

Response to RC1 – Ahston Krajnovich

AR = Authors response

Thank you very much for your constructive review of our manuscript. We have incorporated many of your suggested improvements into our manuscript and believe this has significantly improved its quality and readability.

Most importantly, we have added our reasoning behind the choice of prior parameters and our reasoning behind choosing a known topology graph as our constraint. Additionally, we have revised the mathematical notation throughout the paper to make it consistent and in line with the cited literature and have edited so that definitions of technical terms appear before use of them. The placement of the figures will be subject to the final paper typesetting done by the journal and is thus not final in the current manuscript. We have addressed many of the suggested improvements to language, grammar and figure annotations to improve readability.

Please find all our detailed responses to your comments in the supplementary material, along with both the revised and change-tracked manuscript.

Specific Comments

Title: Consider rephrasing to avoid the repetitive use of the word “using”. I would suggest: “Constraining stochastic 3-D structural geological models with topology information using Approximate Bayesian Computation in GemPy 2.1”

AR: We have changed the title.

Abstract: As the research is built in the GemPy environment, it would be beneficial to highlight it’s usage in the abstract (perhaps at Line 13).

AR: We now mention GemPy in the abstract to improve clarity.

Line 129: Sentence requires revision to be accurate about what the likelihood function represents in Bayes’ theorem. I suggest: “This updating process relies on the use of a likelihood function $p(y|\theta)$, representing the conditional probability of the observed data y given the prior probability of the underlying parameter θ and the theoretical connection to the occurring event.”

AR: We have incorporated your suggestion into the manuscript to improve clarity for the reader.

Line 144: You have reversed the conditional probability described by the likelihood function, which is: the likelihood for observing the data y , given the model based on uncertain parameters θ .

AR: Thank you for pointing out this error, we have switched that around!

Line 147: This is unclear, as likelihood functions are inherently encoding information regarding not just the parameters θ , but also the observations y and the assumed theoretical relationship between θ and y . Consider removing or revising.

AR: We have removed the sentence to avoid confusing the reader.

Section 2.3.2: This section requires additional clarification between "observed data"

and "simulated data". Refer to the treatment of ABC in Gelman et al., 2004 where y is the observed data (observed "summary statistic" in ABC) and y -rep is the simulated data (simulated "summary statistic" in ABC). The use of y -hat to represent the observed summary statistic and y to represent the simulated summary statistic creates additional confusion (as the observed data introduced in Bayes' theorem were defined as y , not y -hat).

AR: Thank you for pointing out this mistake. We have changed the notation to be in line with the literature and our description of Bayes' theorem.

Line 156-157: Perhaps add a reference to (Wood and Curtis, 2004)? (Geological prior information, and its applications to geoscientific problems)

AR: Added reference to provide the reader with additional literature to understand the issue of specifying likelihood functions in geology.

Line 160: Please add an additional clarifying sentence on what the summary statistic is in this work rather than the short parenthetical (to avoid confusion with typical summary statistics like mean, mode, median etc.). Also, a comment: In the proposed (approximate) inference scheme, the new evidence y (or data) is the "summary statistic". So, while the definition of the additional term "summary statistic" to describe " y " is useful for highlighting the approximate nature of ABC, the equivalency of these two terms should be clarified for the reader.

AR: We added additional clarification on what is usually used as summary statistics, and why we use topology graphs when comparing geomodels.

"While summary statistics are often measures such as the mean, mode or median of a model, they tend to be meaningless in summarizing geomodels. In this work we use the geomodel topology graph as a summary statistic of the geomodel to provide a meaningful comparison between geomodels."

Line 162-163: Clarify the 2nd part of the sentence to illustrate that the "observed summary statistic y -hat" is static for the entire geomodel ensemble (i.e., the known, observed topology graph), while "the summary statistic y " is tied to each individual geomodel realization (i.e., a simulated topology graph).

AR: We have fixed the mathematical notation from y to $S(y)$ when referring to the summary statistic. We think this fixes the problem of clarity in this sentence. While we keep the observed topology graph static in this experiment, this is by no means necessary, as we discuss in the paper. We have also improved our discussion of this.

Line 165: Theta-prime has not been introduced. What does it refer to as opposed to theta? I assume you are referring to a single draw from the parameter distribution theta, but please clarify. When relying on mathematical notations from another work (the ones in question here seem to be borrowed from Sadegh and Vrugt, 2014), make sure notations are introduced properly. It also helps to also have a "sanity check" to make sure that the notation used is not confusing with respect to the broader statistical literature (e.g., where the observed data in Bayes theorem are typically represented without a $\hat{\cdot}$ or \prime)

AR: Theta has been introduced in the previous section as the model parameter distributions. We have added an explanation that theta prime is a sample from these distributions.

Section 2.5: Section could be made much more concise to avoid excessive overlap with existing works (seeing as the major contributions of the paper are not focused on novel applications of Shannon entropy).

AR: We have cut detailed explanations of the Shannon entropy and refer the reader to the relevant literature.

Line 227: How and why were the prior uncertainty ranges chosen? Were they considered to be broad, non-informative priors, derived empirically, based on background information or simply assumed by the modeler for the sake of simulation? Same question should be addressed more directly for the Gulfaks case study as well (Line 249-251), where the uncertainties appear to be derived from the referenced work though this is not stated definitively. Also, just a comment: I am quite interested to see how incorporating structural uncertainty (by way of the methods put forth by Pakyuz-Charrier et al., 2018a,b, Roberts et al., 2019 or Krajnovich et al., 2020) would influence the geomodel topology.. Intuitively, there is a high potential for confounding effects on the range of possible geomodel topologies when interface location and interface/fault orientation are varied together!

AR: We have added explanation on how we chose the prior parametrization. As this paper focusses on developing and showcasing a new methodology for constraining uncertain geomodels using topology graphs, prior parametrization has not been a focus of the experiments. We refer the reader to other works, as also mentioned by the reviewer.

Line 246: How was the interface uncertainty applied to the surface points? Independently at each node, or generally to the set of surface points (so as to retain surface shape). From reading into the supplemental codes, it appears that the uncertainty was applied to the group of surface points – but this information needs to also be included in the text for the typical reader. This also applies to the synthetic model, which appears (from the code provided) to have been modeled from similar groups of surface points, though this is not clarified in the text.

AR: We have added the missing description on how the interface point uncertainty has been applied in both the synthetic and the real-world examples.

Line 251: Tying back to the earlier comments on how prior uncertainty ranges were chosen, I believe that “ease of implementation” is somewhat of an inconclusive reasoning. The rest of the sentence provides more meaningful perspective but still could be expanded upon (e.g., what is “simplified uncertainty modeling” in this context?). Please add some more detail.

AR: We have addressed our inconclusive writing and now more clearly describe how the model is perturbed and why we chose this approach.

Line 268: A figure representing this most frequent topology graph from simulation (or other selected simulated topology graphs) would be quite insightful, especially if accompanied by a discussion of their geologic significance (e.g., tying back to points made during the introduction (Line 51), did any simulated topology graphs represent a

compressional rather than extensional tectonic regime?). If length permits of course - perhaps if some figures are combined or suggested section lengths reduced, this could be added.

AR: This could be beneficial, but we think it would distract the reader from the main message of the manuscript: the method itself. Analysing ensembles of topologies for geological setting would require either painstaking manual evaluation of every model or require more research into defining how extension settings can be detected reliably from topology graphs. This would indeed be a very interesting topic, but out of the scope of this research.

Line 287: Since the Jaccard Index used could allow for multiple topologies to be present in the final model ensemble (depending on the rejection threshold used), it would be beneficial to see some exploration of what these possible model topologies looked like (how geologically unrealistic do they get? Are all 675 unused topologies absolutely unrealistic?). Including a discussion of this sort would help guide future works investigating uncertainty of the applied topology information itself (without requiring reproducing the results to show geomodel uncertainty when multiple simulated topologies were present in the final ensemble). See also Comment for Line 268.

AR: Please see our answer to the previous comment.

Line 295: In line with the missing clarification regarding the assumption of the observed topology graph being known without uncertainty, add some clarification behind the reasoning for setting the rejection threshold such that only the applied initial topology remains in the probabilistic geomodel ensemble. Was the goal of empirical testing of thresholds to find the largest threshold which resulted in only a single model topology remaining across the probabilistic geomodel ensemble?

AR: Our aim was indeed to show how to constrain a stochastic geomodel to a topological state --- to allow for the reliable simulation of uncertainty within a single topology state (kind-of like a single conceptual model). We think that more research into how to identify, and thus compare and constrain with, geologically similar geomodels from topology graphs is needed to allow the meaningful relaxation of the error threshold.

Line 297: How does simulation time for ABC-REJ compare to simulation time for the standard MC approach?

AR: We don't believe this comparison is very meaningful, as this is highly dependant on the choice of error, summary statistic etc. That's why we chose to only compare the ABC-REJ with the ABC-SMC – as they are aiming for the same constrained outcome, while the Monte Carlo forward

Line 298: This is a significant improvement in efficiency! Perhaps include a description of acceptance rates from each epoch of SMC, or at least a comparison of the final acceptance rate at the threshold value of 0.025 in SMC for comparison with the rate given for REJ. This information might fit naturally in Figure 12.

AR: We think that acceptance rates for the multiple epochs of the ABC-SMC are difficult to compare to the single acceptance rate of ABC-REJ. One could compare the average, or weighted average, but we believe that the comparison of overall simulation time is more meaningful in this case.

Line 324: If the information applied were non-meaningful (e.g., an incorrect topology graph), the geomodel ensemble would likely still exhibit a reduction in entropy due simply to the convergence of the model realizations towards the single model topology applied. That is, the reduction in uncertainty is arising from the reduction of possible model topologies, not necessarily the meaningfulness of the model topology used in the ABC algorithm.

AR: We fully agree, which is why we highlight the fact that only a meaningful constraint can lead to a meaningful reduction in model uncertainty.

Line 334: It appears that expanding the ABC approach proposed here to incorporate multiple observed topology graphs would not be a matter of "easily scaling". Revise to clarify that the general ABC framework would definitely allow for this, although it would require reparameterizing the current summary statistic and discrepancy measure (distance function), and also possibly changing the simulation method (as mentioned in Line 357-359).

AR: Incorporating multiple observed topology graphs for comparison in the ABC framework would indeed scale easily, as the computational complexity would scale linearly. Thus, a doubling in comparisons would double the Jaccard index computation, which by itself is trivial in terms of computation cost for such small networks in comparison with the computation cost of generating the geomodel in the first place. In the current implementation it would simply requiring looping over a set of topologies and computing the Jaccard index and accepting if one of them is below the allowed error threshold.

Line 335: This would be a good place to bring up again the implications of using the demonstrated ABC approach if there were uncertainty about the observed topology graph.

AR: The uncertainty in observed topology can be addressed via proposing several acceptable topologies or by scaling the error threshold.

Line 345: “: :reducing the parameter dimensionality” – how so? The number of input parameter probability distributions is the same in standard MC or in ABC-REJ/SMC. The computation efficiency improvements arrive from reducing the number of input parameter draws that are run through uncertainty propagation to the 3D geologic model space, which in SMC also allows for reducing the size of the uncertainty space (note, not the parameter dimensionality) iteratively.

AR: We describe here a trade-off that could be made between uncertain geomodel parametrization and probability of subsequently simulated geomodel samples being valid geological models. Increasing geomodel complexity generally requires increased geomodel and statistical model parametrization – which thus scales the parameter space exponentially (see e.g. Betancourt, Michael. "A conceptual introduction to Hamiltonian Monte Carlo" (2017).). This increase in parameter space will drastically increase computational time. Thus, a balance generally needs to be made between model complexity and model parametrization. Our method could allow for lower levels of model parametrization while retaining model complexity by essentially filtering topologically wrong samples (which will inherently become more numerous when stochastically perturbing an under-parametrized complex geomodel).

Line 370: This was not discussed earlier in Section 3.2 when the acceptance rate was initially 0.0059 (0.59%). Does that low acceptance rate warrant reassessing the

prior input uncertainties used in the probabilistic geomodeling? Should be discussed to better frame the current work and guide future work.

AR: It might! It definitely warrants a good look at the prior parametrization. But stochastic geomodels have such inherent topological complexity, as minor changes in the location of interfaces across a fault can have various effects on model topology.

Figure 4: Consider replacing X & Y with N & E to be more intuitive for geoscientists. Applies to all figures of geomodels with labeled axes.

AR: We have chosen non-descript axis labels for these figures as this model is entirely synthetic.

Figure 6: In my opinion, the XZ difference section (and possibly then also XY and YZ sections) from Figure 7 could be appended onto Figure 6 for ease of reference. Also, what do the overlain crosshairs show?

AR: The lines show the locations of the respective other sections. We have added an explanation to the figure description to make this clearer.

Figure 8: Figure does not show (a), (b), (c): : tags. Also, as mentioned in the comment for Line 279, the figure does not show histograms.

AR: We have added subfigure labels a-f and we have removed the outdated reference to histograms to only KDEs.

Figure 10: The significant reduction in model entropy indicates the strong dependence on the initial topology used - this potential source of bias should be addressed. Please discuss the implications of using a rejection threshold which only allows one model topology across the entire final ensemble of geomodel realizations. Since the authors are operating under this (valid) assumption, it needs to be clearly stated earlier that the initial geomodel topology is "known" and treated without uncertainty. See also comments regarding Lines 324, 295, 287 and 268. Also, I believe this figure could be merged with Figure 11.

AR: We have added explanation of this bias in our elaboration on why we chose to constrain with a single topology state in our method description. We hope this clears things up for the reader.

Figure 12: Figure needs correction to show Y-axes labels. Perhaps acceptance rates per epoch would be useful to add as well, as they are tied to the processing efficiency improvement of 10.1x (see comment regarding Line 298).

AR: We have added Y-axes labels and removed then redundant titles. We are not sure how acceptance rate would improve would improve the bar chart. We have chosen not to add it to keep the figure concise.

Response to RC2 – Anonymous Reviewer

AR = Authors response

Thank you very much for your constructive review of our manuscript. We have incorporated many of your suggested improvements into our manuscript and believe this has significantly improved its quality and readability.

We have addressed many of the reviewer's suggestions to improve wording and grammar throughout the manuscript. We have mostly removed the term "likelihood-free" from the manuscript to not confuse the reader, except where we directly reference literature that uses this term. We have elaborated on our choice of prior parametrization, have improved the mathematical notation of the statistical methods used so that the reader can more clearly differentiate between what our priors are and what our constraint is (i.e. the ABC-equivalent to the likelihood). We have not significantly cut the ABC method description, as we believe it will be valuable to the geoscientific reader to read about it in the context of geological modeling.

Please find all our detailed responses to your comments in the supplementary material, along with both the revised and change-tracked manuscript.

General Comments

The reviewer writes: "ABC (both rejection and SMC) approximates likelihood in the following ways: Firstly, exact conformity to the the predictions of the model (θ) is relaxed using a distance measure and threshold. Secondly, if the simulation linking cause (θ) to observation (y) is stochastic then it uses a finite set of MC realisations instead of an integral over all outcomes of the random variables not of interest. Neither of these properties are used in the work presented here."

The reviewer confuses us with their statement: We relax exact conformity by using a distance measure and threshold (we use the Jaccard index as a distance measure between two topology graphs and allow an error threshold). The reviewer further states that "Nothing what separates ABC from the traditional Bayesian approaches is used here.", but then goes on to state that "[...] it is ABC only in the most superficial sense", providing us with a logical contradiction. It would have been helpful for us if the reviewer had provided us with literature references to properly understand their arguments.

We have mostly removed the use of the term "likelihood-free" when referring to ABC, although this is often stated as such in literature. We have also clarified our choice of prior parametrization, as we agree that this will help the reader better understand the experiments despite the focus of the paper on the method itself.

Specific Comments

1) Remove claims all of circumventing or simplifying specification of topological knowledge due to ABC, these are untrue for the constraint presented here. I have highlighted these in the attached annotated pdf along with more detailed comments for each. Lines 60, 150, 155, 165.

AR: We have reworded our description of topological knowledge to make it more clear to the reader how we acquired it (through interpretation). In the supplement the reviewer writes "ABC is not a method for circumventing the subjectivity of specifying priors or likelihoods. It is likelihood-

free only in the sense of a likelihood never being explicitly calculated due to the nature of the approximation built into the samplers.”, **which we agree with. We hope our changes clarify this to the reader, as we never intended to claim that using ABC over conventional Bayesian inference makes the method somehow more objective.**

In the supplement the reviewer further notes that “the implicit probability function you use is simply the one that assigns uniform probability to all theta within the region where $d < \epsilon$.”. But as we *always* accept our parameter samples if they generate a model with a summary statistic within error, we would think that this does not constitute a uniform probability density function, as the area of the curve would be above 1.

2) Remove all uses of the term 'likelihood-free'. There is no place in this article where its use helps clarify how the proposed approach works. Its only effect is as a potential source of misinterpretation. To avoid unneeded additional review rounds I am asking for complete removal and not fixing its use.

AR: We have removed the term “likelihood-free” in the manuscript to avoid confusion of the reader. It remains in two places where we directly refer to literature that describes Bayesian optimization for likelihood-free inference (Gutmann and Corander, 2016) and where we point the reader to literature that specifically describes ABC as a “likelihood-free” method (Marin et al, 2012). Again, it would have been really helpful for us if the reviewer had provided literature references for us to understand their issue with this specific term.

3) Do not refer to the initial connectivity graph as an observation. It is a semi-subjective semi-empirically derived parameter to a subjectively chosen constraint family. Describe briefly how it is obtained and what the reasoning behind using the 'initial' graph was.

AR: We have referred to the initial topology graph as an observation, as this is the terminology used in literature. We understand that this could potentially lead to certain confusion with the reader. We have thus added the clarification to the manuscript, that we treat our subjectively chosen topology graph as the observation in the terminology of ABC.

4) I am not asking that you replace 'likelihood' with 'prior' when referring to your topological constraint. Instead, please state somewhere that you choose to go with this label but that it could also be considered a prior or empirical prior and that the application of these terms is not always clear cut. State that you are simply treating the adjacency graph (y) as an observation.

AR: We already describe the topology graph (adjacency graph) as an observation --- something the reviewer has advised against in their previous comment.

5) Two topology distance measures are defined in section 2.4 which are never used. Since they are never discussed, analysed or compared, they serve no purpose. I suggest you remove them to simplify and shorten the already long paper, but feel free to ignore this suggestion.

AR: They are not used, but we think it serves the purpose of helping the reader think about different possible approaches to constrain geomodels using graph structures --- which we think has value.

6) Remove the mention of fuzzy sets in section 2.5, it is not relevant. Your posterior is probabilistic not fuzzy. It represents degrees of certainty concerning a single underlying truth, not degrees of membership to a category.

AR: We have removed the detailed explanations of information entropy in Section 2.5 and refer the reader to the relevant literature instead. Information entropy is used for assessing the prior and approximate posterior geomodel grids, not the posterior parameter distributions themselves.

7) Several sentences need to be reworded for clarity or readability. I have highlighted these in the annotated pdf. Please try to address most of them.

AR: We have reworded many parts of the manuscript given the suggestions by both reviewers.

References

Marin, J.-M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, 22(6), 1167–1180.

Gutmann, M. U., & Corander, J. (2016). Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models. *Journal of Machine Learning Research*, 17, 1–47.

Constraining stochastic 3-D structural geological models with topology information using Approximate Bayesian Computation in [GemPy 2.1](#)

Alexander Schaaf^{1, 2}, Miguel de la Varga², Florian Wellmann², and Clare E. Bond¹

¹Geology and Petroleum Geology, School of Geosciences, University of Aberdeen, AB24 3UE, UK

²Computational Geoscience and Reservoir Engineering, RWTH Aachen University, Aachen, Germany

Correspondence: a.schaaf@abdn.ac.uk

Abstract. Structural geomodeling is a key technology for the visualization and quantification of subsurface systems. Given the limited data and the resulting necessity for geological interpretation to construct these geomodels, uncertainty is pervasive and traditionally unquantified. Probabilistic geomodeling allows for the simulation of uncertainties by automatically constructing ~~geomodels~~ [geomodel ensembles](#) from perturbed input data sampled from probability distributions. But random sampling of input parameters can lead to construction of geomodels that are unrealistic, either due to modeling artefacts or by not matching known information about the regional geology of the modeled system. We present here a method to incorporate geological information in the form of [known](#) geomodel topology into stochastic simulations to constrain resulting probabilistic geomodel ensembles [using the open-source geomodeling software GemPy](#). Simulated geomodel realisations are checked against topology information using ~~a likelihood-free~~ Approximate Bayesian Computation approach, [to avoid the specification of a likelihood function](#). We demonstrate how we can learn our input data parameter (prior) distributions on topology information in two experiments: (1) A synthetic geomodel using a rejection sampling scheme (ABC-REJ) to demonstrate the approach; (2) A geomodel of a subset of the Gullfaks field in the North Sea, comparing both rejection sampling and a Sequential Monte Carlo sampler (ABC-SMC). ~~We also discuss possible speed-ups of using~~ [Possible improvements to processing speed of up to 10.1x are discussed, focusing on the use of](#) more advanced sampling techniques to avoid [the](#) simulation of unfeasible geomodels in the first place. Results demonstrate the feasibility ~~to use topology of using topology graphs~~ as a summary statistic, to restrict the generation of ~~model ensembles with additional~~ [geomodel ensembles with known](#) geological information and to obtain improved ensembles of probable geomodels [which respect the known topology information and exhibits reduced uncertainty](#) using stochastic simulation methods.

Copyright statement. TEXT

1 Introduction

Structural geomodeling is an elemental part of visualizing and quantifying geological systems (Wellmann and Caumon, 2018). Topology relationships in geological systems (e.g. how layers are connected to each other stratigraphically, or their across-fault connectivity) are important constraints for fundamental geological processes, such as fluid or heat flow (Thiele et al., 2016a, b). Each unique interpretation (model) of a geological setting has a specific topology graph. And as geology is not only an experimental science, but also an interpretive and historical science (Frodeman, 1995), the deduction of the geomodel - often from sparse amounts of data - can inherently lead to numerous potentially valid geological interpretations (Bond et al., 2007), which themselves can lead to equally numerous topology graphs. This aspect is compounded by the complex nature of geological systems and interpretation bias imparted by geoscientists in the explicit creation of geomodels (Bond et al., 2007; Polson and Curtis, 2010; Bond, 2015). It also leads to the creation, and favouring, of specific models that fit expectations and prior knowledge (Baddeley et al., 2004), rather than consideration of the full range of possible models. However, methodologies to create models often focus on the creation of a single deterministic model (Bond et al., 2008) and a lack of systematically considering lack of systematic consideration of data uncertainty (Thore et al., 2002; Tacher et al., 2006; Bardossy and Fodor, 2013). These facts call for the development of alternative approaches. The increasing development of implicit modeling algorithms (Mallet, 2004; Hillier et al., 2014; Laurent et al., 2016) allows for the creation of vast structural geomodel ensembles by making use of interpolation functions, which makes the analysis and visualization of uncertainty using probabilistic simulation approaches possible (Bistacchi et al., 2008; Suzuki et al., 2008; Wellmann et al., 2010; Lindsay et al., 2012; Wellmann and Regenauer-Lieb, 2012; Wellmann, 2013).

~~But the~~ The mathematical nature of implicit modeling, in combination with the use of a probabilistic modeling process, often leads to geologically unsound model realizations and modeling artifacts. Additionally, the modeling algorithms only take a limited set of input data types, e.g. layer interface locations and structural orientation data, which significantly limit limits the amount of geological information that can be included in the modeling process. de la Varga and Wellmann (2016) and Wellmann et al. (2017) showed how Bayesian inference can be used to reduce uncertainty and modeling artifacts in both synthetic and real, implicit, structural geomodel ensembles. Their concept uses supplemental geological information (e.g. layer thicknesses or fault offsets) in the form of likelihood functions to constrain stochastic geomodel ensembles. ~~But the question of how to acquire suitable likelihood functions for specific geological systems and diverse types of prior geological knowledge and reasoning remains.~~ Likelihood functions essentially represent information in a probabilistic mathematical form. This information can be available numerical data, such as information about the range of possible layer thicknesses in a depositional setting. While demonstrated in these specific cases, the general question of how to acquire suitable likelihood functions for specific geological systems and diverse types of prior geological knowledge and reasoning remains.

~~But geological~~ Geological expert knowledge contains much more information that is vital to model creation, such as understanding the geological processes that result in the thickening and thinning of sedimentary deposits and their relative spatial distribution. One key knowledge-based input into geomodeling is the understanding of the kinematic evolution of the rock units into their present configuration. While kinematic modelling software exists (see Groshong et al., 2012; Brandes and

Tanner, 2014, for reviews), it is limited to ‘end-member’ kinematic models’ resulting in geometrical deformations defined by few parameters, and not taking into account a range of other factors, not least of which being the mechanics of the different units (Butler et al., 2018). But we can capture certain kinematics using topology information—for example the across-fault connectivity of layers, where extensional deformation leads to fundamentally different topological relationships than does
5 compressional deformation (see Fig. 1).

We therefore hypothesize that topological information about a geological system can be used as a meaningful constraint for probabilistic 3-D geomodeling outputs.

~~But this~~ This topological information is difficult to incorporate into the mathematical foundations of implicit modeling functions and is highly case-dependant. As the origin of topological information is generally qualitative, obtaining a suitable likelihood function that can be used in a Bayesian inference is considered intractable, apart maybe from time- and cost-consuming expert elicitation (Curtis and Wood, 2004). This work tries instead to approximate the (Bayesian) posterior geomodel ensemble that incorporates both the geological input data and the topology information using an Approximate Bayesian Computation (ABC) approach ~~for a likelihood-free approximation of the posterior~~ to approximate the posterior without specifying a likelihood function directly.
10

15 To test this approach we designed two distinct experiments, one synthetic and one case study:

1. We construct a synthetic fault model and explore its topological uncertainty. We do this by describing our input data not as fixed parameters, but as probability distributions. We then use Monte Carlo sampling to obtain input data realisations from which geomodels are constructed. We then show how a single topology graph can be used as a summary statistic in an ABC-rejection scheme to approximate the posterior model ensemble that honours the added information.
- 20 2. To test the same ABC approach on a real-world dataset, we apply it to a model extracted from a seismic interpretation of the North Sea Gullfaks field. We also explore a more advanced sampling technique to demonstrate possibilities for reducing the computational costs of the method.

In the following section we will give an overview of the applied implicit geomodeling approach, the basic concept of Bayesian inference and its use in probabilistic geomodeling, as well as the ~~idea~~ theory behind Approximate Bayesian Computation. We further describe how we analyze model topology and use it as a summary statistic. We will then introduce, in detail,
25 both the synthetic fault model and the case study, followed by a comprehensive discussion of our findings.

2 ~~Materials and Methods~~ Methodology

2.1 Implicit Geomodeling

Several approaches exist for creating structural geomodels, which can be separated into three main categories: (a) interpolation,
30 (b) kinematic methods and (c) process simulation. The interpolation of surfaces and volumes from spatial data is currently the most widely used approach in geosciences, ~~especially manually~~ typically performed manually by geoscientists, which requires

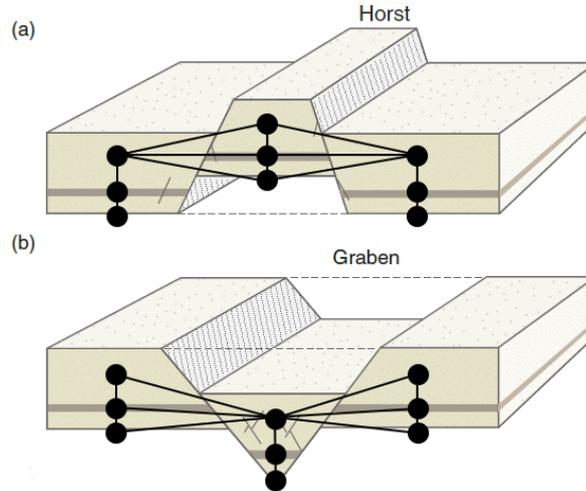


Figure 1. Idealized Horst (a) and Graben (b) structures with topology graph overlay, showing the difference in graph structure for different tectonic settings (modified from Fossen, 2010). The black nodes represent the centroids of the geobodies and the black edges the topology connections, together building a topology graph.

robust knowledge of the geological setting and extensive amounts of data in order to robustly approximate reality. Additionally, highly complex structures such as extensive fault networks and repeatedly folded areas are challenging to recreate using current interpolation methods (Jessell et al., 2014; Wellmann et al., 2016; Laurent et al., 2016).

The open-source, Python-based implicit modeling package *GemPy*¹ (de la Varga et al., 2019) is used here. It is based on the work of Lajaunie et al. (1997) and Calcagno et al. (2008), and allows the interpolation of geological interface position and plane orientation data by using a scalar field method in combination with cokriging (Chilès et al., 2004). For a detailed overview of the algorithm and the functionality of *GemPy*, we refer the reader to de la Varga et al. (2019).

2.2 Geological Topology

Topology, referring to “properties of space that are maintained under continuous deformation, such as adjacency, overlap or separation” (Thiele et al., 2016a; Crossley, 2006), is a highly relevant concept in structural geology, as it provides a useful description of the relations between stratigraphic units across layer interfaces, faults or the contact to an intrusive body. Generally, eight binary topological relationships can exist between three-dimensional objects (Egenhofer, 1990), while a total of 69 relations are possible between *simple* lines, surfaces and bodies (e.g. surfaces without holes; see Zlatanova, 2000). From these eight Egenhofer-Herring relationships, *meets* (i.e. adjacency) is the most relevant one for describing structural and stratigraphic relationships, such as across-fault connectivity of layers (see Fig. 1). The topology relationships of geological models can be represented by an adjacency graph, which represents topological units as individual nodes and their connections by edges (see Fig. 1). The adjacency topology of geological structures is highly dependent on deformation: compressional deformation

¹URL: github.com/cgre-aachen/gempy

leads to different connectivities in the topology graph than does extensional, but even within the same type of deformation they can lead to different topologies—as visualized by the Horst and Graben structures in Figure 1. Not only does the type of deformation have an important influence on the systems topology, but also the quantity—e.g. the fault throw. For an in-depth introduction and discussion of topology in geology see Thiele et al. (2016a) for the fundamental theory and Thiele et al. (2016b) and also Pakyuz-Charrier et al. (2019) for the influence of structural uncertainty on geomodel topology.

2.2.1 Computing geomodel topology

To compute the geomodel topology with the necessary computational efficiency to conduct a feasible stochastic simulation of realistic geomodels, we implemented a topology algorithm using `theano` (Theano Development Team et al., 2016) into the core of `GemPy`. This enables the topology computation to run alongside the geomodel interpolation on graphical processing units (GPUs). As `theano` is a highly optimized linear algebra library, the employed method is mainly focused on utilizing matrix operations for the computation of the geomodel topology. When the implicit geomodel is discretized using a regular grid, it becomes a 3-D matrix of lithology IDs L (Fig. 2a), which we use for the calculation of the geomodel topology. For each geomodel we also have access to the 3-D boolean matrices F_n for each fault, representing the two sides of the respective fault by two ascending consecutive integers (Fig. 2b). Given these two input data, we compute the geomodel topology as follows:

1. The lithology matrix L and the summed fault matrices $\sum_{i=1}^{n_{\text{fault}}} F_i$, where n_{fault} is the total number of faults in the geomodel, are combined into a matrix where each lithology in each fault block is represented by its own unique integer, referred to as the topology labels matrix T (see Fig. 2c):

$$T = L + n_{\text{lith}} \sum_{i=1}^{n_{\text{fault}}} F_i \quad (1)$$

with n_{lith} being the total number of lithology IDs in the geomodel.

2. The topology labels matrix T is then shifted twice (forward and backward) along each axis X, Y and Z. The two resulting shifted matrices S_1 and S_2 along each axis are then subtracted from each other to result in a difference matrix D , in which only the cells along a lithology or fault boundary are non-zero (Fig. 3).
3. The topology labels matrix T is then evaluated at all non-zero cells of D to obtain the two topology labels n_a, n_b of each topological connection (referred to as an edge e) in the geobody, which are stored in a set of unique edges E representing the geomodels topology. For the example shown in Figure 2 and 3 the abbreviated set is $E = \{(0, 4), (0, 5), (0, 1), \dots, (3, 7)\}$.

This method of topology calculation works on regular grids, which imposes a strong bias on the result: if the main lithological and structural features are not aligned with the grid orientation, the resulting topology graph could thus contain (or miss) connections. For a more detailed discussion on the effects of model discretization see Wellmann and Caumon (2018).

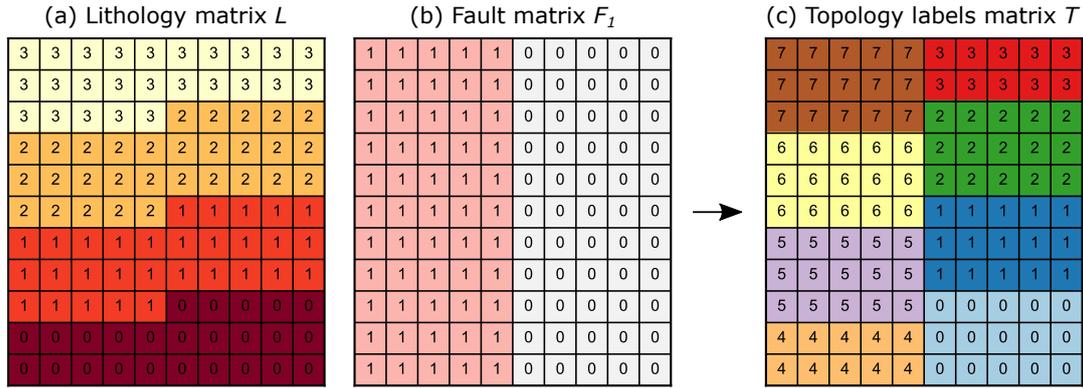


Figure 2. (a) Lithology matrix L of an example 2D geomodel that consists of four layers and a vertical fault in the center; (b) Fault matrix F of the geomodel; (c) Topology labels matrix T of the geomodel.

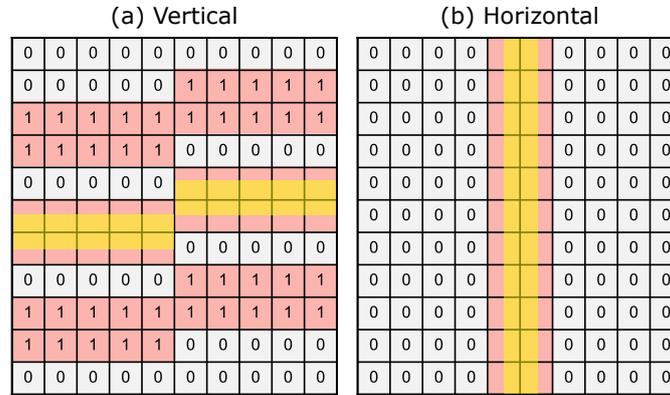


Figure 3. Vertical (a) and horizontal (b) difference matrix D showing all cells (red) in the shifted matrices S_1 and S_2 which are next to the interface between two different layers or of any layers across a fault. The highlighted (yellow) part shows the area in which the implicit interface must be located.

2.3 Stochastic Modelling Approach

2.3.1 Bayesian Inference

Bayesian inference is fundamentally different to the classical frequentist approach of inference. It treats probabilities as *degrees of certainty* of a parameter θ , which is inherently considered to be a random variable itself (Bolstad, 2009; VanderPlas, 2014).

- 5 It is based on *Bayes' theorem* (Eq. 2), which allows ~~to update~~ updating of a given probability - the *prior probability* $p(\theta)$ of a parameter θ - after the occurrence of a connected event (Bolstad, 2009). This updating process relies on the use of a *likelihood function* likelihood function $p(y|\theta)$, representing the ~~probability distribution conditional probability~~ probability distribution conditional probability of the observed data y ~~of the occurring given the prior probability of the underlying paramter θ and the theoretical connection of the occurring~~ of the occurring given the prior probability of the underlying paramter θ and the theoretical connection of the occurring event. It is

used to condition the *prior* into the *posterior distribution* $p(\theta|y)$, which represents the degree of certainty over the parameter θ ~~after given~~ the occurrence of the event and its observed data y .

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta) d\theta} \quad (2)$$

For the use in geomodeling, these parameters can be seen as (de la Varga and Wellmann, 2016; Gelman et al., 2013):

- 5 – **Model parameters** θ : The model-defining parameters (e.g. layer interface positions, dip or fault parameters) used for the interpolation of the geomodel), which can be either deterministic (thus be exactly defined and known) or probabilistic. The latter represent uncertain parameters, which is expressed in the form of probability distributions (e.g. a normal distribution expressing the uncertainty of the vertical subsurface position of a layer interface); ~~We will use θ' as the notation for a sample from these parameter distributions.~~
- 10 – **Observed data** y : Represents additional measurements or observations, which should enhance the model definition by providing additional information with the goal to reduce model uncertainty or enable the comparison of the model to reality (e.g. by comparing geophysical potential-field measurements with the according forward simulation on the basis of a geomodel). In this work we use topology information in the form of a topology adjacency graph as the "observed data";
- 15 – **Likelihood functions** $p(y|\theta)$: These form the relationship between the model parameters θ and the observed data y . Essentially, this function describes the ~~likelihood for the~~ conditional probability for observing the data y given the parameters θ for a given observation y . (e.g. MacKay and Kay, 2003). In the case of structural modeling, this essentially means that we compute the geomodel from the input parameters θ and compare model predictions (e.g. the thickness of a certain layer at a certain position, or topology adjacency graphs), with additional observed data. ~~The likelihood of the parameter θ is then encoded in the likelihood function.~~
- 20

While constructing meaningful likelihood functions for physical properties such as layer thickness or geobody volume from observed data is straight forward (de la Varga and Wellmann, 2016), we have no proper framework to construct them for more abstract or "soft data", such as our understanding of the geological setting, or the topology relationships of our layers across faults or unconformities. For this reason, we chose to ~~pursue a likelihood-free method~~ apply methods to estimate our posterior distributions given abstract geological information without specifying a likelihood function: Approximate Bayesian Computation.

2.3.2 Approximate Bayesian Computation

Geoscientists often have extensive implicit knowledge of the geological settings (e.g. our understanding of the tectonics of a system), but only a limited amount of this knowledge can be incorporated into the geological interpolation function (Wellmann and Caumon, 2018). Additionally, it is often difficult to define formal likelihood functions for geological knowledge, as required for conventional Bayesian inference methods (Wood and Curtis, 2004). A less formal but valid alternative approach

is to approximate the posterior distributions using Approximate Bayesian Computation (ABC) methods. These methods are, also referred to as likelihood-free inference methods (Marin et al., 2012), ABC methods by some (Marin et al., 2012), evaluate the distance of stochastically generated models to our additional data using one or multiple summary statistics S (e.g. model topology), instead of a probabilistic likelihood function. While summary statistics are often measures such as the mean, mode or median of a model, they tend to be meaningless in summarizing geomodels. In this work we use the geomodel topology graph as a summary statistic of the geomodel to provide a meaningful comparison between geomodels.

To obtain the approximate posterior distribution we need to sample from our prior parameter distributions, plug the values samples values θ' into our simulator functions-function y (our geomodeling software), compute the summary statistic $S(y(\theta'))$ (geomodel topology) and evaluate its distance to our observed summary statistic (data) $\hat{y}-S(y)$ (e.g. a geomodel topology graph). The most fundamental sampling scheme for ABC is based on rejection sampling (ABC-REJ; see Algorithm 1), for which the distance between our simulated data $y(\theta')$ (the simulated geomodel) and observed data \hat{y} (initial geomodel) is calculated using a distance function of the summary statistics $d(S(\hat{y}), S(y(\theta')))$ their summary statistics (topology graphs) $d(S(y), S(y(\theta')))$. The simulated model is accepted if the distance is below a user-specified error bound $\epsilon \geq 0$ (Sadegh and Vrugt, 2014), or else rejected. The accepted samples form the approximate posterior. Thus, this method circumvents the need to specify a likelihood function for our additional data, while still approximating the posterior distributions incorporating the information of both our priors and our additional information (Sunnåker et al., 2013). Within this work we use the Jaccard index $(1 - J)$ as a distance function between topology graphs.

Algorithm 1 ABC-REJ

```

for  $i = 0$  to  $N$  do
  while  $d(S(y), S(y(\theta'))) > \epsilon$  do
    Draw sample  $\theta'$  from priors  $p(\theta)$ 
    Simulate geomodel  $y(\theta')$ 
    Compute geomodel topology  $S(y(\theta'))$ 
    Calculate  $d(S(\hat{y}), S(y(\theta')))$   $d(S(y), S(y(\theta')))$ 
  end while
end for

```

A more advanced sampling scheme for ABC is Sequential Monte Carlo sampling (ABC-SMC). In its simplest form it can be seen as an extension of rejection sampling, by chaining rejection sampling simulations together (each referred to as an *epoch*). During the first epoch of rejection sampling, a large error threshold ϵ_1 is used while sampling from the prior distributions $p(\theta)$. The accepted samples, forming the posterior distributions of the first epoch, form the updated priors of the second epoch by replacing the priors with the kernel density estimation $\hat{f}_h(\theta_{accepted})$ of the posterior samples. Iteratively, with every epoch, the error threshold ϵ is reduced to the target value (e.g. $\epsilon = 0$) to obtain the final posterior sample. Thus, every epoch, the sampler 'learns' from the previous epoch by adjusting the prior distributions further towards the posterior distributions. As ABC-REJ tends to suffer from potentially low computational efficiency when using low error thresholds ϵ , the iterative shrinking paired

with adjustment of the prior distributions can potentially obtain the approximate posterior much more quickly. We apply this sampling scheme to our Gullfaks case study to show the potential speed-ups.

Algorithm 2 ABC-SMC

```

for  $\epsilon$  in  $\{\epsilon_1, \epsilon_2, \dots, \epsilon_M\}$  do
  for  $i = 0$  to  $N$  do
    while  $d(S(y), S(y(\theta'))) > \epsilon$  do
      Draw sample  $\theta'$  from priors  $p(\theta)$ 
      Simulate geomodel  $y(\theta')$ 
      Compute geomodel topology  $S(y(\theta'))$ 
      Calculate  $d(S(\hat{y}), S(y(\theta')))$   $d(S(y), S(y(\theta')))$ 
    end while
  end for
  Replace priors  $p(\theta)$  with KDE  $\hat{f}_h(\theta_{accepted})$ 
end for

```

2.4 Topology distance functions

To use geomodel topology as a constraint for probabilistic geomodels in an ABC framework, we need a consistent way of comparing geomodel topologies—i.e. suitable distance functions. We consider here three possible comparison methods:

1. **Presence or absence of defined connections:** As the relational topology information is captured in adjacency graphs, the most fundamental approach is to check if two relevant nodes n_1 and n_2 (e.g. representing two regions in the model) share an edge $e = (n_1, n_2)$ (are adjacent), and if this edge exists in both models. This is the most simple way of comparing specific aspects of relational topology between geomodels. This approach can be viewed as a boolean comparison: *True* if the given edge exists in both models, *False* if not. This also enables the direct comparison of i multiple edges, which would result in a vector of i boolean statements for each comparison $[e_1, e_2, \dots, e_i]$.
2. **Comparing entire graphs:** To compare topology graphs as a whole, Thiele et al. (2016b) describe the use of the Jaccard index (Jaccard, 1912). It can be used to compare the similarity of sets by creating the ratio of the intersection and union of two graphs A and B :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

For two topology graphs A and B , this means we calculate the ratio of edges (representing connected regions) shared in both (intersection: $A \cap B$) and their total combined number of edges (union: $A \cup B$). This ratio can be used to efficiently identify all unique topology graphs in a given ensemble, as only an identical pair of graphs results in a Jaccard index of $J(A, B) = 1$. A comparison using the Jaccard index yields ratios of integers, thus a discrete comparison. This

method also allows specifying a tolerance $0 < \epsilon < 1$ for model acceptance, i.e. to accept models within the range $1 - \epsilon \leq J \leq 1$.

3. **Contact area:** Comparing the number of actual edge pixels (or voxels), representing the area of the contact A_e between two geobodies could yield a more granular comparison that allows to take into accounts trends of the contact size. Thus the ABC error tolerance ϵ could be used to reject geomodels where certain topological contact areas are above and/or below a certain value $A_e - \epsilon_{low} \leq A_e \leq A_e + \epsilon_{high}$.

In this work we demonstrate the second approach, as it allows us to directly compare entire geomodel topologies. We have chosen to compare the simulated results to a single topology graph—the initial geomodel topology. This approach was selected as a base case to demonstrate how the large variations in geomodel topology observed in the stochastic simulation of input data uncertainties in geomodels (see Thiele et al., 2016b) can be constrained to a base topology (i.e. conceptual model). This of course reinforces the bias of the initial base model into the uncertainty simulation, but allows for the reliable exploration of uncertainty of all possible geomodels honoring the topology constraint.

2.5 Quantifying Uncertainty using Shannon Entropy

Stochastic simulations yield vast ensembles of geomodel realizations and their variability (and thus uncertainty) needs to be analyzed and understood. The uncertainty of a single geological entity (e.g. a layer or a fault) can be estimated from its frequency of occurrence in each single geomodel voxel. In order to analyze the whole geomodel uncertainty at once, more sophisticated measures can be applied: the concept of *Shannon entropy* H can be used in a spatial context to evaluate the uncertainty of an entire geomodel ensemble at once, as described by Wellmann and Regenauer-Lieb (2012). ~~Their concept is based on concepts from information theory, derived by Shannon (1948), and further on the concept of fuzziness established by Zadeh (1965) and De Luca and Termini (1972). If applied to a fuzzy set² $f \in [0, 1]$ in a grid, the measure should only be 0 if every grid cell is either 0 or 1 everywhere (thus the grid having no uncertainty anywhere, meaning we are absolutely certain about the lithology at this position), and should have its maximum value when $f = 0.5$ for all grid cells (meaning all outcomes are equally likely, which represents the highest uncertainty possible: every lithology is equally likely to be present at this position). The resulting equation is:-~~

$$H_m = -\frac{1}{N} \sum_{x=1}^N \left[p_m(x) \log_2(p_m(x)) + (1 - p_m(x)) \log_2(1 - p_m(x)) \right]$$

~~where we denote the fuzzy set f as the probability p_m of an outcome $m \in M$ of a cell x , and H_m being the Shannon entropy normalized by the total number of cells N . The average model entropy Average model entropy \bar{H} can also be evaluated by:-~~

$$\bar{H} = -\frac{1}{N} \sum_{x=1}^N H(x)$$

²i.e. a non-binary set with real numbers in-between the two interval boundaries.

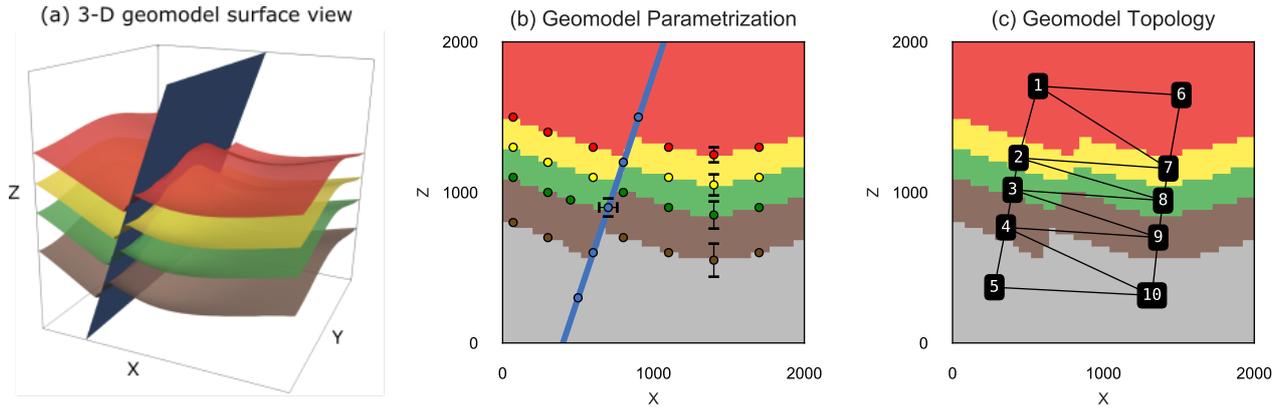


Figure 4. (a) 3D view of the synthetic fault model, with top surfaces of the four lithologies shown and the fault surface in blue; (b) XZ-slice through the center of the discretized model showing partial input data (for visual brevity) and example standard deviations of prior parameters used for the stochastic simulation; (c) Model overlaid with its topology graph used as our summary statistic for the ABC.

~~Which makes the average model entropy collapsed the uncertainty of a geomodel ensemble into a single number. It will be equal to 0 if all cells x have only one possible outcome (no uncertainty), and reaching its maximum when all outcomes are equally likely for all cells of the model (maximum uncertainty).~~

2.6 Experiment Design

5 2.6.1 Synthetic Fault Model

As a proof of concept we show how ABC can be used to incorporate geological knowledge and reasoning into an uncertain synthetic geomodel. This model represents a folded layer cake stratigraphy that is cut by a N-S striking normal fault to represent an idealised reservoir scenario frequently encountered in the energy industry (see Fig. 4a).

The prior parametrization is schematically visualized in Figure 4b and consists of two different kinds of uncertain parameters: (i) vertical location of the layer and fault interfaces and (ii) lateral location of the fault interface, with the specific parametrization displayed in Table 1 in the Appendix. ~~As this work focuses on developing and describing a novel methodology for constraining uncertain geomodels, we have chosen the uncertainty parametrization of the synthetic geomodel entirely subjectively as normal distributions increasing in uncertainty with depth. The uncertainty is applied to each set of surface points to preserve surface shape, individually within each of the two fault blocks. Proper prior parametrization of uncertain geomodels is a vital branch of research on its own (e.g. Pakyuz-Charrier, 2018; Krajnovich et al., 2020) and out of the scope of this work.~~

Two separate simulations were run for this experiment so we can see how topology can constrain an uncertain geomodel compared to the Monte Carlo simulation of ~~uncertainties~~ input parameter uncertainties alone:

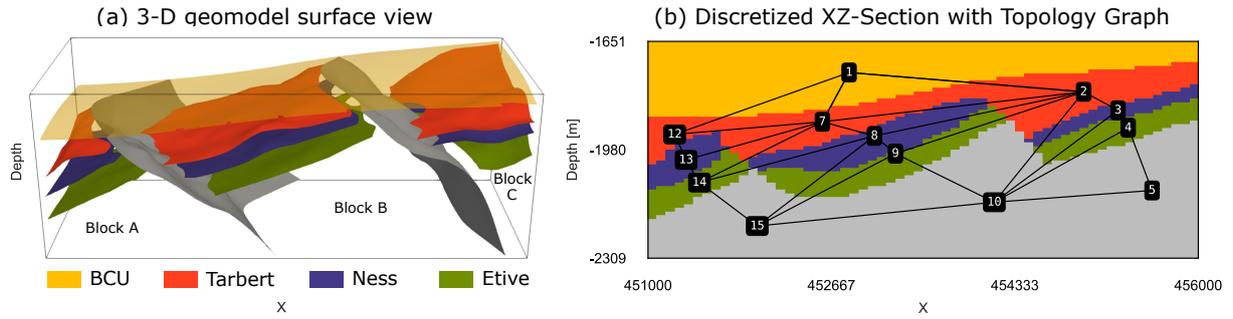


Figure 5. (a) 3D view of the Gullfaks geomodel used as mean prior model in our case study; (b) XZ-section through the discretized geomodel with overlaid [topology oyservtopology](#) graph showing the inter- and intra-fault block relations of geobodies.

1. A Monte Carlo simulation of the prior parameters to evaluate the uncertainty in the resulting geomodel ensemble consisting of 2000 generated models. This represents our 'base case' uncertainty without any [topological](#) constraints.
2. An Approximate Bayesian Computation using the initial model topology graph (see Fig. 4c) to represent our geological knowledge. We are employing a rejection sampling scheme (ABC-REJ) with an error tolerance of $\epsilon = 0$ to obtain 500 generated posterior models. Thus, the resulting posterior geomodel ensemble will contain only samples with matching topology graphs.

2.6.2 Case Study: The Gullfaks Field

To demonstrate the applicability of the method to real datasets we apply it to a model of part of the Gullfaks Field, located in the northern North Sea. The field is located in the western part of the Viking Graben, and consists of the NNE-SSW-trending 10-25 km wide Gullfaks fault block (Fossen and Hesthammer, 1998). For a detailed overview of the regional and structural geology we refer to Fossen and Rørnes (1996); Fossen and Hesthammer (1998); Fossen et al. (2000); Schaaf and Bond (2019).

For the experiment, we constructed a base geomodel (Fig. 5a) founded in an interpretation of the training data set provided with the seismic interpretation software Petrel™. We have chosen a relatively simple subset of the interpretation, containing 2 faults, three horizon tops Tarbert (red), Ness (purple) and Etive (green), and the Base Cretaceous Unconformity (BCU, yellow). To create the geomodel, we exported the corresponding seismic interpretation data from Petrel and imported them into Python. The surface interpretations were then decimated down to 510 surface points and 187 surface orientations, via a target reduction of 80 % per fault block or surface using the VTK-based decimation functionality of `pyvista` (Sullivan and Kaszynski, 2019), to retain the best possible surface shape while allowing fast implicit geomodel construction times in `GemPy`.

The prior parametrization consists of two different kinds of uncertain parameters: (i) vertical location of the layer interfaces for within each fault block; (ii) the lateral location of the fault interfaces. This parametrization is similar to the synthetic fault model (all specifications are listed in Table 2 in the Appendix), [and all sets of surface points within each individual fault block were perturbed together to retain surface shape](#). This parametrization was chosen [due to its ease of implementation and to demonstrate how simplified uncertainty modeling to demonstrate how even a few uncertain parameters in an uncertainty](#)

[modeling workflow](#) can lead to highly uncertain results, especially regarding the topology graphs of the resulting geomodel ensembles in real-world geomodels. We then conducted a sensitivity study of the topological spread with respect to the geomodel resolution. This allowed us to determine the appropriate geomodel resolution necessary for our experiment. Next, we performed three separate simulations to compare different approaches:

- 5 1. A Monte Carlo simulation of the prior uncertainty for 1000 samples, to evaluate the spatial uncertainty and the topological spread of the resulting geomodel ensemble. This serves as our 'base case' uncertainty for comparison with the following two simulations.
2. An ABC-REJ simulation using the initial geomodel topology graph (see Fig. 5b) to represent our geological knowledge. We used an error threshold of $\epsilon = 0.025$ for 1000 accepted posterior samples, as the threshold was small enough to
10 constrain the posterior topology spread to the initial geomodel topology graph.
3. An ABC-SMC simulation using the same initial geomodel topology graph. We ran six SMC epochs using ϵ values of 0.3, 0.2, 0.1, 0.075, 0.05 and 0.025. Each epoch was run for 1000 accepted posterior samples.

3 Results

3.1 Synthetic Fault Model

15 Simulating the uncertainties encoded in the prior parameterization resulted in 100 unique model topologies within the geomodel ensemble of 2000 models, with 18 topology graphs occurring at least ten times and the most frequent 14 making up 90 % of geomodel ensemble topologies. It is also notable that the most frequent topology graph (29.5 %) is not the initial (mean prior) topology graph (15.6 %), but rather represents models where the Shale layer (green) of the foot wall shares an across-fault connection with the Sandstone 2 layer (red) of the hanging wall. The uncertainty of the prior geomodel ensemble is
20 visualized in Figure 6a-c in XZ-, YZ- and XY-sections as Shannon entropy, as described in the methodology. All three sections through the model show clearly the uncertainty of the layer interface position and highest uncertainty around the fault surface. In comparison, applying a single topology graph as a summary statistics to the simulation using ABC leads to significantly reduced uncertainty throughout the geomodel ensemble (see Fig. 6d-f), with average geomodel ensemble entropy being reduced from $\overline{H}_{prior} = 0.44$ down to $\overline{H}_{posterior} = 0.31$, a drop in geomodel uncertainty of nearly 30 %. Visualizing the entropy
25 difference between the prior and the posterior geomodel ensembles shows the highest reduction in entropy for the two inner layer interfaces (see Fig. 7), and not around the fault surface. As expected, constraining the simulation using a single topology graph with an error of $\epsilon = 0$ collapses the number of geomodel ensemble topologies from 100 down to 1.

Figure 8 plots [histograms and their the](#) kernel density estimations (KDE) of the input parameter distributions of prior (grey) and posterior (coloured) samples. The strongest change in mean from prior to posterior distributions occurred for the vertical interface location perturbation priors of Sandstone 2 (red), Shale (green) and Sandstone 1 (brown; see Fig. 8), with the first
30 shifted to higher mean z-values and the latter two shifted deeper by -72 m and -53 m , respectively. Additionally, the initially

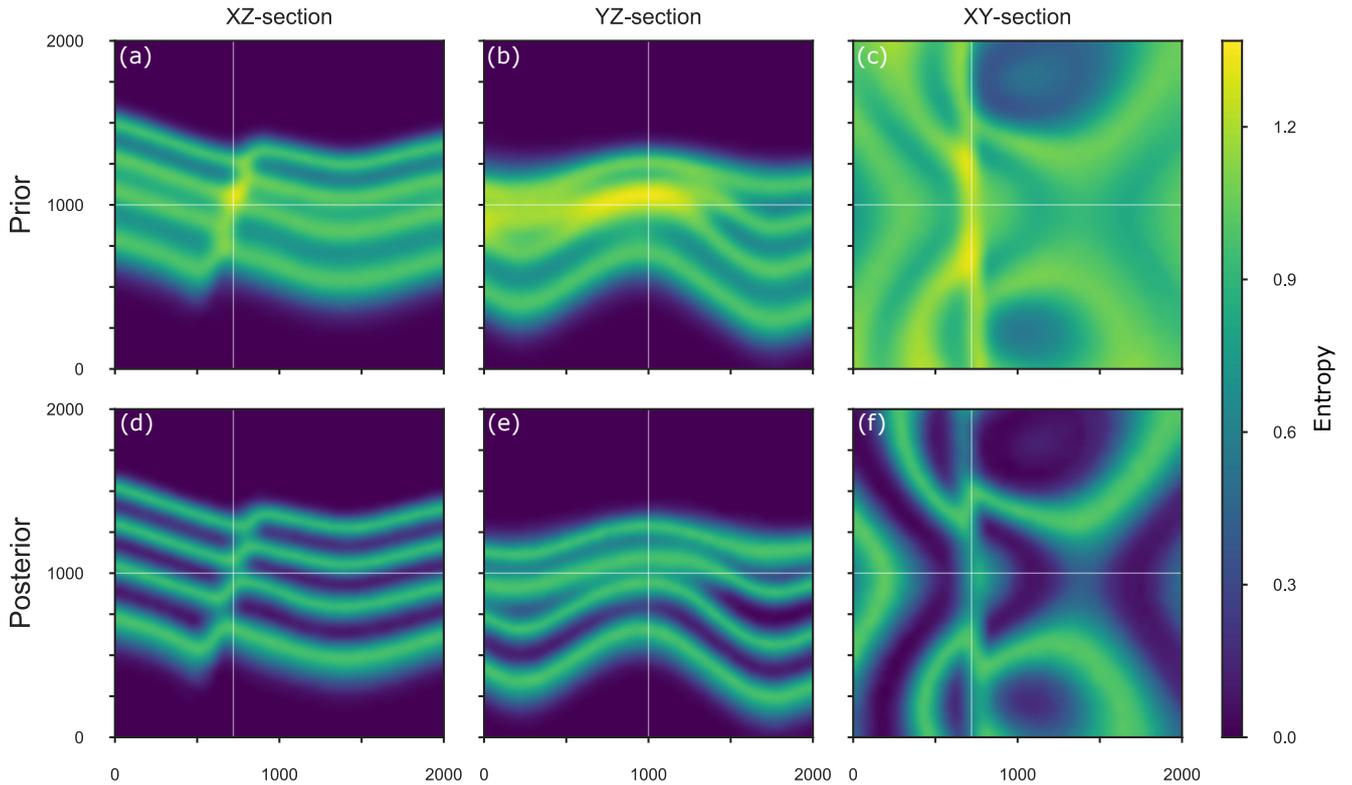


Figure 6. Shannon entropy slices in the XZ- (left), YX- (center) and XY-plane of the prior (top, [a-c](#)) and posterior (bottom, [d-f](#)) geomodel ensemble. [The white lines show the location of respective other cross-sections.](#)

normally distributed prior of Sandstone 1 shows a strong negative skewness of -0.61 in the posterior distribution. Standard deviation for the Siltstone and Shale interface distributions was reduced by roughly 32 % and 40 % respectively. The prior and posterior distributions for the lateral and vertical fault parameter uncertainties show no significant difference (e and f).

3.2 Case Study: The Gullfaks Field

- 5 Forward simulation of the prior uncertainties of the Gullfaks geomodel resulted in 676 unique geomodel topologies within a 1000-model ensemble, with 116 unique topologies occurring more than once. Again, the most frequent topology graph is not the initial (mean prior) topology graph. The uncertainty of a XZ-section of the forward ensemble is visualized in Figure 10a using Shannon entropy. The section illustrates the general trend of uncertainty throughout the forward simulation: we observe highest uncertainty surrounding the two faults in the geomodel, especially around the eastern fault. The area also
- 10 shows increased uncertainty due to the interaction of layer interfaces, the fault and the vertical vicinity of the BCU.

Applying the initial topology graph as a constraining summary statistics using ABC with rejection sampling (ABC-REJ) using a threshold of $\epsilon = 0.025$ (chosen empirically), results in much reduced uncertainty, as exemplified by the entropy section

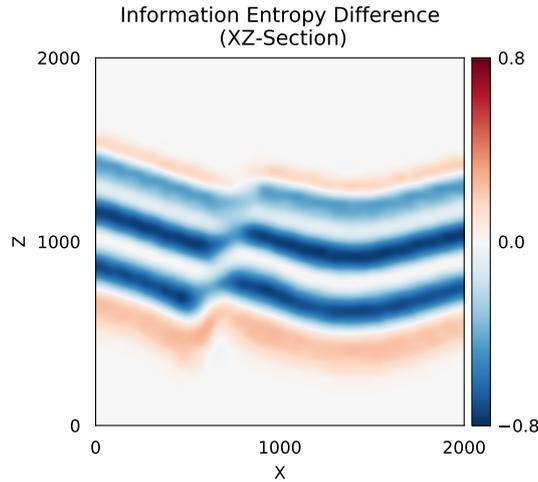


Figure 7. XY-Section of entropy difference between the forward simulated entropy and the approximate posterior entropy. The plot highlights areas where the entropy was reduced (blue), increased (red) and kept constant (white).

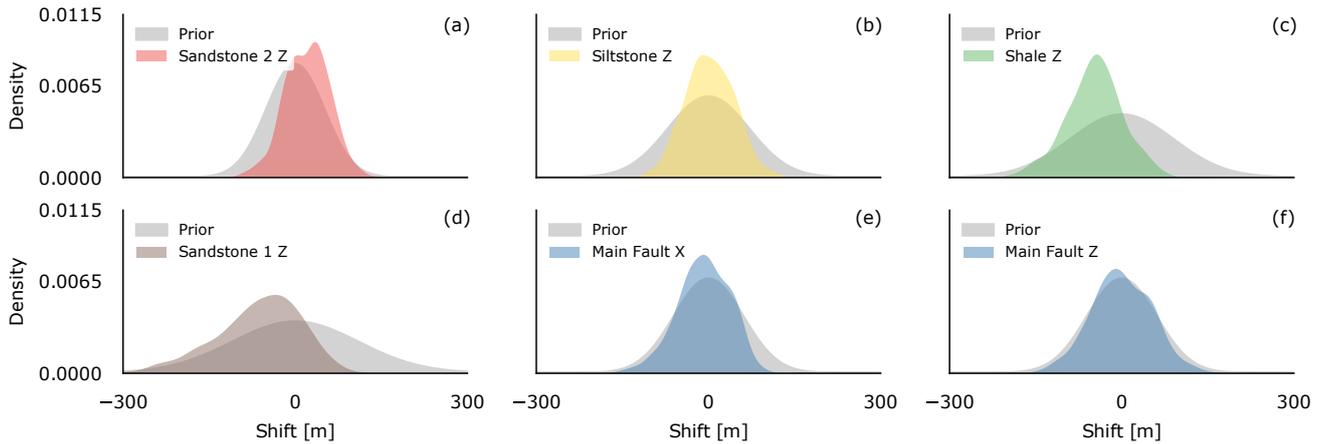


Figure 8. Prior (grey) and posterior (color) kernel density estimations for the different stochastic model parameters for our synthetic fault model.

shown in Figure 10b. At this threshold, the approximate posterior geomodel ensemble contains only the applied initial topology graph. Using rejection sampling with such a strict threshold resulted in a very low acceptance ~~rate of 0.0059~~ of only 0.59 % of simulated geomodels, which required about 40 hours of simulation time to obtain 1000 posterior samples². In contrast, using

²The experiment was run on consumer-grade hardware and leveraging GPU computation: Intel Core i5-8600K @ 3.60GHz, Nvidia GeForce RTX 2070 8GB GDDR6, 16 GB DDR4 RAM @ 2133MHz.

a Sequential Monte Carlo sampling scheme (ABC-SMC) required only 3.96 hours to obtain the same number of posterior samples at the same threshold—a speed-up of 10.1. This includes the five sampling epochs using $\epsilon = \{0.3, 0.2, 0.1, 0.075, 0.05\}$ with 1000 accepted samples each, used to sequentially adapt the priors.

Figure 12a shows the number of unique topologies for forward simulations and each threshold of the ABC-SMC. As we iteratively lower the acceptable threshold during the SMC simulation, the simulated and accepted topologies iteratively converge towards the topology graph we used as our prior geological knowledge. The average geomodel ensemble entropy \bar{H} is also iteratively decreasing from 0.233 for the forward simulation down to 0.112 at $\epsilon = 0.025$ (see Fig. 12b), showing how fixing a probabilistic geomodel to a single topology graph can significantly reduce, or rather significantly constrain, the simulated uncertainty.

Figure 9 shows how the ABC-SMC simulation iteratively affects the probability distributions of selected probabilistic geomodel parameters with decreasing thresholds ϵ . Each row shows the consecutive epochs of the ABC-SMC simulation and corresponds to a specific ϵ . Each column describes a different stochastic parameter in the stochastic model. By applying the initial topology graph of the geomodel as our summary statistics, we can directly see here how the parameter distribution for the BCU (Fig. 9a) shifts its mean μ by 47.4 m upwards and reduces its standard deviation σ by 35.8 % to accommodate our geological knowledge about the geomodel topology. We can observe this effect in the entropy section of the posterior geomodel ensemble as well (Fig. 10b). In Figure 11, we show the difference in entropy between the prior and approximate posterior geomodel ensemble shown in Figure 10, where areas with decreasing entropy values are shown in blue, increasing values in red. We observe here how the BCU moves upward and increases the entropy there, while lowering entropy in the lithologies below. The parameter distributions for Tarbert B (Fig. 9b, red) and Etive B (Fig. 9c, green) show similar behaviour: shifted mean and reduced standard deviation to accommodate the topology information. We see a much stronger reduction in standard deviation for the two faults (Fig. 9d,e): 80.4 % and 80.0 % for Fault A and Fault B, respectively. This is also shown as the strongest reduction in entropy in Figure 11.

4 Discussion

We showed how topology information, as an encoding for important aspects of geological knowledge and reasoning, can be included in probabilistic geomodeling methods in a Bayesian framework. The simulation experiments for our two case studies demonstrated that we are able to approximate posterior distributions to obtain probabilistic geomodel ensembles that honour both our prior parameter knowledge and qualitative geological knowledge. If the applied topological information is meaningful, then the constrained stochastic geomodel ensemble will see a meaningful reduction in uncertainty, and will subsequently allow for more precise model-based estimates and decision-making (Stamm et al., 2019). More importantly, the (approximate) Bayesian approach requires the explicit statement of the geological knowledge (here the topology information) used in the probabilistic geomodel, increasing the transparency of assumptions made during the geomodeling process and any subsequent decisions.

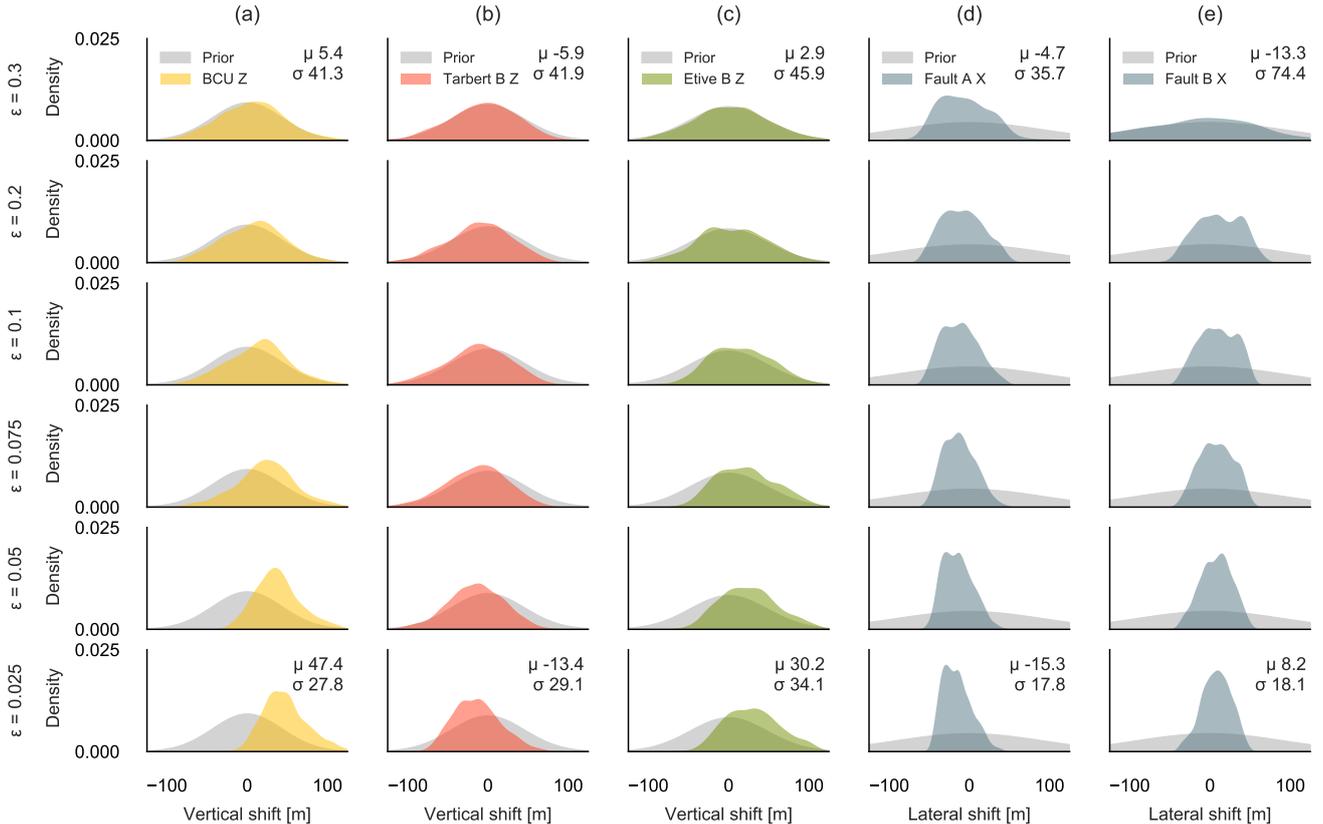


Figure 9. Prior (grey) and posterior (colored) kernel density estimations for selected model parameters (a-e) for the 6 epochs (each row represents an epoch) of the ABC-SMC simulation of the Gullfaks case study, showing how the simulation iteratively approaches the approximate posterior distribution, which shows the possible parameter uncertainty given our topological information. Mean μ and standard deviation σ shown for the first and last epochs.

With our approach, we directly address a scientific challenge raised in recent work by Thiele et al. (2016b), that known topological relationships are frequently not honoured during the probabilistic modeling process, thus potentially invalidating large parts of the resulting geomodel ensemble. Injecting topology information into a Bayesian approach allows us to obtain topologically valid, and hence geologically reasonable, geomodel ensembles. And, although we have only used simple topology information within this study, the demonstrated ABC approach allows to easily scale the amount of topology information used: from simple True-False comparisons of single topology graphs to the use of a whole range of topology graphs and relationships. If a set of acceptable topologies is used, one could for example accept a simulated model if it matches at least one within the error tolerance.

The work of Pakyuz-Charrier et al. (2019) shows how clustering of probabilistic geomodel topologies can be used to differentiate between different modes of topologies. Their approach compares geomodel topologies by describing them as half-

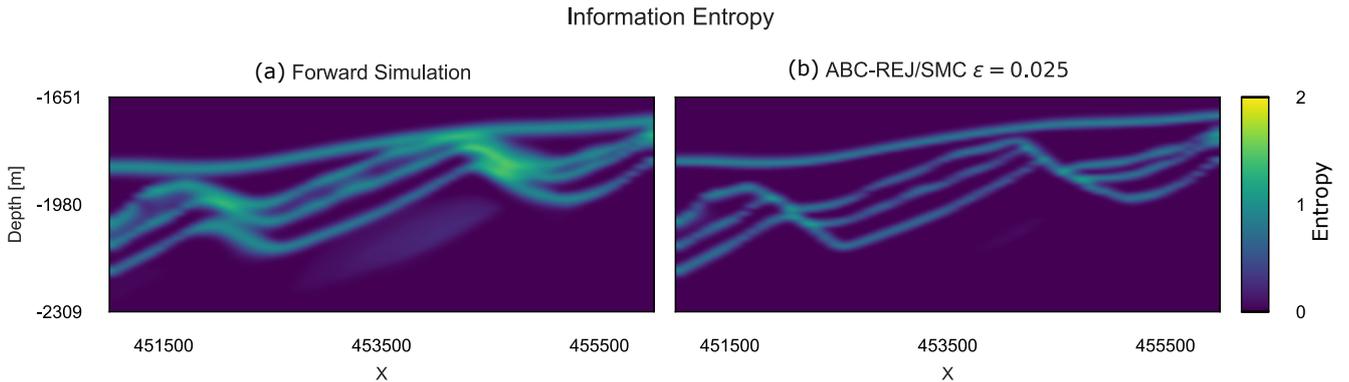


Figure 10. (a) Section of the entropy block of the forward simulation for the prior uncertainty ($H_T = 0.223$); (b) Section of the entropy block of the final epoch ($\epsilon = 0.025$) of the ABC-SMC simulation ($H_T = 0.113$).

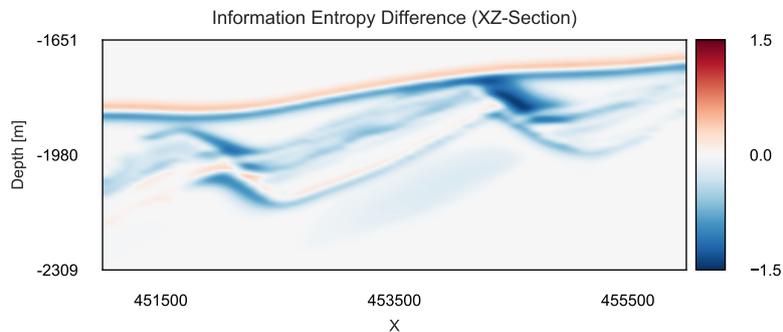


Figure 11. XZ-Section of entropy difference between the forward simulated entropy and the approximate posterior entropy H ($\epsilon = 0.025$). The plot highlights areas where the entropy was reduced (blue), increased (red) and kept constant (white).

vectorized adjacency matrices, resulting in a binary string that can be compared using the Hamming distance (Hamming, 1950). It could be considered as a different distance metric in the ABC approach presented in this work to constrain the simulated probabilistic geomodel. And, while their work focuses on the analysis of existing probabilistic geomodel ensembles, our approach focuses on learning probabilistic geomodels on topology information while reducing the number of required iterations

5 through use of advanced sampling techniques.

As more complex geomodels strongly increase the required parametrization to accurately describe the model domain in a probabilistic framework, constraining them with topological information could help keep this parametrization at computationally feasible levels by reducing the parameter dimensionality, while still obtaining meaningful geomodels (e.g. free of modeling artefacts caused by random perturbations of the limited input data). This would not work using an inefficient rejection sampling

10 scheme (e.g. ABC-REJ), but would rather require the use of "adaptive" sampling algorithms to efficiently explore the posterior

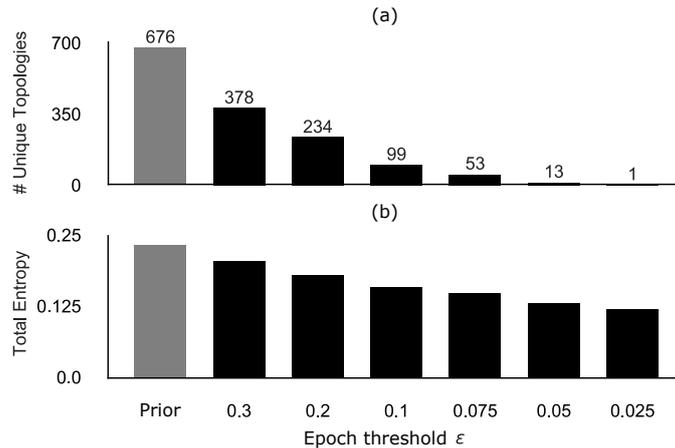


Figure 12. (a) Number of unique topologies within the geomodel ensembles of each SMC epoch, showing the iterative reduction in topological uncertainty throughout the SMC simulation; (b) Average geomodel entropy of the ensembles for each epoch, showing how the reduction of topological uncertainty shown in (a) affects the total geomodel uncertainty.

parameter space without wasting too much computing power on rejected models (e.g. ABC-SMC). In our Gullfaks case study, we have not only shown the efficacy of the method on a real-world example, but demonstrated the stark increase in computational efficiency when using advanced sampling techniques. The SMC sampler used in our work requires manual setting of the acceptance thresholds, which directly influence the algorithm’s efficiency in acquiring samples of the approximate posterior distribution. Adaptive SMC methods automatically tune acceptance thresholds to increase sampling efficiency on-the-fly to minimize computation time and avoid manual (subjective) selection of thresholds (Del Moral et al., 2012).

Sadegh and Vrugt (2014) describe a more complex ABC algorithm based on Differential Evolution Adaptive Metropolis (DREAM-ABC) and demonstrate its much higher efficiency in approximating the posterior. It might be of particular interest for the approximate inference of complex structural geomodels with topology constraints, as it has shown promise to very efficiently explore high-dimensional (read: large amount of prior parameters) and multi-modal parameter spaces. When using multiple topology graphs (which are discrete) in an ABC framework, the posterior parameter space may potentially become multi-modal, which poses significant challenges for traditional Markov Chain-based samplers (Feroz and Hobson, 2008). The approach by Sadegh and Vrugt (2014) is based on combining multiple Markov chains, which natively supports parallel computing and would thus allow for a high scalability of the approach to complex, computationally intensive geomodels.

Alternatively, Bayesian Optimization for likelihood-free inference (BOLFI; Gutmann and Corander, 2016) could be worth considering for complex structural geomodels. The method abstracts the simulator/implicit function into a statistical surrogate model between the priors and the summary statistics and then attempts to minimize their distance, with the potential to significantly reduce the number of needed computations of the geomodel. Overall, the spatial and discrete nature of geomodels and the use of discrete summary statistics poses unique challenges to sampling algorithms, requiring further research to identify algorithms that can confidently converge and minimize the high computational cost of probabilistic 3-D geomodels.

The method demonstrated the effect of topology information on geomodel uncertainty—showing how well the parametrization of a probabilistic geomodel fits our geological assumptions. The acceptance rates during sampling could potentially be used as a proxy for the validity of our assumptions: low acceptance rates could reveal a bad fit between our model and our added geological knowledge and reasoning. Using entropy-difference plots, the effect of geological assumptions on the geomodel uncertainty can be analysed spatially, e.g. how it ~~reduces (or increases)~~ changes around faults and other structures in the geomodel ~~or other summary statistics of the geomodel, such as the gross rock volume of a potential reservoir across all fault blocks (or compartments) of interest~~ ensemble.

Summary

- We have shown how to use Approximate Bayesian Computation to constrain probabilistic geomodels so that the approximate posterior incorporates known topology information.
- The method enables additional geological knowledge and reasoning to be explicitly encoded and incorporated into probabilistic geomodel ensembles, potentially increasing transparency of the modeling assumptions.
- As opposed to standard MC with rejection, the implemented SMC approach makes the use of ABC feasible in realistic settings. Further research into using more advanced sampling schemes could provide additional speed-ups in obtaining the posterior geomodel ensemble, which is especially relevant for computationally more expensive complex geomodels with large parametrizations.

Table 1. Distribution parameters for prior parametrization of the synthetic fault model.

Name	Distribution	μ [m]	σ [m]
Sandstone_2_Z	Normal	0	50
Siltstone_Z	Normal	0	70
Shale_Z	Normal	0	90
Sandstone_1_Z	Normal	0	110
Main_Fault_X	Normal	0	60
Main_Fault_Z	Normal	0	60

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. This research was conducted within the scope of a Total E&P UK-funded postgraduate research project.

Table 2. Distribution parameters for prior parametrization of the Gullfaks case study.

Name	Distribution	μ [m]	σ [m]
BCU Z	Normal	0	43.3
fault3 X	Normal	0	90.9
fault4 X	Normal	0	90.5
tarbert A Z	Normal	0	46.5
tarbert B Z	Normal	0	45.5
tarbert C Z	Normal	0	44.2
ness A Z	Normal	0	48.6
ness B Z	Normal	0	46.7
ness C Z	Normal	0	45.1
etive A Z	Normal	0	50.9
etive B Z	Normal	0	48.1
etive C Z	Normal	0	46.3

Acknowledgements. We would like to thank Total E&P UK in Aberdeen for funding this research. We also thank Fabian Stamm for providing the wonderful synthetic geomodel used in this paper. [We are grateful for the constructive reviews from Ahston Krajnovich and an anonymous reviewer for helping us improve this manuscript.](#)

References

- Baddeley, M. C., Curtis, A., and Wood, R.: An introduction to prior information derived from probabilistic judgements: elicitation of knowledge, cognitive bias and herding, Geological Society, London, Special Publications, 239, 15–27, 2004.
- Bardossy, G. and Fodor, J.: Evaluation of Uncertainties and Risks in Geology: New Mathematical Approaches for Their Handling, Springer Science & Business Media, 2013.
- Bistacchi, A., Massironi, M., Dal Piaz, G. V., Dal Piaz, G., Monopoli, B., Schiavo, A., and Toffolon, G.: 3D Fold and Fault Reconstruction with an Uncertainty Model: An Example from an Alpine Tunnel Case Study, *Computers & Geosciences*, 34, 351–372, <https://doi.org/10.1016/j.cageo.2007.04.002>, 2008.
- Bolstad, W. M.: *Understanding Computational Bayesian Statistics*, John Wiley & Sons, 2009.
- 10 Bond, C., Gibbs, A., Shipton, Z., and Jones, S.: What Do You Think This Is? “Conceptual Uncertainty” in Geoscience Interpretation, *GSA Today*, 17, 4, <https://doi.org/10.1130/GSAT01711A.1>, 2007.
- Bond, C. E.: Uncertainty in structural interpretation: Lessons to be learnt, *Journal of Structural Geology*, 74, 185–200, 2015.
- Bond, C. E., Shipton, Z. K., Gibbs, A. D., and Jones, S.: Structural models: optimizing risk analysis by understanding conceptual uncertainty, *First Break*, 26, 2008.
- 15 Brandes, C. and Tanner, D. C.: Fault-related folding: A review of kinematic models and their application, *Earth-Science Reviews*, 138, 352–370, 2014.
- Butler, R. W., Bond, C. E., Cooper, M. A., and Watkins, H.: Interpreting structural geometry in fold-thrust belts: Why style matters, *Journal of Structural Geology*, 114, 251–273, 2018.
- 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.
- 20 Chilès, J. P., Aug, C., Guillen, A., and Lees, T.: Modelling the Geometry of Geological Units and Its Uncertainty in 3D From Structural Data: The Potential-Field Method, p. 8, 2004.
- Crossley, M. D.: *Essential Topology*, Springer Science & Business Media, 2006.
- 25 Curtis, A. and Wood, R.: Optimal Elicitation of Probabilistic Information from Experts, Geological Society, London, Special Publications, 239, 127–145, <https://doi.org/10.1144/GSL.SP.2004.239.01.09>, 2004.
- 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/https://doi.org/10.5194/gmd-12-1-2019>, 2019.
- 30 De Luca, A. and Termini, S.: A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory, *Information and Control*, 20, 301–312, 1972.
- Del Moral, P., Doucet, A., and Jasra, A.: An Adaptive Sequential Monte Carlo Method for Approximate Bayesian Computation, *Statistics and Computing*, 22, 1009–1020, <https://doi.org/10.1007/s11222-011-9271-y>, 2012.
- 35 Egenhofer, M. J.: Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases3, p. 29, 1990.

- Feroz, F. and Hobson, M. P.: Multimodal Nested Sampling: An Efficient and Robust Alternative to Markov Chain Monte Carlo Methods for Astronomical Data Analyses, *Monthly Notices of the Royal Astronomical Society*, 384, 449–463, <https://doi.org/10.1111/j.1365-2966.2007.12353.x>, 2008.
- Fossen, H.: *Structural Geology*, Cambridge University Press, first edn., 2010.
- 5 Fossen, H. and Hesthammer, J.: *Structural Geology of the Gullfaks Field, Northern North Sea*, Geological Society, London, Special Publications, 127, 231–261, <https://doi.org/10.1144/GSL.SP.1998.127.01.16>, 1998.
- Fossen, H. and Rørnes, A.: Properties of Fault Populations in the Gullfaks Field, Northern North Sea, *Journal of Structural Geology*, 18, 179–190, [https://doi.org/10.1016/S0191-8141\(96\)80043-5](https://doi.org/10.1016/S0191-8141(96)80043-5), 1996.
- Fossen, H., Odinsen, T., Færseth, R. B., and Gabrielsen, R. H.: Detachments and Low-Angle Faults in the Northern North Sea Rift System, Geological Society, London, Special Publications, 167, 105–131, <https://doi.org/10.1144/GSL.SP.2000.167.01.06>, 2000.
- 10 Frodeman, R.: Geological Reasoning: Geology as an Interpretive and Historical Science, *Geological Society of America Bulletin*, 107, 960–996, [https://doi.org/10.1130/0016-7606\(1995\)107<0960:GRGAAI>2.3.CO;2](https://doi.org/10.1130/0016-7606(1995)107<0960:GRGAAI>2.3.CO;2), 1995.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: *Bayesian Data Analysis*, Chapman and Hall/CRC, <https://doi.org/10.1201/b16018>, 2013.
- 15 Groshong, R., Bond, C., Gibbs, A., Ratcliff, R., and Wiltshko, D.: Preface: Structural balancing at the start of the 21st century: 100 years since Chamberlin, *Journal of Structural Geology*, 41, 1–5, 2012.
- Gutmann, M. U. and Corander, J.: Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models, *Journal of Machine Learning Research*, 17, 1–47, 2016.
- Hamming, R. W.: Error detecting and error correcting codes, *The Bell system technical journal*, 29, 147–160, 1950.
- 20 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.
- Jaccard, P.: The Distribution of the Flora in the Alpine Zone.1, *New Phytologist*, 11, 37–50, <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>, 1912.
- Jessell, M., Aillères, L., de Kemp, E., Lindsay, M., Wellmann, F., Hillier, M., Laurent, G., Carmichael, T., and Martin, R.: Next Generation Three-Dimensional Geologic Modeling and Inversion, in: *Building Exploration Capability for the 21st Century*, Society of Economic Geologists, <https://doi.org/10.5382/SP.18.13>, 2014.
- 25 Krajnovich, A., Zhou, W., and Gutierrez, M.: Uncertainty assessment for 3D geologic modeling of fault zones based on geologic inputs and prior knowledge, *Solid Earth*, 11, 1457–1474, 2020.
- Lajaunie, C., Courrioux, G., and Manuel, L.: Foliation Fields and 3D Cartography in Geology: Principles of a Method Based on Potential Interpolation, *Mathematical Geology*, 29, 571–584, <https://doi.org/10.1007/BF02775087>, 1997.
- 30 Laurent, G., Aillères, L., Grose, L., Caumon, G., Jessell, M., and Armit, R.: Implicit Modeling of Folds and Overprinting Deformation, *Earth and Planetary Science Letters*, 456, 26–38, <https://doi.org/10.1016/j.epsl.2016.09.040>, 2016.
- Lindsay, M. D., Aillères, L., Jessell, M. W., de Kemp, E. A., and Betts, P. G.: Locating and Quantifying Geological Uncertainty in Three-Dimensional Models: Analysis of the Gippsland Basin, Southeastern Australia, *Tectonophysics*, 546–547, 10–27, <https://doi.org/10.1016/j.tecto.2012.04.007>, 2012.
- 35 MacKay, D. J. C. and Kay, D. J. C. M.: *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.
- Mallet, J.-L.: Space–Time Mathematical Framework for Sedimentary Geology, *Mathematical Geology*, 36, 1–32, <https://doi.org/10.1023/B:MATG.0000016228.75495.7c>, 2004.

- Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J.: Approximate Bayesian Computational Methods, *Statistics and Computing*, 22, 1167–1180, <https://doi.org/10.1007/s11222-011-9288-2>, 2012.
- Pakyuz-Charrier, E.: Uncertainty Is an Asset: Monte Carlo Simulation for Uncertainty Estimation in Implicit 3D Geological Modelling, PhD Thesis, University of Western Australia, 2018.
- 5 Pakyuz-Charrier, E., Jessell, M., Giraud, J., Lindsay, M., and Ogarko, V.: Topological Analysis in Monte Carlo Simulation for Uncertainty Propagation, *Solid Earth*, 10, 1663–1684, <https://doi.org/https://doi.org/10.5194/se-10-1663-2019>, 2019.
- Polson, D. and Curtis, A.: Dynamics of Uncertainty in Geological Interpretation, *Journal of the Geological Society*, 167, 5–10, <https://doi.org/10.1144/0016-76492009-055>, 2010.
- Sadegh, M. and Vrugt, J. A.: Approximate Bayesian Computation Using Markov Chain Monte Carlo Simulation: DREAM (ABC), *Water Resources Research*, 50, 6767–6787, <https://doi.org/10.1002/2014WR015386>, 2014.
- 10 Schaaf, A. and Bond, C. E.: Quantification of Uncertainty in 3-D Seismic Interpretation: Implications for Deterministic and Stochastic Geomodeling and Machine Learning, *Solid Earth*, 10, 1049–1061, <https://doi.org/https://doi.org/10.5194/se-10-1049-2019>, 2019.
- Shannon, C. E.: A Mathematical Theory of Communication, *Bell System Technical Journal*, 27, 379–423, <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>, 1948.
- 15 Stamm, F. A., de la Varga, M., and Wellmann, F.: Actors, Actions and Uncertainties: Optimizing Decision Making Based on 3-D Structural Geological Models, *Solid Earth Discussions*, pp. 1–35, <https://doi.org/https://doi.org/10.5194/se-2019-57>, 2019.
- Sullivan, C. and Kaszynski, A.: PyVista: 3D Plotting and Mesh Analysis through a Streamlined Interface for the Visualization Toolkit (VTK), *Journal of Open Source Software*, 4, 1450, <https://doi.org/10.21105/joss.01450>, 2019.
- Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C.: Approximate Bayesian Computation, *PLoS Computational Biology*, 9, e1002803, <https://doi.org/10.1371/journal.pcbi.1002803>, 2013.
- 20 Suzuki, S., Caumon, G., and Caers, J.: Dynamic Data Integration for Structural Modeling: Model Screening Approach Using a Distance-Based Model Parameterization, *Computational Geosciences*, 12, 105–119, <https://doi.org/10.1007/s10596-007-9063-9>, 2008.
- Tacher, L., Pomian-Szednicki, I., and Parriaux, A.: Geological Uncertainties Associated with 3-D Subsurface Models, *Computers & Geosciences*, 32, 212–221, <https://doi.org/10.1016/j.cageo.2005.06.010>, 2006.
- 25 Theano Development Team, Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A., Bengio, Y., Bergeron, A., Bergstra, J., Bisson, V., Snyder, J. B., Bouchard, N., Boulanger-Lewandowski, N., Bouthillier, X., de Brébisson, A., Breuleux, O., Carrier, P.-L., Cho, K., Chorowski, J., Christiano, P., Coijmans, T., Côté, M.-A., Côté, M., Courville, A., Dauphin, Y. N., Delalleau, O., Demouth, J., Desjardins, G., Dieleman, S., Dinh, L., Ducoffe, M., Dumoulin, V., Kahou, S. E., Erhan, D., Fan, Z., Firat, O., Germain, M., Glorot, X., Goodfellow, I., Graham, M., Gulcehre, C., Hamel, P., Harlouchet, I., Heng, J.-P., Hidas, B., Honari, S., Jain, A., Jean, S., Jia, K., Korobov, M., Kulkarni, V., Lamb, A., Lamblin, P., Larsen, E., Laurent, C., Lee, S., Lefrançois, S., Lemieux, S., Léonard, N., Lin, Z., Livezey, J. A., Lorenz, C., Lowin, J., Ma, Q., Manzagol, P.-A., Mastropietro, O., McGibbon, R. T., Memisevic, R., van Merriënboer, B., Michalski, V., Mirza, M., Orlandi, A., Pal, C., Pascanu, R., Pezeshki, M., Raffel, C., Renshaw, D., Rocklin, M., Romero, A., Roth, M., Sadowski, P., Salvatier, J., Savard, F., Schlüter, J., Schulman, J., Schwartz, G., Serban, I. V., Serdyuk, D., Shabanian, S., Simon, É., Spieckermann, S., Subramanyam, S. R., Szygowski, J., Tanguay, J., van Tulder, G.,
- 30 Turian, J., Urban, S., Vincent, P., Visin, F., de Vries, H., Warde-Farley, D., Webb, D. J., Willson, M., Xu, K., Xue, L., Yao, L., Zhang, S., and Zhang, Y.: Theano: A Python Framework for Fast Computation of Mathematical Expressions, [arXiv:1605.02688 \[cs\]](https://arxiv.org/abs/1605.02688), 2016.
- Thiele, S. T., Jessell, M. W., Lindsay, M., Ogarko, V., Wellmann, J. F., and Pakyuz-Charrier, E.: The Topology of Geology I: Topological Analysis, *Journal of Structural Geology*, 91, 27–38, 2016a.

- 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>, 2016b.
- Thore, P., Shtuka, A., Lecour, M., Ait-Ettajer, T., and Cognot, R.: Structural Uncertainties: Determination, Management, and Applications, *GEOPHYSICS*, 67, 840–852, <https://doi.org/10.1190/1.1484528>, 2002.
- 5 VanderPlas, J.: Frequentism and Bayesianism: A Python-Driven Primer, arXiv:1411.5018 [astro-ph], 2014.
- Wellmann, J. F.: Information Theory for Correlation Analysis and Estimation of Uncertainty Reduction in Maps and Models, *Entropy*, 15, 1464–1485, <https://doi.org/10.3390/e15041464>, 2013.
- Wellmann, J. F. and Caumon, G.: 3-D Structural Geological Models: Concepts, Methods, and Uncertainties, in: *Advances in Geophysics*, vol. 59, pp. 1–121, Elsevier, 1st edn., 2018.
- 10 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., Thiele, S. T., Lindsay, M. D., and Jessell, M. W.: Pynoddy 1.0: An Experimental Platform for Automated 3-D Kinematic and
 15 Potential Field Modelling, *Geoscientific Model Development*, 9, 1019–1035, <https://doi.org/https://doi.org/10.5194/gmd-9-1019-2016>, 2016.
- 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*, p. SP453.12, <https://doi.org/10.1144/SP453.12>, 2017.
- 20 Wood, R. and Curtis, A.: Geological Prior Information and Its Applications to Geoscientific Problems, *Geological Society, London, Special Publications*, 239, 1.2–14, <https://doi.org/10.1144/GSL.SP.2004.239.01.01>, 2004.
- Zadeh, L. A.: Fuzzy Sets, *Information and Control*, 8, 338–353, [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X), 1965.
- Zlatanova, S.: On 3D Topological Relationships, in: *Proceedings 11th International Workshop on Database and Expert Systems Applications*, pp. 913–919, <https://doi.org/10.1109/DEXA.2000.875135>, 2000.