Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

Solid Earth Discussions
Open Access
EGU

Monte Carlo Simulations for Uncertainty Estimation in 3D Geological Modeling, A Guide for Disturbance Distribution Selection and Parameterization1

Evren Pakyuz-Charrier[1], Mark Lindsay[1], Vitaliy Ogarko[1], Jeremie Giraud[1], Mark Jessell[1]

[1]Centre for Exploration Targeting, The University of Western Australia, 35 Stirling Hwy, Crawley WA 6009 Australia

Correspondence to: Evren Pakyuz-Charrier (evren.pakyuz-charrier@research.uwa.edu.au)

Abstract. Three-dimensional (3D) geological modeling aims to determine geological information in a 3D space using structural data (foliations and interfaces) and topological rules as inputs. They are necessary in any project where the properties of the subsurface matters, they express our understanding of geometries in depth. For that reason, 3D geological models have a wide range of practical applications including but not restrained to civil engineering, oil and gas industry, mining industry and water management. These models, however, are fraught with uncertainties originating from the inherent flaws of the modeling engines (working hypotheses, interpolator's parameterization) combined with input uncertainty (observational-, conceptual- and technical errors). Because 3D geological models are often used for impactful decision making it is critical that all 3D geological models provide accurate estimates of uncertainty. This paper's focus is set on the effect of structural input data uncertainty propagation in implicit 3D geological modeling using GeoModeller API. This aim is achieved using Monte Carlo simulation uncertainty estimation (MCUE), a heuristic stochastic method which samples from predefined disturbance probability distributions that represent the uncertainty of the original input data set. MCUE is used to produce hundreds to thousands of altered unique data sets. The altered data sets are used as inputs to produce a range of plausible 3D models. The plausible models are then combined into a single probabilistic model as a means to propagate uncertainty from the input data to the final model. In this paper, several improved methods for MCUE are proposed. The methods pertain to distribution selection for input uncertainty, sample analysis and statistical consistency of the sampled distribution. Pole vector sampling is proposed as a more rigorous alternative than dip vector sampling for planar features and the use of a Bayesian approach to disturbance distribution parameterization is suggested. The influence of inappropriate disturbance distributions is discussed and propositions are made and evaluated on synthetic and realistic cases to address the sighted issues. The distribution of the errors of the observed data (i.e. scedasticity) is shown to affect the quality of prior distributions for MCUE. Results demonstrate that the proposed workflows improve the reliability of uncertainty estimation and diminishes the occurrence of artefacts.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

1 Introduction

Three-dimensional (3D) geological models are important tools for decision making in geoscience as they represent the current state of our knowledge regarding the architecture of the subsurface. As such they are used in various domains of application such as mining (Cammack 2016; Dominy 2002), oil and gas (Nordahl and Ringrose 2008), infrastructure engineering (Aldiss et al. 2012), water supply management (Prada et al. 2016), geothermal power plants (Moeck 2014), waste disposal (Ennis-King and Paterson 2002), natural hazard management (Delgado Marchal et al. 2015), hydrogeology (Jairo 2013) and archaeology (Vos et al. 2015). By definition, all models contain uncertainty, being simplifications of the natural world (Bardossy and Fodor 2001) linked to errors about their inputs (data and working hypotheses), their processing (model building) and output formatting (discretization, simplification). Reason dictates that these models should incorporate an estimate of their uncertainty, which can be equivalent to their reliability for decision making.

Nearly all the methods proposed in the past five years (Wellmann and Regenauer-Lieb 2012; Lindsay et al. 2012; Jessell et al. 2014a; de la Varga and Wellmann 2016) are based on Monte Carlo simulation uncertainty estimation (MCUE). This approach was introduced to geoscience with the Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley 1992) which is a non-predictive (Camacho et al. 2015) implementation of Bayesian Monte Carlo (BMC). Instead of estimating the uncertainty from a single best-guess model, MCUE (Fig. 1) simulates it by producing a large number of potential models through perturbation of the initial input data, the output models are then merged and/or compared to estimate uncertainty. This can be achieved by replacing each original data input with a probability distribution function (PDF) thought to best represent its uncertainty called a disturbance distribution. The disturbance distributions are then sampled to generate many plausible alternate models in a process called perturbation. In that sense, MCUE can be considered as a form of BMC that is focused on uncertainty propagation.

Several metrics have been used to express the final uncertainty, including information entropy (Shannon 1948; Wellmann 2013; Wellmann and Regenauer-Lieb 2012) and stratigraphic variability (Lindsay et al. 2012). The case for reliable uncertainty estimation in 3D geological modeling has been made repeatedly and this paper aims to further improve several points of MCUE methods at the pre-processing steps (Fig. 1). More specifically, (i) the selection of the PDFs used to represent uncertainties related to the original data inputs and (ii) the parameterization of said PDFs. Section 2 reviews the fundamentals of MCUE methods while sect. 3 addresses PDF selection and parameterization, lastly, sect. 4 expands further into the details of disturbance distribution sampling.

2 MCUE method

Recently developed MCUE-based techniques for uncertainty estimation in 3D geological modeling require the user to define the disturbance distribution for each input data, based on some form of prior knowledge. That is necessary because MCUE is a one-step analysis as opposed to a sequential one: all inputs are perturbed once and simultaneously to generate one of the

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

possible models that will be merged or compared with the others. MCUE is vulnerable to erroneous assumptions about the disturbance distribution in terms of structure (what is the optimal type of disturbance distribution) and magnitude (the dispersion parameters) of the uncertainty of the input data. However, it is possible to post-process the results of an MCUE simulation to compare them to other forms of prior knowledge and update accordingly (Wellmann et al. 2014a).

The MCUE approach is usually applied to geometric modeling engines (Wellmann and Regenauer-Lieb 2012; Lindsay et al. 2013; Jessell et al. 2014a; Jessell et al. 2010), although it can be applied to dynamic or kinematic engines (Wang et al. 2016; Wellmann et al. 2015). This choice is motivated by critical differences between the three approaches, both at the conceptual and practical level (Aug 2004). Geometric modeling engines interpolate features from sparse structural data and topological assumptions (Aug et al. 2005; Jessell et al. 2014a); they require prior knowledge of topology and are computationally affordable (Lajaunie et al. 1997; Calcagno et al. 2008). More specifically, explicit geometric engines require full expert knowledge while implicit ones are based on observed field data and topological constraints (Jessell et al. 2014a). Dynamic modeling engines simulate features based on a set of initial mechanical and geometrical properties and the system's boundary conditions; they are computationally expensive. Kinematic modeling engines define the behavior of features beforehand, this entails knowledge of starting conditions (topology and geometry) and kinematic history (Jessell 1981) as well as computationally inexpensive. The implicit geometric approach is preferred for MCUE because knowledge of initial conditions is nearly impossible to achieve and perfect knowledge of current conditions defeats the purpose of estimating any uncertainty. Implicit geometric modeling engines use mainly three types of inputs: interfaces (3D points), foliations (3D vectors) and topological relationships between geological units and faults (stratigraphic column and fault age relationships). Drillholes and other structural inputs such as fold axes and fold axial planes can also be used. Each data input is assigned to a geological unit and the model is then built according to predefined topological rules in order to solve interpolation ambiguities. The implicit geometric 3D modeling package GeoModeller distributed by Intrepid Geophysics was used as a test platform for this study. The use of this specific software is motivated by its open use of co-Kriging (Appendix C) which is a robust (Matheron 1970; Isaaks and Srivastava 1989; Lajaunie 1990) geostatistical interpolator to generate the models (Calcagno et al. 2008; Guillen et al. 2008; FitzGerald et al. 2009). In addition, GeoModeller allows uncertainty to be safely propagated (Chilès et al. 2004; Aug 2004) as co-Kriging quantifies the intrinsic uncertainty about the interpolation itself. Note that MCUE applied to the co-Kriging interpolator used in GeoModeller is, in effect, equivalent to running a geostatistical simulation. In the next section, a series of improvements are proposed to address the disturbance distribution problem.

3 Distribution types and their parameters

Often, the disturbance distribution used to estimate input uncertainty is the same (same type and same parameterization) for all observations of the same nature (Wellmann et al. 2010; Wellmann and Regenauer-Lieb 2012; Lindsay et al. 2012; Lindsay et al. 2013). Disturbance distribution parameters are defined arbitrarily (Lindsay et al. 2012; Wellmann and Regenauer-Lieb 2012) in most cases. Additionally, uniform distributions have been regularly used as disturbance distribution and expressed as

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

a plus minus range over the location of interfaces (Wellmann et al. 2010; Wellmann 2013) or the dip and dip-direction (Lindsay et al. 2012; Lindsay et al. 2013; Jessell et al. 2014a). Here, propositions are made about the type of disturbance distributions that should be used for MCUE, how to parameterize them and associated possible pitfalls.

3.1 Appropriate distributions for MCUE

The structural data collected to build the model is impacted by many independent random sources of uncertainty (Fig. 1) such as measurement, sampling and observation errors (Bardossy and Fodor 2001; Nearing et al. 2016). Additionally, the uncertainty tied to each measurement is considered to be independent to the others. However, that is not to say that there is no dependence over the measured values themselves. For example, dip measurements along a fault line are expected to be spatially correlated though each measurement is an independent trial in terms of its measurement error. Consequently, MCUE may sample from disturbance distributions independently from one another. Under these conditions, the Central Limit Theorem (CLT) holds true for these data (Sivia and Skilling 2006; Gnedenko and Kolmogorov 1954) as the variance of each source of uncertainty is always defined even if it is unknown. Uncertainty will then be best represented by disturbance distributions that are consistent with the CLT, namely the normal distribution for locations (Cartesian scalar data) and the von Mises-Fisher (vMF) distribution for orientations (spherical vector data) (Davis 2003). The normal distribution is the canonical CLT distribution (i.e. the distribution towards which the sum of random variables tends) defined as

(1)

$$\mathcal{N}(x|\varepsilon,\sigma) = \frac{e^{-(x-\varepsilon)^2/2\sigma^2}}{\sigma\sqrt{2\pi}},$$

where ε and σ are the arithmetic mean and standard deviation, respectively. Note that the normal distribution is conjugate to itself or to Student's t-distribution depending on which parameters are known a priori. That is, a normal prior distribution gives a normal or Student posterior distribution in the Bayesian framework.

The vMF distribution (Fig. 2) is the CLT distribution for spherical data; it is the hyperspherical counterpart to the normal distribution (Fisher et al. 1987) and is used under the same general assumptions for unit vectors on the p-dimensional unit hypersphere $S^{(p-1)}$[1]. The most important property of the vMF distribution is the axial symmetry of the data around the mean direction. The vMF distribution is also the maximum entropy distribution for spherical data and is conjugate to itself (Mardia and El-Atoum 1976). These properties make the vMF distribution appropriate for uncertainty analysis of spherical data (Hornik and Grün 2013). Sampling from the vMF distribution is described in Appendix A. The general probability density of the vMF distribution for S^{p-1} is expressed as follows

[1] Here S^{p-1} denotes the surface of the p-dimensional hypersphere.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

(2)

$$p_{\mathrm{vMF}}(x|\gamma,\kappa) = C_p(\kappa)e^{\kappa\gamma^T x}, \kappa > 0 \text{ and } \|\gamma\| = 1\,,^2$$

where γ^T is the transposed mean direction vector and κ is the concentration respectively. κ is analogous to the inverse of σ for the normal distribution. High κ values denote distributions with low variance (Fig. 2), ultimately leading to a p-dimensional hyperspherical Dirac distribution and $\kappa = 0$ means complete randomness (equivalent to a p-dimensional hyperspherical uniform distribution).

$C_p(\kappa)$ is a normalization constant given by

(3)

$$C_p(\kappa) = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)},$$

where $I_v(\kappa)$ is the modified Bessel function of the first kind at order v, and p the dimensionality of S (p=3 for S^2).

3.2 Disturbance distribution parameterization

Regardless of which type of disturbance distribution is chosen, it is inappropriate to use the same distribution with the exact same parameters for each measurement in many cases including but not restricted to: cases where some data inputs are actually a statistic - such as the mean - that is derived from a sample instead of an actual individual occurrence (Moffat 1988); cases where inputs (at the same location) are samples themselves (Kolmogorov 1950); cases where the magnitude of the uncertainty of measurements may be impacted by the value of the measurement itself (Moffat 1982). Statistics derived from samples (e.g. mean, median) or the actual sample are expected to lead to less dispersed disturbance distributions compared to single observations (Patel and Read 1996; Bewoor and Kulkarni 2009; Bucher 2012; Sivia and Skilling 2006; Davis 2003). Disturbance distributions should be parameterized accordingly to avoid adding artificial uncertainty to the model. As structural data inputs are sparse and often scarce, a Bayesian approach to disturbance distribution parameterization is optimal (Sivia and Skilling 2006). The following demonstration applies to both the normal and the vMF distributions for MCUE models.

Assuming that the uncertainty about a input structural data (locations or orientations) can be described by a distribution G

(4)

$$G = p(x|\mu_{\mathrm{true}}, \vartheta_{\mathrm{true}}),$$

where μ_{true} and $\vartheta_{\mathrm{true}}$ are the true mean and dispersion, respectively. Measured data are a sample $X = \{x_1, \ldots, x_n\}$ of G. The disturbance distribution that should be used for MCUE must take into account prior knowledge about $\vartheta_{\mathrm{true}}$ and the observed data X. This is achieved through a simple application of Bayes' theorem

[2] || || denotes the Euclidean norm.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

(5)

$$p(\mu|X,\vartheta) = \frac{p(X|\mu,\vartheta)p(\mu|\vartheta)}{p(X|\vartheta)} \propto p(X|\mu,\vartheta)p(\mu|\vartheta),$$

where μ and ϑ are the expression of prior knowledge about μ_{true} and ϑ_{true}, respectively. The dispersion ϑ_{true} is expected to be estimated based on rigorous metrological studies which methodology is beyond the scope of this paper. Thus, (5) simplifies to

(6)

$$p(\mu|X) = \frac{p(X|\mu)p(\mu)}{p(X)} \propto p(X|\mu)p(\mu).$$

The prior distribution function $p(\mu)$ expresses prior belief about μ. In this case, $p(\mu)$ is defined as Jeffreys improper prior (Sivia and Skilling 2006) for locations to express complete lack of knowledge about $p(\mu)$

(7)

$$p_{\text{loc}}(\mu) = const,$$

and, for the same reason, as a uniform spherical distribution for orientations

(8)

$$p_{\text{ori}}(\mu) = \frac{1}{4\pi}.$$

The likelihood distribution $p(X|\mu)$ expresses the probability of observing X given μ and is given for all possible values of μ, is obtained with the joint density function for X

(9)

$$p(X|\mu) = \prod_{i=1}^{n} p(x_i|\mu).$$

The posterior predictive distribution $p(\hat{x}|X)$ expresses the theoretical distribution of a new observation given X regardless of μ, it is the target disturbance distribution to be sampled for MCUE and is given by

(10)

$$p(\hat{x}|X) = \int p(\mu|X)p(\hat{x}|X,\mu)\, d\mu\,,$$

where $\hat{x}$ is the element to be sampled. For a normal distribution, the posterior predictive distribution is

(11)

$$p^{\text{N}}(\hat{x}|X) \sim \mathcal{N}\left(x \middle| \mu_0, \sigma^2 + \frac{\sigma^2}{n}\right),$$

where μ_0 and σ are the sample mean and prior standard deviation, respectively. Note, that μ_0 is data contribution only while σ is prior reliable knowledge. For vMF distribution the posterior predictive distribution is given by (Bagchi 1987)

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

(12)

$$p^{\mathrm{vMF}}(\hat{x}|X) = \frac{I_0(W)}{\int I_0(\mathrm{W})\, d\hat{x}},$$

with W defined as

(13)

$$W = \sqrt{k^4R^4 + 2k^3R^2\hat{x}^T\frac{\mu_0}{kR} + k^2}\,.$$

Equation 12 has no closed form solution (Bagchi and Guttman 1988). However, it is possible to double sample to get equivalent results or to use an empirical approximation to increase performance for large samples

(14)

$$p^{\mathrm{vMF}}(\hat{x}|X) \sim vMF(x|vMF(\mu_0, \kappa R), \kappa) \approx vMF\left(x\left|\mu_0, \frac{\kappa R}{1+R}\right.\right),$$

where μ_0 and κ and are the mean direction vector of the sample and prior concentration (Appendix B) of the observed sample. Note that equation 11 and equation 14 can be applied to data recorded as a mean value provided that the size of the sample is known.

In equation 14 μ_0 is given by

(15)

$$\mu_0 = \left(\overset{\cdots}{sin}\phi\overset{\cdots}{cos}\theta, \overset{\cdots}{sin}\phi\overset{\cdots}{sin}\theta, \overset{\cdots}{cos}\phi\right),$$

where

(16)

$$\overset{\cdots}{sin}\phi = \sum_{i=1}^{n} sin\phi_i\,; \overset{\cdots}{sin}\theta = \sum_{i=1}^{n} sin\theta_i\,; \overset{\cdots}{cos}\theta = \sum_{i=1}^{n} cos\theta_i\,,$$

where Φ is the colatitude, θ is the longitude and R is the resultant length of the observed sample.

In equation 14 R is given by

(17)

$$R = \left[\left(\overset{\cdots}{sin}\phi\overset{\cdots}{cos}\theta\right)^2 + \left(\overset{\cdots}{sin}\phi\overset{\cdots}{sin}\theta\right)^2 + \left(\overset{\cdots}{cos}\phi\right)^2\right]^{\frac{1}{2}}.$$

From (11) and (14) it appears clearly that sampling from prior distributions directly will lead to systematic underestimation of dispersion. In turn, this bias will narrow the range of models explored by MCUE and will make the final results look less uncertain than they should be. As MCUE is a one-step analysis it does not refine its results iteratively, inappropriate disturbance distributions will not converge to more appropriate ones during the process: proper parameterization of a disturbance distribution at the beginning of the process is then critical to ensure accurate sampling. Bayesian schemes exist to validate

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

models based on some external observations/assumptions (Fig. 1) that are used to build likelihood functions (de la Varga and Wellmann 2016). However, these schemes are known not to yield good results when incorrect informative priors are used (Freni and Mannina 2010; Morita et al. 2010). Prior uncertainty distributions are then inappropriate disturbance distributions and one should instead sample from the posterior predictive distributions (11, 14) for more accurate results about uncertainty.

3.3 Measurement scedasticity

Scedasticity is defined as the distribution of the error about measured or estimated elements of a random variable of interest (Levenbach 1973). It expresses the relationship between the measured values and their uncertainty. In the case where uncertainty is constant across the variable space the variable is homoscedastic (Fig. 3a), such behavior is commonly observed in gravity surveys (Middlemiss et al. 2016). When uncertainty is not constant throughout the variable space, the variable is called heteroscedastic (Fig. 3b, Fig. 3c). Note that heteroscedastic cases include both structured (Fig. 3b) and unstructured (Fig. 3c) relationships between the measured values and their respective errors. Structured heteroscedastic variables show a clear relationship (e.g. correlation, cyclicality) between the variable and its uncertainty while unstructured ones do not. Structured heteroscedastic behavior is observed in physical modeling (Ogarko and Luding 2012), electrical resistivity tomography (Perrone et al. 2014), magnetotellurics (Thiel et al. 2016; Rawat et al. 2014), airborne gravity and magnetics (Kamm et al. 2015) and controlled-source electromagnetic (Myer et al. 2011) surveys. It is usually possible to transform a structured heteroscedastic variable to a space where it becomes homoscedastic (commonly the log space), perform analysis and transform back to the original space. Unstructured heteroscedastic behavior is common in seismic surveys and impacts inversions (Kragh and Christie 2002; Quirein et al. 2000; Eiken et al. 2005). The heteroscedastic case essentially allows for any level of correlation between the measured values and their uncertainty to be possible (Fig. 4).

The failure to account for scedasticity often implies the assumption of homoscedasticity as this assumption allows for a wider range of statistical methods to be applied. With heteroscedastic data, the results of methods that depend on the assumption of homoscedasticity, such as least squares methods (Fig. 3), give results of much decreased quality (Eubank and Thomas 1993) and this may lead to the validation of incorrect hypotheses. Scedasticity analysis from raw data without prior knowledge is challenging (Zheng et al. 2012) and this topic of research is still being investigated (Dosne et al. 2016). If there is no option for an appropriate transform, it is advisable to perform an empirical analysis of scedasticity beforehand. This is usually achieved through experimental assessment of uncertainty under various conditions (metrological study) of measurement and over the entire range of measured values. The results of such analysis can then be used to define the prior dispersion (ϑ in 5) more accurately as a function of the measurement instead of a constant.

Given the above demonstrations, CLT distributions are a better alternative to estimate input uncertainty, namely the Normal distribution for locations and the vMF distribution for orientations (Fisher et al. 1987). Each data input is expected to carry its own parameterization for disturbance distribution depending on the nature of the input (single measurement, sample, central statistic). Additionally, the parameters of the disturbance distributions are better defined when scedasticity is accounted for. It

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

is worth mentioning that both the Normal distribution and the von Mises-Fisher distribution have a complete range of analytical or approximated solutions for both posterior and posterior predictive distributions (Rodrigues et al. 2000; Bagchi and Guttman 1988). In the next section, disturbance distribution sampling for spherical data (orientations) is discussed.

4 Sampling of orientation data for planar features

In the geoscience, the orientation of planar features such as faults and bedding is described by foliations. These foliations are recorded in the form of dip vectors using the dip dip-direction system. This system is equivalent to a reversed right-hand rule spherical coordinates system. The following covers sampling strategies for such spherical data and demonstrates their impact on MCUE results.

4.1 Artificial heteroscedasticity

Recent research using MCUE (Lindsay et al. 2012; Lindsay et al. 2013; Jessell et al. 2014b; Wellmann and Regenauer-Lieb 2012; de la Varga and Wellmann 2016) use dip and dip-direction values independently (as two scalars) from one another. Though the dip dip-direction system is practical for field operators to record and make sense of even without a computer it is inappropriate for statistics. Geoscientists generally perform statistical analysis on stereographic projections of the dip vectors to the planes. Because stereographic projection involves the transform of dip vectors to pole vectors (normal vector to the plane), it gives a sound representation of the underlying prior uncertainty distribution. The pole transform step is essential to avoid variance distortion (Fisher et al. 1987) as shown in Figure 5. The distortion will increase as the dip of the plane diverges from $\frac{\pi}{4}$ and is maximal for degenerate cases of the dip-direction system such as horizontal and vertical planes (Fig. 5). In the case of an uncertain horizontal plane, dip vectors distribute themselves directly below and about the equator of S^2, following a girdle-like distribution (Fig. 6a, Fig. 6b). Consequently, the resultant length is null and the spherical variance S_s^2 (18) equals unity as the barycenter of all dip vectors is located at the center of S^2.

(18)

$$S_s^2 = 1 - \frac{R}{n},$$

Naive interpretation of S_s^2 may lead one to misinterpret uncertainty to be infinite ($S_s^2 = 1$) and the plane's orientation to be uniformly random where it might be very well constrained in reality. That is so because S_s^2 is a scalar quantity used to represents dispersion for samples of spherical unit vectors. Therefore, it is expected that S_s^2 is ambiguous in some cases. The opposite effect occurs for (sub)vertical planes where S_s^2 will appear to be lower than expected. In Figure 6, the effect of dip vector sampling and pole vector sampling is demonstrated for theoretical cases. Here, the blue clusters have constant point

density and are isotropic, parameterization is easy and reliable for distributions such as von Mises-Fisher (Fig. 6b, Fig. 6d, Fig. 6f) or bounded uniform (Fig. 6a, Fig. 6c, Fig. 6e). Blue clusters always describes the plane's behavior accurately in terms of pole vectors; they are the direct result of pole sampling. Green clusters have varying shapes and may not be modelled appropriately by any existing spherical distribution for all possible cases; they are the result of pole sampling converted back to dips and they describe the plane's behavior accurately in terms of dip vectors. Red clusters have constant point density and are isotropic; they are the result of dip sampling and fail to describe accurately the plane's behavior. Therefore, appropriate sampling based on dip vectors (green) is nearly impossible to achieve without increasing the number of parameters of the distributions to take into account the aforementioned effects (i.e. adding a set of functions to compensate for scedasticity errors as well as boundary effects). For example, in a scenario where dip vectors are used directly to estimate a sample's spherical variance or sample over a disturbance distribution, one may attempt to define separate values for dispersion of dip and dip-direction (Lindsay et al. 2012) in order to compensate for scedastic incoherence. A horizontal plane's uncertainty is then obtained by setting circular variance as null over the dip-direction and as any real positive value over the dip. In addition, some form of boundary control or polarity correction of the dip is necessary to remove incorrect occurrences. Note that this method still does not solve the scedasticity issue entirely, especially for high uncertainty values about the dip and vertical dips. Similarly, if dip vectors are used directly, near vertical planes display uniform random behavior of dip direction (Fig. 6e, Fig. 6f) instead of the expected "bow tie" pattern. This pattern is impossible to model appropriately using CLT spherical distributions as they are unimodal and symmetric.

The use of distributions in MCUE makes it very sensitive to scedasticity over inputs. The uncertainty of a dip vector which is quantified by any dispersion parameter similar to S_s^2 will show non-systematic heteroscedasticity because of variance distortion. A plane dipping at any angle would show increased heteroscedasticity of its uncertainty as the dispersion parameter used to parameterize the underlying distribution increases. Note that uncertain planes show increased heteroscedasticity as their dip diverges from a 45 degrees dip (Fig. 6c, Fig. 6d). Boundary effects also play a role for horizontal and vertical limit cases (Fig. 5, Fig. 6) as the dip, dip-direction system is constrained to $0 - \frac{\pi}{2}$. That of course considerably lowers the quality of any subsequent procedure that relies on proper propagation of uncertainty as these planes are expected to have the lowest uncertainty in terms of dip direction. These impracticalities make it generally better to work on pole vectors rather than dip vectors. Pole vectors mostly eliminate the need for variance correction and allow coherent sampling over a plane's orientation (Fisher et al. 1987). The pole vector transform is widely used in structural geology (Phillips 1960; Wallace 1951; Lisle and Leyshon 2004) through stereographic projection. Therefore, data collected as dip vectors (green clusters, Fig. 6) must be transformed to poles (blue clusters, Fig. 6) for proper estimation of spherical variance. Disturbance distributions should then be defined and sampled based on the pole vectors (mean of blue clusters, Fig. 6), as described in sect. 3.2, instead of the mean dip vector (mean of green clusters, Fig. 6) to avoid distortion (red clusters, Fig. 6). The sample can then be converted back to dip vectors if required.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

4.2 Impact of pole vector sampling versus dip vector sampling

The impact of pole versus dip vector sampling on the results of MCUE is evaluated on a simple synthetic model and on a realistic synthetic model. The simple model is a standard symmetric graben with four horizontal units, it has been chosen for its simplicity and is commonly used as a test case (Wellmann et al. 2014b; de la Varga and Wellmann 2016; Chilès et al. 2004) in MCUE for proof of concepts. The realistic model is a modification of a real demonstration case that is part of the GeoModeller package based on a location near Mansfield, Victoria, Australia. It features a Carboniferous sedimentary basin oriented NW-SE that is in a faulted contact (Mansfield Fault) on its SW edge to a Siluro-Devonian set of older, folded basins. Outcropping units are almost all of the siliceous detritic type, the basement is made of Ordovicio-Cambrian serpentinized sandstone. The original data for the Mansfield model was not altered in any way, instead data based on the Mansfield geological map (Cayley et al. 2006) geophysical map (Haydon et al. 2006) and airborne geophysical survey (Wynne and Bacchin 2009; Richardson 2003) were added to refine it.

The graben model is built using orientations and interfaces only, with 3 interfaces and 3 foliations per unit and 1 interface and 1 foliation per fault (Fig. 7, Fig. 8c). The Mansfield model is built with 281 interface points and 176 foliations over 6 units and 3 faults (Fig. 9, Fig. 10c). For both models, perturbation is performed as described in section 3. For the graben model, units' interfaces are isotropically perturbed over a normal distribution with the mean centered on the original data point and standard deviation of 25. The orientations of the faults are perturbed over a von Mises-Fisher distribution with the original data as the mean vector and concentration of 100 (p95 ~ ±10 degrees) following the recommended pole vector procedure described in sect. 4.1 (Fig. 8a, Fig. 10a) or the dip vector one (Fig. 8b, Fig. 10b). For the Mansfield model, all interfaces and orientations (both for units and faults) are perturbed using the parameterization given for the graben model. The perturbation parameters for orientations were chosen to be compatible with metrological data (Nelson et al. 1987; Stigsson 2016) while perturbation parameters for interfaces were designed to meet observed experimental interface variability (Courrioux et al. 2015; Lark et al. 2014; Lark et al. 2013) and observed GPS uncertainty (Jennings et al. 2010).

The influence of dip vector (Fig. 8b, Fig. 10b) versus pole vector (Fig. 8a, Fig. 10a) sampling of orientations is very noticeable over the output entropy uncertainty models. Dip vector sampling appears to add a layer of artificial "noise" on top of the uncertainty models. The "noise" prevents expected structures of the starting model (Fig. 8c, Fig. 10c) to be easily distinguishable. In cases where the orientation data is more vulnerable to improper sampling error (away from 45° dips) important structures may completely disappear such as the near vertical faults in the graben model (Fig. 8) or the circled areas in Fig. 10. It also appears that areas where low uncertainty would be expected (orange unit in Fig. 10) are the loci of excess uncertainty. These observations support the assertion that pole vector sampling should be favored to improve uncertainty propagation in MCUE.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

5 Discussion

As described in sect. 3.1, CLT distributions should be preferred as prior uncertainty distributions (and disturbance distributions) because they better describe the behavior of uncertainty. However, there may be scenarios where alternatives can offer a better solution. More specifically, the uniform or the Laplace distribution may better describe location uncertainty than the normal distribution. The uniform distribution indicates a lack of constraints as to the prior uncertainty distribution, it is a valid choice when there is little knowledge about data dispersion. The Laplace distribution is appropriate if the measured data abide by the first law of errors instead of the second (Wilson 1923). That is, the Laplace distribution "replaces" the normal distribution. Under the same circumstances, a spherical exponential distribution could be swapped with the vMF distribution. The Kent[3] distribution is also a good candidate to describe orientation uncertainty when the pole vectors of measured orientations appear to be anisotropically distributed on S^2 (Kent and Hamelryck 2005).

In subsection 3.2, it is explicitly assumed that the dispersion of prior uncertainty distributions is a deterministic function. Note that this does not necessarily make this function a constant and it might depend on the observed data. The dispersion function of field measurements (using a compass) of structural data would be expected to be nearly constant. Conversely, the dispersion function of interpreted measurements (using geophysics) would be expected to be dependent on the sensitivity of the intermediary method. Additionally, dispersion functions may be probabilistic as well as deterministic (Bucher 2012). Determinism is a strong assumption when no metrological study was conducted beforehand to assess its plausibility. Such metrological studies involve experimental testing of devices and procedures in order to estimate precision, accuracy, bias, scedasticity or drift about measured data. These estimates can then be compiled into a dispersion function that can be used as input parameter for other purposes, including prior uncertainty distributions for MCUE. Probabilistic dispersion functions imply non-negligible uncertainty onto the dispersion function for prior uncertainty distributions. Uncertainty about dispersion makes the proposed workflow in sect. 3.2 inadequate for disturbance distribution parameterization. Indeed, (5) may not be simplified into (6) anymore and the following statements (7 to 14) would then ignore the probabilistic nature of the dispersion function. Both the normal and the vMF distributions have analytical solutions or good approximations for such cases, the authors recommend the readers to refer to relevant works (Gelman et al. 2014; Bagchi and Guttman 1988) if required. Note that there is significant metrological work about borehole data (Nelson et al. 1987; Stigsson 2016) as opposed to usual structural data such as foliations, fold planes, fold axes or interfaces.

Although Subsection 3.3 makes the case for scedasticity analysis in MCUE, it is left open in this paper. Scedasticity is essentially an untouched subject in geological 3D modeling and it was pointed out to make the geological 3D modeling community aware of this fact and its potentially nefarious influence on MCUE outputs. However, standard metrological studies

3 . The Kent distribution is the spherical analogue to a bivariate normal distribution, it takes an additional concentration parameter along with a covariance matrix. Together, these two parameters allow for any level of elliptic anisotropy on S^2.

can determine scedasticity and include it into a dispersion function to be a parameter of the prior uncertainty distributions (Bewoor and Kulkarni 2009; Bucher 2012).

Section 4 shows the impact of pole vector sampling versus dip vector sampling for orientations that are supposed to describe planar features. The evidence brought at the theoretical (sect. 4.1) and practical (sect. 4.2) levels allows to strongly advocate for the use of pole vectors over dip vectors. This especially applies to MCUE methods where Bayesian post-analysis is performed onto the probabilistic model that results from basic propagation of uncertainty (de la Varga and Wellmann 2016). In this respect, dip vector sampling leads to incorrect highly informative prior distributions which is catastrophic for any Bayesian methods (Morita et al. 2010). Nonetheless, it is worth mentioning that the arguments in sect. 4.1 only apply to dip vectors of a plane and should not be extended to actual vector structural data such as fold axes or lineations. That is so because these data represent linear features (lineations, fold axes, other planar features intersections) for which the concept of a pole does not apply.

As demonstrated in sect. 3 and sect. 4, good prior knowledge about input uncertainty is critical to the propagation of uncertainty in general. This, in turn, makes metrological work a mandatory step to any form of modeling that relies on actual measured data. However, there is another major source of uncertainty that stems from the necessarily imperfect modeling engine itself. Implicit geometric modeling engines (in this case, GeoModeller) use interpolation to draw the contact surfaces of geological units. Therefore, the parameterization of the interpolator may impact results. The co-Kriging interpolator (Appendix C) used in this paper relies on (uncertain) variographic analysis (28) and is natively able to express its own uncertainty (29). Therefore, these sources of uncertainty are expected to be propagated along the input uncertainty as hyperparameter in equation 5.

6 Conclusion

Propagation of uncertainty is the process through which different kinds and sources of uncertainties about the same phenomenon are combined into a single final estimate. MCUE methods seek to achieve propagation of uncertainty using Monte Carlo based systems where input uncertainty is simulated through the sampling of probability distributions called a disturbance distribution. Disturbance distributions are the distributions that normally best represents the uncertainty about the input data. In the context of uncertainty propagation in geological 3D modeling.

This paper discusses the importance of disturbance distribution selection, proposes a simple procedure for better disturbance distribution parameterization and a pole vector based sampling routine for spherical data (orientations) used to represent the geometry of planar features. Pole vector based sampling for spherical data and Bayesian disturbance distribution parameterization are proved - either through demonstration or through experiment - to always be more optimal choices for MCUE applied to implicit 3D geological modeling. Namely, the normal and the vMF distributions are shown to be best candidates for disturbance distributions for location and orientations, respectively. Bayesian parameterization is shown to avoid underestimation of dispersion for disturbance distributions. Which is important as such underestimation artificially decreases the output uncertainty of the 3D geological models. Such underestimation may give a false sense of confidence and

lead to poor decision making. Pole vector sampling is evidenced to be the best alternative because it is guaranteed not to distort the disturbance distributions shape or generate artefacts in the output uncertainty models the way dip vector sampling does.

The proposed framework and methods are compatible with previous MCUE work on 3D geological modeling and can be added easily to existing implementations to improve their accuracy. As MCUE is applicable to all fields where 3D geological models are needed, so is the proposed framework. The primary domains of application are the mining and oil and gas industry at the exploration, development and production steps. In addition, numerous secondary domains of potential application are available to this work, such civil engineering and fundamental research.

7 Data/Code availability

Both the Mansfield and graben GeoModeller models (including the perturbed datasets and series of plausible models) showcased in the present study are available online openly at https://doi.org/10.5281/zenodo.848225 and https://doi.org/10.5281/zenodo.854730 respectively. Instructions on how to use the GeoModeller API can be found at http://www.intrepid-geophysics.com/ig/index.php?page=geomodeller-api.

Although proprietary, the GeoModeller software is available for a fully enabled one-month trial period at http://www.intrepid-geophysics.com/ig/index.php?page=downloads.

8 Appendices

Appendix A: von Mises-Fisher pseudo random number generation

Von Mises-Fisher sampling on the usual sphere is not new (Wood 1994) and this appendix serves as a reminder for the reader. To generate a von Mises-Fisher distributed pseudo random spherical 3D unit vector X_{sphe} on S^2 for a given mean direction μ and concentration κ, define

$$X_{\text{sphe}} = [\phi, \theta, r]. \quad \textbf{(19)}$$

For $\mu = [0, (.), 1]$ the pseudo random vector is given by

$$X_{\text{sphe}} = [arcosW, V, 1], \quad \textbf{(20)}$$

V is given by

$$V \sim U(0, 2\pi), \quad \textbf{(21)}$$

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

U(a,b) is the uniform distribution on[a,b]. W is given by

(22)

$$W = 1 + \frac{1}{\kappa}\left(ln\xi + ln\left(1 - \frac{\xi - 1}{\xi} e^{-2\kappa}\right)\right),$$

where

(23)

$$\xi \sim U(0,1).$$

Note that in equation 22, W is undefined for $\xi=0$ and it should be set to W=-1; in this case. X_{sphe} should then be rotated to be consistent with the chosen μ.

Appendix B: Parameter estimates for von Mises-Fisher

The maximum likelihood estimation $\hat{\mu}$ of μ for a given sample of n unit vectors on S^2 is the mean direction vector

(24)

$$\hat{\mu} = \frac{R}{\|R\|},$$

a simple approximation of the concentration parameter $\hat{\kappa}$ is estimated by (Banerjee et al. 2005)

(25)

$$\hat{\kappa} = \frac{\bar{R}(p - \bar{R}^2)}{1 - \bar{R}^2},$$

where

(26)

$$\bar{R} = \frac{R}{n}.$$

More refined techniques are available to compute this last estimation (Sra 2011) though they do not produce significantly better results for low dimensionality cases (p < 5) with high values of κ. Thus, it is recommended to use the above.

Appendix C: co-Kriging algorithm in GeoModeller

The co-Kriging algorithm used in GeoModeller interpolates a 3D vector field and converts it into a potential (scalar) field (Calcagno et al. 2008; Guillen et al. 2008) that is then contoured to draw interface surfaces. The space between surfaces is defined as belonging to a specific unit based on topological rules. The topological rules are set by (i) the stratigraphic column

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

for units versus units topological rules (ii) the fault network matrix for faults versus faults topological rules (iii) the fault affectation matrix for faults versus units topological rules.

The potential field co-Kriging interpolator is

(27)

$$T^*(p) - T^*(p_0) = \sum_{\alpha=1}^{M} \mu_\alpha \left(T(p_\alpha) - T(p'_\alpha)\right) + \sum_{\beta=1}^{N} \nu_\beta \frac{\partial T}{\partial u_\beta}(p_\beta),$$

where $T^*(p)$-$T^*(p_0)$ is the potential difference at the point p given an arbitrary constant origin point p_0. The weights μ_α and ν_β are the unknowns. M is the number of interfaces and N is 3 times the number of foliations. For practical purposes, the modelled random function $T(.)$ is considered to be affected by a polynomial drift that is deduced from the foliation data (Chilès and Delfiner 2009). The theoretical semi-variogram is obtained through variographic analysis, it is then used to solve equation 27 and is usually of the cubic (28) type (Calcagno et al. 2008)

(28)

$$\gamma(d) = C\left(7\left(\frac{d}{a}\right)^2 - \frac{35}{4}\left(\frac{d}{a}\right)^3 + \frac{7}{2}\left(\frac{d}{a}\right)^5 - \frac{3}{4}\left(\frac{d}{a}\right)^7\right).$$

where d and a are the lag distance and the range, respectively. The theoretical semi-variogram is fit to an empirical semi-variogram (Matheron 1970). In practical cases, the empirical to theoretical semi-variogram fit is never perfect and is mostly parametric. The probability that the potential value estimated at a point $\boldsymbol{x}$ is comprised between t and t' (Aug 2004; Chilès et al. 2004) is given by

(29)

$$P(t \leq T^*(p) - T^*(p_0) < t') = G\left(\frac{t' - t}{\sigma_{CK}(\boldsymbol{x})}\right),$$

where $\sigma_{CK}(\boldsymbol{x})$ is the co-Kriging standard deviation and G is the normal cumulative distribution function. Equation 29 can be used as an uncertainty estimator for the interpolator (27) if and only if both t and t' can be defined adequately as equivalent to the top and bottom of a formation. If it happens to be the case these probabilities can be combined to the final uncertainty model at the merging step (Fig. 1.). However, such definition is not always possible. Note that Kriging can be redefined in the Bayesian framework (Aug 2004; Omre 1987) where its assumptions of normality are considered as prior knowledge and therefore may be challenged/modified (Pilz and Spöck 2008).

9 Competing interests

The authors declare that they have no conflict of interest.

10 Acknowledgements

The authors would like to thank the Geological Survey of Western Australia, the Western Australian Fellowship Program and the Australian Research Council for their financial support. In addition, the authors make special distinction to Intrepid Geophysics for their outstanding technical support.

11 References

Aldiss, D. T., Black, M. G., Entwisle, D. C., Page, D. P., and R.L., T.: Benefits of a 3D geological model for major tunnelling works: an example from Farringdon, east-central London, UK Q J Eng Geol Hydrogeol, 45, 22, 10.1144/qjegh2011-066, 2012.

Aug, C.: Modelisation geologique 3D et caracterisation des incertitudes par la methode du champ de potentiel, PhD, Ecole des Mines de Paris, Paris, 220 pp., 2004.

Aug, C., Chilès, J.-P., Courrioux, G., and Lajaunie, C.: 3D geological modelling and uncertainty: The potential-field method, in: Geostatistics Banff 2004, Springer, 145-154, 2005.

Bagchi, P.: Bayesian analysis of directional data, University of Toronto, Ottawa, Ont: National Library of Canada., 1987.

Bagchi, P., and Guttman, I.: Theoretical considerations of the multivariate von Mises-Fisher distribution, J Appl Stat, 15, 149-169, 1988.

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S.: Clustering on the Unit Hypersphere using von Mises-Fisher Distributions, J Mach Learn Res, 1345–1382, 2005.

Bardossy, G., and Fodor, J.: Traditional and NewWays to Handle Uncertainty in Geology, Nat Ressour Res, 10, 9, 2001.

Beven, K., and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrol Process, 6, 279-298, 1992.

Bewoor, A. K., and Kulkarni, V. A.: Metrology and measurement, McGraw-Hill Education, New Delhi, 2009.

Bucher, J. L.: The metrology handbook, edited by: Bucher, J. L., ASQ Quality Press, United States of America, 2012.

Calcagno, P., Chilès, J. P., Courrioux, G., and Guillen, A.: Geological modelling from field data and geological knowledge, Phys Earth Planet Inter, 171, 147-157, 10.1016/j.pepi.2008.06.013, 2008.

Camacho, R. A., Martin, J. L., McAnally, W., Díaz-Ramirez, J., Rodriguez, H., Sucsy, P., and Zhang, S.: A comparison of Bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modeling, J Am Water Ressour Assoc, 51, 1372-1393, 2015.

Cammack, R.: Developing an engineering geological model in the fractured and brecciated rocks of a copper porphyry deposit, Geol Soc Lond Eng Geol Spec Publ, 27, 93-100, 10.1144/egsp27.8, 2016.

Cayley, R. A., Osborne, C. R., and Vanderberg, A. H. M.: Mansfield 1:50 000 geological map, Geological Survey of Victoria. GeoScience Victoria. Department of Primary Industries., Melbourne, 2006.

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, Orebody Modelling and Strategic Mine Planning, Perth, 22/11/2004, 2004.

Chilès, J. P., and Delfiner, P.: Geostatistics: modeling spatial uncertainty, John Wiley & Sons, New Jersey, 2009.

Courrioux, G., Allanic, C., Bourgine, B., Guillen, A., Baudin, T., Lacquement, F., Gabalda, S., Cagnard, F., Le Bayon, B., and Besse, J.: Comparisons from multiple realizations of a geological model. Implication for uncertainty factors identification, IAMG 2015: The 17th annual conference of the International Association for Mathematical Geosciences, 2015,

Davis, J. C.: Statistics and Data Analysis in Geology, 3rd ed., Wiley, New Jersey, 656 pp., 2003.

de la Varga, M., and Wellmann, J. F.: Structural geologic modeling as an inference problem: A Bayesian perspective, Interpretation, 4, SM1-SM16, 2016.

Delgado Marchal, J., Garrido Manrique, J., Lenti, L., López Casado, C., Martino, S., and Sierra, F. J.: Unconventional pseudostatic stability analysis of the Diezma landslide (Granada, Spain) based on a high-resolution engineering-geological model, 10.1016/j.enggeo.2014.11.002, 2015.

Dominy, S. C. N., Mark A.;Annels,Alwyn E.: Errors and Uncertainty in Mineral Resource and Ore Reserve Estimation: The Importance of Getting it Right, Explor Min Geol, 11, 22, 2002.

Dosne, A.-G., Bergstrand, M., and Karlsson, M. O.: A strategy for residual error modeling incorporating scedasticity of variance and distribution shape, J Pharmacokinet Pharmacodyn, 43, 137-151, 2016.

Eiken, O., Haugen, G. U., Schonewille, M., and Duijndam, A.: D. A Proven Method for Acquiring Highly Repeatable Towed Streamer Seismic Data, in: Insights and Methods for 4D Reservoir Monitoring and Characterization, Society of Exploration Geophysicists and European Association of Geoscientists and Engineers, 209-216, 2005.

Ennis-King, J., and Paterson, L.: Engineering Aspects of Geological Sequestration of Carbon Dioxide, Asia Pacific Oil and Gas Conference and Exhibition Melbourne, Australia, 08/10/2002, 2002.

Eubank, R., and Thomas, W.: Detecting heteroscedasticity in nonparametric regression, J R Stat Soc Ser B Methodol, 145-155, 1993.

Fisher, N. I., Lewis, T., and Embleton, B. J.: Statistical analysis of spherical data, Cambridge university press, Cambridge, 1987.

FitzGerald, D., Chilès, J. P., and Guillen, A.: Delineate Three-Dimensional Iron Ore Geology and Resource Models Using the Potential Field Method, Iron Ore Conference, Perth, 27/07/2009, 2009.

Freni, G., and Mannina, G.: Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution, J Hydrol, 392, 31-39, 2010.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian data analysis, Chapman & Hall/CRC Boca Raton, FL, USA, 2014.

Gnedenko, B., and Kolmogorov, A.: Limit distributions for sums of independent, Amer J Math, 105, 28-35, 1954.

Guillen, A., Calcagno, P., Courrioux, G., Joly, A., and Ledru, P.: Geological modelling from field data and geological knowledge, Phys Earth Planet Inter, 171, 158-169, 10.1016/j.pepi.2008.06.014, 2008.

Haydon, S. J., Skladzien, P. B., and Cayley, R. A.: Parts of Mansfield Alexandra and Euroa 1:100 000 maps: Geological interpretation of geophysical features map., Geological Survey of Victoria. Geoscience Victoria. Deparment of Primary Industries, Melbourne, 2006.

Hornik, K., and Grün, B.: On conjugate families and Jeffreys priors for von Mises–Fisher distributions, J Stat Plan Inference, 143, 992-999, 2013.

Isaaks, E. H., and Srivastava, R. M.: Applied Geostatistics, Oxford University Press, Inc., New York, 561 pp., 1989.

Jairo, N.: Estimation and propagation of parameter uncertainty in lumped hydrological models: A case study of HSPF model applied to luxapallila creek watershed in southeast USA, J Hydrogeol Hydrol Eng, 2013.

Jennings, D., Cormack, S., Coutts, A. J., Boyd, L., and Aughey, R. J.: The validity and reliability of GPS units for measuring distance in team sport specific running patterns, Int J Sport Physiol Perform, 5, 328-341, 2010.

Jessell, M.: Noddy: an interactive map creation package, Unpublished MSc Thesis, University of London, 1981.

Jessell, M., Aillères, L., and de Kemp, E. A.: Towards an integrated inversion of geoscientific data: What price of geology?, Tectonophys, 490, 294-306, 10.1016/j.tecto.2010.05.020, 2010.

Jessell, M., Aillères, L., de Kemp, E. A., Lindsay, M. D., Wellmann, J. F., Hillier, M., Laurent, G., Carmichael, T., and Martin, R.: Next Generation Three-Dimensional Geologic Modeling and Inversion, in: Society of Economic Geologists Special Publication 18, Society of Economic Geologists, 12, 2014a.

Kamm, J., Lundin, I. A., Bastani, M., Sadeghi, M., and Pedersen, L. B.: Joint inversion of gravity, magnetic, and petrophysical data—A case study from a gabbro intrusion in Boden, Sweden, Geophys, 80, B131-B152, 2015.

Kent, J. T., and Hamelryck, T.: Using the Fisher-Bingham distribution in stochastic models for protein structure, Quantitative Biology, Shape Analysis, and Wavelets, 24, 57-60, 2005.

Kolmogorov, A. N.: Foundations of the Theory of Probability, 1950.

Kragh, E., and Christie, P.: Seismic repeatability, normalized rms, and predictability, Lead Edge, 21, 640-647, 2002.

Lajaunie, C.: Comparing Some Approximate Methods for Building Local Confidence Intervals for Predicting Regionalized Variables, Math Geol, 22, 22, 10.1007/BF00890301, 1990.

Lajaunie, C., Courrioux, G., and Manuel, L.: Foliation fields and 3D cartography in geology: principles of a method based on potential interpolation, Math Geol, 29, 571-584, 1997.

Lark, R., Thorpe, S., Kessler, H., and Mathers, S.: Interpretative modelling of a geological cross section from boreholes: sources of uncertainty and their quantification, Solid Earth, 5, 1189, 2014.

Lark, R. M., Mathers, S. J., Thorpe, S., Arkley, S. L. B., Morgan, D. J., and Lawrence, D. J. D.: A statistical assessment of the uncertainty in a 3-D geological framework model, Proc Geol Assoc, 124, 946-958, 10.1016/j.pgeola.2013.01.005, 2013.

Levenbach, H.: The estimation of heteroscedasticity from a marginal likelihood function, J Am Stat Assoc, 68, 436-439, 1973.

Lindsay, M. D., Aillères, L., Jessell, M., de Kemp, E. A., and Betts, P. G.: Locating and quantifying geological uncertainty in three-dimensional models: Analysis of the Gippsland Basin, southeastern Australia, Tectonophys, 546-547, 10-27, 10.1016/j.tecto.2012.04.007, 2012.

Lindsay, M. D., Perrouty, S., Jessell, M., and Aillères, L.: Making the link between geological and geophysical uncertainty: geodiversity in the Ashanti Greenstone Belt, Geophys J Int, 195, 903-922, 10.1093/gji/ggt311, 2013.

Lisle, R. J., and Leyshon, P. R.: Stereographic projection techniques for geologists and civil engineers, Cambridge University Press, Cambridge, 2004.

Mardia, K. V., and El-Atoum, S.: Bayesian inference for the von Mises-Fisher distribution, Biom, 63, 203-206, 1976.

Matheron, G.: La Theorie des Variables Regionalisees et ses Applications, Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau, ENSMP, Paris, 220 pp., 1970.

Middlemiss, R., Samarelli, A., Paul, D., Hough, J., Rowan, S., and Hammond, G.: Measurement of the Earth tides with a MEMS gravimeter, Nat, 531, 614-617, 2016.

Moeck, I. S.: Catalog of geothermal play types based on geologic controls, Renew Sustain Energy Rev, 37, 867-882, 10.1016/j.rser.2014.05.032, 2014.

Moffat, R. J.: Contributions to the theory of single-sample uncertainty analysis, ASME Trans J Fluids Eng, 104, 250-258, 1982.

Moffat, R. J.: Describing the uncertainties in experimental results, Exp Therm Fluid Sci, 1, 3-17, 1988.

Morita, S., Thall, P. F., and Müller, P.: Evaluating the impact of prior assumptions in Bayesian biostatistics, Stat Biosci, 2, 1-17, 2010.

Myer, D., Constable, S., and Key, K.: Broad-band waveforms and robust processing for marine CSEM surveys, Geophys J Int, 184, 689-698, 2011.

Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijs, S. V.: A philosophical basis for hydrological uncertainty, Hydrol Sci J, 61, 1666-1678, 10.1080/02626667.2016.1183009, 2016.

Nelson, R., Lenox, L., and Ward Jr, B.: Oriented core: its use, error, and uncertainty, AAPG Bull, 71, 357-367, 1987.

Nordahl, K., and Ringrose, P. S.: Identifying the representative elementary volume for permeability in heterolithic deposits using numerical rock models, Math Geosci, 40, 753-771, 2008.

Ogarko, V., and Luding, S.: Equation of state and jamming density for equivalent bi-and polydisperse, smooth, hard sphere systems, J Chem Phys, 136, 124508, 2012.

Omre, H.: Bayesian kriging—merging observations and qualified guesses in kriging, Math Geol, 19, 25-39, 1987.

Patel, J. K., and Read, C. B.: Handbook of the normal distribution, CRC Press, New York, 1996.

Perrone, A., Lapenna, V., and Piscitelli, S.: Electrical resistivity tomography technique for landslide investigation: A review, Earth-Science Reviews, 135, 65-82, 2014.

Phillips, F. C.: The use of stereographic projection in structural geology, Edward Arnold, London, 1960.

Pilz, J., and Spöck, G.: Why do we need and how should we implement Bayesian kriging methods, Stoch Env Res Risk Assess, 22, 621-632, 2008.

Prada, S., Cruz, J. V., and Figueira, C.: Using stable isotopes to characterize groundwater recharge sources in the volcanic island of Madeira, Portugal, J Hydrol, 536, 409-425, 10.1016/j.jhydrol.2016.03.009, 2016.

Quirein, J., Hampson, D., and Schuelke, J.: Use of Multi-Attribute Transforms to Predict Log Properties from Seismic Data, EAGE Conference on Exploring the Synergies between Surface and Borehole Geoscience-Petrophysics meets Geophysics, 2000,

Rawat, G., Arora, B., and Gupta, P.: Electrical resistivity cross-section across the Garhwal Himalaya: Proxy to fluid-seismicity linkage, Tectonophys, 637, 68-79, 2014.

Rodrigues, J., Galvão Leite, J., and Milan, L. A.: Theory & Methods: An Empirical Bayes Inference for the von Mises Distribution, Aust N Z J Stat, 42, 433-440, 2000.

Sadegh, M., and Vrugt, J. A.: Approximate bayesian computation using Markov chain Monte Carlo simulation: DREAM (ABC), Water Ressour Res, 50, 6767-6787, 2014.

Shannon, C. E.: A Mathematical Theory of Communication, Bell Syst Tech J, 27, 55, 1948.

Sivia, D. S., and Skilling, J.: Data Analysis A Bayesian Tutorial, 2nd ed., Oxford Science Publications, Oxford University Press, Oxford, 246 pp., 2006.

Sra, S.: A short note on parameter approximation for von Mises-Fisher distributions: and a fast implementation of I s (x), Comput Stat, 27, 177-190, 10.1007/s00180-011-0232-x, 2011.

Stigsson, M.: Orientation Uncertainty of Structures Measured in Cored Boreholes: Methodology and Case Study of Swedish Crystalline Rock, Rock Mech Rock Eng, 49, 4273-4284, 2016.

Thiel, S., Heinson, G., Reid, A., and Robertson, K.: Insights into lithospheric architecture, fertilisation and fluid pathways from AusLAMP MT, ASEG Ext Abstr, 2016, 1-6, 2016.

Vos, P. C., Bunnik, F. P. M., Cohen, K. M., and Cremer, H.: A staged geogenetic approach to underwater archaeological prospection in the Port of Rotterdam (Yangtzehaven, Maasvlakte, The Netherlands): A geological and palaeoenvironmental case study for local mapping of Mesolithic lowland landscapes, Quat Int, 367, 4-31, 10.1016/j.quaint.2014.11.056, 2015.

Wallace, R. E.: Geometry of shearing stress and relation to faulting, J Geol, 59, 118-130, 1951.

Wang, H., Wellmann, J. F., Li, Z., Wang, X., and Liang, R. Y.: A Segmentation Approach for Stochastic Geological Modeling Using Hidden Markov Random Fields, Math Geosci, 10.1007/s11004-016-9663-9, 2016.

Wellmann, J., Thiele, S., Lindsay, M., and Jessell, M.: pynoddy 1.0: an experimental platform for automated 3-D kinematic and potential field modelling, Geosci. Model Dev Discuss, 8, 10011-10051, 2015.

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

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

Wellmann, J. F.: Information Theory for Correlation Analysis and Estimation of Uncertainty Reduction in Maps and Models, Entropy, 15, 1464-1485, 10.3390/e15041464, 2013.

Wellmann, J. F., Finsterle, S., and Croucher, A.: Integrating structural geological data into the inverse modelling framework of iTOUGH2, Comput Geosci, 65, 95-109, 10.1016/j.cageo.2013.10.014, 2014a.

Wellmann, J. F., Lindsay, M. D., Poh, J., and Jessell, M. W.: Validating 3-D Structural Models with Geological Knowledge for Improved Uncertainty Evaluations, Energy Procedia, 59, 374-381, 10.1016/j.egypro.2014.10.391, 2014b.

Wilson, E. B.: First and second laws of error, J Am Stat Assoc, 18, 841-851, 1923.

Wood, A. T.: Simulation of the von Mises Fisher distribution, Commun Stat Simul Comput, 23, 157-164, 1994.

Zheng, H., Xie, J., and Jin, Z.: Heteroscedastic sparse representation based classification for face recognition, Neural Process Lett, 35, 233-244, 2012.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

12 Figures

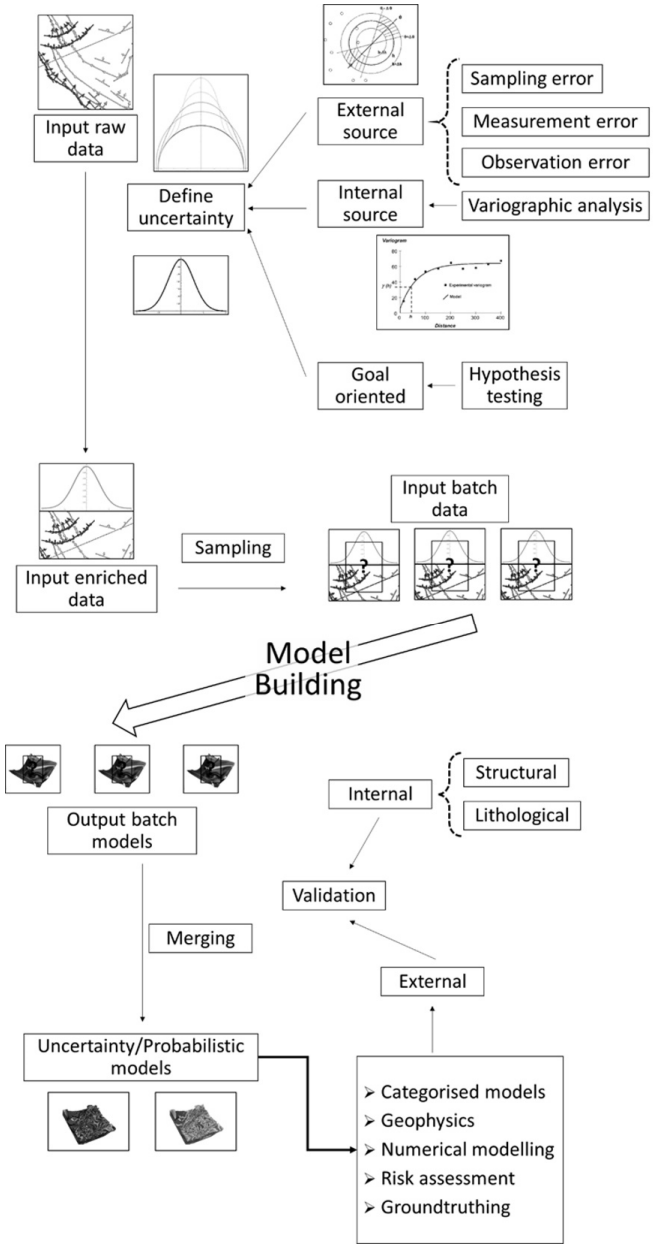

Figure 1: Monte Carlo Uncertainty Estimation procedure workflow.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

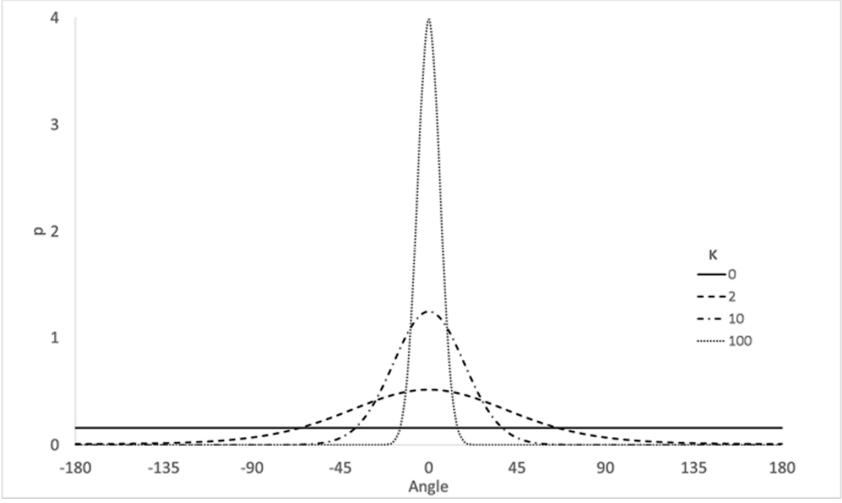

Figure 2: von Mises-Fisher probability distribution function on S1 (p = 2) for various concentrations κ.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

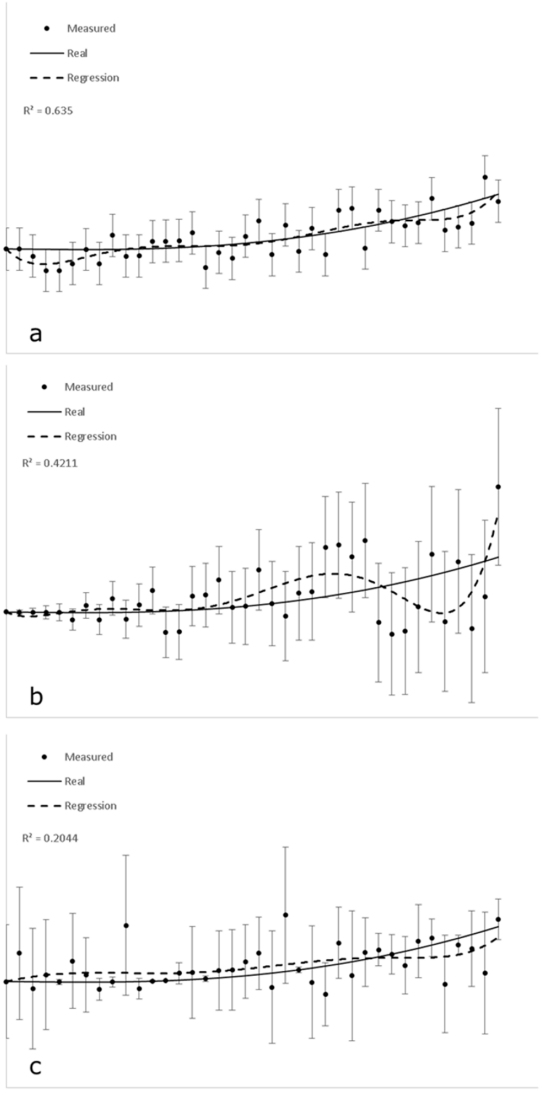

Figure 3: Synthetic examples of different levels of scedasticity of measurements of the same variable. (a) homoscedastic case, (b) structured heteroscedastic case and (c) unstructured heteroscedastic case. Note how the least square polynomial residuals score (R2) is heavily impacted by scedasticity.

5

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

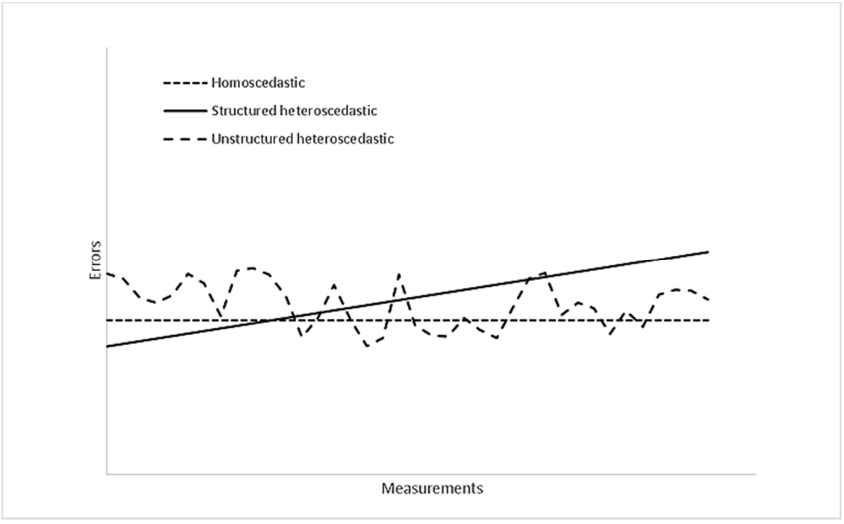

Figure 4: Distribution of errors for the cases described in Figure 3. Homoscedastic case shows constant uncertainty and no relationship of uncertainty to the data. The structured heteroscedastic case has a linear relationship of uncertainty to the data. The unstructured heteroscedastic case demonstrates no obvious relationship of uncertainty to the data and is not constant.

5

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

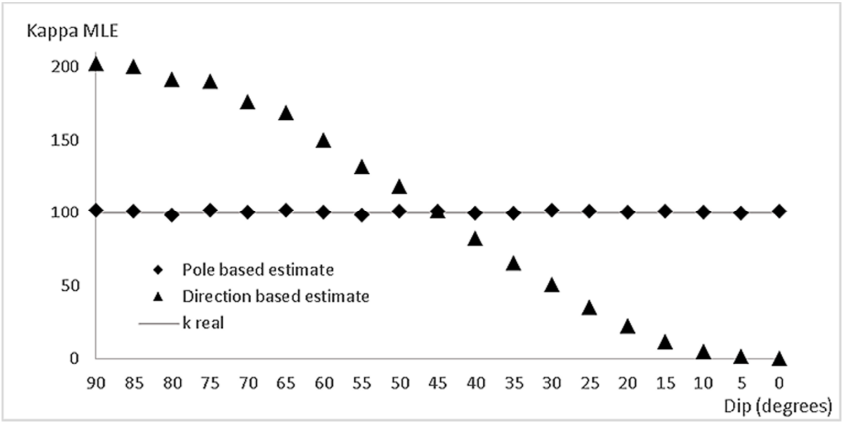

Figure 5: Distortion of the Maximum Likelihood Estimation (MLE) of concentration/spherical variance of a hundred spherical unit vector samples of a size of a thousand individuals drawn from a von Mises-Fisher distribution with κ = 100. Pole based estimate is always consistent with the data while dip based ones either over or underestimate it.

5

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

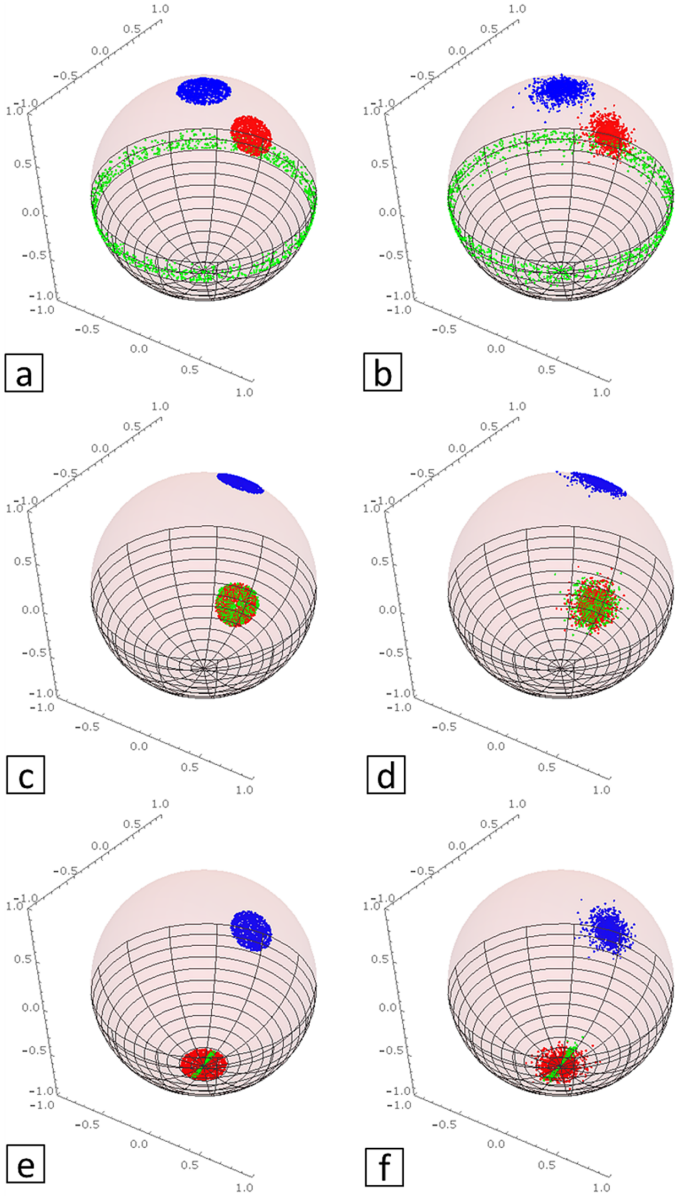

Figure 6: Effect of sampling over dip vectors or pole vectors on bounded uniform spherical distribution at range = 10° (a, c, e) and von Mises-Fisher distribution at κ = 100.0 (b, d, f) for uncertain horizontal planes (a, b), 45° dip planes (c, d) and vertical planes (e, f). See section 4.1 for details.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

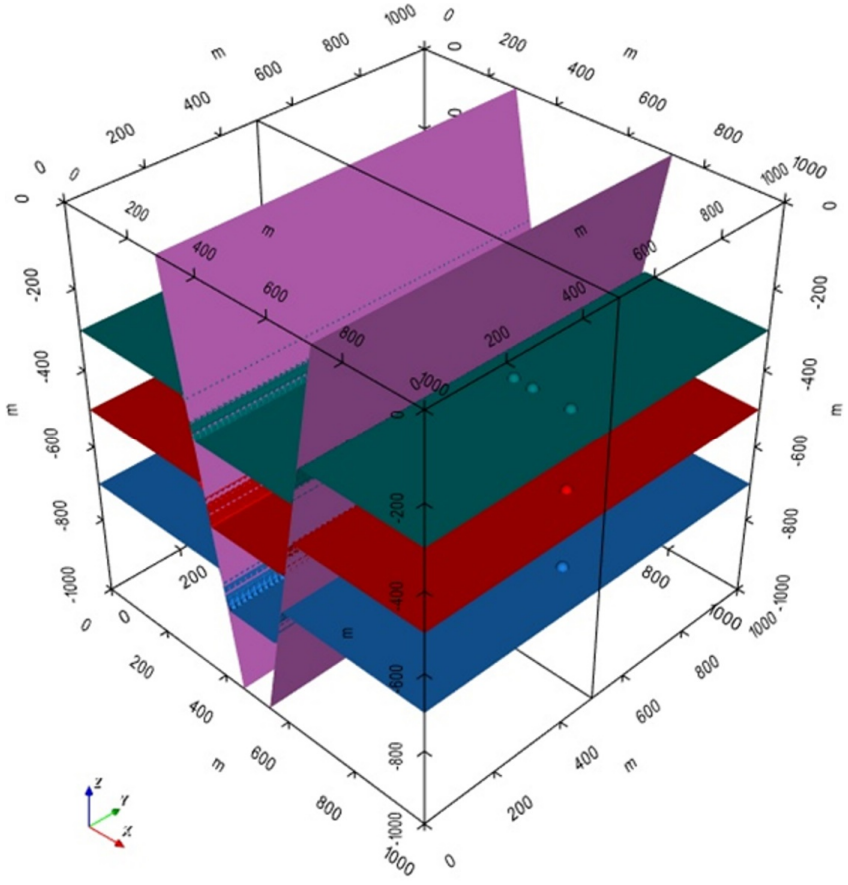

Figure 7: Structural data for the graben model and modelled surfaces for units and faults. Spheres represent interfaces and cones represent pole vectors.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

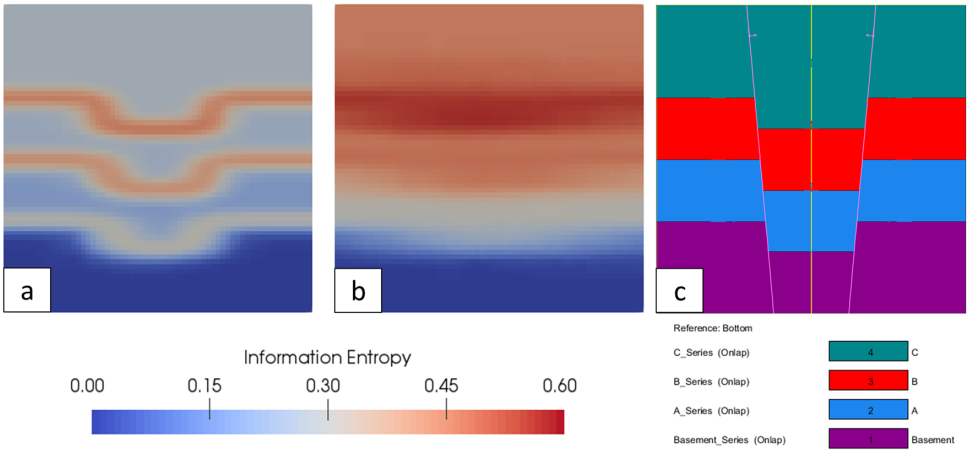

Figure 8: Effect of pole (a) versus dip (b) perturbation for a graben model (c), orientations are perturbed over a vMF distribution with kappa = 100.

5

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

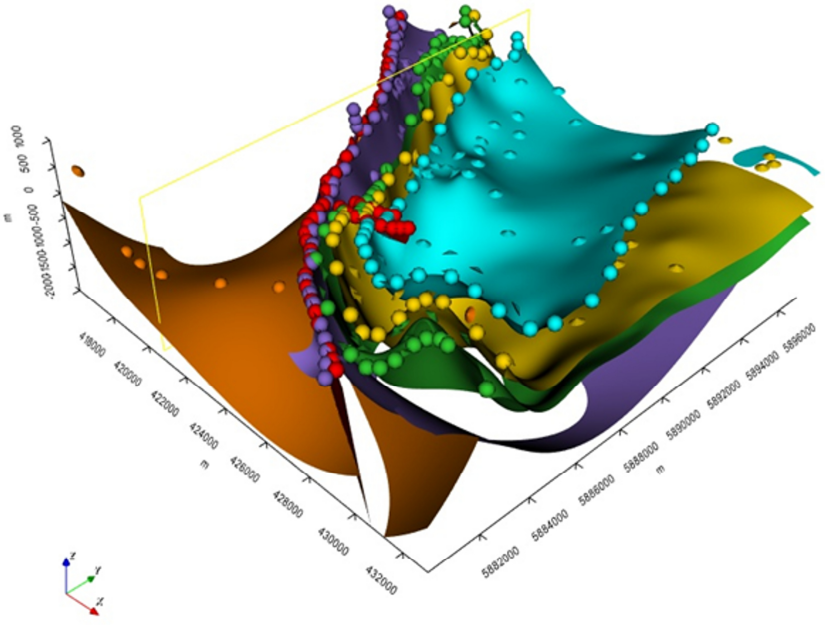

Figure 9: Structural data for the Mansfield model and modelled surfaces for units and faults. Spheres are interfaces and cones are orientations.

Solid Earth Discuss., https://doi.org/10.5194/se-2017-115
Manuscript under review for journal Solid Earth
Discussion started: 17 October 2017

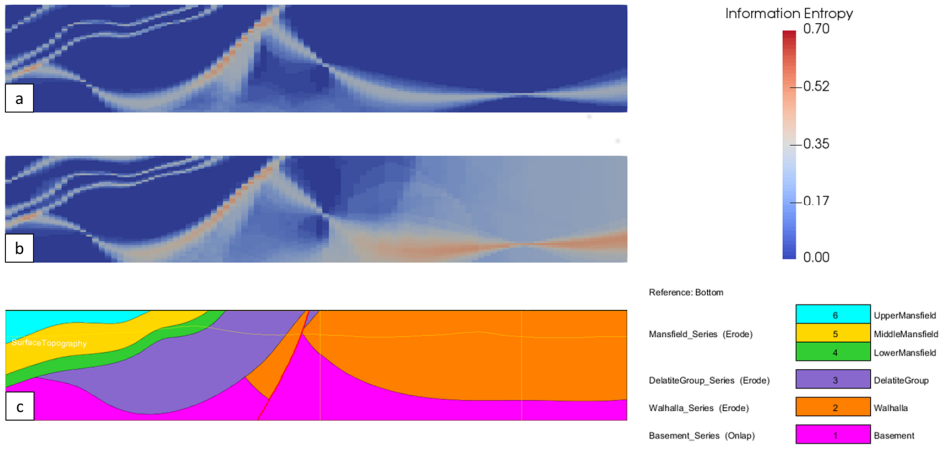

Figure 10: Effect of pole (a) versus dip (b) perturbation on a cross–section of the Mansfield model (c), orientations are perturbed over a von Mises-Fisher distribution with κ = 100.