PROCEEDINGS, 47th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 7-9, 2022
SGP-TR-223

Stochastic Fault Models Constrained to Seismic Observations

Whitney J Trainor-Guitton, Hao Zhang, Carl Hoiland

2365 Mountain Vista Lane #5, Provo, UT 84606

whitney@zanskar.us

Keywords: stochastic, seismic, fault modeling

ABSTRACT

We present a methodology that combines stochastic fault building combined with physics-informed neural network-based seismic modeling. First, we explore all possible fault locations given observed prior information. Next this uncertainty is updated with observed seismic information. This methodology provides probabilistic spatial locations of faults constrained by the seismic observations. We demonstrate this with geologic contact information and seismic observations from an operating geothermal field.

1. INTRODUCTION

Presence of faults provides a proxy to permeability in geothermal reservoirs (Faulds et al., 2015; Siler et al., 2016), which is the basis for understanding both conceptually and concretely, how fluids are circulating. Quantifying the uncertainty in the possible geologic faults and lithologies is key for early and late stages of the geothermal reservoir lifetime: from understanding the conceptual model to reservoir management. Stochastic modeling provides uncertainty measurements on the geologic model (Bond, 2015; Wellmann and Caumon, 2018). Previous stochastic geothermal modeling of active geothermal reservoirs include Tompson et al. (2013) and Chen et al. (2014) focused on hydrothermal flow modeling both alone and coupled with DC electrical resistivity. Pollack et al. (2021) explored the combined uncertainty in gravity, magnetic, and tracer data by modifying the lithology and fault locations. Combining these diverse data could improve the understanding of the reservoir, where one data's limitation is supplemented by another's strength. However, as shown in Pollack et al. (2021), the gravity data did not improve the stochastic modeling results to a substantially lower error. All geophysical methods suffer from non-uniqueness, however potential field methods rely on diffusive phenomena, making it harder for the data to image sharp boundaries.

Seismic has had success in mapping the geometry of the faults in the subsurface, even in the challenging onshore environments in the Basin and Range (Queen et al., 2016; Trainor-Guitton et al., 2019). Although seismic provides more spatially extensive observations of faults in the subsurface, there is considerable uncertainty in both the measurements and the resulting models. Acquisition geometry, limited sources and receivers, uncertainty in the velocity values of the heterogeneous 3D subsurface are just a few of the challenges that contribute to the imperfectness of the seismic method. Stochastic approaches are a viable way to understand these uncertainties, but often the limitation to stochastic modeling is its computational expense. Most stochastic seismic examples are in 2D because 3D can be prohibitive (Chen and Hoversten, 2012).

This paper models 3D uncertainty in lithology and fault boundaries and explores how seismic signals can exclude possible geological structure because of poor fit to observed data. For the geological modeling, we utilize the open-source software GemPy (Varga et al., 2019), and to improve the computational efficiency for computing seismic response on hundreds of geologic models, we use the recently developed deep-learning based EikoNet (Smith et al., 2021). The stochastically generated geologic models are populated with seismic velocities and used to train the EikoNet to efficiently produce RMS misfits of seismic data. Specifically, we model and utilize seismic field observations from Lightning Dock, New Mexico (LDG for short) using a recorded string shot survey.

2. STOCHASTIC GEOLOGICAL & FAULT MODELING WITH GEMPY

Varga et al., (2019) present GemPy: an open-source product for stochastically modeling 3D geologic structures, both faults and horizontal strata. GemPy is Python-based API that utilizes GPU and Bayesian inference frameworks, making stochastic geological modeling and inversions computationally feasible. This section will describe how GemPy was used to calculate the uncertainty in the geologic boundaries of the LDG reservoir.

2.1 LDG Basecase geomodel

The LDG area of interest for modeling encompasses a 5.3km by 5km area and 2.2km in depth from a top elevation of 1400m to a depth of 840m below sea level. This volume is discretized into nx=100, ny=100, and nz=66 cells, such that each cell represents dx=53m, dy=50 and dz=21.2m. To build any geomodel in GemPy, a minimum of two contact points (location in 3D space) and one orientation vector (azimuth and dip) is needed for each "lithoseries," which is how each geologic units are called as GemPy, which utilizes Pandas dataframes.

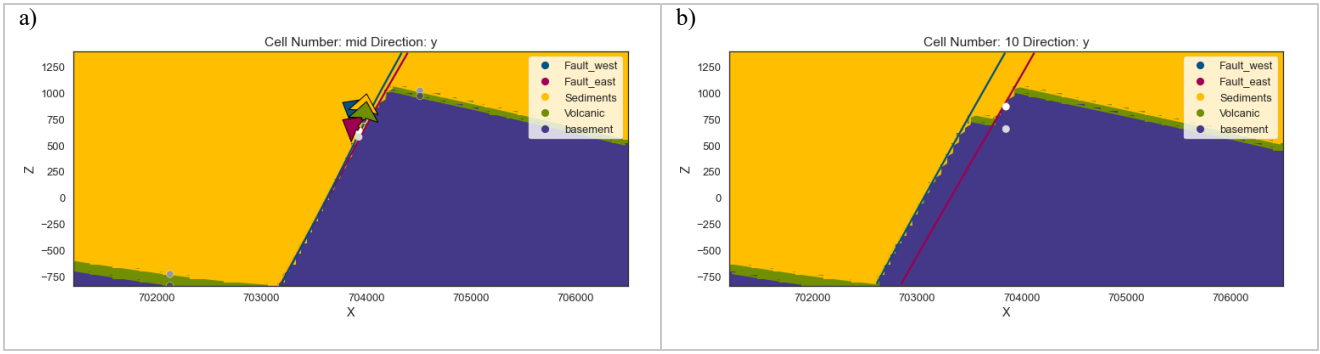


Figure 1: Cross section view of the base geomodel, showing the west (blue) and east (magenta) fault, the sediments (yellow), volcanics (green), and basement (purple). a) shows for yslice=50 and shows the orientation vector for each lithoseries. b) yslice=10

Using universal co-Kriging equations, GemPy utilizes the contacts, orientations, and the lithoseries order and bottom relationships (e.g. erosion, fault), to solve the spatial interpolation. Figure 1 contains two cross sections along the X direction for the LDG basecase model, y_{slice}=50 (middle) and y_{slice}=10 (southern). The two faults overlap as you proceed north in the model area.

2.1 Stochastic geologic modeling: varying the litho-series contacts and orientations

After creating the base model, variations to these XYZ contacts and the orientation dips can be varied via the Markov Chain Monte Carlo (MCMC)-Bayesian package pymc3. By modifying the contact and orientations, we can represent our uncertainty in the actual location and orientations of the geologic units and features. A normal distribution with mean of 0 and a defined sigma is used to modify the initial (basecase) contacts and dips; these new parameters are passed back to GemPy to create a new model "sample" for each modified contact and/or orientation sample. Table 1 contains the first four experiments, where the first four isolate either the Volcanics/Sediments or the Faults when modifying the Z contact location or the dip angle.

Table 1: Different geologic variation experiments, listing which geologic units or features (lithoseries) modified, by what magnitude (sigma) and the average entropy of all 200 samples

	Z contact: Lithoseries & sigma	Dip Angle: Lithoseries & sigma	Average Entropy
Experiment 1	Sediments: 5 meters Volcanic: 5 meters	NA	0.238
Experiment 1b	Fault West: 5 meters Fault East: 5 meters	NA	0.228
Experiment 2	NA	Sediments: 5 degrees Volcanic: 5 degrees	0.159
Experiment 2b	NA	Fault West: 5 degrees Fault East: 5 degrees	0.038
Experiment 3	Sediments: 5 meters Volcanic: 5 meters	Sediments: 5 degrees Volcanic: 5 degrees	0.304

Figure 2 shows both the distribution of the Z contact perturbations (left) and the "chain" of these 200 perturbation values (right). These values are used for the first two experiments (Experiment 1 and 1b in Table 1). Figure 3 contains the distribution (left) and chain (right) of the dip perturbations for the Z component of the orientation pole vector (Gz), where +/-0.1Gz equivalates to 5 degrees in dip. These perturbations are used in third and fourth experiments (Experiment 2 and 2b in Table 1).

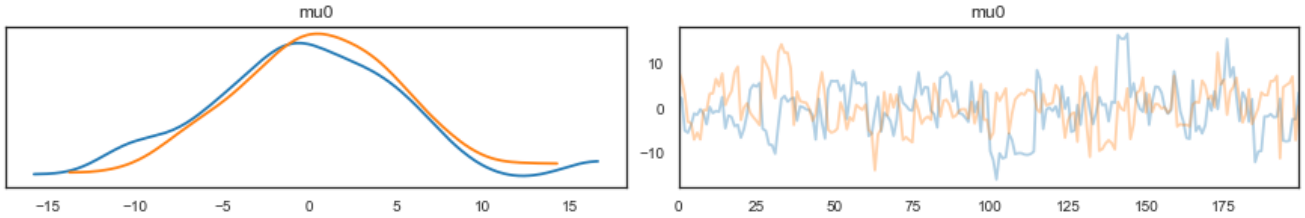


Figure 2: An example of the variability draws for the Z value of contact point for Experiment 1 (Table 1). LEFT: Posterior of variability added to basecase Z value of lithoseries with sigma=5 meters. RIGHT: the variability in series for all 200 samples.

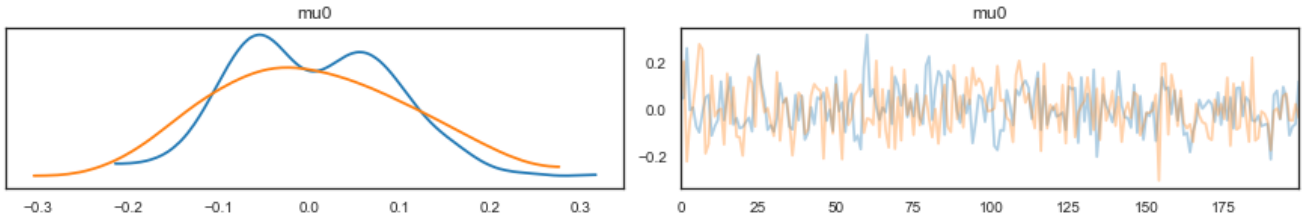


Figure 3: One example of the variability draws for Z component of the orientation vector. LEFT: the posterior probability density function with sigma=0.1 (unitless), which equivalates to +/-5 degrees. RIGHT: the variability in series for all 200 samples.

With 200 perturbation values added to the original contact and orientation poles, 200 new geologic models are generated for each of the experiments in Table 1. Figure 4 demonstrates the probability of sediments for two different cross sections (rows) and two different experiments (columns). The dark blue color represents a zero probability that the sediment lithoseries exists at the location, whereas yellow is 100%. Modifying the Z contact (left side of Figure 4) results in more intermediate probability values compared with modifying the dip angles (right side of Figure 4).

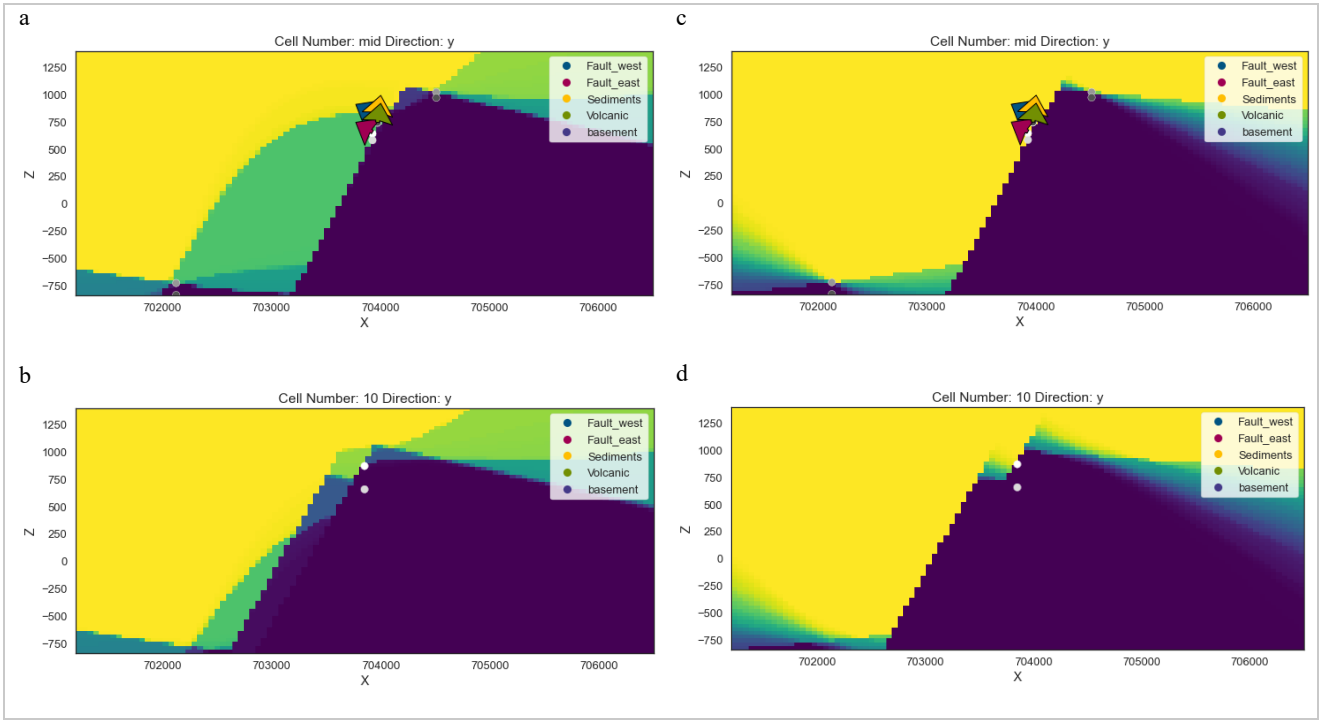


Figure 4: Probability of Sediments cross sections. Left side is Experiment 1 (a: y=50 and b: y=10) where the sediments' and volcanics' Z contacts are varied. Right side is Experiment 2 (c: y=50 and d: y=10) where the orientation angle of the sediments and volcanics are modified.

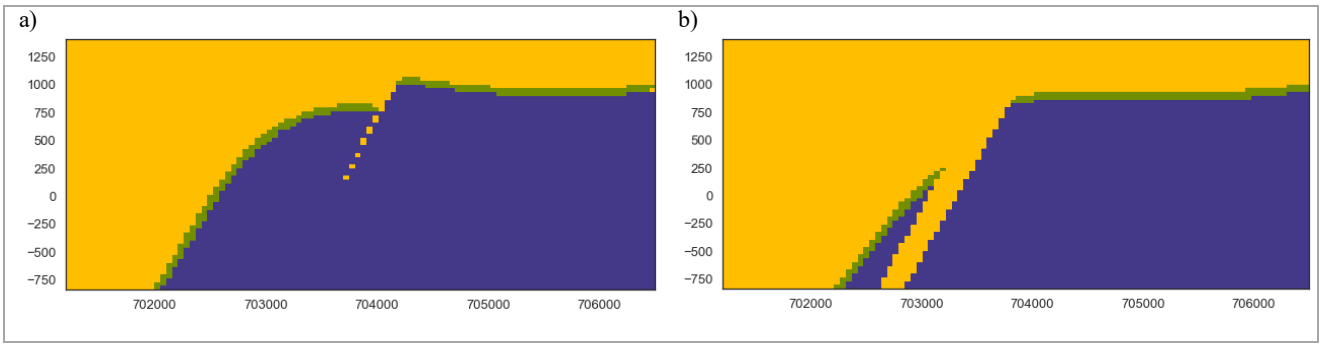


Figure 5: Lithology of sample 50 from Experiment 1 for y_{slice} = 10 (LEFT) and y_{slice} = 50 (RIGHT)

Most models are of similar shapes as the base geomodel shown in Figure 1, which is decipherable from the probability shapes of Figure 4. However, we also include Figure 5 to demonstrate an example model that creates the green, curvilinear probability patterns shown in Figure 4. Both cross sections of Figure 5 are from the same realization (e.g. geologic model number 50 of 200), showing 1) the y_{slice} = 10 and b) further south, y_{slice} = 50. In this case, the modified Z contacts of the Sediments and Volcanics, along with the cokriging solution of the surface, have created rounded geologic interfaces to the west of the two faults. Recall that in this experiment, only the Z contacts are modified and not the dips of any of the lithoseries. If both dips and azimuths were varied, the geologic unit shapes of this realization would be different.

Probability maps, like in Figure 4, are useful but can only demonstrates the probability for one of the lithoseries at a time. To capture a more complete picture of the geologic uncertainty, information entropy is useful. Entropy synthesizes the "disorder" at every location in space by comparing all samples within an experiment. It can informally be thought of the "potential for surprise" of any of the possible discrete outcomes. In other words, 0 means low disorder, low potential for surprise, therefore one geologic unit is dominating at that location. The formula for entropy used in this study is

$$H = -\Sigma_i^{N=3} P_i(x,y,z) log_2 P_i(x,y,z) \quad (1)$$

Where P_i is the probability of each lithoseries (Sediments, Volcanics or Basement) at the location in 3D space (x,y,z) (Wellmann and Caumon, 2018). Given that we have three (N=3) possible geologic classes, entropy is maxed at 1.58 ($log_2 N$). Table 2 shows four possible scenarios, where the first is when the probability of the three class is uniform and thus the entropy is the maximum at 1.58. The next three shift a high probability (0.9) along the three classes, which all produce an entropy low of 0.5. Therefore, the entropy value isn't unique to a specific arrangement of probabilities, just the statistical disorder.

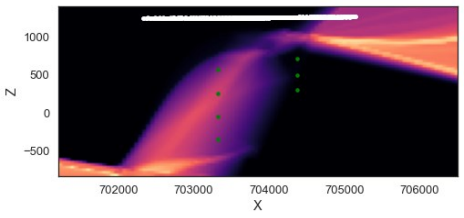
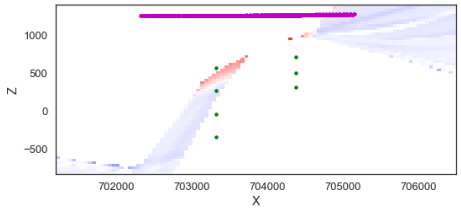
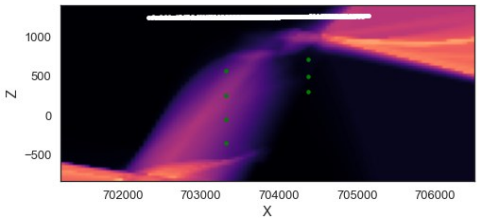
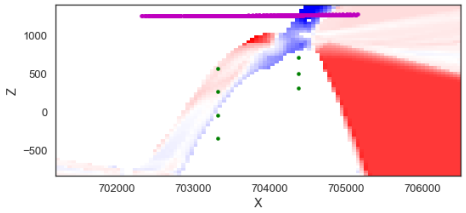
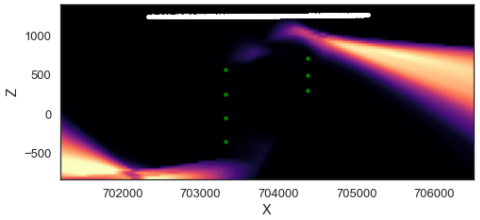
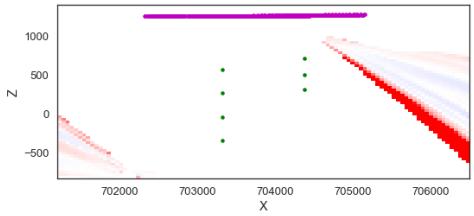
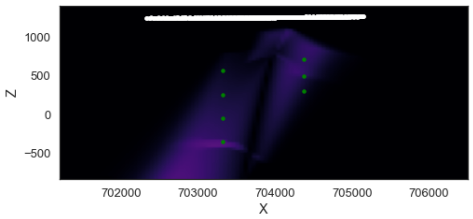
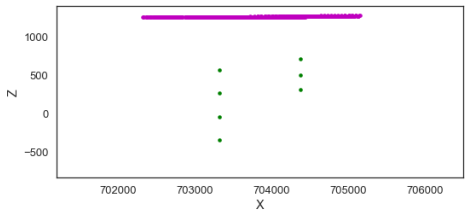
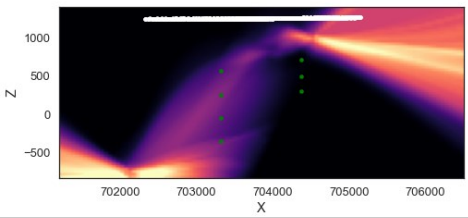
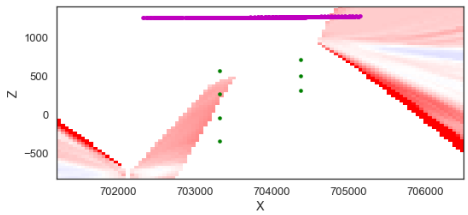
Table 2: Entropy examples

Pr(Θ= Sediments)	Pr(Θ= Volcanics)	Pr(Θ= Basement)	H
0.3	0.3	0.3	1.58
0.9	0.05	0.05	0.5
0.05	0.9	0.05	0.5
0.05	0.05	0.9	0.5

The arithmetic average of all the entropy values for each of the experiments is shown in the last column of Table 1. We see that Experiment 3 has the highest overall average entropy (0.304) which is expected since both the Z contact and dips are modified. Experiment 2b has the lowest (0.034), indicating that the cokriging solutions by modifying *only* the fault dips, are limited.

Table 3 provides cross section visuals of entropy for each of the experiments. These are averages across all y slices. Again, we see how Experiment 2b has very low entropy: the 200 sample models are very similar. Also Experiment 3 displays many areas of high entropy. Additionally plotted in these entropy images are the locations of LDG geophone stations in white and the string shot locations in green. The next section will describe how the seismic modeling is performed.

Table 3: Experiments performed, average entropy of models generated, and reduction in entropy of those fitting to seismic data. The geophone locations are shown in white (entropy) and magenta (entropy reduction), and the string shots in green (both).

	Contact: sigma	Orient-s: sigma	Average Entropy Cross Section: 200 models (0 – 0.4 – 0.8 – 1.2 – 1.4)	Reduction Factor in Average Entropy: seismic RMS < 1 (-0.3 – 0 – +0.3)
Experiment 1	Sediments: 5 meters Volcanic: 5 meters	NA		
Experiment 1b	Fault West: 5 meters Fault East: 5 meters	NA		
Experiment 2	NA	Sediments: 5 degrees anic: 5 degrees		
Experiment 2b	NA	Fault West: 5 degrees Fault East: 5 degrees		
Experiment 3	Sediments: 5 meters Volcanic: 5 meters	Sediments: 5 degrees Volcanic: 5 degrees		

3. EIKONET

This methodology employs the recently developed Eikonet, which solves the Eikonal equation with a deep neural network (Smith et al., 2021). The Eikonal equation is a nonlinear first-order PDE representing a high-frequency approximation to the propagation of waves in heterogeneous media. The equation takes the general form

$$\|\nabla T_{s \to r}\|^2 = \frac{1}{V(\vec{x}_r)^2} = S(\vec{x}_r)^2 \tag{2}$$

Where $\|.\|^2$ is the Euclidean norm, $T_{s \to r}$ is the travel-time from a source location s to receiver location r, V is the velocity of the medium at the receiver location x_r, and S is the slowness at receiver location x_r. respectively. EikoNet aims to accurately predict travel times given source-receiver pairs, but equation 2 has a strong singularity at the source location (Treister and Halder, 2016). To avoid the computational error caused by numerical singularity when the receiver is close to the source, a factored travel-time (Treister and Halder, 2016) is chosen as

$$T_{s \to r} = T_0 \cdot \tau_{s \to r} \tag{3}$$

5

where T_0 is the distance from the source to the receiver, and $\tau_{s\to r}$ is the travel time deviation from a homogeneous unity velocity model (V=1 km/s). Substituting equation 3 into 2, the velocity at the receiver is represented by

$$V(\vec{x}_r) = [T_0^2 \|\nabla_r \tau_{s\to r}\|^2 + 2\tau_{s\to r}(\vec{x}_r - \vec{x}_s) \cdot \nabla_r \tau_{s\to r} + \tau_{s\to r}^2]^{-1/2} \quad (4)$$

A EikoNet (Smith et al., 2020) is composed of four fully connected layers and 10 residual blocks. The prediction of neural network (NN) is actually $\tau_{s\to r}$, and through the back-projection of the NN we can get $\nabla_r \tau_{s\to r}$. Next, using equation 4, the velocity at any receiver can be predicted accordingly. With a mean-squared error loss function for a given velocity model

$$L = \|V - \hat{V}\|^2 \quad (5)$$

EikoNet can be optimized by gradually reducing the loss function given a set of input source-receiver pairs.

To train the EikoNet, locations of 1×10^5 source-receiver pairs are randomly sampled in the given velocity model. This number of source-receiver pairs is guaranteed to sample the study regions enough even for relatively complex models (Smith et al., 2020). The input for NN is represented by $\vec{x} = (\vec{x}_s, \vec{x}_r)$, and the observations are indicated by $V(\vec{x}_r)$. To balance the computational speed and prediction accuracy, 50 iterations and a batch size of 752 are used to update the NN for each velocity model. All the training of NN is on a single Nvidia GeForce RTX 2070.

3.1 Advantages and disadvantages

EikoNet is a mesh-independent physics-informed neural network (Smith et al., 2019). It can learn from and generalize information embraced in the training dataset, and this leads to fast prediction for any new input point close to some in the training dataset. Conversely, for the most typical travel time calculator, finite-difference method (FDM, Capozzoli et al., 2013), every source-receiver pair needs to be calculated at the expense of the same amount of time. The travel time table is not necessary for EikoNet because the travel time solution for any two points in the given method is valid (Smith et al., 2020). This leads to the hard drive storage of the EikoNet being a linear function of the network size (~90 MB for ten residual layers), but that of the travel time table from the FDM dramatically increasing with the number of source-receiver pairs. However, it is time consuming to train EikoNet from scratch. One way to reduce the training time is to utilize transfer learning to allow for quick updates of pre-trained models. Once EikoNet is trained, it is really fast to predict travel times for given inputs (e.g. 0.424 s for 1 million source-point pairs on a single Nvidia Tesla V100; Smith et al., 2020). For example, for the 200 models of Experiment 1b (Table 1), if no pretraining is used, it takes 184 seconds for computing the RMS misfit of each model to the LDG observed data. If a pretrained NN using the basecase velocity model, this decreases to 54 seconds.

3.2 The string shot dataset

The string shot data set consists of 7 source locations downhole that are recorded by 1,206 stations on the surface. Their coverage in X and depth are shown in Table 3. Please see Edwards et al. (2021) for more details on the dataset.

4. RESULTS: ENTROPY REDUCTION OF MODELS WITH RMS<1

Each of the possible geologic models generated from modifying the contacts and dips, are converted to velocity models according to the location of their geologic units. Then the RMS misfit is calculated for observed travel times and the LDG observed travel times. Of the 200 models, only those that have a RMS <1 with the LDG observed data are kept to compute the seismic-constrained entropy. The last column Table 3 demonstrates the factor reduction (in red) of entropy compared to the entropy calculated for all 200 models. There are locations where the entropy of the seismically constrained model ensemble actually higher, but in general it is lower.

Recall that Table 3 is of the average entropy. We can also view the entropy for specific slices within the model and compare the entropy at those slices for the ensemble of only the seismically-constrained models (RMS<1), without any averaging. This is useful when targeting specific locations within the subsurface. Figure 5a shows the entropy at the three middle slices for x, y and z, for all 200 models and for just those with RMS<1. To better visualize the difference, Figure 5b shows the entropy factor reduction. Where there is blue, the entropy hasn't been reduced with the subset of models, so these locations are not best imaged with the current seismic acquisition geometry. However, where it is red, the probabilities of different geologic units have been focused into one unit.

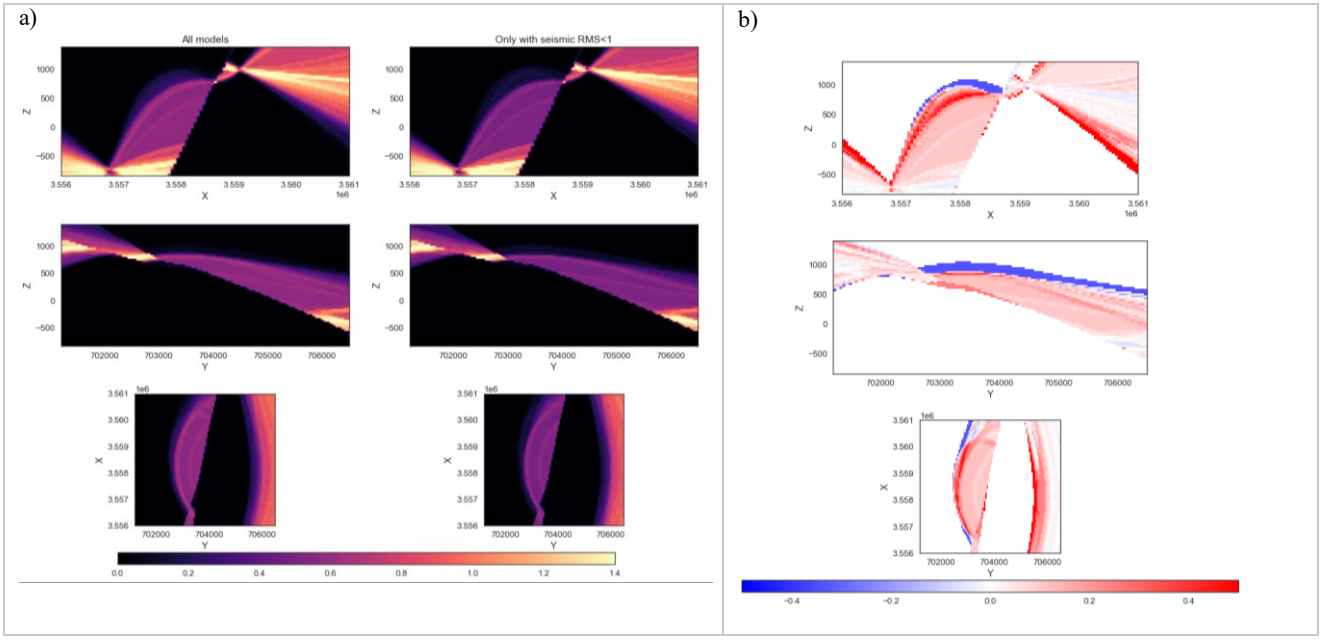


Figure 5: a) Entropy of all models and RMS <1 ensemble for cross sections xslice=50, yslice=50, zslice=33 b) Reduction of entropy for these specific slices.

CONCLUSIONS

This paper presents the combination of stochastic geologic modeling of GemPy along with the efficient seismic modeling of EikoNet. Previous stochastic modeling combined with geophysical observations have not considered 3D seismic modeling, partially because seismic modeling can be more computationally expensive. Additionally, MT and potential field methods have been historically used more for geothermal applications.

We have shown how entropy reduction can identify locations within the 3D subsurface where the seismic observations can reduce geologic uncertainty, in other words, eliminate the likelihood of certain geologic realizations. The development of efficient 3D geological uncertainty evaluation with seismic modeling can lead to more realistic and useful risk tools for both siting new wells and improving brownfield performance, as seismic information can delineate faults and geologic boundaries with more precision than diffuse geophysical methods.

ACKNOWLEDGMENTS

We thank Cryq, who collected the string shot data, for the permission to present this work.

REFERENCES

Bond, C.E., 2015, Uncertainty in structural interpretation: Lessons to be learnt: Journal of Structural Geology, v. 74, p. 185–200, doi: 10.1016/j.jsg.2015.03.003.

Capozzoli, A., C. Curcio, A. Liseno, and S. Savarese, 2013, A comparison of fast marching, fast sweeping fast iterative methods for solution eikonal equation, in Proc. 21st Telecommun. Forum Telfor (TELFOR), p. 685–688.

Chen, J., and Hoversten, G.M., 2012, Joint inversion of marine seismic AVA and CSEM data using statistical rock-physics models and Markov random fields: Geophysics, v. 77, p. R65--R80, doi: 10.1190/geo2011-0219.1.

Chen, M., Tompson, A.F.B., Mellors, R.J., Ramirez, A.L., Dyer, K.M., Yang, X., and Wagoner, J.L., 2014, An efficient Bayesian inversion of a geothermal prospect using a multivariate adaptive regression spline method: Applied Energy, v. 136, p. 619–627, doi: 10.1016/j.apenergy.2014.09.063.

Edwards, J., Hoiland, C; Fleure, T., Sicking, C., McLain, B., Vermilye, J., Witter, J., Tanner, N., Cladouhos, T., 2021. Seismic Imaging of Resonating Fracture Networks at the Lightning Dock Geothermal Field, Hidalgo County, New Mexico: Geothermal Geothermal Resources Council Transactions, Volume 45, Pages 1492-1499.

Faulds, J.E., Hinz, N.H., Coolbaugh, M.F., Shevenell, L.A., Siler, D.L., Craig, M., Hammond, W.C., Kreemer, C., Oppliger, G., Wannamaker, P.E., Queen, J.H., and Visser, C.F., 2015, Integrated Geologic and Geophysical Approach for Establishing Geothermal Play Fairways and Discovering Blind Geothermal Systems in the Great Basin Region , Western USA : A Progress Report: Geothermal Research Council Transactions, v. 39, p. 691–700.

Pollack, A., Cladouhos, T.T., Swyer, M.W., Siler, D., Mukerji, T., and Horne, R.N., 2021, Stochastic inversion of gravity, magnetic, tracer, lithology, and fault data for geologically realistic structural models: Patua Geothermal Field case study: Geothermics, v. 95, p. 102129, doi: 10.1016/j.geothermics.2021.102129.

Queen, J.H., Daley, T.M., Majer, E.L., Nihei, K.T., Siler, D.L., and Faulds, J.E., 2016, Surface Reflection Seismic and Vertical Seismic Profile at Brady ' s Hot Springs , NV , USA: 41st Workshop on Geothermal Reservoir Engineering, p. SGP-TR-209.

Siler, D.L., Hinz, N.H., Faulds, J.E., and Queen, J., 2016, 3D analysis of geothermal fluid flow favorability: Brady's, Nevada, USA: The 41st Workshop on Geothermal Reservoir Engineering, Stanford University, p. SGP-TR-209.

Smith, J.D., Azizzadenesheli, K., and Ross, Z.E., 2021, EikoNet: Solving the Eikonal Equation with Deep Neural Networks: IEEE Transactions on Geoscience and Remote Sensing, v. 59, p. 10685–10696, doi: 10.1109/TGRS.2020.3039165.

Tompson, a F.B., Mellors, R.J., Ramirez, a, Chen, M., Dyer, K., Yang, X., Wagoner, J., and Trainor-Guitton, W.J., 2013, Evaluation of a Geothermal Prospect Using a Stochastic Joint Inversion Modeling Procedure: Geothermal Resources Council Transactions, v. 37, p. 349–356.

Trainor-Guitton, W., Guitton, A., Jreij, S., Powers, H., and Sullivan, B., 2019, 3D Imaging of Geothermal Faults from a Vertical DAS Fiber at Brady Hot Spring, NV USA: Energies, v. 12, doi: 10.3390/en12071401.

Treister, E., and E. Halder, 2016, A fast marching algorithm for the factored eikonal equation, *J. Comput. Phys.*, v. 324, p. 210–225.

Varga, M. de, Schaaf, A., and Wellmann, F., 2019, GemPy 1.0: open-source stochastic geological modeling and inversion: Geoscientific Model Development, v. 12, p. 1–32, doi: https://doi.org/10.5194/gmd-12-1-2019.

Wellmann, F., and Caumon, G., 2018, 3-D Structural geological models: Concepts, methods, and uncertainties: Advances in Geophysics, v. 59, p. 1–121, doi: 10.1016/bs.agph.2018.09.001.