Check for updates

JGR Solid Earth

RESEARCH ARTICLE
10.1029/2021JB021767

Probabilistic Evaluation of Geoscientific Hypotheses With Geophysical Data: Application to Electrical Resistivity Imaging of a Fractured Bedrock Zone

Alex Miltenberger[1], Sebastian Uhlemann[2], Tapan Mukerji[1,3,4], Ken Williams[2], Baptiste Dafflon[2], Lijing Wang[4], and Haruko Wainwright[2]

[1]Department of Geophysics, Stanford University, Stanford, CA, USA, [2]Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [3]Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA, [4]Department of Geological Sciences, Stanford University, Stanford, CA, USA

Key Points:

- Machine learning is integrated into the Popper-Bayes framework to test conceptual geologic hypotheses with geophysical data
- The approach is demonstrated using electrical resistivity tomography (ERT) data from a mountainous watershed to study the geological origins of a zone of highly fractured shale
- Two of six hypotheses are falsified by the ERT data, showing the utility and limitations of ERT data for imaging this fractured bedrock zone

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Miltenberger,
ammilten@stanford.edu

Citation:

Miltenberger, A., Uhlemann, S., Mukerji, T., Williams, K., Dafflon, B., Wang, L., & Wainwright, H. (2021). Probabilistic evaluation of geoscientific hypotheses with geophysical data: Application to electrical resistivity imaging of a fractured bedrock zone. *Journal of Geophysical Research: Solid Earth*, *126*, e2021JB021767. https://doi.org/10.1029/2021JB021767

Received 25 JAN 2021
Accepted 24 AUG 2021

© 2021. American Geophysical Union. All Rights Reserved.

Abstract As climate changes and populations grow, groundwater sustainability is becoming increasingly important. Hydrogeologic models, which are based on a conceptual understanding of the subsurface, are crucial tools for informing decisions. Conceptual models of the subsurface incorporate knowledge of geological processes, and, frequently, observations from geophysical data into a common subsurface parameterization where the parameters may still be uncertain. Many methods exist to test how different conceptual subsurface parameterizations compare to geophysical data, but a frequent problem in hydrogeologic model development occurs when multiple geological phenomena could explain a single subsurface parameterization. In this work, we present a framework for testing geological hypotheses in conditions where a geological feature is observed in geophysical data, but its physical characteristics are uncertain. The framework uses Popper-Bayes methods developed in previous studies, and is applied to study a fractured bedrock zone in a mountainous watershed in southwest Colorado. First, we propose six hypotheses based on the geological history of the watershed. Then, using the proposed Popper-Bayes approach, we demonstrate that two of the hypotheses are inconsistent with the electrical resistivity tomography data. Finally, we discuss the importance of the prior model, and in what other scenarios the framework can be applied to.

Plain Language Summary Groundwater is a major source of fresh water throughout the world, yet much is still unknown about where and how water moves underground. One big reason for this is that it is hard to map out underground flow paths from the surface. One technique to map underground flow paths is to send electrical signals deep into the ground and measure how the ground reacts to the signals. In this work, we show how these measurements can be used to understand the geologic history of an area, which in turn helps to identify underground flow pathways. The data interpretation method we propose in this paper considers multiple potential causes for the formation of underground flow pathways. Then, we rank the scenarios by how consistent they are with the measured data. We demonstrate this method in the Elk Mountains of Central Colorado near a zone of highly fractured bedrock, which water easily flows through compared to the surrounding rock. Through the proposed method we show that two of the six potential causes for the creation of the fracture zone are inconsistent with the electrical data collected in the field.

1. Introduction

Sustainable management of groundwater aquifer systems is becoming increasingly important (Castilla-Rho et al., 2019), as groundwater resources become more heavily utilized in response to climate change and population growth. Key to sustainable groundwater management is the development of groundwater models that predict system behavior especially as concerns future changes in hydrologic processes (Taylor et al., 2013). However, accurate groundwater models are difficult to design in part because of large uncertainty arising from the complexity in subsurface heterogeneity and imperfect conceptual understanding of the subsurface structure (Refsgaard et al., 2012). To address these challenges researchers have developed methods for including conceptual hydrological uncertainty into groundwater model predictions (Reeves et al., 2010; Tsai, 2010; Ye et al., 2010). Additionally, geophysical data are often used to improve conceptual

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

understanding of the subsurface, which aids in the design of hydrogeologic models for groundwater applications (Binley et al., 2015; Linde et al., 2015). In particular, geophysics is powerful in identifying and mapping certain geological features, like faults, sedimentary layers, and buried channels that can have a significant impact on subsurface flow (Christensen et al., 2020; Høyer et al., 2015; Jiang et al., 2019; Knight et al., 2018). However, it is known that the interpretation of geophysical data has significant uncertainty, arising from both measurement errors and the nonuniqueness of geophysical inversion (Malinverno & Briggs, 2004). A key part in developing a conceptual understanding of the subsurface geology is, therefore, the process of developing conceptual geologic hypotheses and testing them against geophysical data in a manner that accounts for various uncertainties such as structure and petrophysical properties (Enemark et al., 2020).

In the last several years, Popper-Bayes methods—which attempt to falsify conceptual hypotheses while accounting for measurement and prior uncertainty (Tarantola, 2006)—have been increasingly used in the Earth sciences (Enemark et al., 2020; Hermans et al., 2015; Park et al., 2013; Phelps et al., 2018; Pirot et al., 2019; Scheidt et al., 2015). In general, Popper-Bayes hypothesis testing starts with the quantitative definition of several hypotheses about the subsurface to be tested with some measured data. Monte Carlo methods are then used to create many realizations of synthetic data, which are then compared to measured data using Bayesian methods and used to estimate the posterior probability of each hypothesis. Most Popper-Bayes studies in the Earth Sciences literature have the goal of determining which subsurface parameterizations, or conceptual models, agree best with measured data. For example, Hermans et al. (2015) used a Popper-Bayes framework to test different spatial patterns of gravel deposition against electrical resistivity tomography (ERT) data, Pirot et al. (2019) test several models that simulate fluvial sediment organization against ground-penetrating radar data, and Enemark et al. (2020) use a combination of seismic refraction and hydrologic data to determine which combinations of subsurface structures and hydrogeologic processes create surface depressions in Australia. While these studies consider the problem of determining which subsurface parameterizations can best explain the data, more can be learned about geologic processes by testing hypotheses within a conceptual subsurface model.

In this work, we consider the problem where a subsurface feature has been identified and can be parameterized by a small number of variables, but its geological origins are unknown. We demonstrate how a Popper-Bayes framework can be used for this class of problems using a case study where a zone of fractured bedrock in a mountainous watershed has been identified through ERT and borehole measurements. The shape of the fractured bedrock zone is related to different geological phenomena, including depositional and faulting processes, which each have different impacts on watershed-scale hydrology. However, uncertainty in the ERT measurements, as well as the nonunique inverse relationship between the ERT data and the subsurface properties, make it challenging to test the geological hypotheses from inverted ERT sections alone.

This paper is organized as follows. In Section 2, we describe the Popper-Bayes framework for updating testing geological hypotheses within a conceptual subsurface model. In Section 3, we describe the study area and data in more detail. In Section 4 we outline six conceptual hypotheses. In Section 5, we apply the Popper-Bayes framework to calculate posterior probabilities of each hypothesis. Finally, in Section 6, we discuss the case study results and the proposed framework in general.

2. Popper-Bayes Hypothesis Testing Methods

The Popper-Bayes method we use in this work is based on previous work by Enemark et al. (2020), Hermans et al. (2015), Pirot et al. (2019), Scheidt et al. (2015), and others. The workflows in each of these studies consists of three parts, shown in Figure 1. First, hypotheses are defined, a prior model is created, and Monte Carlo simulations are used to generate a prior set of geophysical data and corresponding model parameters. Second, dimension reduction is performed on the high-dimensional geophysical data. Third, the posterior distributions of model parameters are efficiently sampled in the low-dimensional space. The workflow in this study follows these three steps, which are described in more detail below.

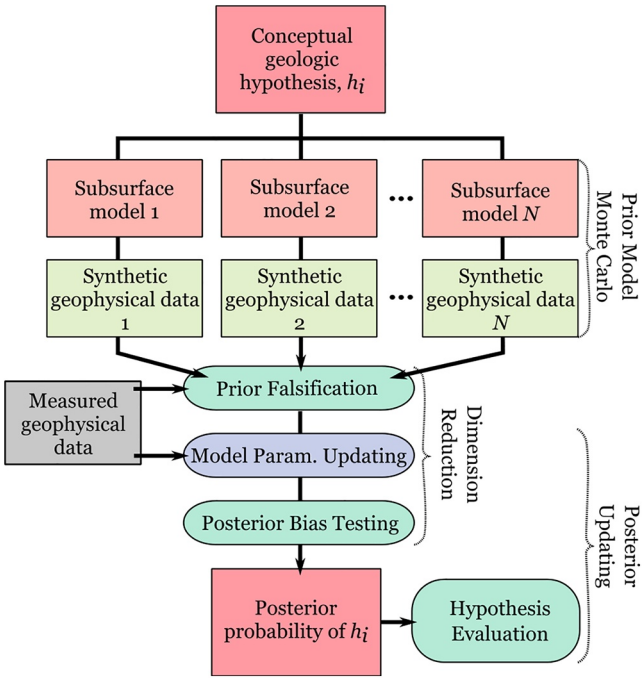


Figure 1. Workflow diagram of a Popper-Bayes hypothesis test, in which the hypothesis is related to an underlying geologic mechanism and the measured data are from a geophysical survey.

2.1. Bayesian Uncertainty Quantification Scheme

The main difference between the problem considered in this work and those presented by Enemark et al. (2020), Hermans et al. (2015), and others, is that in this study we assume the conceptual subsurface model is known, and the hypotheses depend on the values of the subsurface model parameters. The set of parameters used to describe the conceptual subsurface model is denoted by $\boldsymbol{\theta} = [\theta_1,\ldots,\theta_K]^T$, where K is the number of parameters in the model. The parameters can include any variable describing the geological features or petrophysical properties of the subsurface. In addition, the finite set of hypotheses are denoted by $\mathcal{H} = \{h_1,\ldots,h_M\}$, where M is the number of hypotheses. Any set of geophysical data, synthetic or measured, is denoted by $\mathbf{z} = [z_1,\ldots,z_P]^T$ where P is the number of measurements in the data set. The measured, or observed, geophysical data are denoted by $\mathbf{z}_{\text{obs}}$. Therefore, the posterior distribution of the -th hypothesis, h_i, after conditioning on the observed geophysical data can expressed by $p(h_i|\mathbf{z}_{\text{obs}})$. Under these conditions, the posterior distribution of the hypotheses can be approximated by the following equation, derived in Appendix A.

$$p(h_i|\mathbf{z}_{\text{obs}}) \overset{\propto}{\sim} p(h_i)\int p(\boldsymbol{\theta}|h_i)\, p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\boldsymbol{\theta} \tag{1}$$

Based on Equation 1 the two main components of the hypothesis test are: (a) defining the relationship between the hypotheses and model parameters in $p(\boldsymbol{\theta}|h_i)$, and (b) defining the relationship between the measured geophysical data and the model parameters in $p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}})$. The term representing the conditional probability of model parameters given the observation, $p(\boldsymbol{\theta}|\mathbf{z}_{obs})$, can be expanded to separate uncertainty due to measurement error and uncertainty due to the non-uniqueness of geophysical inversion, according to the following expression, derived in Appendix B.

$$p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}}) \approx \int p(\boldsymbol{\theta}|\mathbf{z})\, p(\mathbf{z}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\mathbf{z} \tag{2}$$

where $p(\boldsymbol{\theta}|\mathbf{z})$ encodes the uncertainty associated with geophysical non-uniqueness and $p(\mathbf{z}|\mathbf{z}_{\text{obs}})$ encodes uncertainty and errors in the geophysical measurements.

2.2. Prior Model Monte Carlo

To perform the prior model Monte Carlo simulation, prior distributions are first defined for each conceptual subsurface model parameter (Hermans et al., 2015, 2018; Scheidt et al., 2018). If both the mapping from model parameters to subsurface model, and from subsurface model to geophysical data are deterministic, then $\mathbf{z}$ is deterministically mapped from $\boldsymbol{\theta}$ by the forward model $\mathbf{z} = f(\boldsymbol{\theta})$. Therefore, a set of N prior models and prior geophysical data may be obtained by generating a set of N realizations of $\boldsymbol{\theta}$ and forward modeling the corresponding geophysical responses $\mathbf{z}$.

2.3. Dimension Reduction

In a Popper-Bayes workflow, dimension reduction serves two purposes. First, the prior and measured data can be compared in reduced-dimension space to test whether or not the prior model adequately represents the key features in the geophysical data (Hermans et al., 2015; Scheidt et al., 2015; Tarantola, 2006). This is sometimes referred to as prior falsification. The second use for dimension reduction is efficient posterior updating (Lopez-Alvis et al., 2019; Pirot et al., 2019; Scheidt et al., 2015). The low-dimensional space is useful for posterior sampling because other posterior sampling algorithms, like Markov Chain Monte Carlo, are inefficient for high-dimensional problems (Hermans et al., 2018). In this work, we use different dimension reduction techniques for the prior model testing and posterior updating steps.

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3.1. PCA Dimension Reduction for Prior Falsification

Prior falsification refers to the process of determining whether a prior model can generate synthetic data consistent with the data measured in the field (Scheidt et al., 2015, 2018). If a prior model is falsified through this test that means that the prior is a poor representation of the subsurface. If the prior is not falsified, it does not mean the model is "correct," only that the model can adequately capture the main features of the measured data. To perform a prior falsification test we use a principal component analysis (PCA) approach to compare the prior realizations to the measured geophysical data in a reduced-dimension space, similar to the prior falsification approach used by Hermans et al. (2018). Each prior realization, $\mathbf{z}$, is treated as a sample with P features, where each feature is a single measurement. If we have N prior realizations, then the entire prior set of geophysical data can be stored in an $N \times P$ matrix $\mathbf{X}$. If $P > N$, then after performing PCA on $\mathbf{X}$ we obtain an $N \times N$ matrix of scores $\mathbf{S}$, where each row is a reduced-dimension sample, and a $P \times P$ matrix of eigenvectors $\mathbf{\Phi}$, where each column is an eigenvector of the empirical covariance matrix. The measured geophysical data can then be projected into the principal component (PC) space to test whether or not the measured data contains the same basic features as the forward-modeled data. This is accomplished using

$$\mathbf{S}_{\text{obs}} = (\mathbf{z}_{\text{obs}} - \overline{\mathbf{z}})\mathbf{\Phi}, \tag{3}$$

where $\mathbf{S}_{\text{obs}}$ are the principal component scores for the observed geophysical data and $\overline{\mathbf{z}}$ is the mean vector of the prior realizations of geophysical data. If $\mathbf{S}_{\text{obs}}$ is within the prior distribution of $\mathbf{S}$, for the principal components that explain most of the variance in the prior geophysical data, then the prior is not falsified (Hermans et al., 2018; Scheidt et al., 2015, 2018).

2.3.2. Machine Learning Dimension Reduction for Posterior Updating

In previous work, the dimension reduction used for prior model testing and posterior updating are the same. Hermans et al. (2015) and Scheidt et al. (2015) use distance metrics and multidimensional scaling to compare geophysical data to one another. Enemark et al. (2020), Lopez-Alvis et al. (2019), and Pradhan and Mukerji (2020) use feature extraction techniques to identify summary statistics that represent the key features in their data. If the summary statistic extracted from the data is defined by $g(\mathbf{z})$, then the posterior distribution of model parameters is then estimated via $p(\boldsymbol{\theta}|\mathbf{z}) \approx p(\boldsymbol{\theta}|g(\mathbf{z}))$ (Hermans et al., 2015; Lopez-Alvis et al., 2019).

In this work we leverage machine learning to optimize feature selection. Machine learning has been used in a Popper-Bayes framework before to relate geophysical data to a specific hypothesis, that is, $p(h_i|\mathbf{z})$ (Enemark et al., 2020), and to directly predict reservoir properties from seismic data (Pradhan & Mukerji, 2019). We follow the work of Pradhan and Mukerji (2019) and use machine learning to update model parameters, that is, $p(\boldsymbol{\theta}|\mathbf{z})$ directly from geophysical data, such that $\boldsymbol{\theta}^* = f^{-1}(\mathbf{z})$ where f^{-1} is the machine learning inverse model. Then, the summary statistic used for dimension reduction is simply $g(\mathbf{z}) = f^{-1}(\mathbf{z})$. Any machine learning model that maps $\mathbf{z} \mapsto \boldsymbol{\theta}^*$ will work in this framework, and the approximate posterior distribution of model parameters becomes $p(\boldsymbol{\theta}|f^{-1}(\mathbf{z}))$.

2.4. Posterior Updating

The posterior updating step consists of two parts. In this work, we update the model parameters first, then the hypotheses. The model parameter updating is based on the machine learning inverse model (Section 2.3.2). The hypothesis updating step then requires the specification of $p(\boldsymbol{\theta}|h_i)$. Each step is described in more detail below.

2.4.1. Model Parameter Updating

In the posterior model parameter updating step, we use Equation 2. Using the trained machine learning inverse model, f^{-1}, Equation 2 can be rewritten in the following form (see Appendix B for derivation):

$$p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}}) = \int p(\boldsymbol{\theta}|f^{-1}(\mathbf{z}))\, p(\mathbf{z}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\mathbf{z}. \tag{4}$$

If $\mathbf{z}_{\text{obs}}$ is error free, Equation 4 reduces to $p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}}) = p(\boldsymbol{\theta}|f^{-1}(\mathbf{z}_{\text{obs}}))$, which is the same form used by Hermans et al. (2015), Lopez-Alvis et al. (2019), and Pradhan and Mukerji (2020) to estimate posterior distributions. In these studies, with the exception of Pradhan and Mukerji (2020), the posterior density is estimated

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

through kernel density estimation in the reduced-dimension space (Scheidt et al., 2018). We extend this kernel density estimation method for cases where $\mathbf{z}_{\text{obs}}$ has a defined error distribution by calculating the conditional distribution $p(\boldsymbol{\theta}|f^{-1}(\mathbf{z}))$ for multiple realizations of $\mathbf{z}$ drawn from the measurement distribution $p(\mathbf{z}|\mathbf{z}_{\text{obs}})$. The following sequential sampling algorithm (Hansen et al., 2006) may be used to sample the joint posterior from Equation 4:

1. Estimate $p(\boldsymbol{\theta}, f^{-1}(\mathbf{z}))$ empirically using the prior realizations and the trained machine learning inverse model.
2. Sample one realization of $\mathbf{z}$ from $p(\mathbf{z}|\mathbf{z}_{\text{obs}})$.
3. Estimate and draw one sample from $p(\boldsymbol{\theta}|f^{-1}(\mathbf{z}))$, using the joint distribution from Step 1, and the sample of $\mathbf{z}$ from Step 2.
4. Repeat Steps 2–3 Q times to generate Q samples from the posterior.

If $\boldsymbol{\theta}$ is too high-dimensional ($K \gtrapprox 5$), then the joint distribution $p(\boldsymbol{\theta}, f^{-1}(\mathbf{z}))$ becomes difficult to estimate due to the curse of dimensionality (Scheidt et al., 2018). In such cases, we can simplify the problem by denoting $f^{-1}(\mathbf{z}) = [f_1^{-1}(\mathbf{z}),\ldots,f_K^{-1}(\mathbf{z})]$, and applying the assumption that θ_i, $\theta_{\neg i}$, and $f_{\neg i}^{-1}(\mathbf{z})$, where $\neg i$ denotes all indices not equal to , are conditionally independent given $f_i^{-1}(\mathbf{z})$. Under this assumption the posterior distribution for each parameter can be sampled independently using the following variation of Equation 4 (see Appendix C for derivation):

$$p(\theta_i|\mathbf{z}_{\text{obs}}) = \int p(\theta_i|f_i^{-1}(\mathbf{z}))\, p(\mathbf{z}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\mathbf{z}. \tag{5}$$

Each parameter in $\boldsymbol{\theta}$ can then be sampled independently using the same algorithm described above, using $\boldsymbol{\theta} = \theta_i$, and $f^{-1}(\mathbf{z}) = f_i^{-1}(\mathbf{z})$. The independent sampling algorithm (Equation 5) is much faster than the joint sampling algorithm (Equation 4) because it is faster to sample from K one-dimensional conditional distributions, $p(\theta_i|f_i^{-1})$, than it is to sample from one K-dimensional joint conditional distribution, $p(\boldsymbol{\theta}|f^{-1}(\mathbf{z}))$. However, this speed-up comes at a cost of losing information due to the conditional independence assumption. The loss of information will be greater if there are strong correlations between model parameters in the prior model or in the machine learning inverse model predictions. Posterior samples will also be uncorrelated. In this work we use the independent sampling algorithm (Equation 5).

2.4.2. Testing for Bias in the Model Parameter Posterior Distribution

We define an unbiased uncertainty quantification scheme as one that produces posterior probability distributions accurately reflecting the probability densities of the true model parameters. For example, if the true model parameter value is always equal to the mean of its posterior distribution, then the uncertainty quantification scheme is biased. We use the posterior falsification test method (Pradhan & Mukerji, 2020; Yin et al., 2020) to test for bias using an additional set of prior realizations. For each realization i, we calculate the quantile of the known model parameter, $\boldsymbol{\theta}^{(i)}$, under the posterior distribution of the model parameter conditioned on the forward-modeled geophysical data for the sample, $p(\boldsymbol{\theta}^{(i)}|\mathbf{z}^{(i)})$. The quantile of $\boldsymbol{\theta}^{(i)}$, denoted by α_i, can be calculated using the cumulative density function (CDF) of $p(\boldsymbol{\theta}|z_i)$, such that

$$\alpha_i = F_{\Theta|Z_i}(\boldsymbol{\theta}^{(i)}) = \int_{-\infty}^{\boldsymbol{\theta}^{(i)}} p(\boldsymbol{\theta}|\mathbf{z}^{(i)})\, \mathrm{d}\boldsymbol{\theta} \tag{6}$$

where $F_{\Theta|Z_i}$ is the empirical CDF of $p(\boldsymbol{\theta}|z_i)$. For any randomly sampled distribution the quantiles are uniformly distributed. Therefore, if, for a separate validation set, α is uniformly distributed then we can conclude that the uncertainty quantification scheme is unbiased. Because of the large number of posterior distributions that need to be calculated, the independent posterior sampling scheme (Equation 5) is used to improve computational efficiency.

2.4.3. Hypothesis Updating

To update the hypotheses, we use the marginal posterior distributions of model parameters (Section 5.3.2) and Equation 1. The key component of Equation 1 is defining $p(h_i|\boldsymbol{\theta})$. This formulation is designed for problems where all hypotheses are related to the same set of prior model parameters, such that $p(h_i|\boldsymbol{\theta})$ can be defined from geological knowledge.

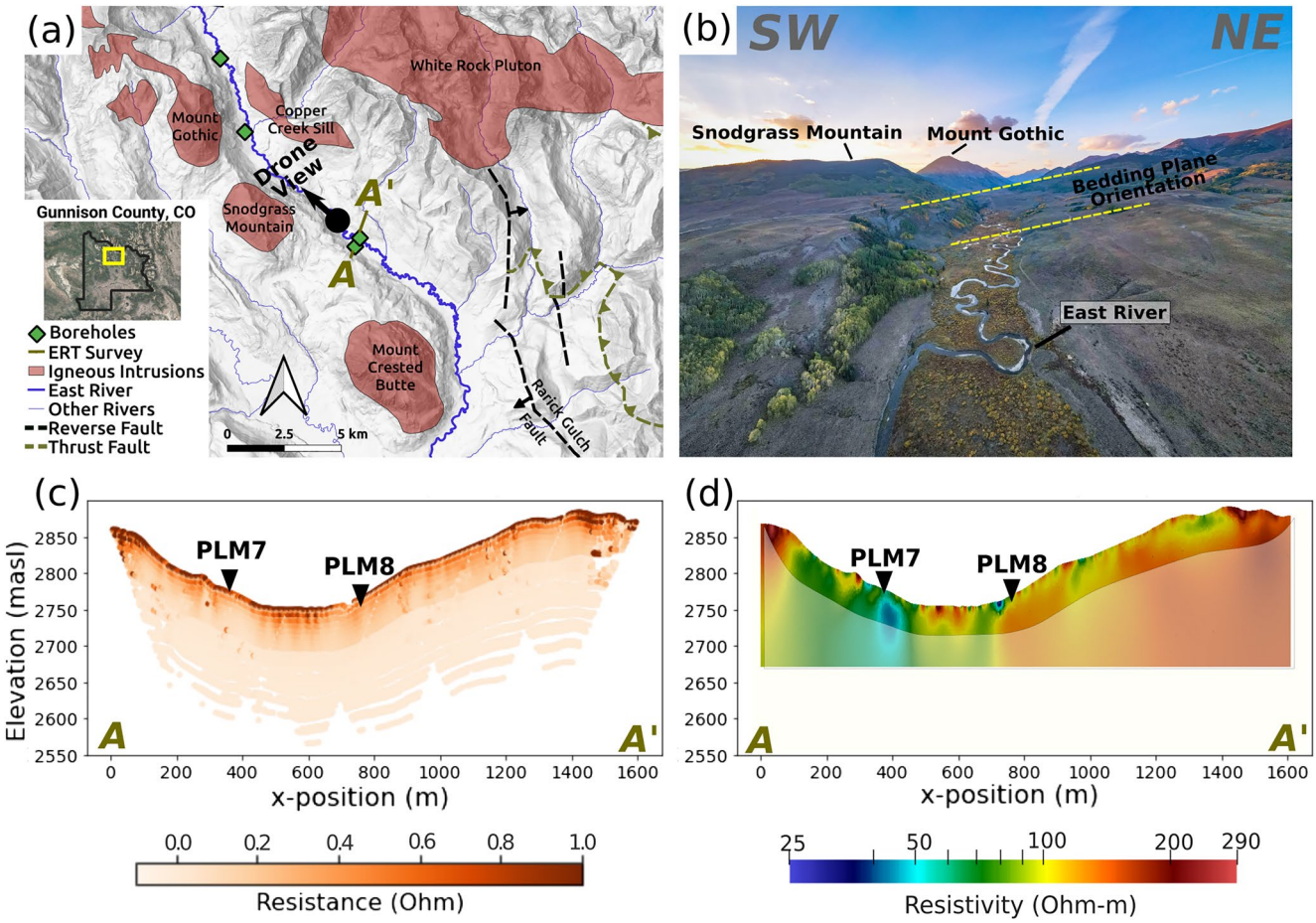


Figure 2. East River Watershed study area and electrical resistivity tomography (ERT) data. (a) a hillshade map showing the East River, igneous intrusions, data sources, and mapped faults. Arrows show the down-thrown side of reverse faults. Sawteeth point to the upthrown side of thrust faults. Cross section A-A' shows the location of ERT data acquisition. Black dot with arrow labeled "Drone View" shows the location of the drone photo in (b). (c) Mean ERT resistance measurements, projected to pseudo-depth and approximate x-position. (d) Inverted resistivity cross section. Shaded region is below the depth of investigation, approximately 75m on average, estimated using the method of Oldenburg and Li (1999).

3. Study Area

The study area, shown in Figure 2a within the mountainous East River Watershed near Crested Butte, Colorado. The East River Watershed has been used as a community watershed for multi-disciplinary research, from ecological to hydrological and atmospheric studies under the Department of Energy-funded Watershed Function Science Focus Area (Hubbard et al., 2018). Multiple hydrological models have been developed in this area (Carroll et al., 2019; Foster & Maxwell, 2019). Geophysical interpretation in this area can improve hydrologic modeling through improved bedrock characterization. In this section we describe the geologic background and the data used in this research for bedrock characterization.

3.1. Geologic Background

The East River Watershed lies within a thick sequence of Cretaceous marine sedimentary rocks, with scattered Oligocene and Miocene intrusive rocks (Streufert et al., 1999). The Oligocene and Miocene intrusives, shown in red in Figure 2a make up the prominent peaks of Mount Crested Butte, Snodgrass Mountain, Gothic Mountain, and others. The data used in this study were collected within the Mancos Shale, a siliceous unit deposited in near-shore marshes and swamps in the Western Interior Seaway (Streufert et al., 1999). During the Laramide Orogeny, with the uplift and closure of the Western Interior Seaway, vast amounts of quartz monzonite and granodiorite were emplaced in sills that make up modern-day Mount

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Crested Butte, Snodgrass Mountain, Mount Gothic, and the Copper Creek Sill (Streufert et al., 1999). The Laramide Orogeny is associated with a compressive stress environment, hence the presence of sills (Stephens et al., 2018) and thrust faults such as the Elk Range Thrust Fault (Tully, 2009), which is in immediate vicinity to the study area.

Following the Laramide Orogeny was the Quaternary glaciations in North America. While the Wisconsin ice sheet never reached southwestern Colorado, significant alpine glaciation occurred (Brugger, 2010). The remnants of the glaciation include the U-shaped valley of the East River and glacial moraines (Brugger, 2010). Post-glacial stress release has caused local tectonic features such as sackungen (Radbruch-Hall et al., 1977; Soldati, 2013) surrounding Mount Crested Butte (Gaskill, 1991). In a more regional context, most of modern-day southwestern Colorado is experiencing an extensional stress environment (Bott & Wong, 1995; Levandowski et al., 2018) in response to complete subduction of the Farallon Plate underneath the west coast of North America (Liu et al., 2008).

3.2. Geophysical Data

An ERT survey with 321 electrodes, spaced approximately 5 m apart over a span of 1,600m, was conducted perpendicular to and across the East River. The locations of the survey is shown in Figure 2a. Data were acquired using the dipole-dipole array, employing dipole lengths (a) of 5, 10, 15, 20, 25, 30, 35, and 40 m, and dipole spacing (n) of 1a to 8a. Data were acquired using a roll-along approach with 112 electrodes, and with a full set of reciprocal measurements (i.e., swapping injection and measurement dipoles) to assess the data error (Tso et al., 2017). The mean of the forward and reciprocal electrical resistance measurement was calculated, and mean resistances with reciprocal errors above 10% were removed from the data set prior to further processing, resulting in 15,371 data points. The measured data are projected to a pseudo-depth of half the lateral distance between the outer-most electrodes, and the lateral distance is estimated by the midpoint of the outer-most electrodes (Figure 2c).

Although the ERT data provides various information of subsurface structure and properties, we focus on a single anomaly in this study. The anomaly is a low-resistance zone that appears around $x = 400$m in the measured ERT data (Figures 2c and 2d). While the projected resistance does not accurately capture the subsurface morphology of the low-resistance zone, the presence of the anomaly indicates the presence of a conductive feature near the surface. This has been confirmed by standard least squares inversion of the data (Figure 2d), which indicates a low-resistive anomaly at this location reaching from the shallow to the deeper subsurface. It is also in agreement with preliminary results of an airborne electromagnetic survey (Figure S1), which indicates a low-resistivity feature that is, perpendicular to this ERT cross-section, extending parallel to the valley axis (Uhlemann et al., 2021). However, there is no surficial evidence of this low-resistivity feature that could help to explain what could be causing it.

3.3. Well Data

There are four boreholes within the East River Watershed that have been drilled and logged. Borehole data include geologic cores, electromagnetic (EM) induction logs, nuclear magnetic resonance (NMR) logs (Figure 3), and more. The ERT survey crosses two of these boreholes, referred to as PLM7 and PLM8. PLM7 was drilled through the low-resistance zone observed in the measured ERT data, and shows a 6m thick surficial layer of quaternary fluvial sediments overlaying the Mancos Shale Bedrock. PLM8 shows a 3 m thick surficial layer of quaternary sediments. EM induction logs show that the resistivity in PLM7 decreases significantly below about 15 m depth, whereas this decrease in resistivity is not observed in PLM8. PLM7, on average, also shows a higher NMR derived porosity than PLM8, though the high-porosity zones tend to occur in discrete packages rather than in large zones. Visual inspection of the cores shows that PLM7 contains more brecciated shale and exhibits more alteration in the uppermost 15 m than PLM8. In contrast, the cores from PLM8 tend to be more competent with more regularly spaced fractures than those from PLM7.

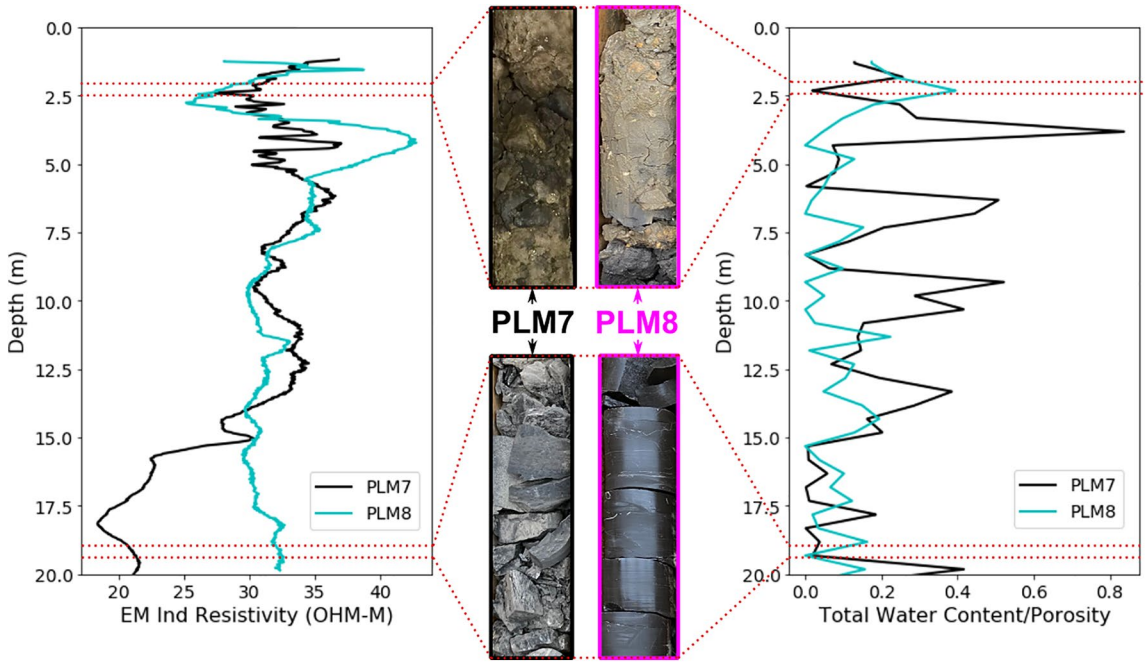


Figure 3. Well data from PLM7 and PLM8, two boreholes drilled along the electrical resistivity tomography (ERT) data acquisition line, with core photographs from a section within the surficial quaternary layer, and within the Mancos Shale bedrock (Left) Inductive electromagnetic (EM) resistivity logs and (right) nuclear magnetic resonance water content logs (Center) core photographs from 4 and 19 m. Nuclear magnetic resonance porosity measurements assume fully saturated rock.

4. Hypotheses for the Origin of a Fractured Mancos Shale Zone

The fractured bedrock zone, observed in the wells and geophysical data as a high-porosity, low-resistivity zone within the Mancos Shale, may have been formed by a variety of processes. Each process has different implications for the morphology and hydraulic properties of the fracture zone, and the watershed as a whole. In this section we describe six hypotheses that may explain the origin of the Mancos Shale fractured bedrock zone, based on the geological conditions that led to the present-day bedrock configuration of the East River Watershed, from the deposition of the Mancos Shale to the end of Quaternary glaciation.

4.1. Sediment Deposition

One possible cause of bedrock fractures is variations in sediment characteristics. The Mancos Shale was deposited in an offshore environment in the Western Interior Seaway (Da Gama et al., 2014). It is generally thought that the Mancos Shale is an offshore continuation of coarser-grained fluvial and deltaic environments (Birgenheier et al., 2017). If, during the depositional process, an event occurred that deposited a layer of sediment that was more prone to brittle fracturing than the surrounding material, then a fracture zone may develop within shale bedding planes. In this area the Mancos Shale dips approximately 10–30° to the southwest (Gaskill, 1991). Therefore, for this hypothesis, the fracture zone is dipping 10–30° to the southwest, as illustrated by hypothesis (i) in Figure 4. Fluids entering the fractured zone at outcrops may quickly spread laterally throughout much of the watershed in this scenario.

4.2. Thrust Faulting

Faulting is another common source of fractures. When stress builds up along a fault, or it slips, the bedrock can become deformed in a zone surrounding the fault and may be more permeable than the surrounding material. During the Laramide Orogeny regional tectonic stresses compressed the rock in the East River Watershed, creating thrust and reverse faults such as the Elk Range Thrust Fault (Tully, 2009). If the fracture zone considered in this study were an unmapped thrust fault, then it most likely dips toward the southwest

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

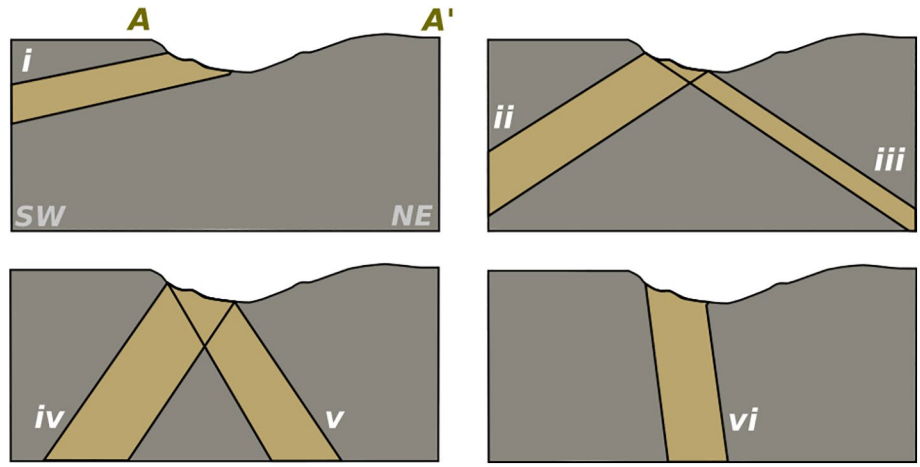


Figure 4. Conceptual illustrations of the difference in fracture zone dip angle for the six different geologic hypotheses. Not to scale. The labels A–A' correspond to the electrical resistivity tomography (ERT) survey shown in Figure 2.

or northeast at 20–40° based on Anderson's theory of faulting (Anderson, 1905). Of course, thrust faults are very complicated and may occur at any angle. However, for the purposes of this study we assume a fracture zone dipping between 20° and 40° is most likely a thrust fault dating back to the Laramide Orogeny. Because thrust faults in this region could theoretically dip in two directions, northeast and southwest, we consider each dip direction a separate hypothesis. The two scenarios consistent with a thrust fault are illustrated by hypotheses (ii-iii) in Figure 4. Fluids entering the fractured zone at an outcrop may be carried laterally and to depth quickly. Furthermore, the existence of a fault implies a previously unmapped offset of the bedrock and intrusive bodies in other parts of the watershed.

4.3. Normal or High-Angle Reverse Fault

Since the end of the Laramide Orogeny southwest Colorado has been experiencing extensional stresses. These extensional stresses have activated normal faults within the vicinity of the East River Watershed (Bott & Wong, 1995). Normal faults, based on Anderson's theory of faulting, generally dip steeper than reverse faults, around 50°–70°. However, there are higher-angle reverse faults in the region, such as the Rarick Gulch Fault just to the southeast of the East River valley (Tully, 2009). The Rarick Gulch Fault is believed to originate during the Laramide Orogeny, but may have been reactivated by the current extensional stress regime. Like with thrust faults, normal and reverse faults could be dipping toward either the northeast or southwest. Therefore, we consider each dip direction a separate hypothesis. The two scenarios consistent with a normal, or high-angle reverse fault, are illustrated by hypotheses (iv–v) in Figure 4. A steep-dipping fracture zone would carry fluids deeper into the subsurface than shallower-dipping fracture zones, and also implies the existence of an active fault and bedrock offsets in other parts of the watershed.

4.4. Sackung

One final hypothesis about the formation of the East River fracture zone is that stress release from glacial retreat created local stress perturbation that resulted in mountain sagging, or sackung (Radbruch-Hall et al., 1977; Soldati, 2013). Sackung features have been observed in many parts of Colorado (Radbruch-Hall et al., 1976). Tension fissures from sackungen are observed in aerial photographs of Mount Crested Butte (Gaskill, 1991), and tend to be sub-vertical fractures that wrap around the high mountain peaks (Savage & Varnes, 1987) and along ridgelines (Gutiérrez-Santolalla et al., 2005). Sackung fissures contrast with regional faults in the fact that normal faults strike in one consistent direction, whereas sackung features are controlled by local topography. While most sackung trenches are less than 50 m wide, the zone of bedrock deformation may extend laterally far beyond the fissure itself, leading to the appearance of a wide fracture zone. Fluids entering a sackung fissure may be routed downwards, but the fissure may terminate or connect with other tension fissures at depth. Because sackungen are difficult to study at depth, not much is known

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

about how deep a sackung fissure may extend, although it has been suggested that they can extend up to 200 m (Tabor, 1971). Hypothesis (vi) in Figure 4 illustrates a potential sackung fissure as it would appear in this study area.

4.5. Prior Probabilities for Each Hypothesis

The major difference between each of the six hypotheses for the formation of the fractured bedrock zone is the expected dip angle (Figure 4). In order from shallowest-dipping to steepest-dipping, the fracture zone could be caused by a brittle sedimentary layer (a), a thrust fault from the Laramide Orogeny dipping to the southwest (b) or northeast (c), a normal fault or reactivated reverse fault dipping to the southwest (d) or northeast (e), or, finally, post-glacial mountain sagging (f).

Three of the six hypotheses are assigned a prior probability of $1/6 = 0.166$, with the exceptions being the post-glacial mountain sagging and the high-angle normal or reverse fault hypotheses. Sackung fissures typically follow the topographic contours of high-relief geographic features (Radbruch-Hall et al., 1977), but airborne electromagnetic data from the East River Watershed show that the fracture zone extends along the East River Valley axis independent of the topography around Mount Crested Butte (Uhlemann et al., 2021). Therefore, we assign the sackung hypothesis a prior probability of 0.1. Other studies of faulting near the East River valley show the existence of a high-angle reverse fault that terminates at the southeast edge of the valley (Streufert et al., 1999), called the Rarick Gulch Fault (Tully, 2009). Because of this prior information, we assign both high-angle normal and reverse fault hypotheses prior probabilities of 0.2.

5. Testing Geoscientific Hypotheses With East River ERT Data

Because the six hypotheses for the formation of the fracture zone are differentiated by their dip angle, we seek to estimate the uncertainty on the dip angle of the fracture zone from the ERT measurements. In this section we apply the Popper-Bayes methods outlined in Section 2 to test the six hypotheses outlined in the previous section.

First, we present a five-parameter prior model that relates the geologic hypotheses to geophysical properties. Then, we attempt to falsify the prior model with the measured geophysical data. Finally, we show that a random forest regression model trained on the prior can be used to predict model parameters from ERT data and quantify the associated uncertainty.

5.1. Prior Model & Monte Carlo

The prior subsurface model we use assumes that the subsurface consists of a homogeneous matrix with an embedded fracture zone. We assume the fracture zone starts 5 m below the surface to account for the surficial quaternary deposits observed in the well data (Section 3.3, Figure 3). The fracture zone is parameterized by its dip angle ϕ, thickness H, and projected outcrop position of the fracture zone midpoint x_o. The background material is assigned an electrical resistivity ρ_b and the fracture zone is assigned an electrical resistivity ρ_f. Therefore, $\boldsymbol{\theta} = [\phi, H, x_o, \rho_f, \rho_b]^T$. A schematic of these parameters is shown in Figure 5. However, note that the only parameter relevant for hypothesis testing is ϕ.

Prior distributions for all five parameters are shown in Table 1, and were estimated via a combination of well logs and the inverted ERT data. The data are simulated using the open-source geophysical software PyGIMLi (Rücker et al., 2017) and the same topography, electrode configuration, and input current as the measured data. The triangular mesh for each realization is generated using tetgen (Si, 2015) with a minimum mesh angle of 22°. The mesh is refined around the electrodes and conforms exactly with the fault zone. A total of 2,000 prior realizations are generated.

5.2. Prior Falsification Test

To determine whether or not the prior is representative of the real ERT data, we test the prior model against the measured data using the PCA-based prior falsification test described in Section 2.3.1. Figure 6 shows the PCA results. Over 58% of the total variance is explained by the first principal component, 4% is explained by

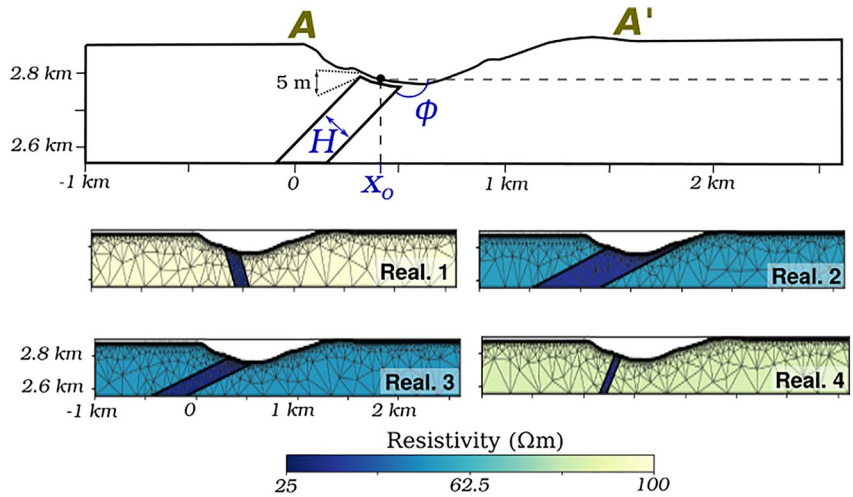


Figure 5. Parameterization of the prior model and four randomly selected prior realizations of the subsurface electrical resistivity models. Triangular meshes are shown along with the resistivity models.

the second component, and the remaining components explain far less of the variance (Figure 6a). About 1,000 principal components are needed to explain 90% of the variance. The measured principal component scores fall well within the range of prior scores up to the 37^{th} principal component (the first 10 are shown in Figure 6b), which accounts for 70% of the total variance. A total of 14 out of 2,000 measured data scores fall outside the range of the prior scores (Figure 6c). The first principal component eigenvector (Figure 6d) is highly correlated with the background resistivity ($r^2 = 0.99$). The dip angle is most correlated with PC 7 (Figure 6e, $r^2 = 0.25$) and PC 8 (Figure 6f, $r^2 = 0.18$). Because the eigenvectors are correlated with the model parameters, and the measured scores fall within the range of the first 37 principal component scores (70% of the variance), we conclude that the prior model cannot be falsified by the measured ERT data.

5.3. Model Parameter Posterior Updating

The posterior updating consists of two steps. First, we must create inverse models, f^{-1}, to relate the measured ERT data to each of the five prior model parameters. In this work we use machine learning models to estimate f^{-1}. Second, we quantify uncertainty in the measured geophysical data and machine learning regression to sample posterior distributions of each prior model parameter.

5.3.1. Relating ERT Data to Fracture Zone Morphology With Machine Learning

The results of the prior model testing show that the prior model adequately reproduces the important features of the measured ERT data. Therefore, to estimate the posterior distribution of model parameters using Equations 1 and 5 we must create a function, f^{-1} that maps $\mathbf{z} \mapsto \hat{\boldsymbol{\theta}}$.

In this work we define f^{-1} as a random forest regression model. The random forest regression is powerful because it is nonlinear, easy to train, and calculates variable importances that can be used to understand how the regressor works. We train the random forest to predict model parameters with a training set of 1,000 prior realizations of model parameters and corresponding forward-modeled ERT data. The remaining 1,000 realizations are used as a test set. A total of five random forests are trained - one for each of the five model parameters. Before training, independent Gaussian noise is added to each realization using the measured standard deviation of ERT measurements, σ_{obs}. Then, each element of $\mathbf{z}$ is normalized with respect to their mean and standard deviation, calculated from

Table 1
Prior Distributions and Sampling Tolerances (Section 2.4) for the Five Model Parameters. $\mathcal{U}$(upper, lower) is a Uniform Distribution

Parameter	Unit	Symbol	Distribution	Sampling Tol., ε
Dip Angle	degree	ϕ	$\mathcal{U}(25,165)$	2.5
Thickness	m	H	$\mathcal{N}(120,40)$	5
Outcrop X-Position	m	x_o	$\mathcal{N}(400,40)$	5
Fracture Zone Resistivity	Ω ·m	ρ_f	$\mathcal{LN}(0.45,20,10)$	5
Background Resistivity	Ω ·m	ρ_b	$\mathcal{LN}(0.8,40,30)$	10

Note. $\mathcal{N}(\mu,\sigma)$ is a normal distribution. $\mathcal{LN}(s,\text{loc},\text{scale})$ is a lognormal distribution as defined in Appendix D.

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

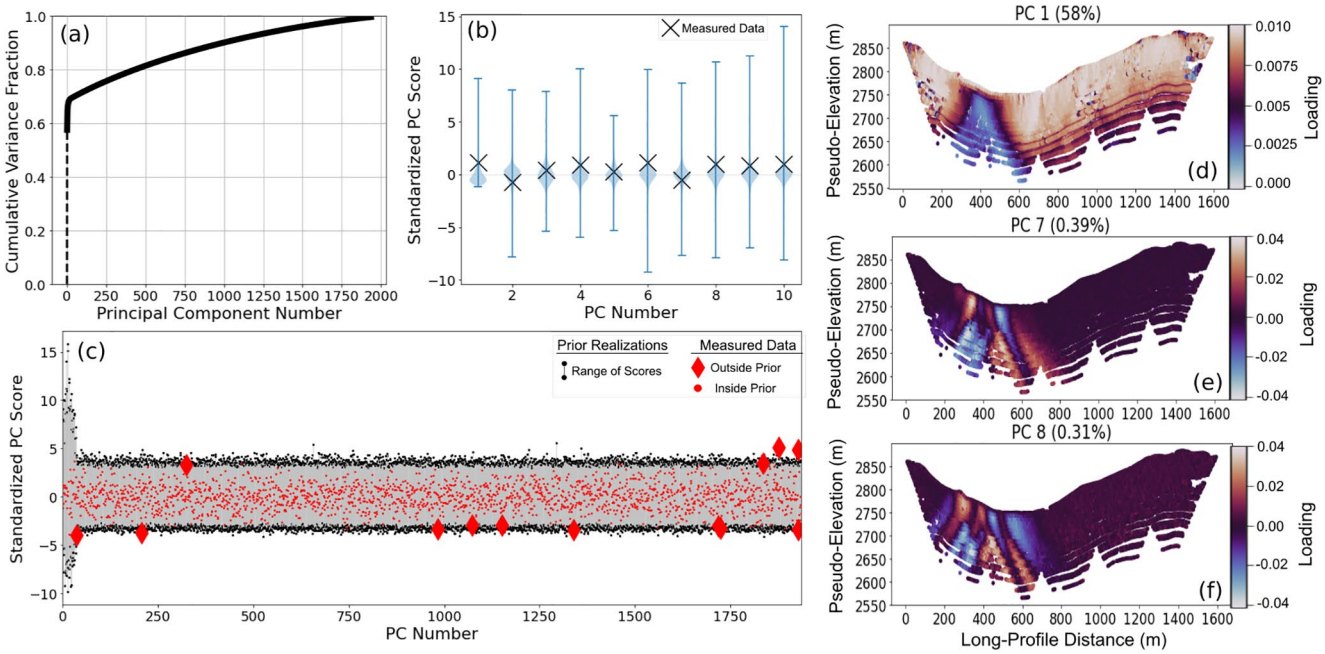


Figure 6. (a) Scree plot showing the cumulative variance explained by each principal component. (b) Violin plots showing the prior distributions of the first 10 principal component (PC) scores and how the measured data compare. (c) Range of all 2,000 prior PC scores, with measured data scores that fall inside and outside the prior (d–f) Eigenvectors for PC 1, PC 7, and PC 8, projected onto the same space as the measured electrical resistivity tomography (ERT) data in Figure 2c.

the training set only. Variable importances, which show which ERT measurements are most informative for each model parameter, are shown in Figure S2. Figure 7 shows the predicted and true values for each of the five prior parameters. Heteroskedasticity is present in the predictions, meaning that $p(\boldsymbol{\theta}|f^{-1}(\mathbf{z})$ cannot be approximated using Gaussian assumptions.

Uncertainty in the measured ERT data is accounted for by sampling 10,000 realizations of measured ERT data from $p(\mathbf{z}|\mathbf{z}_{\text{obs}})$, using an uncorrelated multivariate Gaussian error model with means equal to the reciprocal mean of the ERT measurements, and standard deviations equal to the reciprocal error of the ERT measurements (Uhlemann et al., 2018). Each of the 10,000 realizations from $p(\mathbf{z}|\mathbf{z}_{\text{obs}})$ are normalized with respect to the mean vector and standard deviations from the prior training set. The trained random forests are then used to predict each of the five parameters for each sample of measured ERT data. The mean and 95% confidence intervals for the measured ERT predictions are shown as vertical lines in Figure 7.

5.3.2. Sampling the Posterior Distribution of Model Parameters

We sample the posterior distribution of each model parameter using the independent sampling method (Equation 5). The measured data are filtered using Equation 3 and then back-transformed to the measurement space before posterior sampling. The conditional probability $p(\theta_i|f_i^{-1}(\mathbf{z}))$ (Step 3 in Section 2.4.1) is estimated empirically by searching for prior samples of θ_i which satisfy $|f_i^{-1}(\mathbf{z}| < \varepsilon_i$. The values of ε_i were chosen empirically to be as small as possible without introducing errors from sampling empty conditional sets. One posterior sample is then obtained by randomly selecting one of these conditional samples with uniform probability. The posterior distributions for each model parameter are shown as histograms in Figure 8a. Uncertainty has been reduced for each model parameter, though for some parameters more than others. For example, the posterior distribution of ρ_f has been reduced less than that of x_o or ρ_b. Most importantly, the prior uncertainty for the dip angle has been reduced significantly, showing that the fracture zone is most likely dipping subvertically or to the southeast.

5.3.3. Posterior Estimation With Synthetic Subsurface Heterogeneity

In reality, the subsurface is not homogeneous like the prior model. Although the contrast between the fracture zone and background material is stronger than the background heterogeneity, we investigate the effect

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

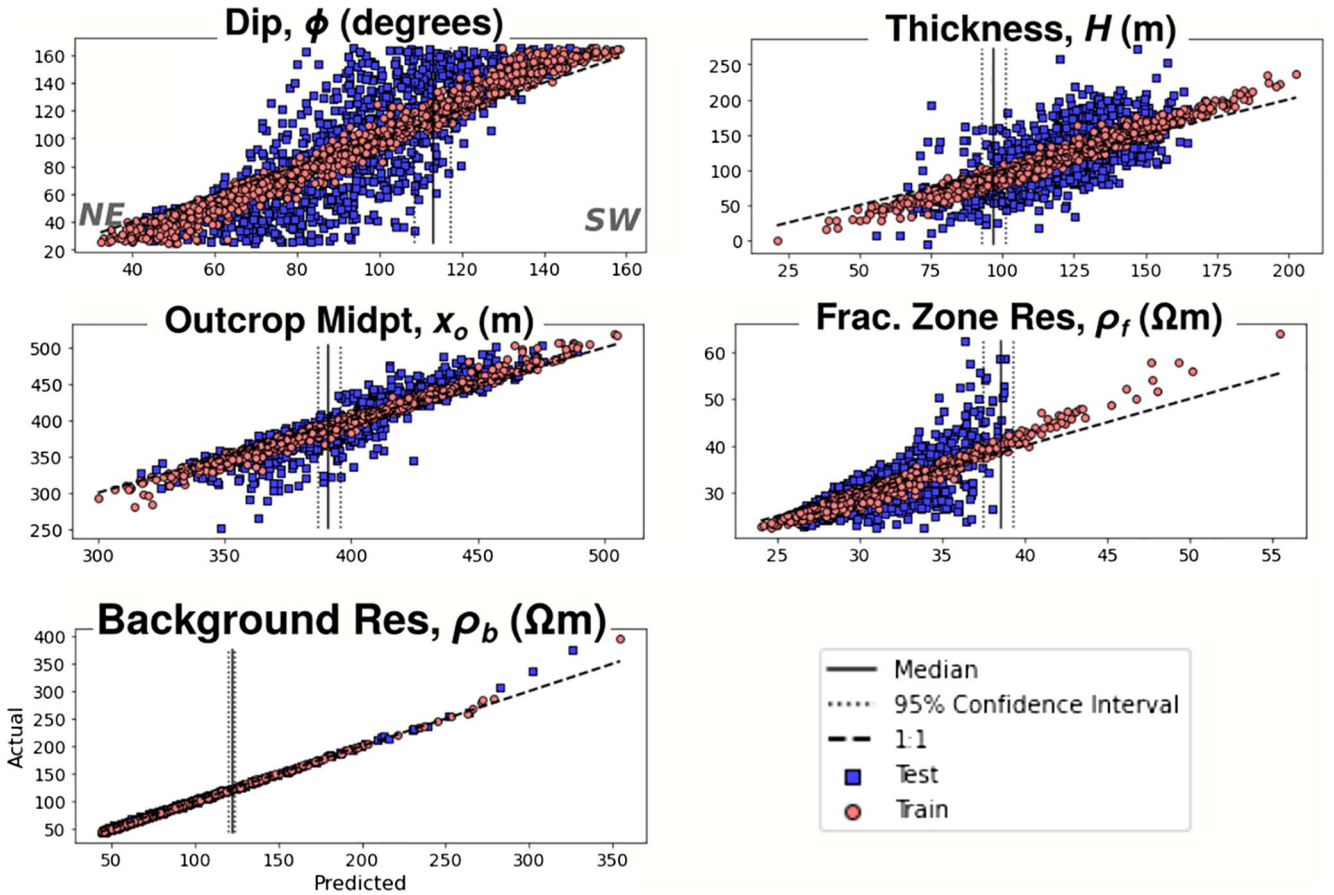


Figure 7. Prediction plots of the five prior parameters by the random forest model trained on the training set of prior realizations. The *x*-axis is the predicted value of the prior parameter from the random forest. The *y*-axis is the true value of the prior parameter used to simulate the electrical resistivity tomography (ERT) data. Pink dots are training set realizations and blue squares are test set realizations. Black vertical lines are random forest predictions from the measured ERT data. Dotted black lines show the 95% confidence interval of predictions based on the observed measurement uncertainty.

of background heterogeneity on posteriors predicted using the homogeneous prior. This is done with two synthetic, heterogeneous subsurface models (Figure 8b), where the heterogeneity is based on the resistivity variability observed in the inverted ERT data (Figure 2d). Both synthetic models have the same model parameters, but different heterogeneity random fields generated using an exponential covariance model. The values for ϕ, H, and x_o used in the synthetic models were taken as the mean values of the respective posterior distributions from the measured data (Figure 8). The mean posterior ρ_f and ρ_b were used as the mean resistivity in the heterogeneous subsurface models, along with variances and correlation lengths estimated through variogram analysis of the inverted ERT data (Figure S3, Text S1).

One synthetic subsurface model contains a highly resistive anomaly next to the fracture zone and near the surface, and the other synthetic subsurface model contains no major anomalies near the fracture zone (Figure 8b). The true model parameters fall within the predicted posterior model parameter distributions for all parameters except ρ_b (Figure 8a).

5.4. Posterior Bias Testing

Posterior bias testing is performed using the procedure outlined in Section 2.4.2 we first generate a validation set of an additional 1,000 prior realizations. These realizations are separate from the 2,000 realizations simulated in Section 5.1. The posterior distributions are estimated using the independent sampling method (Equation 5). The α-quantiles for the set of true model parameters for each realization in the validation set are calculated according to Equation 6. A total of 10,000 samples from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}})$

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

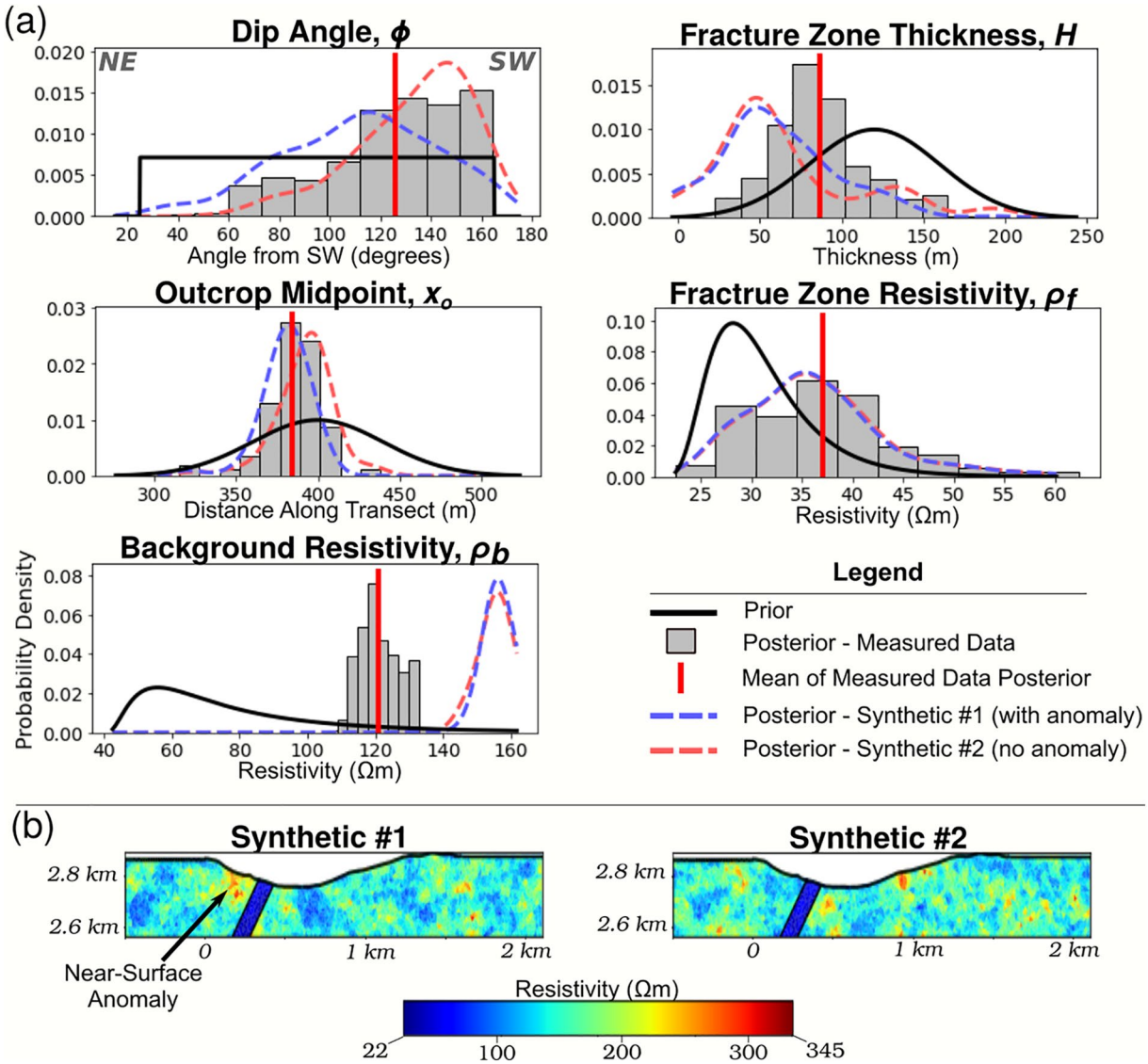


Figure 8. (a) Prior and posterior distributions of the five prior parameters for the measured electrical resistivity tomography (ERT) data and the synthetic ERT data simulated from the two heterogeneous subsurface resistivity models (smoothed using kernel density estimation) shown in (b).

are generated for each posterior distribution. The α distributions (Figure 9) are nearly uniform for ϕ, H, x_0, and ρ_f, implying that there is little posterior bias for these parameters. However, the α distribution for ρ_b is skewed to the center, meaning that the true posterior uncertainty is generally overestimated for these 1,000 synthetic cases.

5.5. Hypothesis Updating

Posterior probabilities for each hypothesis are calculated using Equation 1. The posterior distribution of model parameters, $p(\boldsymbol{\theta} \mid \mathbf{z}_{\text{obs}})$, were calculated in Section 5.3.2 and are shown in Figure 8. The likelihoods of each hypothesis as a function of dip angle, $p(\phi \mid h_i)$, are written in Table 2 and shown visually in Figure 10. Prior probabilities discussed in Section 4.5 are also shown in Table 2. All three of these probability distributions are used together in Equation 1 to update the posterior probabilities of each hypothesis. Overall, the SW-dipping faults and the bedding plane hypotheses have the highest probability, and the NE-dipping faults and the sackung fissure hypotheses have the lowest probabilities (Table 2).

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

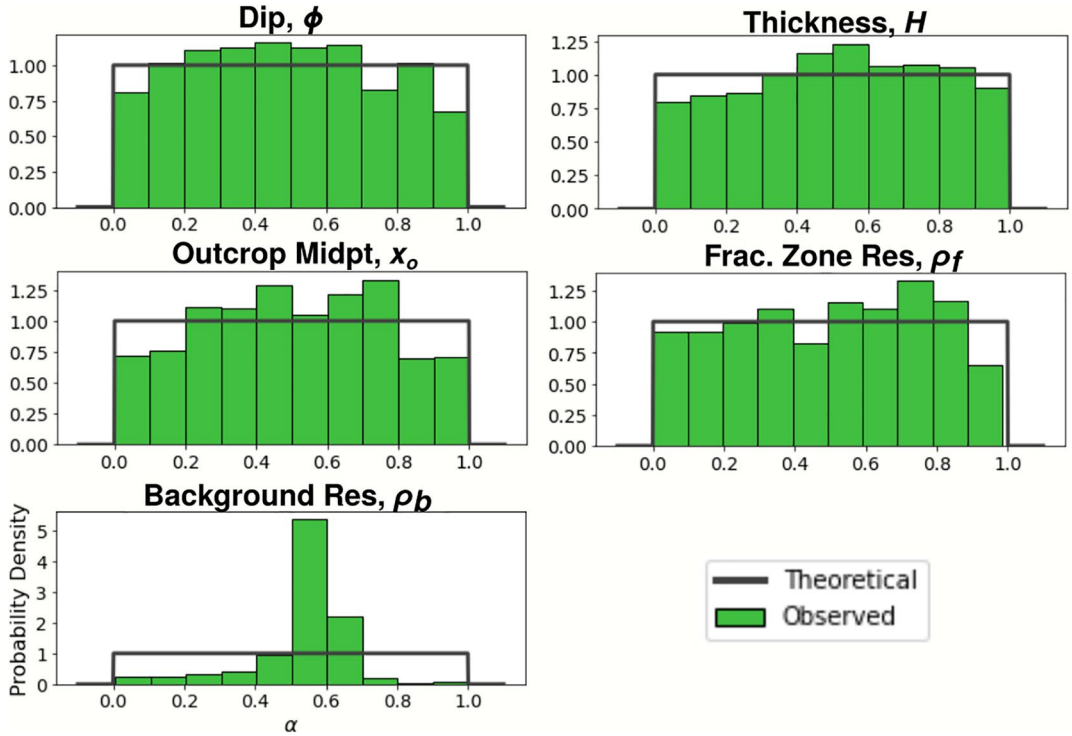


Figure 9. Histograms of posterior quantiles, α_i for each posterior distribution $p(m \mid z_i)$, with the theoretical quantiles for a consistent uncertainty quantification scheme.

6. Discussion

In this section we discuss the prior model and how it compares to the true subsurface. We also discuss the implications of the hypothesis evaluation on our understanding of the East River Watershed geology.

6.1. Assessment of Prior Model

The prior model we use in the paper is a drastic simplification of the true complexity of the subsurface. The shallow layer is assumed to be 5 m thick near the fracture zone based on well observations (Figure 3), though it likely varies along the ERT profile. Also, the shallow layer is assigned the same resistivity as the background material, though it appears to be slightly more conductive than the background material (Figure 2d). However, variability within the shallow layer is not very important for the random forest model predictions of fracture zone parameters (Figure S2).

Table 2
Hypothesis Priors, Likelihood Models, and Posterior Probabilities, Listed in Descending Order of Posterior Probability. Hypothesis Numbers Are Referencing Figure 4

	Hypothesis	Prior	Likelihood	Posterior
ii	Thrust Fault (SW)	0.17	$\mathcal{N}(145,5)$	0.33
iv	Normal/High-Angle Reverse Fault (SW)	0.20	$\mathcal{N}(115,5)$	0.29
i	Bedding Planes	0.17	$\mathcal{N}(160,3)$	0.24
v	Normal/High-Angle Reverse Fault (NE)	0.20	$\mathcal{N}(65,5)$	0.07
vi	Sackung Fissure	0.10	$\mathcal{N}(85,5)$	0.06
iii	Thrust Fault (NE)	0.17	$\mathcal{N}(35,5)$	0.00

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

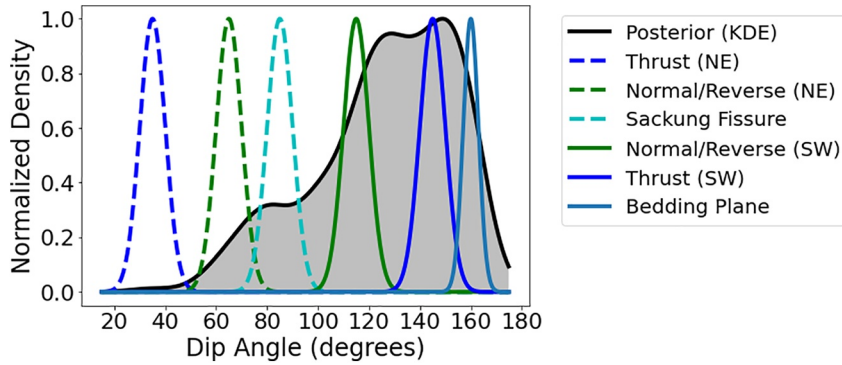


Figure 10. Likelihood functions for each hypothesis in Table 2, plotted along with the posterior distribution of dip angle $p(\phi \mid \mathbf{z}_{\text{obs}})$ approximated by kernel density estimation. All distributions are normalized with respect to their maximum density. Dashed lines show hypotheses with posterior probability below 1/6.

The assumption of homogeneity in the background and fracture zone is another major assumption. However, even with a large anomaly in the heterogeneous synthetic model the model parameter posteriors are consistent with the true model parameters, except for the background resistivity. The relative insensitivity of the model parameter posteriors to heterogeneity is because the contrast of resistivity inside and outside the fracture zone is much more significant than the other forms of heterogeneity present in the measured ERT data (Figures 2 and S1). Furthermore, the fact that less than 2% of measured principal component scores fall outside the prior scores underscores that the contrast between the fracture zone and background material is much greater than other kinds of heterogeneity. Finally, the posterior bias test shows that the posterior distributions for all parameters but the background resistivity are unbiased for models sampled from the prior.

6.2. Evaluating Fracture Zone Origin Hypotheses

The posterior probabilities for the northeast-dipping fault and the sackung fissure hypotheses are significantly lower than the other three hypotheses, which all dip to the southwest (Table 2). The two most probable hypotheses - the southwest-dipping thrust fault and normal/reverse fault - are consistent with the location of the Rarick Gulch fault (Section 4.5), though it was originally assumed to dip toward the northeast instead of the southwest (Tully, 2009). If the fracture zone in this study is related to the Rarick Gulch fault, then our results suggest that the Rarick Gulch fault either changes dip direction in the East River valley, or its dip direction is actually toward the southwest along the entire fault, which would change the interpretation of the geologic structure of the East River Watershed area.

Despite being the third-most probable, the bedding plane hypothesis is also of particular interest because the Mancos Shale is known to have strong internal heterogeneity (Streufert et al., 1999). While outcrop measurements show the Mancos Shale near the ERT data have a dip angle between 10 and 30° SW (Gaskill, 1991), there is significant bedrock deformation in the surrounding region in general. In fact, just one mile away from the study area the Mancos Shale outcrops with vertical beds (Gaskill, 1991). Because of this deformation, it is possible that, despite the dip angle of the outcropped shale, below the surface the dip angle may steepen due to folding. Furthermore, the dip angle is limited in this study to $\phi \leq 165$, or steeper than 15° to the southwest. However, the posterior distribution of dip angle in Figures 8 and 10 demonstrate that a prior including $\phi > 165$ degrees may be consistent with the observed ERT data. However, if the bedding plane hypothesis is correct and $\phi > 165$ degrees, we expect, based on the topographic gradient in the region, that the sedimentary layer would outcrop on the opposite side of the valley as well. There is some evidence for this in the ERT data (Figure 2d) and preliminary airborne electromagnetic data (Figure S1). More investigation into these features is necessary to further evaluate this hypothesis.

Finally, knowing that the fracture zone does not dip to the northeast allows us to optimize future ERT data acquisition. Survey optimization works by identifying electrode configurations that provide the highest

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

resolution of a subsurface target (Uhlemann et al., 2018). Since the fracture zone dips vertically, or to the southwest, we can optimize new ERT surveys to collect higher-resolution data at these locations.

7. Conclusions

In this paper, we have developed a method for testing geoscientific hypotheses using the Popper-Bayes framework for problems where the hypotheses share a common subsurface parameterization, but depend on the values of the parameters. The Popper-Bayes framework accounts for uncertainty in the inverse relationship between geophysical data and the model parameters, as well as error in the geophysical measurements. Monte Carlo methods and machine learning algorithms are key tools for estimating the inverse relationship between model and data, as well as sampling the posterior parameter distributions. We have applied this framework to an ERT data set collected over a zone of fractured bedrock in the Elk Mountains of central Colorado. Our results show that the three hypotheses with northeast-dipping structures have much lower posterior probabilities than the three hypotheses with southwest-dipping structures. The predictions, obtained through a five-parameter homogeneous prior model, are relatively insensitive to internal heterogeneity. These results have improved our understanding of the East River Watershed geology, and future application of these Popper-Bayes methods will help scientists make meaningful interpretations of geophysical data in the context of specific geoscientific hypotheses.

Appendix A: Derivation of the Hypothesis Updating Scheme

The goal of the uncertainty quantification in this study is to calculate, for each hypothesis, the posterior probability $p(h_i|\mathbf{z}_{\text{obs}})$. The hypotheses depend on some latent variable, $\boldsymbol{\theta}$, which in this study is the subsurface model. By marginalizing with respect to $\boldsymbol{\theta}$, we write the posterior as

$$p(h_i|\mathbf{z}_{\text{obs}}) = \int p(h_i|\boldsymbol{\theta},\mathbf{z}_{\text{obs}})\, p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\boldsymbol{\theta}. \tag{A1}$$

Next we apply the conditional independence assumption $p(h_i|\boldsymbol{\theta},\mathbf{z}_{\text{obs}}) = p(h_i|\boldsymbol{\theta})$, implying that if we know the subsurface model parameter $\boldsymbol{\theta}$ then the geophysical data add no additional information. With this assumption, and using Bayes' Rule to rewrite $p(h_i|\boldsymbol{\theta})$ the expression in Equation A1 becomes

$$p(h_i|\mathbf{z}_{\text{obs}}) = \int \frac{p(\boldsymbol{\theta}|h_i)\, p(h_i)}{p(\boldsymbol{\theta})}\, p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\boldsymbol{\theta}. \tag{A2}$$

If we use uniform priors for $p(\boldsymbol{\theta})$, then that probability is a constant for all $\boldsymbol{\theta}$. Therefore, we can drop the term from Equation A2 as it is just a proportionality constant. Also, as $p(h_i)$ does not depend on $\boldsymbol{\theta}$, it can be moved outside the integral. The result is the following expression used for uncertainty quantification in this study.

$$p(h_i|\mathbf{z}_{\text{obs}}) \propto p(h_i) \int p(\boldsymbol{\theta}|h_i)\, p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\boldsymbol{\theta}. \tag{A3}$$

Appendix B: Derivation of the Model Parameter Updating Scheme (Joint)

The posterior distribution of model parameters can be derived as a special case of the generalized inverse problem formalization of Tarantola and Valette (1982), but in this derivation we choose to start from the conditional probability $p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}})$. We marginalize the conditional probability to introduce a variable, $\mathbf{z}$, that represents any possible set of geophysical measurements.

$$p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}}) = \int p(\boldsymbol{\theta}|\mathbf{z},\mathbf{z}_{\text{obs}})\, p(\mathbf{z}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\mathbf{z} \tag{B1}$$

If we make the conditional independence assumption $p(\boldsymbol{\theta}|\mathbf{z},\mathbf{z}_{\text{obs}}) = p(\boldsymbol{\theta}|\mathbf{z})$, then the above equation can be reduced to the following form, which is essentially the form of Jeffrey's rule of conditionalization (Jeffrey, 1983; Pearl, 1988).

$$p(\boldsymbol{\theta}|\mathbf{z}_{\text{obs}}) = \int p(\boldsymbol{\theta}|\mathbf{z})\, p(\mathbf{z}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\mathbf{z} \tag{B2}$$

If we expand the term $p(\boldsymbol{\theta}|\mathbf{z})$ using some model parameter prediction, $\boldsymbol{\theta}^*$, we get

$$p(\boldsymbol{\theta}|\mathbf{z}) = \int p(\boldsymbol{\theta}|\boldsymbol{\theta}^*,\mathbf{z})\, p(\boldsymbol{\theta}^*|\mathbf{z})\, \mathrm{d}\boldsymbol{\theta}^* \tag{B3}$$

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Because $\boldsymbol{\theta}^* = f^{-1}(\mathbf{z})$, $p(\boldsymbol{\theta}^*|\mathbf{z}) = \delta(\boldsymbol{\theta}^* - f^{-1}(\mathbf{z}))$ where $\delta(\cdot)$ is the Dirac delta function. Therefore, the above equation reduces to $p(\boldsymbol{\theta}|\mathbf{z}) = p(\boldsymbol{\theta}|f^{-1}(\mathbf{z}))$. Substituting this into Equation B2 results in the expression used in Equation 4.

Appendix C: Derivation of the Model Parameter Updating Scheme (Independent)

If only marginal posterior distributions for each parameter are desired when sampling from Equation 4, the marginal posterior distribution for parameter can be written as

$$p(\theta_i|\mathbf{z}_{\text{obs}}) = \int_{\Theta_{\neg i}} \int_{\mathcal{Z}} p(\theta_i, |\boldsymbol{\theta}_{\neg i}, f^{-1}(\mathbf{z}))\, p(\boldsymbol{\theta}_{\neg i}|f^{-1}(\mathbf{z}))\, p(\mathbf{z}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\mathbf{z}\, \mathrm{d}\boldsymbol{\theta}_{\neg i}, \tag{C1}$$

where $\boldsymbol{\theta}_{\neg i}$ is a vector of all model parameters except for θ_i. If we assume that θ_i, $\boldsymbol{\theta}_{\neg i}$, and $f_{\neg i}^{-1}(\mathbf{z})$ are conditionally independent given $f_i^{-1}(\mathbf{z})$, then we can rewrite the above expression as

$$p(\theta_i|\mathbf{z}_{\text{obs}}) = \int_{\Theta_{\neg i}} \int_{\mathcal{Z}} p(\theta_i, |f_i^{-1}(\mathbf{z}))\, p(\boldsymbol{\theta}_{\neg i}|f^{-1}(\mathbf{z}))\, p(\mathbf{z}|\mathbf{z}_{\text{obs}})\, \mathrm{d}\mathbf{z}\, \mathrm{d}\boldsymbol{\theta}_{\neg i}, \tag{C2}$$

which can be simplified to Equation 5 to obtain the expression for independent posterior sampling.

Appendix D: Lognormal Distribution Parameterization

We use the lognormal distribution parameterization provided by the Python package SciPy (Virtanen et al., 2020). The probability density of a lognormal distribution over x, with scale parameter s is

$$f(x,s) = f(y(x),s) = \frac{1}{sy(x)\sqrt{2\pi}} \exp\left(-\frac{\log^2(y(x))}{2s^2}\right); \text{ where } y(x) = \frac{x - \text{loc}}{\text{scale}} \tag{D1}$$

Data Availability Statement

The borehole and ERT data can be downloaded from ESS-DIVE (Uhlemann et al., 2020): Geophysical borehole logging data of wells ER-GLS1, ER-GUM1, ER-PLM7, and ER-PLM8 at the East River Watershed, Colorado. Watershed Function SFA. https://doi.org/10.15485/1650355. The analysis code used in this work can be accessed at https://doi.org/10.5281/zenodo.4465155 (Miltenberger, 2021).

Acknowledgments
This material is based upon work supported as part of the Watershed Function Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-AC02-05CH11231. Additional support is from the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-SC0014664. The authors would also like to thank the sponsors of the Stanford Center for Earth Resources Forecasting and the Dean (Prof. S. Graham) of the Stanford School of Earth, Energy and Environmental Sciences for supporting this research. Finally, we thank the reviewers whose comments greatly improved this manuscript.

References

Anderson, E. (1905). The dynamics of faulting. *Transactions of the Edinburgh Geological Society*, *8*(3), 387–402. https://doi.org/10.1144/transed.8.3.387

Binley, A., Hubbard, S. S., Huisman, J. A., Revil, A., Robinson, D. A., Singha, K., & Slater, L. D. (2015). The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. *Water Resources Research*, *51*(6), 3837–3866. https://doi.org/10.1002/2015wr017016

Birgenheier, L. P., Horton, B., McCauley, A. D., Johnson, C. L., & Kennedy, A. (2017). A depositional model for offshore deposits of the lower blue gate member, mancos shale, Uinta Basin, Utah, USA. *Sedimentology*, *64*(5), 1402–1438. https://doi.org/10.1111/sed.12359

Bott, J. D. J., & Wong, I. G. (1995). The 1986 crested butte earthquake swarm and its implications for seismogenesis in colorado. *Bulletin of the Seismological Society of America*, *85*(5), 1495–1500.

Brugger, K. (2010). Climate in the southern Sawatch Range and Elk Mountains, Colorado, USA, during the last glacial maximum: Inferences using a simple degree-day model. *Arctic Antarctic and Alpine Research*, *42*(2), 164–178. https://doi.org/10.1657/1938-4246-42.2.164

Carroll, R. W. H., Deems, J. S., Niswonger, R., Schumer, R., & Williams, K. H. (2019). The importance of interflow to groundwater recharge in a snowmelt-dominated headwater basin. *Geophysical Research Letters*, *46*(11), 5899–5908. https://doi.org/10.1029/2019gl082447

Castilla-Rho, J., Rojas, R., Andersen, M., Holley, C., & Mariethoz, G. (2019). Sustainable groundwater management: How long and what will it take? *Global Environmental Change*, *58*, 101972. https://doi.org/10.1016/j.gloenvcha.2019.101972

Christensen, C. W., Hayashi, M., & Bentley, L. R. (2020). Hydrogeological characterization of an alpine aquifer system in the Canadian Rocky Mountains. *Hydrogeology Journal*, *28*(5), 1871–1890. https://doi.org/10.1007/s10040-020-02153-7

Da Gama, R. O., Lutz, B., Desjardins, P., Thompson, M., Prince, I., & Espejo, I. (2014). Integrated paleoenvironmental analysis of the niobrara formation: Cretaceous western interior seaway, Northern Colorado. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *413*, 66–80. https://doi.org/10.1016/j.palaeo.2014.05.005

Enemark, T., Peeters, L., Mallants, D., Flinchum, B., & Batelaan, O. (2020). A systematic approach to hydrogeological conceptual model testing, combining remote sensing and geophysical data. *Water Resources Research*, *56*(8), e2020WR027578. https://doi.org/10.1029/2020WR027578

Foster, L. M., & Maxwell, R. M. (2019). Sensitivity analysis of hydraulic conductivity and manning's n parameters lead to new method to scale effective hydraulic conductivity across model resolutions. *Hydrological Processes*, *33*(3), 332–349. https://doi.org/10.1002/hyp.13327

Gaskill, D. L. (1991). *Geologic map of the gothic quadrangle, Gunnison County, Colorado*. https://doi.org/10.3133/gq1689

Gutiérrez-Santolalla, F., Acosta, E., Ríos, S., Guerrero, J., & Lucha, P. (2005). Geomorphology and geochronology of sackung features (uphill-facing scarps) in the central spanish pyrenees. *Geomorphology*, *69*(1), 298–314. https://doi.org/10.1016/j.geomorph.2005.01.012

Hansen, T. M., Journel, A. G., Tarantola, A., & Mosegaard, K. (2006). Linear inverse Gaussian theory and geostatistics. *Geophysics*, *71*(6), R101–R111. https://doi.org/10.1190/1.2345195

Hermans, T., Nguyen, F., & Caers, J. (2015). Uncertainty in training image-based inversion of hydraulic head data constrained to ert data: Workflow and case study. *Water Resources Research*, *51*(7), 5332–5352. https://doi.org/10.1002/2014wr016460

Hermans, T., Nguyen, F., Klepikova, M., Dassargues, A., & Caers, J. (2018). Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using Bayesian evidential learning. *Water Resources Research*, *54*(4), 2931–2948. https://doi.org/10.1002/2017wr022135

Høyer, A.-S., Jørgensen, F., Sandersen, P., Viezzoli, A., & Møller, I. (2015). 3d geological modelling of a complex buried-valley network delineated from borehole and aem data. *Journal of Applied Geophysics*, *122*, 94–102.

Hubbard, S. S., Williams, K. H., Agarwal, D., Banfield, J., Beller, H., Bouskill, N. et al. (2018). The east river, Colorado, watershed: A mountainous community testbed for improving predictive understanding of multiscale hydrological–Biogeochemical dynamics. *Vadose Zone Journal*, *17*(1), 180061. https://doi.org/10.2136/vzj2018.03.0061

Jeffrey, R. (1983). *The logic of decision*. University of Chicago Press.

Jiang, Z., Mallants, D., Peeters, L., Gao, L., Soerensen, C., & Mariethoz, G. (2019). High-resolution paleovalley classification from airborne electromagnetic imaging and deep neural network training using digital elevation model data. *Hydrology and Earth System Sciences*, *23*(6), 2561–2580. https://doi.org/10.5194/hess-23-2561-2019

Knight, R., Smith, R., Asch, T., Abraham, J., Cannia, J., Viezzoli, A., & Fogg, G. (2018). Mapping aquifer systems with airborne electromagnetics in the central valley of California. *Groundwater*, *56*(6), 893–908. https://doi.org/10.1111/gwat.12656

Levandowski, W., Herrmann, R. B., Briggs, R., Boyd, O., & Gold, R. (2018). An updated stress map of the continental United States reveals heterogeneous intraplate stress. *Nature Geoscience*, *11*(6), 433–437. https://doi.org/10.1038/s41561-018-0120-x

Linde, N., Renard, P., Mukerji, T., & Caers, J. (2015). Geological realism in hydrogeological and geophysical inverse modeling: A review. *Advances in Water Resources*, *86*, 86–101. https://doi.org/10.1016/j.advwatres.2015.09.019

Liu, L., Spasojević, S., & Gurnis, M. (2008). Reconstructing farallon plate subduction beneath North America back to the late cretaceous. *Science*, *322*(5903), 934–938. https://doi.org/10.1126/science.1162921

Lopez-Alvis, J., Hermans, T., & Nguyen, F. (2019). A cross-validation framework to extract data features for reducing structural uncertainty in subsurface heterogeneity. *Advances in Water Resources*, *133*, 103427. https://doi.org/10.1016/j.advwatres.2019.103427

Malinverno, A., & Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical bayes and empirical bayes. *Geophysics*, *69*(4), 1005–1016. https://doi.org/10.1190/1.1778243

Miltenberger, A. (2021). *Fracture zone Monte Carlo source code and data analysis scripts*. https://doi.org/10.5281/zenodo.4465155

Oldenburg, D. W., & Li, Y. (1999). Estimating depth of investigation in dc resistivity and ip surveys. *Geophysics*, *64*(2), 403–416. https://doi.org/10.1190/1.1444545

Park, H., Scheidt, C., Fenwick, D., Boucher, A., & Caers, J. (2013). History matching and uncertainty quantification of facies models with multiple geological interpretations. *Computational Geosciences*, *17*(4), 609–621. https://doi.org/10.1007/s10596-013-9343-5

Pearl, J. (1988). Chapter 2 - Bayesian inference. In J. Pearl (Ed.), *Probabilistic reasoning in intelligent systems* (pp. 29–75). Morgan Kaufmann. https://doi.org/10.1016/b978-0-08-051489-5.50008-4

Phelps, G., Scheidt, C., & Caers, J. (2018). Exploring viable geologic interpretations of gravity models using distance-based global sensitivity analysis and kernel methods. *Geophysics*, *83*(5), G79–G92. https://doi.org/10.1190/geo2017-0742.1

Pirot, G., Huber, E., Irving, J., & Linde, N. (2019). Reduction of conceptual model uncertainty using ground-penetrating radar profiles: Field-demonstration for a braided-river aquifer. *Journal of Hydrology*, *571*, 254–264. https://doi.org/10.1016/j.jhydrol.2019.01.047

Pradhan, A., & Mukerji, T. (2019). *Seismic Bayesian evidential learning: Estimation and uncertainty quantification of sub-resolution reservoir properties*.

Pradhan, A., & Mukerji, T. (2020). Seismic Bayesian evidential learning: Estimation and uncertainty quantification of sub-resolution reservoir properties. *Computational Geosciences*, *24*(3), 1121–1140. https://doi.org/10.1007/s10596-019-09929-1

Radbruch-Hall, D. H., Varnes, D. J., & Colton, R. B. (1977). Gravitational spreading of steep-sided ridges ("sackung") in colorado. *Journal of Research of the U. S. Geological Survey*, *5*(3), 369–363.

Radbruch-Hall, D. H., Varnes, D. J., & Savage, W. Z. (1976). Gravitational spreading of steep-sided ridges ("sackung") in western United States. *Bulletin of the International Association of Engineering Geology*, *13*(1), 23–35. https://doi.org/10.1007/BF02634754

Reeves, D. M., Pohlmann, K. F., Pohll, G. M., Ye, M., & Chapman, J. B. (2010). Incorporation of conceptual and parametric uncertainty into radionuclide flux estimates from a fractured granite rock mass. *Stochastic Environmental Research and Risk Assessment*, *24*(6), 899–915. https://doi.org/10.1007/s00477-010-0385-0

Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., & Troldborg, L. (2012). Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. *Advances in Water Resources*, *36*, 36–50. https://doi.org/10.1016/j.advwatres.2011.04.006

Rücker, C., Günther, T., & Wagner, F. M. (2017). Pygimli: An open-source library for modelling and inversion in geophysics. *Computers & Geosciences*, *109*, 106–123. https://doi.org/10.1016/j.cageo.2017.07.011

Savage, W. Z., & Varnes, D. J. (1987). Mechanics of gravitational spreading of steep-sided ridges ("sackung"). *Bulletin of the International Association of Engineering Geology*, *35*(1), 31–36. https://doi.org/10.1007/BF02590474

Scheidt, C., Jeong, C., Mukerji, T., & Caers, J. (2015). Probabilistic falsification of prior geologic uncertainty with seismic amplitude data: Application to a turbidite reservoir case. *Geophysics*, *80*(5), M89–M12. https://doi.org/10.1190/geo2015-0084.1

Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying uncertainty in subsurface systems*. Wiley.

Si, H. (2015). Tetgen, a delaunay-based quality tetrahedral mesh generator. *ACM Transactions on Mathematical Software*, *41*(2), 1–36. https://doi.org/10.1145/2629697

Soldati, M. (2013). Deep-seated gravitational slope deformation. In P. T. Bobrowsky (Ed.), *Encyclopedia of natural hazards* (pp. 151–155). Springer Netherlands. https://doi.org/10.1007/978-1-4020-4399-4_86

Stephens, T. L., Walker, R. J., Healey, D., Bubeck, A., & England, R. W. (2018). Mechanical models to estimate the paleostress state from igneous intrusions. *Solid Earth*, *9*, 847–858. https://doi.org/10.5194/se-9-847-2018

Streufert, R. K., Eakins, W., & Hemborg, H. T. (1999). *Geology and mineral resources of gunnison county (Geology and Mineral Resources No. RS-37)*. Colorado Geological Survey, Division of Minerals and Geology, Department of Natural Resources.

Tabor, R. W. (1971). Origin of ridge-top depressions by large-scale creep in the Olympic Mountains, Washington. *GSA Bulletin*, *82*(7), 1811–1822. https://doi.org/10.1130/0016-7606(1971)82[1811:OORDBL]2.0.CO;2

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tarantola, A. (2006). Popper, bayes and the inverse problem. *Nature Physics*, *2*(8), 492–494. https://doi.org/10.1038/nphys375

Tarantola, A., & Valette, B. (1982). Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics*, *20*(2), 219–232. https://doi.org/10.1029/rg020i002p00219

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., et al. (2013). Ground water and climate change. *Nature Climate Change*, *3*(4), 322–329. https://doi.org/10.1038/nclimate1744

Tsai, F. T.-C. (2010). Bayesian model averaging assessment on groundwater management under model structure uncertainty. *Stochastic Environmental Research and Risk Assessment*, *24*(6), 845–861. https://doi.org/10.1007/s00477-010-0382-3

Tso, C.-H. M., Kuras, O., Wilkinson, P. B., Uhlemann, S., Chambers, J. E., Meldrum, P. I., et al. (2017). Improved characterisation and modelling of measurement errors in electrical resistivity tomography (ERT) surveys. *Journal of Applied Geophysics*, *146*, 103–119. https://doi.org/10.1016/j.jappgeo.2017.09.009

Tully, J. E. (2009). *Structural interpretation of the elk range thrust system, western Colorado*.

Uhlemann, S., Carr, B., Dafflon, B., & Williams, K. (2020). Geophysical borehole logging data of wells er-gls1, er-gum1, er-plm7, and er-plm8 at the east river watershed, Colorado. *Watershed Function SFA*, *1*. https://doi.org/10.15485/1650355

Uhlemann, S., Dafflon, B., Wainwright, H., Williams, K., Minsley, B., Zamudio, K., & Hubbard, S. (2021). Revealing covariance of below and above-ground variability across a mountainous watershed using machine learning and airborne em data. *Science Advances*.(under review)

Uhlemann, S., Wilkinson, P. B., Maurer, H., Wagner, F., Johnson, T., & Chambers, J. (2018). Optimized survey design for electrical resistivity tomography: Combined optimization of measurement configuration and electrode placement. *Geophysical Journal International*, *214*, 108–121. https://doi.org/10.1093/gji/ggy128

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0 contributors (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Ye, M., Pohlmann, K. F., Chapman, J. B., Pohll, G. M., & Reeves, D. M. (2010). A model-averaging method for assessing groundwater conceptual model uncertainty. *Groundwater*, *48*(5), 716–728. https://doi.org/10.1111/j.1745-6584.2009.00633.x

Yin, Z., Strebelle, S., & Caers, J. (2020). Automated monte carlo-based quantification and updating of geological uncertainty with borehole data (autobel v1.0. *Geoscientific Model Development*, *13*, 651–672. https://doi.org/10.5194/gmd-13-651-2020

21699356, 2021, 9, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021JB021767, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License