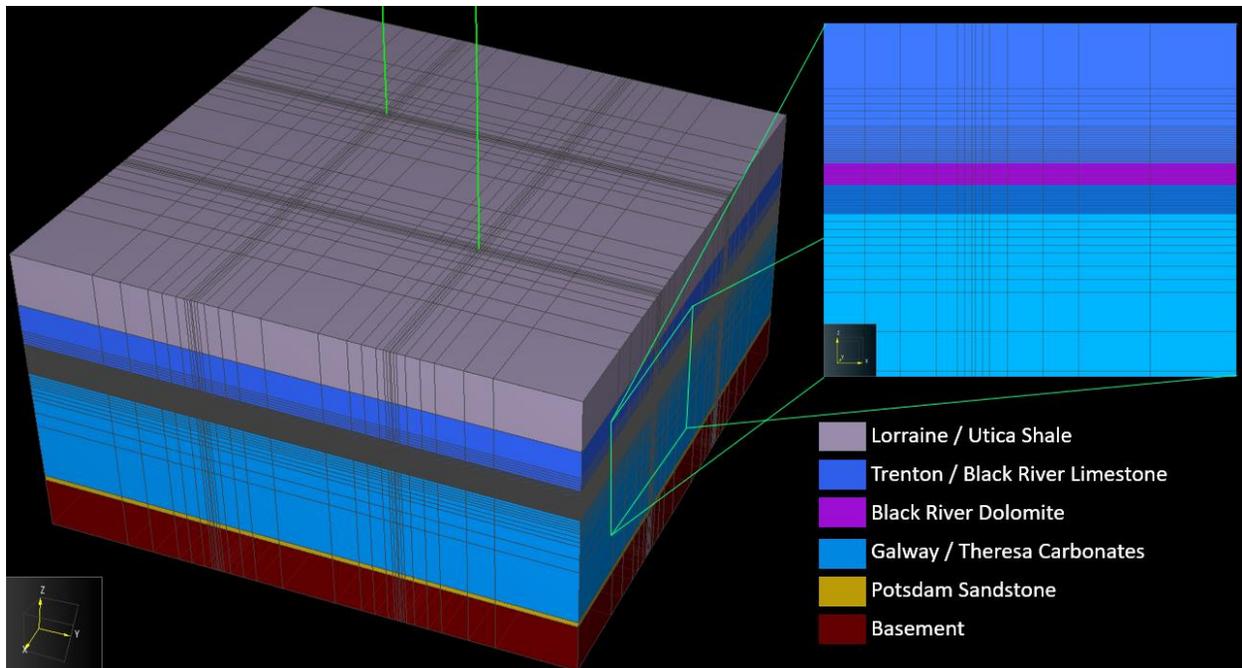
directed to derivations in Gringarten et al. (1975) and the implementation in the GEOPHIRES software (Beckers and McCabe, 2018) for further details and for solutions to these equations in the Laplace domain.

#### **4.4.2 Numerical Reservoir Model TOUGH2**

The TOUGH2 (Transport of Unsaturated Groundwater and Heat, version 2) software is a commonly used thermal-hydraulic numerical model for geothermal reservoir simulations in porous and fractured media (Pruess et al., 2012). We utilize the TOUGH2 EOS1 module (freshwater and heat) to model geothermal reservoir temperature and pressure over time. In this study, TOUGH2 is used to model the Trenton-Black River sedimentary reservoir and surrounding rocks as porous media; fracture flow is not modeled with TOUGH2 in this study. An adaptive timestep with a minimum of 100 seconds was used in numerical simulations. A maximum of 8 iterations per timestep was selected to achieve a relative error tolerance of  $1E-5$ .

In TOUGH2, we assume that each formation is a homogeneous porous medium that has anisotropic permeability (vertical vs. horizontal). This assumption provides a best-case scenario because heterogeneous porous media is more likely to result in flow channeling that reduces the lifetime of a geothermal system (e.g. Sinclair et al., 2018; Huang et al., 2017). Heterogeneities in permeability are considered between formations.

The simulation grid used for TOUGH2 reservoir simulations of the Trenton-Black River (TBR) reservoir is provided in Figure 4.5. Grid cell density is concentrated near the wells and the TBR formation. Horizontal grid cell spacings range from 5 m at the well locations to 200 m in the center of the reservoir. The injection and extraction well are separated by 1 km. The volume of rock in the grid is 2 km x 2 km x 1.14 km.



**Figure 4.5:** Modeled geologic strata and simulation grid in the TOUGH2 numerical model. Grid cells are colored by the rock lithology: blue: carbonate, gray: shale, yellow: sandstone, red: basement. Wells are shown as vertical green lines extending from the top of the grid. The injection well is in the top left finely grid area.

#### 4.4.2.1 Initial Conditions

Simulations were allowed 5,000 years to reach a natural steady state before running geothermal reservoir production scenarios in TOUGH2. Initial values of temperatures at depth are specified using the geothermal gradients in Table 4.1. Brine density samples from regional wells were used to estimate a pressure profile with depth below Cornell. We assume that the brine density is constant with depth, and pressure is hydrostatic with depth. Hydrostatic pressure conditions have been observed in central New York deep boreholes (e.g., Auburn geothermal borehole, Lynch and Castor, 1983). Oilfield brine composition data (Dresel, 1985), disposal well data (Waller et al., 1978), and the produced waters database from the USGS (Siegel et al., 1990; Skeen, 2010; Lynch and Castor, 1983; Matsumoto et al., 1996) were used to evaluate brine density to inform the pressure profile. These datasets are from counties in north-central

Pennsylvania and south-central New York. Based on data from 56 wells, we assume a hydrostatic pressure profile with pore fluid density of 1180 kg/m<sup>3</sup>.

For simplicity, we assume that the water table is located at the surface. The depth to the water table in 110 wells in counties south and southeast of Ithaca in New York State ranges from less than 12 ft to as much as 800 ft with an average of about 120 ft (Williams, 2005). Given the assumed density of brine, the difference in the pressure profile for water at the surface versus 800 ft is about 2 MPa at the depth of the TBR reservoir.

Using these assumptions, the pressure profile is calculated using Equation 4.8

$$P(z) = \rho_b g z + P_{atm} \quad [4.8]$$

where  $P(z)$  is the formation pressure as a function of depth below ground surface,  $z$ ,  $g$  is the acceleration of gravity,  $\rho_b$  is the density of brine, and  $P_{atm}$  is the atmospheric pressure. We assume that the local gravity is 9.81 m/s<sup>2</sup>, and an atmospheric pressure of 1 atm. For the Trenton-Black River target depth of about 2270 m, the estimated pressure is about 26 MPa.

#### 4.4.2.2 Boundary Conditions

The surface boundary condition is a fixed temperature and a fixed pressure at the top of the Lorraine / Utica Shale. The values selected are consistent with the estimated temperature and pressure profiles presented in Table 4.1 and Equation 4.8, respectively. No flow boundary conditions are used for the sides of the simulation grid. The basal boundary condition is the heat flow as calculated using Equation 4.1.

### 4.4.3 Extraction Well Borehole Heat Transfer

The above geothermal reservoir models describe fluid temperature over time at the depth of the reservoir. To estimate temperatures at the ground surface, we use an analytical solution for wellbore heat transfer in the extraction well (Ramey, 1961). The injection well temperature gain is not considered, but this would increase the lifetime of the geothermal system. Simplifications made to compute wellbore heat transfer include: 1) using an average geothermal gradient computed from the surface to the extraction well depth, 2) assuming any wellbore casing thermal resistance is negligible compared to rock thermal resistance. Using these assumptions, the extraction wellbore heat loss over time is computed using Equation 4.9

$$T(z, t) = - \left( \left( T(z, 0) - T_{\text{prod}}(t) \right) - G(z - A) + \left( T_{\text{prod}}(t) - GA - T(z, 0) \right) e^{-\frac{z}{A}} \right),$$

$$A = \frac{\dot{m}C_{p_w} f(t)}{2\pi K_r}, \quad f(t) = -\ln(r_o) + 0.5 \ln \left( \frac{K_r t \eta}{\rho_r C_{p_r}} \right) + \ln(2) - 0.29$$

[4.9]

where  $A$  is a heat transfer coefficient,  $\dot{m}$  is the mass flow rate of water,  $f(t)$  is the line-source transient heat conduction time function as a function of time  $t$ ,  $r_o$  is the outer diameter of the wellbore,  $\eta$  is the capacity factor of the geothermal system,  $G$  is the average geothermal gradient,  $T_{\text{prod}}(t)$  is the production temperature from the geothermal reservoir model, and all other terms are as previously defined. For this paper, we assume a system capacity factor that allows for 2 to 10 days of maintenance or down time, on average, in a year.

#### **4.4.4 Model Implementation and Uncertainty Analysis**

The analytical reservoir models and the wellbore heat transfer model are implemented in the GEOPHIRES software (Beckers et al., 2018). We made modifications to GEOPHIRES for Monte Carlo analysis of uncertain geologic properties and temperature gradients. A summary of the input values and probability distributions used in GEOPHIRES analytical models is provided in Table 4.3. For numerical reservoir modeling, we evaluated the sensitivity of our results to different temperature profiles, corresponding to the coolest 5<sup>th</sup> percentile, median, and warmest 5<sup>th</sup> percentile in Table 4.1.

**Table 4.3:** Summary of parameters used in GEOPHIRES analytical reservoir models. Probability distributions are listed for those variables that were selected randomly within Monte Carlo simulations. Triangular distributions list the lower bound, mode, and upper bound. Normal distributions list the mean and standard deviation. Lognormal distributions list the real space mean and standard deviation. Beta distributions list the lower tail shape parameter, the upper tail shape parameter, lower bound, and upper bound.

Variable	Trenton-Black River Plug Flow	Trenton-Black River Parallel Fractures	Basement Parallel Fractures	Notes and Sources
Reservoir Depth (km)	2.27 - 2.30	2.27 - 2.30	3 - 3.5, or 3.5 - 4	
Well Orientation in Reservoir	Vertical	Horizontal	Horizontal	
Reservoir or Fracture Height (m)	30	30	500	
Reservoir or Fracture Width (m)	1000	1000	500	
Reservoir or Well Lateral Length (m)	1000	1000	1000	
Fracture Separation (m)	NA	20	fixed at 30 m, or Triangular: 30, 30, 200	Adirondack Mountains
Fracture Width (mm)	NA	0.5	0.5	Camp and Jordan (2017), Adirondack Mountains
Reservoir Impedance (GPa-s/m <sup>3</sup> )	Triangular: 0.05, 0.15, 0.5	Triangular: 0.05, 0.15, 0.5	Triangular: 0.05, 0.15, 0.5	Camp et al. (2018) regional reservoir productivity.
Reservoir Rock Density (kg/m <sup>3</sup> )	Normal: 2800, 40	Normal: 2800, 40	Triangular: 2550, 2730, 3200	Local well logs, Simmons (1964)
Reservoir Rock Porosity (-)	Lognormal: 0.08, 0.046	NA	NA	Local well logs
Reservoir Rock Heat Capacity (J/kg-K)	Triangular: 900, 930, 940	Triangular: 900, 930, 940	Triangular: 800, 825, 850	Roberson and Hemingway (1995)
Reservoir Rock Thermal Conductivity (W/m-K)	Triangular: 1.92, 2.91, 3.9	Triangular: 1.92, 2.91, 3.9	Normal: 2.83, 0.36	Cornell University (2016a), matches assumptions in Smith (2019, Ch. 3)
Surface Temperature (°C)	Triangular: 8, 10, 12	Triangular: 8, 10, 12	Triangular: 8, 10, 12	Gass (1983), SMU Geothermal Lab (2016), matches assumptions in Smith (2019, Ch. 3)
Geothermal Gradient (°C/km)	0 – 1.5 km: Triangular: 26.5, 29.5, 33.7 1.5 – 2.8 km: Triangular: 23.7, 24.4, 25 2.8 – 4 km: Triangular: 16.5, 17, 17.5	0 – 1.5 km: Triangular: 26.5, 29.5, 33.7 1.5 – 2.8 km: Triangular: 23.7, 24.4, 25 2.8 – 4 km: Triangular: 16.5, 17, 17.5	0 – 1.5 km: Triangular: 26.5, 29.5, 33.7 1.5 – 2.8 km: Triangular: 23.7, 24.4, 25 2.8 – 4 km: Triangular: 16.5, 17, 17.5	Smith (2019, Ch. 3)
Utilization System Capacity Factor (-)	Beta(4,2), 0.97 - 0.994	Beta(4,2), 0.97 - 0.994	Beta(4,2), 0.97 - 0.994	Allows for two to ten days on average per year for maintenance.
Number of Monte Carlo Replicates	1000	1000	1000	

## 4.5 Production Scenarios

The sensitivity of reservoir heat production results to fluid flow rate and injection fluid temperature was evaluated for several production scenarios using analytical models. A subset of these models was also used in numerical simulations of the Trenton-Black River porous media reservoir. A summary of the production scenarios is provided in Table 4.4. Details for the selection of these values and other production parameters are described in Sections 4.5.1 – 4.5.3. The selected capacity factor for the system is discussed in Section 4.5.4.

**Table 4.4:** Summary of production scenarios for geothermal reservoir simulations.

<b>Production Scenario</b>	<b>Well Flow Rate (kg/s)</b>	<b>Injection Fluid Temperature (°C)</b>	<b>Well Inner Diameter (in)</b>
1	30	20	6.2
2	30	30	6.2
3	30	50	6.2
4	50	20	6.2
5	50	30	6.2
6	50	50	6.2
7	70	20	8.5
8	70	30	8.5
9	70	50	8.5

### 4.5.1 Well Flow Rates and Diameters

Flow rates of 30 kg/s, 50 kg/s, and 70 kg/s were selected based on common values for operating geothermal systems (Snyder et al., 2017). Wellbore diameters were selected so that the speed through pipes would be at most 3 m/s, which is the target circulation speed through the Cornell campus district heating system. The wellbore pipe size for 30 kg/s and 50 kg/s scenarios used a “small” inner diameter of 6.2” and the 70 kg/s scenario used a “large” inner diameter of 8.5” (Lowry et al., 2017).

### **4.5.2 Injection Temperatures**

The temperatures of injected fluid were selected based on proposed surface utilization requirements for the Cornell campus. According to S.M. Beyers (personal communication, 2018), a reinjection temperature of 50 °C is the current district heating system return temperature for campus facilities, and this temperature would be appropriate if heat pumps were not used to extract additional energy. If heat pumps are used, 30 °C represents an optimal geothermal district heating system coefficient of performance (COP) of about 10. A temperature of 20 °C represents a lower limit of injection temperatures for which the heat-pump-assisted geothermal system COP would be more efficient than using the current natural gas-based system (COP of 3.5 - 4).

For numerical modeling using TOUGH2, the injected temperature of freshwater is calculated over time according to the specified freshwater enthalpy. The enthalpy of water is a function of pressure and temperature, both of which vary over time during production from the geothermal reservoir. The enthalpy specification used in this analysis resulted in an effective injection temperature of about 23.5 °C for numerical models, instead of the 20 °C that was used in analytical models. Therefore, the production results from numerical models cannot be compared exactly to the analytical model results, but each numerical model result may be compared to each other numerical model result.

### **4.5.3 Well Separation Distance**

The distance between the injection well and the extraction well is limited by the area available on the Cornell site to space the wells and to directionally drill from the site. We evaluate 1 km well separation at depth for vertical wells in the TBR reservoir. For the parallel fractures models, 1 km horizontal wells are used, and the injection well is located below the

extraction well. In the TBR reservoir, the vertical well separation distance is the 30 m thickness of the reservoir. In basement reservoirs the vertical well separation distance is 500 m.

#### **4.5.4 Capacity Factor**

The capacity factor considers the fraction of the year that energy is extracted by the geothermal system. We assume capacity factors of 0.970 to 0.994, corresponding to a system being down for maintenance two to ten days on average per year.

### **4.6 Results**

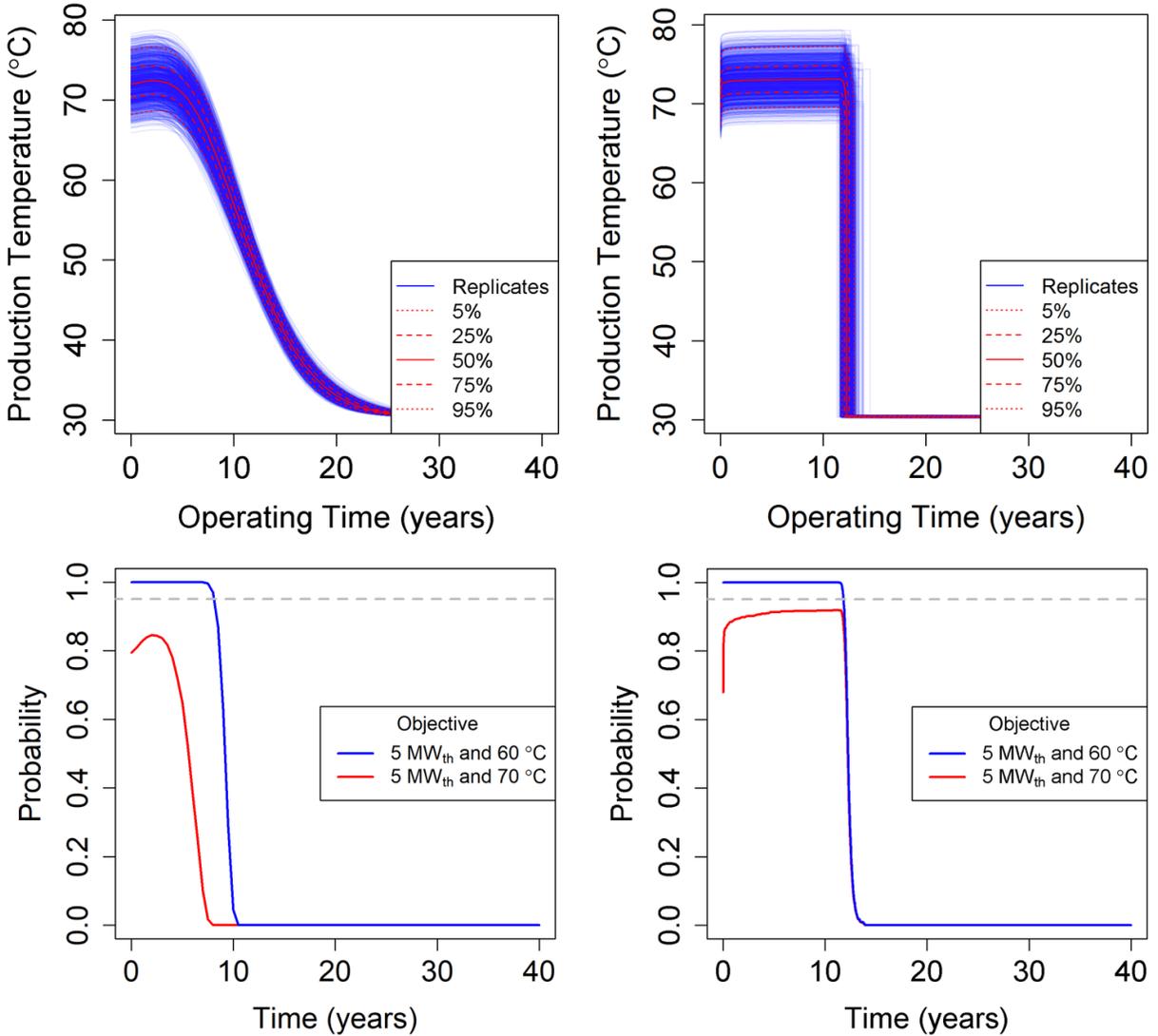
The results for the Trenton-Black River (TBR) reservoir simulations are presented in Section 4.6.1 and Section 4.6.2, followed by results for basement reservoirs at 3 – 3.5 km depth and 3.5 – 4 km depth in Section 4.6.3.

#### **4.6.1 Trenton-Black River Analytical Model Results**

The temperatures of produced fluids estimated from analytical modeling of the Trenton-Black River reservoir are shown in Figure 4.6 for an injection temperature of 30 °C and a pumping rate of 50 kg/s. Figure 4.6 illustrates temperature-time curves that are characteristic of each analytical model. For the parallel fractures model with vertical flow, rock temperature at the top of the fracture is initially cooler than the rock temperature at the bottom of the fracture. Over time, rock at the top of the fracture is warmed to at most the rock temperature at the bottom of the fracture. For the porous media plug flow model, a sudden drop in the produced temperature occurs once thermal breakthrough of the cold thermal front reaches the extraction well. For both geothermal reservoir models, extraction well borehole heat transfer cools the water along the

length of the extraction well; the effect of wellbore cooling decreases over time as the extraction borehole warms.

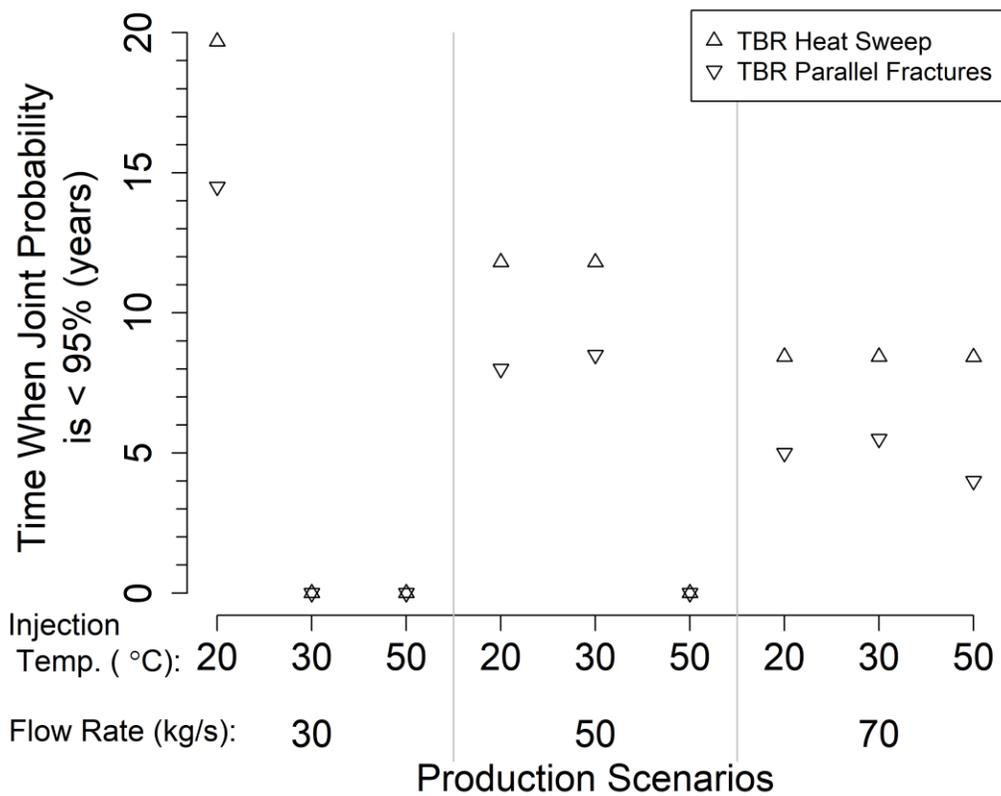
Figure 4.6 also shows the range of production temperatures that result from uncertainty in the initial temperature conditions (Figure 4.1, Table 4.1). For the plug flow model, uncertainty in the produced temperature decreases after thermal breakthrough as a result of the majority of the thermal energy having been extracted from the reservoir. For both models, uncertainty in the time to thermal breakthrough is right skewed with a range of about 2 years. Based on the Monte Carlo replicates, the joint probability of meeting the 60 °C and 5 MW<sub>th</sub> minimum requirements for the Cornell system falls below 95% after 8 - 12 years of operation for this production scenario. It is not possible to achieve a 95% probability of meeting 70 °C and 5 MW<sub>th</sub> requirements for standard temperature Cornell facilities with this production scenario.



**Figure 4.6:** Top Row: Production temperature over time for the Trenton-Black River reservoir using the parallel fractures model (left) and the heat sweep model (right). The injection temperature is 30 °C and the flow rate is 50 kg/s. Each blue line provides the results of a single Monte Carlo replicate. Selected quantiles over time are provided in red. Bottom row: The probability of meeting the listed objectives over time. The horizontal dashed gray line is located at 95% probability.

Figure 4.7 summarizes the joint probability of producing fluids that meet both the 60 °C and the 5 MW<sub>th</sub> heating objectives for each of the production scenarios and TBR analytical reservoir models. We consider the necessary condition for successful operation of the geothermal system to be a high joint-probability of meeting the objectives, and therefore identify an estimate

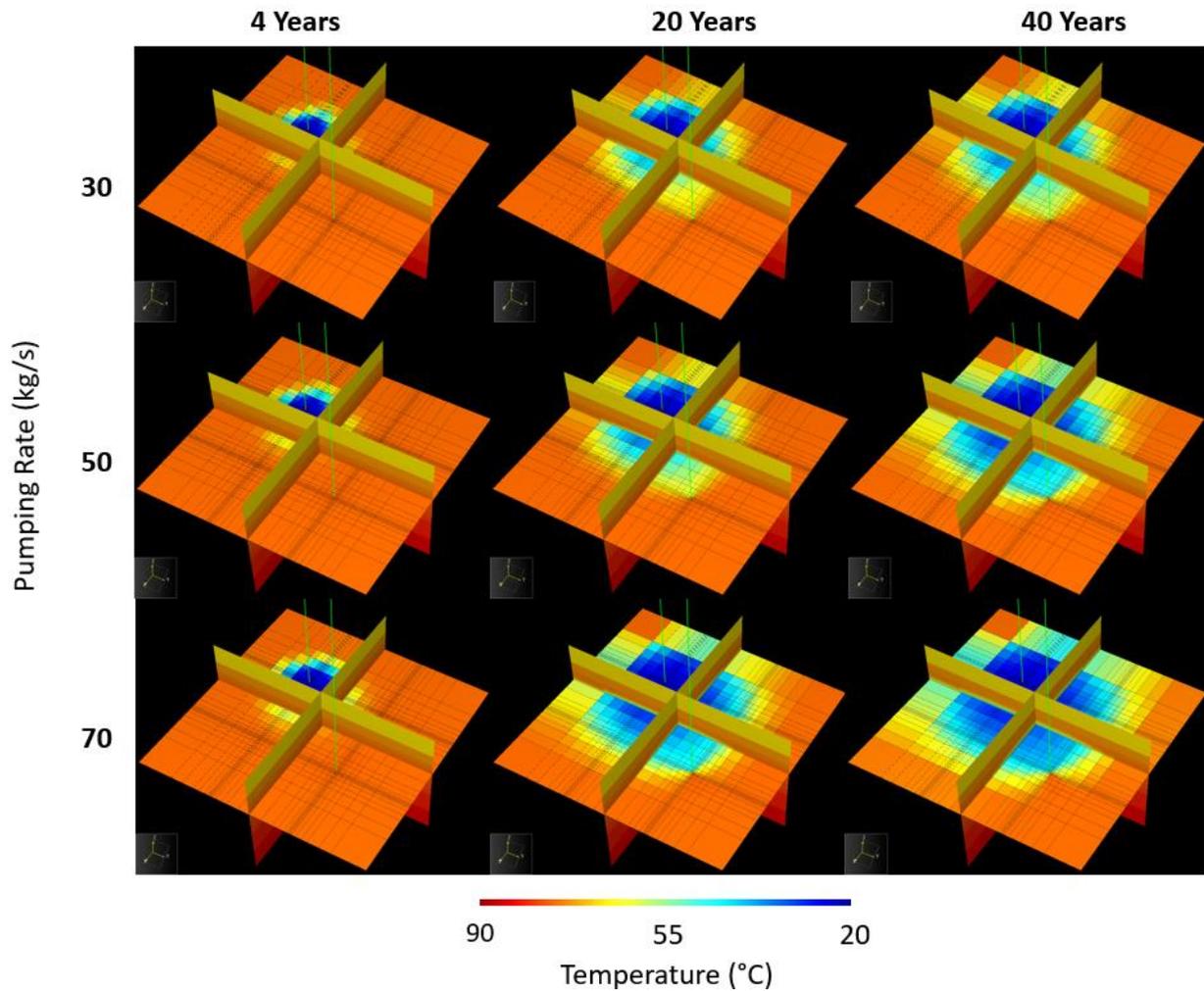
of the useful life span of the geothermal system to be the time at which the joint probability of meeting the objectives is less than 95%. If these analytical models bound the possible flow geometries of a reservoir, the range in useful life estimates for a production scenario may be interpreted as an estimate of the time that thermal breakthrough is likely to begin. Using these criteria, a flow rate of 30 kg/s with an injection temperature of 20 °C dominates the other scenarios in expected useful life. We evaluate this scenario in numerical modeling, along with other 20 °C injection scenarios to demonstrate the sensitivity of the numerical simulation results to flow rate.



**Figure 4.7:** The time after geothermal production began at which the joint probability of a Trenton-Black River production scenario meeting a temperature of 60 °C and heat production of 5 MW<sub>th</sub> declines to less than 95%. Results are shown for the heat sweep porous media analytical model (upward pointing triangles) and parallel fractures model (downward pointing triangles). Where both models estimate similar performance, the overlap of the two symbols makes a star shape.

#### 4.6.2 Trenton-Black River Numerical Model

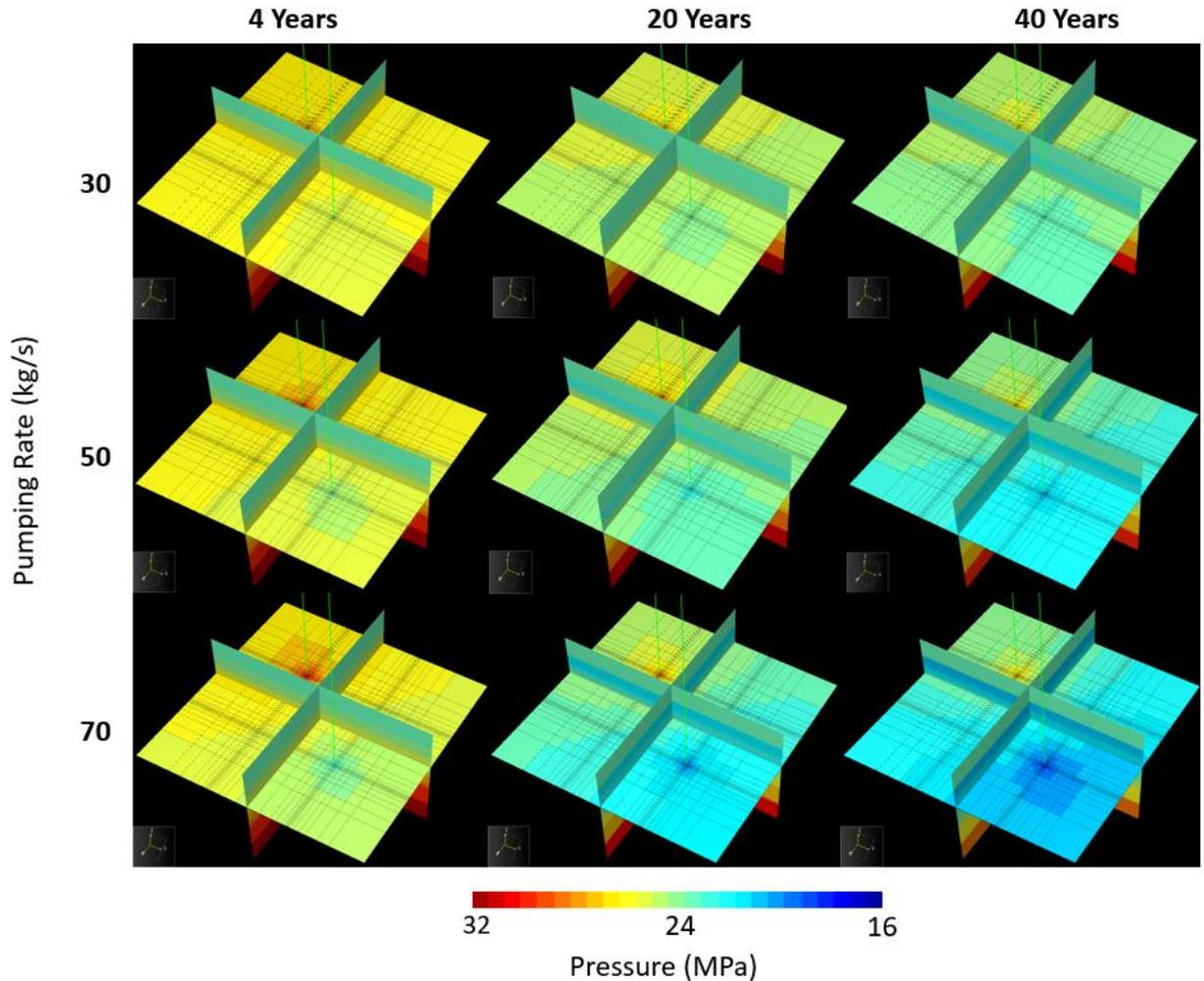
Results from TOUGH2 numerical modeling of the TBR rock temperature over 40 years of production from porous media flow are provided in Figure 4.8. These plots are for the warmest 5<sup>th</sup> percentile of temperature, which provides patterns of temperature and pressure in the reservoir over time that are similar to the other two initial rock temperature percentiles evaluated. In contrast to the analytical models, wells in numerical models allow for fluids to migrate in flow paths other than straight-line paths to the wells, which results in longer times until thermal breakthrough occurs. Some caprock thermal recharge is apparent after 20 years of production, which also contributes to longer times until thermal breakthrough compared to analytical models. Flow rates of 30 kg/s result in fluid migration mostly within the ~700 m between the edges of the simulation grid and the injection well. Flow rates of 70 kg/s will sweep a larger volume of rock, but will also result in shorter times to thermal breakthrough from the injection well, as observed in analytical models.



**Figure 4.8:** Rock temperature over 40 years of production from the Trenton-Black River (TBR) reservoir for the condition of flow through matrix porosity for the warmest 5<sup>th</sup> percentile initial rock temperature conditions. A 23.5 °C injection temperature is used with well flow rates of 30 kg/s, 50 kg/s, and 70 kg/s. The surface parallel with the x-y plane is located in the center of the TBR reservoir. The surfaces parallel with the x-z and y-z planes are located in the center of the y and x axes, respectively. The grid orientations shown in this figure are the same as in Figure 4.5 and Figure 4.9. Wells (vertical green lines) are injection in the top left zone and extraction in the lower right zone.

Plots of the rock pressure over time are provided in Figure 4.9. Near the injection well, pressure is initially high where water is forced into the system, which for the 70 kg/s flow rate causes at most about 6 MPa increase in pressure relative to the ambient pressure of 26 MPa. Camp and Jordan (2017) estimated that the total pressure needed to hydroshear optimally

oriented pre-existing fractures in regional TBR reservoirs as  $46.8 \pm 5$  MPa. All simulated pressures are well below this threshold. Therefore, the pressure increases that result from fluid injection are not expected to be large enough to significantly change the flow geometry near the injection well for any of the evaluated scenarios.



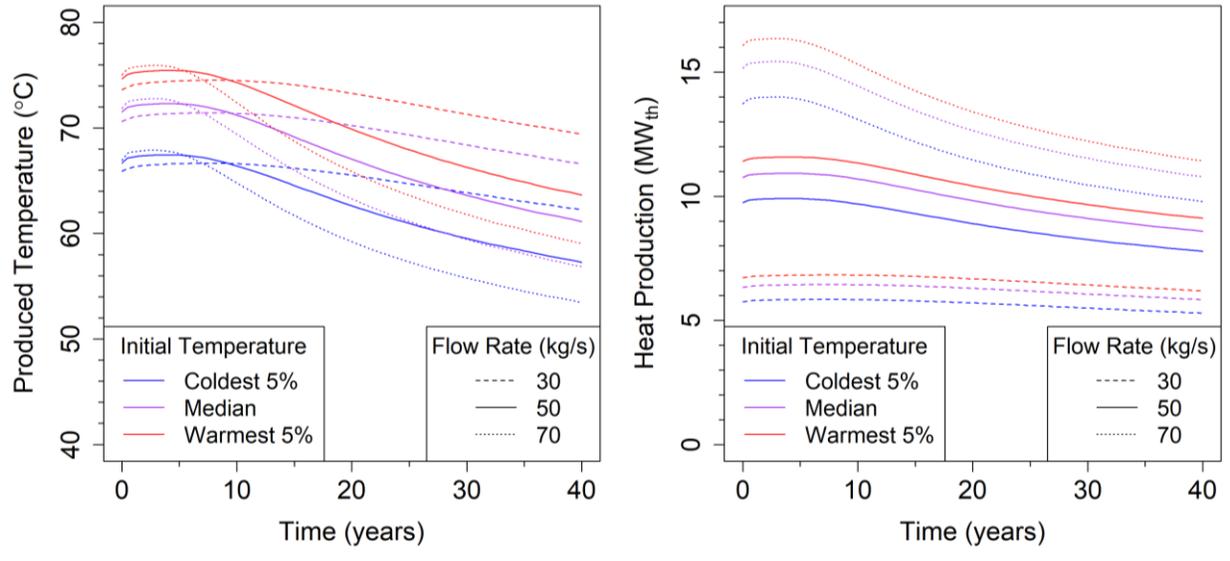
**Figure 4.9:** Pressure over 40 years of production from the Trenton-Black River (TBR) reservoir for the warmest 5<sup>th</sup> percentile initial rock temperature conditions. A 23.5 °C injection temperature is used with well flow rates of 30 kg/s, 50 kg/s, and 70 kg/s. The surface parallel with the x-y plane is located in the center of the TBR reservoir. The surfaces parallel with the x-z and y-z planes are located in the center of the y and x axes, respectively. The grid orientations shown in this figure are the same as in Figure 4.5 and Figure 4.8. Wells (vertical green lines) are injection in the top left zone and extraction in the lower right zone.

#### 4.6.2.1 Produced Temperatures

Figure 4.10 provides the produced temperature profiles for numerical models after considering the effects of extraction well borehole heat transfer. The initial temperature drops due to wellbore heat transfer range from about 1 °C for 70 kg/s flows to about 3 °C for 30 kg/s flows. Faster fluid flow allows for less heat transfer time. The initial timesteps illustrate the effect of drawing water from deeper parts of the reservoir with greater pumping rates. Pumping rates also have a clear impact on the time to thermal breakthrough, as shown previously with analytical model results (Figure 4.7). Pumping rates of 50 and 70 kg/s result in temperature declines of about 5 °C and 10 °C, respectively, after 20 years of operation (Figure 4.10). Pumping at 30 kg/s results in a temperature decline beginning after about 15 years of production, and only 1 °C total decline after 20 years of operation (Figure 4.10). If the TBR permeable zone is constrained to a 4 km<sup>2</sup> spatial area, as modeled, pumping at 30 kg/s would be attractive for long-term operation of the system, as the heat production and temperature objectives are more likely to be met over a long useful life.

For many of the numerically modeled production scenarios, heat production meets or exceeds the target objectives for at least 40 years. Produced temperatures for only three of the scenarios do not exceed 60 °C for 40 years. Two of these scenarios are for flow rates of 70 kg/s, which would supply heat greater than the 12 MW<sub>th</sub> campus baseload even for the coolest 5<sup>th</sup> percentile estimate of temperature, and therefore would likely not be necessary to meet Cornell's heating objectives (Figure 4.10). If the median temperature was met or exceeded by a Cornell geothermal system, using a 50 kg/s flow rate could supply baseload energy requirements for standard temperature facilities with 70 °C supply for a minimum of 12 years. Given the uncertainty in the initial rock temperatures, it seems unlikely that the TBR reservoir would

provide temperatures sufficient for high temperature facilities with 80 °C supply temperature. To meet the heating demands of those buildings a deeper reservoir target would be needed, such as the proposed basement rock targets.

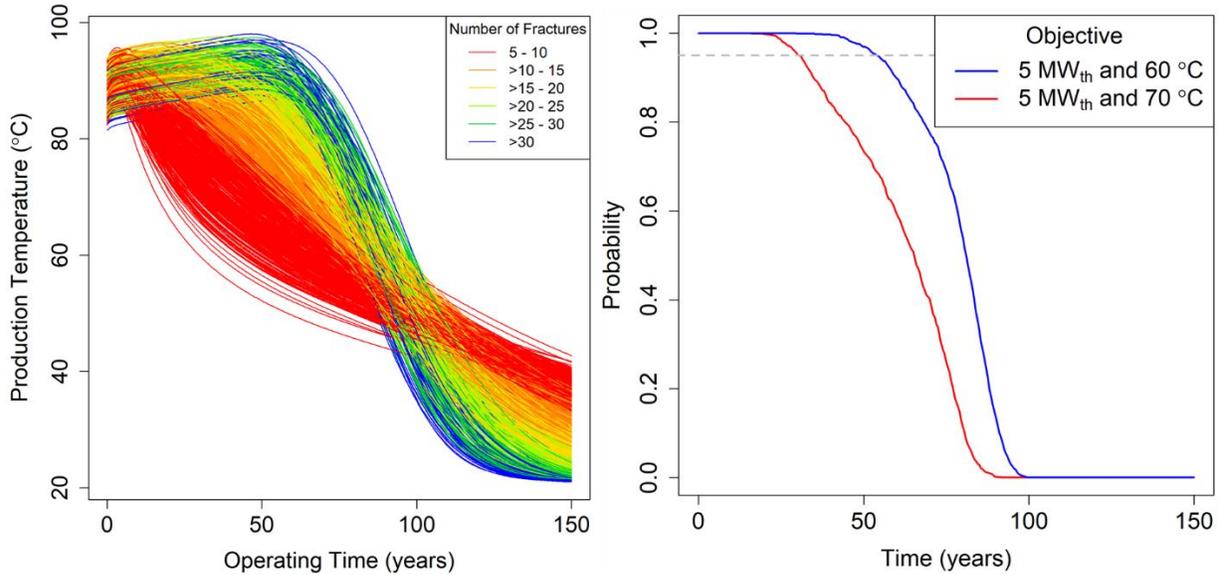


**Figure 4.10:** Estimated heat production and temperature over time for the Trenton-Black River for injection of 23.5 °C fluid. The initial rock temperature percentiles were selected based on the estimated temperatures at depth (Figure 4.1).

### 4.6.3 Basement

Results of the Monte Carlo simulations of a fracture-dominated basement reservoir at 3.0 – 3.5 km depth are shown in Figure 4.11 for a well flow rate of 50 kg/s and 20 °C fluid injection temperature. The time to thermal breakthrough is highly dependent on fracture spacing, as demonstrated in other studies (Gringarten et al., 1975). Using the results for variable fracture spacing in Figure 4.11, one can decide how many fractures would be required for meeting the objectives of a geothermal system. Unfortunately, the temperature-time trajectories are similar in the first 5-10 years regardless of the number of fractures. For the conditions evaluated in Figure 4.11, fracture spacing greater than about 50 m may not provide desirable long-term production before thermal breakthrough. Even so, considering fracture spacing variability, 50 years of

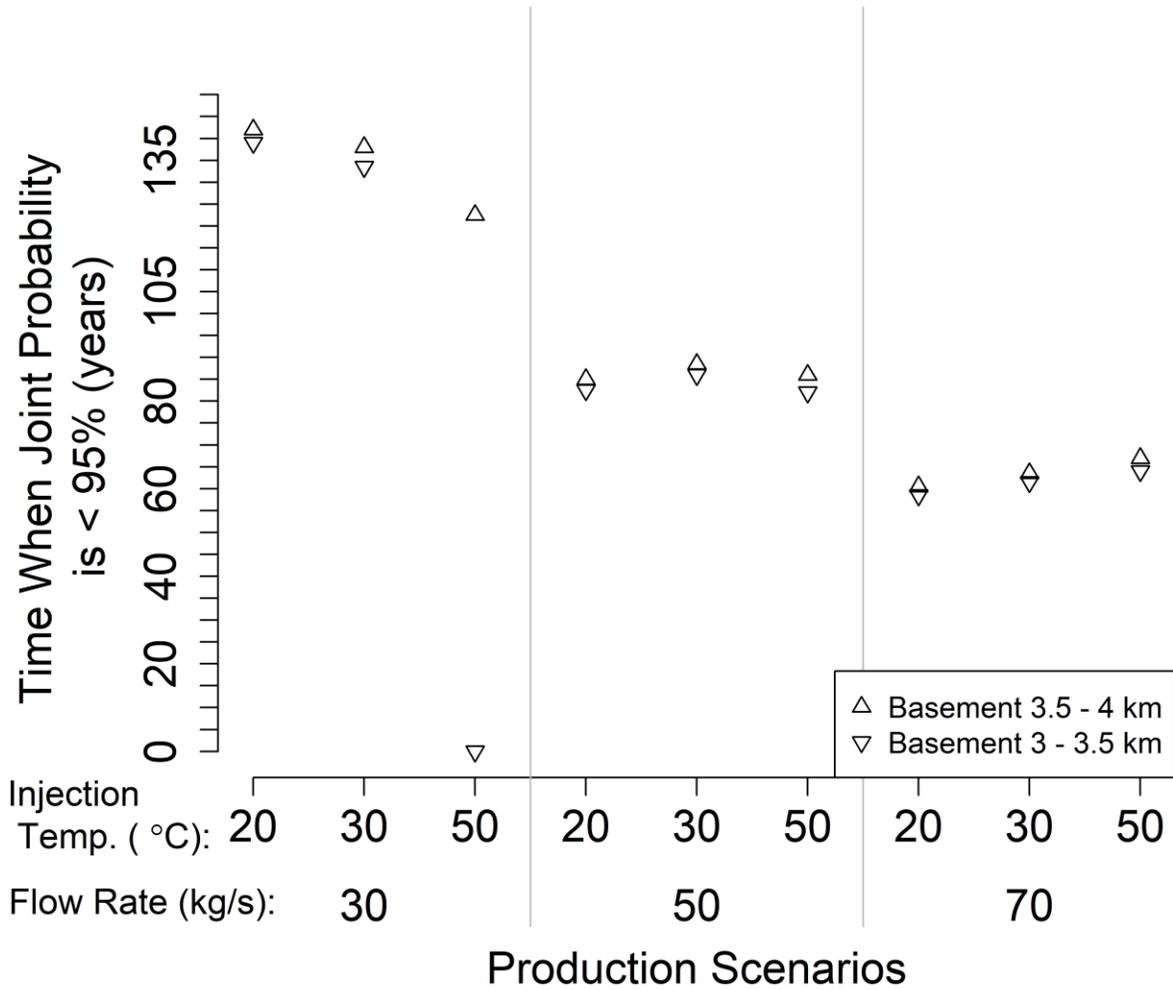
operation could be expected with at least 95% probability of 5 MW<sub>th</sub> heat production at temperatures greater than 60 °C, and 25 years could be expected at temperatures greater than 70 °C. All of the Monte Carlo replicates are initially sufficient for “high-temperature facilities” with 80 °C supply temperatures at Cornell. The replicates with the fewest fractures fall below 80 °C within 10 years, although they remain above 60 °C for a minimum of about 25 years.



**Figure 4.11:** Multiple parallel fractures temperature production results for basement rock at 3 – 3.5 km depth for 20 °C injection at 50 kg/s. Left: Each line provides the results of a single Monte Carlo replicate. Lines are colored by the number of fractures, which ranged from 5 to 33 over a 1 km horizontal well length. Right: The probability of producing fluid that meets the specified objectives over time.

Figure 4.12 presents the joint probability of producing fluids that meet both the 60 °C and the 5 MW<sub>th</sub> heating objectives using basement reservoirs, conditional on having a 30 m fracture spacing. If such spacings do not occur naturally for basement rocks below Cornell, it may be possible to use enhanced geothermal systems to engineer such a fracture system. For basement rock at 3 – 3.5 km depth and 3.5 – 4 km depth, nearly all scenarios evaluated meet the heating objectives for Cornell. If these conditions could be achieved, drilling to 3.5 – 4 km depth seems

to have little value, given the small difference in estimated useful life for deeper wells, whereas the deeper wells could cost an additional \$5 million total to drill (Lowry et al., 2017).



**Figure 4.12:** The time at which the joint probability of a basement reservoir production scenario meeting a temperature of 60 °C and heat production of 5 MW<sub>th</sub> becomes less than 95%. Results are shown for rocks at 3.5 – 4 km depth (upward pointing triangles) and for 3 – 3.5 km depth (downward pointing triangles). Both reservoirs were modeled using a parallel fractures model.

#### 4.7 Discussion

This paper presents uncertainty analysis methods for simple reservoir models that may be used for pre-drilling evaluations of reservoir feasibility. These analyses can be used for probabilistic assessments of meeting heating objectives that can inform economic and risk

decisions about drilling a well, or the value of obtaining additional data that could improve the precision of the models (e.g. Trainor-Guitton et al., 2013).

The probabilities estimated for any reservoir meeting the target objectives are conditional on the assumed reservoir flow geometries. The flow geometry will be understood only vaguely in the planning phase of most geothermal projects, so we used simple analytical models that could bound possible reservoir performance for porous media and fracture-dominated systems. It is likely that a real geothermal reservoir will have complexity that is not captured in our analyses. Nevertheless, our methods of uncertainty analysis for analytical models are not limited to simpler reservoir flow geometries; other models of geothermal reservoirs including heterogeneous discrete fracture networks (e.g. Fox et al., 2016; Doe et al. 2014), or fractured porous media (e.g. Abbasi et al., 2019) may be evaluated using the same methodologies.

For numerical modeling, flow channeling (e.g. Watanabe et al., 2010) that results from heterogeneous porous media within a formation was not evaluated. The formation for which spatially correlated heterogeneous permeability would likely be most important is the Black River dolomite. Variograms of the vertical spatial correlation in Black River rock porosity for the wells with corrected porosity logs are provided in Appendix D Section 4. For the Duddleston well that was used to approximate high productivity dolomite, the correlation range is less than 10 ft (3 m), which suggests a heterogeneous reservoir on the scales modeled in our analysis (2 m grid cells in numerical modeling). Thus, flow channeling may be important to evaluate for this reservoir in future studies. We also did not evaluate the impact of chemical reactions that alter the pore space distribution during production and have been shown to be important for carbonate reservoirs, such as the Trenton-Black River (e.g. Pandey and Chaudhuri, 2017). Additionally, because the TBR reservoir is known to produce gas regionally, future work should evaluate

numerical models that include multiphase flow of gas and water and compare the results of the analysis to those presented in this paper.

We consider the performance of a single well pair for the purpose of ranking reservoir production scenarios. Re-drilling, multi-well production, or rotation of wells over time to allow thermal recharge may improve the performance of the system (e.g. Fox et al., 2013), and can be considered in economic assessments. We also do not consider variable rock volumes that result from different well spacings, or from uncertainty in geologic knowledge. For the estimated uncertainty in the depth to the basement of  $\pm 200$  m (T.E. Jordan, personal communication, 2018), about  $\pm 3.5$  °C change in the temperatures at the top of the basement are expected, and about  $\pm 1$  °C would result from the uncertainty on depth to the Trenton-Black River reservoir. Given that basement rocks provided temperature estimates in excess of the 80 °C high temperature facility supply requirement for Cornell, such uncertainty in depth to this reservoir should minimally impact decision making for basement reservoirs. For Trenton-Black River reservoirs, the most likely produced temperatures are close to the standard temperature facility supply temperature of 70 °C. Using the multiple parallel fractures model, we evaluated a shallower TBR scenario in which the depth to the top of the TBR was located at 2170 m instead of 2270 m. For the production scenario in Figure 4.6, the length of time that reservoir production achieved 95% joint probability of meeting Cornell's minimum heating objectives was reduced by about half a year. The joint probability of meeting standard temperature facilities heating requirements was reduced from a maximum of about 85% (Figure 4.6) to about 50%. These appreciable reductions in performance suggest that considering the reservoir depth (i.e. geologic column uncertainties) as variable across reservoir simulations would be valuable to better inform decisions derived from both analytical and numerical reservoir models.

Specific to the Cornell campus, if the hypothetical TBR reservoir has the assumed 4 km<sup>2</sup> permeable zone area over a 30 m thickness, results show that production temperatures and heat production could meet the objectives of the Cornell campus district heating system for several decades for even the coolest 5<sup>th</sup> percentile of estimated rock temperatures. If instead the reservoir is confined to a smaller 1 km<sup>2</sup> area, TBR useful production could be as short as 15 years for the most favorable production scenario. That favorable scenario used a 30 kg/s flow rate with 20 °C injection temperature, for which heat production and temperature objectives are estimated to be met for a minimum of about 15 years with 95% probability. There is a chance that a naturally permeable zone in the TBR does not exist below Cornell. The probability of a thinner reservoir or no permeable zone was not considered in this study because it is assumed that reservoir enhancement methods could be used to generate the assumed permeable thickness. If an enhanced geothermal system is unable to provide sufficient flow, basement rocks may be targeted.

If basement rocks have flow geometries representative of parallel vertical fractures with 30 m spacing, the length of time that these reservoirs are able to meet Cornell's heating objectives is expected to be about 8 times greater than a Trenton-Black River (TBR) reservoir with a similar geometry. For the Cornell site, well costs for basement rocks at 3 - 3.5 km depth are expected to be only 2 - 3 times more than TBR well costs. Therefore, developing a basement rock reservoir is economically more attractive than developing the TBR reservoir, if the assumed fracture flow geometry can be obtained in basement rocks.

Both the Trenton-Black River and basement reservoirs have potential to meet Cornell district heating objectives. Future work may consider an integrated analysis and optimization of a surface utilization system with the more favorable geothermal reservoir production scenarios

presented in this paper. The results in this paper constrain the surface utilization scenarios that may be viable for each reservoir target. For example, Gustafson et al. (2019) present optimal well flow rates for the Cornell system for fixed reservoir production temperatures and reinjection temperatures. They show that to meet exactly 20% of the Cornell campus demand, for a supply temperature of 70 °C and reinjection of 30 °C, the optimal well flow rate is about 30 kg/s. That production scenario is unlikely to result in a sufficiently attractive useful life span using the TBR reservoir (Figure 4.7), so alternative utilization designs may be needed.

#### **4.8 Conclusions**

This paper presents methods of uncertainty analysis for geothermal reservoir models of porous media and fracture-dominated systems. Those stochastic methods support probabilistic decision making in the planning phase of a geothermal project. Uncertainty in the available geologic properties and temperatures may be large in this phase, and we evaluate the effect of such uncertainty on meeting direct-use heating objectives for the Cornell University main campus. We propagate uncertainty in available geologic and temperature data through analytical geothermal reservoir models to evaluate the joint probability that a reservoir-production scenario combination will achieve a produced temperature of at least 60 °C and heat production of at least 5 MW<sub>th</sub>. Results from analytical reservoir models were used as a screening method with which to inform the selection of more favorable production scenarios for modeling with more computationally demanding numerical models.

Propagating uncertainties in the data through reservoir simulations allows for probabilistic forecasts of the reservoir's ability to meet heating demands and inform decisions regarding surface utilization operations. For the Cornell case study, uncertainty in the initial

sedimentary reservoir target rock temperatures was important to consider in the probability of meeting desired building supply temperatures. For hotter basement rocks, uncertainty in the initial rock temperature was unimportant to meet Cornell's campus objectives. Both sedimentary and basement targets were found to have viable production scenario combinations that could meet the needs of the district heating system, conditional on the reservoir flow geometries we evaluated. Future work can address an integrated subsurface-surface geothermal system for economic analysis and optimization using the results provided in this paper.

#### **4.9 Acknowledgements**

This research was funded in part by the U.S. Department of Energy under award number DE-EE0008103.

The author would like to thank several coauthors of this work: Teresa Jordan, Jood Al Aswad, J. Olaf Gustafson, Jefferson Tester, and Jerry Stedinger. The authors thank several additional members of the Cornell University Deep Direct Use project team, particularly Steve Beyers, Tasnuva Ming Khan, and Frank Horowitz. We also thank Tom Pasquini for sharing data collected on basement rock in the Adirondack Mountains.

#### **4.10 Disclaimer**

The information, data, or work presented herein was funded in part by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

#### **4.11 Code and Data Availability, and Software Credits**

Numerical modeling was completed using the TOUGH2 software, version 2.0 (Pruess et al., 2012) as implemented in PetraSim (Thunderhead Engineering, 2018). Analytical reservoir models and the wellbore heat transfer model were completed using the GEOPHIRES software available on Github (Beckers and McCabe, 2018; Beckers, 2016) in Python version 2.7.15. Modifications were made to the GEOPHIRES software to complete Monte Carlo analysis, and add additional reservoir modeling functionalities. The modified software is available upon written request to Jared Smith (jds485@cornell.edu).

Statistical analyses were completed using R version 3.5.0 (R Core Team, 2018) and the packages vioplot (Adler, 2005), doParallel (Microsoft Corporation and S. Weston, 2017), abind (Plate and Heiberger, 2016), dataframes2xls (van Steen, 2016), readxl (Wickham and Bryan, 2018), and Hmisc (Harrell Jr. et al., 2018), sp (Pebesma and Bivand, 2005), rgdal (Bivand, Keitt, and Rowlingson, 2017), rgeos (Bivand and Rundel, 2017), raster (Hijmans, 2017), and GISTools (Brunsdon and Chen, 2014).

## REFERENCES

- Abbasi, M., M. Mansouri, A. Daryasafar, and M. Sharifi. (2019). Analytical model for heat transfer between vertical fractures in fractured geothermal reservoirs during water injection. *Renewable Energy*, 130. Pp. 73 - 86.
- Abdulagatov, I.M., Z. Z. Abdulagatova, S. N. Kallaev, A. G. Bakmaev, and P. G. Ranjith. (2014). Thermal-diffusivity and heat-capacity measurements of sandstone at high temperatures using laser flash and DSC methods. *Int J Thermophys*. DOI 10.1007/s10765-014-1829-4
- Adler, D.. (2005). vioplot: Violin plot. R package version 0.2. <http://wsopuppenkiste.wiso.uni-goettingen.de/~dadler>
- Akar, S. and K.R. Young. (2015). Assessment of new approaches in geothermal exploration decision making. *Proceedings of the 40<sup>th</sup> Workshop on Geothermal Reservoir Engineering*. Stanford, California. SGP-TR-204.
- Aliyu, M.D., and H.-P. Chen. (2017). Sensitivity analysis of deep geothermal reservoir: Effect of reservoir parameters on production temperature. *Energy*, 129. Pp. 101 - 113.
- Asai, P., P. Panja, J. McLennan, and J. Moore. (2018). Performance evaluation of enhanced geothermal system (EGS): Surrogate models, sensitivity study and ranking key parameters. *Renewable Energy*, 122. Pp. 184 - 195.
- Bassiouni, Z.. (1994). Theory, Measurement, and Interpretation of Well Logs. Henry L. Doherty Memorial Fund of AIME, Society of Petroleum Engineers. Richardson, TX.
- Beckers, K.J.H.F.. (2016). Low-temperature geothermal energy: Systems modeling, reservoir simulation, and economic analysis. Cornell University Ph.D. Thesis. 272 p.
- Beckers, K. and K. McCabe. (2018). GEOPHIRES. Online GitHub Repository. <https://github.com/kfbeckers/GEOPHIRES>.
- Bedre, M. and B. Anderson. (2012). Sensitivity analysis of low-temperature geothermal reservoirs: Effect of reservoir parameters on the direct use of geothermal energy. *GRC Transactions*, 36. Pp. 1255 - 1262.
- Bivand, R., T. Keitt, and B. Rowlingson. (2017). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.2-16. <https://CRAN.R-project.org/package=rgdal>
- Bivand, R. and C. Rundel. (2017). rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.3-26. <https://CRAN.R-project.org/package=rgeos>
- Blackwell, David, M. Richards, Z. Frone, J. Batir, A. Ruzo, R. Dingwall, and M. Williams. (2011). Temperature at depth maps for the conterminous US and geothermal resource estimates, *GRC Transactions*, 35. Pp. 1545 - 1550.
- Breede, K., K. Dzebisashvili, X. Liu, and G. Falcone. (2013). A systematic review of enhanced (or engineered) geothermal systems: past, present and future. *Geothermal Energy*, 1(1). 4. <https://doi.org/10.1186/2195-9706-1-4>
- Brunsdon, C. and H. Chen. (2014). GISTools: Some further GIS capabilities for R. R package version 0.7-4. <https://CRAN.R-project.org/package=GISTools>
- Camp, E.. (2017). Repurposing Petroleum Reservoirs for Geothermal Energy: A Case Study of the Appalachian Basin: Cornell University, Ph.D. Dissertation, p. 1–219.
- Camp, E., and T. Jordan. (2017). Feasibility study of repurposing Trenton--Black River gas fields for geothermal heat extraction, southern New York: *Geosphere*, 13. p. GES01230-1–14.

- Camp, E.R., T.E. Jordan, M.J. Hornbach, and C.A. Whealton. (2018). A probabilistic application of oil and gas data for exploration stage geothermal reservoir assessment in the Appalachian Basin. *Geothermics*, 71. p. 187-199.
- Carter, K.M., and D.J. Soeder. (2015). Reservoir porosity and permeability. In Patchen, D.G. and K.M. Carter (eds.). A geologic play book for Utica Shale Appalachian basin exploration, Final report of the Utica Shale Appalachian basin exploration consortium. p. 141-159, Available from: <http://www.wvgs.wvnet.edu/utica>.
- Carter, L.S., S.A. Kelley, D.D. Blackwell, and N.D. Naeser. (1998). Heat flow and thermal history of the Anadarko Basin, Oklahoma. *AAPG bulletin*, 82. p. 291–316.
- Cheng, W.-L., C.-L. Wang, Y.-L. Nian, B.-B. Han, and J. Liu. (2016). Analysis of influencing factors of heat extraction from enhanced geothermal systems considering water losses. *Energy*, 115. Pp. 274 - 288.
- Cornell. (1959). Guide book geology of the Cayuga Lake Basin. *New York State Geological Association*, 31. p. 1 – 36.
- Cornell University. (2016a). Appalachian Basin Play Fairway Analysis: Improvements in 2016 to Thermal Quality Analysis in Low-Temperature Geothermal Play Fairway Analysis (GPFA-AB) [data set]. Retrieved from <https://gdr.openei.org/submissions/879>
- Cornell University. (2016b). Natural sedimentary reservoirs data geothermal play fairway analysis 2016 revision [data set]. Retrieved from <https://gdr.openei.org/submissions/881>
- Cornell University. (2017). Final Report: Low Temperature Geothermal Play Fairway Analysis for the Appalachian Basin. Retrieved from <https://gdr.openei.org/submissions/899>
- Daniilidis, A., B. Alpsy, and R. Herber. (2017). Impact of technical and economic uncertainties on the economic performance of a deep geothermal heat system. *Renewable Energy*, 114. Pp. 805 - 816.
- Doe, T., R. McLaren, and W. Dershowitz. (2014). Discrete fracture network simulations of enhanced geothermal systems. *Proceedings of the 39<sup>th</sup> Workshop on Geothermal Reservoir Engineering*. Stanford, CA. SGP-TR-202.
- Dresel, P.E., 1985, The geochemistry of oilfield brines from western Pennsylvania: Pennsylvania State University, State College, PA, 237 p.
- Frone, Z., and D. Blackwell. (2010). Geothermal map of the northeastern United States and the West Virginia thermal anomaly. *GRC Transactions*, 34. Pp. 339 - 344.
- Fox, D.B., D. Sutter, K.F. Beckers, M.Z. Lukawski, D.L. Koch, B.J. Anderson, and J.W. Tester. (2013). Sustainable heat farming: Modeling extraction and recovery in discretely fractured geothermal reservoirs. *Geothermics*, 46. Pp. 42 - 54.
- Fox, D.B., D.L. Koch, and J.W. Tester. (2016). An analytical thermohydraulic model for discretely fractured geothermal reservoirs. *Water Resour. Res.*, 52. Pp. 6792–6817. doi:10.1002/2016WR018666
- Garg, S.K., and J. Combs. (2015). A reformulation of USGS volumetric “heat in place” resource estimation method. *Geothermics*, 55. Pp. 150 - 158.
- Gass, T.E.. (1982). The geothermal heat pump. *GRC Bulletin*, 11. Pp. 3 - 8.
- Goumas, M.G., V.A. Lygerou, and L.E. Papayannakis. (1999). Computational methods for planning and evaluating geothermal energy projects. *Energy Policy*, 27. Pp. 147 - 154.
- Grant, M. A., J. Clearwater, J. Quinão, P.F. Bixley, and M. Le Brun. (2013). Thermal stimulation of geothermal wells: A review of field data. *Proceedings of the 38<sup>th</sup> Workshop on Geothermal Reservoir Engineering*. Stanford, CA. 7 p.

- Gringarten, A.C., P.A. Witherspoon, and Y. Ohnishi. (1975). Theory of heat extraction from fractured hot dry rock. *Journal of Geophysical Research*, 80(8). Pp. 1120-1124.
- Gudmundsson, J.S., D.H. Freeston, and P.J. Lienau. (1985). The Lindal diagram. *GRC Transactions*, 9. Pp. 15 - 19.
- Gustafson, J.O., J.D. Smith, S.M. Beyers, J.A. Al Aswad, T.E. Jordan, and J.W. Tester. (2018). Earth source heat: Feasibility of deep direct use of geothermal energy on the Cornell campus. *GRC Transactions*, 42.
- Gustafson, J.O., J.D. Smith, S.M. Beyers, J.A. Al Aswad, T.E. Jordan, J.W. Tester, and T.M. Khan. (2019). Risk reduction in geothermal deep direct-use development for district heating: A Cornell University case study. *Proceedings of the 44st Workshop on Geothermal Reservoir Engineering*, Stanford, CA, February 11-13.
- Hadi, J., P. Quinlivan, G. Ussher, O. Alamsyah, B. Promono, and A. Masri. (2010). Resource risk assessment in geothermal greenfield development: An economic implications. *Proceedings World Geothermal Congresss 2010*. Bali, Indonesia.
- Harrell, F.E. Jr., et al. (2018). Hmisc: Harrell Miscellaneous. R package version 4.1-1. <https://CRAN.R-project.org/package=Hmisc>
- Held, S., A. Genter, T. Kohl, T. Kölbl, J. Sausse, and M. Schoenball. (2014). Economic evaluation of geothermal reservoir performance through modeling the complexity of the operating EGS in Soultz-sous-Forêts. *Geothermics*, 51. Pp. 270 - 280.
- Hijmans, R.J. (2017). raster: Geographic Data Analysis and Modeling. R package version 2.6-7. <https://CRAN.R-project.org/package=raster>
- Horowitz, F. G., J.D. Smith, and C.A. Wheaton. (2015). One dimensional conductive geothermal Python code. Retrieved August 1, 2017, from <https://bitbucket.org/geothermalcode/onedimensionalgeothermalheatconductionmodel.git>
- Huang, W., W. Cao, and F. Jiang. (2017). Heat extraction performance of EGS with heterogeneous reservoir: A numerical evaluation. *International Journal of Heat and Mass Transfer*, 108. Pp. 645 - 657.
- Jennings, J. and F. Lucia. (2003). Predicting permeability from well logs in carbonates with a link to geology for interwell permeability mapping. *SPE Reserv. Eval. Eng.*, 6. pp. 215–226. doi:10.2118/84942-PA
- Johnsson, M.J.. (1986). Distribution of maximum burial temperatures across northern Appalachian Basin and implications for Carboniferous sedimentation patterns. *Geology*, 14. Pp. 384 - 387.
- Jordan, T.E., K. Tamulonis, V. Lin, and A. Pérez. (2012). Task 4: Target Horizon Geology and Physical-Chemical Characteristics in Jacobi, R., et al., CO2 Sequestration: Capacity, Security, & Enhanced Gas Recovery in Central New York State: Final Technical Report to New York State Energy Research and Development Authority.
- Kolkas, M.M., and G.M. Friedman. (2007). Brine disposal in deep geologic formations of the Cambro-Ordovician (Sauk sequence) of New York: implications for new salt-cavern gas storage reservoirs. *Northeastern Geology and Environmental Sciences*, 29. p. 121.
- Lachenbruch, A.H.. (1970). Crustal temperature and heat production: Implications of the linear heat-flow relation. *J. of Geophys. Res.*, 75(17). Pp. 3291 - 3300.
- Lidiak, E.G., and W.J. Hinze. (1993). Grenville Province in the subsurface of Eastern United States. In J.C. Reed Jr., M.E. Bickford, R.S. Houston, P.K. Link, D.W. Rankin, P.K. Sims, and W.R. VanSchmus (eds.). Precambrian: Conterminous U.S.. The geology of North America, (C-2). Boulder, CO. Geological Society of America. Pp. 335 - 461.

- Liu, S., J.C. Lu, D.W. Kolpin, and W.Q. Meeker. (1997). Analysis of environmental data with censored observations. *Environ. Sci. Technol.*, 31. Pp. 3358 - 3362.
- Lowry, T.S., J.T. Finger, C.R. Carrigan, A. Foris, M.B. Kennedy, T.F. Corbett, C.A. Doughty, S. Pye, and E.L. Sonnenthal. (2017). Reservoir maintenance and development task report for the DOE geothermal technologies office GeoVision study. Sandia Report SAND2017-9977. 81 p.
- Lucia, F.J.. (2007). Carbonate Reservoir Characterization, Second. ed. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-540-72742-2
- Lukawski, M.Z., R.L. Silverman, and J.W. Tester. (2016). Uncertainty analysis of geothermal well drilling and completion costs. *Geothermics*, 64. Pp. 382 - 391.
- Lynch, R.S., and Castor, T.P., 1983, Auburn Low-temperature Geothermal Well Final Report: NYSERDA report 84-18
- Manning, C. E., and S. E. Ingebritsen. (1999). Permeability of the continental crust: Implications of geothermal data and metamorphic systems. *Rev. Geophys.*, 37(1), pp. 127 – 150. doi:10.1029/1998RG900002.
- Matsumoto, M.R., Atkinson, J.F., Bunn, M.D., and Hodge, D.S., 1996, Disposal/Recovery Options for Brine Waters from Oil and Gas Production in New York State: NYSERDA, v. Report 96-4, p. 1–168.
- McLelland, J.M., B.W. Selleck, and M.E. Bickford. (2010). Review of Proterozoic evolution of the Grenville Province, its Adirondack outlier, and the Mesoproterozoic inliers of the Appalachians. In R.P. Tollo, M.J. Bartholomew, J.P. Hibbard, and P.M. Karabinos (eds.). From Rodinia to Pangea: The Lithotectonic Record of the Appalachian Region. Geological Society of America Memoir, 206. Pp. 1 - 29. doi: 10.1130/2010.1206(02).
- McLelland, J.M., B.W. Selleck, and M.E. Bickford. (2013). Tectonic evolution of the Adirondack Mountains and Grenville Orogen inliers within the USA. *Geoscience Canada*, 40. Pp. 318 - 352.
- Microsoft Corporation and S. Weston (2017). doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.11. <https://CRAN.R-project.org/package=doParallel>
- Millington, R.J. and J.P. Quirk. (1961). Permeability of porous solids. *Trans. Faraday Soc.*, 57. Pp. 1200-1207.
- New York State Museum. (2014). FEI\_Whiteman\_core\_analysis\_030120.xls. file available through the New York State Museum (last accessed September 2014).
- Oakley, J.. (2004). Estimating percentiles of uncertain computer code outputs. *Appl. Statist.*, 53(1). Pp. 83 - 93.
- Olasolo, P., M.C. Juárez, J. Olasolo, M.P. Morales, and D. Valdani. (2016). Economic analysis of enhanced geothermal systems (EGS). A review of software packages for estimating and stimulating costs. *Applied Thermal Engineering*, 104. Pp. 647 - 658.
- Pandey, S.N., and A. Chaudhuri. (2017). The effect of heterogeneity on heat extraction and transmissivity evolution in a carbonate reservoir: A thermo-hydro-chemical study. *Geothermics*, 69. Pp. 45 - 54.
- Pebesma, E.J. and R.S. Bivand. (2005). Classes and methods for spatial data in R. R News 5 (2), <https://cran.r-project.org/doc/Rnews/>.
- Plate, T., and R. Heiberger (2016). abind: Combine Multidimensional Arrays. R package version 1.4-5. <https://CRAN.R-project.org/package=abind>

- Pruess, K., C. Oldenburg, and G. Moridis. (2012). TOUGH2 user's guide, version 2. Earth Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley. LBNL-43134.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Ramalingam, A., and S. Arumugam. (2012). Experimental study on specific heat of hot brine for salt gradient solar pond application. *International Journal of ChemTech Research*, 4(3). Pp. 956 - 961.
- Ramey, H.J. Jr.. (1961). Wellbore heat transmission. SPE 36<sup>th</sup> Annual Fall Meeting. April, 1962, pp. 427 - 435.
- Rauch, C., K. Barrie, S.C. Collins, M.J. Hornbach, and C. Brokaw. (2018). Heat flow and thermal conductivity measurements in the northeastern Pennsylvania Appalachian Basin depocenter. *AAPG Bulletin*, 102(11). Pp. 2155 - 2170.
- Robertson, E.C. and B.S. Hemingway. (1995). U.S. Department of the Interior Geological Survey Open File Report 95-622
- Rowan, E.L.. (2006). Burial and thermal history of the Central Appalachian Basin, based on three 2-D models of Ohio, Pennsylvania, and West Virginia. USGS Open File Report 2006-1019.
- Selvadurai, A.P.S., M.J. Boulon, and T.S. Nguyen. (2005). The permeability of an intact granite. *Pure Appl. Geophys.* 162. Pp. 373–407. DOI 10.1007/s00024-004-2606-2
- Sener, A.C., and J.R. van Dorp. Evolution of technical and economical decision making in geothermal energy projects. *GRC Transactions*, 29. Pp. 475 - 481.
- Siegel, D.I., R.J. Szustakowski, and S. Frape. (1990). Regional appraisal of brine chemistry in the Albion Group sandstones (Silurian) of New York: Pennsylvania and Ohio. *Association of Petroleum Geochemical Explorationists Bulletin*, 6. p. 66–77.
- Simmons, G.. (1964). Gravity survey and geological interpretation, northern New York: *Geological Society of America Bulletin*, 75. p. 81–98.
- Sinclair, L.K., J.D. Smith, D.L. Koch, J.W. Tester, and J.F.H Thompson. (2018). Hydraulic conductivity and residence time in fractured rock: Florence Copper in situ leaching case study. *CIM Journal*, 9(3). Pp. 151 - 164.
- Skeen, J.C.. (2010). Basin analysis and aqueous chemistry of fluids in the Oriskany Sandstone, Appalachian Basin, USA, M.S. Thesis, West Virginia State University, Morgantown, WV.
- Smith, J.D.. (2016). Analytical and geostatistical heat flow modeling for geothermal resource reconnaissance applied in the Appalachian Basin. Cornell University. M.S. Thesis.
- Smith, L.B.. (2006). Origin and reservoir characteristics of Upper Ordovician Trenton-Black River hydrothermal dolomite reservoirs in New York. *AAPG Bulletin*, 90. p. 1691.
- Smith, L., C. Lugert, S. Bauer, B. Ehgartner, R. Nyahay. (2005). Final report: Systematic technical innovations initiative brine disposal in the Northeast. NETL Report DE-FC26-01NT41298.
- Smith, L., R. Nyahay, and B. Slater. (2010). Integrated Reservoir Characterization of the Subsurface Cambrian and Lower Ordovician Potsdam, Galway and Theresa Formations in New York. Albany, NY, New York State Energy Research and Development Authority.

- Snyder, D.M., K.F. Beckers, K.R. Young, and B. Johnston. (2017). Analysis of geothermal reservoir and well operational conditions using monthly production reports from Nevada and California. *GRC Transactions*, 41.
- Stutz, G.R., E. Shope, G.A. Aguirre, J. Batir, Z. Frone, M. Williams, T.J. Reber, C.A. Whealton, J.D. Smith, M.C. Richards, D.D. Blackwell, J.W. Tester, J.R. Stedinger, and T.E. Jordan. (2015). Geothermal energy characterization in the Appalachian Basin of New York and Pennsylvania. *Geosphere*, 11(5). Pp. 1291 - 1304.
- Thunderhead Engineering. (2018). PetraSim User Manual. 137 p.
- Trainor-Guitton, W.J., Ramirez, A., Ziagos, J., Mellors, R., Roberts, J., 2013. An initial value of information (VOI) framework for geophysical data applied to the exploration of geothermal energy. *Proceedings of the 38<sup>th</sup> Workshop on Geothermal Reservoir Engineering*. Stanford, CA. 8 p.
- Valentino, B. (2016). Cornell University internal report: Petrological and Geophysical Analysis of the Subsurface Basement Rocks in Central New York (Mohawk Valley and Finger Lakes Regions)
- Valentino, D.W., J.D. Valentino, J.R. Chiarenzelli, and R.W. Inclima. (2016). Faults and fracture systems in the basement rocks of the Adirondack Mountains, New York. Ch. 8 in C. Northrop, J. Chiarenzelli, and B. Selleck (eds.). *The Adirondack Journal of Environmental Studies*, Vol 21. Kelly Adirondack Center, Union College, New York. pp. 101 - 118.
- van Steen, G.. (2016). dataframes2xls: Write Data Frames to Xls Files. R package version 0.4.7. <https://CRAN.R-project.org/package=dataframes2xls>
- Waller, R.M., J.T. Turk, and R.J. Dingman. (1978). Potential effects of deep-well waste disposal in Western New York. US Geol. Surv. Professional Paper 1053. p. 1–39.
- Walsh, S.D.C., N. Garapati, A.M.M. Leal, and M.O. Saar. (2017). Calculating thermophysical fluid properties during geothermal energy production with NESS and Reaktoro. *Geothermics*, 70. Pp. 146 - 154.
- Waples, D.W., and J.S. Waples. (2004). A review and evaluation of specific heat capacities of rocks, minerals, and subsurface fluids. Part 1: Minerals and nonporous rocks. *Natural Resources Research*, 13(2). Pp. 97 - 122.
- Watanabe, N., W. Wang, C.I. McDermott, T. Taniguchi, and O. Kolditz. (2010). Uncertainty analysis of thermo-hydro-mechanical coupled processes in heterogeneous porous media. *Comput. Mech.*, 45. Pp. 263–280. doi:10.1007/s00466-009-0445-9
- Wickham, H. and J. Bryan (2018). readxl: Read Excel Files. R package version 1.1.0. <https://CRAN.R-project.org/package=readxl>
- Wickstrom, L.H., J.D. Gray, and R.D. Stieglitz. (1992). Stratigraphy, structure, and production history of the Trenton Limestone (Ordovician) and adjacent strata in northwestern Ohio. Ohio Division of Geological Survey, Report of Investigations No. 143. 78 p.
- Williams, J.H.. (2010). Evaluation of well logs for determining the presence of freshwater, saltwater, and gas above the Marcellus Shale in Chemung, Tioga, and Broome Counties, New York. USGS SIR 2010-5224. 35 p.
- Witter, J.B., W.J. Trainor-Guitton, and D.L. Siler. Uncertainty and risk evaluation during the exploration stage of geothermal development: A review. *Geothermics*, 78. Pp. 233 - 242.
- Zeng, Y.-C., Z. Su, and N.-Y. Wu. (2013). Numerical simulation of heat production potential from hot dry rock by water circulating through two horizontal wells at Desert Peak geothermal field. *Energy*, 56. Pp. 92 - 107.

## CHAPTER 5

### CONCLUSIONS AND FUTURE RESEARCH OPPORTUNITIES

#### **5.0 Summary of Thesis Research Progression**

This thesis presents methods for exploratory spatial data analysis to identify discordant observations in geothermal datasets, and methods for uncertainty propagation for geothermal resource assessment and reservoir models. Chapters 2, 3, and 4 in aggregate consider the entire phase of geothermal resource exploration, and address how the developed methods may be used to inform decisions in the geothermal project planning phase.

In the exploration phase, oftentimes limited data are available or the data are low-quality because they were not collected for the purpose of geothermal energy projects. These issues can result in large uncertainties in values computed for geothermal variables in resource assessment models. Chapter 2 describes exploratory spatial data analysis procedures that identify discordant observations in datasets of geothermal variables that are derived from available temperature datasets. Removing discordant observations improved the precision of spatial autocorrelation estimates for the derived geothermal variables.

Chapter 3 describes a stochastic framework to propagate uncertainty in those estimated geothermal variables and uncertainty in geologic property values through two geothermal resource assessment models. These models 1) compute uncertainty distributions for temperatures as a function of depth across the region of interest, and 2) use those temperature-depth distributions to compute distributions of stored thermal energy within rock volumes that may serve as geothermal reservoirs. A sensitivity analysis of each model informs which of the input variables' values would be most useful to refine to reduce uncertainty in the output variables. For

the Appalachian Basin region of interest, viewing the results of the sensitivity analysis spatially revealed that the most important input variable to refine depended on location within the basin.

Chapter 4 uses the resulting temperature-depth uncertainty distributions from Chapter 3 in a stochastic evaluation of geothermal reservoir models. These models estimate the distribution of heat production from potential geothermal reservoirs at specific sites. Because stochastic methods are used for resource assessment models and geothermal reservoir models, they can be used to propagate uncertainties in geothermal variables computed in the exploration phase of a geothermal project to probabilistic risk assessments that inform geothermal project planning. The application in Chapter 4 evaluated the probability of potential geothermal reservoirs to meet direct-use heating objectives for the Cornell University campus in Ithaca, NY.

## **5.1 Summary of Major Conclusions**

Geothermal resource assessments in sedimentary basins rely on the abundance of low-quality oil and gas well bottom-hole temperature (BHT) measurements to supplement often limited high-quality temperature-depth information within a region of interest. Chapter 2 of this thesis presents exploratory spatial data analysis (ESDA) procedures that detect discordant observations in datasets of geothermal variables derived from BHTs. The surface heat flow is the variable considered for that analysis, but other variables, such as the temperature at a specific depth, may also be considered. The ESDA methods aim to improve the precision of parameters estimated in spatial regressions of the derived geothermal variables. For example, in Chapter 2, such regressions would use the values of surface heat flow computed at well locations to estimate the surface heat flow across the region of interest. Because spatial regression is the intended use of the derived heat flow values at well locations, measures of surface heat flow

spatial autocorrelation are used as statistical diagnostics to evaluate the effect of applying the ESDA procedures to the heat flow values at well locations.

For the Appalachian Basin BHT dataset, applying an ESDA procedure that evaluated the deviation of heat flow values from the median value within a local neighborhood was useful to determine thresholds for BHT measurement depths to be deep enough to represent conduction-dominated heat transfer. Removing heat flow values that were computed from BHTs measured at shallow depths resulted in significant improvements to the precision of surface heat flow spatial autocorrelation estimates. Applying an ESDA procedure for local spatial outlier detection, and removing points detected as potential outliers, further stabilized the autocorrelation estimates. These results support that ESDA procedures with associated spatial autocorrelation statistical diagnostics should be a standard component of geothermal resource assessments. The statistical diagnostics also provide objective means to evaluate hydrogeological hypotheses about the depths of shallow groundwater aquifers and geophysical hypotheses about the heat flow values in a region. Such hypotheses have been used without statistical support in other geothermal resource assessment studies to remove shallow BHT measurements that were likely affected by groundwater advection, and to remove observations that did not conform to geophysical expectations.

For the Appalachian Basin, evaluating spatial correlation of the surface heat flow within geologic sub-regions of the basin was important to reveal non-stationary spatial correlation structures that should be modeled in spatial regressions. With the analyses completed, it is unclear if differences in spatial correlation structure result primarily from physical heterogeneities in basin characteristics, or are primarily a result of differences in the quality of BHT datasets throughout the basin. Regardless, this finding suggests that computing spatial

correlation within sub-regions of a basin should be a routine practice in geothermal resource assessments to evaluate potential non-stationarity in the spatial correlation structures of the data.

Chapter 3 presents stochastic methods to propagate uncertainty in spatial predictions of the surface heat flow through two geothermal resource assessment models. For the Appalachian Basin, the effect of a non-stationary spatial correlation structure was particularly important to compare the favorability of sites based on the uncertainty in stored thermal energy contained within reservoirs across a region where projects may be proposed. Propagating spatial prediction uncertainties in the surface heat flow through to estimates of stored thermal energy was useful to discover areas within reservoirs that have more precise energy estimates as a result of where wells were drilled near the reservoir.

Sensitivity analyses of the two geothermal resource assessment models inform which of the temperature and geologic variables would be most useful to refine to reduce uncertainty in the geothermal resource estimates for sites of interest. To reduce uncertainty in calculations of the surface heat flow at well locations in the Appalachian Basin, acquisition of reliable temperature measurements or thermal conductivity data trade off as the most useful. Generally, BHTs are most useful to acquire when they are measured at shallower depths, and thermal conductivities are most useful to acquire when BHTs are measured at deeper depths. To reduce uncertainty in calculations of stored thermal energy, refining estimates of reservoir thicknesses and temperature within a reservoir trade off as most useful. Other geologic variables, such as the rock porosity, reservoir depth, rock density and rock heat capacity had relatively little effect on the uncertainty in stored thermal energy for this Appalachian Basin dataset.

Chapter 4 presents an evaluation of heat production from Cornell campus reservoirs in sedimentary and basement rocks. Probabilistic assessment of analytical geothermal reservoir

models was useful to select production scenarios that were more likely to meet direct-use heating objectives. Numerical models were useful to evaluate reservoir fluid flow complexities that analytical models did not consider. The uncertainties in initial rock temperature and the reservoir flow geometry (porous media vs. fracture-dominated) were important considerations to determine the success of a geothermal project at Cornell. Uncertainty in reservoir flow geometry was evaluated as scenarios through use of different reservoir models. A stochastic evaluation of permeability networks and fracture spacings could be valuable for a more complete representation of reservoir flow geometry uncertainty in the resource production at this site.

## **5.2 Future Research Opportunities**

This thesis presents methods that target the resource exploration and project planning phases of geothermal projects. The following two sections present future research opportunities for each phase that build off of the methods and findings presented in this thesis.

### **5.2.1 Geothermal Resource Exploration Phase**

The exploratory spatial data analysis (ESDA) procedures in Chapter 2 treat the estimates of surface heat flow at well locations as constants. These point estimates of heat flow actually have uncertainty due to uncertainty in the BHT measurement and uncertainties in geologic property values that are required to compute the surface heat flow using the heat conduction model described in Chapter 3. Future research can evaluate ESDA procedures that account for the uncertainty in the point estimates at each well relative to its local neighbors. Results from the Chapter 3 sensitivity analysis of the surface heat flow at well locations suggest that considering the uncertainty in surface heat flow point estimates within ESDA procedures will matter more

for those wells whose local neighbors have BHT measurements taken at substantially different depths. In these regions, the effect of the geologic property uncertainties (e.g. thermal conductivity values) and BHT correction uncertainty on the surface heat flow estimates will be different among neighboring wells, which can influence whether or not a point is identified as being spatially discordant. Because of the observed effect of BHT measurement depth on the surface heat flow uncertainty, future research can evaluate different definitions of spatial neighborhoods that consider BHT depth as an additional criterion.

The uncertainties in the surface heat flow at well locations were also not considered in the laterally stratified kriging geostatistical spatial regression that estimated values of the Appalachian Basin surface heat flow used in Chapter 3. Future research can evaluate spatial regressions that consider the uncertainty in the estimated surface heat flow at well locations as measurement errors. This may be used to obtain kriging prediction uncertainties that account for these errors. The surface heat flow map used in Chapter 3 also did not consider predictor variables that could explain the observed variability in the surface heat flow across the Appalachian Basin. Results from Chapter 3 sensitivity analyses suggest that geologic region and depth to basement rocks affect the values of surface heat flow in the Appalachian Basin. Thus, to further explain variability of the surface heat flow and to reduce prediction uncertainty, future research can evaluate spatial regressions that consider such basin characteristics.

Sensitivity analyses of the surface heat flow at well locations and the stored thermal energy within potential geothermal reservoirs inform which input variables are most useful to obtain to reduce uncertainty at different phases of geothermal project exploration. Future research should be devoted to economic analysis of both the financial resources and time required to collect and analyze additional data, and to improve geothermal resource assessment

models. With such an analysis, optimization of the data to collect and models to refine to maximize uncertainty reduction while minimizing time and financial expenses would be particularly useful to guide research efforts for each site in a region of interest.

Sensitivity analyses at well locations suggested that reducing uncertainty in temperature data would be most beneficial to reduce uncertainty in geothermal resource estimates for many locations across the Appalachian Basin. This could be accomplished through acquisition of new and reliable data, or by improvements to existing models. Acquiring new equilibrium temperature-depth profiles or re-logging temperature-depth profiles in existing wells would likely be more valuable than acquiring additional BHTs. The high-quality information at a single site could be used to update BHT corrections within the neighborhood of that site. Future research that can improve existing stochastic models may evaluate including uncorrected BHT values as minimum bounds of temperature at their depth of measurement. Those uncorrected BHTs could be treated as “greater than” censored observations in resource assessment models.

Uncertainties in the thicknesses of sedimentary formations were important contributors to uncertainty in the surface heat flow and temperatures at depth for Appalachian Basin wells. Sedimentary rock thicknesses in this study were assigned using generalized geologic columns that were defined for broad spatial areas of the Appalachian Basin. Using well log data available throughout the Appalachian Basin, continuous surfaces of formation tops may be estimated using regressions. These surfaces will likely improve site-specific estimates of the formation thicknesses, as well as provide more accurate uncertainty estimates that are derived from the regression models.

The assessment of stored thermal energy in Chapter 3 was limited in this study to potential geothermal reservoirs whose spatial extents were defined primarily by existing

hydrocarbon plays. The formations corresponding to these potential geothermal reservoirs extend beyond those plays, and may be viable geothermal reservoirs even if they were not attractive for hydrocarbon development. Future research can evaluate the spatial extent of formations beyond their hydrocarbon plays to extend the spatial coverage of geothermal resource assessment.

### **5.2.2 Geothermal Project Planning Phase**

The reservoir models evaluated in Chapter 4 did not consider uncertainty in the thickness of the target geothermal reservoirs. Given that reservoir thickness was found to be an important contribution to the uncertainty of stored thermal energy in Chapter 3, future assessments of Cornell reservoirs should evaluate the effect of uncertain reservoir thickness on the probability of meeting direct-use heating objectives.

The probabilistic analysis of reservoir models also did not model the utilization of heat in the Cornell University district heating system. For project risk assessments, the subsurface models used in Chapter 4 may be used in an integrated surface-subsurface geothermal systems analysis. Future research can then optimize production scenarios for the probability of meeting relevant heating objectives, as well as the probability of meeting any economic and environmental objectives for the project.

APPENDIX A  
PERFORMANCE TESTS FOR THE ASYMMETRIC BOXPLOT OUTLIER CRITERION  
PRESENTED IN CHAPTER 2

The asymmetric boxplot outlier criterion provided in Equation 2.1 depended on a parameter,  $k$ , that sets the bounds of the outlier criterion. Typically, one chooses an outlier criterion based on a type I error: the probability of identifying a point as an outlier even though it is drawn from a specified distribution. In this example, the null hypothesis is that the data are distributed according to the distribution specified in Table A.1, Table A.2, and Figure A.1. Type I error rates in Table A.1 were calculated based on Monte Carlo analysis from 100,000 replicates of sample size 25 from the specified distribution. Type I error rates in Table A.2 were calculated based on perfect knowledge of the distribution parameters, which simulates a large sample size. Figure A.1 illustrates the type I error as a function of  $k$  for several null distributions and sample sizes.

For the recommended value of  $k = 3$ , if the data are normally distributed and one has a small sample size ( $n = 25$ ), one would expect to identify about 3% of the dataset as outliers (Table A.1). Thicker tailed distributions, such as the Student  $t$ , will have higher identification rates. Thin-tailed Beta(1,1) (uniform distribution) and Beta(2,2) have high type I error rates in small sample sizes (Table A.1) compared to large sample sizes (Table A.2).

For skewed distributions, selection of an outlier criterion may be informed by the mean squared error (MSE) of the mean, which results from asymmetrically removing extreme outlier values. For regression, if rogue data are expected, error may be beneficial if it results from the removal of rogue observations. If the removed outliers are not rogue, large MSE would be undesirable. We calculate the MSE of the mean based on applying our asymmetric boxplot to

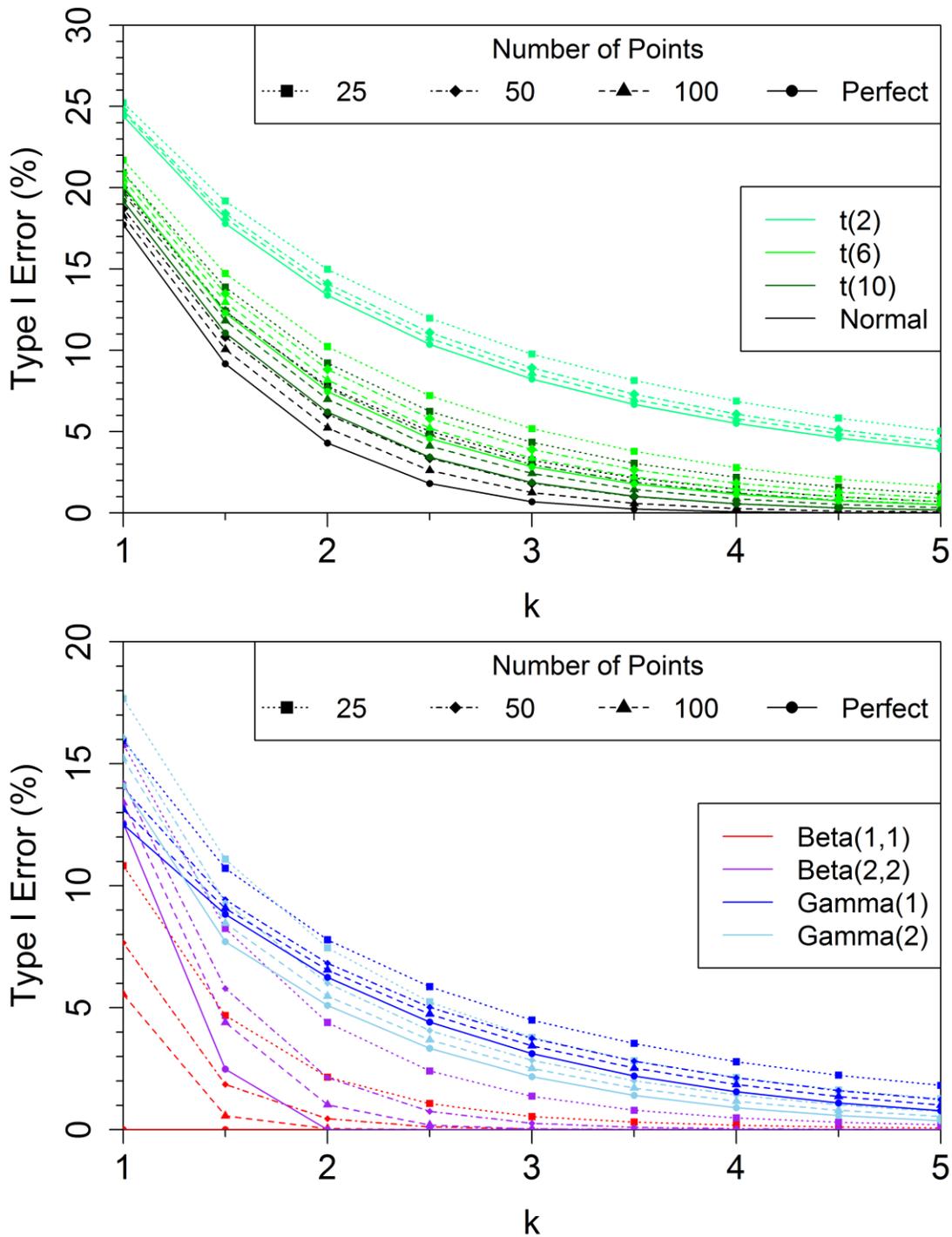
10,000 Monte Carlo replicates of sample size 50 from a lognormal distribution with parameters specified in Figure A.2. The MSE tends to be larger for more skewed distributions and for smaller values of  $k$ . Relatively large skew is needed for the MSE to become appreciable.

**Table A.1:** Type I error (%) for the asymmetric boxplot outlier criterion based on 100,000 replicates of sample size 25. Beta(1,1) is the uniform distribution. Gamma(1) is the exponential distribution. The argument for the gamma distribution is the shape parameter. The argument for the Student t distribution is the shape parameter (degrees-of-freedom), which controls the thickness of the tails. Upper and lower bounds used to define outliers are based on Equation 2.1 for the selected value of  $k$ .

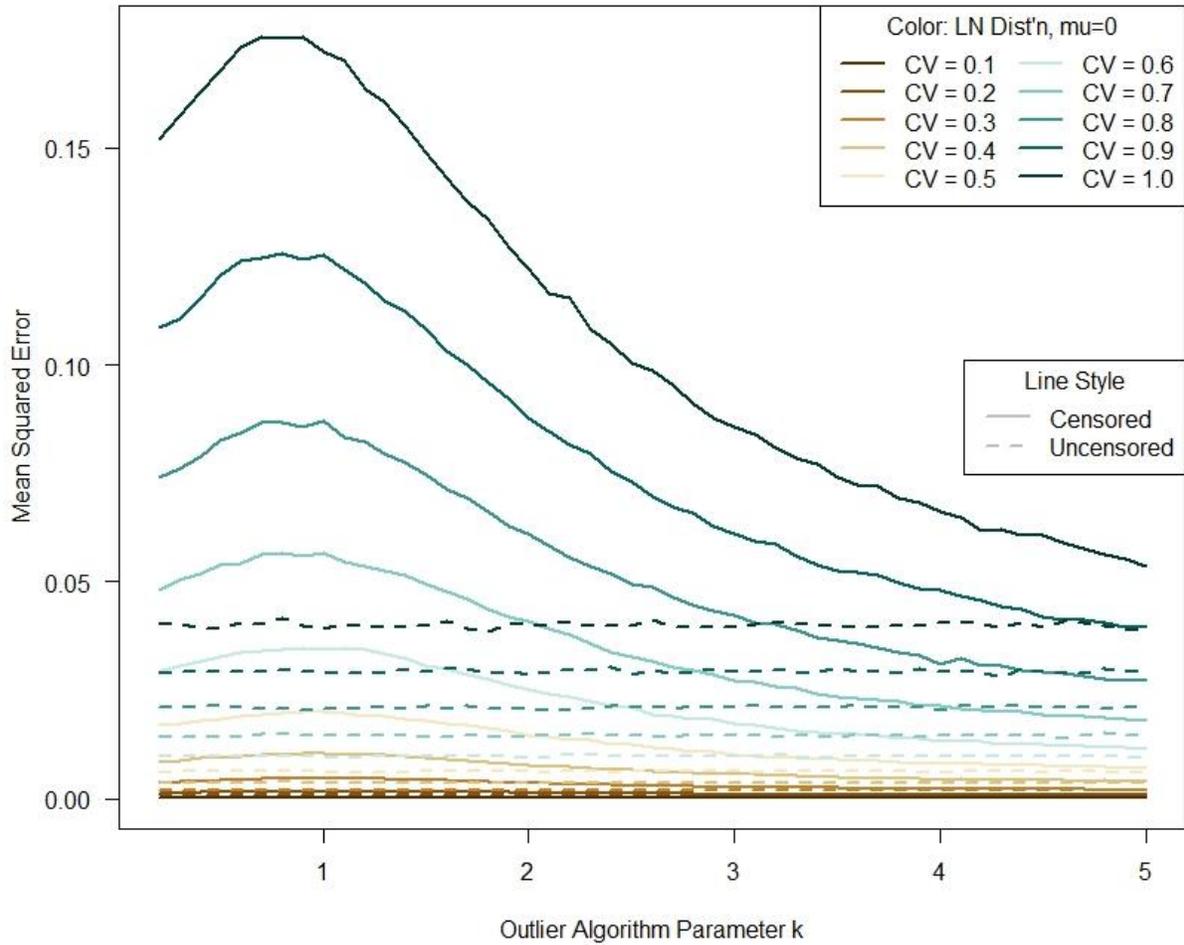
$k$	Distribution							
	Normal	Student t (10)	Student t (6)	Student t (2)	Beta (1, 1)	Beta (2, 2)	Gamma (1)	Gamma (2)
1.0	19.82	20.92	21.70	25.22	10.82	15.79	15.95	17.67
1.5	12.39	13.88	14.73	19.18	4.68	8.25	10.72	11.09
2.0	7.79	9.23	10.23	14.98	2.16	4.40	7.79	7.45
2.5	5.02	6.25	7.22	11.98	1.08	2.41	5.87	5.24
3.0	3.25	4.35	5.18	9.77	0.54	1.38	4.49	3.78
3.5	2.16	3.06	3.79	8.15	0.31	0.80	3.54	2.82
4.0	1.47	2.19	2.79	6.88	0.18	0.48	2.79	2.14
4.5	0.99	1.58	2.08	5.83	0.11	0.30	2.24	1.63
5.0	0.71	1.17	1.62	5.04	0.07	0.20	1.82	1.27

**Table A.2:** Type I error (%) for the asymmetric boxplot outlier criterion based on perfect knowledge of the parameters (large sample). Beta(1,1) is the uniform distribution. Gamma(1) is the exponential distribution. The argument for the gamma distribution is the shape parameter. The argument for the Student t distribution is the shape parameter (degrees-of-freedom), which controls the thickness of the tails. Upper and lower bounds used to define outliers are based on Equation 2.1 for the selected value of  $k$ , and the quantiles are calculated from the population distribution.

$k$	Distribution							
	Normal	Student t (10)	Student t (6)	Student t (2)	Beta (1, 1)	Beta (2, 2)	Gamma (1)	Gamma (2)
1.0	17.73	19.19	20.13	24.41	0.00	12.57	12.50	14.10
1.5	9.18	11.08	12.30	17.80	0.00	2.49	8.84	7.71
2.0	4.30	6.21	7.48	13.40	0.00	0.00	6.25	5.09
2.5	1.82	3.43	4.58	10.37	0.00	0.00	4.42	3.34
3.0	0.70	1.88	2.84	8.23	0.00	0.00	3.13	2.18
3.5	0.24	1.03	1.79	6.67	0.00	0.00	2.21	1.41
4.0	0.07	0.57	1.15	5.51	0.00	0.00	1.56	0.91
4.5	0.02	0.32	0.76	4.62	0.00	0.00	1.10	0.58
5.0	0.01	0.18	0.51	3.92	0.00	0.00	0.78	0.37



**Figure A.1:** Type I error (%) as a function of  $k$  for the asymmetric boxplot outlier criterion based on 100,000 replicates of the symbolized number of points. Upper and lower bounds used to define outliers are based on Equation 2.1 for the selected value of  $k$  for finite sample size, and the quantiles are calculated from the population distribution for perfect information (large sample). Beta(1,1) is the uniform distribution. Gamma(1) is the exponential distribution. The argument for the gamma distribution is the shape parameter. The argument for the Student t distribution is the shape parameter (degrees-of-freedom), which controls the thickness of the tails.

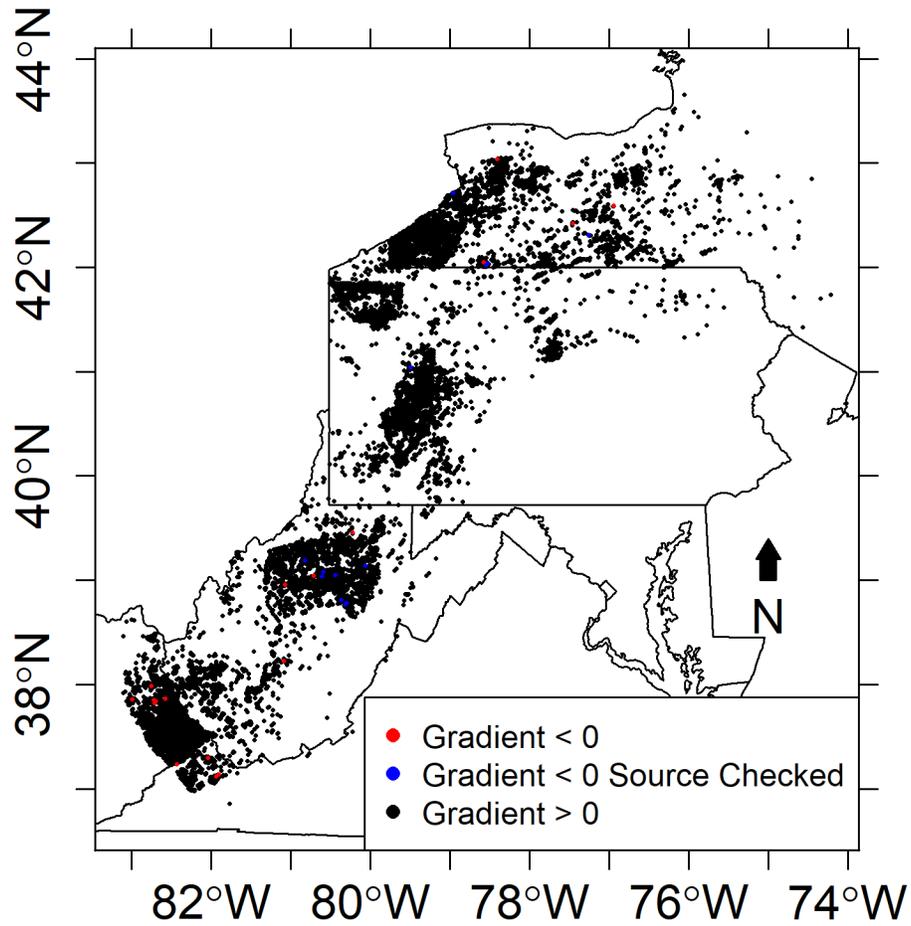


**Figure A.2:** Real-space mean squared error (MSE) of the mean as a function of  $k$  for the asymmetric boxplot outlier criterion based on 10,000 replicates of size 50 from a lognormal distribution with the colored real-space coefficient of variation (CV) and log-space mean of 0. The MSE is provided before and after outlier censoring using the asymmetric boxplot.

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

1) Locations of records with negative geothermal gradients

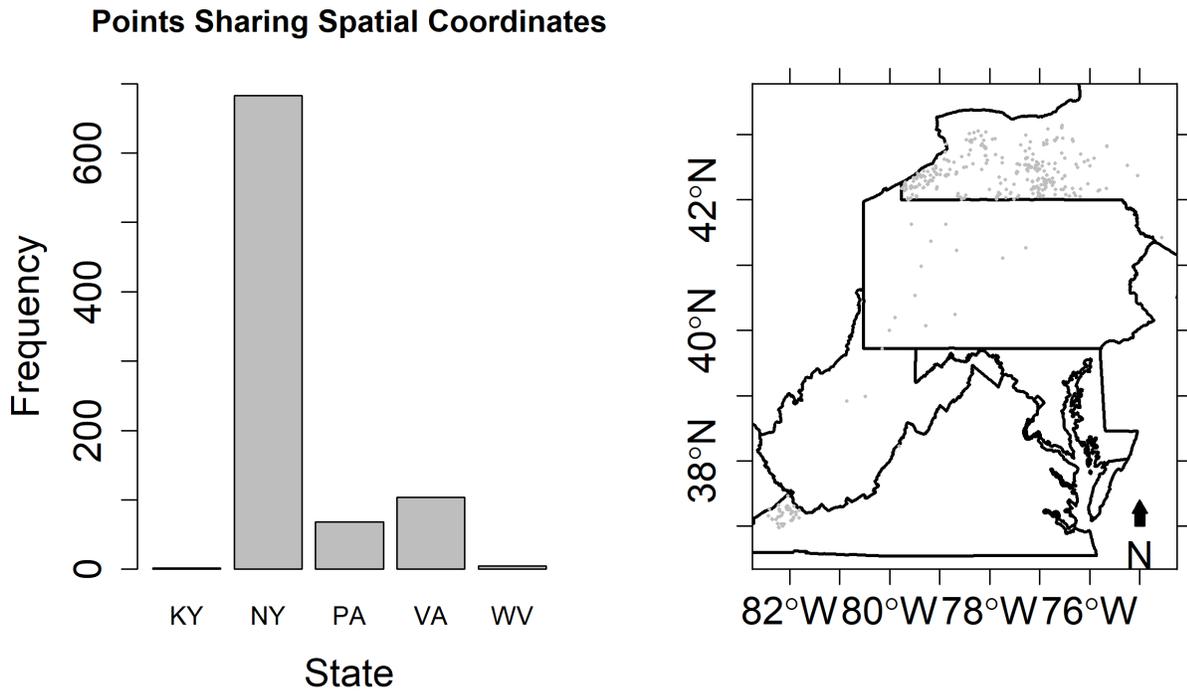


**Figure B.1:** Locations of records with negative geothermal gradients (red), and records for which the original source was inspected (Table 2.1, Table B.1). All other records are plotted in black.

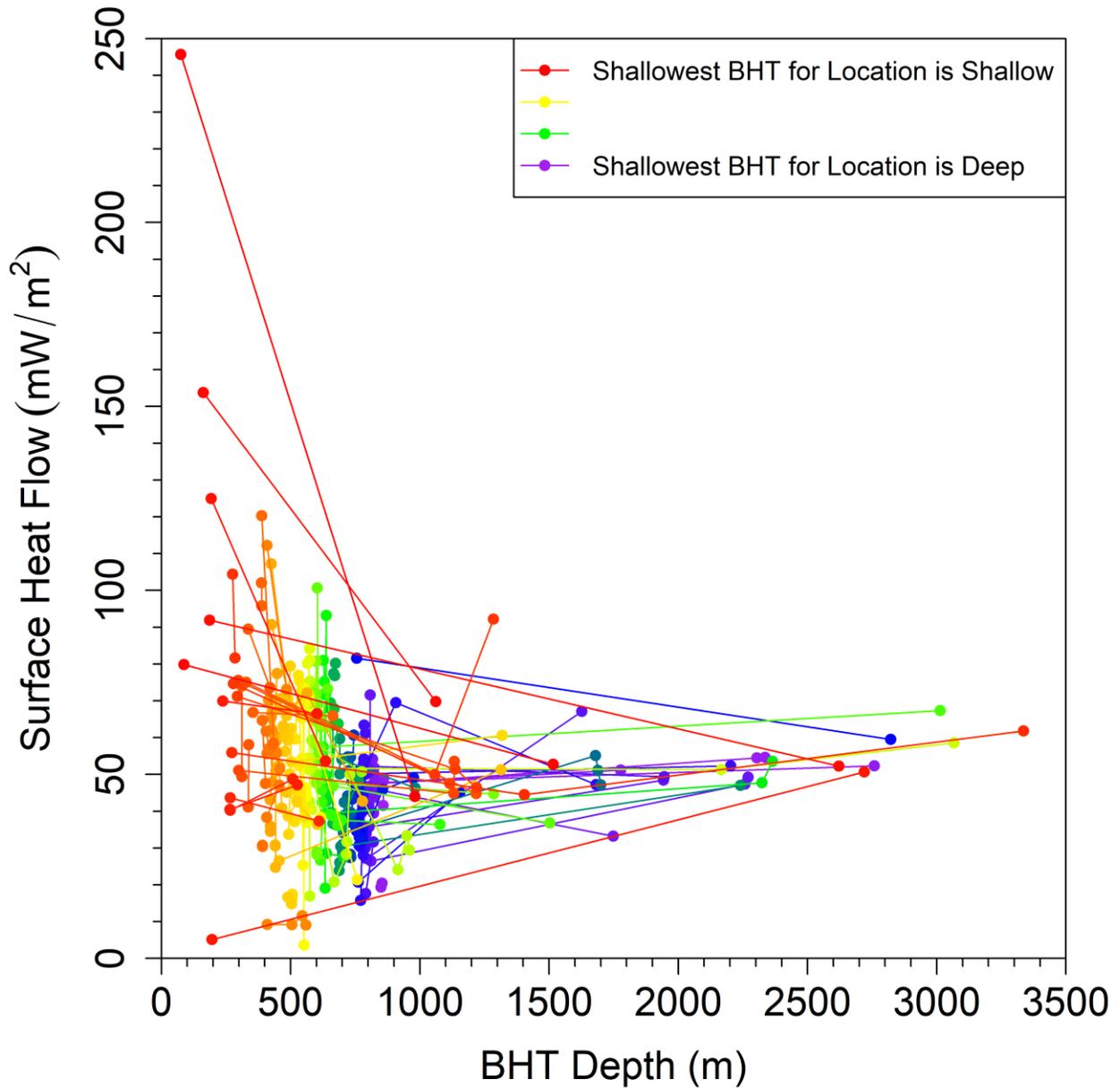
**Table B.1:** State and API numbers for inspected records (blue points in Figure B.1).

State	API
New York	31101215510001
	31009248890000
	31009242440000
	31029069280000
Pennsylvania	37031226260000
West Virginia	47001008240000
	47017037660000
	47041037850000
	47041037910000
	47097026580000
	47097025860000
	47097024970000
	47097025750000
47041016130000	

## 2) Records sharing spatial coordinates



**Figure B.2:** Histogram showing by state the number of records that are in locations (on the map) with multiple BHTs.



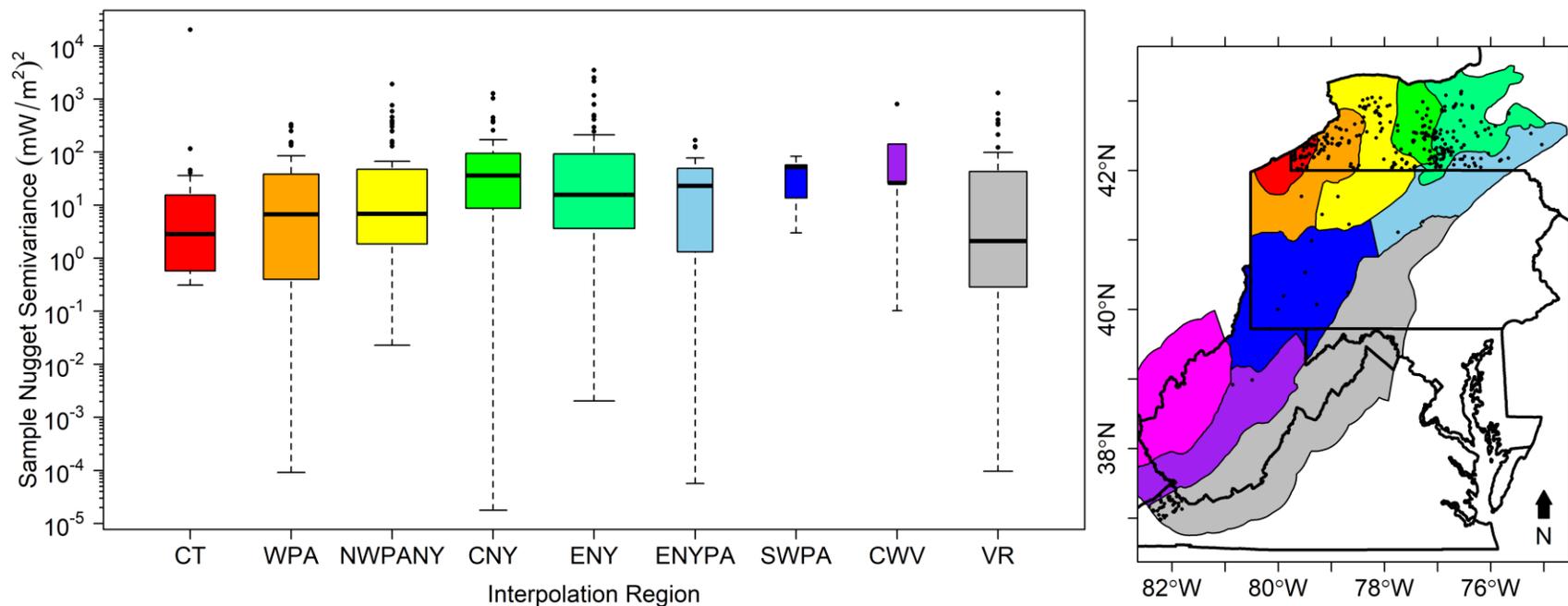
**Figure B.3:** Surface heat flow vs. BHT depth of measurement for locations with multiple BHTs. Lines connect surface heat flow values for the same spatial location. The lines and points are colored by the depth of the shallowest BHT measurement for the location. Generally, the surface heat flow values become more consistent with increasing depth.

### 3) Boxplots of the sample nugget semi-variance for records in the same spatial location

Records in the same spatial location may be used to estimate the nugget semi-variance, or the noise in the dataset at very small separation distances. The sample nugget semi-variance is calculated using Equation B.1 (e.g. Cressie, 1988)

$$s_n^2 = \frac{(Q_{s,i} - Q_{s,j})^2}{2} \quad [\text{B.1}]$$

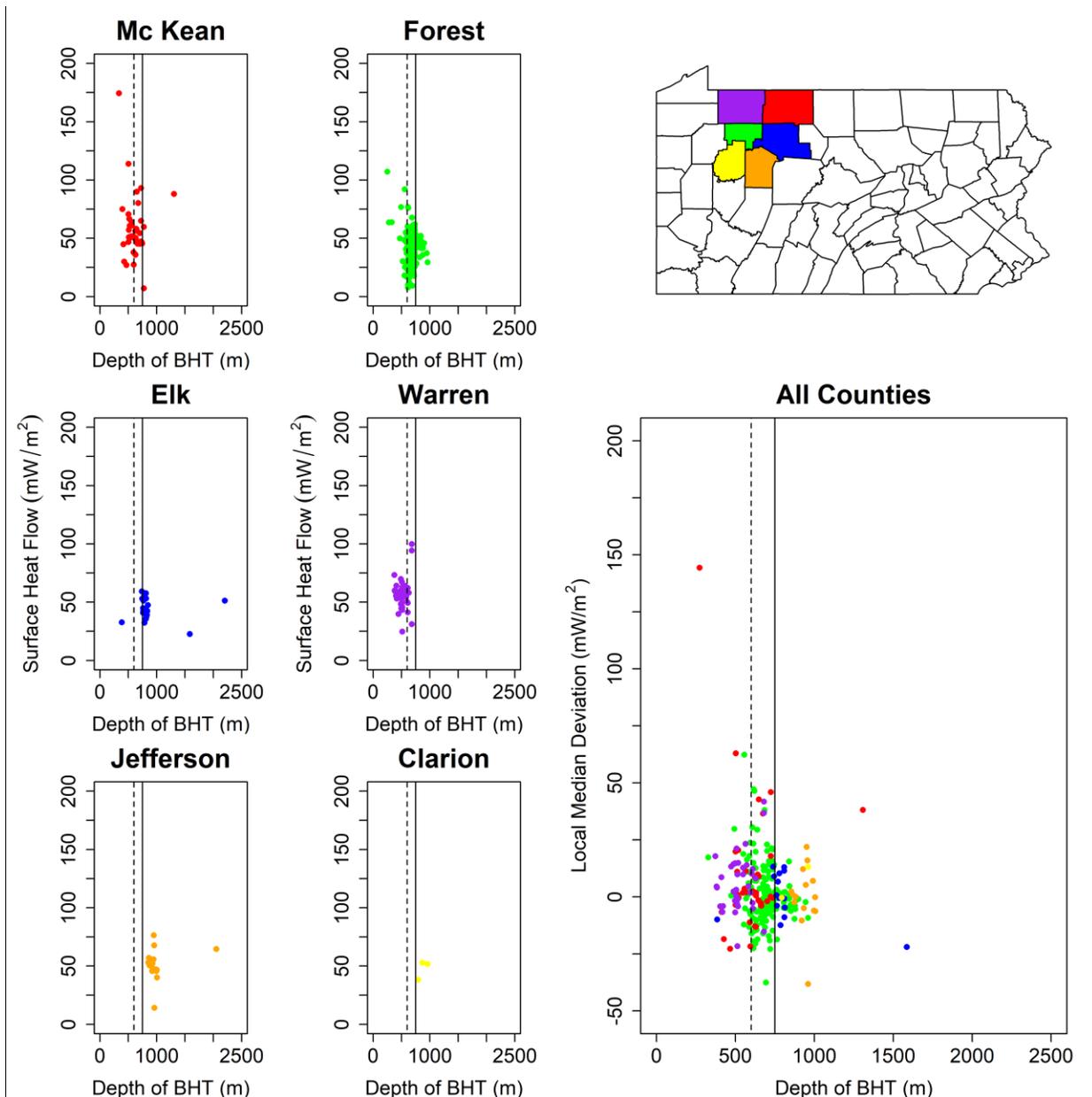
where  $s_n^2$  is the sample nugget semi-variance,  $Q_s$  is the surface heat flow, and  $i$  and  $j$  are each a surface heat flow value for a single spatial location. Some locations had more than two BHT measurements, which resulted in  $(N \cdot [N-1])/2$  sample nugget semi-variances, where  $N$  is the number of multiple measurements at the spatial location. For these locations, the average of the sample nugget semi-variances that were calculated using Equation B.1 was reported for that location. The average sample nugget semi-variances are plotted as boxplots within their geologic regions in Figure B.4. For an estimator of the true nugget semi-variance in these geologic regions, the median value is recommended because this computation occurs before the correction and removal of potentially rogue observations. The smallest median nugget semi-variance is about  $3 \text{ (mW/m}^2\text{)}^2$ , which results in a surface heat flow difference of about  $2.5 \text{ mW/m}^2$  for values in the same spatial location.



**Figure B.4:** Boxplots of the average sample nugget semi-variance for records in the same spatial location within each geologic (interpolation) region. The width of the boxplot is proportional to the square root of the number of records used in the geologic region. Geologic region abbreviations: CT – Chautauqua, NY, CNY – central New York, CWV – central West Virginia, ENY – eastern New York, ENYPA – eastern New York and Pennsylvania, NWPANY, northwestern Pennsylvania and New York, SWPA – southwestern Pennsylvania, WPA – western Pennsylvania, and VR – Valley and Ridge.

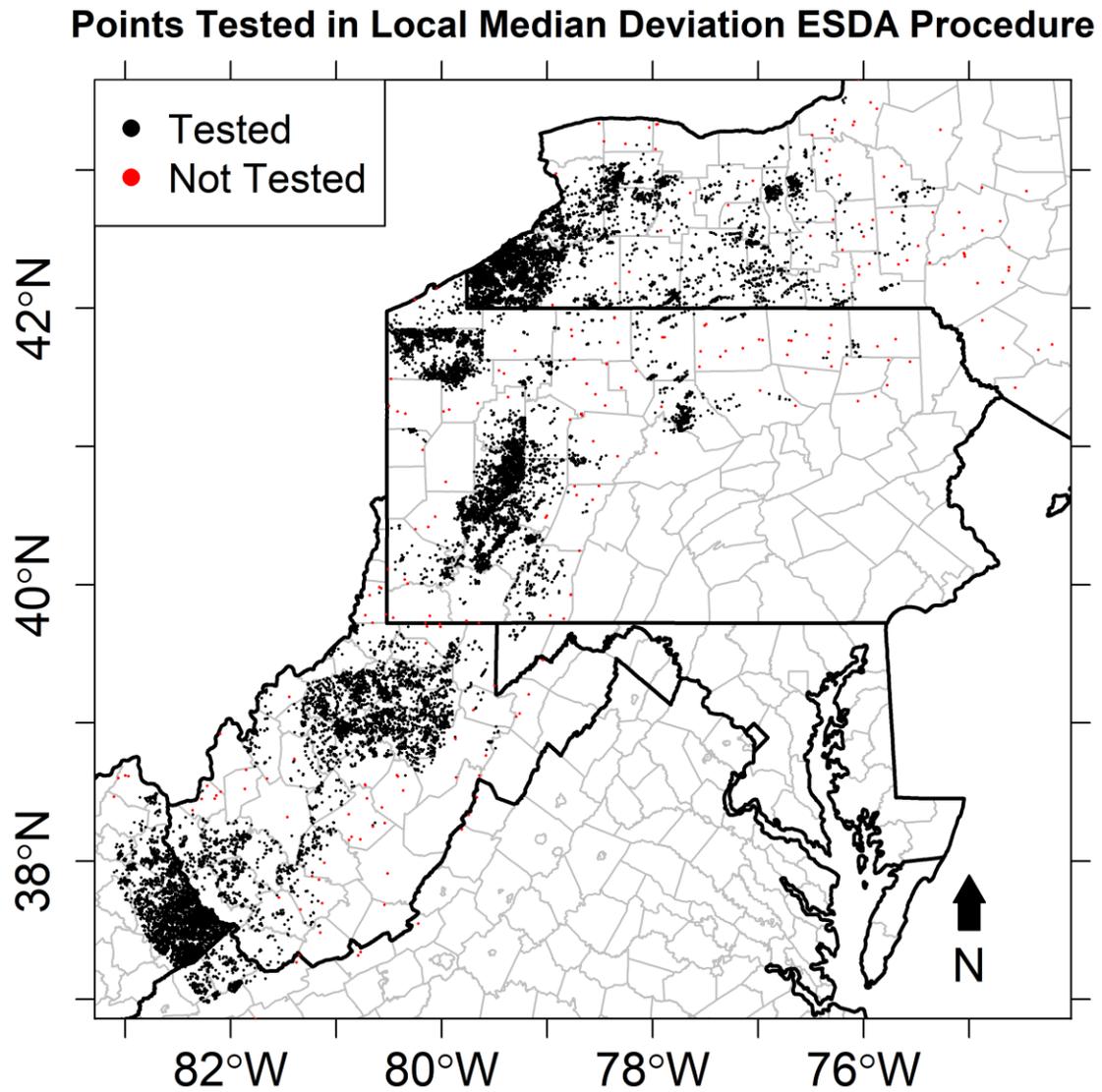
#### **4) Northwestern Pennsylvania minimum BHT depth analysis**

Data gaps caused by the 1000 m minimum depth threshold in northwestern Pennsylvania will create spatial gaps for geothermal resource assessments. The BHT measurements taken between 600 m and 1000 m in this area were examined for consistency in the surface heat flow value. A plot of the surface heat flow versus BHT depth for this area of Pennsylvania is provided in Figure B.5. Elk, Jefferson, and Clarion counties have surface heat flow values that are fairly consistent deeper than 600 m. Warren and McKean counties have some high surface heat flow values between 600 m and 750 m. Forest County has a relatively large spread in the surface heat flow between 600 m and 750 m, but seems to be more consistent with neighboring Elk County deeper than 750 m. As a result, the minimum depth for BHTs in this area was set to 750 m. Some of the remaining high and low values may be screened in the spatial outlier analysis. Reducing the minimum depth in this area added 70 records back to the dataset.



**Figure B.5:** Local median deviation for the surface heat flow values for wells in Figure 2.4 in northwestern Pennsylvania. The colors of points match the counties on the inset map. The vertical dashed line is the 600 m minimum depth cutoff, and the solid line is the 750 m proposed minimum depth for this area.

5) Points tested in the local median deviation ESDA procedure



**Figure B.6:** Points that were not tested in the local median deviation analysis for the surface heat flow (red). All other points were tested.

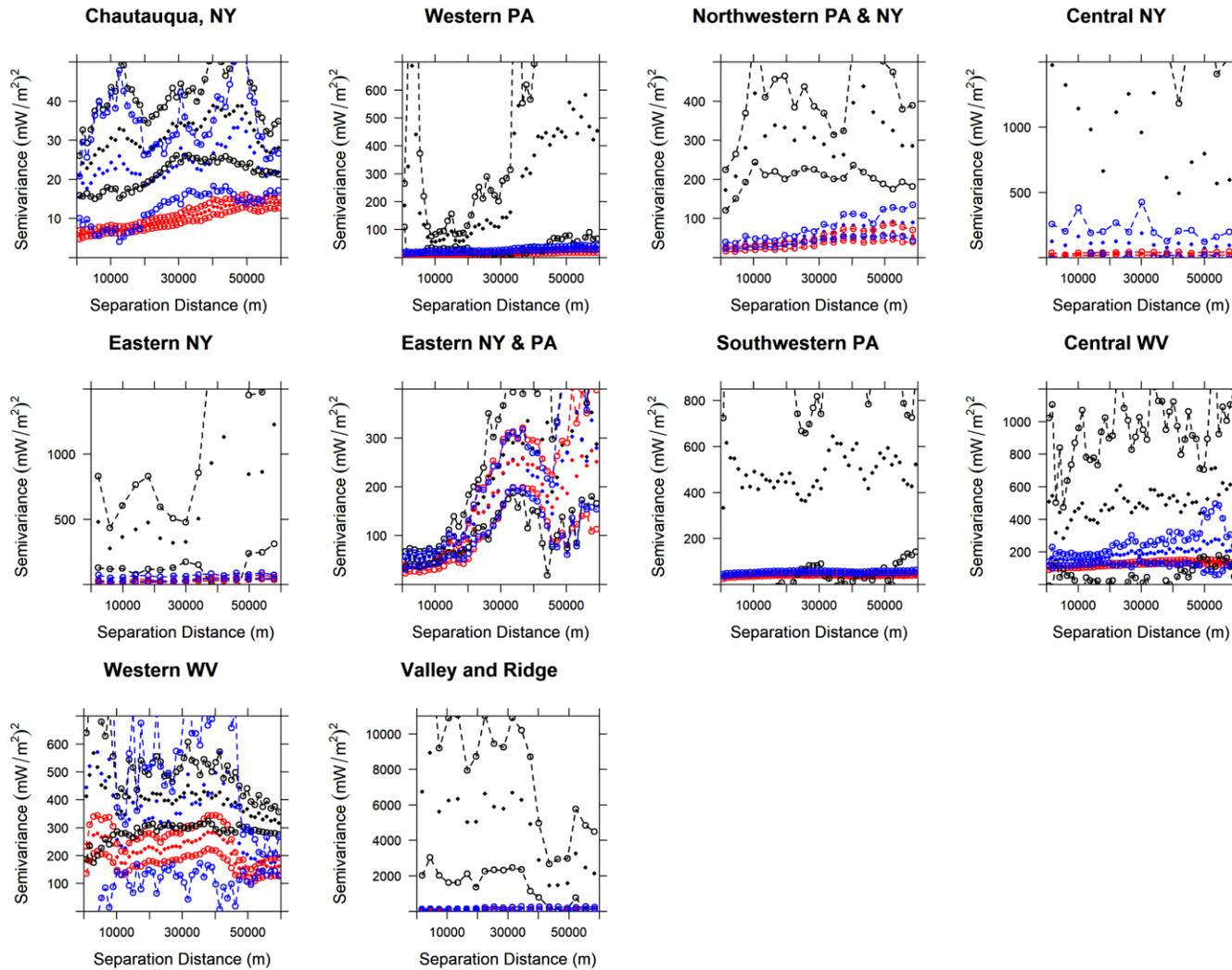
**6) Comparison of semi-variance estimates for ESDA procedures**

Table B.2 provides average semi-variance estimates for selected separation distance lags within each geologic region. Jackknife approximate 95% confidence intervals for all average semi-variance estimates are provided in Figure B.7 for each dataset provided in Table B.2, and Figure B.8 for Dataset 2 and Dataset 3 in Table B.2. Confidence intervals reveal that the uncertainty in the semi-variance estimates tends to decrease after each ESDA procedure is applied.

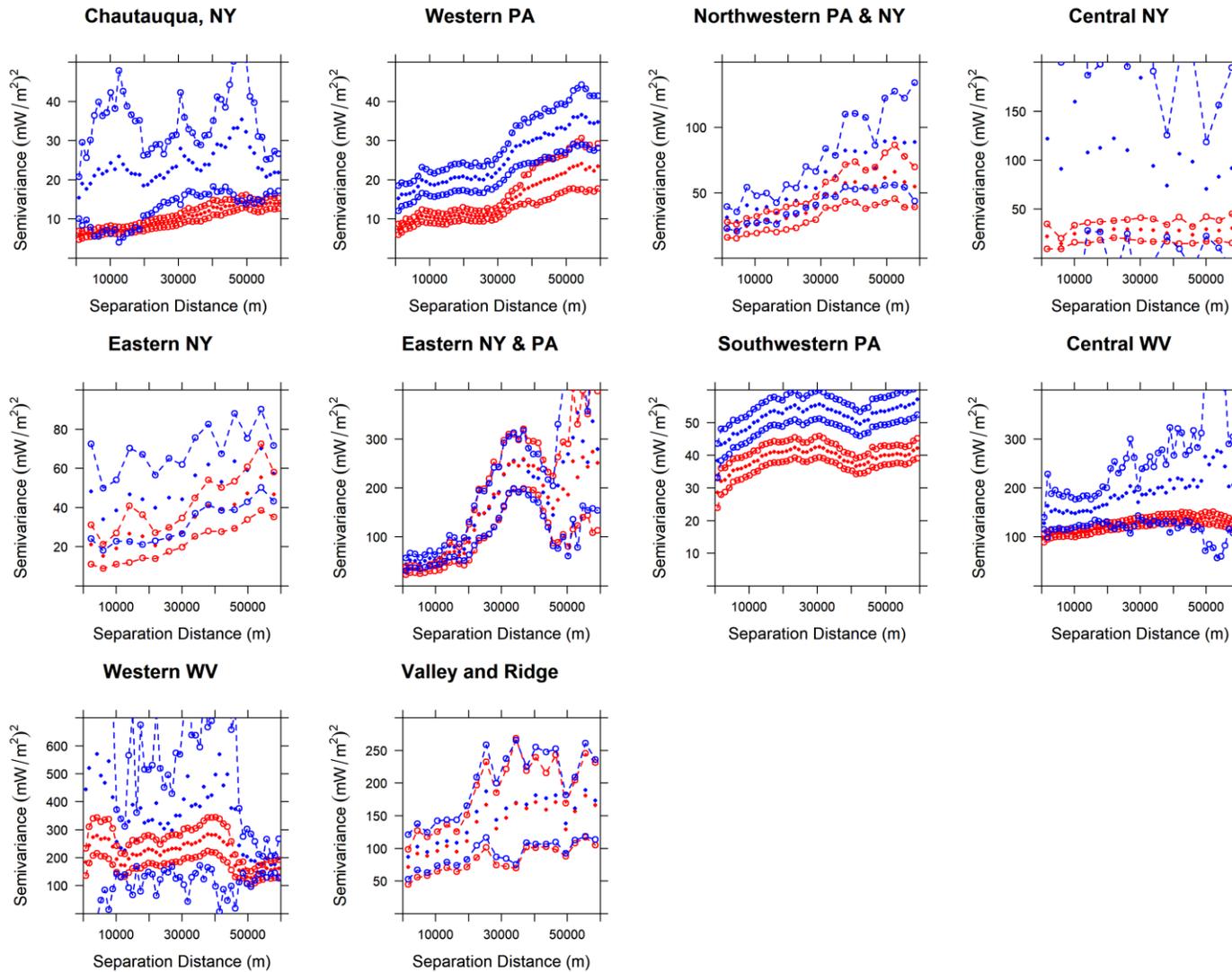
**Table B.2:** Estimates of mean semi-variance for selected separation distance lags after each major ESDA procedure: Dataset 1) pre-ESDA, Dataset 2) data shallower than 1 km removed, and Dataset 3) post-ESDA with outliers removed. Separation distance lags for each geologic region were selected based on the closest lags to 2.5 km, 5 km, 10 km, and 20 km.

<b>Geologic Region</b>	<b>Separation Distance (m)</b>	<b>Semi-variance for Dataset 1 (mW/m<sup>2</sup>)<sup>2</sup></b>	<b>Semi-variance for Dataset 2 (mW/m<sup>2</sup>)<sup>2</sup></b>	<b>Semi-variance for Dataset 3 (mW/m<sup>2</sup>)<sup>2</sup></b>
Chautauqua, NY	3030	22.8	17.6	6.3
	5410	26.2	21	6.4
	10204	30.5	24.2	7.3
	19797	26.9	18.5	7.9
Western PA	2987	687	16.2	8.7
	5415	158	17	9.2
	10202	56.1	18.8	10.4
	19801	73.2	20.8	11.4
Northwestern PA & NY	1467	172	31.1	21.8
	4604	207	28.1	20.6
	10465	420	37.2	25.4
	19494	332	44.9	30.4
Central NY	1849	1477	122	21.9
	6100	1323	90.9	14.7
	10120	1142	160	24.7
	17949	662	122	29.2
Eastern NY	2268	479	48.1	21.0
	5953	276	33.9	15.1
	9924	364	38.5	19.1
	21921	352	39.7	20.6

Eastern NY & PA	2289	57.6	51.7	34.2
	5259	52.3	46.0	30.5
	9750	56.2	52.7	37.9
	20224	129	97.0	88.5
Southwestern PA	3021	550	43.7	32.4
	5413	502	46.6	35.4
	10209	491	48.6	37.1
	19801	446	52.8	41
Central WV	3033	317	152	105
	5418	284	152	108
	10208	465	148	107
	19802	458	166	122
Western WV	3019	567	706	272
	5412	452	493	265
	10202	349	256	194
	19803	399	331	230
Valley and Ridge	1796	6741	87	71
	4431	8919	102	92
	10492	6238	108	96
	19556	5036	124	111



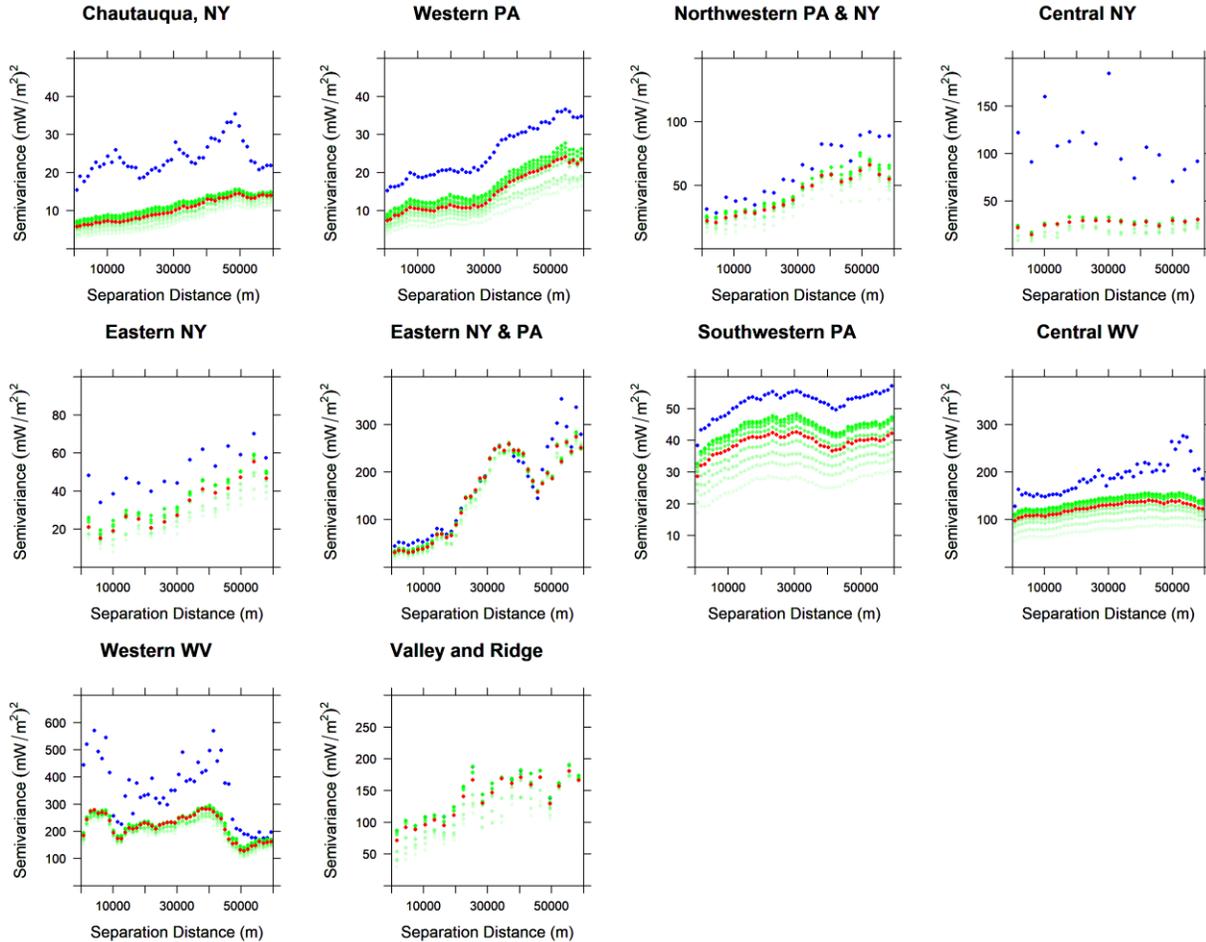
**Figure B.7:** Average semi-variance of the surface heat flow (solid points) with jackknife 95% confidence intervals (open points connected with dashed lines) within the geologic regions provided in Figure 2.8. Colors represent: 1) dataset pre-ESDA procedures (black), 2) dataset post-minimum depth threshold and pre-spatial outlier detection (blue), and 3) dataset post-ESDA with spatial outliers removed (red). Jackknife 95% confidence intervals are estimated following the procedure described in Shafer and Varljen (1990).



**Figure B.8:** Average semi-variance of the surface heat flow (solid points) with jackknife 95% confidence intervals (open points) within the geologic regions provided in Figure 2.8. Colors represent: 1) dataset post-minimum depth threshold and pre-spatial outlier detection (blue), and 2) dataset post-ESDA with spatial outliers removed (red). Jackknife 95% confidence intervals are estimated following the procedure described in Shafer and Varljen (1990).

## 7) Sensitivity of the Spatial Outlier Detection Results to the Choice of $k$ in Equation 2.1

Figure B.9 illustrates that our sample semi-variogram results are relatively insensitive to the choice of asymmetric boxplot outlier criterion,  $k$ , for values of  $k$  ranging from 1 to 5. For most regions, the values spanned by the semi-variograms for different values of  $k$  are similar to the confidence intervals for  $k = 3$  in Figure B.8.



**Figure B.9:** Average semi-variance of the surface heat flow within the geologic regions provided in Figure 2.8. This figure illustrates the sensitivity of the spatial outlier detection results to the value of the asymmetric boxplot outlier criterion parameter,  $k$ , in Equation 2.1. Colors represent: 1) dataset resulting from  $k = \text{infinity}$  (blue), 2) dataset resulting from  $k = 3$  (red), 3) dataset resulting from  $k = 1, 1.5, 2, \dots, 5$  (light green to dark green).

## APPENDIX C

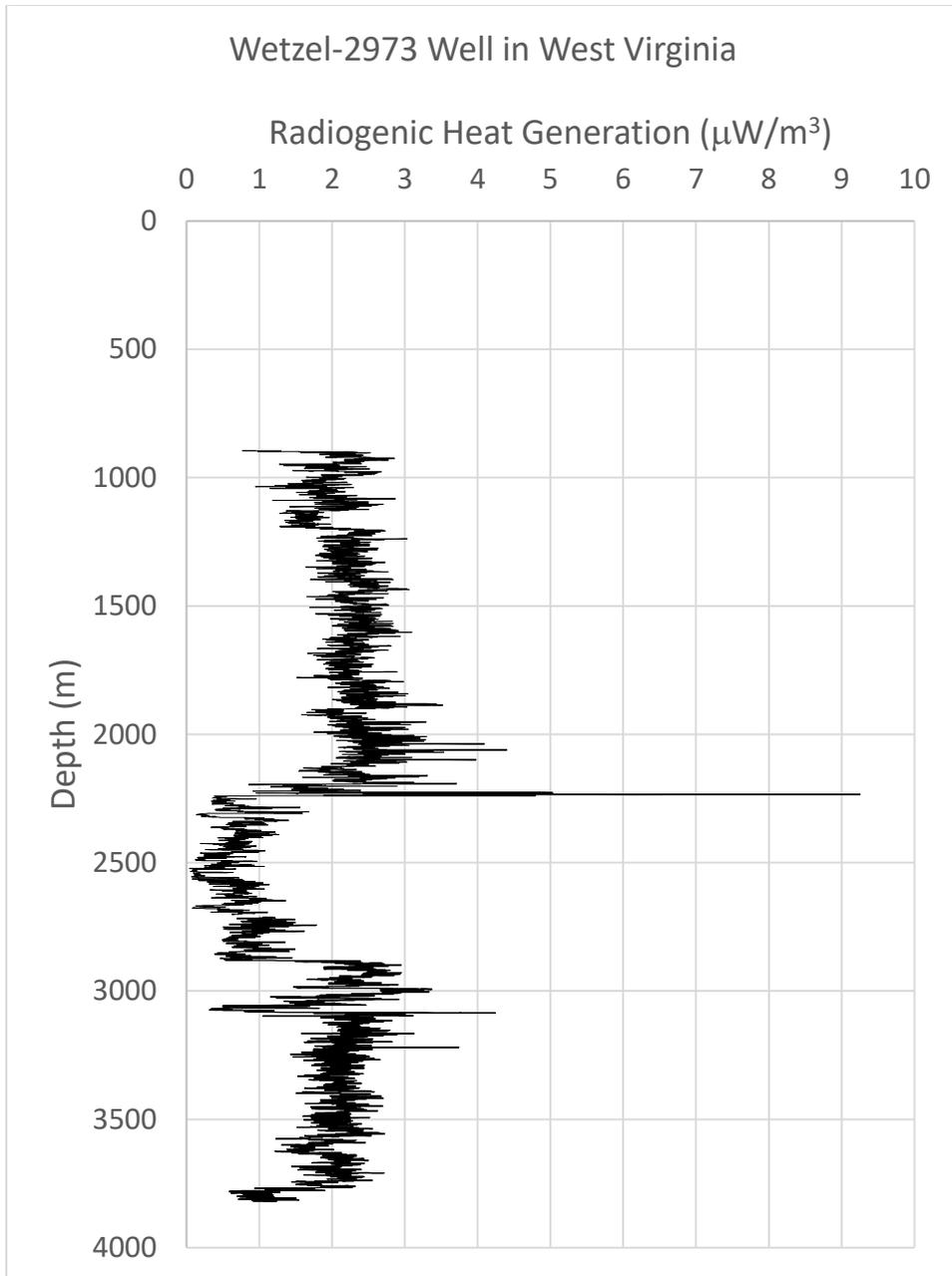
### SUPPLEMENTARY MATERIAL FOR CHAPTER 3

#### 1) Well Spectral Gamma Ray Logs

Radiogenic heat generation may be calculated from spectral gamma ray logs, which measure the concentration of uranium, thorium, and potassium in rocks. Bucker and Rybach (1996) present an equation to estimate the heat generation rate using these element concentrations:

$$A = 1E^{-5} * \rho_r * (9.52C_U + 2.56C_{Th} + 3.48C_K) * 1000 \left[ \frac{\mu W}{m^3} \right]$$

where  $A$  is the heat generation rate,  $\rho_r$  is the density of the rock ( $g/cm^3$ ),  $C_U$  is the uranium concentration (ppm),  $C_{Th}$  is the thorium concentration (ppm), and  $C_K$  is the potassium concentration (%). Using this equation, the radiogenic heat generation rate for sedimentary rocks for a well in West Virginia was calculated (Figure C.1). Siliciclastic rocks tend to average about  $2 \frac{\mu W}{m^3}$ , and carbonate rocks (e.g. Onondaga Limestone, Salina Group) tend to average about  $0.7 \frac{\mu W}{m^3}$ . This well log shows radiogenic heat generation rates greater than  $5 \frac{\mu W}{m^3}$  for the Marcellus Shale (7292 ft. [2222 m]). The range of named stratigraphic units is from the Big Injun Group (top: 2544 ft. [775 m]) to the Point Pleasant Formation (top: 12267 ft. [3739 m]).



**Figure C.1:** Radiogenic heat generation rate with depth estimated from a spectral gamma ray log for Wetzel-2973 well in West Virginia (API: 4710302973; WVGES, 2019).

## 2) Tuscarora Sandstone Potential Reservoir in Morgantown, WV

The Tuscarora Sandstone (e.g. Avary, 1996) is a target formation for use as a geothermal reservoir for the West Virginia University campus in Morgantown, WV. The geologic properties of the Tuscarora Sandstone near Morgantown are being evaluated as part of a U.S. Department of Energy-funded feasibility study for deep direct-use of geothermal energy. The spatial extent of the Tuscarora reservoir was estimated based on knowledge of local faults and structural features, and an assumed 5 km maximum pumping distance from the campus. An analysis of five wells within 30 km of Morgantown provided in McCleery et al. (2018) estimated the thickness of the Tuscarora near Morgantown as 122 m, with a vertical depth to the formation top of about 3058 m below ground surface. It is unclear from available data if this entire thickness would be productive as a geothermal reservoir for Morgantown. The uncertainty level assigned to the Tuscarora thickness in this paper is 1, which states that  $\pm 20\%$  of the mean thickness defines the lower and upper bounds of a symmetric triangular distribution. Two of three wells for which complete Tuscarora reservoir thickness was available near Morgantown have thicknesses within this interval, and the other thickness is 4 meters thinner.

McDowell (2018) collected visual porosity estimates for 29 thin sections spanning 19 m of the Tuscarora Sandstone in the Clay 513 well (API 4701500513) in Clay County, WV. A thickness of 19 m is a smaller scale than the thickness of the Tuscarora, and over this thickness there are a variety of features that induce localized porosity contrasts, including stylolites, burrows, and fractures. For this dataset, the average low porosity is about 1.3%, and the average high porosity is about 4.7% with an overall average of about 3%. We assigned an average porosity of 3% for the reservoir analysis in this study. Given the range in observed porosity for a single well, we assigned porosity an uncertainty level of 2 for this reservoir.

### 3) Spatial Data Processing Notes

In the Appalachian Basin reservoir dataset (available in Cornell University, 2016b), reservoir number 82 (RsvNum = 82) in New York was self-intersecting. There were several nodes of polygons located near the border of PA that overlapped, resulting in a multi-part feature with several distinct spatial areas corresponding to a single attribute in the database. Nodes were deleted to make a single part feature that was not self-intersecting.

The centers of each 1 km<sup>2</sup> grid cell in the surface heat flow raster files (available in Cornell University, 2016a) were computed and clipped to the reservoir polygon dataset. A “many to one” spatial join was used to join the information from potentially many reservoirs to a single grid cell point. To accomplish this task, a new record was created for each reservoir-grid cell pairing. A total of 16,708 pairs were created to compute stored thermal energy. These points are uniquely identified by the reservoir number and grid cell number pairing.

## APPENDIX D

### SUPPLEMENTARY MATERIAL FOR CHAPTER 4

#### 1) Justification for Selected Geologic Properties

This section presents justifications and details for the assumptions made to select geologic properties for geothermal reservoir simulations. The first subsection presents generic values for specific heat capacity, pore compressibility, pore expansivity, and tortuosity. The second subsection presents formation-specific values for heat capacity and other variables.

##### *Generic Values*

##### *Rock Specific Heat Capacity*

The source used for generic values of rock specific heat capacity is Robertson and Hemingway (1995). They present equations for the heat capacity of minerals and rocks as a function of temperature. We used their equations for limestone and dolomite to estimate heat capacity.

Limestone: We used the temperature (T) correlation for Salem Limestone in Robertson and Hemingway (1995). That equation for heat capacity ( $C_p$ ) is:

$$C_p = -2.78857 + 17.08 \cdot 10^{-4} T - 15.36498 \cdot 10^{-4} T^2 + 81.0669 \cdot T^{-0.5}$$

Example calculations used for our formations:

Trenton:  $T = 70 + 273.15$  K,  $C_p = 870$  J/kg-K

Black River: Assumed to be 10 J/kg-K more than the Trenton limestone because of the increase in temperature and possibility of more dolomitic limestone.

Beekmantown, Galway, and Theresa: 880 J/kg-K for  $T = 350$  K.

Dolomite: We used the temperature (T) correlation for Dolomite in Robertson and Hemingway (1995). That equation for heat capacity ( $C_p$ ) is:

$$C_p = (547.9 - 0.1676*T + 2.84E+06*T^{-2} - 6548*T^{-0.5} + 7.708E-05*T^2) / 184.41*1000$$

Example calculation used in our study for the Black River:

$$\text{Black River: } T = 75 + 273.15 \text{ K, } C_p = 930 \text{ J/kg-K}$$

### *Pore Compressibility, Pore Expansivity, and Tortuosity*

These variables are used for numerical modeling using TOUGH2. Pore compressibility affects how the pore volume changes as the pressure changes (e.g. as a result of pumping). The pore expansivity affects how pore volume changes as a function of temperature (e.g. as a result of cooling near a geothermal injection well). Tortuosity describes how a fluid flows through a rock matrix. Values for these variables are not directly available from datasets acquired for the formations deeper than the Utica Shale. We selected a value of 0 for these variables in our study, which is the default value in TOUGH2 (Pruess et al., 2012). These parameters are not often examined in published studies that include sensitivity analyses of numerical thermal-hydraulic geothermal reservoir models. Generally, these parameters affect the porosity of the rock, to which thermal energy production is often found to be relatively insensitive; however, hydraulic and thermal stresses could result in opening of pre-existing fractures or fracture generation, especially near the injection well (e.g. thermal reservoir stimulation in Grant et al., 2013). We do not consider such thermal-mechanical stresses in this study because site-specific data with which to estimate *in-situ* stresses at the depth of our formations are limited.

Setting the tortuosity to 0 in TOUGH2 results in using the Millington and Quirk (1961) relationship to compute tortuosity (Pruess et al., 2012). The Millington and Quirk (1961) relationship estimates tortuosity from rock porosity, which we estimate from well log data.

## ***Formation-Specific Values***

### *Lorraine / Utica Shale*

The depth to the top of the Lorraine/Utica shale was estimated based on the thicknesses of these formations in local well logs, and a formation top interpolation by J.A. Al Aswad (personal communication, 2018). The estimated depth below Cornell was 1860 m, with a thickness of 200 m.

Heat capacity of shale is assumed to follow the relationship provided in Waples and Waples (2004) for heat capacity,  $C_p$ , as a function of grain density,  $\rho$  :

$$C_p = 1.0263 * \exp(0.2697\rho) / \rho .$$

This equation provides the heat capacity for shale at 20 °C. Waples and Waples (2004) provide a plot of the effect of temperature on common minerals, which shows that the change in heat capacity from minerals at 20 °C to minerals at 70 °C is about 30 – 50 J/kg-K. Using the Utica Shale grain density measured in our selected well logs (Table D.4), we assume an average heat capacity of 830 J/kg-K.

### *Trenton Limestone*

The depth of the top of the Trenton formation was estimated from an interpolation of tops selected from local well logs (J.A. Al Aswad, personal communication, 2018).

### *Black River Dolomite*

The depth to the Trenton-Black River target reservoir was estimated by three independent methods (J.A. Al Aswad, personal communication, 2018; T.E. Jordan, personal communication, 2018) that interpolated tops from local wells (wells listed in Table D.1). The estimated depths

below Cornell were 2240 m and 2300 m using a subset of six wells near Cornell, and 2210 m with error of about 70 m using a larger subset of wells in the region. We assumed a depth of 2270 m below Cornell for this study, which is within the error of the shallowest estimation. Given the error in the estimation methods, a difference in  $\sim 1.5$  °C could result for the temperature at the top of this formation based on the estimated geothermal gradients.

For permeable Black River dolomite, we used 250 mD average permeability corresponding to 7% porosity (based on the Duddleston well log porosity [Table D.10] and the average value from local Black River reservoirs [Cornell University, 2016]). The permeability value is near the average predicted values of the regression estimates for maximum and 90-degree permeability presented in this study (Equation 4.2).

Permeability values observed in local Trenton-Black River oil and gas reservoirs ranged from 0.01 mD to an upper detection limit of 14,590 mD, and averaged 4,680 mD (Camp and Jordan, 2017). Vertical permeability was orders of magnitude less variable, ranging from essentially impermeable to 58.2 mD, and averaging at 2.6 mD (Camp and Jordan, 2017). We assume the average value of 2.6 mD for vertical permeability in this study.

Density values were selected based on the top 25<sup>th</sup> quantile values for Rehebein and Duddleston wells, and the upper tail of the Stevenson well (Table D.9).

Reservoir impedance was based on the estimated Reservoir Productivity Index (RPI) calculated for local Trenton-Black River reservoirs about 50 km southwest of Ithaca (Camp and Jordan, 2017; Cornell University, 2016b). The estimated impedance values (inverse of the RPI) range from 0.05 to 0.5 GPa-s/m<sup>3</sup>. The most likely value was about 0.15 GPa-s/m<sup>3</sup>.

### *Black River Limestone*

In this paper, the depth to the top of this formation below Cornell is constrained by the assumed 30 m thickness of permeable Black River dolomite.

Black River limestone is assigned porosity of 1% based on well logs (Table D.10). Using the censored regression relationships presented in this paper (Equation 4.2), we assigned horizontal permeability of 0.5 mD and vertical permeability of 0.0005 mD.

A rock density of 2700 kg/m<sup>3</sup> on average was assumed based on Trenton and Black River well logs (Table D.6, Table D.9). A greater value is used compared to Trenton limestone because the rock is more likely to be dolomitized.

### *Upper Beekmantown Group*

The depth to the top of this formation was estimated by an interpolation of formation tops in local wells by J.A. Al Aswad (personal communication, 2018).

### *Galway / Theresa*

The depth to the top of the Galway formation was estimated by J.A. Al Aswad (personal communication, 2018) as 2560 m below Cornell. The 220 m thickness for this formation group below Cornell was estimated from the Grund well, which is about 8 km from Cornell (Table D.1). The location of the Grund well and Cornell are estimated to be in an area of similar structural complexity within the region (Jordan et al., 2012; Cornell, 1959).

The Galway / Theresa formation group permeability ranges from 0.75 to 3.7 mD (Smith et al., 2005). Camp (2017) uses 2.6 mD on average for New York State, which we adopt in this study.

The porosity of these formations observed in local well logs is at most 4%, and commonly less than 1%. We assume an average of 1%.

### *Potsdam*

The depth of the top of the Potsdam formation below Cornell was estimated based on the expected thickness of the Potsdam formation. J. Al Aswad (2018) estimated a 20 m thickness of this formation below Cornell, which is similar to the thickness of the Potsdam formation in local wells that reached basement rocks.

The Potsdam sandstone is assigned the thermal conductivity value for sandstone from Carter et al. (1998). This value is similar to the value measured in the St. Lawrence lowlands (Rauch et al., 2018).

The Potsdam formation porosity is generally less than 5%, and permeability ranges from below detection limit ( $< 0.001$  mD) to 0.02 mD (Kolkas and Friedman, 2007). Based on the air permeability data reported in Kolkas and Friedman (2007), the average permeability is close to 0.002 mD. Based on Waller et al. (1978), the permeability in the vertical direction is likely about an order of magnitude smaller than the lateral direction.

A heat capacity value of 860 J/kg-K is assumed based on 80 °C sandstone in Abdulagatov et al. (2014).

## *Basement*

The depth to the top of basement rocks below Cornell was estimated from two independent analyses of deep well logs near the Cornell campus (T.E. Jordan, 2018; J.A. Al Aswad, 2018). Estimated depths ranged from 2760 m to 2865 m, with approximately 200 m of uncertainty in the estimations. We assume a basement depth of 2800 m below Cornell, which is close to the average of these predictions, and is within the uncertainty of each method.

The range of density values for Precambrian basement rocks were obtained from Simmons (1964, Table 1). Because we do not know the specific rock that underlies Cornell, and a variety of lithologies have been observed in New York State, we assume the range of densities observed by Simmons (1964) in the Adirondacks is representative of rocks that could underlie Cornell. Simmons (1964) state that the average density of Adirondack rocks is likely between 2720 – 2780 kg/m<sup>3</sup>, and values in Simmons (1964) Table 1 indicate that larger values of density are less likely than smaller values.

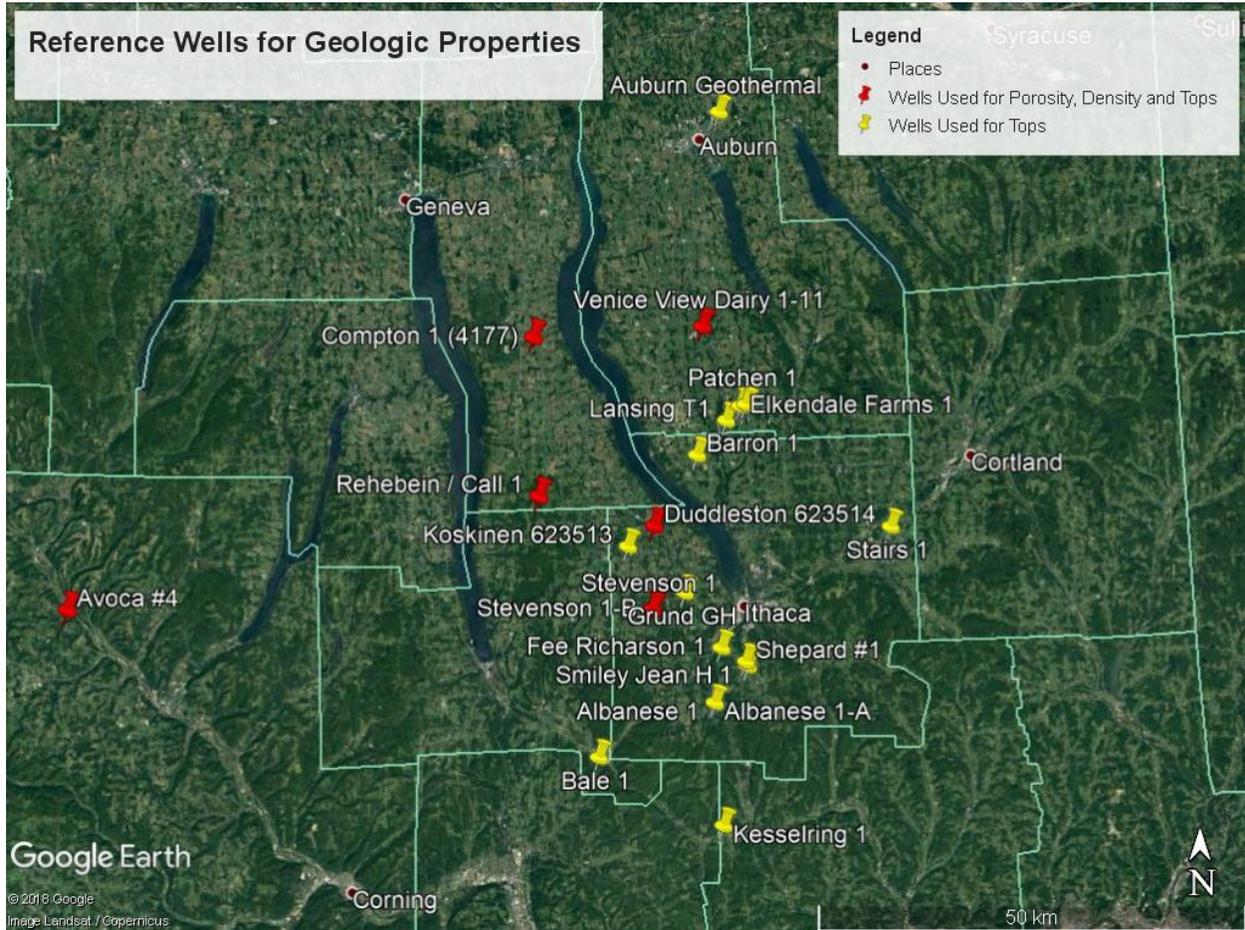
Porosity and permeability for Precambrian matrix rock was assumed to be similar to measurements taken on granite and gneiss. We assumed an average porosity of 1% and an average permeability of 1  $\mu$ D (Selvadurai et al., 2005; Manning and Ingebritsen, 1999).

The assumed thermal conductivity value of 2.83 W/m-K (Cornell University, 2016a) is similar to measurements taken by Southern Methodist University on Adirondack biotite granite gneiss, which ranged from 2.4 – 3.3 W/m-K.

The heat capacity value was estimated from Robertson and Hemingway (1995). At the temperatures expected for basement rock below Cornell, the heat capacity ranges from about 800 – 850 J/kg-K. We assume a value of 825 J/kg-K as a most likely value.

## 2) Reference Wells for Geologic Properties

The locations of reference wells used in this study to estimate geologic properties are provided in Figure D.1. Table D.1 provides the information needed to retrieve the well logs from the New York Empire State Organized Geologic Information System (ESOGIS) database.



**Figure D.1:** Locations of reference wells for lithologic properties and well logs near the Cornell project site. Wells with yellow pinpoint were used to inform formation tops below Cornell. Wells with red pinpoint were used for formation tops, density, and porosity information. Ithaca, the location of Cornell, is shown on the map.

**Table D.1:** Deep wells with density (+) and porosity (\*) logs near the Cornell project site. Some of these wells were also used to estimate formation tops at the Cornell project site (<sup>T</sup>). Wells with density and porosity logs that were processed and used in this study have the Well Name in bold and italics. The symbols indicating data types provided by wells are printed next to well names.

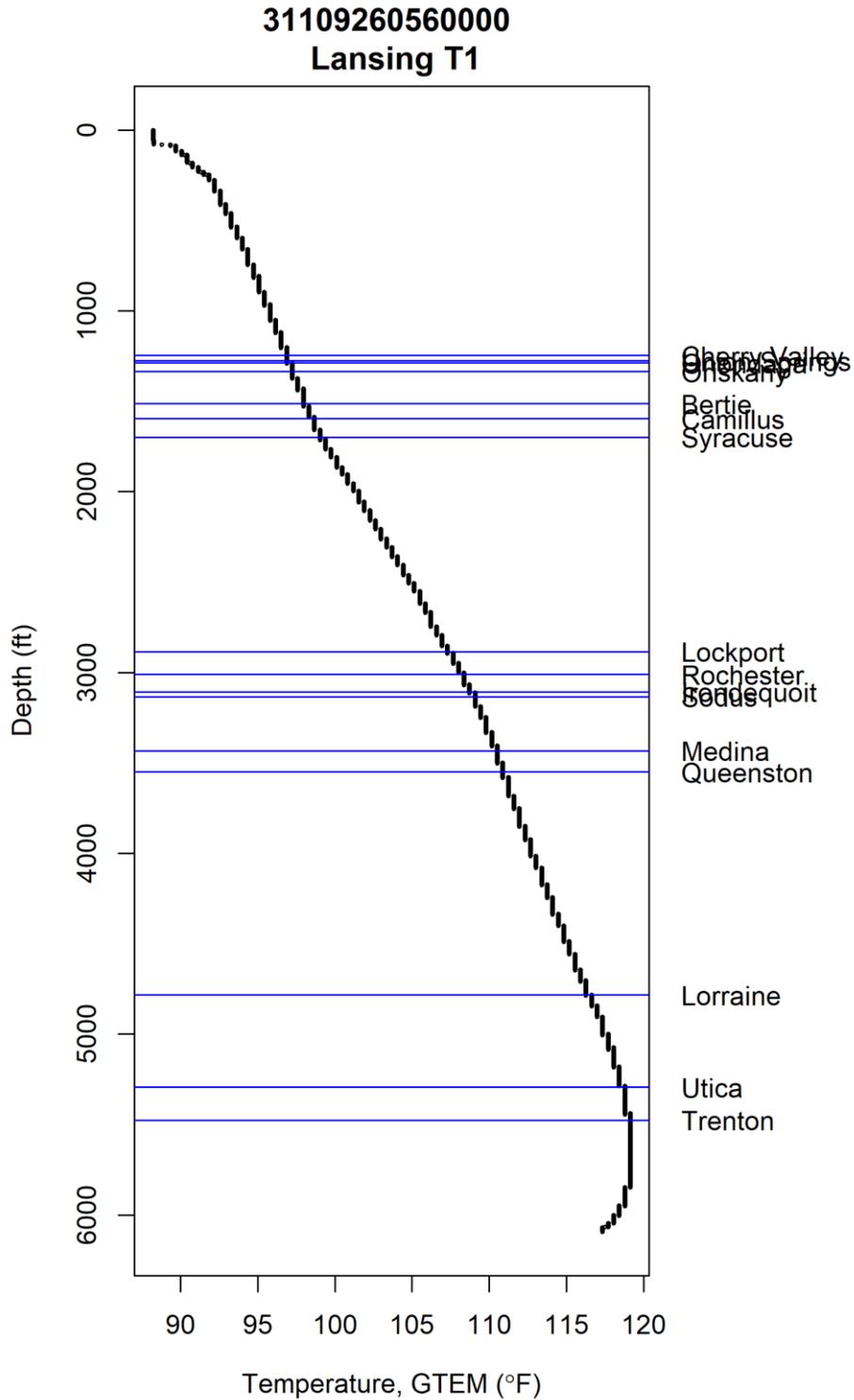
API Number	Well Name	County	Distance from Ithaca (km)	Deepest Formation
31109044670000	Fee-Richarson 1 + <sup>T</sup>	Tompkins	7	Rose Run
31109039730000	Shepard 1 * <sup>T</sup>	Tompkins	8	Precambrian
31109041300000	Grund GH	Tompkins	8	Galway
31109040070000	Smiley Jean H 1 *	Tompkins	9	Little Falls
31109229980000	<b><i>Stevenson 1</i></b> *+ <sup>T</sup>	Tompkins	12	Tribes Hill
31109229980100	<b><i>Stevenson 1-A</i></b> *+ <sup>T</sup>	Tompkins	12	Tribes Hill
31109229980200	<b><i>Stevenson 1-B</i></b> *+ <sup>T</sup>	Tompkins	12	Black River
31109229980400	<b><i>Stevenson 1-D</i></b> *+ <sup>T</sup>	Tompkins	12	Black River
31109227670000	<b><i>Duddleston 623514</i></b> *+ <sup>T</sup>	Tompkins	14	Little Falls
31109229970000	Albanese 1 *+	Tompkins	14	Trenton
31109229970100	Albanese 1-A *+	Tompkins	14	Trenton
31109229970200	Albanese 1-B *+	Tompkins	14	Trenton
31109227530000	Koskinen 623513 *+ <sup>T</sup>	Tompkins	15	Black River
31109260390000	Barron 1 *+ <sup>T</sup>	Tompkins	17	Black River
31109217160000	Stairs 1 *+	Tompkins	18	Little Falls
31109260560000	Lansing T1 *+ <sup>T</sup>	Tompkins	20	Black River
31011238400000	Patchen 1 *+	Cayuga	22	Black River
31011239820000	Elkendale Farms 1 *+	Cayuga	22	Trenton
31097214950000	Bale 1 *+	Schuyler	26	Precambrian
31015004430000	Kesselring 1 <sup>T</sup>	Chemung	27	Galway
31109227890000	<b><i>Rehebein / Call 1</i></b> *+ <sup>T</sup>	Tompkins	27	Galway
31011161200000	<b><i>Venice View Dairy</i></b> *+ <sup>T</sup>	Cayuga	31	Precambrian
31109204460000	<b><i>Compton 1</i></b> *+	Seneca	39	Precambrian
31011214690000	Auburn Geothermal *+	Cayuga	56	Precambrian
31101216240000	<b><i>Avoca 4</i></b> *+	Steuben	80	Precambrian

### 3) Non-Equilibrium Temperature Logs from Wells Near Cornell

Several wells near Cornell have temperature logs. Most of these logs are likely not in thermal equilibrium, as their log dates suggest that they were taken shortly after drilling was completed. Nevertheless, the logs support a conduction-dominated geothermal setting below likely water table depths in this region, and the logs show changes in the geothermal gradient where thermal conductivity of geologic formations is likely to change. The well log data are available on the New York Empire State Organized Geologic Information System (ESOGIS). Wells logs may be accessed using the information provided in Table D.2. Some of these wells are likely deviated, so isothermal portions at the end of some logs may be a result of directional drilling. Several of the wells produced gas, which likely resulted in cooler temperatures measured at the deeper portion of those logs. The temperature log in Figure D.2 illustrates all of the aforementioned effects.

**Table D.2:** Wells with temperature logs near Cornell. The log name is as listed in the digital LAS file on the ESOGIS database, or the end of the raster file name on the ESOGIS database.

API Number	Well Name	Log Name
31011900010000	Auburn Geothermal	TEMP
31011238400000	Patchen 1	GTEM
31101216240000	Avoca 4	TEMP
31109260390000	Barron 1	GTEM
31109260560000	Lansing T1	GTEM
31015004430000	Kesselring 1	Raster file TEM-01



**Figure D.2:** Temperature digital log from the Lansing T1 well located about 20 km north of Cornell. GTEM is the measured temperature. Blue lines indicate the locations of estimated formation tops, as listed in the ESOGIS database. Names of the formations are listed to the right of the blue lines.

#### 4) Processing of Well Log Porosity and Density Data

For the formation groups listed in Table 4.2, formation tops were selected for six wells with digital logs that allowed for corrections to be completed for gas and shale. Summary statistics for density and the computed effective porosity values are provided in the following tables. The well logs used and the log names are provided for each formation. Sufficient information with which to characterize the observed spatial heterogeneities in formation properties were not readily available with the datasets collected in this study.

The equations used to compute effective porosity, which adjusts neutron porosity log values for the effects of shale and gas, are provided in Bassiouni (1994). These equations included: the shale index (Eq. 15.2), shale volume with Larionov correction for older rocks (Eq. 15.6), density porosity (Eq. 15.15), gas presence (Eq. 16.2), and correction for gas presence (Eq. 16.24). Using these corrections, the effective porosity was calculated using Eq. 15.20b. These equations are reproduced here for reference:

##### *Shale Index*

$$I_{shale} = \frac{\gamma_{log} - \gamma_{clean}}{\gamma_{shale} - \gamma_{clean}}$$

Where  $I_{shale}$  is the shale index,  $\gamma_{log}$  is the gamma ray log measurement,  $\gamma_{clean}$  is the value the gamma ray log would measure in the cleanest region of a particular formation (without shale contamination), and  $\gamma_{shale}$  is the value the gamma ray log would measure in shale. Because we did not know what the gamma ray values for the cleanest values would be, the minimum gamma ray value measured for a formation was used instead. Similarly, the maximum gamma ray value was used as a proxy for the response for shale.

### ***Shale Volume with Larinov Correction***

$$V_{shale} = 0.33(2^{2I_{shale}} - 1)$$

Where  $V_{shale}$  is the shale volume and  $I_{shale}$  is the shale index.

### ***Density Porosity***

$$\phi_D = \frac{\rho_{matrix} - \rho_{bulk}}{\rho_{matrix} - \rho_{fluid}}$$

Where  $\phi_D$  is the density porosity,  $\rho_{matrix}$  is the assumed matrix density of the formation without any contamination from shale,  $\rho_{bulk}$  is the log-measured density, and  $\rho_{fluid}$  is the assumed pore fluid density. For this analysis, we assumed standard (Bassiouni, 1994) matrix densities for sandstone (2.65 g/cm<sup>3</sup>) limestone (2.71 g/cm<sup>3</sup>) and dolomite (2.88 g/cm<sup>3</sup>). We assumed brine with a density of 1.19 g/cm<sup>3</sup> for the pore fluid.

### ***Gas Presence***

$$\phi_D > \phi_{log}$$

Where  $\phi_D$  is the density porosity measured in a gas-bearing formation without any contamination from shale, and  $\phi_{log}$  is the log-measured neutron porosity.

### ***Correction for Gas Presence***

$$\phi_{corr} = \left( \frac{\phi_{log}^2 + \phi_D^2}{2} \right)^2$$

Where  $\phi_{corr}$  is the porosity corrected for gas presence,  $\phi_D$  is the density porosity measured in a gas-bearing formation without any contamination from shale, and  $\phi_{log}$  is the log-measured

neutron porosity. If gas was not assumed based on the gas presence criterion above, the simple average of  $\phi_D$  and  $\phi_{log}$  was used as the corrected porosity.

### ***Effective Porosity***

$$\phi_{eff} = \phi_{corr} + V_{shale} * \phi_{shale}$$

Where  $\phi_{eff}$  is the effective porosity,  $\phi_{log}$  is the log-measured neutron porosity,  $V_{shale}$  is the volume of shale, and  $\phi_{shale}$  is the neutron-measured porosity corresponding to shale. For these calculations, we assume that the neutron porosity of shale is 6%, as observed for the Utica shale in Wickstrom et al. (1992). Sensitivity of the calculations to the selection of shale neutron porosity was evaluated for all formations using 0% (no shale removal), 6% and 12%. The range of mean effective porosity from 0% to 12% shale censoring was generally less than 1%. The effect of the selected shale neutron porosity value on one-dimensional vertical spatial correlation of porosity (see Trenton and Black River variograms in Figure D.4 and Figure D.6) seemed important for some wells to filter out noise and arrive at a clearer spatial correlation signal. Spatial correlation was not considered in reservoir modeling in this study, but these results could be used to inform the selection of vertical density and porosity correlation in future studies.

## Utica

**Table D.3:** Wells used to estimate Utica density.

Well	Name	Longitude	Latitude	Log Names	Top Depth (ft)
3101116120	Venice View Dairy 1	-76.568	42.72023	RHOB	4700
3109721495	Bale 1	-76.7136	42.27001	RHOB	8140
3110121624	Avoca 4	-77.4672	42.42005	RHOB	6535.9
311092299801	Stevenson 1-A	-76.6384	42.42457	RHOZ	6452.67
3110926056	Lansing T1	-76.534	42.62213	RHO8	4800

Only density logs were used for the Utica shale because information needed to correct the porosity logs was unavailable for our datasets.

**Table D.4:** Quantile estimates of density ( $\text{g/cm}^3$ ) for the Utica formation from wells in Table D.3.

Quantile	3101116120	3109721495	3110121624	311092299801	3110926056
0.01	2.64	2.61	2.31	2.64	2.64
0.025	2.65	2.61	2.41	2.65	2.65
0.05	2.65	2.62	2.45	2.67	2.66
0.1	2.66	2.64	2.50	2.67	2.67
0.25	2.67	2.67	2.59	2.69	2.68
0.5	2.68	2.69	2.65	2.70	2.70
0.75	2.69	2.71	2.68	2.72	2.72
0.9	2.70	2.72	2.70	2.73	2.74
0.95	2.71	2.73	2.71	2.74	2.75
0.975	2.71	2.73	2.72	2.74	2.76
0.99	2.71	2.74	2.74	2.75	2.77

**Trenton**

**Table D.5:** Wells used to estimate Trenton density and porosity.

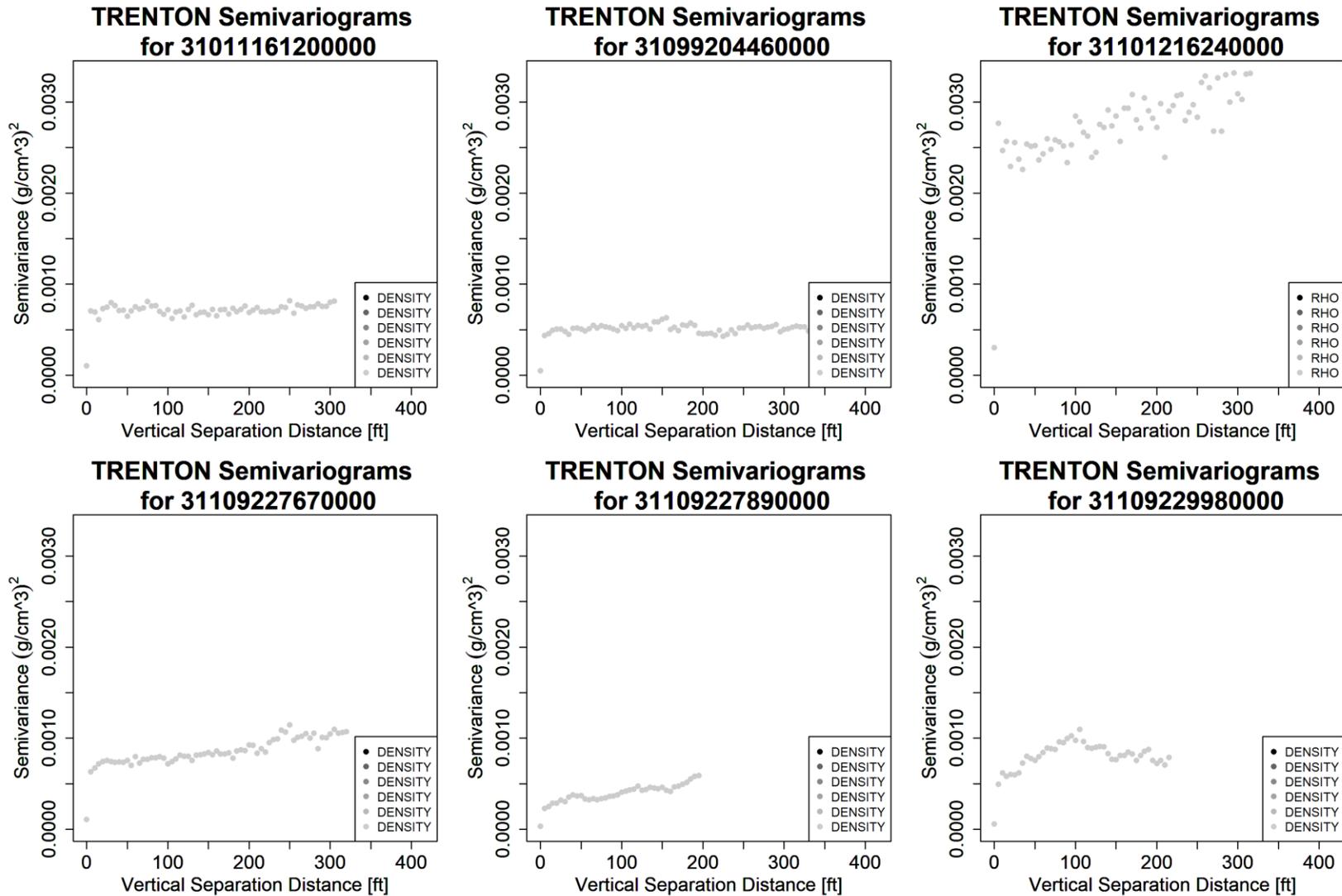
Well	Name	Long	Lat	Log Names	Top Depth (ft)
3101116120	Venice View Dairy 1	-76.56798	42.7202	GR, NPFI, RHO	5309
3109920446	Compton 1	-76.80819	42.7086	GR, NPFI, RHO	4534
3110121624	Avoca 4	-77.46718	42.4201	GR, NPFI, RHO	7171
3110922767	Duddleston 623514	-76.63581	42.512	GR, NPFI, RHO	6296
3110922789	Rehebein/Call 1 A	-76.27994	42.5429	GR, NPFI, RHO	6415.75
3110922998	Stevenson 1	-76.63839	42.4246	GR, NPFI, RHO	7216.17

**Table D.6:** Quantile estimates of density (g/cm<sup>3</sup>) for the Trenton formation from wells in **Table D.5**.

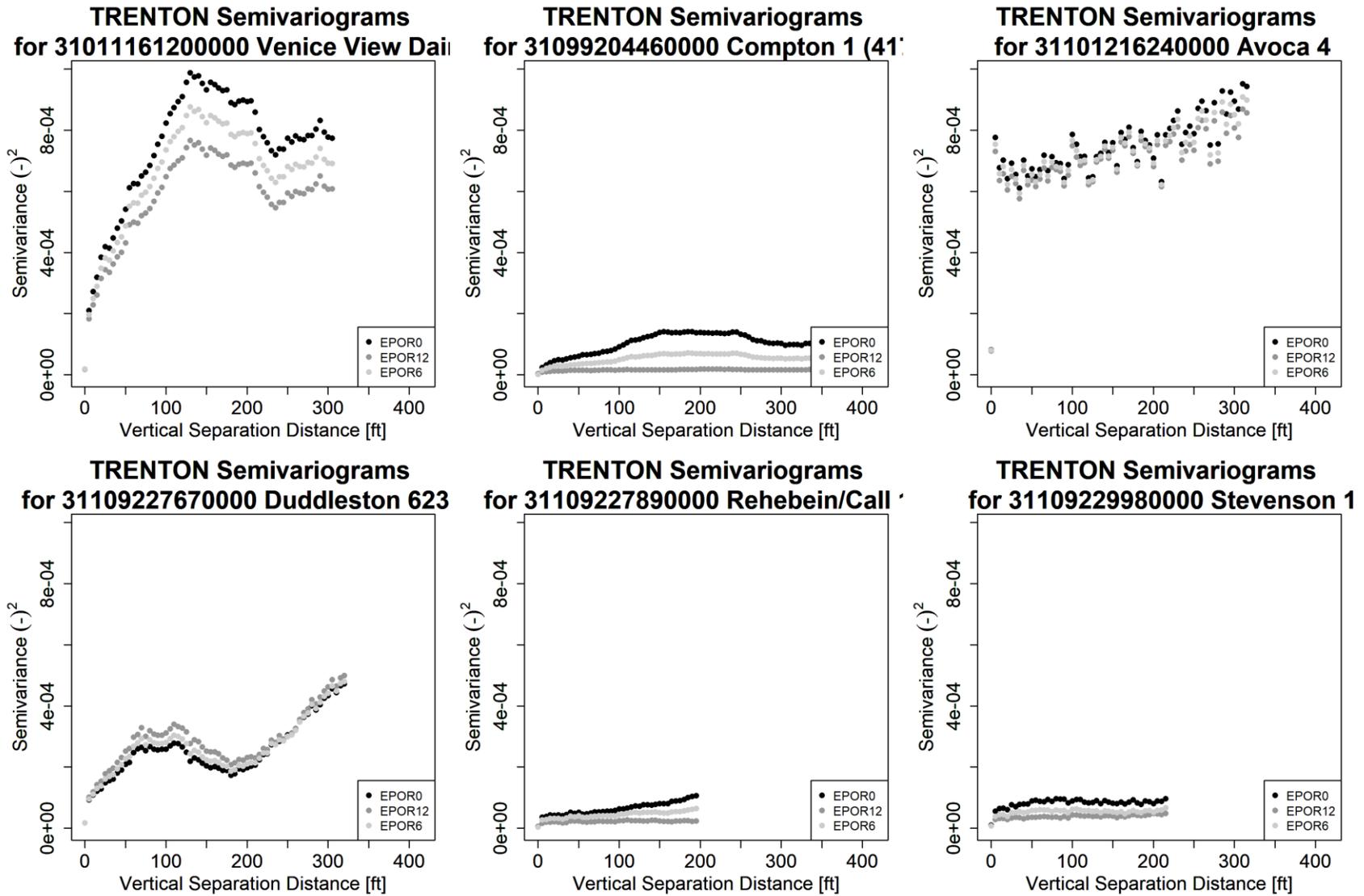
Quantile	3101116120	3109920446	3110121624	3110922767	3110922789	3110922998
0.01	2.56	2.61	2.40	2.58	2.65	2.62
0.025	2.61	2.66	2.50	2.60	2.65	2.63
0.05	2.63	2.67	2.54	2.62	2.66	2.65
0.1	2.65	2.69	2.56	2.64	2.66	2.66
0.25	2.67	2.70	2.60	2.66	2.68	2.68
0.5	2.68	2.71	2.64	2.67	2.70	2.70
0.75	2.70	2.71	2.66	2.69	2.71	2.72
0.9	2.71	2.72	2.68	2.70	2.72	2.73
0.95	2.71	2.73	2.69	2.71	2.73	2.74
0.975	2.72	2.73	2.70	2.72	2.74	2.75
0.99	2.73	2.74	2.71	2.75	2.75	2.77

**Table D.7:** Quantile estimates of (decimal) porosity for the Trenton formation from wells in **Table D.5**.

Quantile	3101116120	3109920446	3110121624	3110922767	3110922789	3110922998
0.01	0.0021	0.00005	0.0014	0.0082	0.0001	0.0001
0.025	0.0027	0.0001	0.0046	0.0101	0.0002	0.0002
0.05	0.0037	0.0003	0.0098	0.0126	0.0005	0.0004
0.1	0.0052	0.0006	0.0163	0.0148	0.0009	0.0008
0.25	0.0113	0.0015	0.0280	0.0203	0.0024	0.0020
0.5	0.0244	0.0030	0.0402	0.0305	0.0060	0.0046
0.75	0.0454	0.0080	0.0593	0.0437	0.0123	0.0086
0.9	0.066	0.017	0.075	0.063	0.019	0.016
0.95	0.082	0.022	0.090	0.069	0.023	0.021
0.975	0.095	0.027	0.110	0.078	0.025	0.028
0.99	0.109	0.032	0.160	0.088	0.034	0.039



**Figure D.3:** Vertical one-dimensional variograms for Trenton density for wells in Table D.5. The density values were nearly identical for the well logs used in each well, so they plot on top of each other.



**Figure D.4:** Vertical one-dimensional variograms for Trenton porosity for wells in Table D.5. Results for three neutron porosity (EPOR) shale censoring levels are provided: 0%, 6%, and 12%.

## Black River

**Table D.8:** Wells used to estimate Black River density and porosity.

Well	Name	Long	Lat	Log Names	Top Depth (ft)
3101116120	Venice View Dairy 1	-76.56798	42.7202	GR, NPFI, RHO	5920
3109920446	Compton 1	-76.80819	42.7086	GR, NPFI, RHO	5367
3110121624	Avoca 4	-77.46718	42.4201	GR, NPFI, RHO	7808
3110922767	Duddleston 623514	-76.63581	42.512	GR, NPFI, RHO	6941
3110922789	Rehebein/Call 1 A	-76.27994	42.5429	GR, NPFI, RHO	6808.91
3110922998	Stevenson 1	-76.63839	42.4246	GR, NPFI, RHO	7646.95

### *Dolomite*

For density, we used the 75<sup>th</sup> percentile estimates from Rehebein and Duddleston, and the upper tail of Stevenson. For porosity, only Duddleston was used to represent a productive reservoir.

**Table D.9:** Quantile estimates of density (g/cm<sup>3</sup>) for the Black River dolomite from wells in Table D.8.

Quantile	3110922767	3110922789	3110922998
0.01	2.31	2.67	2.73
0.025	2.40	2.68	2.74
0.05	2.48	2.68	2.76
0.1	2.54	2.70	2.79
0.25	2.65	2.72	2.83
0.5	2.72	2.74	2.86
0.75	2.78	2.81	2.90
0.9	2.82	2.85	2.94
0.95	2.84	2.87	2.96
0.975	2.86	2.87	2.97
0.99	2.87	2.88	2.99

**Table D.10:** Quantile estimates of (decimal) porosity for Black River dolomite from wells in Table D.8.

Quantile	3110922767	3110922789	3110922998
0.01	0.013	0.0002	0.0001
0.025	0.024	0.0008	0.0004
0.05	0.029	0.001	0.001
0.1	0.034	0.002	0.002
0.25	0.047	0.005	0.005
0.5	0.069	0.009	0.012
0.75	0.097	0.013	0.021
0.9	0.143	0.019	0.034
0.95	0.171	0.027	0.043
0.975	0.208	0.034	0.053
0.99	0.242	0.041	0.059

***Limestone***

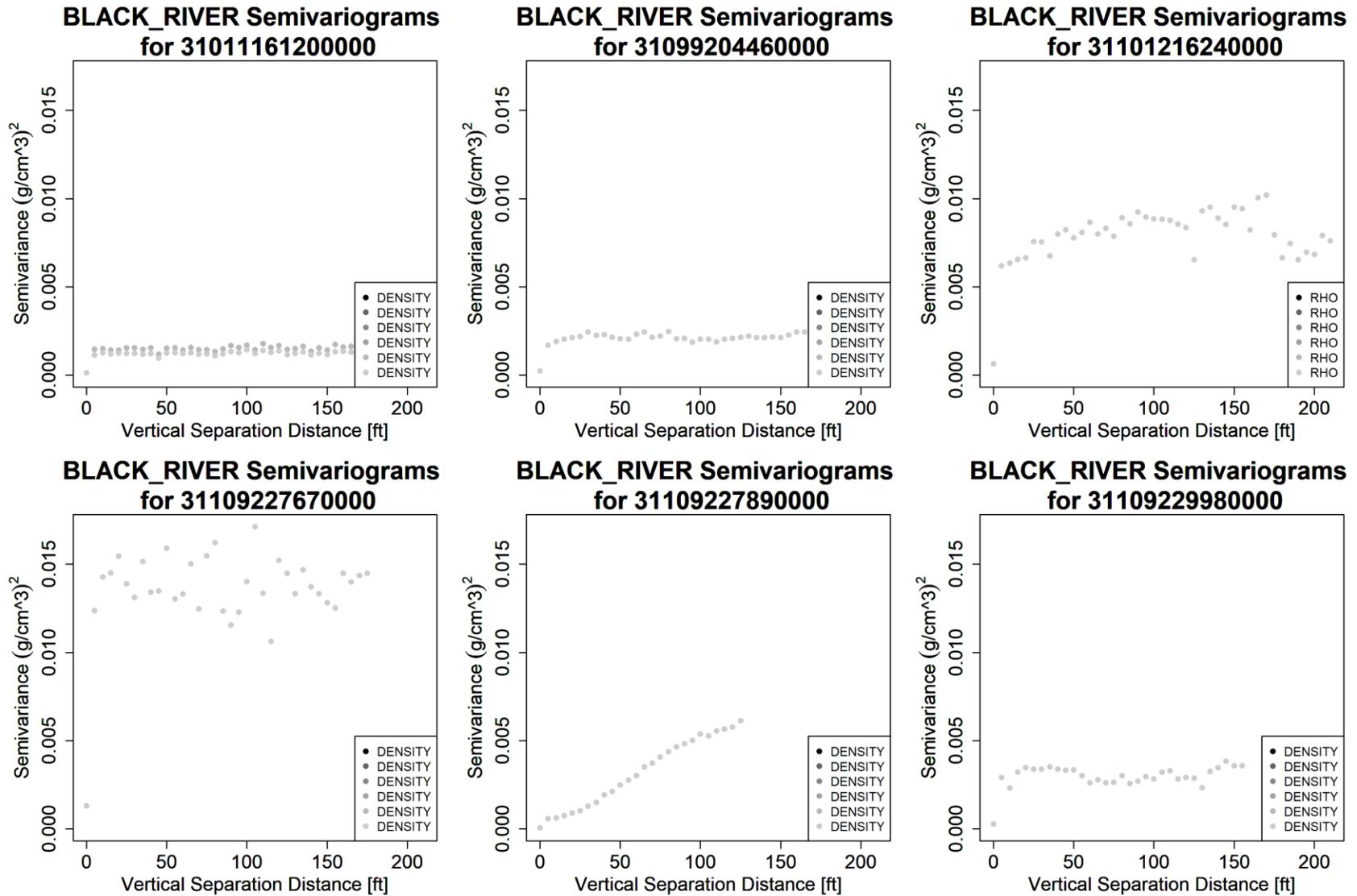
All wells except Duddleston were used to estimate Black River limestone.

**Table D.11:** Quantile estimates of density (g/cm<sup>3</sup>) for the Black River limestone from wells in Table D.8.

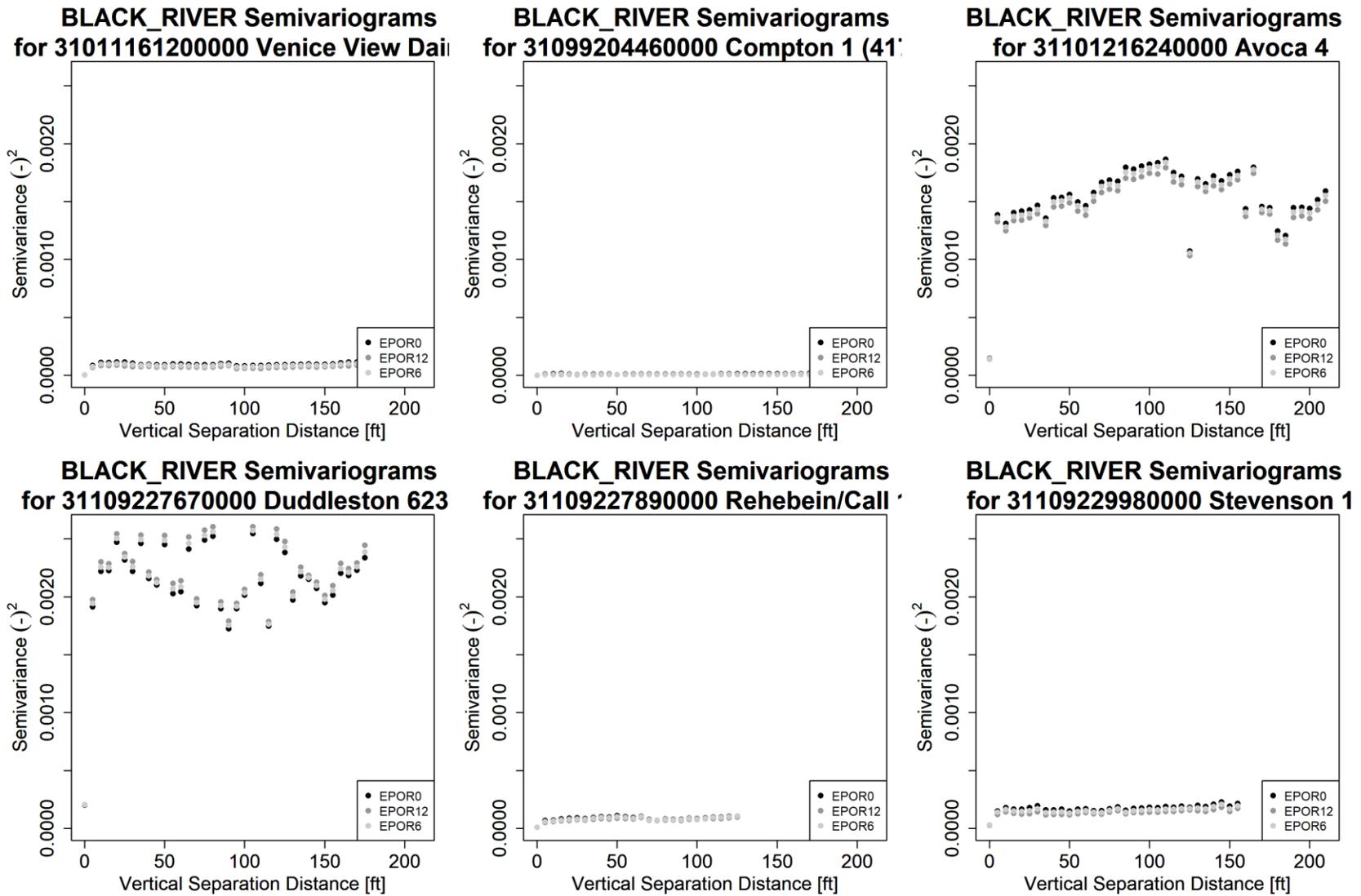
Quantile	3101116120	3109920446	3110121624	3110922789	3110922998
0.01	2.54	2.60	2.24	2.67	2.73
0.025	2.59	2.66	2.49	2.68	2.74
0.05	2.64	2.68	2.59	2.68	2.76
0.1	2.66	2.70	2.64	2.70	2.79
0.25	2.68	2.73	2.68	2.72	2.83
0.5	2.70	2.77	2.71	2.74	2.86
0.75	2.72	2.79	2.73	2.81	2.90
0.9	2.74	2.82	2.76	2.85	2.94
0.95	2.75	2.83	2.77	2.87	2.96
0.975	2.76	2.84	2.78	2.87	2.97
0.99	2.78	2.85	2.80	2.88	2.99

**Table D.12:** Quantile estimates of (decimal) porosity for Black River limestone from wells in Table D.8.

Quantile	3101116120	3109920446	3110121624	3110922789	3110922998
0.01	0.00004	0.00008	0.00019	0.00022	0.00014
0.025	0.00009	0.00012	0.00043	0.00078	0.00038
0.05	0.00015	0.00022	0.00071	0.00121	0.00076
0.1	0.00035	0.00040	0.00169	0.00189	0.00160
0.25	0.00095	0.00105	0.00499	0.00465	0.00501
0.5	0.00201	0.00188	0.01337	0.00892	0.01201
0.75	0.0039	0.0030	0.0276	0.0133	0.0209
0.9	0.0063	0.0042	0.0443	0.0187	0.0341
0.95	0.017	0.0050	0.061	0.027	0.043
0.975	0.034	0.0054	0.108	0.034	0.053
0.99	0.065	0.013	0.218	0.041	0.059



**Figure D.5:** Vertical one-dimensional variograms of Black River density well logs for wells in Table D.8. The density values were nearly identical for the well logs used in each well, so they plot on top of each other.



**Figure D.6:** Vertical one-dimensional variograms for Black River porosity for wells in Table D.8. Results for three neutron porosity (EPOR) shale censoring levels are provided: 0%, 6%, and 12%.

## Upper Beekmantown Group

**Table D.13:** Wells used to estimate Beekmantown Group density and porosity.

Well	Name	Long	Lat	Log Names	Top Depth (ft)
3101116120	Venice View Dairy 1	-76.56798	42.7202	GR, NPFI, RHO	6282
3109920446	Compton 1	-76.80819	42.7086	GR, NPFI, RHO	5762
3110121624	Avoca 4	-77.46718	42.4201	GR, NPFI, RHO	8231
3110922767	Duddleston 623514	-76.63581	42.512	GR, NPFI, RHO	7298
3110922789	Rehebein/Call 1 A	-76.27994	42.5429	GR, NPFI, RHO	7064.3
3110922998	Stevenson 1	-76.63839	42.4246	GR, NPFI, RHO	7958.78

**Table D.14:** Quantile estimates of density ( $\text{g/cm}^3$ ) for the Beekmantown Group from wells in Table D.13.

Quantile	3101116120	3109920446	3110121624	3110922767	3110922789	3110922998
0.01	2.62	2.10	2.66	2.27	2.72	2.50
0.025	2.64	2.64	2.70	2.35	2.73	2.51
0.05	2.66	2.70	2.72	2.45	2.73	2.72
0.1	2.70	2.73	2.74	2.54	2.73	2.75
0.25	2.72	2.76	2.77	2.65	2.75	2.77
0.5	2.75	2.79	2.79	2.74	2.76	2.82
0.75	2.77	2.81	2.81	2.78	2.80	2.87
0.9	2.79	2.85	2.83	2.80	2.82	2.92
0.95	2.80	2.89	2.84	2.81	2.82	2.94
0.975	2.81	2.91	2.85	2.84	2.84	2.95
0.99	2.82	2.93	2.85	2.86	2.84	2.96

**Table D.15:** Quantile estimates of (decimal) porosity for the Beekmantown Group from wells in Table D.13.

Quantile	3101116120	3109920446	3110121624	3110922767	3110922789	3110922998
0.01	0.0007	0.0001	0.0028	0.0040	0.0006	0.0007
0.025	0.0010	0.0002	0.0095	0.0175	0.0009	0.0013
0.05	0.0015	0.0003	0.0139	0.0262	0.0021	0.0018
0.1	0.0023	0.0006	0.0189	0.0285	0.0048	0.0026
0.25	0.0057	0.0019	0.0268	0.0379	0.0092	0.0069
0.5	0.0114	0.0045	0.0358	0.0619	0.0166	0.0171
0.75	0.018	0.009	0.049	0.105	0.027	0.032
0.9	0.037	0.017	0.062	0.157	0.049	0.048
0.95	0.060	0.023	0.073	0.183	0.053	0.077
0.975	0.069	0.053	0.083	0.218	0.054	0.145
0.99	0.083	0.091	0.094	0.255	0.054	0.147

**Galway/Theresa Carbonates and Rose Run Sandstone**

**Table D.16:** Wells used to estimate Galway/Theresa density and porosity.

Well	Name	Long	Lat	Log Names	Top Depth (ft)
3101116120	Venice View Dairy 1	-76.56798	42.7202	GR, NPHI, RHO	6605
3109920446	Compton 1	-76.80819	42.7086	GR, NPHI, RHO	6028
3110121624	Avoca 4	-77.46718	42.4201	GR, NPHI, RHO	8766.63

This group of formations consisted of carbonate and sandstone rocks. The Avoca and Venice View wells had relatively more Rose Run Sandstone than the Compton well, which resulted in higher values of density. For the purpose of geothermal reservoir modeling in Black River dolomite nearly 300 m shallower, heat production results are not likely to be sensitive to changes in density and porosity in this formation.

**Table D.17:** Quantile estimates of density (g/cm<sup>3</sup>) for the Galway/Theresa from wells in Table D.16.

Quantile	3101116120	3109920446	3110121624
0.01	2.55	2.53	2.53
0.025	2.57	2.55	2.55
0.05	2.58	2.56	2.56
0.1	2.59	2.57	2.57
0.25	2.61	2.60	2.62
0.5	2.63	2.62	2.75
0.75	2.72	2.63	2.82
0.9	2.77	2.65	2.84
0.95	2.79	2.66	2.85
0.975	2.80	2.69	2.86
0.99	2.81	2.74	2.87

**Table D.18:** Quantile estimates of (decimal) porosity for the Galway/Theresa from wells in Table D.16.

Quantile	3101116120	3109920446	3110121624
0.010	0.00002	0.00002	0.001
0.025	0.00005	0.00005	0.004
0.050	0.0001	0.0001	0.007
0.100	0.0002	0.0002	0.012
0.250	0.001	0.0005	0.021
0.500	0.002	0.001	0.034
0.750	0.006	0.003	0.051
0.900	0.014	0.006	0.076
0.950	0.020	0.015	0.091
0.975	0.028	0.021	0.098
0.990	0.036	0.028	0.109

## Potsdam

**Table D.19:** Wells used to estimate Potsdam density and porosity.

Well	Name	Long	Lat	Log Names	Top Depth (ft)
3101116120	Venice View Dairy 1	-76.56798	42.7202	GR, NPHI, RHO	7194
3110121624	Avoca 4	-77.46718	42.4201	GR, NPHI, RHO	9654

Few data points were available for the Venice View Dairy well after corrections were made to adjust for the effects of gas and shale. It is possible that apparent “shale” in the Potsdam formation is actually feldspar, as the correction method used could not distinguish between these rock types. As a result, we did not use the Venice View Dairy data to estimate porosity and density for the Potsdam formation. The density data for Venice View Dairy are provided for reference only.

**Table D.20:** Quantile estimates of density (g/cm<sup>3</sup>) for the Potsdam from wells in Table D.19.

Quantile	3101116120	3110121624
0.01	2.44	2.52
0.025	2.47	2.54
0.05	2.59	2.55
0.1	2.60	2.57
0.25	2.61	2.59
0.5	2.63	2.62
0.75	2.66	2.64
0.9	2.75	2.68
0.95	2.79	2.70
0.975	2.80	2.73
0.99	2.81	2.75

**Table D.21:** Quantile estimates of (decimal) porosity for the Potsdam from wells in Table D.19.

Quantile	3110121624
0.01	0.0001
0.025	0.0002
0.05	0.0006
0.1	0.002
0.25	0.005
0.5	0.015
0.75	0.027
0.9	0.040
0.95	0.051
0.975	0.056
0.99	0.070

## 5) Fracture Orientations in Basement Rocks of the Adirondack Mountains

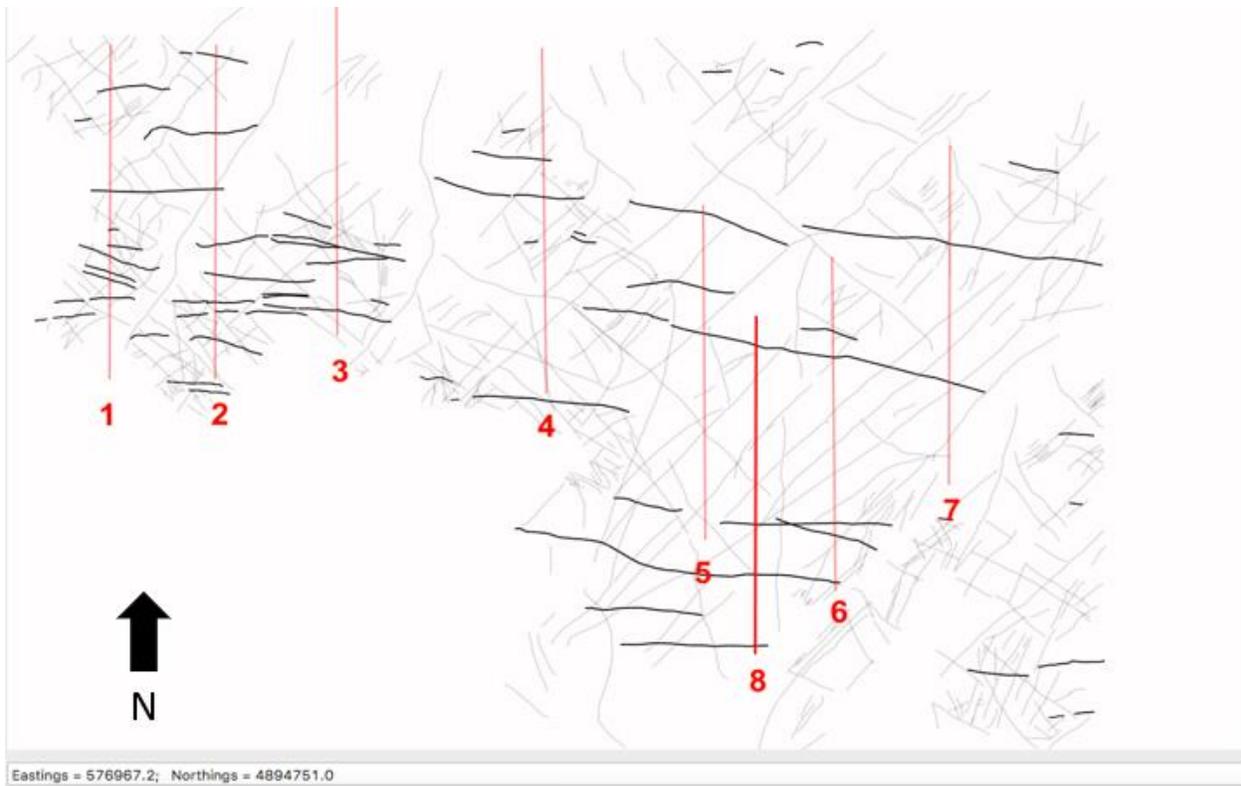
A LiDAR dataset covering a 15 km x 20 km area of the Adirondack Mountains near Mount Marcy was analyzed by Tasnuva Ming Khan (personal communication, 2018) for large-scale fracture spacings. Fracture spacings were estimated for three fracture orientations using linear scanlines that were approximately normal to the orientation of the fractures in plan view. Orientations considered were 1)  $45^\circ \pm 15^\circ$ , 2)  $95^\circ \pm 15^\circ$ , and 3)  $135^\circ \pm 15^\circ$  clockwise from North. Scanlines were sampled using the Scanline application developed by R. Allmendinger (2018, provided via personal communication). Summaries of the scanline data and their locations are presented in the following tables and figures. The length of scanlines was selected to be approximately 7 km long. In the sampled area, denser vegetation cover in the region of scanline #5 in Figure D.9 may introduce noise in detecting fractures.



**Figure D.7:** Scanlines (red) used to sample fractures with orientations of  $45^\circ \pm 15^\circ$  from North (black). Fractures with other orientations are shown in gray.

**Table D.22:** Scanline fracture spacings and fracture lengths estimated from the scanlines in Figure D.7.

Line #	Scanline Length (m)	Average Fracture Spacing (m)	Fracture Spacing Std. Deviation (m)	Number of Fracture Segments	Mean fracture length (m)
1	7106	1468	910	8	1783
2	7104	1060	438	10	4095
3	7103	677	304	18	3900
4	7105	815	319	12	4921
5	7106	1076	735	12	4561



**Figure D.8:** Scanlines (red) used to sample fractures with orientations of  $95^\circ \pm 15^\circ$  from North (black). Fractures with other orientations are shown in gray.

**Table D.23:** Scanline fracture spacings and fracture lengths estimated from the scanlines in Figure D.8.

Line #	Scanline Length (m)	Average Fracture Spacing (m)	Fracture Spacing Std. Deviation (m)	Number of Fracture Segments	Mean fracture length (m)
1	7020	625	475	14	1228
2	7020	849	331	14	1589
3	6910	448	321	6	1896
4	7261	782	NA	2	1496
5	7020	1269	260	4	3944
6	7020	1319	961	8	3934
7	7107	2901	NA	2	6185
8	7063	2101	916	6	4825



**Figure D.9:** Scanlines (red) used to sample fractures with orientations of  $135^\circ \pm 15^\circ$  from North (black). Fractures with other orientations are shown in gray.

**Table D.24:** Scanline fracture spacings and fracture lengths estimated from the scanlines in Figure D.9. Scanline 5 was sampled in a region with denser vegetation cover, which could affect the ability of LiDAR to show fracture details.

Line #	Scanline Length (m)	Average Fracture Spacing (m)	Fracture Spacing Std. Deviation (m)	Number of Fracture Segments	Mean fracture length (m)
1	7023	417	256	16	1160
2	7025	420	305	22	1497
3	7008	789	609	14	1673
4	7023	916	1012	14	4081
5	7041	2064	1305	6	4400
6	7179	664	375	10	1556

## 6) Censored Regression Parameter Significance Levels

**Table D.25:** Parameter estimates, standard errors, and significance levels for the censored regression on maximum permeability presented in Equation 4.2.

<b>Parameter</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
Intercept	-0.07415	1.2	NA
log(porosity)	3.441	0.704	1.03E-6
Log standard deviation	1.27	0.180	1.49E-12

**Table D.26:** Parameter estimates, standard errors, and significance levels for the censored regression on 90° permeability presented in Equation 4.2.

<b>Parameter</b>	<b>Estimate</b>	<b>Standard Error</b>	<b>p-value</b>
Intercept	-0.9357	1.017	NA
log(porosity)	2.651	0.564	2.63E-6
Log standard deviation	1.107	0.164	1.28E-11