PROCEEDINGS, 41st Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 22-24, 2016
SGP-TR-209

A Framework for Robust Analysis and Visualization of Geothermal Prospectivity

Dylan R. Harp[a], Youzuo Lin[b], William Glassley[c], David E. Dempsey[d], Satish Karra[e], Mark Person[f], and Richard Middleton[g]

[a,b,d,e,g] Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM, 87545

[c] Earth System Sciences, LLC, Santa Fe, NM, 87506

[d] Currently at University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

[f] New Mexico Institute of Mining and Technology801 Leroy Pl, Socorro, NM 87801

[a]dharp@lanl.gov, [b]ylin@lanl.gov, [c]bill.glassley@geo.au.dk, [d]d.dempsey@auckland.ac.nz, [e]satkarra@lanl.gov, [f]markaustinperson@gmail.com, [g]rsm@lanl.gov

Keywords: robustness, prospectivity, framework

ABSTRACT

We describe a framework to synthesize geothermal data streams by joint inversion. Geothermal production robustness, a non-probabilistic metric of geothermal prospectivity, is defined as the amount that uncertain model parameters can deviate from nominal, best-fit values and still produce simulations that meet geothermal production criteria. Larger parameter deviations indicate greater robustness in geothermal prospectivity. Results are automatically presented as 3D robustness maps using the open-source visualization software ParaView allowing interactive interpretation of the results.

1. INTRODUCTION

Determining the robustness of decisions to develop, explore, or enhance a geothermal resource from available data is a challenge. A variety of expertise is required to collect, process, and analyze hydrological, geological, geochemical, and geophysical indicators of geothermal prospectivity. The individuals tasked with making geothermal development decisions cannot be experts in all areas of geothermal sciences and technology. Therefore, geothermal energy development decisions require the synthesis and distillation of these disparate geothermal indicators into an accessible format. This requires additional forms of expertise involving data visualization, uncertainty quantification, model calibration, and decision analysis.

Geothermal data streams come from many sources with varying degrees of fidelity. For example, direct borehole temperatures are often considered high fidelity measurements and provide hard constraints on hydrothermal reservoir models. Geophysical surveys, such as seismic and MT, requiring a degree of expert interpretation, are considered to be of lower fidelity but often have higher spatial coverage providing less-precise large-scale constraints. Regional geological and gravity surveys often provide locations of faults and pyroclastic rock formations. Geochemical data such as dissolved silica and Na-Ca-K provide estimates of maximum reservoir temperatures. Robust decision analysis in geothermal exploration and development requires a synthesis of these data streams.

The development of approaches to jointly invert geothermal data streams is a current research challenge. The predominant approach has been to use some form of selective sampling scheme based on Bayesian inference to identify model realizations that are consistent with available data streams. For example, Jardani and Revil (2009) invert permeabilities using borehole temperature and self-potential data. Mellors et al. (2013) describe a framework for joint inversion of geothermal data streams using Bayesian inference through a Markov Chain Monte Carlo (MCMC) sampling scheme. McCalman et al. (2014) describe another framework that uses Bayesian inference to invert multiple geothermal data streams.

Similar to previously described approaches above, we utilize sampling schemes to collect parameter combinations. However, while the sampling scheme may use Bayesian inference to direct the sampling towards more “likely” parameter combinations, Bayesian inference is not a requirement in our framework. We have applied the framework simply using Latin Hypercube Sampling as well. Of course, sampling economy will increase with directed sampling schemes. Our approach uses concepts from information gap theory (Ben-Haim, 2001; Harp and Vesselinov, 2013) to post-process the sampling. The approach produces a non-probabilistic, set-based, spatially defined metric of robustness, or geothermal prospectivity. The concept behind our analysis approach is to quantify the robustness (i.e., confidence against failure) of alternative geothermal decisions without requiring probabilistic assumptions.

2. METHOD

Given the opacity of the Earth, the expense of exploratory drilling, and the limited ability to uniquely interpret geophysical surveys, geothermal data streams are often sparse, ambiguous, and/or low-resolution. Geothermal data streams are often highly uncertain and are better described as containing serious gaps in information. We therefore utilize a non-probabilistic approach to facilitate geothermal development decisions in cases involving gaps in information, which, in our estimation is true in many if not most geothermal investigations.

Probabilistic Bayesian inference, implemented through algorithms such as MCMC (Gilks, 2005), is an extremely powerful approach to selectively sample parameter combinations to a joint posterior distribution that produces simulations that are consistent with observations. Prior distributions of parameters containing beliefs or existing information are updated using the data to produce a joint posterior distribution. In cases where little or no information exist to define a prior, it is conventional to use a uniform distribution. In information theory (Reza, 1961), a uniform distribution has maximum entropy, indicating that the random variable has maximum uncertainty, and hence minimal information. This information-theoretic logic is often applied to justify the use of uniform prior distributions when there is a lack of information. In fact, a uniform prior is commonly referred to as an uninformative or non-informative prior. A problem arises when this information-theoretic logic is applied to cases where there are legitimate gaps in information. The uniform distribution does not represent a gap in information, but a lack of certainty. In fact, the probability for each possible value of the random variable is precisely defined as equal to other values in a uniform distribution; hence, incapable of representing gaps in information. Bayesian inference results are sensitive to prior distribution (Berger, 1990; Lavine, 1991; Vanpaemel, 2010; Müller, 2012). If they were not, their use as a means to incorporate existing beliefs and information would be flawed. Attempts to obtain robust Bayesian inferences have been developed that require evaluating the results of multiple priors (Berger et al., 1994; Insua and Ruggeri, 2012), a strategy that can lead to infeasible computational cost.

An alternative approach in cases with recognized information limitations is to use concepts from information gap theory. We have implemented such an approach into the framework and describe it below.

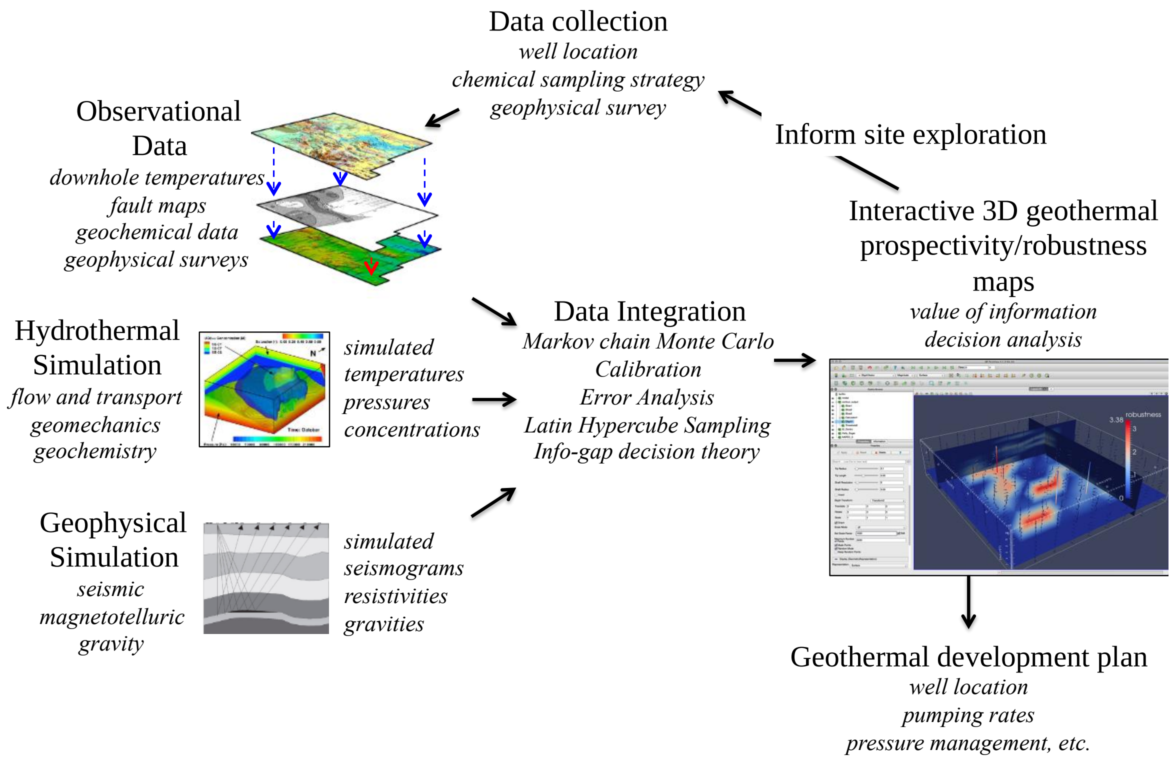

Figure 1: Framework flow diagram.

2.1 Framework

The workflow of the framework is illustrated in Figure 1. The workflow is similar to existing frameworks (Mellors et al., 2013; McCalman et al., 2014) in that simulations are compared with available data. A significant difference in our framework is the non-probabilistic analysis of these comparisons. After a sampling scheme has collected an ensemble of simulations that are consistent with observations, geothermal robustness (prospectivity) is calculated as described below. Results, including the most likely (i.e. most consistent with observations) and robustness metric are automatically opened in ParaView for 3D interactive exploration. If the results provide decision makers with enough confidence in the viability of the potential geothermal system, the results can be used to develop a geothermal development plan. Otherwise, the results may be used to determine where and what types of additional data should be acquired.

Hydrothermal and geophysical simulators are incorporated in the framework as Python functions. Execution of non-Python simulators (compiled executables or simulators written in other interpreted languages) is handled by system calls from the Python function. Creation of external model input files is facilitated in a similar fashion to other model analysis packages using template files (e.g., PEST (Doherty et al., 2006) and Dakota (Adams et al., 2009)).

The framework is comprised of a Python module written specifically for this framework, that leverages other freely available, open-source Python modules. The main leveraged Python module is the Model Analysis ToolKit (MATK; http://matk.lanl.gov). MATK facilitates model analysis setup (e.g. model function, parameter and observation definitions), external simulator input file generation, access to Python model analysis algorithms, and concurrent execution of simulations. Other leveraged Python modules, either directly or indirectly through MATK, include NumPy (Van Der Walt et al., 2011), SciPy (Jones et al., 2014), emcee (Foreman-Mackey et al., 2013), lmfit (Newville et al., 2014), and pyvtk (https://github.com/pearu/pyvtk). The framework outputs analysis results (e.g. geothermal robustness) in VTK format using pyvtk and facilitates the automatic execution, data loading and visualization formatting of the open-source visualization tool ParaView (Hendersen et al, 2004).

2.2 Models

The framework is simulator agnostic, simply expecting that all models are python functions. Automated input file generation is facilitated using MATK functionality. System calls can be handled using existing Python modules (e.g. the *system* method in the *os* module or the *call* or *popen* methods in the *subprocess* module). The model functions are expected to return a list or dictionary of simulation results. An example of this type of function can be viewed at http://matk.lanl.gov/example_fehm.html.

To date, two hydrothermal reservoir simulators have been used in the framework: the Finite Element Heat and Mass transfer code (FEHM; https://fehm.lanl.gov; Zyvoloski, 2007), via the PyFEHM Python module (http://pyfehm.lanl.gov), and PFLOTRAN code (http://www.pflotran.org; Lichtner et al., 2013), via the PyFLOTRAN Python module (https://bitbucket.org/satkarra/pyflotran-release-v1.0.0). PyFEHM and PyFLOTRAN facilitate setting up models in the form of Python function. FEHM is a finite-volume multi-phase, multi-component flow and transport simulator that has been developed at Los Alamos National Laboratory for over 30 years. FEHM was originally developed for the Hot Dry Rock test site at Fenton Hill, New Mexico (Kelkar et al, 1986). PFLOTRAN is a massively parallel subsurface and reactive transport simulator. PFLOTRAN is also a finite-volume multi-phase, multi-component simulator. Both FEHM and PFLOTRAN can be used in the framework to produce simulations that can be compared with observations of temperatures, pressure, and chemical concentrations.

Efforts have been undertaken to incorporate a seismic model into the framework. Geophysical methods can be used to measure the physical properties that are sensitive to temperature and fluid content of the geothermal reservoir. According to the relationship between the physical parameters and the geothermal reservoir, geophysical methods can be further differentiated into two categories: indirect methods (seismic methods, magnetic methods, etc.) and direct methods (geoelectrical methods etc.). Indirect methods are those methods inferring geophysical properties that reveal the characteristics of a geothermal reservoir, while the direct methods yield information on parameters that are influenced by the geothermal activity. The major advantage of indirect methods over the direct method is its accuracy and its capability to delineate the detailed image of the subsurface structure.

The seismic module can be incorporated with other modules within the framework. We can employ the seismic method to infer various geophysical properties within the subsurface, such as seismic-wave velocities, density, etc. As an example, fault zone location is an important piece of information in characterization of the geothermal reservoirs. It provides guidance for operators to locate geothermal resources and site wells. Incorporation of a seismic module allows obtaining inversion results of parameters defining fault zone location and extent using seismic measurements.

2.3 Prospectivity definition

Since the robustness metric is non-probabilistic, it does not quantify a probability of success. It is our belief that in many geothermal analyses, the quantity and quality of information will not justify such claims. Utilizing concepts from information-gap theory (Ben-Haim, 2001; http://info-gap.com), the approach is based on nested sets of uncertainty as opposed to distributions of uncertainty. The robustness of a spatial location within the model domain is calculated relative to the parameter combination that is most consistent with observations (e.g., smallest sum-of-squared residuals, most likely, etc.). These parameters are considered the nominal parameters, in other words, the parameters that a decision would be based upon. The robustness is calculated as the normalized amount that all parameters can deviate from nominal, and the simulated temperature and production at the location still exceed the desired quantities. Robustness therefore quantifies confidence against failure based on the extent that our nominal parameters can be incorrect and the decision is still a good one. There is an implicit requirement in the analysis that a decision is to be made, as the uncertainty is quantified relative to the nominal decision parameter values. While the distinction of information gap analysis versus optimal or Bayesian decision analysis may seem esoteric, it does represent a significant paradigm shift in the manner in which uncertainty involving gaps in information are handled in a decision analysis.

The following provides a formal definition of our geothermal robustness metric. If we consider $\widetilde{\boldsymbol{\theta}}$ as the vector of nominal parameters, $\boldsymbol{\sigma_\theta}$ the vector of parameter standard deviations, and α as the uncertainty parameter, often referred to as the horizon of uncertainty, we can define an expanding set of uncertain parameter combinations as

$$\boldsymbol{\theta} \in \widetilde{\boldsymbol{\theta}} \pm \alpha \boldsymbol{\sigma_\theta}, \alpha \geq 0$$

Based on this definition of nested expanding sets of uncertain parameter combinations, we can define associated sets of system responses as

$$S_T\left(\alpha, \widetilde{\boldsymbol{\theta}}, \boldsymbol{x}\right) = \left\{T(\boldsymbol{\theta}, \boldsymbol{x}) \colon \forall\, \boldsymbol{\theta} \in \widetilde{\boldsymbol{\theta}} \pm \alpha \boldsymbol{\sigma_\theta}\right\}, \alpha \geq 0$$

and

$$S_Q\left(\alpha, \widetilde{\boldsymbol{\theta}}, \boldsymbol{x}\right) = \left\{Q(\boldsymbol{\theta}, \boldsymbol{x}) \colon \forall\, \boldsymbol{\theta} \in \widetilde{\boldsymbol{\theta}} \pm \alpha \boldsymbol{\sigma_\theta}\right\}, \alpha \geq 0$$

where $T(\boldsymbol{\theta}, \boldsymbol{x})$ and $Q(\boldsymbol{\theta}, \boldsymbol{x})$ are the simulated temperature and production, respectively, at location $\boldsymbol{x}$ given parameter vector $\boldsymbol{\theta}$, and variable α defines the horizon of uncertainty.

Linguistically, the robustness $\hat{\alpha}$ can be defined as

$$\hat{\alpha}(\boldsymbol{x}) = \max\{\alpha\text{: minimum requirements are always satisfied}\}$$

Using the set definitions above, the temperature and production robustness can be defined as

$$\hat{\alpha}_T(\boldsymbol{x}) = \max\left\{\alpha\text{: } \left(\min S_T(\alpha, \widetilde{\boldsymbol{\theta}}, \boldsymbol{x})\right) \geq T_C\right\}$$

and

$$\hat{\alpha}_Q(\boldsymbol{x}) = \max\left\{\alpha\text{: } \left(\min S_Q(\alpha, \widetilde{\boldsymbol{\theta}}, \boldsymbol{x})\right) \geq Q_C\right\}$$

respectively, where T_C and Q_C are the specified minimum temperature and production required for a viable geothermal system. The total robustness can be calculated as the minimum of the temperature and production robustness's as

$$\hat{\alpha}(\boldsymbol{x}) = \min\left(\hat{\alpha}_T(\boldsymbol{x}), \hat{\alpha}_Q(\boldsymbol{x})\right)$$

2.4 Analysis description

The preliminary development of the framework used borehole temperature data from the Salton Sea Geothermal Field (Tiedeman et al., 2011). Other modeling conceptualizations are simple abstractions of a hypothetical geothermal resource solely developed to aid in the development of the framework. While the modeling scenario was integral in the development of the framework, the results are not to be taken as informative for the Salton Sea Geothermal Field. A more detailed analysis incorporating geologic information from the field can be found in Mellors et al. (2013).

As mentioned above, the analysis assumes that directly measured temperatures are available from three boreholes, represented as T_i, where $i = 1,\ldots 57$ is the measurement index including measurements from all three boreholes. The lithostratigraphy is assumed to be known and represented by zonation heterogeneity in the model with three zones. Permeabilities are assumed to be uniform within each zone but with uncertain values, represented as k_i, $i = 1,2,3$, where i indicates the zone. It is assumed that the basal heat flow is heterogeneous, but that its distribution is unknown. Basal heat flow at each node along the bottom of the model are parameterized as uncertain parameters represented by Q_i, $i = 1,\ldots,12$, where i indicates the bottom node index. Therefore, the parameter vector contains the uncertain zonal permeabilities and basal heat flows as $\boldsymbol{\theta} = \{\boldsymbol{k}, \boldsymbol{Q}\}$, where $\boldsymbol{k}$ and $\boldsymbol{Q}$ are vector representations of k_i and Q_i, respectively. We developed an FEHM hydrothermal model parameterized with $\boldsymbol{\theta}$.

We obtained an ensemble of parameter combinations biased towards those producing FEHM simulated temperatures ($\widehat{T}_i, i = 1, \ldots, 57$) that are more consistent with measured borehole temperatures using an MCMC algorithm (Foreman-Mackey et al., 2013). We then perform production simulations at each node of the model with each parameter combination of the ensemble. For computational efficiency, we use the Theis solution (Theis, 1935) as our production model. The materials surrounding each location determine Theis solution properties. A numerical simulator could be used in this step at significantly greater computational expense. Geothermal production performance metrics are defined as $Q(\Delta t_p, \boldsymbol{x}_p) \geq Q_C = 0.01$ kg/s and $T(t = 0, \boldsymbol{x}_p) \geq T_C = 100$℃ where Δt_p = 0 - 20 years and $\boldsymbol{x}_p$ is the production location. Therefore, it is required that the geothermal system can be produced at 0.01 kg/s or greater for 20 years and that the current temperature is at least 100℃.

The nominal parameter values, $\widetilde{\boldsymbol{\theta}}$, were determined as the parameter combination most consistent with measurements (i.e. lowest sum-of-squared errors, $\Sigma_i^N(\hat{T}_i - T_i)^2$). Parameter uncertainties must be standardized in the calculation of the robustness metric described above. We use the parameter standard deviations across the ensemble, $\boldsymbol{\sigma_\theta} = \{\sigma_{k_1}, \sigma_{k_2}, \sigma_{k_3}, \sigma_{q_1}, \ldots, \sigma_{q_{12}}\}$, to standardize the parameter uncertainties. The ensemble production simulations are then used to determine $S_T(\alpha, \widetilde{\boldsymbol{\theta}}, \boldsymbol{x})$ and $S_Q(\alpha, \widetilde{\boldsymbol{\theta}}, \boldsymbol{x})$, which allow the calculation of $\hat{\alpha}_T(\boldsymbol{x})$ and $\hat{\alpha}_Q(\boldsymbol{x})$, respectively, and ultimately $\hat{\alpha}(\boldsymbol{x})$.

3. RESULTS

The framework exports results to VTK files and automatically loads the results in ParaView as shown in Figure 1.

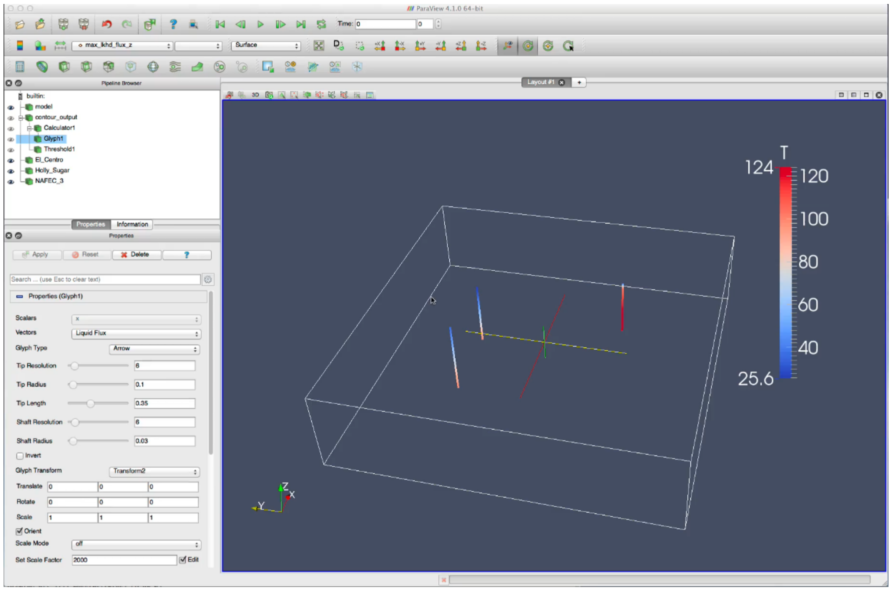

Figure 1: Screenshot of automatically loaded results in ParaView. Vertical lines represent measured borehole temperatures in Celsius.

Model results from the most likely scenario are also loaded automatically in ParaView. The flow vectors for the most likely scenario are shown in Figure 2 as arrows colored by the magnitude of their vertical component.

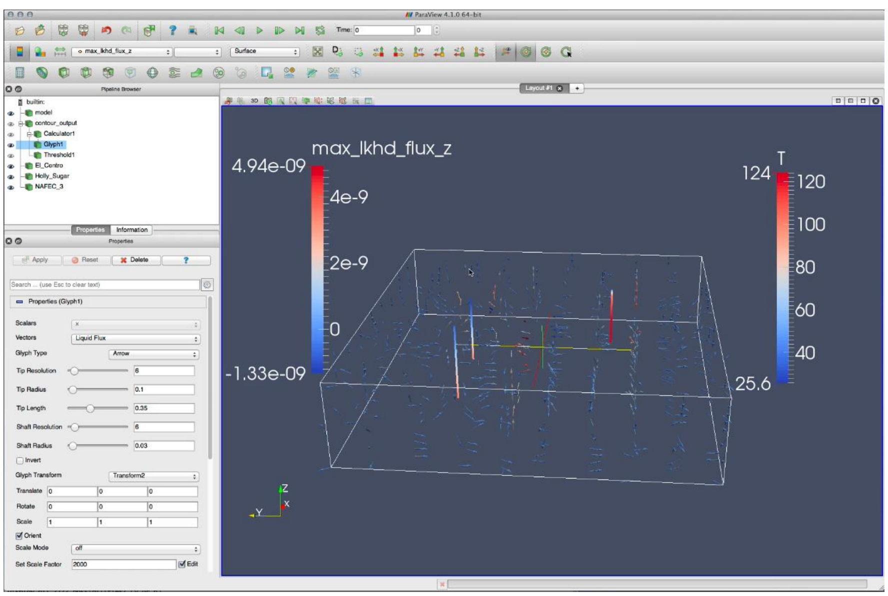

Figure 2: Flow vectors for the most likely scenario are presented along with the measured borehole temperatures colored by their vertical component in kg/s.

Surfaces extracted from the model colored by robustness are presented in Figure 3 with the more robust locations for geothermal prospectivity are red, less robust locations are blue. The measured borehole temperatures provide reference.

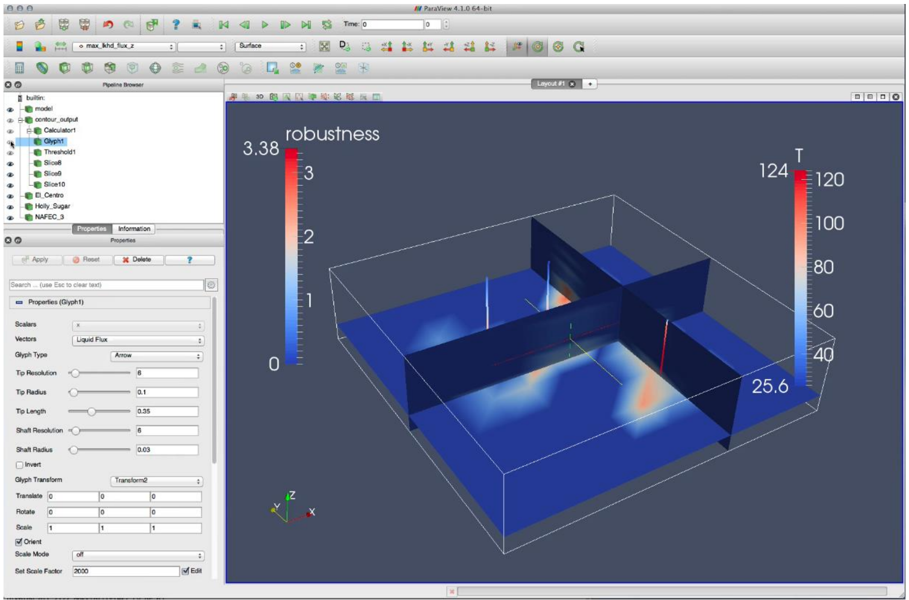

Figure 3: Model cross-sections colored by geothermal robustness with measured borehole temperatures for reference.

4. SUMMARY

We developed a framework for geothermal decision analysis and visualization. The framework is novel in its use of a non-probabilistic metric of prospectivity, defined as geothermal robustness, and its integration with open-source visualization software. The basic functionality of the framework, including directed sampling (identification of a set of parameter combinations consistent with observations), production simulations, calculation of geothermal robustness, and automatic visualization have been described and demonstrated.

REFERENCES

McCalman, L., O'Callaghan, S.T., Reid, A., Shen, D., Carter, S., Krieger, L., Beardsmore, G., Bonilla, E.V. and Ramos, F.T., Distributed Bayesian Geophysical Inversions, *Proceedings,* 39th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2014).

Mellors, R.J., Ramirez, A., Tompson, A., Chen, M., Yang, X., Dyer, K., Wagoner, J., Foxall, W. and Trainor-Guitton, W. Stochastic Joint Inversion of a Geothermal Prospect. *Proceedings*, 38th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California (2013).

Jardani, A., and A. Revil.: Stochastic joint inversion of temperature and self-potential data, *Geophysical Journal International* **179,** (2009), 640-654.

Gilks, W. R. (2005). Markov chain Monte Carlo. John Wiley & Sons, Ltd.

Reza, F. M. (1961). An introduction to information theory. Courier Corporation.

Ben-Haim, Y. (2001). Information-gap decision theory: decisions under severe uncertainty. Academic Press.

Harp, Dylan R., and Velimir V. Vesselinov: Contaminant remediation decision analysis using information gap theory. *Stochastic Environmental Research and Risk Assessment* **27**, no. 1 (2013), 159-168.

Doherty, John, L. Brebber, and P. Whyte.: PEST: Model-independent parameter estimation. *Watermark Computing*, Corinda, Australia (1994).

Adams, Brian M., W. J. Bohnhoff, K. R. Dalbey, J. P. Eddy, M. S. Eldred, D. M. Gay, K. Haskell, Patricia D. Hough, and L. P. Swiler.: DAKOTA, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: version 5.0 user's manual. *Sandia National Laboratories*, Tech. Rep. SAND2010-2183 (2009).

Van Der Walt, Stefan, S. Chris Colbert, and Gael Varoquaux.: The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering* **13**, no. 2 (2011): 22-30.

Jones, Eric, Travis Oliphant, and Pearu Peterson.: SciPy: Open source scientific tools for Python. (2014).

Foreman-Mackey, Daniel, David W. Hogg, Dustin Lang, and Jonathan Goodman.: emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific* **125**, no. 925 (2013): 306-312.

Newville, Matthew, Till Stensitzki, Daniel B. Allen, and Antonino Ingargiola.: LMFIT: non-linear least-square minimization and curve-fitting for Python. *Zenodo*, http://dx. doi. org/10.5281/zenodo (2014).

Henderson, Amy, Jim Ahrens, and Charles Law.: The ParaView Guide. Clifton Park, NY: *Kitware*, (2004).

Zyvoloski, G.: FEHM: A control volume finite element code for simulating subsurface multi-phase multi-fluid heat and mass transfer, *Los Alamos Unclassified Report* LA-UR-07-3359 (2007).

Lichtner, P. C., G. E. Hammond, C. Lu, S. Karra, G. Bisht, B. Andre, R. T. Mills, and J. Kumar: PFLOTRAN User Manual. *Tech. rep., 2013*. 159, 167, (2013).

Kelkar, Sharad, George Zyvoloski, and Zora Dash.: Pressure testing of a high temperature naturally fractured reservoir. *Proceedings*, 11th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California (1986).

Lavine, Michael.: Sensitivity in Bayesian statistics: the prior and the likelihood, *Journal of the American Statistical Association* **86**, no. 414 (1991), 396-399.

Vanpaemel, Wolf.: Prior sensitivity in theory testing: An apologia for the Bayes factor, *Journal of Mathematical Psychology* **54**, no. 6 (2010), 491-498.

Müller, Ulrich K.: Measuring prior sensitivity and prior informativeness in large Bayesian models, *Journal of Monetary Economics* **59**, no. 6 (2012), 581-597.

Berger, James O.: Robust Bayesian analysis: sensitivity to the prior, *Journal of statistical planning and inference* **25**, no. 3 (1990), 303-328.

Berger, J.O., Moreno, E., Pericchi, L.R., Bayarri, M.J., Bernardo, J.M., Cano, J.A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B. and Dasgupta, A.,: An overview of robust Bayesian analysis. *Test*, **3,** no. 1 (1994), 5-124.

Insua, David Ríos, and Fabrizio Ruggeri, eds.: Robust Bayesian Analysis. Vol. 152. Springer Science & Business Media, (2012).

Tiedeman, A., S. Bjornstad, S. Alm, L. Frazier, D. Meade, C. Page, M. Lazaro, J. Woolford, and B. Crowder.: Intermediate depth drilling and geophysical logging results at superstition mountain, Naval Air Facility El Centro, California, *GRC Transactions* **35** (2011).

Theis, Charles V.: The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using ground-water storage, *Transactions, American Geophysical Union*, **16**, (1935), 519-524.