Authors' response to RC1 (Alexander Schaaf's) interactive comment on "Uncertainty assessment for 3D geologic modeling of fault zones based on geologic inputs and prior knowledge".

Reviewer's comments are shown in italicized fonts, and the authors' responses and changes to the manuscript are highlighted with callouts. The general comments are shown and addressed first, then the specific comments are addressed by line/figure number.

1 General comments

The paper presents a step forward in the uncertainty-aware modeling of subsurface faults in structural geomodels. The authors make use of Monte Carlo simulations to simulate uncertainty of fault zones based on a specific fault zone parameterization (surface traces, vertical termination surfaces, structural orientation and fault zone thickness) using a proprietary software suite. The authors elaborate the use of anisotropic spherical distributions for parameterizing orientation data for uncertainty simulation, which is a valuable contribution. The manuscript is overall well structured, except for a few re-arrangements necessary to increase readability (detailed in the specific comments). The authors give proper credit to related work and clearly indicate their own contribution. The title clearly reflects the contents of the paper. The figures presented will require some work to improve legibility and to avoid confusion of the reader.

But the authors appear to be confusing their simulation approach: They introduce MCUP (i.e. Monte Carlo simulation) in the methodology and properly parameterize their stochastic geomodel using probability distributions. But they then erroneously describe that they use Markov Chain Monte Carlo (MCMC) sampling. MCMC sampling is used for exploring the posterior space, which does not exist in a Monte Carlo simulation (i.e. MCUP). As the probability space is known in a Monte Carlo simulation, it needs no exploration. In a Monte Carlo simulation we only don't know how the combination of samples effect the output of the simulator function (the geomodeling software), thus we randomly sample (Monte Carlo sampling) from the parameter distributions to create a geomodel ensemble that shows us how the uncertainty in the input parameters effects the geomodel output. Luckily, to my knowledge, the used probabilistic programming framework pymc3 defaults to Monte Carlo sampling when no likelihood function is given (and thus no Bayesian inference can be conducted). Thus the authors appear to have accidentally conducted the simulations they wanted to do (MCUP/Monte Carlo sampling). The use of trace plots (as in Figure 6 and 8) for Monte Carlo simulation results is meaningless though (and potentially misleading), as no sampler is being used that requires determination of convergence. As, luckily, the presented simulation results appear to be valid MCUP results, the authors only need to change their writing accordingly, without the need for re-running simulations.

In its current state, mixing up the terminology of Bayesian inference and MCUP, I can not recommend the paper to be accepted. But if the authors fix their method descriptions and discussions of the results to fit the MCUP simulations they actually conducted, I believe this could become a valid scientific contribution that is worth publishing in Solid Earth.

Authors' answer:

The authors truly appreciate the detailed clarifying comments, especially regarding the description of the simulation approach. The authors acknowledge that there was confusion regarding how the simulation approach was described, primarily in the erroneous usage of the term Markov Chain Monte Carlo (MCMC) when the method applied in the study is in fact simply Monte Carlo sampling to explore the set of input probability distributions. The authors wish to reaffirm that the methodology employed in the study is intended to be that which the reviewer identified: the use of Monte Carlo sampling to explore the prior uncertainty space of geologic modeling inputs. As is, the analysis performed in the study does not contain any use of Bayesian inference via MCMC (i.e., no likelihood functions were defined or used), nor was this an intended description of the methodology employed. The authors have taken steps throughout the text to remove any erroneous descriptions of the methodology and clarify the intended use of Monte Carlo sampling. A broad overview of the changes made to rectify this error include (i) removal of any mention of MCMC or posterior distributions and replacement with appropriate terminology and (ii)

the removal of trace plots and a rethinking of the simulation quality assessment section.

2 Specific comments

2.1 1 - Introduction

2.1.1 *The Introduction of the paper needs to clearly state the scope of the study / hypothesis to be tested or explored.*

Authors' answer:

The two paragraphs previously in lines 52-68 have been incorporated fluidly into the introduction to specify the scope and motivation of the study.

Changes to the manuscript:

Text moved to the end of Section 1 – Introduction. . . This study expands the use of ~~MCUP~~ probabilistic geomodeling to a new aspect of geologic modeling – fault zones, or the localized volume of fractured and displaced rock surrounding a finite fault surface, typically composed of a fault core and a damage zone (Caine et al., 1996; Childs et al., 2009; Peacock et al., 2016; Choi et al., 2016). Fault zones introduce regions of altered geotechnical strength and hydraulic permeability into the surrounding in-tact rockmass and are therefore of major importance to geological engineering projects that rely on accurate assessments of subsurface rock properties (e.g., tunnels, mines). While faults have been the focus of a significant amount of recent geologic modeling research (Røe et al., 2014; Cherpeau et al., 2010; Cherpeau and Caumon, 2015; Aydin and Caers, 2017), these works have focused on modeling fault surfaces directly rather than modeling the 3D geometry of fault zones. Detailed modeling of the 3D geometry of fault zones can improve the understanding of faults' impacts on geotechnical and reservoir engineering projects due to the fact that variations in fault zone thickness or composition can greatly alter the mechanical and hydrological behavior of a fault, e.g., its sealing potential (Caine et al., 1996; Fredman et al., 2008; Manzocchi et al., 2010). ~~By an in-depth search of the literature, as of yet there is no dedicated approach to characterizing the uncertainty of fault zones in 3D geologic models.~~ Building on the existing literature on understanding the uncertainties about faults in the subsurface (Choi et al., 2016; Shipton et al., 2019; Torabi et al., 2019b), this study develops a novel, dedicated approach to leveraging probabilistic geomodeling to characterize the uncertainty of fault zones using 3D geologic models.

Fault zones may be irregular in shape, creating complex geometries which are difficult to characterize quantitatively (Torabi et al., 2019a, b). Peacock et al. (2016) provide a detailed list of the various types of damage zones and intersecting fault networks that comprise the general term "fault zone". ~~The inherent complexity of fault zone structure makes their precise modeling intractable in an automated MCUP formulation.~~ The inherent complexity of fault zone structure makes their precise modeling intractable in an automated geologic modeling application, such as that required by probabilistic geomodeling. A simplified approach to modeling fault zones in 3D geologic models is developed in this study based on the key elements defining fault zone geometry at a practical level of detail.

2.2 2 - Model implementation

2.2.1 *L52 - Both paragraph (lines 52-68) need to incorporated into the introduction as they define the scope and motivation of the study.*

Authors' answer:

See above comment regarding reorganizing the content of the Introduction.

2.3 3 - Probability distributions for MCUP

2.3.1 *L138 - How do you evaluate the likelihood of the proposal step in a Markov Chain during MCUP? This is only possible when doing a Bayesian inference, not a Monte Carlo uncertainty propagation, as you don't have any likelihood function.*

Authors' answer:

The erroneous mention of MCMC sampling algorithms has been removed and edited to specify the use of Monte Carlo sampling algorithms. This eliminates the question of "How do you evaluate the likelihood of the proposal step in a Markov Chain during MCUP?" by clarifying that only Monte Carlo sampling is performed, and not Bayesian inference.

Changes to the manuscript:

~~Simulation of scalar data is straightforward and well-established through the use of Markov-chain Monte Carlo (MCMC) sampling algorithms, easily accessible through the open source Python package *PyMC3* (Salvatier et al., 2016).~~ Simulation of scalar data is straightforward and well-established through the use of Monte Carlo sampling algorithms, easily accessible through the open source Python package *PyMC3* (Salvatier et al., 2016).

2.3.2 *Overall the description of simulation/sampling should be moved into Section 3.2.*

Authors' answer:

The description of simulation and sampling has been incorporated into Section 3.2.

Changes to the manuscript:

The paragraph referenced has been moved to Section 3.2: Simulation of scalar data is straightforward and well-established through the use of Markov-chain Monte Carlo (MCMC) sampling algorithms, easily accessible through the open source Python package *PyMC3* (Salvatier et al., 2016). The *PyMC3* library has been demonstrated as a platform for performing MCUE of 3D geologic models (de la Varga and Wellmann, 2016; Schneeberger et al., 2017), and has even been implemented in the open source geologic modeling platform GemPy (de la Varga et al., 2019). An additional consideration in the case of continuous data types is the distinction between scalar and vectorial data (e.g., structural orientations). A probability distribution describing orientation data resides on the surface of a unit-sphere in 3D, and can be characterized using spherical probability distributions (Fisher et al., 1987; Mardia and Jupp, 2000). The benefit of using spherical probability distributions to describe structural orientation uncertainty in 3D geologic modeling is clearly stated by Pakyuz-Charrier et al. (2018b), and their application in MCUE formulations continues to develop (Pakyuz-Charrier et al., 2018b, a; Carmichael and Ailleres, 2016). To remain concise, the following section focuses on the new contributions made to the use of spherical probability distributions utilizing the *R-fast* open source package available in the R language (Papadakis et al., 2018).

2.3.3 *L140 - The paper de la Varga & Wellmann (2016) uses pymc to conduct a Bayesian inference - thus not MCUP.*

Authors' answer:

While the referenced study in question includes the use of the MCUP approach for probabilistic geomodeling, the authors realize the lack of clarity due to the referenced paper's focus on *PyMC*'s capability to perform Bayesian inference. The sentence has been reworded and expanded to emphasize the capability of *PyMC3* to perform Monte Carlo sampling outside of a

Bayesian inference, fitting its application in the current study.

Changes to the manuscript:

~~The *PyMC3* library has been demonstrated as a platform for performing MCUP of 3D geologic models (de la Varga and Wellmann, 2016; and has even been implemented in the open source geologic modeling platform GemPy (de la Varga et al., 2019).~~ The *PyMC3* library is designed to facilitate Bayesian inference using computational sampling algorithms, though the inclusion of likelihood functions is not required thereby allowing for utilization of the package functions for Monte Carlo sampling alone. The use of *PyMC3* has been demonstrated successfully in the context of 3D geologic modeling by de la Varga and Wellmann (2016); Schneeberger et al. (2017), and its implementation in *Theano* has allowed for seamless integration with the open source geologic modeling platform GemPy (de la Varga et al., 2019). This study focuses solely on the step of probabilistic geomodeling based on 3D geologic modeling inputs, leveraging only the Monte Carlo sampling capabilities of *PyMC3*.

2.4 3.2 - Simulation

2.4.1 *A more adequate name for the section would be "Sampling".*

Authors' answer:

The section name has been changed to Sampling.

2.5 3.3 - Rotation

2.5.1 *This section is part of sampling and should be merged into Section 3.2*

Authors' answer:

The section has been appended to Section 3.2 - Sampling.

2.6 4.2 - Surface trace

2.6.1 *L291 - It is unclear to me what the "approximate geographical error of known landmarks" is.*

Authors' answer:

The section containing this sentence has been revised and expanded to clearly define the sources of uncertainty affecting the fault trace and the methods with which they are quantified - including the geographical error arising from the use of a historic geologic map.

Changes to the manuscript:

The uncertainty affecting the surface fault trace results in changes in the trace location and shape. Independent perturbations of the trace's endpoints are applied and linearly propagated along the fault trace to arrive at a smoothly varied location and shape. ~~A normal distribution characterizing the uncertainty about the location of each trace endpoint is parameterized from the joint uncertainty stemming from fault zone centerline definition, digitization error and geographical errors in addition to a random direction of shift obtained from a uniform distribution.~~ The three primary sources of uncertainty are quantified using the available information listed in respective order: average fault zone thickness, published metrological studies (Zhong-Zhong, 1995) and approximate geographical error of known landmarks (e.g., mountain tops). ~~Additional details on the method of~~

~~perturbing the location and shape of the surface trace based on the joint effect of these sources of uncertainty are available in the code supplement.~~

A bounded uniform distribution is parameterized to simulate a random direction of perturbation for each trace endpoint due to geographical error (i.e., drafting and georeferencing error). The normal distribution representing the total bound on geographical error is converted to respective $\hat{x}$ and $\hat{y}$ components using the directional cosine of the angle sampled from the uniform distribution. This conversion to unit components is used similarly with the fault zone centerline definition uncertainty and digitization uncertainty using the acute angle θ between the orientation of the fault trace with the northing and easting directions. An additional logical check for the strike quadrant of the surface trace is required to implement this approach.

The individual sources of uncertainty affecting the surface trace endpoint locations are combined into a derived distribution using a deterministic function to determine the total uncertainty affecting the location of each endpoint, given by Eq. 1.

$$\begin{aligned} P(\hat{x}|\sigma_{centerline},\sigma_{dig},\sigma_{geo},\theta) &= cos(\theta)\Big(N(0,\sigma_{centerline})+N(0,\sigma_{dig})\Big)+N(0,\sigma_{geo})sin\Big(U(0,2\pi)\Big), \\ P(\hat{y}|\sigma_{centerline},\sigma_{dig},\sigma_{geo},\theta) &= sin(\theta)\Big(N(0,\sigma_{centerline})+N(0,\sigma_{dig})\Big)+N(0,\sigma_{geo})cos\Big(U(0,2\pi)\Big) \end{aligned} \tag{1}$$

The average fault zone thickness was used to characterize the fault zone centerline definition uncertainty affecting each surface trace endpoint. The geographical error was calculated to be approximately 40 meters based on the average distance measured between known landmarks (e.g., mountain tops) on the geologic map and modern satellite imagery data. For both of these sources of uncertainty, the maximum error range described is treated as a 95% confidence interval, allowing a normal distribution to be parameterized with a mean of zero and a standard deviation equal to $maximumerror/3.92$. The digitization error for a 1:12,000 map was represented by a normal distribution with a standard deviation of 3.666 m based on (Zhong-Zhong, 1995).

2.7 4.3 - Vertical termination depth

2.7.1 *L312 - What is a deterministic distribution? Do you mean a derived distribution? Or an empirically parametrized distribution? A distribution should be by definitiv nondeterministic.*

Authors' answer:

The phrase "deterministic distribution" was used following the *PyMC3* function terminology pm.Deterministic(), which is used for combining multiple stochastic variables (i.e., distributions) using a deterministic function. As the reviewer notes, this function is in fact empirically parameterized, and the more appropriate name "empirically derived distribution" – also known as a deterministic model combining multiple stochastic variables in *PyMC3* terminology – has been substituted. The description of the use of this style of distribution has been modified to clarify the reliance on combining several stochastic variables using an empirically derived deterministic function.

Changes to the manuscript:

~~The vertical termination depth is sampled from a deterministic distribution combining uncertainty about the aspect ratio, fault trace persistence and prior information of the fault elevation and average dip.~~ Sampling the uncertainty of the fault zone vertical termination depth involves combining multiple probability distributions using a deterministic function to generate an empirically derived probability distribution. In the derived distribution for fault zone vertical termination depth, f_{length} and $Aspectratio$ are characterized as independent probability distributions (respectively) and combined using a deterministic function based on the empirically derived description of 3D fault surface geometry, $z_{term} = z_{outcrop} - f_{height} * sin(\theta); f_{height} = \frac{f_{length}}{Aspectratio}$. In this manner, the vertical termination depth (z_{term}) is calculated by converting the fault height to the vertical height using the average dip angle (θ) and subtracting this from the average elevation of the fault outcrop ($z_{outcrop}$).

2.8 4.5 - Simulation quality assessment

2.8.1 *L334 - Without a likelihood function you can't use a MCMC sampler, as you are unable to evaluate the step proposals.*

2.8.2 *L336 - In an MCUP simulation, you have no posterior uncertainty space, as you are not using any likelihood function. You are mixing up terminology of MCUP and Bayesian inference. Again, MCMC sampling is only possible with a likelihood function (thus not in MCUP).*

Authors' answer:

The entire Section 4.5 has been revised in depth to remove erroneous inclusions of the terminology MCMC sampler and to remove the erroneous use of trace plots in assessing the quality of Monte Carlo simulation for exploring the prior uncertainty space.

Changes to the manuscript:

The quality of probabilistic simulation ~~relies on primarily~~ is a product of the size of the uncertainty space, the simulation method used and the number of samples drawn. For any simulation, the realizations generated can be plotted in the data space and visually examined for appropriate coverage and shape (termed a realization plot). For ~~non-spherical~~ scalar data types, ~~use of MCMC simulation methods creates trace plots and posterior~~ histograms of the Monte Carlo draws provide an intuitive method for independently assessing the quality of simulation for each input. ~~, which provide an intuitive method for independently assessing the quality of simulation for each input. Uniformity of the trace plot and width of the 95% highest posterior density (HPD) indicate, respectively, the convergence of the Markov Chain sampling and the thorough exploration of the posterior uncertainty space.~~ Visual analysis of the shape of the histogram compared to the expected shape of the distribution and a comparison between the input distribution parameters (e.g., mean and standard deviation for a normal distribution) and their values calculated from the samples can quickly determine whether the samples drawn have sufficiently explored the uncertainty space. Figure 6 shows an example of the realization plot~~, trace plots and posterior~~ sample histograms generated for the simulation of vertical termination depths from Section 4.3. This figure allowed for identifying a strong tailing behavior in the output realizations, leading to a reparameterization discussed in Section 6.2.

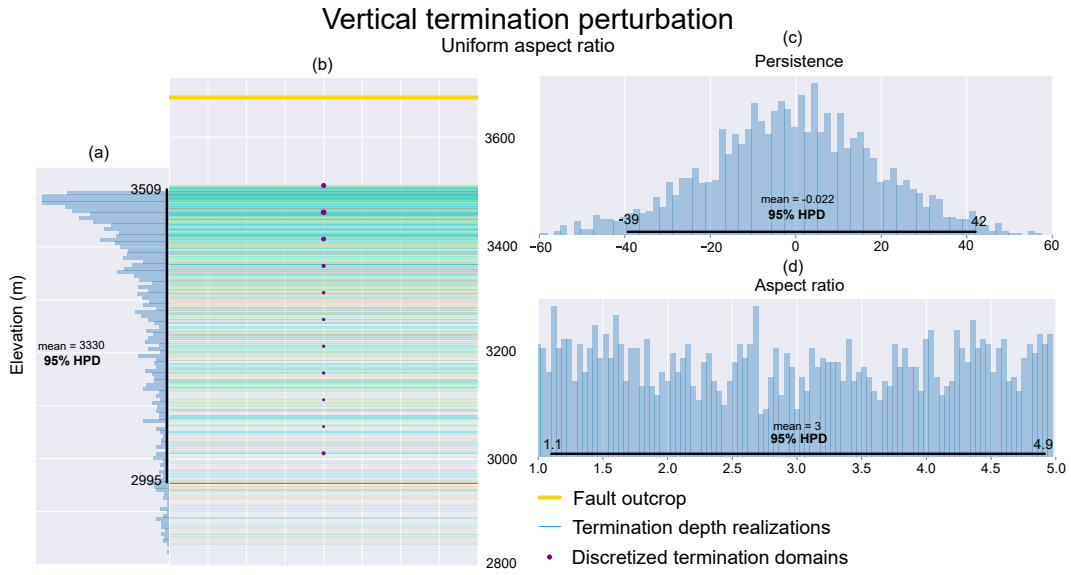

Figure 1. ~~Realizations and Monte Carlo analysis results (trace plot and posterior histograms)~~ Visualization of Monte Carlo samples and associated geologic input realizations from perturbation of the fault zone vertical termination depth based on a uniform distribution of fault aspect ratio. The 95% highest ~~posterior~~ predictive density is overlain on the ~~posterior~~ histograms of the Monte Carlo samples.

~~Trace plots are not available for the spherical data simulations due to reliance on the acceptance-rejection simulation method,~~ ~~while posterior~~ For spherical data simulations, histograms may be replaced by Exponential Kamb contouring (Vollmer, 1995) or Rose diagrams to visualize the density of sampled poles across the surface of the unit sphere (as projected onto a lower-hemisphere projection). This visual assessment provides a semi-quantitative evaluation of the shape and distribution of the ~~posterior spherical probability distribution~~ sampled structural orientations. Additionally, a recalculation of the eigenvector decomposition from the set of simulated samples provides a measure of the accuracy of the posterior distribution with respect to the input ~~orientation~~ parameter values. Tools for generating figures for simulation quality assessment are provided and detailed in the input perturbation script.

Based on the assessment of simulation quality and consideration of compounding factors during uncertainty propagation, the MCUE formulation for the single fault model was run for a number of various realization counts (100, 300, 500, 1,000, 2,000 and 3,000). The processing time generally increases linearly with realization count, reaching many hours to several days for high realization counts on the single fault mock model containing 2.5 million cells. The vast majority of processing time is taken up by the model updating and block model calculation in Leapfrog. For the single fault mock model with 1,000 realizations and 2.5 million cells the sampling benchmark time was 87 seconds while the model processing benchmark time was 38.5 hours. This study is intended to introduce and expand on the use of MCUE formulations for specific geologic modeling problems, and work regarding optimizing the efficiency of model processing is not a focus. The experiments conducted do highlight the need to understand (i) the realization requirement for exploring modeling inputs independently and its relationship to the size of the independent uncertainty spaces, (ii) the interactions of various, related parameters during the uncertainty propagation step and (iii) identification of a balance between final model resolution, coverage, complexity and processing time.

2.9 6.2 - Model parameterization

2.9.1 *L388 - The use of "posterior distribution" is false, as you are doing MCUP, not a Bayesian inference.*

Authors' answer:

In line with the general comments regarding erroneous usage of terms from Bayesian inference (MCMC, posterior distributions), the description here has been revised to appropriately describe the prior predictive model that was explored using Monte Carlo sampling.

Changes to the manuscript:

However, the ~~posterior~~ empirically derived distribution of vertical termination depths resulting from a bounded uniform parameterization of fault aspect ratio showed a strong tailing effect (right skewed).

2.10 6.3 - Parameter relationships

2.10.1 *L411 - The meaning of the entire paragraph is unclear to me and needs to be revised.*

Authors' answer:

The specified paragraph on addressing the observed relationships among geologic modeling input parameters using the MCUP formulation has been revised clarify the authors' stance on why and how these parameter relationships arise in the MCUP formulation. In essence, the paragraph is intended to highlight the potential for undersampling the geologic model uncertainty space when considering overlapping uncertainty envelopes of individual model inputs.

Changes to the manuscript:

~~Relationships between modeling inputs also arise in different ways, for example the vertical termination depth and structural orientation uncertainty envelopes overlap heavily in the combined model uncertainty (Figure 5) leading to undersampling of the model uncertainty space when the independent uncertainty envelopes are combined. Similar behavior is observed when comparing orientation perturbations to fault zone thickness where thinner fault zones require finer orientation perturbations to fully populate the uncertainty space of the 3D geologic model.~~ Despite performing a thorough exploration of each, independent parameter's uncertainty during Monte Carlo sampling (Section 4.5), undersampling of the combined geologic model uncertainty space can still occur during uncertainty propagation. An example of this arises when considering the vertical termination depth and structural orientation. Truncation of fault zone realizations at any given termination interval effectively reduces the number of realizations available for sampling the full range of structural orientation uncertainty at deeper intervals. This is evidenced in Figure 5(b) by the increasing prevalence of "stair-stepping" artefacts in the combined model uncertainty with depth.

2.10.2 *L419 - Gibbs sampling is not applicable to MCUP, as no likelihood is used.*

Authors' answer:

The erroneous mention of Gibbs sampling has been removed. The section has been revised to better highlight the potential for exploring the input uncertainty space using joint distributions among various parameters believed to be correlated.

Changes to the manuscript:

A treatment of these relationships through ~~the use of Gibbs sampling or other conditional methods of Monte Carlo simulation~~ parameterizing previously independent input probability distributions using a joint distribution (and an appropriate sampling scheme) could potentially generate more realistic and efficient assessments of model uncertainty.

3 Figures

3.1 Figure 2

The dotted volume texture makes annotations for uncertainty extremely hard to read. The same goes for there fault zone signature/texture. I'd highly recommend removing as much texture as possible from the plot to improve legibility.

"The Visual Display of Quantitative Information" by Edward Tufte provides ample of additional reasons for reducing distracting "ink" from scientific visualizations and is well worth a read :-)

Authors' answer:

The suggested changes have been made to the fault zone schematic figure, clearing away non-informative ink and emphasizing the illustrated uncertainty envelopes. Thank you very much to the reviewer for the kind suggestion on better practices for graphical display.

Changes to the manuscript:

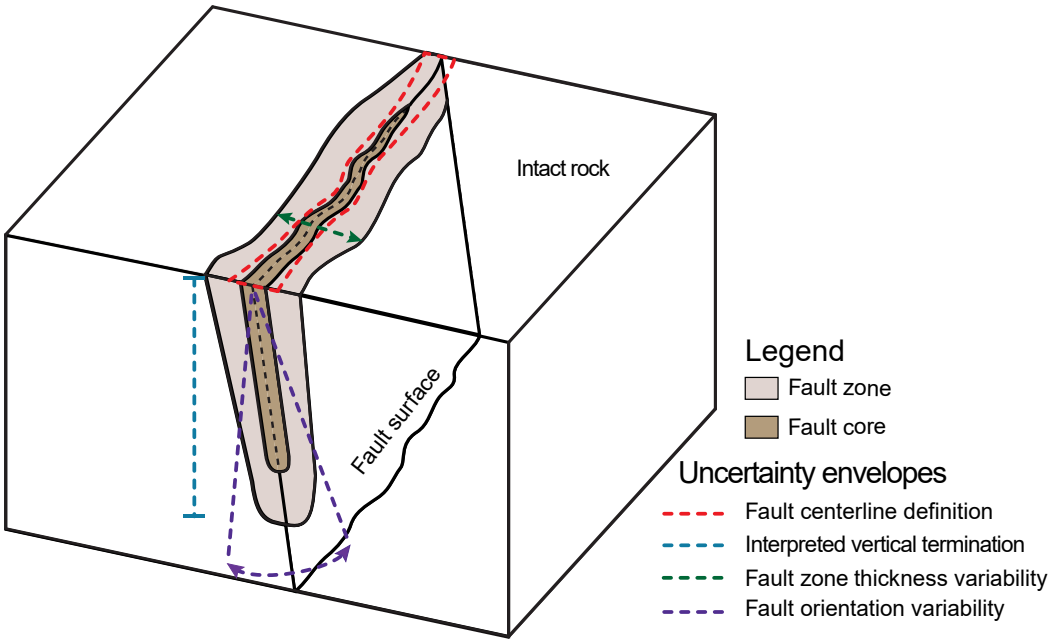

Figure 2. A schematic showing ~~the sources of uncertainty~~ possible uncertainty envelopes about the four geologic modeling inputs used to characterize the 3D geometry of a fault zone in the subsurface. Modified from Krajnovich et al. (2020).

3.2 Figure 4

Highlighting of fault traces is really difficult to see. I highly recommend making this figure more legible to the reader by removing visual complexity: e.g. remove coloring of geological map in the background.

Authors' answer:

The figure has been edited to make the fault traces bolder and to provide a stronger contrast between the fault traces and the geological map in the background.

Changes to the manuscript:

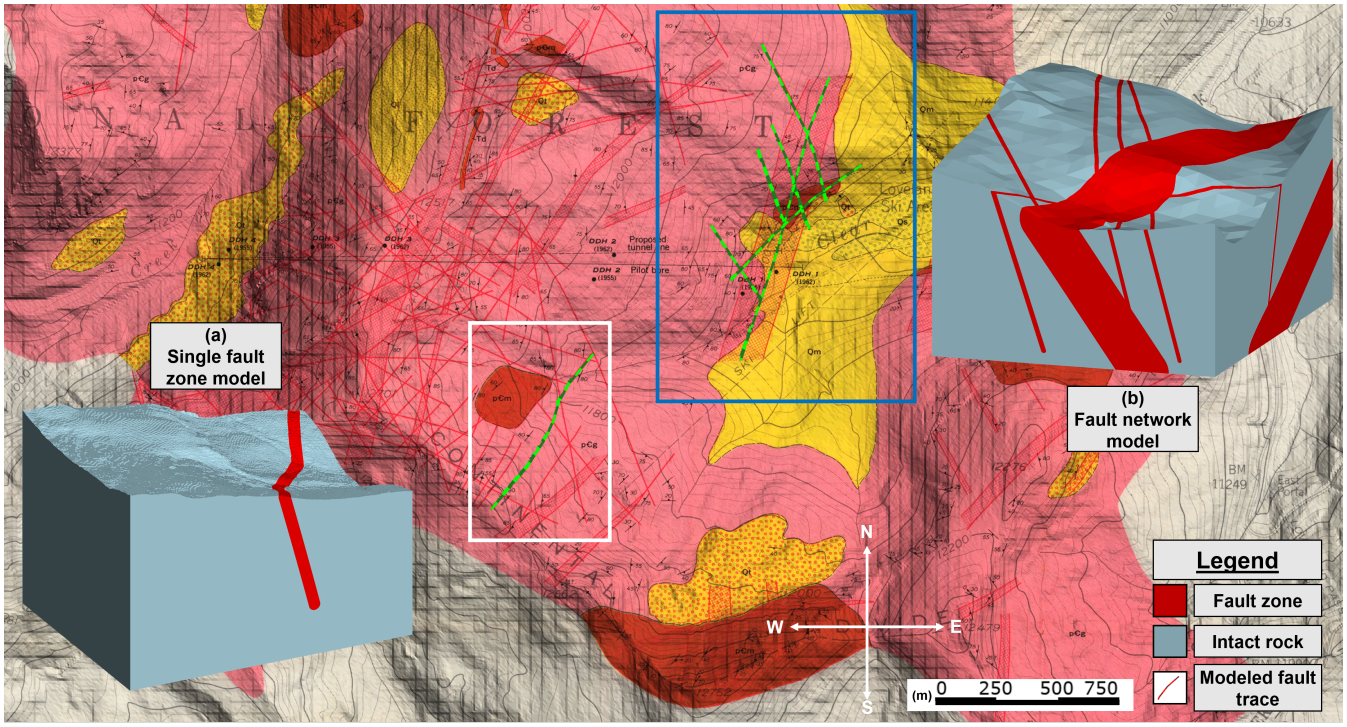

Figure 3. 1:12,000 geologic map from Robinson et al. (1974) showing mapped fault zones of varying widths. The white rectangle and associated overlay (a) show the single fault model while the blue rectangle and associated overlay (b) show the fault network model. Fault trace(s) used for modeling are highlighted within each rectangle as green polylines.

3.3 Figure 5

- *legend is barely legible - please increase text size*
- *entropy plot of the fault zone thickness barely shows any uncertainty. If your discretization is not fine enough to resolve the simulated uncertainties, then is it worth incorporating into you model?*

Authors' answer:

The legend text size has been increased to match the average text size of the paper's text.

The reviewer's comment is interesting and insightful, and is addressed briefly in the preceding section. While it is quite clear that the fault zone thickness uncertainty is largely insignificant in this single fault model (where the fault zone thickness was 8 m, vs. the 5 m block size), the authors observed that the wider fault zone present in the second, fault network model

would contribute more heavily to the model uncertainty. This reasoning, and the desire to present general recommendations to modeling the uncertainty of fault zones, led the authors to leave the results as is for the single fault model.

Changes to the manuscript:

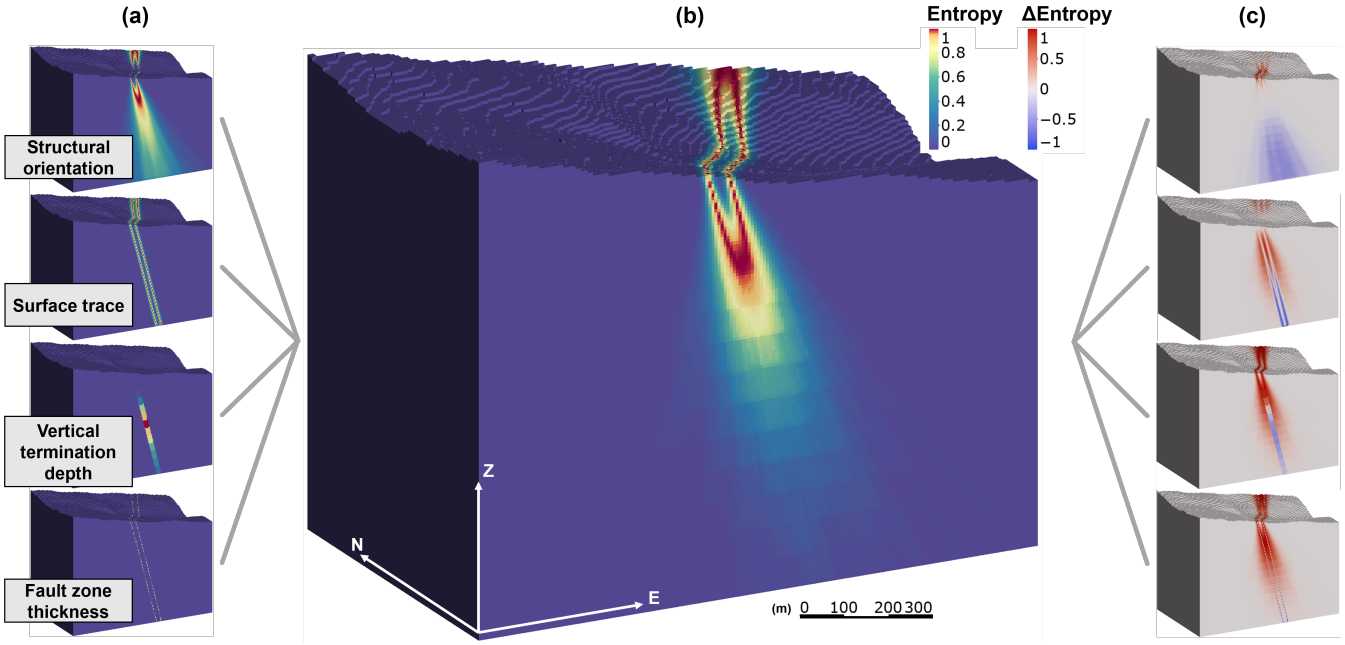

Figure 4. Block models showing information entropy quantified from (a) independent modeling inputs and (b) combined modeling inputs. The difference between the combined geologic model uncertainty and each independent modeling input is shown in (c), where blue values indicate that the independent modeling input showed greater entropy than the combined model uncertainty.

3.4 Figure 6

The use of trace plots is only useful if evaluating convergence of (e.g.) Markov chains. MCUP uses Monte Carlo simulation, thus the use of trace plots serves no purpose and is confusing. Also the rug plot on the left size shows the same information as the histogram of the vertical termination depth in the lower right. I'd recommend just using the histogram to demonstrate that you've sampled enough samples.

Authors' answer:

In line with this specific comment and the general comments above, the erroneous referrals to MCMC methods and the use of trace plots have been eliminated from the paper. They have been replaced by the appropriate discussion of assessing the exploration of the input uncertainty space using graphical representations of realizations and histograms of the Monte Carlo samples.

Changes to the manuscript:

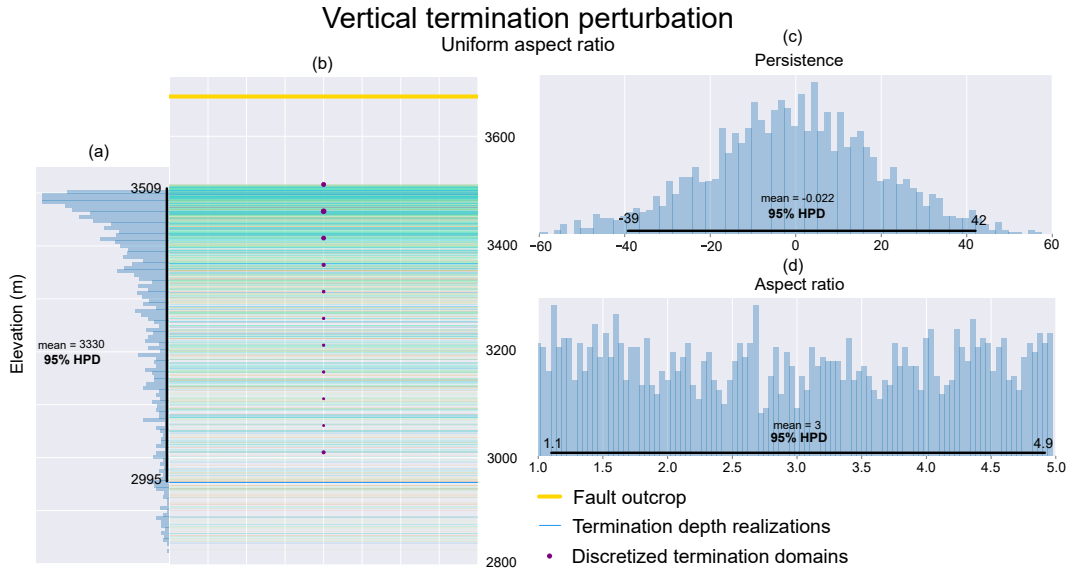

Figure 5. ~~Realizations and Monte Carlo analysis results (trace plot and posterior histograms)~~ Visualization of Monte Carlo samples and associated geologic input realizations from perturbation of the fault zone vertical termination depth based on a uniform distribution of fault aspect ratio. The 95% highest ~~posterior~~ predictive density is overlain on the ~~posterior~~ histograms of the Monte Carlo samples.

3.5 Figure 8

Authors' answer:

See above comment.

Changes to the manuscript:

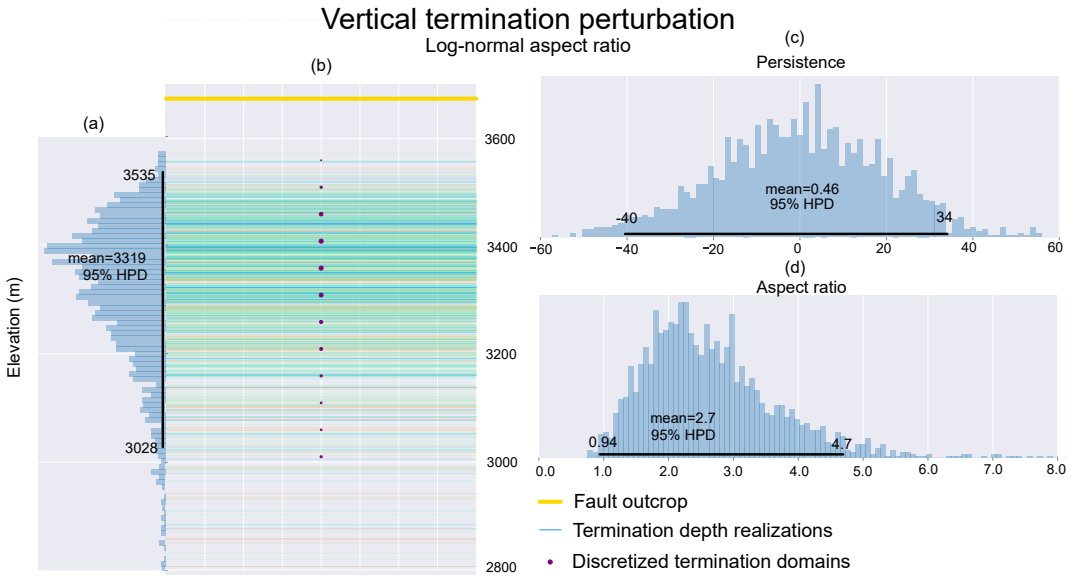

Figure 6. ~~Realizations and Monte Carlo analysis results (trace plot and posterior histograms)~~ Visualization of Monte Carlo samples and associated geologic input realizations from perturbation of the fault zone vertical termination depth, reparameterized using a log-normal distribution of fault aspect ratio. The 95% highest ~~posterior~~ predictive density is overlain on the ~~posterior~~ histograms of the Monte Carlo samples.

References

Aydin, O. and Caers, J. K.: Quantifying structural uncertainty on fault networks using a marked point process within a Bayesian framework, Tectonophysics, 712-713, 101–124, https://doi.org/10.1016/j.tecto.2017.04.027, 2017.

Caine, J. S., Evans, J. P., and Forster, C. B.: Fault zone architecture and permeability structure, Geology, 24, 1025–1028, https://doi.org/10.1130/0091-7613(1996)024<1025:FZAAPS>2.3.CO;2, 1996.

Carmichael, T. and Ailleres, L.: Method and analysis for the upscaling of structural data, Journal of Structural Geology, 83, 121–133, https://doi.org/10.1016/j.jsg.2015.09.002, 2016.

Cherpeau, N. and Caumon, G.: Stochastic structural modelling in sparse data situations, Petroleum Geoscience, 21, 233–247, https://doi.org/10.1144/petgeo2013-030, 2015.

Cherpeau, N., Caumon, G., and Lévy, B.: Stochastic simulations of fault networks in 3D structural modeling, Comptes Rendus - Geoscience, 342, 687–694, https://doi.org/10.1016/j.crte.2010.04.008, 2010.

Childs, C., Manzocchi, T., Walsh, J. J., Bonson, C. G., Nicol, A., and Schöpfer, M. P.: A geometric model of fault zone and fault rock thickness variations, Journal of Structural Geology, 31, 117–127, https://doi.org/10.1016/j.jsg.2008.08.009, 2009.

Choi, J. H., Edwards, P., Ko, K., and Kim, Y. S.: Definition and classification of fault damage zones: A review and a new methodological approach, Earth-Science Reviews, 152, 70–87, https://doi.org/10.1016/j.earscirev.2015.11.006, 2016.

de la Varga, M. and Wellmann, J. F.: Structural geologic modeling as an inference problem: A Bayesian perspective, Interpretation, 4, SM1–SM16, https://doi.org/10.1190/INT-2015-0188.1, 2016.

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

Fisher, N., Lewis, T., and Embleton, B.: Statistical analysis of spherical data, Cambridge University Press, 1987.

Fredman, N., Tveranger, J., Cardozo, N., Braathen, A., Soleng, H., Røe, P., Skorstad, A., and Syversveen, A. R.: Fault facies modeling: Technique and approach for 3-D conditioning and modeling of faulted grids, AAPG Bulletin, 92, 1457–1478, https://doi.org/10.1306/06090807073, 2008.

Krajnovich, A., Zhou, W., and Gutierrez, M.: Uncertainty Assessment of Fault Zones in 3D Geologic Models of Mountain Tunnels (in press), in: Word Tunnel Congress, International Tunneling and Underground Space Association, 2020.

Manzocchi, T., Childs, C., and Walsh, J. J.: Faults and fault properties in hydrocarbon flow models, Geofluids, 10, 94–113, https://doi.org/10.1111/j.1468-8123.2010.00283.x, 2010.

Mardia, K. V. and Jupp, P. E.: Directional Statistics, John Wiley & Sons, London, https://doi.org/10.1002/9780470316979, 2000.

Pakyuz-Charrier, E., Giraud, J., Ogarko, V., Lindsay, M., and Jessell, M.: Drillhole uncertainty propagation for three-dimensional geological modeling using Monte Carlo, Tectonophysics, 747-748, 16–39, https://doi.org/10.1016/j.tecto.2018.09.005, 2018a.

Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J., and Jessell, M.: Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modeling, a guide for disturbance distribution selection and parameterization, Solid Earth, 9, 385–402, https://doi.org/10.5194/se-9-385-2018, 2018b.

Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., Borboudakis, G., Burkardt, J., Zou, C., and Lakiotaki, K.: Package 'Rfast', https://rfast.eu, 2018.

Peacock, D. C., Nixon, C. W., Rotevatn, A., Sanderson, D. J., and Zuluaga, L. F.: Glossary of fault and other fracture networks, Journal of Structural Geology, 92, 12–29, https://doi.org/10.1016/j.jsg.2016.09.008, 2016.

Robinson, C. S., Lee, F. T., Scott, J. H., Carroll, R. D., Hurr, R. T., Richards, D. B., Mattei, F. A., Hartmann, B. E., and Abel, J. F.: Engineering Geologic, Geophysical, Hydrologic, and Rock-Mechanics Investigations of the Straight Creek Tunnel Site and Pilot Bore, Colorado, Tech. rep., United States Geological Survey, Washington, D.C., https://doi.org/10.3133/pp815, 1974.

Røe, P., Georgsen, F., and Abrahamsen, P.: An Uncertainty Model for Fault Shape and Location, Mathematical Geosciences, 46, 957–969, https://doi.org/10.1007/s11004-014-9536-z, 2014.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C.: Probabilistic programming in Python using PyMC3, PeerJ Computer Science, 2, 1–24, https://doi.org/10.7717/peerj-cs.55, 2016.

Schneeberger, R., de la Varga, M., Egli, D., Berger, A., Kober, F., Wellmann, F., and Herwegh, M.: Methods and uncertainty-estimations of 3D structural modelling in crystalline rocks: A case study, Solid Earth, 8, 987–1002, https://doi.org/10.5194/se-2017-47, 2017.

Shipton, Z., Roberts, J., Comrie, E., Kremer, Y., Lunn, R., and Caine, J.: Fault fictions : cognitive biases in the conceptualization of fault zones, Geological Society Special Publications, 2019.

Torabi, A., Alaei, B., and Libak, A.: Normal fault 3D geometry and displacement revisited: Insights from faults in the Norwegian Barents Sea, Marine and Petroleum Geology, 99, 135–155, https://doi.org/10.1016/j.marpetgeo.2018.09.032, 2019a.

Torabi, A., Johannessen, M. U., and Ellingsen, T. S. S.: Fault Core Thickness: Insights from Siliciclastic and Carbonate Rocks, Geofluids, 2019, 1–24, https://doi.org/10.1155/2019/2918673, 2019b.

Vollmer, F. W.: C program for automatic contouring of spherical orientation data using a modified Kamb method, 21, 31–49, https://doi.org/10.1016/0098-3004(94)00058-3, 1995.

Zhong-Zhong, C.: The estimation of digitizing error and its propagation results in GIS and application to habitat mapping, Ph.D. thesis, University of Massachusetts Amherst, https://scholarworks.umass.edu/dissertations/AAI9524686, 1995.

Authors' response to RC2 (Florian Wellmann's) interactive comment on "Uncertainty assessment for 3D geologic modeling of fault zones based on geologic inputs and prior knowledge".

Reviewer's comments are shown in italicized fonts, and the authors' responses and changes to the manuscript are highlighted with callouts. The general comments are shown and addressed first, then the specific comments are addressed by line/figure number.

1 General comments

In the article "Uncertainty assessment for 3D geologic modeling of fault zones based on geologic inputs and prior knowledge", the authors present an application of a probabilistic geological modelling approach to assess uncertainties in fault zone models.

The work fits into the active field of probabilistic geological modelling and uncertainty assessment in 3-D structural models. The main contribution is the detailed consideration of different fault zone parameters, combined with the evaluation of different types of spherical distributions. In addition, the method is tested in a case study to investigate fault zone uncertainty in a Precambrian crystalline setting with two focus areas: one investigating a single fault zone, and another model of a fault network. Both examples are well chosen to test the application in a realistic setting.

Interesting is specifically the description of sources of uncertainty for different faults zone parameters. Although some of the choices are clearly debatable (e.g. the fault aspect ratio), the authors state the problem of a lack of sufficient information - and this also implicitly highlights the relevance to, at least, consider these sources of uncertainty in a generated geological model, as done here.

The document is supported with code (written in Python and R), including a fully reproducible example for input file generation. The code is hosted on GitHub, and substituted with a license, ensuring the possibility for future use and adaptation. A brief suggestion: it would be good to provide a requirements file and/or more detailed installation instructions (using conda environments or a docker container), as the packages rpy2 and pymc3 are not part of common python distributions. The current version, used to generate results in this manuscript, are furthermore stored in a snapshot on Zeonodo, with a DOI.

One critical aspect in the manuscript is the description of the probabilistic model itself. There are several aspects that need to be adjusted (or clarified) before the manuscript can be considered for publication:

- *The authors seem to mix MC draws from prior distributions (and the representation in models, i.e. the prior predictive models) with a full Bayesian inference. The tool that is employed, pymc, is fully capable of complete Bayesian inference methods, but as I see it (and I looked carefully, even in the code), there is no Bayesian model used here - even if the authors mention "likelihood" (but, I think, mixing terminology here a bit, see comments below) and "MCMC" in line 334. To my best interpretation, this is not the case here - if I am wrong, then please describe more clearly. Otherwise, the entire description on the probabilistic model needs a careful revision.*
- *The statistical parameters of the model are only broadly described, and the reader is referred to the online code. Not every interested reader can be expected to go through a submitted python code to understand and evaluate the scientific details and I strongly suggest to include details about the distributions (including hyperparamters) here in the text. Especially the deterministic transformation for fault depth termination would have been relevant, as it would have made a mistake more obvious: the (as I see it) wrong interpretation of the "tailing behaviour" in figure 6.*
- *The entire description in section 4.5 "Simulation quality assessment" needs to be revised: you can not present draws as traces, if they are not from a (sequential) MCMC chain, i.e.: you do not explore the posterior space and, therefore, you do not draw samples from a posterior distribution (if my interpretation is correct, see above) - the samples you obtain are simply draws from the (independent) prior distributions.*

– *Combining the previous two points: the interpretation of the tailing behaviour is, as I see it, wrong. It is not an effect of a posterior, but an effect of a derived distribution (or, in pymc lingo, a deterministic model combining multiple stochastic variables). This is quite obvious in your code (InputUncertaintyQuantification.py, lines 495ff):*

```
""" Define the uniform or log-normal aspect ratio distribution """
if aspect_logn == 1:
    aspect_dist = pm.Lognormal('Aspect ratio', mu = term_mu, sd = term_sd)
else:
    aspect_dist = pm.Uniform('Aspect ratio',lower=aspect_min,upper=aspect_max)

""" Define the distribution for variability of persistence """
persistence_dist = pm.Normal('Persistence', mu = 0, sd = persistence_std)

""" Deterministic distribution to compute termination depth based on above distributions
zterm_dist = pm.Deterministic('Vertical termination',zo-np.sin(dip*pi/180)*(persistences
```

You do not sample from the posterior distribution, but combine two stochastic variables with a deterministic function. This is a very different concept (and completely fine, but needs adjustment in description).

Authors' answer:

The authors truly appreciate the detailed constructive comments from the referee. The authors have taken several steps to address the misuse of terminology regarding Markov Chain Monte Carlo (MCMC) algorithms and posterior distributions. As the referee noted, there is and was no intention of a Bayesian model being used in the study. The manuscript has been edited throughout to replace erroneous mentions of terms related to MCMC and Bayesian inference (e.g., MCMC sampling, posterior distributions) to proper terminology for the Monte Carlo sampling that was performed in the study.

As for the description of the statistical parameters of the Monte Carlo model, sections of text that were previously withheld for the sake of brevity have been reintroduced to Sections 4.1-4.4 to describe in detail the parameterization of and rationale behind the distributions explored using Monte Carlo sampling.

Changes to the manuscript:

. . . Text added to Section 4.1, Structural orientation: As the dataset from Robinson et al. (1974) lacks specific measurements of fault zone orientations to generate a Bingham distribution directly using the orientation matrix (Fisher et al., 1987), the parameters of the Bingham distribution were assigned manually to generate a distribution covering the expected range of orientations. Varying the parameterization of the Bingham distribution ultimately resulted in determining an appropriate parameterization for the fault zone in the single fault model using an input orientation of 75^o/123^o (dip/dip azimuth), maximum eigenvalue of $\lambda_1 = 200$ (matching the observed azimuth variation of the surface trace) and intermediate eigenvalue of $\lambda_2 = 22.5$ (providing an approximate +/-20^o dip angle variation).

. . . Text added to Section 4.2, Surface trace: First, a bounded uniform distribution is parameterized to simulate a random direction of perturbation for each trace endpoint due to geographical error (i.e., drafting and georeferencing error). The normal distribution representing the total bound on geographical error is converted to respective $\hat{x}$ and $\hat{y}$ components using the directional cosine of the angle sampled from the uniform distribution. This conversion to unit components is used similarly with the fault zone centerline definition uncertainty and digitization uncertainty using the acute angle θ between the orientation of the fault trace with the northing and easting directions. An additional logical check for the strike quadrant of the surface trace is required to implement this approach.

The three individual sources of uncertainty affecting the surface trace endpoint locations are combined into a derived distribution using a deterministic function to determine the total uncertainty affecting the location of each endpoint, given by Eq. 1.

$$P(\hat{x}|\sigma_{centerline},\sigma_{dig},\sigma_{geo},\theta) = cos(\theta)\Big(N(0,\sigma_{centerline}) + N(0,\sigma_{dig})\Big) + N(0,\sigma_{geo})sin\Big(U(0,2\pi)\Big),$$
$$P(\hat{y}|\sigma_{centerline},\sigma_{dig},\sigma_{geo},\theta) = sin(\theta)\Big(N(0,\sigma_{centerline}) + N(0,\sigma_{dig})\Big) + N(0,\sigma_{geo})cos\Big(U(0,2\pi)\Big) \quad (1)$$

The average fault zone thickness was used to characterize the fault zone centerline definition uncertainty affecting each surface trace endpoint. The geographical error was calculated to be approximately 40 meters based on the average distance measured between known landmarks (e.g., mountain tops) on the geologic map and modern satellite imagery data. For both of these sources of uncertainty, the maximum error range described is treated as a 95% confidence interval, allowing a normal distribution to be parameterized with a mean of zero and a standard deviation equal to $maximumerror/3.92$. The digitization error for a 1:12,000 map was represented by a normal distribution with a standard deviation of 3.666 m based on (Zhong-Zhong, 1995).

. . . Text added to Section 4.3, Vertical termination depth: Sampling the uncertainty of the fault zone vertical termination depth involves combining multiple probability distributions using a deterministic function to generate an empirically derived probability distribution. In the derived distribution for fault zone vertical termination depth, f_{length} and $Aspectratio$ are characterized as independent probability distributions and combined using a deterministic function based on the empirically derived description of 3D fault surface geometry, $z_{term} = f_{height} * sin(\theta); f_{height} = \frac{f_{length}}{Aspectratio}$. In this manner, the vertical termination depth (z_{term}) is calculated by converting the fault height to the vertical height using the average dip angle (θ) and subtracting this from the average elevation of the fault outcrop ($z_{outcrop}$).

f_{length} is characterized as a normal distribution with a mean of zero and a standard deviation of 20 meters based on assumed deviance of the surface trace length vs. the true tip-tip fault length. $Aspectratio$ was characterized using both a uniform and a log-normal distribution, respecting the expected maximum and minimum values of 1 and 5. Section 6.2 explores the impact of these two different parameterizations on the shape of the derived distribution for vertical termination depth.

. . . Text added to Section 4.4, Fault zone thickness: For the sake of visualization in the single fault example, a normal distribution characterizing fault zone thickness used a conservative parameterization of $\mu = 30m$ and $\sigma = 2m$.

Relating to the above point regarding erroneous usage of terms from Bayesian inference, Section 4.5 has been reworked extensively to provide an appropriate description of how the study went about assessing the quality of the exploration of the input uncertainty space from Monte Carlo sampling.

The authors agree that the interpretation of the tailing behavior is a result of the use of an empirically derived distribution (previously referred to as a "deterministic distribution"), and not a result of posterior analysis. The wording has been adjusted to clearly state this. However, the authors chose to retain the interpretation of tailing behavior in the vertical termination depth. The authors believe it highlights the possibility of unexpected uncertainty envelope shapes when using empirically derived probability distributions.

The authors have also updated the README file included with the code published on Github to provide a clearer and more comprehensive snapshot of the dependencies required for use of the input uncertainty quantification script.

A note on terminology: I know that "MCUP" has been used in some manuscripts in descent years. However, the concept itself (i.e. the sampling part, Monte Carlo sampling from a distribution and observing the propagation of uncertainty) is not new at all and in widespread use since the mid 20th century. I don't think it is particularly helpful to use a new term for a well established approach, and, on the contrary, will lead to unnecessary confusion for all researchers who are looking at the content and who

are not familiar with the term - as it, at first, pretends to be something novel. I am confident that all appearances of MCUP and easily be reformatted without any loss of information and without making the document significantly longer:

- *I would suggest to refer to the approach itself as "probabilistic modeling" (or "probabilistic geomodeling", when in the clear context of the geomodeling). All references to previous work can be kept, because this is what all the previous works also do.*
- *Instead of "MCUP formulation", simply refer to the setup or definition of the probabilistic model, where stochastic variables are defined by distributions (with corresponding hyperparameters). Adjusting the manuscript in this way will make the content a lot more accessible and comparable to other approaches, also outside the field of geomodeling itself.*

Authors' answer:

The authors agree with the referee's well thought out recommendation regarding the extraneous use of abbreviations when referring to the concept of probabilistic geomodeling through exploration of the geologic model input uncertainty space using Monte Carlo sampling. Changes have been made throughout the paper to refer to the method as 'probabilistic geomodeling', following a proper introduction of what the term means in the context of the study. This includes changing the term "MCUP formulation" to "probabilistic model".

More details on the workflow are also required. The provided python scripts are an important foundation, but they are only capable of producing the input data set for the geological model. A "Leapfrog back-end support" is mentioned in line 244. Is this method also available for other researchers? It is surely the case that the provided input script can be used to generate input data sets for many types of modeling approaches (even though surely not for all, line 259), but it hides a bit the fact that the forward modeling methods (which are generally far more complex) are often implemented in commercial software and often not accessible.

Authors' answer:

The authors acknowledge that the description of the automated workflow implemented in Leapfrog with custom support is sparse, and have supplemented additional text to detail the process of automated model updating. The authors believe that the automated model updating that was implemented for this study is in fact rather straightforward, in the sense that it follows the same series of modeling steps that a user in Leapfrog would follow if they wished to create n realizations of their own fault zone model. The authors believe that the updated text clearly illustrates this concept to the reader.

The method implemented in Leapfrog is available to other researchers on the basis that they contact the developers of Leapfrog (Seequent) independently to acquire access to the unique functionality (which is built on top of a default Leapfrog installation). The product is not currently commercially available and was designed with the supervision of the authors to accomplish the specific goals of the current study. The authors of this study are not developers of Leapfrog, and are therefore not privy to the specific code written in the Leapfrog development environment to accomplish the automated model updating. Rather, the authors worked in collaboration with the developers of Leapfrog to guide them in implementing our own requirements for automated model updating. Communication between the authors and the developers of Leapfrog provided a sufficient level of transparency in how the code was developed, although the specific code cannot be released to the public as it is built directly within the Leapfrog engine.

Changes to the manuscript:

Model realization creation is handled by custom Leapfrog Works back-end support developed for this study to allow for automated updating of geologic modeling parameters from an initial model containing input fault zones. The initial model must be created in Leapfrog using the workflow provided in Figure 1, and naming conventions for the geologic model, fault zones, polylines and termination surfaces being specified in a user-generated text file. The custom version of Leapfrog Works

uses this text file in conjunction with the Leapfrog model and input realization files to automatically step through creation of each geologic model realization based on the simulated data, following the workflow from Figure 1. Model realizations generated in this manner are automatically evaluated onto a grid of cells defined by the text file for subsequent analysis. The method put forward by Wellmann and Regenauer-Lieb (2012) is implemented to calculate the probability of occurrence of fault zone lithology in each cell. The probability of occurrence is then used to compute information entropy to describe the uncertainty of fault zones in the geologic model. In a binary system (e.g., fault zone vs. intact rock), information entropy is maximal when the probability of occurrence of a fault zone is 50%, which as discussed in Krajnovich et al. (2020) can introduce potential for misinterpretation of the geologic model uncertainty envelope if an inappropriate colormap is used. The method implemented in Leapfrog can be made available to other researchers on the basis that they contact the developers of Leapfrog (Seequent) independently to inquire about access to the unique functionality (which is built on top of a default Leapfrog installation).

2 Specific comments

2.1 Line 14: *Note that the presented results are not a sensitivity analysis, but more a (visual) comparison of propagated uncertainties. A sensitivity analysis typically relates to a quantity of interest.*

Authors' answer:

The authors recognize that the term "sensitivity analysis" typically refers to a specific method of analysis performed on a probabilistic model, and the usage has been revised to match the "visual comparison of the contribution of independent input uncertainties to the geologic model uncertainty" performed in the study.

Changes to the manuscript:

The MCUP formulation developed is applied to a simple geologic model built from historically available geologic mapping data ~~to assess the~~ allowing for a visual comparison of the independent ~~sensitivity~~ contributions of each modeling input on the combined model uncertainty, revealing that vertical termination depth and structural orientation uncertainty dominate model uncertainty at depth while surface trace uncertainty dominates model uncertainty near the ground surface.

2.2 Line 23: *Would be suitable to include here the reference Wellmann Caumon (2018) (not fishing for citations here, but as you reference it anyway later...).*

Authors' answer:

The reference to Wellmann and Caumon, 2018 has been added here as the authors agree that it is a fundamental work towards the growing use of 3D geologic models for prediction and communication of subsurface geology. Instead of Line 23, it was added in the reference of Line 22.

Changes to the manuscript:

~~Three-dimensional (3D) geologic models are becoming the state of the art for the prediction and communication of subsurface geology in a wide range of projects (Turner and Gable, 2007) including. . .~~ Three-dimensional (3D) geologic models are becoming the state of the art for the prediction and communication of subsurface geology in a wide range of projects (Turner and Gable, 2007; Wellmann and Caumon, 2018) including. . .

2.3 Line 38: *See comments above for terminology (MCUP)*

Authors' answer:

Changes to the use of the "MCUP" and "MCUP formulation" terminology have been made throughout the paper in line with the general comment addressed above.

Changes to the manuscript:

~~Recently, a Monte-Carlo simulation, input uncertainty propagation method, commonly known as Monte Carlo simulation for uncertainty propagation (MCUP), has grown into a widely used method for 3D geologic model uncertainty assessment (Wellmann and Caumon, 2018).~~ Recently, the well established Monte-Carlo simulation method has been adopted into a widely used method for 3D geologic model uncertainty assessment by way of uncertainty propagation of geologic model inputs into the 3D geologic model space (Wellmann and Caumon, 2018). The method, henceforth termed "probabilistic geomodeling", focuses on the impact of uncertainty in geologic modeling inputs on a 3D geologic model by generating a set of model realizations based on perturbations in selected modeling inputs, sampled using Monte Carlo simulation algorithms. ~~MCUP~~ Probabilistic geomodeling is flexible, allowing for a wide variety of uncertainty sources affecting various geologic modeling inputs to be quantified by the user and propagated into the 3D geologic model.

2.4 Line 63: *I am quite sure that there are studies on uncertainties in fault zone characterization. The topic of uncertainties about the characterization of fault zones has been discussed recently in a paper by Shipton et al. (2019, Fault Fictions: Cognitive biases in the conceptualization of fault zones.). But you probably refer to the treatment in a probabilistic geomodel? Please specify.*

Authors' answer:

The authors acknowledge the lack of clarity when referring to characterizing the uncertainties present in understanding fault zone structure. The section has been revised to specify the application of studied uncertainties about fault zone structure to a probabilistic geomodeling workflow.

Changes to the manuscript:

~~By an in-depth search of the literature, as of yet there is no dedicated approach to characterizing the uncertainty of fault zones in 3D geologic models.~~ Building on the existing literature on understanding the uncertainties about faults in the subsurface (e.g., Choi et al., 2016; Shipton et al., 2019; Torabi et al., 2019a), this study develops a novel, dedicated approach to leveraging probabilistic geomodeling to characterize the uncertainty of fault zones using 3D geologic models.

2.5 Line 67: *Here, for example, no need to refer exclusively to MCUP - this is true for any automated modeling approach.*

Authors' answer:

The reference to complex modeling of fault zone structures has been revised to refer to any automated geologic modeling workflow, rather than only the previously mentioned MCUP approach.

Changes to the manuscript:

~~The inherent complexity of fault zone structure makes their precise modeling intractable in an automated MCUP formulation.~~ The inherent complexity of fault zone structure makes their precise modeling intractable in an automated geologic modeling application, such as that required by probabilistic geomodeling.

2.6 Line 78: ***RBFs are not really a rivaling approach - mostly another form for the kernel, otherwise a lot of similarity. Also, the first paper on RBF's in geomodeling dates back to 2002 (Cowan, on which Leapfrog is based), so also not "so" recent.***

Authors' answer:

The sentence intended to illustrate that RBF based implicit modeling approaches are comparable to the more common co-kriging based methods has been revised to properly respect the literature behind RBF based implicit modeling.

Changes to the manuscript:

~~Use of RBFs in implicit geologic modeling has developed in recent years to rival co-kriging based methods (Hillier et al., 2014, 2017).~~ Implicit geologic modeling using RBFs is comparable in quality to modeling using popular co-kriging approaches (Cowan et al., 2003; Hillier et al., 2014, 2017).

2.7 Line 105: ***Suggestion: avoid "prior knowledge" when referring to likelihood functions, as these are different concepts. "Additional geological knowledge or observations" would make more sense.***

Authors' answer:

The sentence has been revised as suggested to avoid confusion between "prior knowledge" and likelihood functions in the context of a Bayesian model.

Changes to the manuscript:

The second method, operating within a Bayesian inference scheme, is to incorporate ~~the prior knowledge as a likelihood function~~ additional geological knowledge or observations to validate – or rather, as Tarantola (2006) states, invalidate – model realizations.

2.8 Fig. 2 caption: ***here, you actually consider different parameters of the probabilistic model (fault zone thickness, vertical termination, etc.) - not the sources of uncertainty? Please clarify or adjust.***

Authors' answer:

The sentence has been revised to match the figure's content, which is a depiction of possible uncertainty envelopes about the four geologic inputs used in 3D modeling of fault zones.

Changes to the manuscript:

A schematic showing ~~the sources of uncertainty~~ possible uncertainty envelopes about the four geologic modeling inputs used to characterize the 3D geometry of a fault zone in the subsurface. Modified from Krajnovich et al. (2020).

2.9 Line 118: ***You can only accommodate information in a form that can be included as a stochastic variable into the specific interpolation approach (see also treatment in Wellmann Caumon, 2018). It is possible to include additional information in the form of likelihood functions (e.g. de la Varga Wellmann, 2016), but this requires a full integration of modeling and a formulation in a Bayesian framework.***

Authors' answer:

The authors agree with the referee's comment regarding what kind of information can be incorporated into the MCUP formulation. The sentence has been revised to highlight that any additional observations that can be accommodated into the modeling workflow are required to be of the same format as the four inputs considered in the study, i.e., inputs to the implicit modeling interpolation.

Changes to the manuscript:

While formulated for a case with limited data, the developed ~~MCUP formulation~~ probabilistic geomodeling approach allows for accommodating additional observations of the fault zone geologic modeling inputs (e.g., via a modern outcrop study) ~~without significant reparameterization~~ through a straightforward reparameterization of the input probability distributions to include the new data.

2.10 Line 138: *See above about terminology: you do not use MCMC here, so also referring to it here should be removed (even if the statement is true). If you want to mention MC, then simply MC sampling algorithms will do.*

Authors' answer:

The erroneous mention of MCMC sampling algorithms has been removed and edited to specify the use of Monte Carlo sampling algorithms.

Changes to the manuscript:

~~Simulation of scalar data is straightforward and well-established through the use of Markov-chain Monte Carlo (MCMC) sampling algorithms, easily accessible through the open source Python package *PyMC3* (Salvatier et al., 2016).~~ Simulation of scalar data is straightforward and well-established through the use of Monte Carlo sampling algorithms, easily accessible through the open source Python package *PyMC3* (Salvatier et al., 2016).

2.11 Line 312: *I don't understand this sentence here, as the same is true for all other continuous interfaces in the model, which are finally mapped onto a discrete mesh. Can be removed, in my opinion.*

Authors' answer:

The sentence has been revised to clarify the idea that the implementation of fault zone vertical terminations in the developed fault zone modeling approach relies on a series of pre-defined horizontal termination surfaces. Unlike the discretization of the implicit scalar field isovalues onto meshes (which is handled by the Leapfrog modeling engine), this step requires the user to pre-generate termination surfaces at specifically chosen depths in the subsurface prior to performing probabilistic geomodeling steps.

Changes to the manuscript:

Sampled vertical termination depths ~~are continuous, and~~ are subsequently discretized onto the pre-defined termination surfaces ~~in the geologic model~~ created during the initial geologic modeling step (Figure 1(b)). The ~~scale at which~~ interval of these termination surfaces ~~are defined is determined~~ can be adjusted based on the end user needs of the geologic model; 50 meter intervals were used in this study to balance illustrative quality with model processing time.

2.12 Section 4.5: ***the entire section has to be removed or adjusted, as a posterior analysis does not make sense (see above), even if arviz produces nice plots. You can, of course, show samples of the model realizations (left side) and it is also fine to discuss the histogram to discuss the effect of the derived distribution. But any notion of posterior analysis and the plot of the chains are misleading.***

Authors' answer:

The entire Section 4.5 has been revised in depth to remove erroneous terminology of an MCMC sampler and to remove the erroneous use of trace plots in assessing the quality of Monte Carlo simulation for exploring the prior uncertainty space.

Changes to the manuscript:

The quality of probabilistic simulation ~~relies on primarily~~ is a product of the size of the uncertainty space, the simulation method used and the number of samples drawn. For any simulation, the realizations generated can be plotted in the data space and visually examined for appropriate coverage and shape (termed a realization plot). For ~~non-spherical~~ scalar data types, ~~use of MCMC simulation methods creates trace plots and posterior~~ histograms of the Monte Carlo draws provide an intuitive method for independently assessing the quality of simulation for each input. ~~, which provide an intuitive method for independently assessing the quality of simulation for each input. Uniformity of the trace plot and width of the 95% highest posterior density (HPD) indicate, respectively, the convergence of the Markov Chain sampling and the thorough exploration of the posterior uncertainty space.~~ Visual analysis of the shape of the histogram compared to the expected shape of the distribution and a comparison between the input distribution parameters (e.g., mean and standard deviation for a normal distribution) and their values calculated from the samples can quickly determine whether the samples drawn have sufficiently explored the uncertainty space. Figure 1 shows an example of the realization plot~~, trace plots and posterior~~ sample histograms generated for the simulation of vertical termination depths from Section 4.3. This figure allowed for identifying a strong tailing behavior in the output realizations, leading to a reparameterization discussed in Section 6.2.

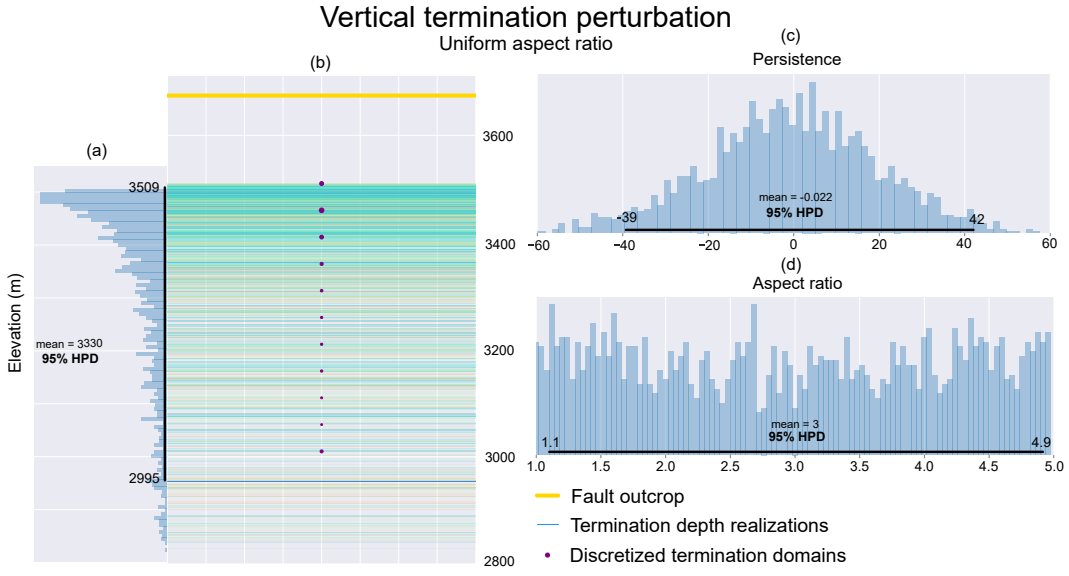

Figure 1. ~~Realizations and Monte Carlo analysis results (trace plot and posterior histograms)~~ Visualization of Monte Carlo samples and associated geologic input realizations from perturbation of the fault zone vertical termination depth based on a uniform distribution of fault aspect ratio. The 95% highest ~~posterior~~ predictive density is overlain on the ~~posterior~~ histograms of the Monte Carlo samples.

~~Trace plots are not available for the spherical data simulations due to reliance on the acceptance-rejection simulation method, while posterior~~ For spherical data simulations, histograms may be replaced by Exponential Kamb contouring (Vollmer, 1995) or Rose diagrams to visualize the density of sampled poles across the surface of the unit sphere (as projected onto a lower-hemisphere projection). This visual assessment provides a semi-quantitative evaluation of the shape and distribution of the ~~posterior spherical probability distribution~~ sampled structural orientations. Additionally, a recalculation of the eigenvector decomposition from the set of simulated samples provides a measure of the accuracy of the posterior distribution with respect to the input ~~orientation~~ parameter values. Tools for generating figures for simulation quality assessment are provided and detailed in the input perturbation script.

Based on the assessment of simulation quality and consideration of compounding factors during uncertainty propagation, the MCUE formulation for the single fault model was run for a number of various realization counts (100, 300, 500, 1,000, 2,000 and 3,000). The processing time generally increases linearly with realization count, reaching many hours to several days for high realization counts on the single fault mock model containing 2.5 million cells. The vast majority of processing time is taken up by the model updating and block model calculation in Leapfrog. For the single fault mock model with 1,000 realizations and 2.5 million cells the sampling benchmark time was 87 seconds while the model processing benchmark time was 38.5 hours. This study is intended to introduce and expand on the use of MCUE formulations for specific geologic modeling problems, and work regarding optimizing the efficiency of model processing is not a focus. The experiments conducted do highlight the need to understand (i) the realization requirement for exploring modeling inputs independently and its relationship to the size of the independent uncertainty spaces, (ii) the interactions of various, related parameters during the uncertainty propagation step and (iii) identification of a balance between final model resolution, coverage, complexity and processing time.

2.13 Line 350: *Even for this high resolution, this seems like a long computation time. Can you comment on the relative timing for sampling (i.e. your script) and the runtime of the model using Leapfrog?*

Authors' answer:

The runtime of the model generation in Leapfrog is approximately three orders of magnitude longer than the timing for sampling (38 hrs vs. 1.5 minutes). The focus of the study was admittedly on the input uncertainty quantification, and efficiency of the automated model updating process was not sought after at this stage. Benchmark statistics on an example run of sampling and model generation have been introduced into the text to clarify this.

Changes to the manuscript:

The processing time generally increases linearly with realization count, reaching many hours to several days for high realization counts on the single fault mock model containing 2.5 million cells. The vast majority of processing time is taken up by the model updating and block model calculation in Leapfrog. For the single fault mock model with 1,000 realizations and 2.5 million cells the sampling benchmark time was 87 seconds while the model processing benchmark time was 38.5 hours.

2.14 Line 388: *It is not a posterior distribution, but a derived distribution.*

Authors' answer:

Correct, and the appropriate revision has been made in line with this specific comment and the earlier general comments.

Changes to the manuscript:

However, the ~~posterior~~ empirically derived distribution of vertical termination depths resulting from a bounded uniform parameterization of fault aspect ratio showed a strong tailing effect (right skewed).

2.15 Line 393: ***Also: typically correlations between parameters are unknown. Effect can (partly) be mitigated through an implementation in a full Bayesian framework with appropriate likelihood functions.***

Authors' answer:

The authors agree with the referee's comments regarding lack of knowledge about the possible correlations between modeling parameters. The authors however believe that the suggested commentary fits naturally into the next Section 6.3 - Parameter relationships. The following sentence has been appended onto the end of Section 6.2 to smoothly transition into that discussion.

Changes to the manuscript:

This reparameterization highlights the key strength – and susceptibility – of MCUP formulations, the reliance on a user defined characterization of input uncertainty. Again, it is necessary to reiterate that the modeler must take into consideration not only field observations and theoretical prior knowledge when assessing a geologic modeling uncertainty formulation, but also their informed expectation of what is geologically realistic for their chosen modeling problem. Furthermore, reparameterizing individual aspects of the probabilistic model may prove to be insufficient due to the presence of inherently unknown relationships between the chosen model parameters.

2.16 Line 405: ***Which artefacts do you refer to? Please highlight or show difference plot, nothing too obvious.***

Authors' answer:

The phrase artefacts has been replaced with the more accurate concept of a skewed uncertainty envelope. The figure in question has also been replaced with a closer view of the uncertainty envelopes, highlighting the skewing (i.e., tendency of realizations towards more vertical dips) suspected due to disagreement between surface trace and structural orientation inputs when sampling orientations from wide, isotropic spherical distributions. A difference plot is also included to highlight the slight skewing of the geologic model uncertainty envelope towards steeper dipping fault zones when parameterized using wide, isotropic spherical distributions.

Changes to the manuscript:

Deviations in the modeled fault surface from the input orientations can occur when the two inputs are significantly different, typically arising in Leapfrog Works by way of overestimation of fault surface dip by up to 10^o when the surface trace azimuth (i.e., average normal to the fault surface trace) and global trend dip azimuth differed by greater than 20^o (Krajnovich et al., 2020). Comparing two geologic models generated with orientations sampled from anisotropic and isotropic Bingham distributions with equivalent maximum uncertainty ranges (Figure 9) showed ~~artefacts~~ a skewing of the ~~present in the~~ geologic model uncertainty envelope when generated from the isotropic Bingham distribution. The results show a consistent skew towards near-vertical dips of the modeled fault zone realizations. This issue is alleviated when the structural orientation is parameterized with an anisotropic Bingham distribution, allowing for increased variability in the dip angle without compromising the certainty of the dip azimuth.

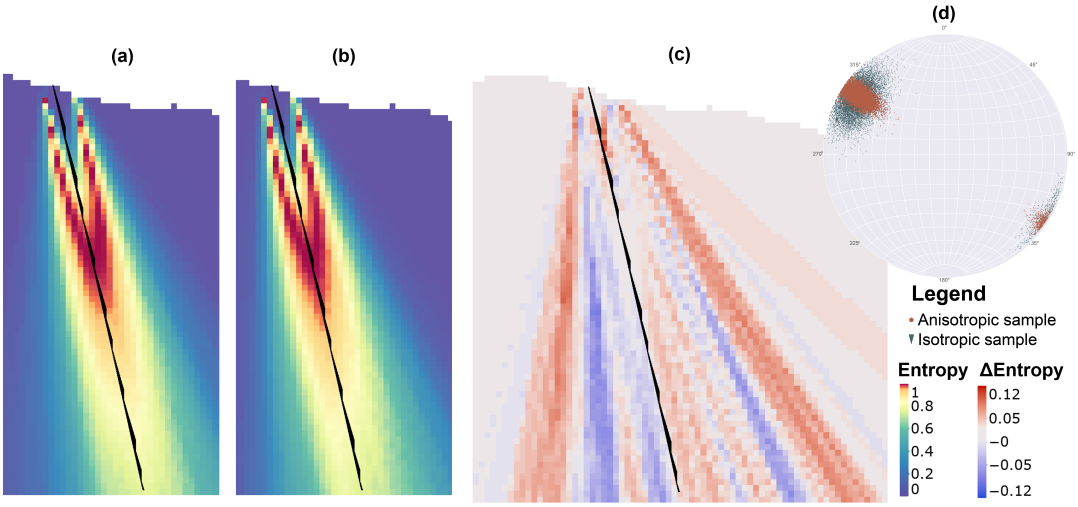

Figure 2. Visual comparison of geologic model entropy generated using anisotropic (a) and isotropic (b) Bingham distributions. The entropy difference between these two models (c) highlights slight skewing of the uncertainty envelope towards steeper dipping fault zones when parameterized using samples from an isotropic Bingham distribution (d). Entropy difference is shown such that negative values indicate higher entropy in the isotropic orientation model.

2.17 Line 420: *Gibbs sampling is a variant of an MCMC approach, again: should not be referred to here. Do you simply mean appropriate sampling schemes from a joint distribution? Same difficulty concerning correlations as in line 393.*

Authors' answer:

The erroneous reference to Gibbs sampling has been removed, and the correction that the referee identified has been added to the section.

Changes to the manuscript:

A treatment of these relationships through ~~the use of Gibbs sampling or other conditional methods of Monte Carlo simulation~~ parameterizing previously independent input probability distributions using a joint distribution (and an appropriate sampling scheme) could potentially generate more realistic and efficient assessments of model uncertainty. While the correlation between parameters in a probabilistic model is typically unknown, in some cases – such as the correlation between fault trace azimuth and structural orientation azimuth – the correlation can be assumed based on available prior geologic knowledge of the parameter's real-world relationships.

2.18 Line 440: *Authors, I assume?*

Authors' answer:

Correct.

Changes to the manuscript:

The author~~s~~ acknowledge~~s~~ that in reality fault zone geometry includes horizontal terminations.

2.19 Line 472: ***The difficulty here: you would need to implement the forward model completely into the modeling framework (if information on the basis of the generated model should be considered). Is this possible with the approach used here (running Leapfrog as an integrated part of the Python script)? Please clarify.***

Authors' answer:

The authors appreciate the referee's useful insights into the future of this work. Full integration of the probabilistic model (i.e., the "MCUP formulation") into the geologic modeling framework (i.e., Leapfrog) – or vise versa – is currently not possible. Without entering into too much detail on yet unfinished work, the authors believe that Bayesian inference can be leveraged to implement additional information in the form of subsurface observations of fault zones. The anticipated Bayesian model formulation would use information from the preliminary probabilistic geomodel as the prior uncertainty, validated against new subsurface observations by way of a misfit-style likelihood function, building on the approach published by Schneeberger et al. (2017). As this future work is only in a preliminary stage of planning at the time of this writing, the authors believe its detailed treatment lies outside of the scope of the current publication.

The sentence has been revised to introduce a clearer stance on the possible uses of Bayesian inference following the probabilistic geomodeling laid out in this study.

Changes to the manuscript:

Future work stemming from this ~~preliminary modeling~~ input-based probabilistic geomodeling formulation may include incorporating new information in a Bayesian inference scheme to further refine the geologic model, either following the methodology introduced by de la Varga and Wellmann (2016) to infer additional information about the model parameters (requiring a full integration of the probabilistic modeling with the automated geologic modeling approach) or by Schneeberger et al. (2017) to validate the initial model in light of its generated uncertainty.

3 Figures

overall good quality, but labels are generally far too small. Block diagrams also need orientation (at least North arrow) and axes labels should also be included.

Authors' answer:

The necessary changes to figures throughout the text have been made as suggested by the referee.

References

Cowan, M. W., Beatson, R. K., Ross, H. J., Fright, W. R., McLennan, T. J., Evans, T. R., Carr, J. C., Lane, R. G., Bright, D., Gillman, A., Oshust, P., and Titley, M.: Practical Implicit Geological Modelling, 2003.

de la Varga, M. and Wellmann, J. F.: Structural geologic modeling as an inference problem: A Bayesian perspective, Interpretation, 4, SM1–SM16, https://doi.org/10.1190/INT-2015-0188.1, 2016.

Fisher, N., Lewis, T., and Embleton, B.: Statistical analysis of spherical data, Cambridge University Press, 1987.

Hillier, M. J., Schetselaar, E. M., de Kemp, E. A., and Perron, G.: Three-Dimensional Modelling of Geological Surfaces Using Generalized Interpolation with Radial Basis Functions, Mathematical Geosciences, 46, 931–953, https://doi.org/10.1007/s11004-014-9540-3, 2014.

Hillier, M. J., Kemp, E. A. D., Schetselaar, E. M., Hillier, M. J., Kemp, E. A. D., and Schetselaar, E. M.: Implicit 3-D modelling of geological surfaces with the Generalized Radial Basis Functions (GRBF) algorithm, Geological Survey of Canada, Open File 7814, pp. 1–15, 2017.

Krajnovich, A., Zhou, W., and Gutierrez, M.: Uncertainty Assessment of Fault Zones in 3D Geologic Models of Mountain Tunnels (in press), in: Word Tunnel Congress, International Tunneling and Underground Space Association, 2020.

Robinson, C. S., Lee, F. T., Scott, J. H., Carroll, R. D., Hurr, R. T., Richards, D. B., Mattei, F. A., Hartmann, B. E., and Abel, J. F.: Engineering Geologic, Geophysical, Hydrologic, and Rock-Mechanics Investigations of the Straight Creek Tunnel Site and Pilot Bore, Colorado, Tech. rep., United States Geological Survey, Washington, D.C., https://doi.org/10.3133/pp815, 1974.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C.: Probabilistic programming in Python using PyMC3, PeerJ Computer Science, 2, 1–24, https://doi.org/10.7717/peerj-cs.55, 2016.

Schneeberger, R., de la Varga, M., Egli, D., Berger, A., Kober, F., Wellmann, F., and Herwegh, M.: Methods and uncertainty-estimations of 3D structural modelling in crystalline rocks: A case study, Solid Earth, 8, 987–1002, https://doi.org/10.5194/se-2017-47, 2017.

Vollmer, F. W.: C program for automatic contouring of spherical orientation data using a modified Kamb method, 21, 31–49, https://doi.org/10.1016/0098-3004(94)00058-3, 1995.

Wellmann, J. F. and Regenauer-Lieb, K.: Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models, Tectonophysics, 526-529, 207–216, https://doi.org/10.1016/j.tecto.2011.05.001, 2012.

Zhong-Zhong, C.: The estimation of digitizing error and its propagation results in GIS and application to habitat mapping, Ph.D. thesis, University of Massachusetts Amherst, https://scholarworks.umass.edu/dissertations/AAI9524686, 1995.

Uncertainty assessment for 3D geologic modeling of fault zones based on geologic inputs and prior knowledge

Ashton Krajnovich[1], Wendy Zhou[1], and Marte Gutierrez[2]

[1]Department of Geology and Geological Engineering, Colorado School of Mines, 1516 Illinois St, Golden, CO, 80401, USA
[2]Department of Civil and Environmental Engineering, Colorado School of Mines, 1012 14th St, Golden, CO, 80401, USA

Correspondence: Ashton Krajnovich (akrajnov@mines.edu)

Abstract. Characterizing the zone of damaged and altered rock surrounding a fault surface is highly relevant to geotechnical and ~~reservoir~~ geo-environmental engineering works in the subsurface. Evaluating the uncertainty associated with 3D geologic modeling of these fault zones is made possible using the popular and flexible input-based uncertainty propagation approach to geologic model uncertainty assessment – termed ~~Monte Carlo Simulation for Uncertainty Propagation (MCUP)~~probabilistic geomodeling. To satisfy the automation requirements of ~~MCUP~~ probabilistic geomodeling while still preserving the key geometry of fault zones in the subsurface, a clear and straightforward modeling approach is developed based on four geologic inputs used in implicit geologic modeling algorithms (surface trace, structural orientation, vertical termination depth and fault zone thickness). The rationale applied to identifying and characterizing the various sources of uncertainty affecting each input are explored and provided using open-source codes. In considering these sources of uncertainty, a novel model formulation is implemented using prior geologic knowledge (i.e., empirical and theoretical relationships) to parameterize modeling inputs which are typically subjectively interpreted by the modeler (e.g., vertical termination depth of fault zones). Additionally, the application of anisotropic spherical distributions to modeling disparate levels of information available regarding a fault zone's dip azimuth and dip angle is demonstrated, providing improved control over the structural orientation uncertainty envelope. The ~~MCUP formulation~~ probabilistic geomodeling approach developed is applied to a simple fault zone geologic model built from historically available geologic mapping data ~~to assess the independent sensitivity~~ allowing for a visual comparison of the independent contributions of each modeling input on the combined model uncertainty, revealing that vertical termination depth and structural orientation uncertainty dominate model uncertainty at depth while surface trace uncertainty dominates model uncertainty near the ground surface. The method is also successfully applied to a more complex fault network model containing intersecting major and minor fault zones. The impacts of the model parameterization choices, the fault zone modeling approach and the effects of fault zone interactions on the final geologic model uncertainty assessment are discussed.

1 Introduction

Three-dimensional (3D) geologic models are becoming the state of the art for the prediction and communication of subsurface geology in a wide range of projects ~~(Turner and Gable, 2007)~~ (Turner and Gable, 2007; Wellmann and Caumon, 2018) including regional geologic characterization (Stafleu et al., 2012; Waters et al., 2015); natural resource exploration ~~(Zhou et al., 2007; Zhou, 2009;~~

(Zhou et al., 2007; Zhou, 2009; Anderson et al., 2014; Zhou et al., 2015); structural geology ~~(Bond et al., 2015)~~(Bond et al., 2015; Aillères ; geotechnical site characterization (Thum and De Paoli, 2015; Zhu et al., 2013); geophysics ~~(Guillen et al., 2008; Høyer et al., 2015)~~ (Guillen et al., 2008; Høyer et al., 2015; Anderson et al., 2014); hydrology (Watson et al., 2015); and mining (Wellmann et al., 2018; Yang et al., 2019). The recent widespread adoption of flexible, implicit 3D geologic modeling algorithms (Cowan et al., 2003; Calcagno et al., 2008; Guillen et al., 2008; Jessell et al., 2014; Hillier et al., 2014, 2017) is leading the field of 3D geologic modeling away from the creation of static models based on a single, best interpretation and towards stochastic geologic modeling with quantified uncertainty (Caumon, 2010). Understanding the uncertainty of a 3D geologic model provides not only a measure of model quality to an end user (Turner and Gable, 2007; Walker et al., 2003; Stamm et al., 2019), but also aids the geologist during model creation by analyzing the quality of input data and highlighting the impacts of subjective prior knowledge and interpretations (Bond, 2015; Wood and Curtis, 2004; Jessell et al., 2018). As the use of 3D geologic modeling continues to grow, novel methods for assessing the uncertainty of various aspects of geologic models is pertinent.

Because a single model conveys no information regarding its uncertainties (Wellmann and Caumon, 2018), multiple realization approaches are becoming a popular method for assessing geologic model uncertainty (Wellmann et al., 2010; Wellmann and Regenauer-Lieb, 2012; Pakyuz-Charrier et al., 2018a, b; Pakyuz-charrier et al., 2019; Jessell et al., 2014, 2018; Lindsay et al., 2013; de la Varga and Wellmann, 2016; de la Varga et al., 2019; Schweizer et al., 2017; Thiele et al., 2016; Yang et al., 2019; Schneeberger et al., 2017). Recently, ~~a~~ the well established Monte-Carlo simulation ~~, input uncertainty propagation method , commonly known as Monte Carlo simulation for Uncertainty Propagation (MCUP) (Pakyuz-Charrier et al., 2018a, b), has grown~~ method has been adopted into a widely used method for 3D geologic model uncertainty assessment ~~(Wellmann and Caumon, 201~~ by way of uncertainty propagation of geologic model inputs into the 3D geologic model space (Wellmann and Caumon, 2018). The method, henceforth termed "probabilistic geomodeling", focuses on the impact of uncertainty in geologic modeling inputs on a 3D geologic model by generating a set of model realizations based on perturbations in selected modeling inputs, sampled using Monte Carlo simulation algorithms. ~~MCUP~~ Probabilistic geomodeling is flexible, allowing for a wide variety of uncertainty sources affecting various geologic modeling inputs to be quantified by the user and propagated into the 3D geologic model.

While growing in popularity, the field of geologic model uncertainty assessment remains a developing one and the application of ~~MCUP~~ probabilistic geomodeling to new, practical problems requires unique model formulations. The development of novel ~~MCUP formulations~~ probabilistic geomodeling approaches to address specific aspects of 3D geologic modeling will lead to growth in the field not only by broadening the usability of the method, but also by advancing the understanding of the method's strengths and limitations. In addition to assessing the uncertainty of a single geologic model, ~~MCUP~~ probabilistic geomodeling using Monte Carlo sampling naturally fits into Bayesian inference schemes (de la Varga and Wellmann, 2016; Salvatier et al., 2016; Scalzo et al., 2019; Thiele et al., 2019), allowing for future refinement of model uncertainty as new information is made available.

2 ~~Model implementation~~

This study expands the use of ~~MCUP~~ probabilistic geomodeling to a new aspect of geologic modeling – fault zones, or the localized volume of fractured and displaced rock surrounding a finite fault surface, typically composed of a fault core and a damage zone (Caine et al., 1996; Childs et al., 2009; Peacock et al., 2016; Choi et al., 2016). Fault zones introduce regions of altered geotechnical strength and hydraulic permeability into the surrounding in-tact rockmass and are therefore of major importance to geological engineering projects that rely on accurate assessments of subsurface rock properties (e.g., tunnels, mines). While faults have been the focus of a significant amount of recent geologic modeling research (Røe et al., 2014; Cherpeau et al., 2010; Cherpeau and Caumon, 2015; Aydin and Caers, 2017), these works have focused on modeling fault surfaces directly rather than modeling the 3D geometry of fault zones. Detailed modeling of the 3D geometry of fault zones can improve the understanding of faults' impacts on geotechnical and reservoir engineering projects due to the fact that variations in fault zone thickness or composition can greatly alter the mechanical and hydrological behavior of a fault, e.g., its sealing potential (Caine et al., 1996; Fredman et al., 2008; Manzocchi et al., 2010). ~~By an in-depth search of the literature, as of yet there is no~~ Building on the existing literature on understanding the uncertainties about faults in the subsurface (Choi et al., 2016; Shipton et al., 2019; Torabi et al., 2019b), this study develops a novel, dedicated approach to ~~characterizing~~ leveraging probabilistic geomodeling to characterize the uncertainty of fault zones ~~in~~ using 3D geologic models.

Fault zones may be irregular in shape, creating complex geometries which are difficult to characterize quantitatively (Torabi et al., 2019a, b). Peacock et al. (2016) provide a detailed list of the various types of damage zones and intersecting fault networks that comprise the general term "fault zone". The inherent complexity of fault zone structure makes their precise modeling intractable in an automated ~~MCUP formulation~~geologic modeling application, such as that required by probabilistic geomodeling. A simplified approach to modeling fault zones in 3D geologic models is developed in this study based on the key elements defining fault zone geometry at a practical level of detail.

2 Model implementation

The proposed workflow for modeling fault zones is provided in Figure 1. The workflow combines observations from a geologic map with prior knowledge from the literature to approximate the 3D geometry of subsurface fault zones. The implicit geologic modeling software Leapfrog Works, a software specially designed to support creation of subsurface geological models, is used in this study. The 3D fault zone is modeled in Leapfrog Works from four inputs – surface trace (polyline), structural orientation (dip/dip azimuth vector), fault zone thickness (scalar distance function) and vertical termination depth (discretized surfaces). The modeling approach preserves the essential 3D geometry of fault zones in the subsurface while providing sufficient generalization to fit into an ~~MCUP formulation for uncertainty assessment~~automated implicit modeling workflow for uncertainty propagation.

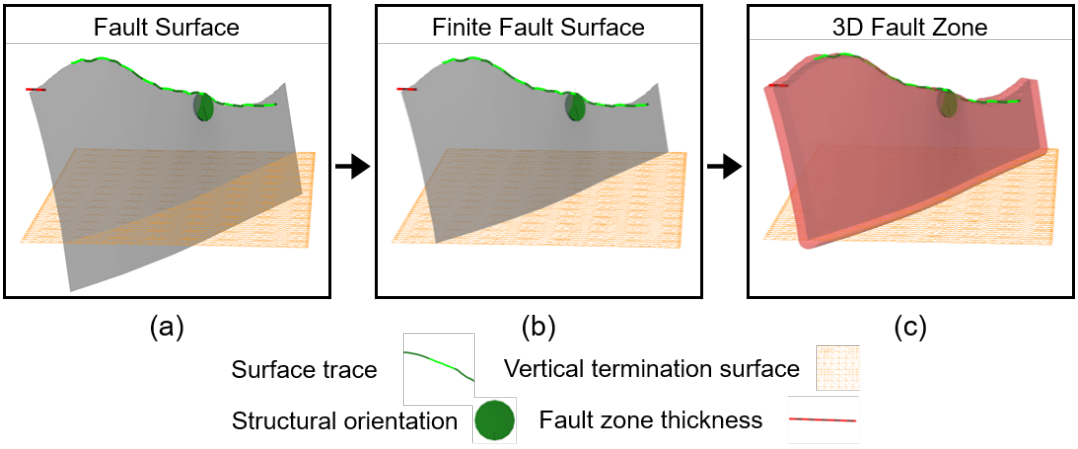

Figure 1. The proposed fault zone modeling workflow implemented includes (a) modeling the central fault surface from a polyline and structural orientation, (b) terminating the fault surface on a predefined vertical termination surface and (c) defining the 3D fault zone volume using a distance function from the central fault surface (Krajnovich et al., 2020a).

The popular implicit geologic modeling method was chosen for modeling the uncertainty of fault zones due to its ability to directly incorporate structural orientation data to modeling geologic structures. Leapfrog Works uses radial basis functions (RBFs) to efficiently interpolate the scalar fields describing implicit geologic surfaces (Seequent, 2014). ~~Use of RBFs in implicit geologic modeling has developed in recent years to rival~~ Implicit geologic modeling using RBFs is comparable in quality to modeling using popular co-kriging ~~based methods (Hillier et al., 2014, 2017)~~approaches (Cowan et al., 2003; Hillier et al., 2014, 2017) .

2.1 ~~MCUP formulation~~ Probabilistic geomodel setup

~~The MCUP formulation~~ Setting up the probabilistic geomodel begins with the careful selection of key geologic modeling inputs for perturbation (Figure 1). The set of geologic modeling inputs are characterized using probability distributions chosen and parameterized based on the believed and/or observed uncertainties in the input variables. Monte Carlo simulations independently explore the uncertainty space of each input, generating a set of input realizations which are propagated into 3D geologic models through the use of an automated implicit geologic modeling algorithm. From the set of geologic model realizations, the commonly used Shannon information entropy metric (Shannon, 1948) allows for quantifying the uncertainty about modeled geologic structures (Wellmann and Regenauer-Lieb, 2012). This overview is provided merely to introduce the reader to the general idea of ~~MCUP~~ probabilistic geomodeling as applied to implicit geologic modeling; the reader is referred to the book by Wellmann and Caumon (2018) for a more thorough review of the ~~MCUP~~ probabilistic geomodeling and implicit modeling methods' conceptual bases.

The same flexibility that allows ~~MCUP~~ probabilistic geomodeling to be effectively formulated for nearly any geologic modeling problem is also a potential susceptibility ~~of MCUP – that~~ the model formulation and input uncertainties must be

predefined by the user. This can lead to potential over- or under-estimation and biases in the uncertainty assessment performed with ~~MCUP~~ probabilistic geomodeling due to inappropriate selection of modeling inputs or incorrect parameterization of input probability distributions (de la Varga and Wellmann, 2016; Wellmann and Caumon, 2018; Pakyuz-Charrier et al., 2018b). Following the school of thought reviewed by Nearing et al. (2016), the modeler must ask questions along the lines of:

- What inputs control the geometry of the modeled structure?
- How can these inputs be defined probabilistically?
- What information is available to characterize each source of input uncertainty?

For hard data inputs (e.g., observed contacts, structural orientation measurements), the answers to these questions are relatively well-established using measures of variance and deviation to directly characterize uncertainties (Caers, 2011; Wellmann and Caumon, 2018). On the other hand, for subjective inputs to geologic models (e.g., interpreted fault terminations), there are generally two approaches with which the subjective uncertainties can be characterized. The first is describing and quantifying the uncertainty associated with prior knowledge by utilizing believed theoretical or empirical relationships to quantify the reasoning behind a subjective geologic interpretation (Wood and Curtis, 2004). The second method, operating within a Bayesian inference scheme, is to incorporate ~~the prior knowledge as a likelihood function to validate –~~ additional geological knowledge or observations to validate – or rather, as ~~Tarantola (2006)~~ Tarantola (2006) states, invalidate ~~–~~ – model realizations. This study focuses on the first method of applying prior knowledge from published structural geology literature (Torabi et al., 2019a) to parameterize the reasoning behind subjective inputs used for ~~modeling fault zonesin the MCUP formulation~~probabilistic geomodeling of fault zones. This approach is demonstrated effectively in this study when considering the vertical termination depth of fault surfaces (Section 4.3).

Having identified the geologic modeling inputs controlling the geometry of the modeled fault zone and appropriate methods for defining these inputs probabilistically, Figure 2 shows ~~the careful consideration of the various~~ an illustration depicting possible uncertainty envelopes due to key sources of uncertainty affecting each input.

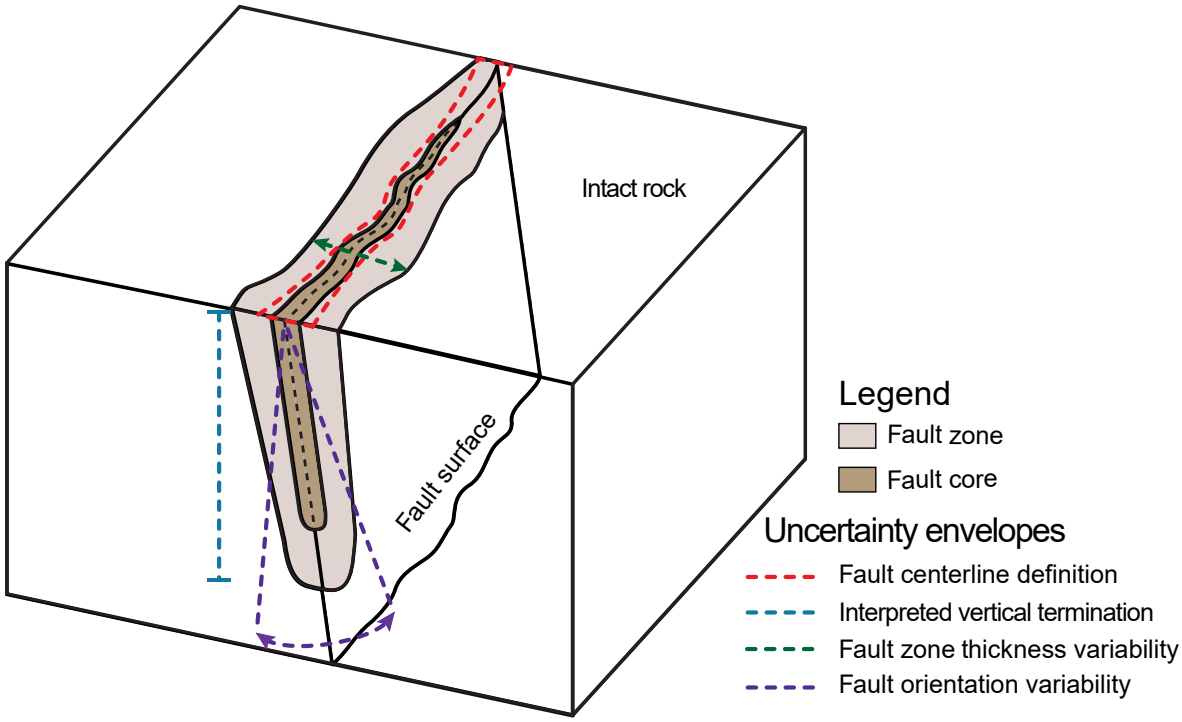

Figure 2. A schematic showing expected uncertainty envelopes due to the specified sources of uncertainty ~~influencing~~ regarding characterizing the 3D geometry of a fault zone in the subsurface. Modified from Krajnovich et al. (2020a).

In the 3D geologic model, the geometry and extent of fault zones in the subsurface often are approximated from limited information available. While useful for conceptualizing subsurface geology and guiding future investigations, the lack of detailed data contributes significant sources of uncertainty to the 3D geologic model (Figure 2). This study focuses on and provides guidelines for performing realistic uncertainty assessments when creating 3D geologic models of fault zones from limited, preliminary investigation data (e.g., a geologic map), further demonstrating how prior knowledge is used to characterize uncertainty about inputs which are typically subjective (i.e., vertical termination depth). While formulated for a case with limited data, the developed ~~MCUP formulation~~ probabilistic geomodeling approach allows for accommodating additional observations of the fault zone geologic modeling inputs (e.g., via a modern outcrop study) ~~without significant reparameterization~~ through a straightforward reparameterization of the input probability distributions to include the new data.

Moving forward, Section 3 details the appropriate selection and parameterization of probability distributions to accurately characterize input uncertainties for various the data types used in 3D geologic modeling of fault zones. Section 4 highlights the considerations for characterizing each modeling input's uncertainty, and references the Python code written for performing and assessing the input perturbations on a single fault zone model built from a geologic map in the Rocky Mountains of Colorado, USA. Section 5 demonstrates the ~~MCUP formulation~~ probabilistic geomodeling approach applied to a more complex, fault network model to assess how the method scales and investigate the interaction of intersecting fault zones. Section 6 dis-

cusses the observed results of each model, highlighting key contributions of the ~~MCUP formulation~~ probabilistic geomodeling approach including the impact of different uncertainty parameterizations and guidelines for future model refinement. Section 7 reviews the ~~MCUP~~ probabilistic geomodeling implementation for fault ~~zone modeling~~ zones and reiterates the importance of the rationale used, concluding with recommendations for future work.

3 Probability distributions ~~for MCUP~~

The ~~MCUP formulation~~ probabilistic geomodeling approach developed requires additional advancements in the selection and parameterization of probability distributions used for characterizing the uncertainty of objective and subjective geologic modeling inputs. The selection of an appropriate probability distribution type from the various distributions available for modeling involves the consideration of two factors: the data type and the level of knowledge about the input. Geologic data may be discrete (e.g., lithological categories) or continuous (e.g., thickness), with various probability distributions available to characterize both data types (e.g., Normal, Uniform, Log-normal, Binomial – see Gelman et al. (2004)). All of the modeling inputs used for modeling the 3D geometry of fault zones are described using continuous data types.

~~Simulation of scalar data is straightforward and well-established through the use of Markov-chain Monte Carlo (MCMC) sampling algorithms, easily accessible through the open source Python package *PyMC3* (Salvatier et al., 2016). The *PyMC3* library has been demonstrated as a platform for performing MCUP of 3D geologic models (de la Varga and Wellmann, 2016; Schneeberger , and has even been implemented in the open source geologic modeling platform GemPy (de la Varga et al., 2019).~~ An additional consideration in the case of continuous data types is the distinction between scalar and vectorial data (e.g., structural orientations). A probability distribution describing orientation data resides on the surface of a unit-sphere in 3D, and can be characterized using spherical probability distributions (Fisher et al., 1987; Mardia and Jupp, 2000). The benefit of using spherical probability distributions to describe structural orientation uncertainty in 3D geologic modeling is clearly stated by Pakyuz-Charrier et al. (2018b), and their application in ~~MCUP formulations~~ probabilistic geomodeling continues to develop (Pakyuz-Charrier et al., 2018b, a; Carmichael and Ailleres, 2016). To remain concise, the following section focuses on the new contributions made to the use of spherical probability distributions utilizing the *R-fast* open source package available in the R language (Papadakis et al., 2018).

3.1 Spherical probability distributions

Fault orientations are vectors described by dip and dip azimuth components. Stereographic projection is often used to describe the fault plane using its pole (i.e., normal), defined by a unit vector or a trend and plunge. For handling orientation data in ~~MCUP formulations~~the probabilistic geomodel, distributions of the Fisher-Bingham family (Bingham, 1964; Kent, 1982) provide a wide variety of choices for modeling varying degrees of uncertainty. Pakyuz-Charrier et al. (2018b) recently showed that scalar distributions are inadequate for modeling the uncertainty of structural orientation data, providing an example using the von-Mises Fisher (vMF) distribution (spherical analogue to the isotropic bivariate normal distribution). This research continues the exploration into the use of spherical distributions in ~~MCUP~~ probabilistic geomodeling by implementing the more general

Bingham and Kent distributions (Fisher et al., 1987) from the Fisher-Bingham family to characterize anisotropic uncertainty of structural orientations used in 3D geologic modeling.

In structural geology, the use of spherical distributions to understand the uncertainty about structural orientation measurements is well established (Mardia, 1981; Cheeney, 1983; Davis and Titus, 2017; Roberts et al., 2019). The open source Orient and Stereonet softwares by Vollmer (2018) and Allmendinger (2015) provide uncertainty estimates of structural data using spherical statistics, while Davis and Titus (2017) and Roberts et al. (2019) used spherical distributions to estimate confidence intervals about uncertain structural orientation measurements of folds and foliations. The analysis of anisotropic orientation uncertainty in structural geology is well established with Zhou and Maerz (2002), Peel et al. (2001), Carmichael and Ailleres (2016) and Davis and Titus (2017) applying it to joint set identification, structural data clustering and foliation-lineation characterization. While Pakyuz-Charrier et al. (2018a) showed clearly that the dip angle and dip azimuth should not be simulated independently as scalar values, it is apparent that the uncertainty affecting each of these aspects of the structural orientation need not be equal.

Two spherical distributions were reviewed in this study: the Bingham and Kent distributions. As opposed to the isotropic vMF distribution, the Bingham and Kent distributions are capable of modeling potentially more realistic anisotropic uncertainty envelopes. This study explores a novel application of anisotropic spherical distributions in ~~MCUP~~probabilistic geomodeling: characterizing subjective bias in the structural orientation uncertainty of fault zones (i.e., differing levels of information regarding the dip angle and dip azimuth of faults modeled in implicit 3D geologic models). The distributions, typically characterized from a series of input measurements, can also be characterized by directly controlling the input parameters themselves. This method allows for the modeller to assume the size and shape of the structural orientation uncertainty envelope in the ~~MCUP formulation~~probabilistic geomodel setup.

The Bingham distribution is an antipodally symmetric distribution for axial data, defined explicitly in R^3 ($p = 3$) by a set of orthogonal eigenvectors ($\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}$) and corresponding eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3 = 0$). The eigenvector and eigenvalue pairs respectively detail the direction and degree of maximum, intermediate and minimum variance of the Bingham distribution. The distribution is described by Eq. 1 where $A = diag(\lambda_1, \lambda_2, \lambda_3)$ and $c(A)$ is the corresponding normalization constant (Fallaize and Kypraios, 2016). Setting $\lambda_3 = 0$ merely ensures that the distribution shows maximum variance in the axial direction (i.e., across the unit sphere), allowing λ_1 and λ_2 to fully control the shape of the distribution when projected onto the lower hemisphere.

$$P(\boldsymbol{x}|A) = \exp\left(-\sum_{i=1}^{p-1} \lambda_i x_i^2\right) \frac{1}{c(A)} \qquad c(A) = \int_{\boldsymbol{x} \in S^{p-1}} \exp\left(-\sum_{i=1}^{p-1} \lambda_i x_i^2\right) dS^{p-1}(\boldsymbol{x}) \tag{1}$$

The Kent distribution (or Fisher-Bingham 5-parameter distribution) is a more generalized form of the Bingham distribution, defined for vectorial data focused around a known mean vector with an anisotropic uncertainty envelope. It is characterized by a mean vector ($\boldsymbol{x} = x_1, x_2, x_3$), the concentration parameter κ and an ovalness parameter β. Its density is described by Eq. (2) for $\kappa \geq 0, \beta \geq 0$ where Ω is an orthogonal $p \times p$ matrix that can be likened to the eigenvector matrix used in simulating the Bingham distribution. The reader is referred to Appendix C of Pakyuz-Charrier et al. (2018a) for a more thorough explanation

of the parameterization of the Kent distribution.

$$P(\boldsymbol{x}|\Omega,\kappa\beta) = C_3(\kappa,\beta)\exp\left(\kappa\omega_1^T\boldsymbol{x} + \beta\left[(\omega_2^T\boldsymbol{x})^2 - (\omega_3^T\boldsymbol{x})^2\right]\right) \quad C_3(\kappa,\beta) = 2\pi\sum_j^{\infty}\frac{\Gamma\left(j+\frac{1}{2}\right)}{\Gamma(j+1)}\beta^{2j\left(\frac{\kappa}{2}\right)-2j-\frac{1}{2}}I_{2j+\frac{1}{2}}(\kappa) \tag{2}$$

When modeling structural geologic orientation data, the distinction between axial (i.e., undirected) and vectorial (i.e., directed) data is irrelevant following the application of lower-hemisphere stereographic projection. The stereographic projection is applicable because in reality the orientation of geologic structures is defined by an axis laying within the plane of the structure. Regardless of the method of characterizing and sampling the orientation data, a stereographic projection to the lower-hemisphere will return the conventional down-dip orientations as defined by dip/dip-azimuth or right-hand-rule systems.

3.2 ~~Simulation~~ Sampling

Sampling of scalar data is straightforward and well-established through the use of Monte Carlo sampling algorithms, easily accessible through the open source Python package *PyMC3* (Salvatier et al., 2016). The *PyMC3* library is designed to facilitate Bayesian inference using computational sampling algorithms, though the inclusion of likelihood functions is not required, thereby allowing for utilization of the package functions for Monte Carlo sampling alone. The use of *PyMC3* has been demonstrated successfully in the context of 3D geologic modeling by de la Varga and Wellmann (2016) and Schneeberger et al. (2017), and its implementation in *Theano* has allowed for seamless integration with the open source geologic modeling platform GemPy (de la Varga et al., 2019). This study focuses solely on the step of probabilistic geomodeling based on 3D geologic modeling inputs, leveraging only the Monte Carlo sampling capabilities of *PyMC3*.

Simulating samples from spherical distributions requires dedicated algorithms separate from those used for scalar data types, due to the transformation between rectangular and spherical coordinates creating non-uniform areas of angular trend and plunge increments on the unit-sphere. Several solutions have been demonstrated to simulate random samples from spherical distributions which are comprehensively documented in Kent et al. (2018). Simulation algorithms for distributions of the Fisher-Bingham family have been implemented by Papadakis et al. (2018) in an open-source R package *Rfast*. Open-source tools for statistical simulation in the R and Python environments (including their combined usage through the *rypy2* package), provide convenient, well-documented tools for applying established statistical techniques to novel fields in geoscience.

The algorithm for simulating random points from the Bingham and Kent distributions included in *Rfast* uses the acceptance-rejection method, inspired by Kent et al. (2013) and Fallaize and Kypraios (2016). The method uses a Central Angular Gaussian (CAG) distribution as an envelope to approximate the Bingham distribution. The algorithm uses only the first two eigenvalues for identifiability, resulting in the need to use a rotation to align the sampled points with the desired orientation. The rotation of data sampled from spherical distributions to any new set of axes is possible due to the rotation-independence of the dispersion of spherical distributions.

3.3 ~~Rotation~~

Two rotations using the Euler-Rodrigues formula (Dai, 2015) are useful for properly aligning the data simulated using the *Rfast* algorithms. The first, necessary for the Bingham distribution, interchanges two axes by a rotation of π about an axis defined by the sum of the two axes to be interchanged, $\boldsymbol{k} = \boldsymbol{v_1} + \boldsymbol{v_2}$. The second rotation is necessary with either the Bingham or Kent distribution to correct for the arbitrary alignment of the orientation uncertainty envelope (Figure 3(a)), which is not desired when characterizing anisotropic uncertainty of dip angle and dip azimuth of a fault. This occurs in the ~~MCUP formulation~~ probabilistic geomodel setup when characterizing the distributions from input parameters directly rather than through eigen-decomposition or maximum likelihood estimation. An effective approach to rotating anisotropic spherical distributions was developed which rotates iteratively by small increments about the input orientation vector until an accuracy threshold on the desired alignment of the $\boldsymbol{e_3}$ plane is satisfied (Figure 3(b)).

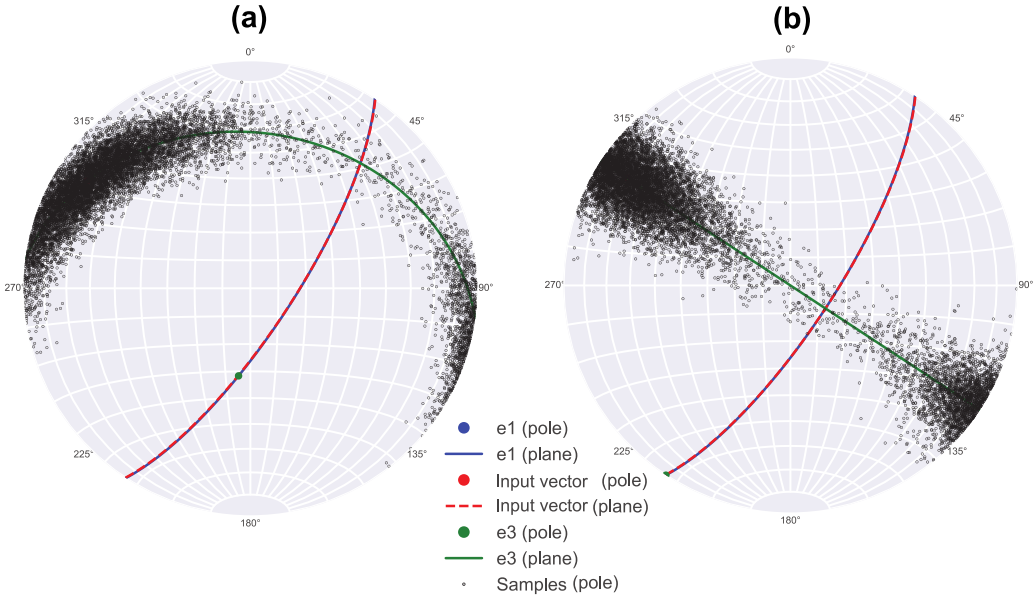

Figure 3. (a) Random $\boldsymbol{e_3}$ orientation of anisotropic Bingham distribution resulting from simulation and (b) rotated distribution with properly aligned $\boldsymbol{e_3}$ orientation.

4 Uncertainty assessment

After determining the appropriate probability distributions to use in the ~~MCUP formulation~~probabilistic geomodel, the uncertainty of chosen geologic modeling inputs is quantified based on geologic data and prior knowledge available. This essential step requires a thorough consideration of the types of uncertainty and methods for quantification available for each modeling

input. To minimize potentially unwanted bias in the ~~MCUP formulation~~probabilistic geomodel, careful attention must be paid to understanding the geologic nature of the objective and subjective sources of uncertainty affecting each modeling input.

Two models were generated for this study from data derived from a historical geologic map in the Rocky Mountains of Colorado, USA (Figure 4). The first model contains a single fault zone (Figure 4(a)) and is analyzed to clearly illustrate the impacts of different geologic modeling inputs and uncertainty parameterizations on a 3D geologic model. The second model contains a network of major and minor fault zones (Figure 4(b)) and is provided to demonstrate how the methodology developed can be generalized to more complex models and to discuss the implications of fault zone interactions on geologic model uncertainty. The models follow the workflow illustrated in Figure 1 to simplify the 3D geometry of fault zones to satisfy the flexibility and automation required ~~by MCUP~~for uncertainty propagation.

The geology in the project area consists of uplifted, Precambrian crystalline igneous and metamorphic rocks. The most recent period of tectonic activity occurred from c. ~70-40 Ma. during the Laramide Orogeny which formed the modern day Rocky Mountains. During this time, brittle faulting occurred pervasively throughout the study area as part of the regional Loveland-Berthoud Pass Fault Zone which passes just east of the study site, where it trends NNE for nearly 50 km (Lovering, 1935). A large number of fault zones of varying widths cross through the study area and were mapped at a 1:12,000 scale by Robinson et al. (1974). The geologic map (Figure 4) contains approximate traces and boundaries of fault zones in the study area, showing fault zones with thickness ranging from 5-50 ft (~1.5-15 m) as lines and fault zones greater than 50 ft (~15 m) thick as hatched zones. Robinson et al. (1974) provided contoured stereonets of all faults mapped in the area, revealing that the majority of fault zones mapped dip steeply to the east/northeast. However, specific information on the dip angle of individual faults is missing from the geologic map. In addition to this significant source of orientation uncertainty, the historic map also includes significant geographical uncertainty resulting from georeferencing and drafting errors.

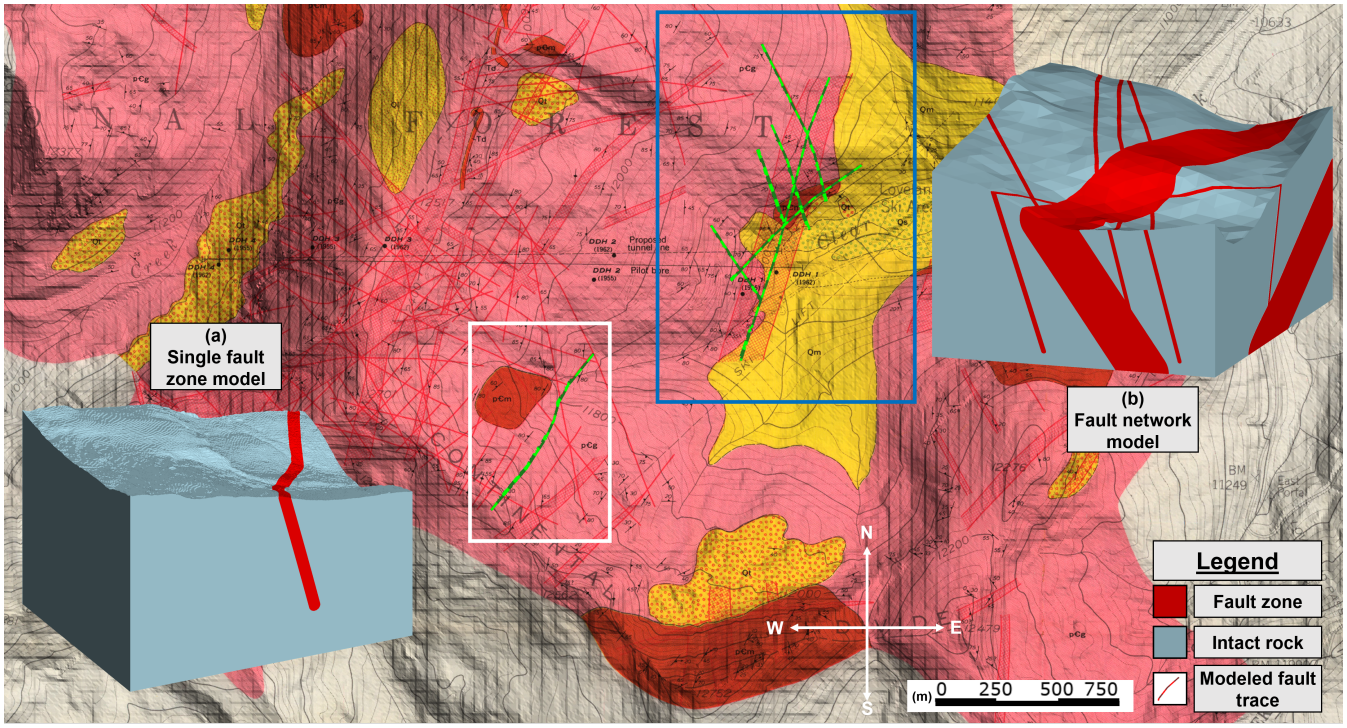

Figure 4. 1:12,000 geologic map from Robinson et al. (1974) showing mapped fault zones of varying widths. The white rectangle and associated overlay (a) show the single fault model while the blue rectangle and associated overlay (b) show the fault network model. Fault trace(s) used for modeling are highlighted within each rectangle as green polylines.

Model realization creation is handled by custom Leapfrog Works back-end support developed for this study to allow for automated updating of geologic modeling parameters from an initial model containing input fault zones. The initial model must be created in Leapfrog using the workflow provided in Figure 1, and naming conventions for the geologic model, fault zones, polylines and termination surfaces being specified in a user-generated text file. The custom version of Leapfrog Works uses this text file in conjunction with the Leapfrog model and input realization files to automatically step through creation of each geologic model realization based on the simulated data, following the workflow from Figure 1. Model realizations generated in this manner are automatically evaluated onto a grid of cells defined by the text file for subsequent analysis. The method put forward by Wellmann and Regenauer-Lieb (2012) is implemented to calculate the probability of occurrence of fault zone lithology in each cell. The probability of occurrence is then used to compute information entropy to describe the uncertainty of fault zones in the geologic model. In a binary system (e.g., fault zone vs. intact rock), information entropy is maximal when the probability of occurrence of a fault zone is 50%, which as discussed in Krajnovich et al. (2020a) can introduce potential for misinterpretation of the geologic model uncertainty envelope if an inappropriate colormap is used. The method implemented in Leapfrog can be made available to other researchers on the basis that they contact the developers

of Leapfrog (Seequent) independently to inquire about access to the unique functionality (which is built on top of a default Leapfrog installation).

A set of 1,000 realizations for each modeling input were propagated into the 3D geologic model independently and compared with the combined geologic model uncertainty assessment from all four modeling inputs (Figure 5). The significance of orientation and vertical termination depth uncertainty on the combined model uncertainty is clearly apparent at depth, while the uncertainty of the fault zone near the ground surface is dominated by surface trace uncertainty. Uncertainty about the fault zone thickness appears to be largely overshadowed by that of the surface trace, which is considered to be a consequence of the significant georeferencing and drafting errors arising from the use of the historic geologic map (Section 6.1).

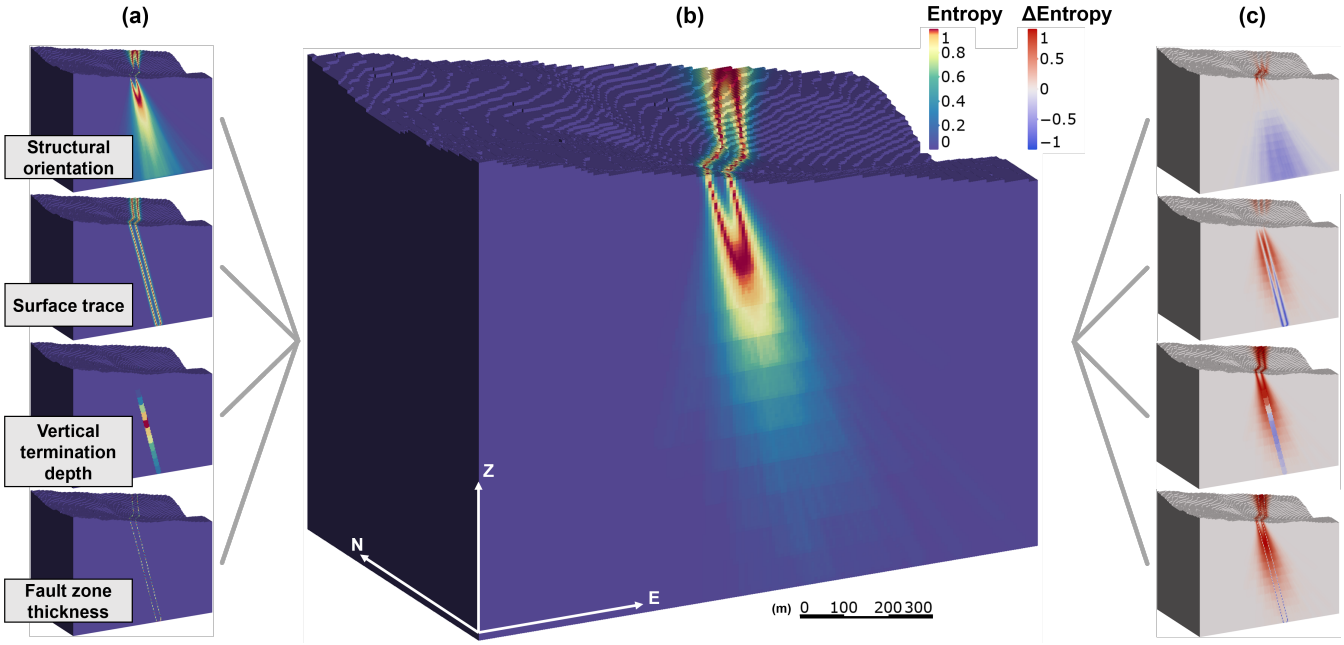

Figure 5. Block models showing information entropy quantified from (a) independent modeling inputs and (b) combined modeling inputs. The difference between the combined geologic model uncertainty and each independent modeling input is shown in (c), where blue values indicate that the independent modeling input showed greater entropy than the combined model uncertainty.

The following sections provide a detailed description of the methods implemented for parameterizing each geologic modeling input propagated into the model shown in Figure 5. The input perturbation script, which is compatible with virtually any 3D geologic modeling software, has been developed using open source code of the R and Python languages and is published at the code and data supplement. The interested reader is suggested to refer to the script alongside Sections 4.1-4.4 for exact results, figures and parameters of the ~~MCUP formulation~~probabilistic geomodel setup.

4.1 Structural orientation

The structural orientation of a fault zone varies along its surface and is implemented in modeling as a single dip azimuth and dip angle vector applied to the implicit modeling interpolant. Natural variability of the fault surface and error of individual measurements contribute objective uncertainty to the orientation used for modeling (Whitmeyer et al., 2019; Stigsson, 2016). Additional, subjective sources of uncertainty are present that may affect the fault orientation in non-random ways (Bond, 2015), including the application of prior knowledge (e.g., regional structural analysis) or measurement bias from difficulty interpreting fault slip surfaces.

Stigsson (2016) shows that the objective uncertainty of measurement imprecision often underestimates the true uncertainty associated with a geologic structure's orientation, suggesting that a more robust characterization of structural uncertainty should consider potential sources of subjective bias. In the example of modeling fault zones from a geologic map (Figure 4), anisotropic uncertainty untold by any single measurement is present due to the inconsistent information available regarding fault zone dip azimuth and dip angle. An anisotropic Bingham distribution is simulated to capture the uncertainty of the input orientation data from the geologic map. Observed variability in the surface trace and the regional structural analysis provided in Robinson et al. (1974) were used to establish the uncertainty space used in simulation. Section 6.3 continues the discussion regarding the use of an anisotropic structural orientation uncertainty envelope.

As the dataset from Robinson et al. (1974) lacks specific measurements of fault zone orientations to generate a Bingham distribution directly using the orientation matrix (Fisher et al., 1987), the parameters of the Bingham distribution were assigned manually to generate a distribution covering the expected range of orientations. Varying the parameterization of the Bingham distribution ultimately resulted in determining an appropriate parameterization for the fault zone in the single fault model using an input orientation of $75^o/123^o$ (dip/dip azimuth), maximum eigenvalue of $\lambda_1 = 200$ (matching the observed azimuth variation of the surface trace) and intermediate eigenvalue of $\lambda_2 = 22.5$ (providing an approximate +/-20^o dip angle variation).

4.2 Surface trace

The surface trace of a fault zone is a polyline along the topography which follows an approximate central fault surface. In ideal conditions, the fault trace would follow the centerline of the fault zone and reach along the whole length of the fault, ending at the fault tip points. However, in reality, the centerline of the fault zone can rarely be determined exactly (Childs et al., 2009). Additionally, when extracting the surface trace from a scanned geologic map, digitization error and geographical errors are likely to be present as well. In the proposed ~~MCUP formulation~~probabilistic geomodel setup, the fault centerline inherits uncertainty from its possible position within the perceived fault zone as well as from the relevant geographical errors.

The uncertainty affecting the surface fault trace results in changes in the trace location and shape. Independent perturbations of the trace's endpoints are applied and linearly propagated along the fault trace to arrive at a smoothly varied location and shape. ~~A normal distribution characterizing the uncertainty about the location of each trace endpoint is parameterized from~~

~~the joint uncertainty stemming from fault zone centerline definition, digitization error and geographical errors in addition to a random direction of shift obtained from a uniform distribution.~~

~~The three~~ The three primary sources of uncertainty are quantified using the available information listed in respective order: average fault zone thickness, published metrological studies (Zhong-Zhong, 1995) and approximate geographical error of known landmarks ~~. Additional details on~~ (e.g., mountain tops).

First, a bounded uniform distribution is parameterized to simulate a random direction of perturbation for each trace endpoint due to geographical error (i.e., drafting and georeferencing error). A normal distribution representing the total bound on geographical error is converted to respective $\hat{x}$ and $\hat{y}$ components using the directional cosine of the ~~method of perturbing the location and shape of the surface trace~~ angle sampled from the uniform distribution. This conversion to unit components is used similarly with the fault zone centerline definition uncertainty and digitization uncertainty using the acute angle θ between the orientation of the fault trace with the northing and easting directions. An additional logical check for the strike quadrant of the surface trace is required to implement this approach.

The three individual sources of uncertainty affecting the surface trace endpoint locations are combined into a derived distribution using a deterministic function to determine the total uncertainty affecting the location of each endpoint, given by Eq. 3.

$$\begin{aligned} P(\hat{x}|\sigma_{centerline},\sigma_{dig},\sigma_{geo},\theta) &= cos(\theta)\Big(N(0,\sigma_{centerline}) + N(0,\sigma_{dig})\Big) + N(0,\sigma_{geo})sin\Big(U(0,2\pi)\Big), \\ P(\hat{y}|\sigma_{centerline},\sigma_{dig},\sigma_{geo},\theta) &= sin(\theta)\Big(N(0,\sigma_{centerline}) + N(0,\sigma_{dig})\Big) + N(0,\sigma_{geo})cos\Big(U(0,2\pi)\Big) \end{aligned} \tag{3}$$

The average fault zone thickness was used to characterize the fault zone centerline definition uncertainty affecting each surface trace endpoint. The geographical error was calculated to be approximately 40 meters based on the ~~joint effect~~ average distance measured between known landmarks (e.g., mountain tops) on the geologic map and modern satellite imagery data. For both of these sources of uncertainty~~are available in the code supplement.~~, the maximum error range described is treated as a 95% confidence interval, allowing a normal distribution to be parameterized with a mean of zero and a standard deviation equal to $maximumerror/3.92$. The digitization error for a 1:12,000 map was represented by a normal distribution with a standard deviation of 3.666 m based on (Zhong-Zhong, 1995).

4.3 Vertical termination depth

In implicit geologic modeling, by default, all faults extend through the entire model domain. To arrive at a more realistic approximation of the 3D fault geometry, a series of surfaces at fixed elevations are defined and used to terminate faults. The depth at which a fault is expected to terminate is based on prior knowledge obtained from past works into approximating the 3D geometry of faults (Walsh and Watterson, 1988; Nicol et al., 1996; Schultz and Fossen, 2001; Torabi et al., 2019a). These works established empirical relationships for fault surface geometry using an aspect ratio of fault length along strike from tip to tip (f_{length}, henceforth length) vs. fault height along dip (f_{height}, henceforth height): $Aspect\ ratio = \frac{f_{length}}{f_{height}}$. The fault surface aspect ratio, while highly variable (ranging from ~1.5-16 in the sedimentary basin rocks studied by Torabi et al. (2019a)), has been demonstrated to be an effective measure for quantifying the 3D geometry of fault surfaces. While studies

into the aspect ratio of faults in crystalline rock are scarce and typically focus on major thrust faults, a reasonable assumption is that the aspect ratio will lie in the lower range of that for sedimentary basin rocks (Torabi et al., 2019a) due to the lack of mechanical stratigraphy; an aspect ratio of 1 to 5 was assumed to be appropriate for ~~the MCUP formulation in~~ this study.

The vertical termination depth used in modeling inherits uncertainty from the aspect ratio and persistence of the fault trace used. Lacking specific information on the location of fault tip points, an assumption of length must be made using the persistence of a fault trace at the surface as a proxy for the fault length. However, the endpoints of the fault trace may not be the true tip points due to artificial truncation by overburden or lack of outcrop. Lacking a detailed study of fault tip points, this uncertainty can be characterized by averaging the fault trace lengths of a number of faults of the same group (i.e., similar orientation).

~~The~~ Sampling the uncertainty of the fault zone vertical termination depth ~~is sampled from a deterministic distribution combining uncertainty about the aspect ratio, fault trace persistence and prior information of the fault elevation and average dip~~ involves combining multiple probability distributions using a deterministic function to generate an empirically derived probability distribution. In the derived distribution for fault zone vertical termination depth, f_{length} and $Aspect\,ratio$ are characterized as independent probability distributions and combined using a deterministic function based on the empirically derived description of 3D fault surface geometry, $z_{term} = f_{height} * sin(\theta)$; $f_{height} = \frac{f_{length}}{Aspect\,ratio}$. In this manner, the vertical termination depth (z_{term}) is calculated by converting the fault height to the vertical height using the average dip angle (θ) and subtracting this from the average elevation of the fault outcrop ($z_{outcrop}$). Sampled vertical termination depths are ~~continuous, and are~~ subsequently discretized onto the pre-defined termination surfaces ~~in the geologic model. The scale at which termination surfaces are defined is determined~~ created during the initial geologic modeling step (Figure 1(b)). The interval of these termination surfaces can be adjusted based on the end user needs of the geologic model; 50 meter intervals were used in this study to balance illustrative quality with model processing time.

f_{length} is characterized as a normal distribution with a mean of zero and a standard deviation of 20 meters based on assumed deviance of the surface trace length vs. the true tip-tip fault length. $Aspect ratio$ was characterized using both a uniform and a log-normal distribution, respecting the expected maximum and minimum values of 1 and 5. Section 6.2 explores the impact of these two different parameterizations on the shape of the derived distribution for vertical termination depth.

4.4 Fault zone thickness

In the ideal case, a detailed study of fault zone thickness at outcrop provides a measure of the average thickness and its variability. Given the thickness range of 5-50 ft (1.5-15 m) on fault zones mapped as lines in Figure 4, a conservative estimate on fault zone thickness in the case study project area was characterized using a normal distribution with $\mu = 30m$ and $\sigma = 2m$.

While not directly applied in the model shown in Figure 5, when lacking direct observations, fault zone thickness can also be approximated based on prior knowledge of a fault's estimated displacement using an established displacement to thickness (D:T) relationship appropriate for the project's geology (Torabi et al., 2019b; Childs et al., 2009). This approach would inherit uncertainty from the subjective interpretation of the fault's historical displacement and the empirical derivation of the fault's D:T relationship. With a measure of displacement, a power-fit relationship allows for approximating the fault zone thickness.

The input perturbation script includes functions for exploring the use of the D:T relationship based on a linearized power law function provided by Torabi et al. (2019b), with curve-fitting parameters $\log_{10}(b)$ and m to approximate fault zone thickness (f_t) from displacement (f_d), and a modifier $f_{CoreVsZone}$ used to model different sections of the fault zone based on the work by Childs et al. (2009). Uncertainty arises in the parameters of the D:T power-fit relationship, $\log_{10}(b)$ and m and the assumed fault displacement.

4.5 Simulation quality assessment

The quality of probabilistic simulation ~~relies on primarily~~ is a product of the size of the uncertainty space, the simulation method used and the number of samples drawn. For any simulation, the realizations generated can be plotted in the data space and visually examined for appropriate coverage and shape (termed a realization plot). For ~~non-spherical~~ scalar data types, ~~use of MCMC simulation methods creates trace plots and posterior histograms , which~~ histograms of the Monte Carlo draws provide an intuitive method for independently assessing the quality of simulation for each input. ~~Uniformity of the trace plot and width of the 95% highest posterior density (HPD) indicate, respectively, the convergence of the Markov Chain sampling and the thorough exploration of the posterior~~ Visual analysis of the shape of the histogram compared to the expected shape of the distribution and a comparison between the input distribution parameters (e.g., mean and standard deviation for a normal distribution) and their values calculated from the samples can quickly determine whether the samples drawn have sufficiently explored the uncertainty space. Figure 6 shows an example of the realization plot ~~, trace plots and posterior~~ sample histograms generated for the simulation of vertical termination depths from Section 4.3. This figure allowed for identifying a strong tailing behavior in the output realizations, leading to a reparameterization discussed in Section 6.2.

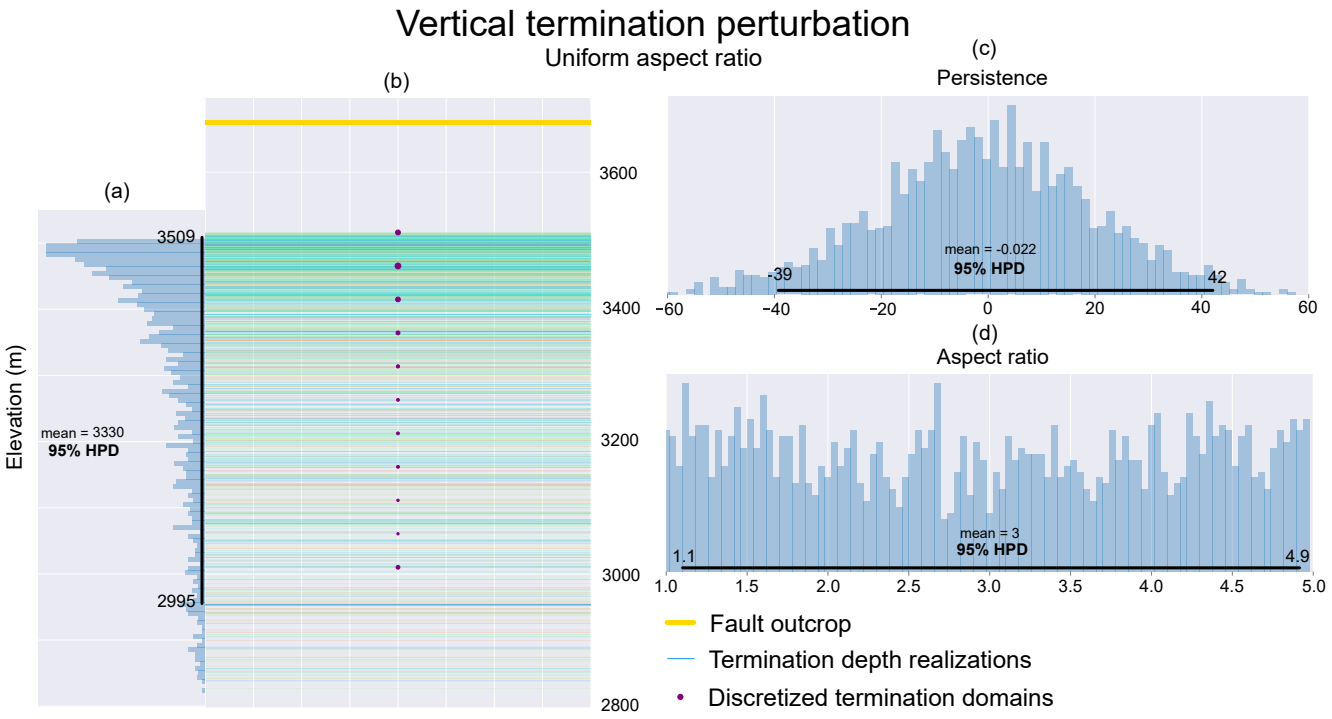

Figure 6. ~~Realizations and~~ Visualization of Monte Carlo ~~analysis results (trace plot~~ samples and ~~posterior histograms)~~ associated geologic input realizations from perturbation of the fault zone vertical termination depth based on a uniform distribution of fault aspect ratio. The 95% highest ~~posterior~~ predictive density is overlain on the ~~posterior~~ histograms of the Monte Carlo samples.

~~Trace plots are not available for the~~ For spherical data simulations~~due to reliance on the acceptance-rejection simulation method, while posterior~~ , histograms may be replaced by Exponential Kamb contouring ~~Exponential Kamb contouring~~ (Vollmer, 1995) or Rose diagrams to visualize the density of sampled poles across the surface of the unit sphere (as projected onto a lower-hemisphere projection). This visual assessment provides a semi-quantitative evaluation of the shape ~~of the posterior spherical probability distribution~~and distribution of the sampled structural orientations. Additionally, a recalculation of the eigenvector decomposition from the set of simulated samples provides a measure of the accuracy of the ~~posterior distribution~~ distribution of samples with respect to the input ~~orientation~~ parameter values. Tools for generating figures for simulation quality assessment are provided and detailed in the input perturbation script.

Based on the assessment of simulation quality and consideration of compounding factors during uncertainty propagation, ~~the MCUP formulation for the~~ probabilistic geomodeling of the single fault model was run for a number of various realization counts (100, 300, 500, 1,000, 2,000 and 3,000). The processing time generally increases linearly with realization count, reaching many hours to several days for high realization counts on the single fault mock model containing 2.5 million cells. The vast majority of processing time is taken up by the model updating and block model calculation in Leapfrog. For the single fault mock model with 1,000 realizations and 2.5 million cells the sampling benchmark time was 87 seconds while the model

processing benchmark time was 38.5 hours. This study is intended to introduce and expand on the use of ~~MCUP formulations~~ probabilistic geomodeling for specific geologic modeling problems, and work regarding optimizing the efficiency of model processing is not a focus. The experiments conducted do highlight the need to understand (i) the realization requirement for exploring modeling inputs independently and its relationship to the size of the independent uncertainty spaces, (ii) the interactions of various, related parameters during the uncertainty propagation step and (iii) identification of a balance between final model resolution, coverage, complexity and processing time.

5 Fault network model

The fault network model shown in Figure 4(b) contains a major fault zone (thickness = ~175 m) with five minor fault zones (thickness $\leq$ 16 m) branching out of it. The parameterization of modeling inputs was conducted following the methods described in Section 4. The input orientations of each fault were determined based on the measured surface trace strike, and a dip angle assigned initially by a static random value based on the data published by Robinson et al. (1974). The resultant uncertainty model is shown in Figure 7, with cross-sections highlighting the intersection of three minor fault zones with the major fault zone. The modeling formulation scaled effectively to the more complex fault network model, which included two deterministic horizontal terminations of minor faults into the major fault zone.

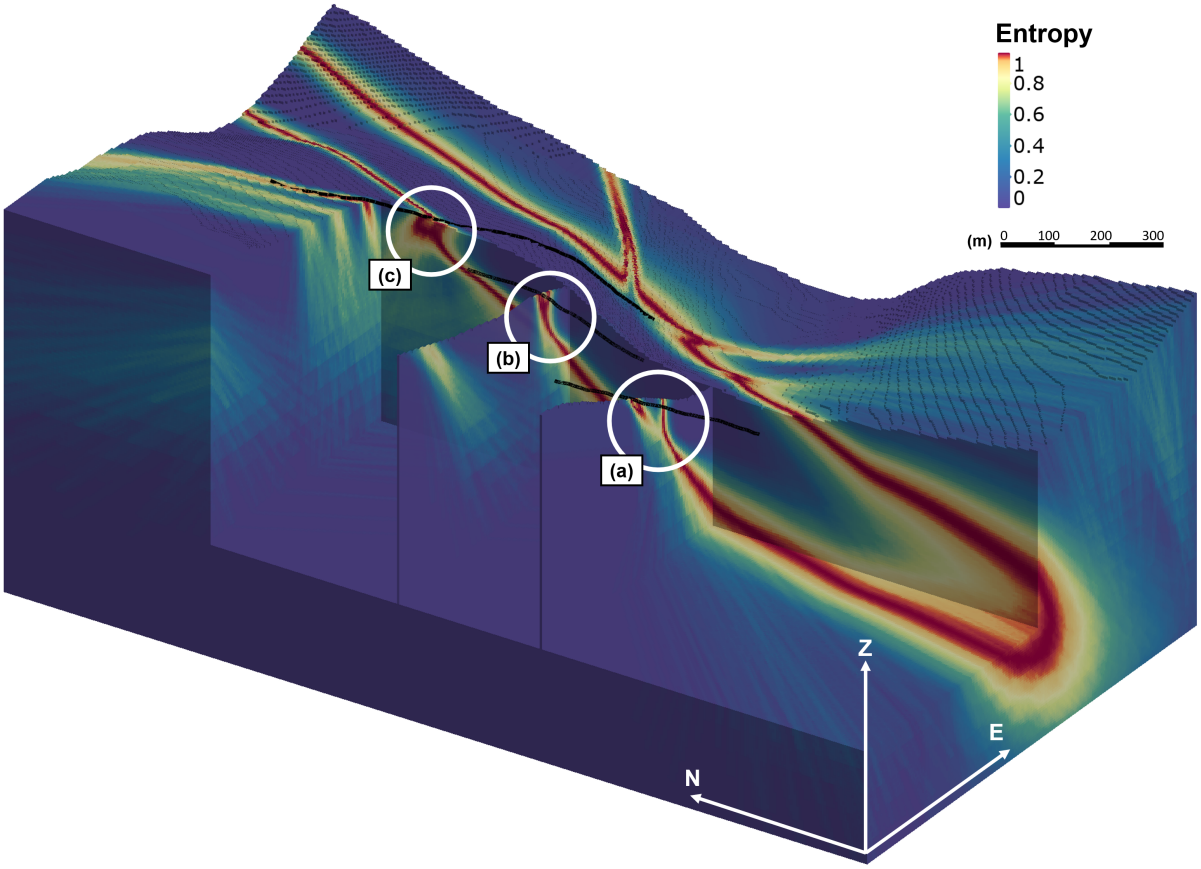

Figure 7. Combined geologic model entropy results for the fault network model, highlighting three intersection points between minor fault zones with the major fault zone: (a) approaching major fault zone boundary, (b) bordering major fault zone boundary and (c) overlapping boundaries.

Focusing on the intersection of fault zones reveals the interaction between fault zone uncertainty envelopes. In Figure 7, the slice (a) shows the uncertainty envelope of the major fault zone being deflected by a zone of lower entropy as the nearby minor fault zone approaches it. Slices (b) and (c) show the uncertainty envelopes of major and minor fault zones merging, showing how the deflection from (a) transfers into a hot spot of high uncertainty at the peak intersection point (overlapping boundaries). Just as the uncertainty of a single fault zone is maximal at its boundaries, the overlapping of two fault zone boundaries produces an entropy hot spot.

6 Discussion

6.1 Historic dataset sensitivity

The observed sensitivity of the model uncertainty to the surface trace perturbation may be due to the relative significance of geographical sources of uncertainty in the case of modeling from a historic geologic map. These errors include inherent error in the geologic map stemming from its accuracy and georeferencing errors arising from the conversion between geographical and projected coordinate systems. Georeferencing typically uses a rubber-sheet method, moving, rotating and stretching the map – in an ArcGIS platform – to optimally translate from the geographical coordinate system of the map (Lat/Long) to the projected coordinate system required by modeling (UTM Northing and Easting), and the error typically reported by these methods may underestimate the true error associated with the georeferenced map. Quantifying the true geographical error is difficult, but can be approximated by comparing the location of known landmarks (e.g., intersection of roads, mountain tops, corners of buildings) between the georeferenced map and modern, digital datasets. From Figure 4, a maximum bound on the error in the location of known features across the geologic map was approximated to be 40 meters by visual examination of the georeferenced map. While the same methodology would apply to a surface trace obtained from modern mapping methods (e.g., global positioning systems), it is apparent that the contribution to model uncertainty could be drastically reduced. Reapplying the methodology to a modern dataset may highlight different sensitivity.

6.2 Model reparameterization

Based on the inherent variability of fault aspect ratios, initially a bounded uniform distribution was determined to be appropriate (Figure 6(a)). However, the ~~posterior~~ empirically derived distribution of vertical termination depths resulting from a bounded uniform parameterization of fault aspect ratio showed a strong tailing effect (right skewed). This could be interpreted as being unrealistic, arising as an ~~artifact~~ artefact due to the shortening of vertical termination depth intervals for equivalent increases in fault aspect ratio. A reparameterization using a custom log-normal distribution of aspect ratio respecting similar maximum bounds of 1 and 5 is illustrated in Figure 8(b), showing a significant reduction in the tailing effect present in the vertical termination depth realizations. This reparameterization highlights the key strength ~~– and susceptibility – of MCUP formulations~~ -- and susceptibility — of probabilistic geomodeling based on geologic modeling inputs, the reliance on a user defined characterization of input uncertainty. Again, it is necessary to reiterate that the modeler must take into consideration not only field observations and theoretical prior knowledge when assessing a geologic modeling uncertainty formulation, but also their informed expectation of what is geologically realistic for their chosen modeling problem. Furthermore, reparameterizing individual aspects of the probabilistic model may prove to be insufficient due to the presence of inherently unknown relationships between the chosen model parameters.

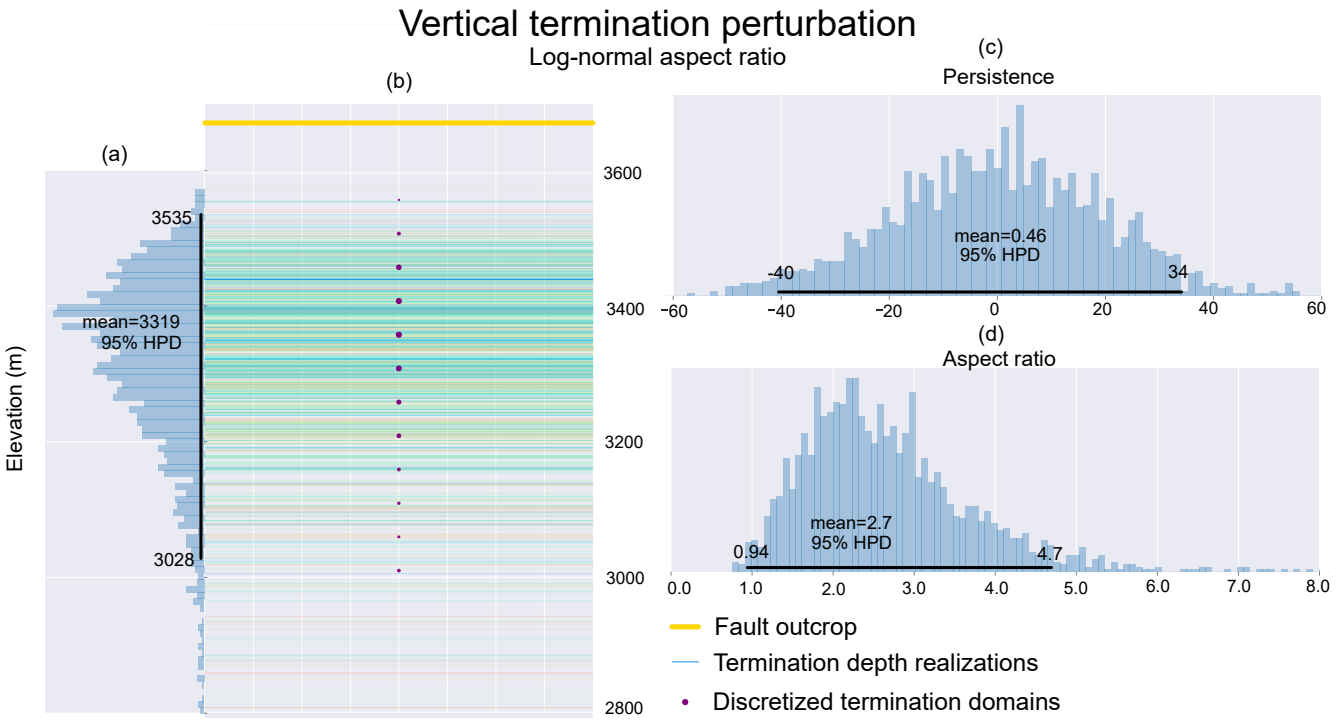

Figure 8. ~~Realizations and~~ Visualization of Monte Carlo ~~analysis results (trace plot~~ samples and ~~posterior histograms)~~ associated geologic input realizations from perturbation of the fault zone vertical termination depth, reparameterized using a log-normal distribution of fault aspect ratio. The 95% highest ~~posterior~~ predictive density is overlain on the ~~posterior~~ histograms of the Monte Carlo samples.

6.3 Parameter relationships

While many ~~MCUP formulations~~ proabilistic geomodeling approaches – including this study – have focused on independent perturbations of geologic modeling inputs, relationships between modeling inputs are apparent in the modeling formulation developed. Intuitively, there is a relationship between the average orientation of a fault's surface trace and the structural orientation applied to the fault surface modeling interpolant's global trend. Deviations in the modeled fault surface from the input orientations can occur when the two inputs are significantly different, typically arising in Leapfrog Works by way of overestimation of fault surface dip by up to 10~~degrees~~ ^{o} when the surface trace azimuth (i.e., average normal to the fault surface trace) and global trend dip azimuth differed by greater than 20^o (Krajnovich et al., 2020a). Comparing two geologic models generated with orientations sampled from anisotropic and isotropic Bingham distributions with equivalent maximum uncertainty ranges (Figure ~~??) showed artifacts present in~~ 9) showed a skewing of the geologic model uncertainty envelope when generated from the isotropic Bingham distribution. The results show a consistent skew towards near-vertical dips of the modeled fault zone realizations. This issue is alleviated when the structural orientation is parameterized with an anisotropic Bingham distribution, allowing for increased variability in the dip angle without compromising the certainty of the dip azimuth.

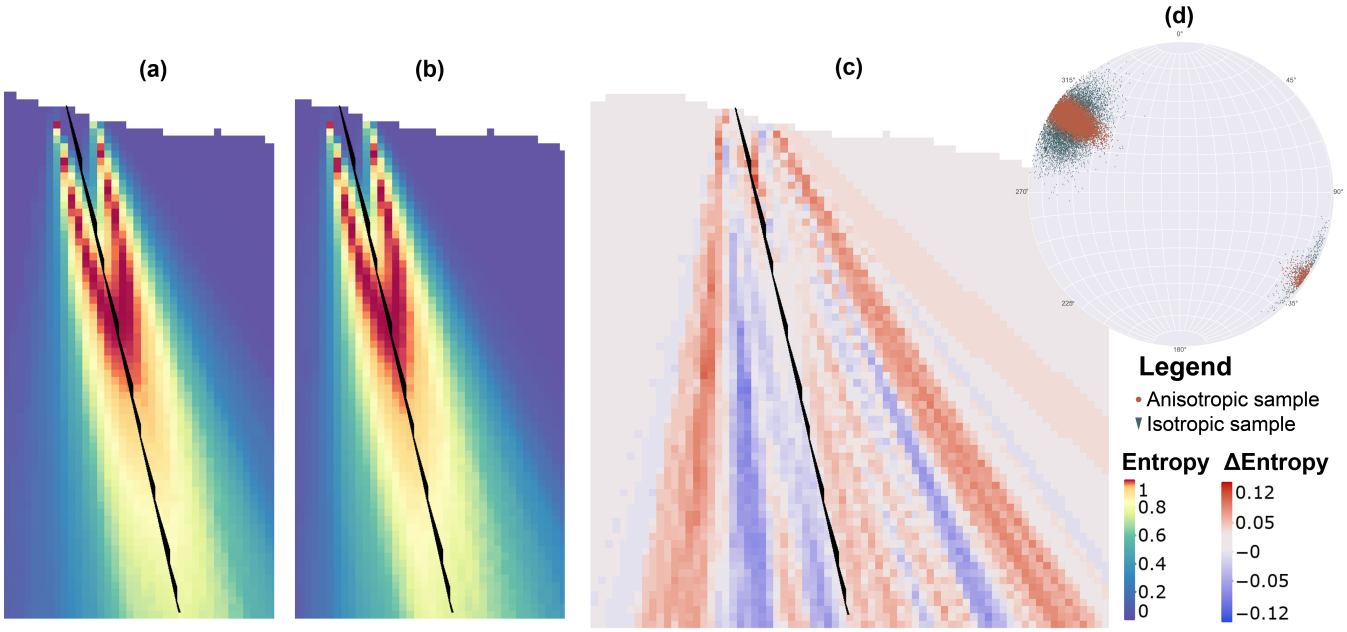

Figure 9. ~~Block models generated from 1,000 realizations~~ Visual comparison of geologic model entropy generated using anisotropic (a) ~~an anisotropic Bingham distribution~~ and isotropic (b) ~~an isotropic~~ Bingham ~~distribution~~distributions. The entropy difference between these two models (c) ~~Comparison~~ highlights slight skewing of ~~10,000 orientation~~ the uncertainty envelope towards steeper dipping fault zones when parameterized using samples ~~collected~~ from ~~anisotropic and~~ an isotropic Bingham ~~distributions covering the same range of maximum uncertainty~~ distribution (~~$\lambda_2 = 22.5$~~d). Entropy difference is shown such that negative values indicate higher entropy in the isotropic orientation model.

The observed interaction between the surface trace and structural orientation inputs to the implicit 3D geologic model suggests that coupling and correlation may be present between different geologic modeling inputs.

~~Relationships between modeling inputs also arise in different ways, for example the vertical termination depth and structural orientation uncertainty envelopes overlap heavily in the combined model uncertainty (Figure 5)leading to~~ Despite performing a thorough exploration of each, independent parameter's uncertainty during Monte Carlo sampling (Section 4.5), undersampling of the combined geologic model uncertainty space ~~when the independent uncertainty envelopes are combined. Similar behavior is observed when comparing orientationperturbations to fault zone thickness where thinner fault zones require finer orientation perturbations to fully populate the uncertainty space of the 3D geologic model .~~ can still occur during uncertainty propagation. An example of this arises when considering the vertical termination depth and structural orientation. Truncation of fault zone realizations at any given termination interval effectively reduces the number of realizations available for sampling the full range of structural orientation uncertainty at deeper intervals. This is evidenced in Figure 5(b) by the increasing prevalence of "stair-stepping" artefacts in the combined model uncertainty with depth.

Coupling of the surface trace and structural orientation or structural orientation and vertical termination depth is not implemented in the current formulation due to inconsistencies in the sampling methods which required independent sampling of spherical and non-spherical distributions. Increasing the number of model realizations mitigated some of the effects of input correlations, though this comes at the expense of increased processing time. A treatment of these relationships through ~~the use of Gibbs sampling or other conditional methods of Monte Carlo simulation~~ parameterizing previously independent input probability distributions using a joint distribution (and an appropriate sampling scheme) could potentially generate more realistic and efficient assessments of model uncertainty. While the correlation between parameters in a probabilistic model is typically unknown, in some cases – such as the correlation between fault trace azimuth and structural orientation azimuth – the correlation can be assumed based on available prior geologic knowledge of the parameter's real-world relationships. However, until standard methods for simulating spherical and scalar data types in a single system are made available, creative model parameterizations – such as constraining the dip azimuth uncertainty to observed surface trace variability using an anisotropic spherical distribution – can circumvent some of the issues associated with these relationships.

6.4 Anisotropic spherical distributions

For modeling anisotropic uncertainty of structural orientation data, the Bingham and Kent distributions generate practically equivalent uncertainty envelopes, differing primarily in their parameterization. For the Bingham distribution, variations in the magnitudes of λ_1 and λ_2 allow for independently varying the size of the uncertainty space in orthogonal directions while for the Kent distribution variations in the values of κ and β can generate distributions with different overall size and ellipticity. To achieve independent uncertainty ranges of the dip angle and dip azimuth, both distributions require a series of rotations (described in Section ~~??~~3.2) to properly align the distribution such that the major ellipse axis is aligned with a great circle with 90^o dip (Figure 3). Once aligned, the range of the Bingham distribution in the dip azimuth and dip angle directions can easily be varied independently through direct changes in the magnitudes of λ_1 and λ_2. The Kent distribution, however, introduced difficulty in varying these uncertainties independently due to coupling of the κ and β parameters. Compared to the Kent distribution, the Bingham distribution was observed to be more efficient at modeling strongly girdle-shaped distributions, which as discussed in Section 4.1 can be particularly useful for the limited data available from a geologic map. For these reasons, the Bingham distribution was chosen for the analysis in this study, though methods for simulating and rotating the Kent distribution are still provided for thoroughness.

6.5 Additional complexity for fault zone geometry

The ~~author acknowledges~~ authors acknowledge that in reality fault zone geometry includes horizontal terminations. The mechanisms affecting the location of horizontal terminations are numerous and varied, including fault zone anastomosing, abrupt termination in intact rock and false trace termination due to obscuring by overburden. While several works have investigated the nature of fault-fault terminations using stochastic simulation (Aydin and Caers, 2017; Cherpeau and Caumon, 2015), due to the presence of other poorly defined sources of uncertainty, the placement of horizontal terminations in this study remained

deterministic. Future work supplementing the limited dataset used in this study with detailed outcrop studies will be required for defining the nature and uncertainty of horizontal fault zone terminations.

Aside from horizontal fault terminations, fault zones present other interesting complexities that introduce additional levels of refinement for the developed modeling workflow. For example, internal fault zone composition is heterogeneous and modeling the different fault zone components (fault core, transition zone and damage zone) and could be implemented using the developed ~~MCUP formulation~~ methods if desired. Asymmetry of fault zone structure between the hanging wall and footwall has also been documented by Choi et al. (2016), which would require a new method of defining the distance function for generating the 3D fault zone volume. Similarly, variations in fault zone thickness along the area of the central fault surface are realistic, though would similarly require a new method for defining the distance function. These research questions enter into the realm of implicit geologic modeling theory, and are introduced merely to shed light on where refinements in model creation can benefit the modeling of fault zone structure.

7 Conclusions

~~The flexible MCUP~~ The flexible probabilistic geomodeling method should continue to be leveraged to model novel problems in geologic modeling, such as the uncertainty of fault zones in 3D geologic models based on limited data from a historic geologic map and available prior knowledge. ~~MCUP formulations developed~~ The setup of the probabilistic geomodel should make full use of open source statistical packages in the R and Python languages, many of which are experiencing ongoing development at the time of writing. While a unique modeling case was presented, the rationale applied to assessing uncertainty of poorly-constrained, preliminary geologic models sheds light on new ways to implement varied types of information in ~~MCUP~~ probabilistic geomodeling formulations.

Practical guidelines for developing ~~MCUP formulations~~ probabilistic geomodels are also provided, reinforcing (i) the importance of developing a clear yet robust modeling workflow for the structure of interest, (ii) consideration of varied sources of geologic uncertainty and (iii) creatively exploring the methods available for characterizing both objective and subjective modeling inputs. Prior knowledge in the form of established empirical and theoretical relationships from structural geology present an opportunity to quantify and parameterize geologic modeling inputs that are usually interpretive, allowing for their inclusion in ~~automated MCUP formulations~~the probabilistic geomodel. The Bingham distribution, while only moderately impactful on model uncertainty when comparing anisotropic and isotropic parameterizations, is recommended to replace the vMF distribution for modeling structural orientations due to the increased flexibility of its parameterization. While the Bingham distribution was preferred in this study, the use of the Kent distribution appears to be practically equivalent.

Future work stemming from this ~~preliminary modeling~~ input-based probabilistic geomodeling formulation may include incorporating new information in a Bayesian inference scheme to further refine the geologic model, either following the methodology introduced by de la Varga and Wellmann (2016) to infer additional information about the model parameters (requiring a full integration of the probabilistic modeling with the automated geologic modeling approach) or by Schneeberger et al. (2017) to validate the initial model in light of its generated uncertainty. Additionally, refining the modeling of fault zone internal struc-

ture and variability is recommended not only to further the usefulness of 3D geologic models in practical applications (e.g., subsurface construction, fluid flow), but to also expand the understanding of the geometry and characteristics of these complex geologic structures.

Code and data availability. The necessary code and data for generating realizations of the geologic modeling inputs (including example results) is available at https://github.com/ajkran2/Geologic-Model-Input-Uncertainty-Characterization and https://doi.org/10.5281/zenodo.3676094 (Krajnovich et al., 2020b). Version updates, if applicable, will be made available via GitHub.

Author contributions. Ashton Krajnovich is the primary researcher on this project responsible for the literature review, data acquisition, geologic modeling, code development and writing. Dr. Wendy Zhou is the primary research advisor providing close supervision, feedback, guidance and revisions. Dr. Marte Gutierrez is the research co-advisor and is a primary contributor to forming the overarching research questions that this research attempts to address.

Competing interests. The author's declare that no competing interests are present.

Acknowledgements. This research is funded by the United States Department of Transportation University Transportation Center for Underground Transportation Infrastructure, Grant No. 69A3551747118. The authors would also like to gratefully acknowledge the outstanding support from Dr. Rose Pearson from Seequent for dedicating her time to develop automated generation of geologic model realizations and block model evaluations inside the Leapfrog Works software.

References

Ailleres, L., Jessell, M., de Kemp, E., Caumon, G., Wellmann, F., Grose, L., Armit, R., Lindsay, M., Giraud, J., Brodaric, B., Harrison, M., and Courrioux, G.: Loop - Enabling 3D stochastic geological modelling, ASEG Extended Abstracts, 2019, 1–3, https://doi.org/10.1080/22020586.2019.12072955, 2019.

Allmendinger, R. W.: Modern Structural Practice, A structural geology laboratory manual for the 21st Century, 1.9.0 edn., 2015.

Anderson, E. D., Zhou, W., Li, Y., Hitzman, M. W., Monecke, T., Lang, J. R., and Kelley, K. D.: Three-dimensional distribution of igneous rocks near the Pebble porphyry Cu-Au-Mo deposit in southwestern Alaska: Constraints from regional-scale aeromagnetic data, Geophysics, 79, B63–B79, https://doi.org/10.1190/geo2013-0326.1, 2014.

Aydin, O. and Caers, J. K.: Quantifying structural uncertainty on fault networks using a marked point process within a Bayesian framework, Tectonophysics, 712-713, 101–124, https://doi.org/10.1016/j.tecto.2017.04.027, 2017.

Bingham, C.: Distributions on the sphere and on the projective plane, Phd, Yale University, 1964.

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

Bond, C. E., Johnson, G., and Ellis, J. F.: Structural model creation: the impact of data type and creative space on geological reasoning and interpretation, Geological Society, London, Special Publications, 421, 83–97, https://doi.org/10.1144/SP421.4, 2015.

Caers, J.: Modeling Uncertainty in the Earth Sciences, John Wiley Sons, Ltd, https://doi.org/10.1002/9781119995920, 2011.

Caine, J. S., Evans, J. P., and Forster, C. B.: Fault zone architecture and permeability structure, Geology, 24, 1025–1028, https://doi.org/10.1130/0091-7613(1996)024<1025:FZAAPS>2.3.CO;2, 1996.

Calcagno, P., Chilès, J. P., Courrioux, G., and Guillen, A.: Geological modelling from field data and geological knowledge. Part I. Modelling method coupling 3D potential-field interpolation and geological rules, Physics of the Earth and Planetary Interiors, 171, 147–157, https://doi.org/10.1016/j.pepi.2008.06.013, 2008.

Carmichael, T. and Ailleres, L.: Method and analysis for the upscaling of structural data, Journal of Structural Geology, 83, 121–133, https://doi.org/10.1016/j.jsg.2015.09.002, 2016.

Caumon, G.: Towards stochastic time-varying geological modeling, Mathematical Geosciences, 42, 555–569, https://doi.org/10.1007/s11004-010-9280-y, 2010.

Cheeney, R. F.: Statistical methods in geology for field and lab decisions, George Allen & Unwin (Publishers) Ltd, London, 1 edn., 1983.

Cherpeau, N. and Caumon, G.: Stochastic structural modelling in sparse data situations, Petroleum Geoscience, 21, 233–247, https://doi.org/10.1144/petgeo2013-030, 2015.

Cherpeau, N., Caumon, G., and Lévy, B.: Stochastic simulations of fault networks in 3D structural modeling, Comptes Rendus - Geoscience, 342, 687–694, https://doi.org/10.1016/j.crte.2010.04.008, 2010.

Childs, C., Manzocchi, T., Walsh, J. J., Bonson, C. G., Nicol, A., and Schöpfer, M. P.: A geometric model of fault zone and fault rock thickness variations, Journal of Structural Geology, 31, 117–127, https://doi.org/10.1016/j.jsg.2008.08.009, 2009.

Choi, J. H., Edwards, P., Ko, K., and Kim, Y. S.: Definition and classification of fault damage zones: A review and a new methodological approach, Earth-Science Reviews, 152, 70–87, https://doi.org/10.1016/j.earscirev.2015.11.006, 2016.

Cowan, M. W., Beatson, R. K., Ross, H. J., Fright, W. R., McLennan, T. J., Evans, T. R., Carr, J. C., Lane, R. G., Bright, D., Gillman, A., Oshust, P., and Titley, M.: Practical Implicit Geological Modelling, 2003.

Dai, J. S.: Euler-–Rodrigues formula variations, quaternion conjugation and intrinsic connections, Mechanism and Machine Theory, 92, 144–152, https://doi.org/10.1016/j.mechmachtheory.2015.03.004, 2015.

Davis, J. R. and Titus, S. J.: Modern methods of analysis for three-dimensional orientational data, Journal of Structural Geology, 96, 65–89, https://doi.org/10.1016/j.jsg.2017.01.002, 2017.

de la Varga, M. and Wellmann, J. F.: Structural geologic modeling as an inference problem: A Bayesian perspective, Interpretation, 4, SM1–SM16, https://doi.org/10.1190/INT-2015-0188.1, 2016.

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

Fallaize, C. J. and Kypraios, T.: Exact Bayesian inference for the Bingham distribution, Statistics and Computing, 26, 349–360, https://doi.org/10.1007/s11222-014-9508-7, 2016.

Fisher, N., Lewis, T., and Embleton, B.: Statistical analysis of spherical data, Cambridge University Press, 1987.

Fredman, N., Tveranger, J., Cardozo, N., Braathen, A., Soleng, H., Røe, P., Skorstad, A., and Syversveen, A. R.: Fault facies modeling: Technique and approach for 3-D conditioning and modeling of faulted grids, AAPG Bulletin, 92, 1457–1478, https://doi.org/10.1306/06090807073, 2008.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian Data Analysis, https://doi.org/10.1007/s13398-014-0173-7.2, 2004.

Guillen, A., Joly, A., Ledru, P., Courrioux, G., and Calcagno, P.: Geological modelling from field data and geological knowledge Part II. Modelling validation using gravity and magnetic data inversion, Physics of the Earth and Planetary Interiors, 171, 158–169, https://doi.org/10.1016/j.pepi.2008.06.014, 2008.

Hillier, M. J., Schetselaar, E. M., de Kemp, E. A., and Perron, G.: Three-Dimensional Modelling of Geological Surfaces Using Generalized Interpolation with Radial Basis Functions, Mathematical Geosciences, 46, 931–953, https://doi.org/10.1007/s11004-014-9540-3, 2014.

Hillier, M. J., Kemp, E. A. D., Schetselaar, E. M., Hillier, M. J., Kemp, E. A. D., and Schetselaar, E. M.: Implicit 3-D modelling of geological surfaces with the Generalized Radial Basis Functions (GRBF) algorithm, Geological Survey of Canada, Open File 7814, pp. 1–15, 2017.

Høyer, A. S., Jørgensen, F., Sandersen, P. B. E., Viezzoli, A., and Møller, I.: 3D geological modelling of a complex buried-valley network delineated from borehole and AEM data, Journal of Applied Geophysics, 122, 94–102, https://doi.org/10.1016/j.jappgeo.2015.09.004, 2015.

Jessell, M., Aillères, L., De Kemp, E., Lindsay, M., Wellmann, J. F., Hillier, M., Laurent, G., Carmichael, T., and Martin, R.: Next Generation Three-Dimensional Geologic Modeling and Inversion, Society of Economic Geologists, https://doi.org/10.5382/SP.18.13, 2014.

Jessell, M., Pakyuz-Charrier, E., Lindsay, M., Giraud, J., and de Kemp, E.: Assessing and Mitigating Uncertainty in Three-Dimensional Geologic Models in Contrasting Geologic Scenarios, in: Metals, Minerals, and Society, Society of Economic Geologists (SEG), https://doi.org/10.5382/SP.21.04, 2018.

Kent, J. T.: The Fisher-Bingham Distribution on the Sphere, Journal of the Royal Statistical Society, 44, 71–80, https://www.jstor.org/stable/2984712, 1982.

Kent, J. T., Ganeiber, A. M., and Mardia, K. V.: A new method to simulate the Bingham and related distributions in directional data analysis with applications, pp. 1–16, http://arxiv.org/abs/1310.8110, 2013.

Kent, J. T., Ganeiber, A. M., and Mardia, K. V.: A New Unified Approach for the Simulation of a Wide Class of Directional Distributions, Journal of Computational and Graphical Statistics, 27, 291–301, https://doi.org/10.1080/10618600.2017.1390468, 2018.

Krajnovich, A., Zhou, W., and Gutierrez, M.: Uncertainty Assessment of Fault Zones in 3D Geologic Models of Mountain Tunnels (in press), in: Word Tunnel Congress, International Tunneling and Underground Space Association, 2020a.

Krajnovich, A., Zhou, W., and Gutierrez, M.: ajkran2 / Geologic-Model-Input-Uncertainty-Characterization(supplement to Solid Earth Discussions Manuscript 'se-2020-21'), https://doi.org/10.5281/zenodo.3676094, 2020b.

Lindsay, M. D., Jessell, M. W., Ailleres, L., Perrouty, S., de Kemp, E., and Betts, P. G.: Geodiversity: Exploration of 3D geological model space, Tectonophysics, 594, 27–37, https://doi.org/10.1016/j.tecto.2013.03.013, 2013.

Lovering, T. S.: Geology and Ore Deposits of the Montezuma Quadrangle, Colorado, Tech. rep., United States Geological Survey, Washington, D.C., https://doi.org/10.3133/pp178, 1935.

Manzocchi, T., Childs, C., and Walsh, J. J.: Faults and fault properties in hydrocarbon flow models, Geofluids, 10, 94–113, https://doi.org/10.1111/j.1468-8123.2010.00283.x, 2010.

Mardia, K.: Directional Statistics in Geosciences, Communications in Statistics - Theory and Methods, 10, 1523–1543, https://doi.org/10.1080/03610928108828131, 1981.

Mardia, K. V. and Jupp, P. E.: Directional Statistics, John Wiley & Sons, London, https://doi.org/10.1002/9780470316979, 2000.

Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijs, S. V.: A philosophical basis for hydrological uncertainty, Hydrological Sciences Journal, 61, 1666–1678, https://doi.org/10.1080/02626667.2016.1183009, 2016.

Nicol, A., Watterson, J., Walsh, J. J., Childs, C., and Group, F. A.: The shapes, major axis orientations and displacement patterns of fault surfaces, 18, 235–248, https://doi.org/10.1016/S0191-8141(96)80047-2, 1996.

Pakyuz-Charrier, E., Giraud, J., Ogarko, V., Lindsay, M., and Jessell, M.: Drillhole uncertainty propagation for three-dimensional geological modeling using Monte Carlo, Tectonophysics, 747-748, 16–39, https://doi.org/10.1016/j.tecto.2018.09.005, 2018a.

Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J., and Jessell, M.: Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modeling, a guide for disturbance distribution selection and parameterization, Solid Earth, 9, 385–402, https://doi.org/10.5194/se-9-385-2018, 2018b.

Pakyuz-charrier, E., Jessell, M., Giraud, J., Lindsay, M., and Ogarko, V.: Topological Analysis in Monte Carlo Simulation for Uncertainty Estimation, Solid Earth Discussions, https://doi.org/10.5194/se-2019-78, 2019.

Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., Borboudakis, G., Burkardt, J., Zou, C., and Lakiotaki, K.: Package 'Rfast', https://rfast.eu, 2018.

Peacock, D. C., Nixon, C. W., Rotevatn, A., Sanderson, D. J., and Zuluaga, L. F.: Glossary of fault and other fracture networks, Journal of Structural Geology, 92, 12–29, https://doi.org/10.1016/j.jsg.2016.09.008, 2016.

Peel, D., Whiten, W. J., Mclachlan, G. J., David, P., William, J. W., and Geoffrey, J. M. L.: Fitting Mixtures of Kent Distributions to Aid in Joint Set Identification, Journal of the American Statistical Association, 96, 56–63, https://doi.org/10.1198/016214501750332974, 2001.

Roberts, N. M., Tikoff, B., Davis, J. R., and Stetson-Lee, T.: The utility of statistical analysis in structural geology, Journal of Structural Geology, 125, 64–73, https://doi.org/10.1016/j.jsg.2018.05.030, 2019.

Robinson, C. S., Lee, F. T., Scott, J. H., Carroll, R. D., Hurr, R. T., Richards, D. B., Mattei, F. A., Hartmann, B. E., and Abel, J. F.: Engineering Geologic, Geophysical, Hydrologic, and Rock-Mechanics Investigations of the Straight Creek Tunnel Site and Pilot Bore, Colorado, Tech. rep., United States Geological Survey, Washington, D.C., https://doi.org/10.3133/pp815, 1974.

Røe, P., Georgsen, F., and Abrahamsen, P.: An Uncertainty Model for Fault Shape and Location, Mathematical Geosciences, 46, 957–969, https://doi.org/10.1007/s11004-014-9536-z, 2014.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C.: Probabilistic programming in Python using PyMC3, PeerJ Computer Science, 2, 1–24, https://doi.org/10.7717/peerj-cs.55, 2016.

Scalzo, R., Kohn, D., Olierook, H., Houseman, G., Chandra, R., Girolami, M., and Cripps, S.: Efficiency and robustness in Monte Carlo sampling for 3-D geophysical inversions with Obsidian v0.1.2: Setting up for success, Geoscientific Model Development, 12, 2941–2960, https://doi.org/10.5194/gmd-12-2941-2019, 2019.

Schneeberger, R., de la Varga, M., Egli, D., Berger, A., Kober, F., Wellmann, F., and Herwegh, M.: Methods and uncertainty-estimations of 3D structural modelling in crystalline rocks: A case study, Solid Earth, 8, 987–1002, https://doi.org/10.5194/se-2017-47, 2017.

Schultz, R. A. and Fossen, H.: Displacement-length scaling in three dimensions: The importance of aspect ratio and application to deformation bands, Journal of Structural Geology, 24, 1389–1411, https://doi.org/10.1016/S0191-8141(01)00146-8, 2001.

Schweizer, D., Blum, P., and Butscher, C.: Uncertainty assessment in 3-D geological models of increasing complexity, Solid Earth, 8, 515–530, https://doi.org/10.5194/se-8-515-2017, 2017.

Seequent: Unearthing 3D Implicit Modelling, http://www.implicit-modelling.com/, 2014.

Shannon, C.: A Mathematical Theory of Communication (Part I), Bell System Technical Journal, 27, 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x, 1948.

Shipton, Z., Roberts, J., Comrie, E., Kremer, Y., Lunn, R., and Caine, J.: Fault fictions : cognitive biases in the conceptualization of fault zones, Geological Society Special Publications, 2019.

Stafleu, J., Menkovic, A., Doornenbal, H., and Vernes, R.: Digital Geological Models of the Netherlands, https://www.academia.edu/3152159/Digital_Geological_Models_of_the_Netherlands, 2012.

Stamm, F. A., Varga, M. D., and Wellmann, F.: Actors, actions and uncertainties: Optimizing decision making based on 3-D structural geological models, Solid Earth, 10, 2015–2043, https://doi.org/10.5194/se-10-2015-2019, 2019.

Stigsson, M.: Orientation Uncertainty of Structures Measured in Cored Boreholes: Methodology and Case Study of Swedish Crystalline Rock, Rock Mechanics and Rock Engineering, 49, 4273–4284, https://doi.org/10.1007/s00603-016-1038-5, 2016.

Tarantola, A.: Popper, Bayes and the inverse problem, Nature Physics, 2, 492–494, 2006.

Thiele, S. T., Jessell, M. W., Lindsay, M., Wellmann, J. F., and Pakyuz-Charrier, E.: The topology of geology 2: Topological uncertainty, Journal of Structural Geology, 91, 74–87, https://doi.org/10.1016/j.jsg.2016.08.010, 2016.

Thiele, S. T., Grose, L., Cui, T., Cruden, A. R., and Micklethwaite, S.: Extraction of high-resolution structural orientations from digital data: A Bayesian approach, Journal of Structural Geology, 122, 106–115, https://doi.org/10.1016/j.jsg.2019.03.001, 2019.

Thum, L. and De Paoli, R.: 2D and 3D GIS-based geological and geomechanical survey during tunnel excavation, Engineering Geology, 192, 19–25, https://doi.org/10.1016/j.enggeo.2015.03.013, 2015.

Torabi, A., Alaei, B., and Libak, A.: Normal fault 3D geometry and displacement revisited: Insights from faults in the Norwegian Barents Sea, Marine and Petroleum Geology, 99, 135–155, https://doi.org/10.1016/j.marpetgeo.2018.09.032, 2019a.

Torabi, A., Johannessen, M. U., and Ellingsen, T. S. S.: Fault Core Thickness: Insights from Siliciclastic and Carbonate Rocks, Geofluids, 2019, 1–24, https://doi.org/10.1155/2019/2918673, 2019b.

Turner, A. and Gable, C.: A review of geological modeling, Tech. rep., 2007.

Vollmer, F. W.: C program for automatic contouring of spherical orientation data using a modified Kamb method, 21, 31–49, https://doi.org/10.1016/0098-3004(94)00058-3, 1995.

Vollmer, F. W.: Orient 3.8.0 Spherical projection and orientation data analysis software user manual, www.frederickvollmer.com, 2018.

Walker, W., Harremoës, P., Rotmans, J., Van Der Sluijs, J., M.B.A., V. A., Janssen, P., and Krayer Von Krauss, M.: Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support, Integrated Assessment, 4, 5–17, https://doi.org/10.1076/iaij.4.1.5.16466, 2003.

Walsh, J. J. and Watterson, J.: Analysis of the relationship between displacements and dimensions of faults, Journal of Structural Geology, 10, 239–247, https://doi.org/10.1016/0191-8141(88)90057-0, 1988.

Waters, C., Terrington, R., Cooper, M., Raine, R., and Thorpe, S.: The construction of a bedrock geology model for the UK: UK3D_v2015, Tech. rep., British Geological Survey, http://nora.nerc.ac.uk/id/eprint/512904/1/UK3DMetadataReport{_}Submissioncopy.pdf, 2015.

Watson, C., Richardson, J., Wood, B., Jackson, C., and Hughes, A.: Improving geological and process model integration through TIN to 3D grid conversion, Computers and Geosciences, 82, 45–54, https://doi.org/10.1016/j.cageo.2015.05.010, 2015.

Wellmann, F. and Caumon, G.: 3-D Structural geological models: Concepts, methods, and uncertainties, vol. 59, Elsevier Inc., 1 edn., https://doi.org/10.1016/bs.agph.2018.09.001, 2018.

Wellmann, J. F. and Regenauer-Lieb, K.: Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models, Tectonophysics, 526-529, 207–216, https://doi.org/10.1016/j.tecto.2011.05.001, 2012.

Wellmann, J. F., Horowitz, F. G., Schill, E., and Regenauer-Lieb, K.: Towards incorporating uncertainty of structural data in 3D geological inversion, Tectonophysics, 490, 141–151, https://doi.org/10.1016/j.tecto.2010.04.022, 2010.

Wellmann, J. F., de la Varga, M., Murdie, R. E., Gessner, K., and Jessell, M.: Uncertainty estimation for a geological model of the Sandstone greenstone belt, Western Australia – insights from integrated geological and geophysical inversion in a Bayesian inference framework, Geological Society, London, Special Publications, 453, 41–56, https://doi.org/10.1144/SP453.12, http://sp.lyellcollection.org/lookup/doi/10.1144/SP453.12, 2018.

Whitmeyer, S. J., Pyle, E. J., Pavlis, T. L., Swanger, W., and Roberts, L.: Modern approaches to field data collection and mapping: Digital methods, crowdsourcing, and the future of statistical analyses, Journal of Structural Geology, 125, 29–40, https://doi.org/10.1016/j.jsg.2018.06.023, 2019.

Wood, R. and Curtis, A.: Geological prior information, and its application to geoscientific problems, in: Geological Prior Information: Informing Science and Engineering, pp. 1–14, https://doi.org/10.1144/GSL.SP.2004.239.01.01, 2004.

Yang, L., Hyde, D., Grujic, O., Scheidt, C., and Caers, J.: Assessing and visualizing uncertainty of 3D geological surfaces using level sets with stochastic motion, Computers and Geosciences, 122, 54–67, https://doi.org/10.1016/j.cageo.2018.10.006, 2019.

Zhong-Zhong, C.: The estimation of digitizing error and its propagation results in GIS and application to habitat mapping, Ph.D. thesis, University of Massachusetts Amherst, https://scholarworks.umass.edu/dissertations/AAI9524686, 1995.

Zhou, W.: An Outlook of GIS Applications in Mineral Resource Estimation, in: Gold Mining: Formation and Resource Estimation, Economics and Environmental Impact, edited by Corral, M. D. and Earle, J. L., pp. 33–62, Nova Science Publishers, Inc., New York, 2009.

Zhou, W. and Maerz, N. H.: Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes, Computers Geosciences, 28, 827 – 839, https://doi.org/https://doi.org/10.1016/S0098-3004(01)00111-X, 2002.

Zhou, W., Chen, G., Li, H., Luo, H., and Huang, S. L.: GIS application in mineral resource analysis–A case study of offshore marine placer gold at Nome, Alaska, Computers and Geosciences, 33, 773–788, https://doi.org/10.1016/j.cageo.2006.11.001, 2007.

Zhou, W., Minnick, M. D., Mattson, E. D., Geza, M., and Murray, K. E.: GIS-based geospatial infrastructure of water resource assessment for supporting oil shale development in Piceance Basin of Northwestern Colorado, Computers and Geosciences, 77, 44–53, https://doi.org/10.1016/j.cageo.2015.01.007, 2015.

Zhu, L.-f., Li, M.-j., Li, C.-l., Shang, J.-g., Chen, G.-l., Zhang, B., and Wang, X.-f.: Coupled modeling between geological structure fields and property parameter fields in 3D engineering geological space, Engineering Geology, 167, 105–116, https://doi.org/10.1016/j.enggeo.2013.10.016, 2013.