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.*

45   – *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.
60   *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
65 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
70 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:

75   **. . . 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), maxi-
80 mum 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° 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
85 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.

90

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.

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

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

(1)

95

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

100 distribution to be parameterized with a mean of zero and a standard deviation equal to $maximum error/3.92$. 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 termi-
105 nation 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 $Aspect ratio$ 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
110 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. $Aspect ratio$ 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
115 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 $\sigma = 2m$.

120 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
125 (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.

130 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
135 *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
185 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
190 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

195

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:

200   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:

205

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.

210

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:

215   The reference to Wellmann and Caumon, 2018 has been added here as the authors agree that it is a fundamental work to-
wards 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:

220

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 be-
coming 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.

270 **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:

275 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:

280 ~~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.*

285 **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.

290 **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.

295 **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 300 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 305 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.*

310 **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.
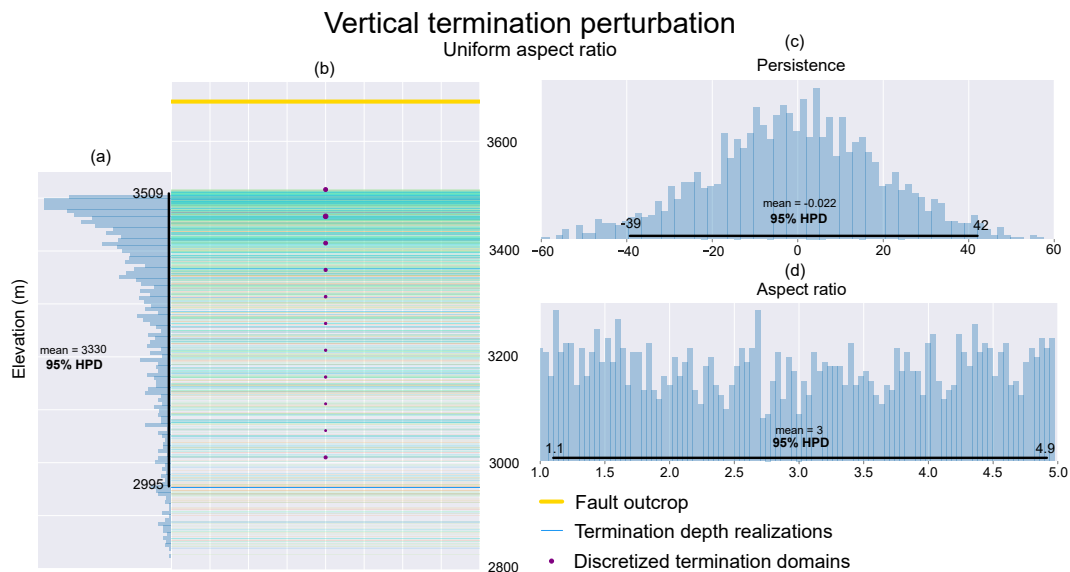

Figure 1. ~~Realizations and Monte Carlo analysis results (trace plot and posterior histograms)~~ Visualization of Monte Carlo samples and associated geologic input realizations from perturbation of the fault zone vertical termination depth based on a uniform distribution of fault aspect ratio. The 95%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

420   **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 model-
425   ing 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:

430   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, reparame-
terizing individual aspects of the probabilistic model may prove to be insufficient due to the presence of inherently unknown
435   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
440   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 in-
puts 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.

445

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° when the surface trace azimuth
450   (i.e., average normal to the fault surface trace) and global trend dip azimuth differed by greater than 20° (Krajnovich et al.,
2020). Comparing two geologic models generated with orientations sampled from anisotropic and isotropic Bingham distribu-
tions 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 parameter-
455   ized with an anisotropic Bingham distribution, allowing for increased variability in the dip angle without compromising the
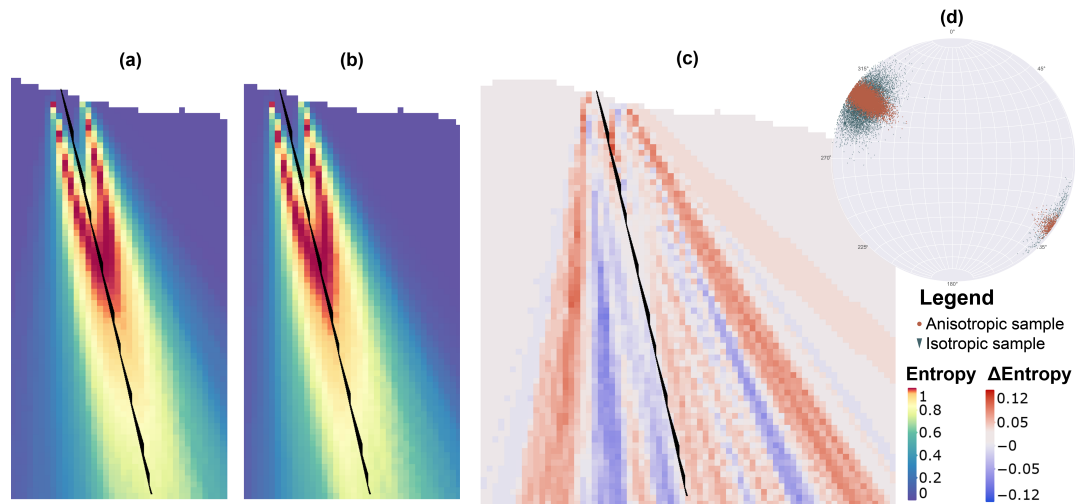certainty of the dip azimuth.

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:

485

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

490 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.

495 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:

500 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.

505 **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:

510

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

13

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.