Solid Earth Discuss.,
https://doi.org/10.5194/se-2019-57-RC1, 2019

Interactive comment on "Actors, actions and uncertainties: Optimizing decision making based on 3-D structural geological models" *by* Fabian Antonio Stamm et al.

Anonymous Referee #1

Received and published: 13 May 2019

This is an interesting paper, I think that the use of loss functions in this setting is a useful development and one with considerable potential, and I think that the paper is to be welcomed because of this. That said, the account of the statistical formulation of the author's model is very unclear and needs considerable improvement.

In section 2.2 the authors present their synthetic setting. This can be modelled, by an implicit modelling algorithm, from some set of input values. I assume that these input values are notional values of the z coordinate for a particular contact at each of a set of locations $\{x, y\}$ in addition to notional values of dip direction and angle. The authors state that they can sample 'deviation values' (I assume they mean errors)

for these observations which are then added to the 'known' values. With one major caveat (below), the application of the implicit modelling algorithm to each of a series of 'observations' with stochastic error would generate a series of models of the underlying state of affairs. The caveat is that the authors give no indication of how they specify the joint distribution of errors, i.e. the joint distribution for the errors in z at locations x_1, y_1 and x_2, y_2. One can assume only that they treat these errors as independent, but that is most implausible in any geological setting, and could have dramatic consequences for the resulting model output. There are several studies, including some cited in this paper, which have considered the spatial dependence of model errors, and the authors should give this careful consideration in revising their work.

The authors then attempt to set their work in a Bayesian context. However, they do not make this at all explicit, and it is very difficult to understand or assess what they have done. A Bayesian analysis requires some model, for which some set of parameters are unknown and are treated as random variables to reflect this uncertainty. Prior probability distributions express the modeller's subjective view, prior to examining the data under consideration, about the values of these parameters. Likelihood values can be computed for proposed values of the parameters, and this, possibly in conjunction with some sampling method such as MCMC, is then used to characterize the posterior distribution of the parameters, given the data. Posterior distributions, or samples from them, could then be used, for example, to estimate the expected value of a loss function under a particular decision rule.

While the paper starts with some very generalized statements about Bayesian methods (Section 2.4), we are nowhere given a sufficiently clear account of their model system and how it is expressed in Bayesian form. The jump from 2.4 to 2.4.1 is much too large. It is not possible for the reader to extract from section 2.4.1 a sufficiently clear account of what the authors have done to be able to reproduce it in a comparable setting, or to be confident that it has been done correctly. For example, the authors refer to 'layer thickness likelihoods'. I assume that the thickness of a layer is a model parameter (assumed to be the same everywhere in space?). It is not made clear how the likelihood is defined, given a set of notional observations. Are the observations treated as independent? Is this reasonable? I would certainly be concerned if the thicknesses of the separate layers are treated as independent random variables. This does not strike me as implausible and, as assumptions go, it is likely to be a sensitive one in this particular context.

In short, I think that the authors need to start again, aiming at the simplest and clearest account they can give of their modelling framework so as to exemplify the loss function concept to best advantage. At present this gets lost in vague and hand-waving description.

A couple of minor points. On page 6 line 10 you refer to a negatively skew normal distribution of mean zero used to generate errors of the throw for your fault. One would expect such a distribution to have a median value larger than the mean, and there statement that most of the values produced are negative is curious. Please clarify.

Second, the statement of the loss functions from page 11 onward should be clarified. As stated Equations (4)–(7) are identical upto a constant value. I assume that you actually introduce the constants in (5) onward to introduce different degrees of asymmetry. This should be expressed in a rigorous way, indicating the conditions over which the constant is introduced. With some care a parsimonious but rigorous notation could be developed.

Interactive comment on Solid Earth Discuss., https://doi.org/10.5194/se-2019-57, 2019.