Bayesian geological and geophysical data fusion for the construction and uncertainty quantification of 3D geological models

Hugo K. H. Olierook[1], Richard Scalzo[2], David Kohn[3], Rohitash Chandra[2,4], Ehsan Farahbakhsh[2,4], Chris Clark[1], Steven M. Reddy[1], R. Dietmar Müller[4]

[1]School of Earth and Planetary Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia
[2]Centre for Translational Data Science, University of Sydney, NSW 2006 Sydney, Australia
[3]Sydney Informatics Hub, University of Sydney, NSW 2006 Sydney, Australia
[4]EarthByte Group, School of Geosciences, University of Sydney, NSW 2006 Sydney, Australia

Correspondence to: Hugo K. H. Olierook (hugo.olierook@curtin.edu.au)

Abstract. Traditional approaches to develop 3D geological models employ a mix of quantitative and qualitative scientific techniques, which do not fully provide quantification of uncertainty in the constructed models and fail to optimally weight geological field observations against constraints from geophysical data. Here, we demonstrate a Bayesian methodology to fuse geological field observations with aeromagnetic and gravity data to build robust and detailed 3D models in a 13.5 × 13.5 km region of the Gascoyne Province, Western Australia. Our approach is validated by comparing 3D model results to independently-constrained geological maps and cross-sections produced by the Geological Survey of Western Australia. By fusing geological field data with magnetics and gravity surveys, we show that 89% of the modelled region has >95% certainty for a particular geological unit. The boundaries between geological units are characterized by narrow regions with <95% certainty, which are typically 400–1000 m wide at the Earth's surface and 500–2000 m wide at depth. Beyond ~4 km depth, the model requires geophysical survey data with longer wavelengths (e.g., active seismic) to constrain the deeper subsurface. Our results show that surface geological observations fused with geophysical survey data yield robust 3D geological models with narrow uncertainty regions at the surface and shallow subsurface, which will be especially valuable for mineral exploration and the development of 3D geological models under cover.

1 Introduction

Surface mapping and subsurface interpretation of geological units are essential requirements for mineral exploration, and the wider geological and geophysical community. Geological units and their associated boundaries may share similar geological histories or may be juxtaposed to one another via unconformities or structural discontinuities such as faults or suture zones. Accurately positioning geological units and their boundaries is fundamental for, and not limited to, constraining plate reconstructions (Cawood and Korsch, 2008; Merdith et al., 2017), defining stratigraphy (Gradstein et al., 2012), and successful mineral and petroleum exploration (Dentith and Mudge, 2014; Selley and Sonnenberg, 2014). In order to conceive a comprehensive model of both the surface and subsurface geology, a combination of geological mapping, geophysical

interpretation, sample analysis and prior knowledge are used. Although all of these ingredients are important, contemporary workflows to incorporate them all tend to develop at best only a handful of possibly biased solutions that either neglect, or incompletely account for, the uncertainty associated with geological or geophysical interpretation, as well as the knowledge of how far to extrapolate data derived from sample analysis. This shortcoming of traditional geological model-building is exacerbated in regions with thick sedimentary or regolith 'cover' because the uncertainty poses a significant impediment to understanding the nature of the subsurface. In a future where exploration under cover has been recognized as vitally important for the mineral exploration sector (McFadden et al., 2012), developing geological models with accounted uncertainty is pivotal.

To develop robust geological models, it is important to quantify the uncertainty on the position and configuration of geological units. Previous work addressing the uncertainty problem included employing fuzzy logic and information entropy approaches to build semi-quantitative 3D geological models (Abedi and Norouzi, 2012; Joly et al., 2012; Wellmann and Regenauer-Lieb, 2012). However, these approaches still require a significant degree of human decision making into how to fuse disparate geoscientific datasets. Other approaches characterize uncertainty by generating ensembles of 3D models through perturbations of a set of underlying descriptive geometric parameters (de la Varga et al., 2018; Giraud et al., 2019; Lindsay et al., 2013; Pakyuz-Charrier et al., 2018a; Pakyuz-Charrier et al., 2018b). However, these approaches still largely elide the question of how the joint distribution of such parameters is meant to be derived. A fully quantitative and informative 3D geological model will fuse available constraints in a probabilistically rigorous fashion. Bayesian inference provides a suitable framework for doing this by using Markov chain Monte Carlo (MCMC) sampling methods for estimation and uncertainty quantification of free parameters. Previous studies using Bayesian inference in the geosciences have primarily focused on (i) geophysical joint inversions (Bosch et al., 2006; Giraud et al., 2017; Shen et al., 2013), (ii) fluid flow through permeable reservoirs for groundwater, hydrocarbon or carbon dioxide storage applications (Oladyshkin et al., 2013; Refsgaard et al., 2012; Seifert et al., 2012; Ye et al., 2010), and (iii) geomorphologic and climate evolution (Chandra et al., 2018; Hapke and Plant, 2010; Pall et al., 2018), and (iv) fusion of structural geology data with geophysical datasets (Grose et al., 2018; Jessell et al., 2014; Jessell et al., 2010; Pakyuz-Charrier et al., 2018b; Wellmann et al., 2018). Despite a clear need for Bayesian fusion of solid Earth geological and geophysical datasets (Jessell et al., 2014; Jessell et al., 2018; Jessell et al., 2010), there is still relatively little work in developing Bayesian-inferred 3D geological models, particularly detailed models at the local and camp-scale in a mineral exploration context.

The Obsidian software package (McCalman et al., 2014; Reid et al., 2013) provides a platform for MCMC sampling of the posterior probability distribution for a parameterized model of the 3D structure and physical properties of geological formations. Here, the free parameters are estimated by MCMC with the sampling taking into account both the estimated prior probability of the existence of any particular formation, and the likelihood of that configuration producing all available geophysical survey data in the modelled region. Obsidian's major strength is an efficient distributed implementation of an adaptive parallel-tempered MCMC algorithm (Miasojedow et al., 2013) capable of sampling from distributions that may have

multiple local modes, such as those arising in inversion problems with limited prior information. Obsidian was originally designed for deep (~1–5 km) geothermal energy applications in sedimentary basins, which includes the ability to fuse airborne or surface geophysical surveys (e.g., aeromagnetics, gravity, magnetotellurics, temperature) with laterally-sparse geophysical drill hole data (e.g., geological unit depths, bore hole temperature, density) and drill hole geological units as prior points. Although previous iterations of the Obsidian software package could not fuse geological field observations made on the Earth's surface with geophysical survey data, relatively little amendment to the program is required to make this possible. The ability to incorporate surface geological data is vital for surface and near-surface applications (< 1 km). Such applications include the mineral exploration sector but also extend to any igneous or metamorphic ('hard-rock') terranes. Unlike sparse drill hole data, surface geological observations provide high-resolution lateral constraints on 2D surface geological models that, together with geophysical survey data, permit the development of robust 3D geological models.

In this contribution, we extend the Obsidian software package to enable coupling of (i) airborne magnetic and gravity survey data (using petrophysical priors derived from surface samples) with (ii) geological field observations that inform the configuration of lithostratigraphic units at discrete points on the Earth's surface. We demonstrate the validity of our techniques by building models of a 13.5 × 13.5 km subsection of the Gascoyne Province, Western Australia (Fig. 1), that is rich in data diversity and coverage, and comparing the model results to surface geological maps and interpreted cross-sections produced by the Geological Survey of Western Australia (GSWA; Fig. 2; Johnson et al., 2012). These maps were made primarily from geological mapping with input of geophysical only utilized in regions of cover, such as is present in the southern part of the area (Fig. 2a). There are a few datasets available in the region that are not utilized in our model. (1) There are only a few ~10 m-deep drill holes in the southwestern corner, so drill hole data is omitted as it does not add further detail than surface observations provide. With a lack of drill hole data, our contribution is able to address the impact of solely surficial geological data on the model accuracy. The lack of prior information is particularly useful for applications such as greenfields mineral exploration or tectonic analysis of hard rock terranes without drillholes would benefit from this understanding. (2) Our study excludes the use of structural geological data because other workers have recently focussed on this problem (e.g., Pakyuz-Charrier et al., 2018b; Wellmann et al., 2018) and because Obsidian cannot yet incorporate structural data. Again, in a greenfields exploration context, this is a realistic scenario. (3) The single 2D active seismic line immediately to the west of our model (Fig. 1) is not utilized in a Bayesian framework because the vast majority of hard rock terranes do not have seismic data coverage. Despite these shortcomings, (Wellmann et al., 2018)the model of the chosen study area is still particularly suitable for the future of mineral exploration as it exhibits a significant portion of recent sedimentary and regolith cover, which makes certain areas inaccessible for recording geological basement observations directly but possible to infer using Bayesian techniques applied to geophysical measurements.

2 Background

2.1 Geological setting of the Gascoyne Province

The Gascoyne Province, and the wider Capricorn Orogen, record the protracted amalgamation of the West Australian Craton and subsequent intracontinental tectonothermal activity (Fig. 1, Fig. 2). Two main events are thought to contribute to forming the West Australian Craton. First, the ca. 2195–2145 Ma Ophthalmia Orogeny sutured the Glenburgh Terrane, comprised of the Halfway Gneiss, to the Pilbara Craton (Krapež et al., 2017; Rasmussen, 2005). The deposition of the Moogie Metamorphics was associated with the Ophthalmia Orogeny, deposited into a foreland basin that formed in a response to the Glenburgh-Pilbara collision (Johnson et al., 2013). Second, the ca. 2005–1950 Ma Glenburgh Orogeny then amalgamated the combined Pilbara Craton-Glenburgh Terrane with the Yilgarn Craton to form the West Australian Craton (Johnson et al., 2013; Olierook et al., 2018). The Glenburgh Orogeny was associated with two major Andean-type granitoid formations, the Dalgaringa and Bertibubba Supersuites, and several subduction-related basins (Johnson et al., 2011; Olierook et al., 2018). After unification, the Capricorn Orogen experienced at least five intracontinental tectonomagmatic events, each decreasing in severity of tectonic character and magmatism (Johnson et al., 2017). The first two events, the 1830–1780 Ma Capricorn Orogeny and 1680–1620 Ma Mangaroon Orogeny, were both associated with significant granitoid magmatism of the Moorarie and Durlacher Supersuites, respectively (Sheppard et al., 2010a; Sheppard et al., 2005). Deposition of the Leake Springs Metamorphics was also concurrent with the early stages of the Capricorn Orogeny. Later events were predominantly amagmatic but were still associated with up to amphibolite-facies metamorphism and hydrothermal activity (Korhonen et al., 2017; Sheppard et al., 2007). Both suturing and intracontinental tectonic events have developed a pervasive east–west striking structural fabric in the Gascoyne Province that has compartmentalized the region into several geological zones that share tectonic characteristics (Sheppard et al., 2010a). In the south, zone and formation boundaries trend NE–SW whereas major structures are oriented NW–SE in the north, yielding a wedge-shaped geometry for the Gascoyne Province (Fig. 1, Fig. 2). Compared to the rest of the Capricorn Orogen, the Gascoyne Province is relatively well exposed but there are still significant areas covered by recent regolith and sediment that hamper mineral exploration.

2.2 Bayesian inversion and inference

Inverse problems aim to recover the causal factors that produced a set of observations (Mosegaard and Tarantola, 1995; Sambridge and Mosegaard, 2002). For geological and geophysical applications, the objective of inverse problems is to recover the subsurface properties such as density and magnetic susceptibility from surface-based geophysical survey measurements of gravitational acceleration and magnetic field strength. However, geophysical surveys cannot yield unique solutions of the subsurface petrophysical properties. Thus, there are an infinite number of subsurface petrophysical measurement configurations that would produce the same survey readings (Sambridge, 1999; Sambridge and Mosegaard, 2002). Given that there is no unique solution, there is no reason to prefer one model over another without introducing constraints on what form the model should take (Parker, 1977; Sambridge and Mosegaard, 2002). Ways to introduce such

constraints include 3D geometry inversion (Fullagar et al., 2008; Guillen et al., 2008), level-set inversions (Bijani et al., 2017; Zheglova et al., 2018) and (cross-)gradient regularization (Giraud et al., 2019; Scholl et al., 2016). However, these techniques are deterministic, yielding a single geological-geophysical inverse model that represents only one scenario. (Bijani et al., 2017; Zheglova et al., 2018)

An alternative approach is to use sampling over all possible models in a probabilistic Bayesian context, which provides a more systematic approach towards uncertainty quantification and the incorporation of prior constraints. The Bayesian paradigm departs from traditional deterministic inversions in two main respects. First, it encapsulates prior knowledge about the geological model (before the data are taken into account) in terms of a probability distribution $P(\theta)$ over model parameters θ, called the prior. The probability $P(D|\theta)$ of the data D conditioned on a set of model parameters, called the likelihood, takes the place of the usual misfit function in a traditional inversion. Given these two elements, the posterior distribution $P(\theta|D)$ of the model parameters given the data is given by Bayes' rule:

$$P(\theta|D) \propto P(D|\theta)\, P(\theta). \qquad \text{- Eq. 1}$$

Second, in Bayesian reasoning the full posterior distribution is the quantity of interest, rather than a point estimate (as in maximum likelihood or maximum a posteriori) or, in some cases, a point estimate with a parameter covariance matrix describing the shape of a single posterior mode.

MCMC methods represent the posterior distribution by drawing a finite number of representative samples of parameter sets. Although computationally expensive (Sen and Stoffa, 1996; Tarantola and Valette, 1982), they have the advantages of not needing to calculate the unknown constant of proportionality in Eq. 1, and of representing the exact posterior in the limit of large numbers of samples, rather than some parametric approximation to the posterior. To do this, samples (proposals) are drawn from a target distribution by constructing a Markov chain for which the desired posterior is the invariant distribution (Hastings, 1970; Kass et al., 1998; Metropolis et al., 1953; Raftery and Lewis, 1996; van Ravenzwaaij et al., 2018). Convergence criteria determine when to stop sampling that, for example, could be a predetermined number of samples or an assessment of the behaviour of the likelihood function. However, for complex and large-scale 3D inversion problems, convergence can be challenging due to the large number of free parameters that need to be sampled effectively in limited computation time (i.e., high dimensionality; Sen and Stoffa, 1996).

For multimodal posteriors, the Markov chain can become trapped in a single mode and cannot fully explore the posterior distribution, making sampling much less efficient or causing the chain to converge to the wrong distribution. Parallel tempering (PT) is a sophisticated MCMC method (PT-MCMC, Fig. 3) that aims to increase the efficiency of the exploration of multimodal posterior distributions (Geyer, 1993; Hukushima and Nemoto, 1996; Sambridge, 2013). Parallel tempering uses a number of

replicas of the original sampling method, where the replicas are created at different 'temperatures' (Brooks et al., 2011; Earl and Deem, 2005) by rescaling the likelihood probability density function (Sambridge, 2013). High temperature replicas sample a smoother (flatter) version of the likelihood function in order to 'escape' from local minima and provide global exploration features (Fig. 3). Conversely, low temperature replicas provide local exploration capabilities. Hence, with parallel tempering, there is a delicate balance between global and local exploration (Earl and Deem, 2005; Sambridge, 2013). During sampling, the replicas are able to exchange their configurations, typically between neighbouring replicas via the Metropolis–Hastings proposal (Sambridge, 2013). Ultimately, this improves the mixing of Markov chain (i.e., thorough exploration of space and convergence to the target distribution) and efficiency of convergence (Sambridge, 2013).

2.3 Obsidian software package for joint geophysical inversion

The Obsidian software package was originally designed for geothermal exploration in sedimentary basins. Here, we provide a brief overview of the salient features that are important for our inverse problem. For a detailed background of Obsidian, the reader is referred to Ramos et al. (2012), Reid et al. (2013), McCalman et al. (2014), Beardsmore et al. (2016) and Scalzo et al. (in review). For completeness, we briefly describe the critical elements of the inversion framework here.

Obsidian's 3-D model consists of N layers with petrophysical properties (in this case, mass density ρ_i and magnetic susceptibility $\chi_{M,i}$) that are spatially constant within each layer. The depth to the top of layer i at the jth control point (x_j,y_j) on a regularly spaced rectangular grid across the modeled area is labeled ζ_{ij}. Let θ denote the totality of these parameters.

In each iteration of the Obsidian workflow, the petrophysical properties for this parametrized 3-D model are discretized onto a grid of voxels. Each layer depth $z_i(x,y)$ across the modeled area is interpolated from the ζ_{ij}, using a two-dimensional Gaussian process regression with a square exponential covariance kernel, where the covariance length in each direction matches the lateral spacing between control points. The constraint $z_i(x,y) \leq z_{i+1}(x,y)$ for each layer i is enforced on the discretized grid, setting the thickness of layer i at location (x,y) to $D_z(x,y) = \max(0, z_{i+1}(x,y) - z_i(x,y))$. This allows layers with no coverage in a particular region to "pinch out" to zero thickness; each layer interface is differentiable, except possibly at the lateral boundary of a pinched-out region. The rock properties beyond the horizontal boundaries are padded with constant values continuous with the properties evaluated on the boundary. The expected geophysical sensor readings are then forward-modeled, using rectangular prism approximations for the gravitational field (Li and Oldenburg, 1998) and total magnetic anomaly (Li and Oldenburg, 1998, 1996). Lithostratigraphic observations are forward-modeled by sampling the discretized model at the surface. Obsidian is also capable of forward-modeling other sensors, including 1-D magnetotellurics, temperature, and reflection seismology, although seismic measurements may be more efficiently incorporated as priors on the depth to each layer.

The model prior and likelihood are then calculated using the given parameter values and the forward-modeled data. The prior distribution $P(\rho_i,\chi_{M,i})$ for mass density and magnetic susceptibility is multivariate Gaussian, informed by petrophysical measurements from the literature. The priors $P(\zeta_{ij})$ on the control points for the depth to layer i are also Gaussian, with user-defined mean and variance if drill cores and/or seismic interpretation are available. The likelihood models for the various datasets are described in more detail in section 3.4 below.

3 Materials and Methods

3.1 3D geological model parameterization

The construction of the parameterized 3D geological model in the chosen 13.5 × 13.5 km area in the Gascoyne Province involves three types of data: a hierarchical construction of layers using available 2D seismic data interpretation (Fig. 1b; Johnson et al., 2013) and two types of point-based measurements — magnetic and density data — on hand samples from Aitken et al. (2014). A two-dimensional seismic survey was conducted in 2011 (Johnson et al., 2013). The surface position of the seismic line is immediately to the west of our study area and, thus, probably cross-cuts the same geological units that are present in our area (Fig. 1a). The seismic interpretation of Johnson et al. (2013), aided by geochronological data, directly informs that the stratigraphy in the 3D geological model, from oldest to youngest, are: (i) the ca. 2550–2430 Ma Halfway Gneiss (Johnson et al., 2017), (ii) ca. 2210–2150 Ma Moogie Metamorphics (Martin and Morris, 2010), (iii) ca. 1840–1810 Ma Leake Spring Metamorphics and ca. 1830–1780 Ma Moorarie Supersuite (Johnson et al., 2011; Sheppard et al., 2010b), (iv) ca. 1690–1660 Ma Durlacher Supersuite (Piechocka et al., 2017; Sheppard et al., 2005), and (v) ca. 995–900 Ma Thirty Three Supersuite (Piechocka et al., 2017; Sheppard et al., 2010b). In the chosen 13.5 × 13.5 km area, only the Halfway Gneiss and Durlacher Supersuite are areally significant, comprising ~59% and ~35% of the interpreted area by GSWA (Fig. 2). Other areally-minor units in the 13.5 × 13.5 km area include the Leake Spring Metamorphics (~3%), Moogie Metamorphics (~2%), Moorarie Supersuite (<1%) and Thirty Three Supersuite (<1%). Only the two volumetrically-major units are modelled in this study as the other units appear primarily near the surface (Johnson et al., 2011), and are also present only in areas smaller than our model can resolve (see next paragraph). Resolving finer-scale features is out of the scope of this contribution.

The square area is tiled with a 5 × 5 grid of control points defining the boundary between the Halfway Gneiss and Durlacher Supersuite. Including mass density and magnetic susceptibility defined for both rock layers, the model has 29 parameters in total. A Gaussian prior with mean zero and standard deviation of 5 km is placed over the depth to the interface, reflecting only vague prior knowledge about the location of the boundary, and allowing either formation to be accessible at the surface. This parametrized model is discretized with resolution of 500 m during the forward model calculations for the potential fields through the prism approximation; the parametrization can be directly queried at the surface for lithostratigraphic measurements so that no spatial discretization is needed.

3.2 Petrophysical data

Two types of petrophysical data, magnetic susceptibility and density, were collected on hand samples (Aitken et al., 2014) to link to aeromagnetic and gravity data, respectively (Fig. 4). A relatively low number of magnetic (n = 104) and density (n = 103) samples across the entire Gascoyne Province were available (Aitken et al., 2014). Thus, the sample mean and covariance of magnetic and density measurements for each geological unit were used to inform a multivariate Gaussian prior, acknowledging that spatial differences in magnetic and density distributions cannot be captured in this contribution.

3.3 Geophysical and geological data

Two geophysical survey data are employed, namely aeromagnetic (Fig. 5a) and gravity data (Fig. 5b), that are forward-modelled to correspond to sample-based magnetic susceptibility and density data, respectively. These types of geophysical surveys were already available for incorporation in the Obsidian framework. In addition to the geophysical data, field-based geological units observations are incorporated into Obsidian (Fig. 5c).

Aeromagnetic data in the study area (Fig. 5a) utilized a subsection of the 1995/96 Bangemall Survey directed by GSWA from latitudes 23.5–26.0° S and longitudes 115.0–120.0° E (Geological Survey of Western Australia, 1996). The Bangemall aeromagnetic data were flown at a 7.5 m sample interval, 500 m flight line spacing and a mean terrain clearance of 60 m. The final magnetic intensity in nT has the following corrections from the raw data: (i) the 1990 IGRF model removed and a base value of 54940 nT added, (ii) diurnal correction applied, with a base value of 55220 nT, (iii) parallax correction of 0.4 fiducial applied, (iv) levelled using tie line information, and (v) tie lines force levelled to flight lines. The original horizontal datum of the Bangemall Survey was the AGD84, projected using AMG zone 50, but this was converted to the WGS84, zone 50 S for each data point. The flight path vector data were explicitly favoured over the post-processed raster data to avoid introducing correlations. Very finely, regularly-sampled raster data are often preferred in order to apply fast Fourier transform inversion techniques. However, resampling or interpolating non-gridded data onto such a grid results in correlations between the gridded data points, which can lead to biases and incorrect results in probabilistic inversions if not explicitly accounted for (Scalzo et al., in review).

Gravity anomaly data in the study area (Fig. 5b) were derived from the 2010 Gascoyne North and Gascoyne South surveys directed by GSWA from latitudes 23.5–26.0° S and longitudes 115.2–118.5 ° E (Mathews and Jecks, 2010). The Gascoyne North and South gravity data were acquired at a ground-based nominal station spacing of 2500 m in a square grid configuration. The final complete spherical cap Bouguer anomaly in μms^{-2} had the following corrections from the raw data: (i) correction of remanent drift, typically less than 0.05 μms^{-2} hr^{-1}, (ii) computation of Bouguer anomaly using a modified spreadsheet

developed by M. Bacchin of Geoscience Australia, (iii) spherical cap Bouguer anomaly computation relative to the Australian Absolute Gravity Datum 2007, and (iv) terrain correction using the AUSGEOID09 vertical coordinate reference frame. The original horizontal datum was GDA 94, which is equivalent to WGS84. Similar to the magnetic data, the located surface point-based data were explicitly favoured over the post-processed raster data.

Lithostratigraphic unit observations made at the surface were obtained from the Western Australian Rocks (WAROX) database, available from GSWA (Fig. 2, Fig. 5c). Each spatially-referenced sample point records an observation of rock type coupled to an interpreted geological unit. Designation of a particular geological unit is informed from petrographic, geochemical and geochronological knowledge obtained on a subset of WAROX data. For example, the assignment of a geological unit is near certain where U–Pb crystallization ages are available or where whole-rock major and trace element geochemistry has been collected. In the chosen 13.5 × 13.5 km study area, only one sample has geochronological and geochemical information (Fig. 5c). However, there are more than 100 age and over 500 samples with geochronological and geochemical data, respectively, in the Gascoyne Province from the same geological units that are present in the 13.5 × 13.5 study area (e.g., Johnson et al., 2017). All samples with U–Pb ages and/or geochemical data in the Gascoyne Province also have petrographic and/or hand sample descriptions, which can be used to inform the geological unit for geological surface observations where only hand sample or petrographic descriptions are available. Even though inference of geological units from similar petrographic and hand sample descriptions is relatively robust, it may be in error. Thus, we have accounted for this potential uncertainty in geological field observations (see section 3.4 for further details). All our observations were taken at the surface but could be readily used where geological unit observations could be made in the subsurface (i.e., via drill hole information). The formation that is observed at the surface defines the value of a given field observation.

3.4 Likelihood models

Given the forward-modeled predictions $f_s(g(\theta))$ for the observations D_s of a geophysical sensor s based on a discretized 3-D model $g(\theta)$, the likelihood is:

$$\mathrm{P}(D_s \,|\theta) \;=\; t_{2\alpha s}\,([f_s(g(\theta)) \;-\; D_s)] \;\times\; \sqrt{\frac{\alpha s}{\beta s}}) \qquad \text{- eq. 2}$$

(McCalman et al., 2014; Scalzo et al., in review), where $t_\nu(x)$ denotes the Student's-*t* distribution with ν degrees of freedom. The *t* distribution has fairly thick tails compared to the Gaussian distribution. It arises in this case by assuming that each sensor has Gaussian noise with unknown variance, and by averaging the underlying Gaussian likelihood over all possible values of this noise variance for each sensor (see Appendix A). The hyperparameters α_s and β_s describe the shape of an inverse-gamma distribution IG(x; α_s, β_s) that forms the prior on the noise variance for each sensor; β_s provides an overall

variance scale, while α_s controls the tails of the distribution, with smaller α_s corresponding to a heavier tail towards large possible variance. Integrating out the noise variance in this way helps to speed up sampling. We choose $\alpha_s = 1$, $\beta_s = 0.2$ for each sensor, producing a 95% credible interval for the noise variance of 0.05 < sigma < 3 (as a fraction of the sample variance of the data) with 95% credibility.

In a similar manner, we formulate a likelihood for categorically-distributed lithostratigraphic observations. Supposing that these observations have some unknown probability p of being correct (that is, that the true underlying rock formation corresponds to the geologist's interpretation) in the absence of corroborating geochronological or geochemical information, we consider an underlying binomial likelihood,

$$P(k|n,p) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} p^k (1-p)^{n-k}, \qquad \text{- eq. 3}$$

where k is the number of successes, n is the number of trials, p is the probability of success for each trial and Γ is the gamma function. The left fraction is the binomial coefficient and is more commonly written as $\binom{n}{k}$ but the gamma function representation generalizes to non-integers. In practice p may not be precisely known, but can be constrained by a prior distribution. To make averaging over the unknown value of p analytically tractable, we use a beta distribution B(p; α_s, β_s) as a prior over p, where in this case α_s corresponds to a prior weight towards higher p, and β_s to a prior weight towards lower p. Elicitation of a suitable prior from geologists with field experience suggests $\alpha_s = 20$, $\beta_s = 1$, resulting in a 95% credible lower limit of 85% on p. Integrating over the unknown value of p then leads to a beta-binomial likelihood:

$$P(k|n,\alpha,\beta) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \int_0^1 p^k (1-p)^{n-k} \left[\frac{p^\alpha (1-p)^\beta}{B(\alpha,\beta)}\right] dp \qquad \text{- eq. 4.1}$$

$$P(k|n,\alpha,\beta) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \frac{\Gamma(k+\alpha)\Gamma(n-k+\beta)}{\Gamma(n+\alpha+\beta)} + \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \qquad \text{- eq. 4.2}$$

(see appendix A for full derivation of formula). This captures overdispersion resulting from unresolved sources of variation in the observation errors, such as distinct geologists interpreting rock formations differently (Gelman et al., 2013).

3.5 Markov chain Monte Carlo sampling

We use PT-MCMC to explore the parameter space, which is defined by our 3D geological model parameterization. PT-MCMC operates by running a number of Markov chains in parallel, each sampling a modified version of the posterior:

$$P_i(\theta|D) \propto \mathrm{P}(D|\theta)^{\beta_i} P(\theta), \qquad \text{- eq. 5}$$

where the inverse temperatures { β_i } form a decreasing sequence $1 = \beta_0 > \beta_1 > \beta_2 > \ldots > \beta_N > 0$. Thus only the chain with $\beta_0 = 1$ is sampling from the true posterior distribution, whereas the other chains explore versions of the problem with reduced influence from the data (a chain sampling in the limit $\beta_i \rightarrow 0$ is sampling from the prior). Two chains are allowed to swap their most recent states θ and θ' with probability:

$$P_{swap} = min(1, [\frac{P(D|\theta\prime)}{P(D|\theta)}]^{\beta\prime-\beta} \times \frac{P(\theta\prime)}{P(\theta)}). \qquad \text{- eq. 6}$$

This method of sharing information among chains enables chain 0 can explore parts of parameter space visited by the other chains while still guaranteeing convergence to the true posterior distribution. At cessation of sampling only the samples from chain 0 are used.

Obsidian uses the adaptive version of the PT-MCMC algorithm described by Miasojedow et al. (2013), which continuously adapts both the scale of proposals within a chain and the temperature differences between chains in order to satisfy a target acceptance rate for swaps (we use 0.25; Atchad'e et al., 2011). The maximum allowed scale for adjustments is decreased to zero over time, ensuring that the adaptive algorithm will still produce samples from the desired posterior distribution. Obsidian's implementation is distributed, allowing a large number of cores to be used efficiently for sampling (Beardsmore et al., 2016; McCalman et al., 2014).

The Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) is used to sample parameters associated with each chain's target distribution. This algorithm makes a proposal θ' chosen from a proposal distribution $q(\theta'|\theta)$, which is then accepted with probability

$$P_{\text{accept}} = min(1, \frac{P(D|\theta\prime)}{P(D|\theta)} \times \frac{P(\theta\prime)}{P(\theta)} \times \frac{q(\theta|\theta\prime)}{q(\theta\prime|\theta)}). \qquad \text{- eq. 7}$$

If the proposal is accepted, it is added to the chain; if it is rejected, a copy of the previous state is added to the chain. We use the preconditioned Crank–Nicolson MCMC proposal (Cotter et al., 2013; Hu et al., 2017; Rudolf and Sprungk, 2018):

$$\theta' = \sqrt{(1-\eta^2)}\,\theta_n + \eta u, \qquad \text{- eq. 8}$$

where θ_n is the state of the chain at iteration n, $\eta \in (0,1)$ is a step-size parameter, and $u \sim P(\theta)$ is a random vector drawn from the (multivariate Gaussian) prior. When $\eta \ll 1$, the proposal reduces to a random walk, while for $\eta \sim 1$ the proposal approaches a draw from the prior. This proposal was originally developed for use in inversions on function spaces, but it may also speed convergence as a within-chain proposal in high-dimensional parallel tempering problems (see Scalzo et al., in review).

3.6 Experiment Design

We run the Obsidian PT-MCMC with six parallel tempering temperature ladders (stacks), where each stack consists of twelve PT-MCMC chains. The likelihood of each chain is raised to the power of a different temperature in the ladder and the lowest temperature chain is the unnormalized likelihood. Samples are only collected for the lowest temperature chain. This setup has enough chains in each ladder to ensure geometric spacing between temperatures on the ladder, confirmed by empirical examination of the ladder, and has enough stacks to ensure sufficient confidence in the convergence diagnostics, specifically the potential scale reduction factor diagnostic (see section 4.1). The experiment was run on an area of 13.5 x 13.5 km for 96 hours to ensure convergence for all rock property and control point parameters.

4 Results

4.1 Convergence diagnostics

Convergence diagnostics aid in evaluating whether the MCMC sampling has converged (i.e., whether sampling is occurring from the target distribution; Gelman et al., 2013). Convergence of the control point parameters occurred after only 12 hours but the rock property parameters required approximately half of the total 96-hour run time to reach convergence. Several techniques are listed here to confirm that our model outputs are statistically-valid, including (i) trace plots of the MCMC samples (Fig. 6a, b, Table 1), (ii) autocorrelation times and effective sample size (Fig. 6c, Table 1), (iii) potential scale reduction factor (Table 1), and (iv) Geweke score (Geweke, 1992) (Fig. 7).

Trace plots for modelled density and magnetic susceptibility show that: (a) chains initialized at different initial states have similar posterior densities, and (b) chains mix well, i.e., they sufficiently explore the support of the posterior distributions as determined by the parameters' respective priors (first column in each panel of Fig. 6). The Halfway Gneiss and Durlacher Supersuite have modelled densities of 2.72 ± 0.12 and 2.67 ± 0.12 g cm^{-3} (uncertainties quoted at two standard deviations [2σ] here and throughout), respectively, and average modelled $\log_{10}$ magnetic susceptibilities of –3.65 ± 0.57 and –2.60 ± 0.07 (2σ), respectively (second column in each panel of Fig. 6). Compared to the Halfway Gneiss, a lower magnetic mean susceptibility and larger variance in magnetic susceptibility for the Durlacher Supersuite agrees with the prior density and magnetic measurements for both formations (Fig. 4). The lack of difference between the modelled densities of the Halfway Gneiss and Durlacher Supersuite (at 2σ) are also in agreement with the priors (Fig. 4).

The integrated autocorrelation time (IACT) is an estimate of the number of successive MCMC samples between statistically independent samples from the posterior distribution. It is given by:

$$\tau = \mathrm{sum}\{-\infty, \infty\} P(\tau), \qquad \text{- eq. 9}$$

where P(τ) is the normalized discrete autocorrelation function of the parameter values treated as a time series; a white noise time series, in which every sample is statistically independent of all the others, will have τ = 1. Since any MCMC algorithm works by making proposals that adjust a chain's current location, the posterior samples they generate will necessarily be correlated in time (measured by sample index); this inefficiency is balanced by the ability of MCMC to preferentially sample regions of the parameter space with high posterior probability. Lower values of the IACT are desirable, since the statistical power of a MCMC chain of given finite length will vary inversely with the IACT. In practice, complex, high-dimensional posterior distributions can result in IACTs in the hundreds of thousands of samples (Ruggeri et al., 2015).

There are approximately 1.5 million total samples for each chain. However, samples from MCMC are correlated (third column in each panel of Fig. 6), which reduces the number of independent samples (i.e., the effective sample size) from the posterior distribution (Gelman et al., 2013). The MCMC autocorrelation is on the order of 1 in 12,000 to 1 in 14,000 independent samples per MCMC proposal for rock property parameters and 1 in 1,000 to 1 in 15,000 independent samples for control point parameters (Table 1). This means that there are approximately 105–129 and 100–1,000 independent samples for the rock property and control point parameters, respectively.

The potential scale reduction factor (PSRF), also known as the Gelman-Rubin statistic or $\hat{R}$ (Gelman and Rubin, 1992), assesses convergence by comparing the variance between means of multiple chains relative to the average of the variance within chains to show how much an estimator of the marginal posterior variance will decrease as the number of samples increases (Brooks and Gelman, 1998; Cowles and Carlin, 1996). Suppose we have m chains with n samples each of a given scalar parameter (i.e. projection of the parameter set onto one of the variables of interest), and let μ_i and σ^2_i denote the mean and variance of samples from each chain with 1 <= i <= m. The PSRF is then given by

$$\hat{R} = \left(n - \frac{1}{n}\right) + \left(\frac{1}{n} + \frac{1}{mn}\right)\frac{B}{W}, \qquad \text{- eq. 10}$$

where B/n = var(μ_i) is the sample variance of the m within-chain means, and W = <σ^2_i> is the sample mean of the m within-chain variances.If the diagnostic is close to 1 then limited reductions in variance can be made from further sampling and the sampling has likely converged to the target distribution. All of our rock property and control point parameters have $\hat{R}$ = 1.02–1.03 and $\hat{R}$ = 1.00–1.06, respectively (Table 1), which indicate convergence on the basis that the Gelman-Rubin statistic is less than the threshold of 1.10 (Gelman et al., 2013).

The convergence of the target distribution may also be evaluated using the Geweke score (Geweke, 1992), which is a z-score diagnostic that compares the mean of subsets of samples from the start and end of the MCMC chains. Using the samples from a single parameter sampled by a single chain, we calculate

$$z = \sqrt{(\mu_{\text{initial}} - \mu_{\text{final}})/(\sigma^2{}_{\text{initial}} + \sigma^2{}_{\text{final}})} \qquad \text{- eq. 11}$$

where the initial $n_{initial}$ samples from the chain have mean $\mu_{initial}$ and variance $\sigma^2{}_{initial}$, and the final μ_{final} samples from the chain have mean μ_{final} and variance $\sigma^2{}_{final}$. We use $n_{initial} = n_{final} = 0.1n$ where n is the total chain length. A heuristic for convergence are Geweke scores between –2 and +2, indicating normality of the difference in means (Cowles and Carlin, 1996). We see convergence of our chains in terms of this diagnostic with Geweke scores of and –1 to +1 and –0.5 to 1.5 for the density and magnetic susceptibility parameters, respectively, except for two chains (chain 2, density for both formations) which show large deviations (Fig. 7).

4.2 Residuals from forward models

Aeromagnetic and gravity models from forward models are broadly comparable to their measured counterparts (Fig. 8a, d).

Aeromagnetic models effectively identify the NW–SE strike of magnetic lineaments in the northern half of the modelled volume (Fig. 8) that would be predicted from geological maps (Fig. 2). The NE-trending elongate unit in the southern half of the map, corresponding to the Durlacher Supersuite (Fig. 2), also shows limited discrepancies between modelled and measured data (Fig. 8). Aeromagnetic residuals display an approximately Gaussian distribution with a mean of 0 nT (i.e., equivalent to measured aeromagnetic data) and a range of $^{+358}_{-317}$ nT (2σ), which covers approximately 21% of the total magnetic range (Fig. 8). The ~21% residual standard deviation is comparable to the standard deviation range of magnetic susceptibility values (Fig. 4). Only one region in the northwestern portion of the map has significantly higher magnetic field strength than modelled (Fig. 8).

Modelled gravity effectively identifies the long-wavelength, N–S trending structure in the measured data but there are significant positive residuals in the south-west corner and negative residuals in the central-eastern portion of the modelled area (Fig. 2, Fig. 8). Gravity residuals are positively skewed with a mean residual of 0 and a standard deviation of $^{+3.96}_{-2.36}$ mGal (2σ, 20% of the total gravity range; Fig. 8). The gravity residual range of ~20% is comparable to the standard deviation of density data (Fig. 4).

The forward modelled field observations have a posterior probability of success of approximately 80% (Fig. 9a–c). Five of 141 actual Halfway Gneiss observations have been misclassified as belonging to the Durlacher Supersuite, while most of the actual Durlacher Supersuite observations have been misclassified as belonging to Halfway Gneiss (Fig. 9d–f). All the misclassifications occur within 1 km of boundaries between geological units, particularly in the southeast and eastern parts of the 13.5 × 13.5 km area (Fig. 9e).

4.3 Probability density of layer locations

Voxelized posterior distributions of the modelled volume reveal a strong probability contrast between regions of high certainty (defined as >95%) at the surface (Fig. 10). The modelled volume shows that the Durlacher Supersuite occupies the northeastern section of the region and an ellipsoidal inlier towards the southern extent of the 13.5 × 13.5 km map (Fig. 10b). The remainder of the map shows Halfway Gneiss. Approximately along the Chalba Shear Zone (Fig. 10a), ~300–1000 m widths of <95% certainty separate regions of >95% certainty (Fig. 10b). The uncertainty on the boundary between the ellipsoidal Durlacher Supersuite inlier and the Halfway Gneiss is constrained differently in different parts of the model. The horizontal distance for the boundary between the Durlacher Supersuite and Halfway Gneiss (<95%) is relatively tightly constrained along the NW and SW margins (~450 m), moderately constrained along the SE margin (~750–1100 m) and poorly constrained towards the east (up to 2350 m; Fig. 10).

At depth, sub-vertical unit boundaries are maintained as informed by the prior (Figs. 1, 2; seismic interpretation of Johnson et al., 2013). The modelled cross-section yields dips of 85° near the surface, and progressively reducing in inclination to ~72° at 4 km depth. In the cross-section, the horizontal distance between regions of >95% certainty between the Halfway Gneiss or Durlacher Supersuite becomes progressively more diffuse, from ~420 m to 1060 m between the surface and 4 km depth, respectively (Fig. 11). This translates to a percentage decrease in horizontal confidence of ~250%. In other parts of the 3D model, regions of <95% certainty can be as wide as ~2500 m at depth (Fig. 11b).

5 Discussion

5.1 Validity of 3D models and comparison to geological maps and cross-sections.

The fusion of geological field observations with gravity and magnetic data are valid on a statistical basis, including showing: (i) modelled petrophysical properties comparable with the prior (Figs. 4, 6), (ii) adequate yield of independent samples (Fig. 6), (iii) sufficient exploration of the parameter space ($\hat{R} < 1.1$, Table 1), and (iv) convergence as indicated by Geweke scores between –1 and +1 (Fig. 7). Aeromagnetic, gravity and field observations show mean residuals of ~0 with 2σ tails that are a maximum of ~20% of the total dataset range (Fig. 8c,f). Aeromagnetic residuals are spatially uncorrelated except for a small

region in the NW corner of the 13.5 × 13.5 area (Fig. 8b), indicating that the model captures relevant variation on the length scales of interest. Gravity residuals are systematically positive in the south, west and north, and negative in the east (Fig. 8e). This is primarily a function of a long-wavelength gravity response that is probably attained from the deep subsurface (Johnson et al., 2013), whereas the model aims to capture the shallow subsurface model (<5 km). Field observation misclassifications are only found within 1 km of geological boundaries. The discrepancies between field observations and modelled geological units may have resulted from the presence of other geological units (particularly those that are highly magnetic). Additionally, more petrophysical data for the Halfway Gneiss and Durlacher Supersuite inside the modelled area may have yielded better priors for the geophysical surveys, which in turn would have corroborated better with the position of geological field observations. Ultimately, the data residuals are sufficiently small to have yielded a reliable model output despite minor discrepancies.

The voxelized posterior distributions of the modelled volume are visually comparable to geological maps (Johnson et al., 2012)and interpreted cross-sections (Johnson et al., 2012) made by GSWA (Fig. 2, Fig. 10). At the surface, the NW–SE striking Chalba Shear Zone boundary between the Halfway Gneiss and Durlacher Supersuite and the ellipsoidal inlier of Durlacher Supersuite are effectively captured in the models, with predominantly <1 km of <95% confidence regions separating >95% certainty domains (Fig. 10). However, there is an additional ~1 km wide NW–SE spur of Durlacher Supersuite immediately south of the main portion of the Chalba Shear Zone that is not captured in the models (Fig. 10a). Additionally, a thin sliver of mapped Durlacher Supersuite that encroaches the map in the NW section of the map (at ~7255500 mN) is modelled as Halfway Gneiss (Fig. 10), which explains why an abnormally high magnetic residual is present there (Fig. 8). Both of these discrepancies are probably a result of the aeromagnetic data integrating the magnetic response of Halfway Gneiss at depth. For example, in the GSWA cross-section across the NW–SE spur of Durlacher Supersuite, this NW–SE spur is interpreted to be underlain by Halfway Gneiss at depths below 2 km (Fig. 10).

In three-dimensions, the model maintains the sub-vertical to steeply-dipping regions of <95% certainty (i.e., geological boundaries; Fig. 11). This is particularly well viewed in the X–Y cross-section, where the posterior distributions reveal sub-vertical dips (>85°) that are comparable to the sub-vertical dips measured in the field and propagated into interpreted cross-sections (Fig. 2, Fig. 10). The modelled inclinations at 4 km depth are shallower (72°) than those interpreted by GSWA, which maintain dips of >85° at 4 km (Fig. 2). Seismic interpretation data indicates that the Chalba Shear Zone is dipping at ~65° at 4 km depth (Fig. 2), more comparable to our modelled dips that those interpreted from geological mapping. However, with the lack of drill hole data, it is difficult to know exactly whether the dips obtained from seismic interpretations, geological cross-section interpretation or modelled posterior distributions are correct. Despite these small discrepancies, the broad architecture of the model maintains the framework inferred from geological maps and cross-sections.

Another important output is that the modelled posterior distributions reveal that the Durlacher Supersuite is separated into two domains at the surface and shallow subsurface, one NE of the Chalba Shear Zone and the other as an ellipsoidal inlier, with a ~2.5 km-wide spur of >95% confidence Halfway Gneiss separating the two regions (Fig. 10). This model output is corroborated at the surface by geological mapping across the region (Fig. 2) but it was difficult to know whether this spur of Halfway Gneiss between the two Durlacher Supersuite domains continued at depth or was truncated in the near subsurface. Our results indicate that the spur of Halfway Gneiss continues until at least 4 km as assumed from geological mapping.

5.2 Implications and limitations for quantification of uncertainty in 3D geological models

To develop robust 3D geological models, fusion of geological and geophysical data in a fully probabilistic (Bayesian) method are vital for pure (e.g., plate reconstructions) and applied geological problems (e.g., mineral exploration). (de la Varga et al., 2018)

At the surface and near-surface (1 km), our model results are highly similar to independently-constrained geological maps and interpreted cross-sections (Fig. 2, Fig. 10), which are useful for mineral exploration applications that rarely exceed economic deposit depths of 1 km (McFadden et al., 2012). This similarity is in spite that the prior distribution for the depth to the geological contact was Gaussian with standard deviation 5 km at each control point (i.e., very permissive). Any further assumptions about depth to contact were uninformative by design to encapsulate the uncertainty in formulating initial models in areas about which little prior information is known. We emphasize that in well-studied areas our approach is unrealistic but for the purposes of mineral exploration in partially-covered terranes this approach is effective in defining geological boundaries. At the surface, distances between domains of >95% confidence rarely exceed 1 km, although the eastern part of the ellipsoidal Durlacher Supersuite inlier are as wide as ~2350 m. The minimum horizontal distance of <95% uncertainty at geological boundaries appears is ~400 m, which appears to be inherently linked to the line spacing of the aeromagnetic survey. Given that the gravity survey and geological field observations are far more widely spaced, and therefore have less control on the model outputs, the aeromagnetic data distribution is probably the dominant control on the width of uncertain regions. For geological mapping applications (particularly in the mineral exploration sector), geological mapping in these uncertain regions (if outcrop is available) and/or high-resolution geophysical surveys across these small regions of uncertainty provide targeted and cost-effective methods of yielding better 3D geological models. Where such regions are under cover and drilling is required to establish formation contacts, our results could also aid in constraining which areas should be drilled first to maximize information gain.

For deep applications (e.g., depth to sedimentary basement or depth to Moho), our models require Bayesian incorporation of additional geophysical and geological data, such as active seismic (Johnson et al., 2013), passive seismic (Zhu and Kanamori, 2000) or deep drill hole geological observations (e.g., petroleum wells; Beardsmore et al., 2016; McCalman et al., 2014). The

incorporation of structural measurement at the surface and in drill core (e.g., faults, folds) could also aid in in informing the prior, particularly when seismic data is unavailable, to provide geologically-feasible models at the surface and shallow subsurface. Other geophysical surveys (e.g., magnetotellurics, radiometrics) could significantly improve the model certainty by identifying other variables in which geological units can have different rock properties.

The similarity of geophysical responses from different geological units in terranes that are broadly granitic (e.g., Halfway Gneiss and Durlacher Supersuite) has meant that the time to reach convergence is significantly greater than studies with units that display vastly different rock properties (Beardsmore et al., 2016; McCalman et al., 2014). The limiting factor is the ability to explore very high-dimensional posteriors that result from a large-scale non-parametric model (i.e. the number of control points at a given resolution scales exponentially with area). Our modelled area is 13.5 × 13.5 km, which is useful for local-scale mineral exploration or detailed geological mapping, but may not be useful for reconnaissance-scale mineral exploration or terrane-scale geological modelling. Although our convergence times are not prohibitive for up-scaling the model to, for example 100 × 100 km, computational time becomes difficult for developing 3D geological models for significantly larger areas (e.g., the entire Gascoyne Province). Incorporation of other data types (see above) may be part of the solution but these all rely on hand sample petrophysical measurements, which are not routinely collected, let alone reported in the geosciences. To solve the paucity of petrophysical data compared to geophysical surveys, the MCMC sampler could be modified to a reversible jump scheme (Green, 1995; Sambridge, 2013), which is able to define the number of layers, sampling over rock categories to define a baseline prior irrespective of available rock property data.

An important limitation relates to the confidence of geological field observations. This study has simplified the probability distribution of each field observation to a single beta-binomial distribution, when different supporting data (age, geochemistry, sample descriptions) will provide different likelihoods. Probability distributions of samples that have, for example, age and geochemical data, should be significantly more confident than samples that only have hand sample descriptions. However, the exact range of probabilities to ascribe to these samples still requires some user input. To make this process Bayesian and fully remove operator bias in assigning probabilities to field observations, an independent study will need to be conducted that purely assesses the likelihood of geological field observations, taking into account information like age data, geochemistry data and sample descriptions.

Another limitation of the current model is that only two geological units are modelled. In this study, the rationale for such a simplification is that the Halfway Gneiss and Durlacher Supersuite comprise >90% of the surface geological units (Fig. 2) but this will rarely be the case in other problems. Integration of volumetrically-minor geological units may be vitally important with respect to mineral exploration or unravelling tectonic histories (e.g., Li-bearing pegmatites; Kesler et al., 2012). A major impediment to effectively modelling these volumetrically-minor units is the line spacing for different geophysical surveys and the potential paucity of geological field observations. In areas that are covered by shallow regolith

or sedimentary cover, the problem of modelling volumetrically-minor units is exacerbated due to unavailability of geological surface measurements. Here, the only solution is to have drill hole geological information. Integration of drill hole geological units is already possible to build into the modelling process but moderately deep (>100 m) drill hole data is lacking for the modelled part of the Gascoyne Province.

The Obsidian package – originally designed for sedimentary basins – inherently has limitations when applied to more complex geological geometries as are common in metamorphic terranes. Incorporation of Obsidian's advanced sampling methods with more sophisticated 3D modeling methods such as the implicit surface approach used by GemPy (de la Varga et al., 2018) can improve future geological modelling.

1 Conclusions

Bayesian integration of geological field observations with geophysical survey data yield statistically-reliable and geologically-plausible 3D models at the surface and shallow subsurface (<4 km). Approximately 89% of the model area has >95% certainty. Regions of <95% certainty are found exclusively within 1 km of mapped or inferred geological boundaries. The widths and positions of regions with <95% certainty are primarily a consequence of lack of geophysical, petrophysical or geological data. Our results indicate that these widths of these uncertain regions can be reduced by targeted geophysical surveys, petrophysical data collection and/or geological mapping. The integration of drill hole geological data and geophysical surveys with higher wavelengths (e.g., active seismic) are required to model deeper into the Earth's crust. Ultimately, the fusion of surface geological observations with geophysical data yield robust 3D geological models with narrow uncertainty regions at the surface and shallow subsurface that will be especially valuable for mineral exploration and the development of 3D geological models under cover. We anticipate that this framework provides a foundation for future applications in igneous and metamorphic terranes, particularly in the mineral exploration sector and especially in exploration under cover (McFadden et al., 2012).

Code and data availability

Aeromagnetic survey (Geological Survey of Western Australia, 1996) and gravity (Mathews and Jecks, 2010) data and metadata is freely available from the Geophysical Archive Data Delivery System (www.geoscience.gov.au/geophysical-data-delivery). Petrophysical data are from Aitken et al. (2014). Geological field observations are available from WAROX, a database managed by the Geological Survey of Western Australia (GSWA). Geological field observation data may be directly requested from GSWA.

Model code is stored on GitHub (DOI: 10.5281/zenodo.2580422)

Author contributions

HKHO, RC, CC, SR and RDM designed the project. RS and DK developed the experiments, model code and performed the simulations. HKHO and RS prepared the manuscript with contributions from all co-authors.

Acknowledgements

This manuscript benefited significantly from the detailed comments of two anonymous reviewers, G. Laurent and the editorial handling of Juan Alcalde. G.A. Houseman is thanked for discussions and comments on an early version of this manuscript. This research was funded by the Science and Industry Endowment Fund as part of The Distal Footprints of Giant Ore Systems: UNCOVER Australia Project (RP04-063) — Capricorn Distal Footprints.

Competing interests

The authors declare that they have no conflict of interest.

References

Abedi, M. and Norouzi, G.-H.: Integration of various geophysical data with geological and geochemical data to determine additional drilling for copper exploration, Journal of Applied Geophysics, 83, 35-45, 2012.

Aitken, A. R. A., Joly, A., Dentith, M. C., Johnson, S. P., Thorne, A. M., and Tyler, I. M.: 3D architecture, structural evolution, and mineral prospectivity of the Gascoyne Province: Report 123, Geological Survey of Western Australia, Perth, Australia, 2014. 2014.

Andrieu, C. and Thoms, J.: A tutorial on adaptive MCMC, Statistics and computing, 18, 343-373, 2008.

Atchad'e, Y., Fort, G., Moulines, E., and Priouret, P.: Adaptive Markov chain Monte Carlo: theory and methods. In: Bayesian Time Series Models, Barber, D., Cemgil, A. T., and Chiappa, S. (Eds.), Cambridge University Press, 2011.

Backus, G.: Inference from Inadequate and Inaccurate Data, I, Proceedings of the National Academy of Sciences, 65, 1-7, 1970.

Beardsmore, G., Durrant-Whyte, H., McCalman, L., O'Callaghan, S., and Reid, A.: A Bayesian inference tool for geophysical joint inversions, ASEG Extended Abstracts, 2016, 1-10, 2016.

Bijani, R., Lelièvre, P. G., Ponte-Neto, C. F., and Farquharson, C. G.: Physical-property-, lithology-and surface-geometry-based joint inversion using Pareto Multi-Objective Global Optimization, Geophysical Journal International, 209, 730-748, 2017.

Bosch, M., Meza, R., Jiménez, R., and Hönig, A.: Joint gravity and magnetic inversion in 3D using Monte Carlo methods, GEOPHYSICS, 71, G153-G156, 2006.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L.: Handbook of markov chain monte carlo, CRC press, 2011.

Brooks, S. P. and Gelman, A.: General methods for monitoring convergence of iterative simulations, Journal of computational and graphical statistics, 7, 434-455, 1998.

Cawood, P. A. and Korsch, R. J.: Assembling Australia: Proterozoic building of a continent, Precambrian Research, 166, 1-35, 2008.
Chandra, R., Müller, R. D., Deo, R., Butterworth, N., Salles, T., and Cripps, S.: Multi-core parallel tempering Bayeslands for basin and landscape evolution, arXiv preprint arXiv:1806.10939, 2018. 2018.
Cotter, S. L., Roberts, G. O., Stuart, A. M., and White, D.: MCMC methods for functions: modifying old algorithms to make them faster, Statistical Science, 2013. 424-446, 2013.
Cowles, M. K. and Carlin, B. P.: Markov chain Monte Carlo convergence diagnostics: a comparative review, Journal of the American Statistical Association, 91, 883-904, 1996.
de la Varga, M., Schaaf, A., and Wellmann, F.: GemPy 1.0: open-source stochastic geological modeling and inversion, 2018. 2018.
Dentith, M. and Mudge, S. T.: Geophysics for the mineral exploration geoscientist, Cambridge University Press, 2014.
Earl, D. J. and Deem, M. W.: Parallel tempering: Theory, applications, and new perspectives, Physical Chemistry Chemical Physics, 7, 3910-3916, 2005.
Fullagar, P. K., Pears, G. A., and McMonnies, B.: Constrained inversion of geologic surfaces— pushing the boundaries, The Leading Edge, 27, 98-105, 2008.
Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M., and Stephenson, J.: Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, Marine and Petroleum Geology, 26, 525-535, 2009.
Gelman, A. and Rubin, D. B.: Inference from iterative simulation using multiple sequences, Statistical science, 7, 457-472, 1992.
Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian data analysis, Chapman and Hall/CRC, 2013.
Geological Survey of Western Australia: Bangemall Basin, Job Number 2048, Perth, Western Australia, 1996.
Geweke, J.: Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bayesian Statistics, University Press, Minneapolis, 1992.
Geweke, J.: Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA, 1991.
Geyer, W.: The importance of suppression of turbulence by stratification on the estuarine turbidity maximum, Estuaries and Coasts, 16, 113-125, 1993.
Giraud, J., Lindsay, M., Ogarko, V., Jessell, M., Martin, R., and Pakyuz-Charrier, E.: Integration of geological uncertainty into geophysical inversion by means of local gradient regularization, Solid Earth, 10, 193-210, 2019.
Giraud, J., Pakyuz-Charrier, E., Jessell, M., Lindsay, M., Martin, R., and Ogarko, V.: Uncertainty reduction through geologically conditioned petrophysical constraints in joint inversionConditioned petrophysical constraints, Geophysics, 82, ID19-ID34, 2017.
Gradstein, F. M., Ogg, J. G., Schmitz, M. D., and Ogg, G. M., (Eds.): The Geologic Time Scale 2012 2-Volume Set (Vol. 2), Elsevier, Oxford, U.K., 2012.
Green, P. J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika, 82, 711-732, 1995.
Grose, L., Laurent, G., Aillères, L., Armit, R., Jessell, M., and Cousin-Dechenaud, T.: Inversion of Structural Geology Data for Fold Geometry, Journal of Geophysical Research: Solid Earth, 123, 6318-6333, 2018.
Guillaume, B., Ian, L., and Youssef, M.: Bayesian inverse problems with Monte Carlo forward models, Inverse Problems and Imaging, 7, 81-105, 2013.

Guillen, A., Calcagno, P., Courrioux, G., Joly, A., and Ledru, P.: Geological modelling from field data and geological knowledge: Part II. Modelling validation using gravity and magnetic data inversion, Physics of the Earth and Planetary Interiors, 171, 158-169, 2008.

Hapke, C. and Plant, N.: Predicting coastal cliff erosion using a Bayesian probabilistic model, Marine Geology, 278, 140-149, 2010.

Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, 1970. 1970.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G.: Bayesian nonparametrics, Cambridge University Press, 2010.

Hu, Z., Yao, Z., and Li, J.: On an adaptive preconditioned Crank–Nicolson MCMC algorithm for infinite dimensional Bayesian inference, Journal of Computational Physics, 332, 492-503, 2017.

Hukushima, K. and Nemoto, K.: Exchange Monte Carlo method and application to spin glass simulations, Journal of the Physical Society of Japan, 65, 1604-1608, 1996.

Jessell, M., Aillères, L., De Kemp, E., Lindsay, M., Wellmann, J. F., Hillier, M., Laurent, G., Carmichael, T., and Martin, R.: Next generation three-dimensional geologic modeling and inversion, Society of Economic Geologists Special Publication, 18, 261-272, 2014.

Jessell, M., Pakyuz-Charrier, E., Lindsay, M., Giraud, J., and De Kemp, E.: Assessing and Mitigating Uncertainty in Three-Dimensional Geologic Models in Contrasting Geologic Scenarios, Society of Economic Geologists Special Publication, 21, 63-74, 2018.

Jessell, M. W., Ailleres, L., and de Kemp, E. A.: Towards an integrated inversion of geoscientific data: What price of geology?, Tectonophysics, 490, 294-306, 2010.

Johnson, S. P., Korhonen, F. J., Kirkland, C. L., Cliff, J. B., Belousova, E. A., and Sheppard, S.: An isotopic perspective on growth and differentiation of Proterozoic orogenic crust: From subduction magmatism to cratonization, Lithos, 268–271, 76-86, 2017.

Johnson, S. P., Sheppard, S., Groenewald, P. B., and Farrell, T. R.: Yinnetharra, WA Sheet 2148, Geological Survey of Western Australia, 1:100 000 Geological series, 2012.

Johnson, S. P., Sheppard, S., Rasmussen, B., Wingate, M. T. D., Kirkland, C. L., Muhling, J. R., Fletcher, I. R., and Belousova, E. A.: Two collisions, two sutures: Punctuated pre-1950 Ma assembly of the West Australian Craton during the Ophthalmian and Glenburgh Orogenies, Precambrian Research, 189, 239-262, 2011.

Johnson, S. P., Thorne, A. M., Tyler, I. M., Korsch, R. J., Kennett, B. L. N., Cutten, H. N., Goodwin, J., Blay, O., Blewett, R. S., Joly, A., Dentith, M. C., Aitken, A. R. A., Holzschuh, J., Salmon, M., Reading, A., Heinson, G., Boren, G., Ross, J., Costelloe, R. D., and Fomin, T.: Crustal architecture of the Capricorn Orogen, Western Australia and associated metallogeny, Australian Journal of Earth Sciences, 60, 681-705, 2013.

Joly, A., Porwal, A., and McCuaig, T. C.: Exploration targeting for orogenic gold deposits in the Granites-Tanami Orogen: Mineral system analysis, targeting model and prospectivity analysis, Ore Geology Reviews, 48, 349-383, 2012.

Kaipio, J. and Somersalo, E.: Statistical and computational inverse problems, Springer Science & Business Media, 2006.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M.: Markov chain Monte Carlo in practice: a roundtable discussion, The American Statistician, 52, 93-100, 1998.

Kesler, S. E., Gruber, P. W., Medina, P. A., Keoleian, G. A., Everson, M. P., and Wallington, T. J.: Global lithium resources: Relative importance of pegmatite, brine and other deposits, Ore Geology Reviews, 48, 55-69, 2012.

Korhonen, F. J., Johnson, S. P., Wingate, M. T. D., Kirkland, C. L., Fletcher, I. R., Dunkley, D. J., Roberts, M. P., Sheppard, S., Muhling, J. R., and Rasmussen, B.: Radiogenic heating and craton-margin plate stresses as drivers for intraplate orogeny, Journal of Metamorphic Geology, 35, 631-661, 2017.

Krapež, B., Müller, S. G., Fletcher, I. R., and Rasmussen, B.: A tale of two basins? Stratigraphy and detrital zircon provenance of the Palaeoproterozoic Turee Creek and Horseshoe basins of Western Australia, Precambrian Research, 294, 67-90, 2017.

Li, Y. and Oldenburg, D. W.: 3-D inversion of gravity data, Geophysics, 63, 109-119, 1998.

Li, Y. and Oldenburg, D. W.: 3-D inversion of magnetic data, Geophysics, 61, 394-408, 1996.

Lindsay, M. D., Jessell, M. W., Ailleres, L., Perrouty, S., de Kemp, E., and Betts, P. G.: Geodiversity: Exploration of 3D geological model space, Tectonophysics, 594, 27-37, 2013.

Malinverno, A.: Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, Geophysical Journal International, 151, 675-688, 2002.

Martin, D. M. and Morris, P. A.: Tectonic setting and regional implications of ca 2.2 Ga mafic magmatism in the southern Hamersley Province, Western Australia, Australian Journal of Earth Sciences, 57, 911-931, 2010.

Mathews, L. R. and Jecks, M. J.: Gascoyne Gravity Survey, Report Number R2010006, Atlas Geophysics, Morley, Western Australia, 1-93 pp., 2010.

McCalman, L., O'Callaghan, S. T., Reid, A., Shen, D., Carter, S., Krieger, L., Beardsmore, G., Bonilla, E. V., and Ramos, F. T.: Distributed Bayesian geophysical inversions, Stanford University, Stanford, California, February 24-26, 20142014, 1-11.

McFadden, P., Barnicoat, A., Blewett, R. S., Collet, D., Collins, W., Dransfield, M., Fitzgerald, D., Haydon, R., Hough, R., Hronsky, J., McCuaig, C., McWilliams, M., Müller, R. D., Whiting, T., and Woodgate, P.: Searching the Deep Earth: a vision for exploration geoscience in Australia, Australian Academy of Science, Canberra, Australia, 47 pp., 2012.

Merdith, A. S., Collins, A. S., Williams, S. E., Pisarevsky, S., Foden, J. D., Archibald, D. B., Blades, M. L., Alessio, B. L., Armistead, S., Plavsa, D., Clark, C., and Müller, R. D.: A full-plate global reconstruction of the Neoproterozoic, Gondwana Research, 50, 84-134, 2017.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, The journal of chemical physics, 21, 1087-1092, 1953.

Miasojedow, B., Moulines, E., and Vihola, M.: An adaptive parallel tempering algorithm, Journal of Computational and Graphical Statistics, 22, 649-664, 2013.

Mosegaard, K. and Tarantola, A.: Monte Carlo sampling of solutions to inverse problems, Journal of Geophysical Research: Solid Earth, 100, 12431-12447, 1995.

Mosegaard, K. and Vestergaard, P. D.: A simulated annealing approach to seismic model optimization with sparse prior information, Geophysical Prospecting, 39, 599-611, 1991.

Oladyshkin, S., Class, H., and Nowak, W.: Bayesian updating via bootstrap filtering combined with data-driven polynomial chaos expansions: methodology and application to history matching for carbon dioxide storage in geological formations, Computational Geosciences, 17, 671-687, 2013.

Oldenburg, D. W. and Li, Y.: Inversion for applied geophysics: A tutorial, Investigations in geophysics, 13, 89-150, 2005.

Olierook, H. K. H., Sheppard, S., Johnson, S. P., Occhipinti, S. A., Reddy, S. M., Clark, C., Fletcher, I. R., Rasmussen, B., Zi, J.-W., Pirajno, F., LaFlamme, C., Do, T., Ware, B., Blandthorn, E., Lindsay, M., Lu, Y.-J., Crossley, R. J., and Erickson, T. M.: Extensional episodes in the Paleoproterozoic Capricorn Orogen, Western Australia, revealed by petrogenesis and geochronology of mafic–ultramafic rocks, Precambrian Research, 306, 22-40, 2018.

Pakyuz-Charrier, E., Giraud, J., Ogarko, V., Lindsay, M., and Jessell, M.: Drillhole uncertainty propagation for three-dimensional geological modeling using Monte Carlo, Tectonophysics, 747-748, 16-39, 2018a.

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

Pall, J., Chandra, R., Azam, D., Salles, T., Webster, J. M., and Cripps, S.: BayesReef: A Bayesian inference framework for modelling reef growth in response to environmental change and biological dynamics, arXiv preprint arXiv:1808.02763, 2018. 2018.

Parker, R. L.: Understanding inverse theory, Annual Review of Earth and Planetary Sciences, 5, 35-64, 1977.

Piechocka, A. M., Gregory, C. J., Zi, J.-W., Sheppard, S., Wingate, M. T. D., and Rasmussen, B.: Monazite trumps zircon: applying SHRIMP U–Pb geochronology to systematically evaluate emplacement ages of leucocratic, low-temperature granites in a complex Precambrian orogen, Contributions to Mineralogy and Petrology, 172, 63, 2017.

Raftery, A. E. and Lewis, S. M.: Implementing mcmc, Markov chain Monte Carlo in practice, 1996. 115-130, 1996.

Ramos, F. T., Bonilla, E., McCalman, L., O'Callaghan, S., Reid, A., Uther, W. T. B., Sambridge, M., and Rawling, T.: Bayesian data fusion for geothermal exploration, 2012. 2012.

Rasmussen, B.: Radiometic dating of sedimentary rocks: the application of diagenetic xenotime geochronology, Earth-Science Reviews, 68, 197-243, 2005.

Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., and Troldborg, L.: Review of strategies for handling geological uncertainty in groundwater flow and transport modeling, Advances in Water Resources, 36, 36-50, 2012.

Reid, A., Simon Timothy, O. C., Bonilla, E. V., McCalman, L., Rawling, T., and Ramos, F.: Bayesian Joint Inversions for the Exploration of Earth Resources, 2013 2013, 2877-2884.

Rocca, P., Benedetti, M., Donelli, M., Franceschini, D., and Massa, A.: Evolutionary optimization as applied to inverse scattering problems, Inverse Problems, 25, 123003, 2009.

Rudolf, D. and Sprungk, B.: On a generalization of the preconditioned Crank–Nicolson Metropolis algorithm, Foundations of Computational Mathematics, 18, 309-343, 2018.

Ruggeri, P., Irving, J., and Holliger, K.: Systematic evaluation of sequential geostatistical resampling within MCMC for posterior sampling of near-surface geophysical inverse problems, Geophysical Journal International, 202, 961-975, 2015.

Sambridge, M.: Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble, Geophysical Journal International, 138, 727-746, 1999.

Sambridge, M.: A parallel tempering algorithm for probabilistic sampling and multimodal optimization, Geophysical Journal International, 196, 357-374, 2013.

Sambridge, M. and Mosegaard, K.: Monte Carlo methods in geophysical inverse problems, Reviews of Geophysics, 40, 3-1, 2002.

Sambridge, M. S. and Compston, W.: Mixture modeling of multi-component data sets with application to ion-probe zircon ages, Earth and Planetary Science Letters, 128, 373-390, 1994.

Scalzo, R., Kohn, D., Olierook, H. K. H., Houseman, G., Chandra, R., and Cripps, S.: Efficiency and robustness in Monte Carlo sampling of 3-D geophysical inversions with Obsidian v0.1.2: Setting up for success, Geoscientific Model Development, in review. arXiv:1812.00318v00311, in review.

Scholl, C., Neumann, J., Watts, M. D., Hallinan, S., and Mulè, S.: Geologically constrained 2D and 3D airborne EM inversion through cross-gradient regularization and multi-grid efficiency, ASEG Extended Abstracts, 2016, 1-6, 2016.

Seifert, D., Sonnenborg, T. O., Refsgaard, J. C., Højberg, A. L., and Troldborg, L.: Assessment of hydrological model predictive ability given multiple conceptual geological models, Water Resources Research, 48, 2012.
Selley, R. C. and Sonnenberg, S. A.: Elements of petroleum geology, Academic Press, 2014.
Sen, M. K. and Stoffa, P. L.: Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion, Geophysical Prospecting, 44, 313-350, 1996.
Sen, M. K. and Stoffa, P. L.: Global optimization methods in geophysical inversion, Cambridge University Press, 2013.
Shafer, G.: Belief functions and parametric models, Journal of the Royal Statistical Society. Series B (Methodological), 1982. 322-352, 1982.
Shen, W., Ritzwoller, M. H., Schulte-Pelkum, V., and Lin, F.-C.: Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, Geophysical Journal International, 192, 807-836, 2013.
Sheppard, S., Bodorkos, S., Johnson, S. P., Wingate, M. T. D., and Kirkland, C. L.: The Paleoproterozoic Capricorn Orogeny: Intracontinental reworking not continent–continent collision, Geological Survey of Western Australia, Report 108, 33p, 2010a.
Sheppard, S., Fletcher, I. R., Rasmussen, B., Zi, J.-W., Muhling, J. R., Occhipinti, S. A., Wingate, M. T. D., and Johnson, S. P.: A new Paleoproterozoic tectonic history of the eastern Capricorn Orogen, Western Australia, revealed by U–Pb zircon dating of micro-tuffs, Precambrian Research, 286, 1-19, 2016.
Sheppard, S., Johnson, S. P., Wingate, M. T. D., Kirkland, C. L., and Pirajno, F.: Explanatory notes for the Gascoyne Province, Geological Survey of Western Australia, Perth, Western Australia, 2010b.
Sheppard, S., Occhipinti, S. A., and Nelson, D. R.: Intracontinental reworking in the Capricorn Orogen, Western Australia: the 1680 – 1620 Ma Mangaroon Orogeny, Australian Journal of Earth Sciences, 52, 443-460, 2005.
Sheppard, S., Rasmussen, B., Muhling, J. R., Farrell, T. R., and Fletcher, I. R.: Grenvillian-aged orogenesis in the Palaeoproterozoic Gascoyne Complex, Western Australia: 1030–950 Ma reworking of the Proterozoic Capricorn Orogen, Journal of Metamorphic Geology, 25, 477-494, 2007.
Tarantola, A. and Valette, B.: Generalized nonlinear inverse problems solved using the least squares criterion, Reviews of Geophysics, 20, 219-232, 1982.
van Ravenzwaaij, D., Cassey, P., and Brown, S. D.: A simple introduction to Markov Chain Monte–Carlo sampling, Psychonomic bulletin & review, 25, 143-154, 2018.
Wellmann, J. F., de la Varga, M., Murdie, R. E., Gessner, K., and Jessell, M.: Uncertainty estimation for a geological model of the Sandstone greenstone belt, Western Australia – insights from integrated geological and geophysical inversion in a Bayesian inference framework, Geological Society, London, Special Publications, 453, 41-56, 2018.
Wellmann, J. F. and Regenauer-Lieb, K.: Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models, Tectonophysics, 526–529, 207-216, 2012.
Ye, M., Pohlmann, K. F., Chapman, J. B., Pohll, G. M., and Reeves, D. M.: A model-averaging method for assessing groundwater conceptual model uncertainty, Groundwater, 48, 716-728, 2010.
Zheglova, P., Lelièvre, P. G., and Farquharson, C. G.: Multiple level-set joint inversion of traveltime and gravity data with application to ore delineation: A synthetic study, GEOPHYSICS, 83, R13-R30, 2018.
Zhu, L. and Kanamori, H.: Moho depth variation in southern California from teleseismic receiver functions, Journal of Geophysical Research: Solid Earth, 105, 2969-2980, 2000.

Figures

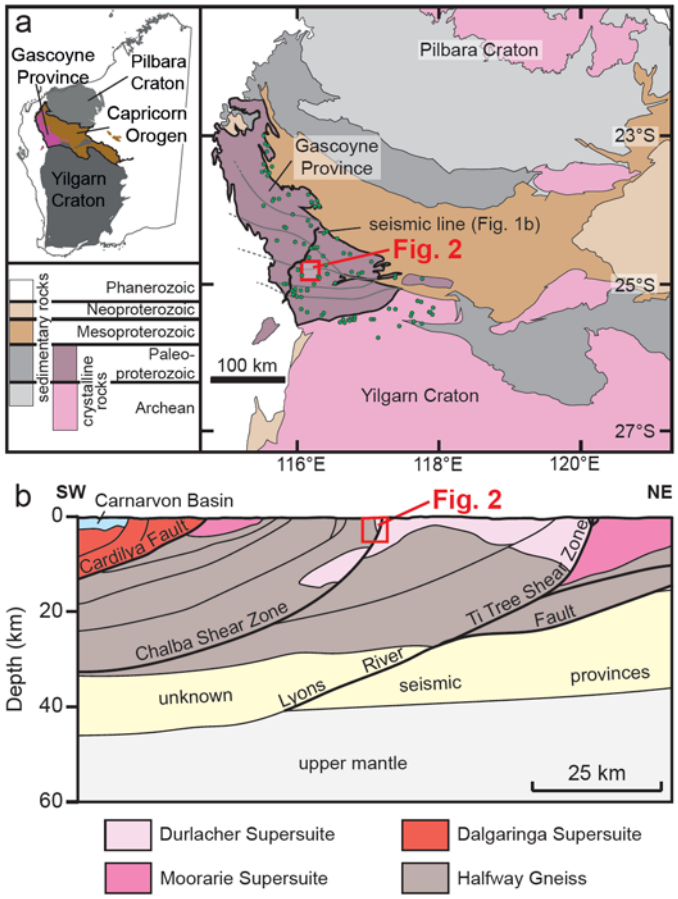

Fig. 1: (a) Geological map of the West Australian Craton, modified from Sheppard et al. (2016), showing location of seismic section and petrophysical data (green circles) from Aitken et al. (2014). (b) Interpretation of part of seismic line 10GA-CP2, after Johnson et al. (2013). The modelled region in this study is shown on both the map and seismic section (cf. Fig. 2).

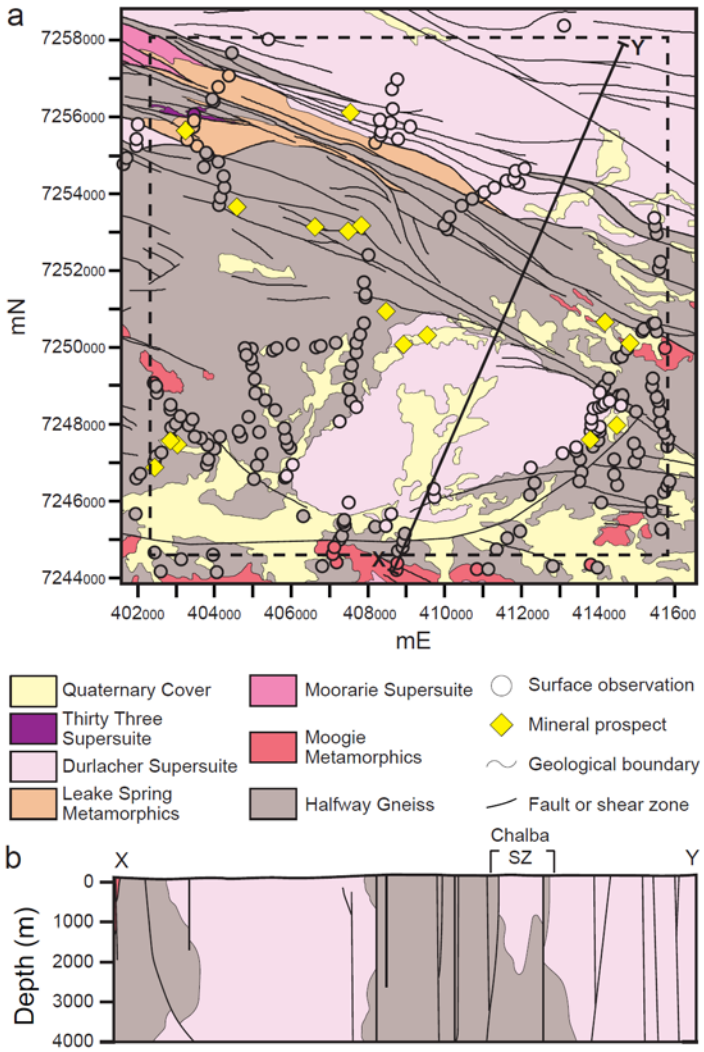

Fig. 2: (a) Detailed geological map of a 15 × 15 km portion of the Gascoyne Province with the same centre as the modelled 13.5 × 13.5 km area (dashed area), showing geological units, structural discontinuities (faults/shear zones), geological surface observations and mineral prospects and deposits. (b) Cross-section through detailed geological map, where vertical and horizontal scale are 1:1. Map and cross-section compiled using 1:100,000 geological maps from the Geological Survey of Western Australia.

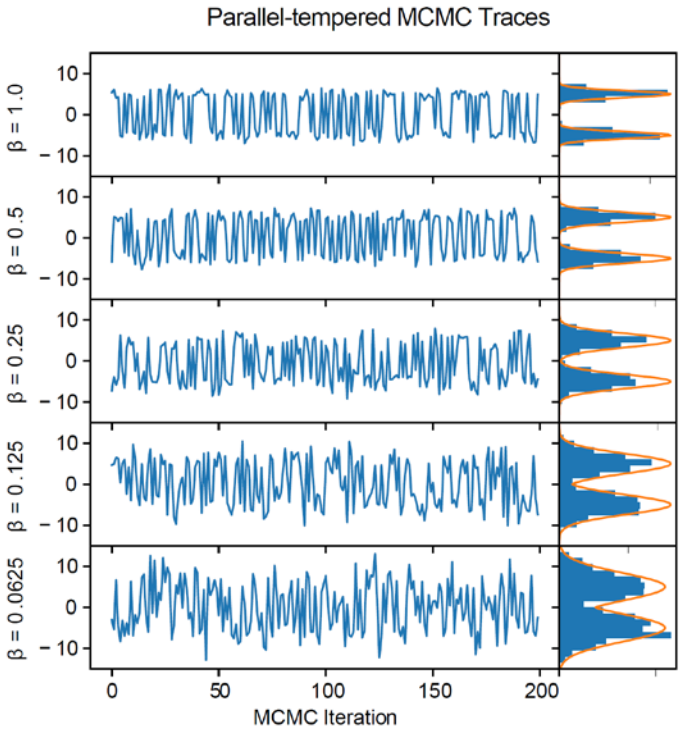

Fig. 3: Illustration of how parallel-tempered Markov chain Monte Carlo explores multi-modal distributions. Trace plots are shown for sampling of a mixture of two well-separated Gaussians, with each trace corresponding to a different value of the inverse temperature parameter β. In the chain with the lowest value of β, the modes are explored freely; proposals that swap states between adjacent chains on the β ladder enable exchange of information to the β = 1 chain, allowing it to explore both Gaussian modes readily.

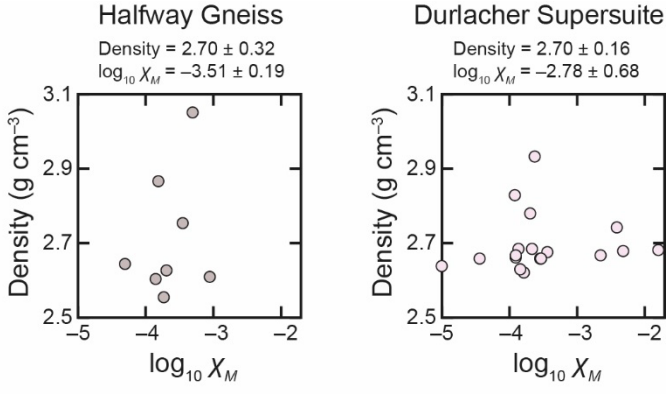

Fig. 4: Measured density and magnetic susceptibility ($\log_{10} \chi_M$) for modelled geological units: (a) Halfway Gneiss, (b) Durlacher Supersuite. Mean density and susceptibility errors are quoted at 2σ uncertainty.

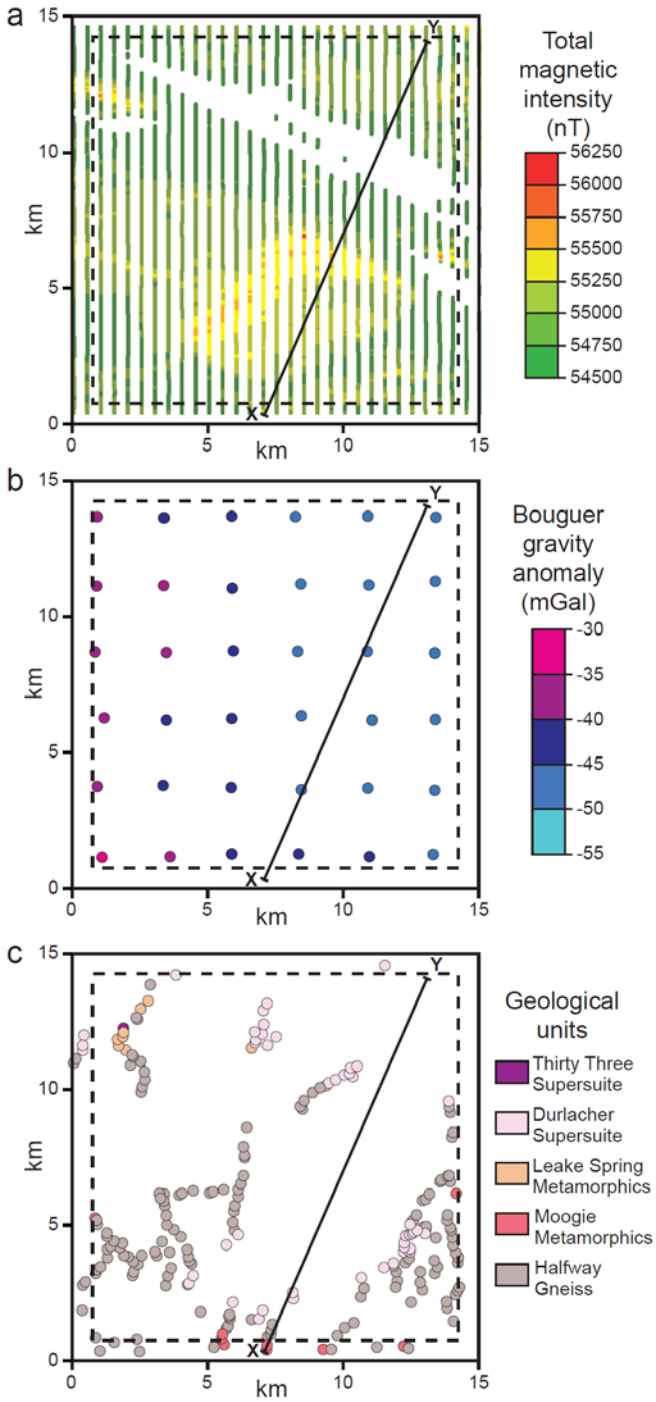

Fig. 5: Measured geophysical survey data and geological field observations for 15 × 15 km area, with dashed line representing modelled 13.5 × 13.5 km area. (a) Aeromagnetic data, showing locations of measured data on flight lines. (b) Gravity data, showing locations of ground-based measuring stations. (c) Geological field observations. Note the paucity of units other than the Halfway Gneiss and Durlacher Supersuite. X–Y line correspond to cross-section line in Fig. 2.

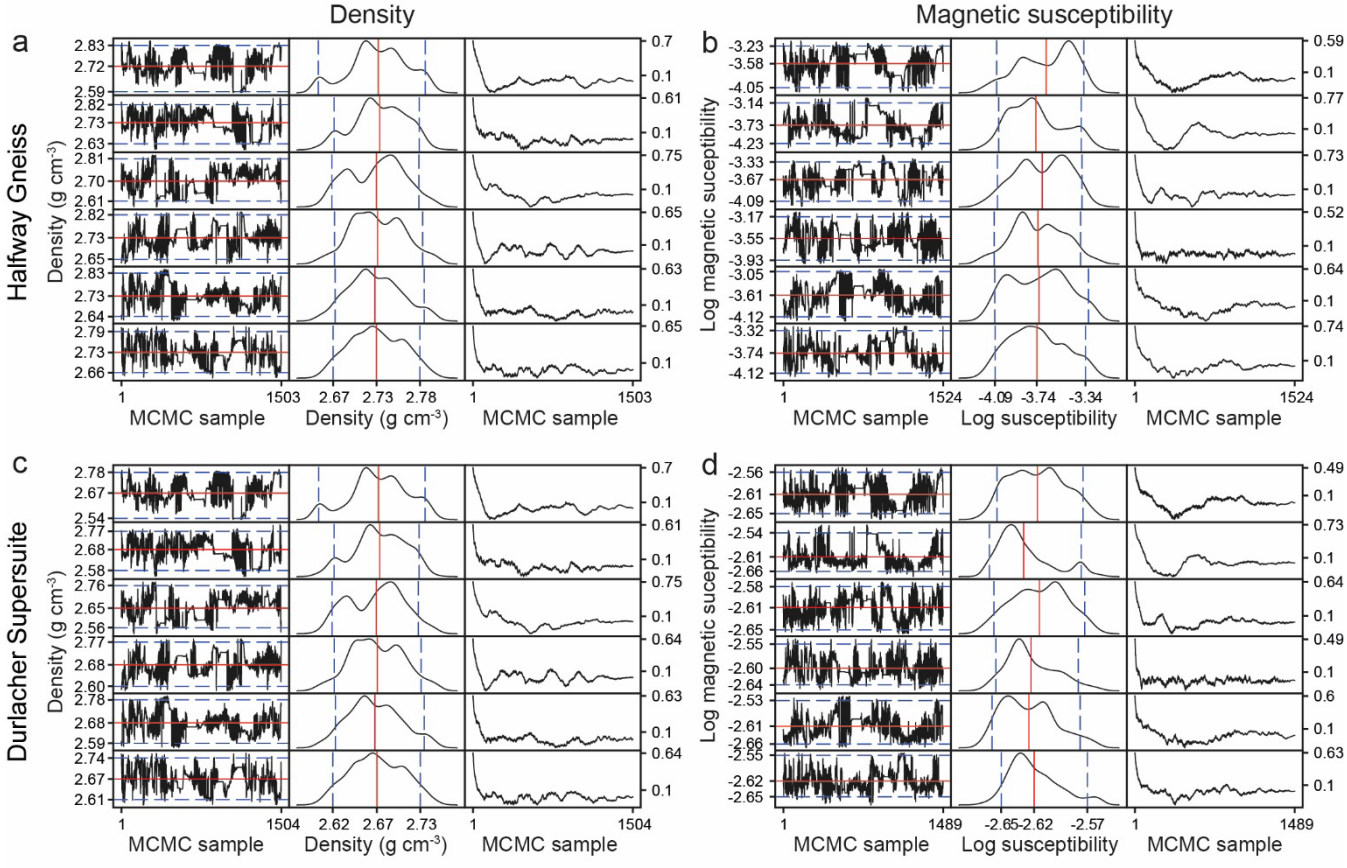

Fig. 6: MCMC diagnostics for the modelled Halfway Gneiss (a,b) and Durlacher Supersuite (c,d), for density (a,c) and magnetic susceptibility (b,d). Column 1 in each panel shows six of the twelve lowest temperature chains. Column 2 in each panel shows the distribution of petrophysical properties per chain. The red line and blue lines in columns 1 and 2 are the mean and 2σ, respectively. Column 3 shows the autocorrelation time from the beginning of each chain to the end. Columns 1 and 3 MCMC iterations are thinned by 1000 (i.e., total number of samples is approximately 1.5 million per chain).

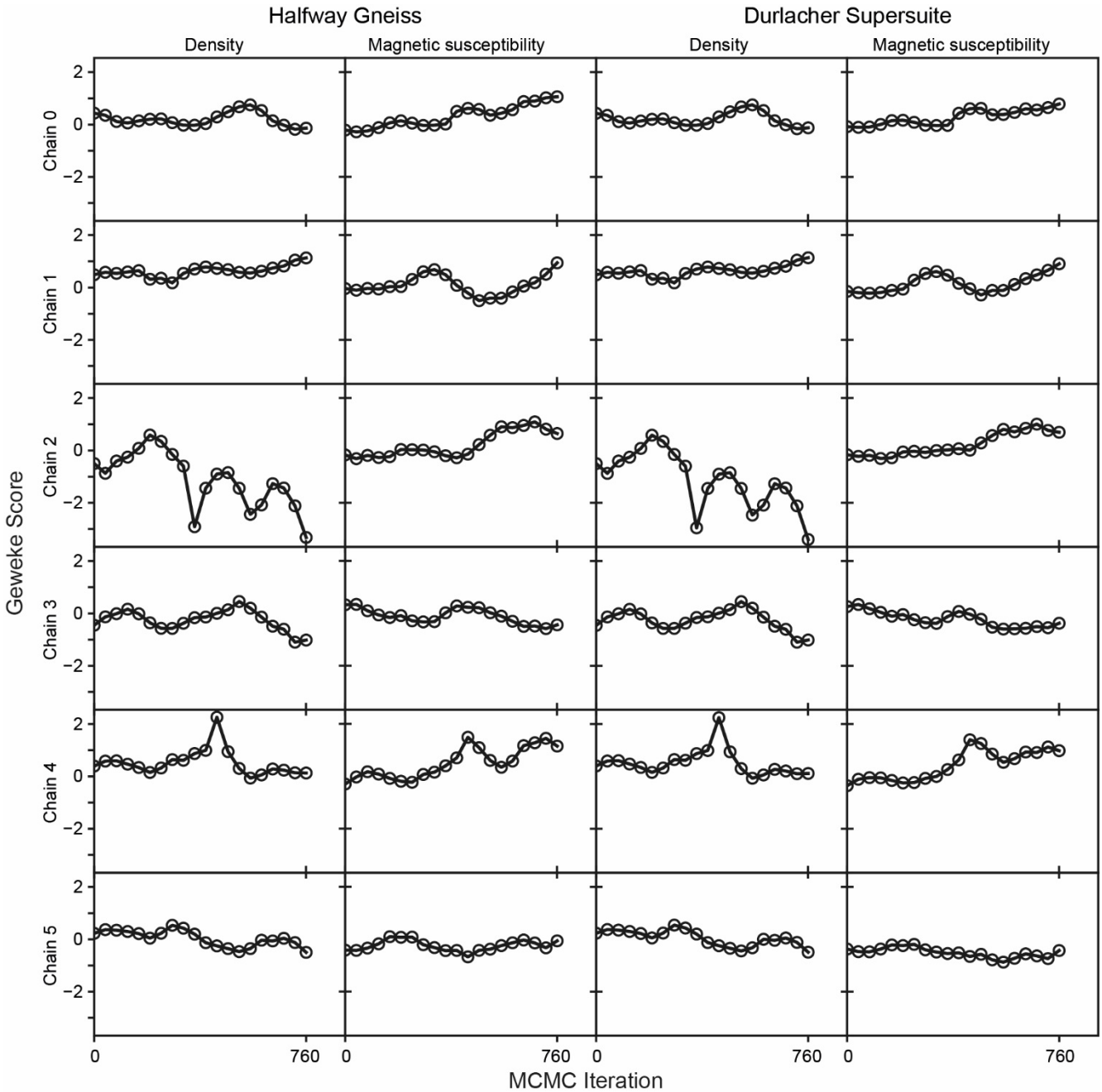

Fig. 7: Geweke scores for modelled densities and magnetic susceptibilities for the Halfway Gneiss and Durlacher Supersuite, shown for six chains.

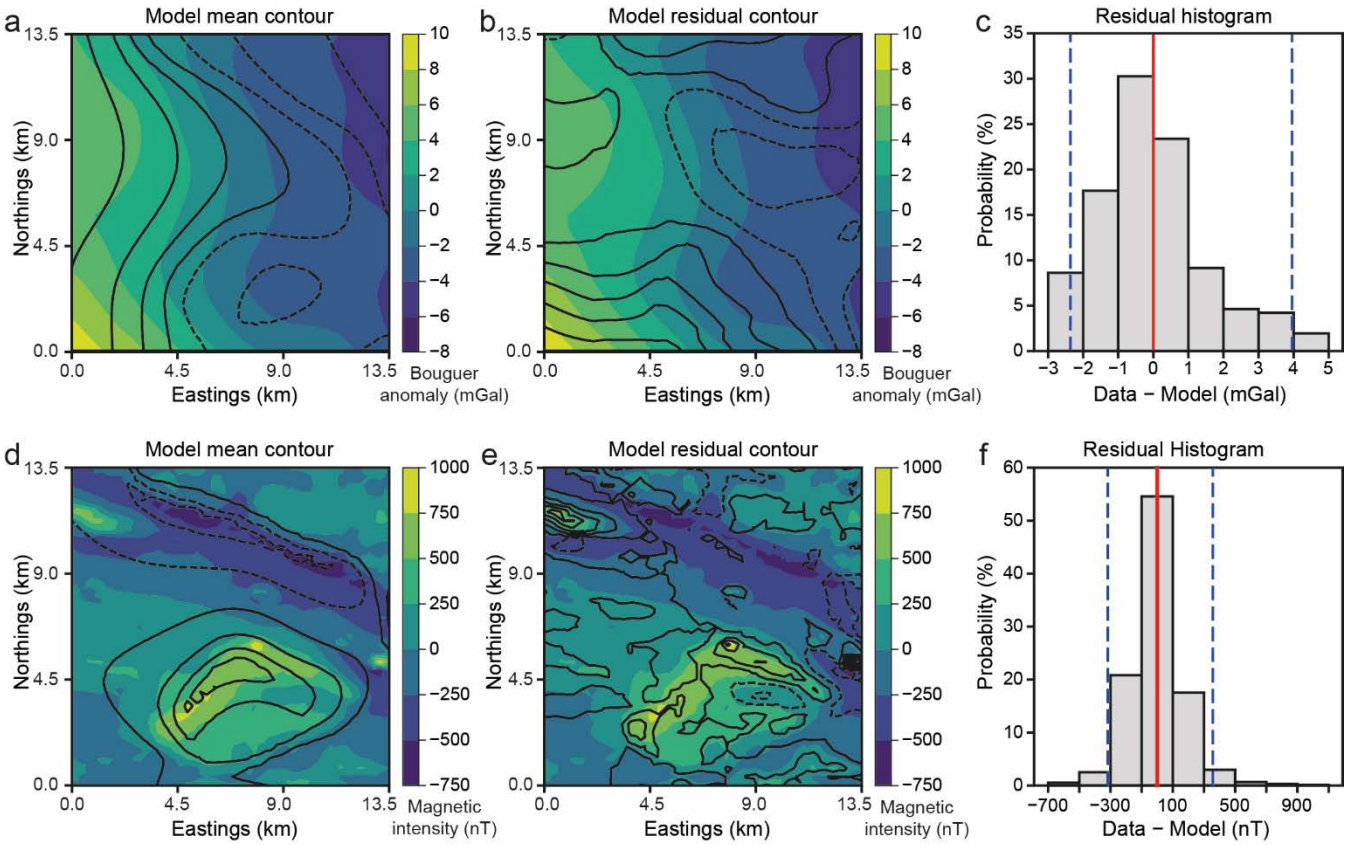

Fig. 8: Modelled Bouguer anomaly and magnetic intensity. Modelled mean contours of (a) Bouguer anomaly and (d) magnetic intensity compared to interpolated mean colored data. Modelled residual (i.e., data – model) contours for (b) Bouguer anomaly and (e) magnetic intensity compared to interpolated mean colored data. In a–d, contour lines are in 2 mGal and 250 nT increments for gravity and magnetic intensity, respectively, where solid lines ≥ 0 and dashed lines < 0. Histograms of residuals for (c) Bouguer anomaly and (f) magnetic intensity.

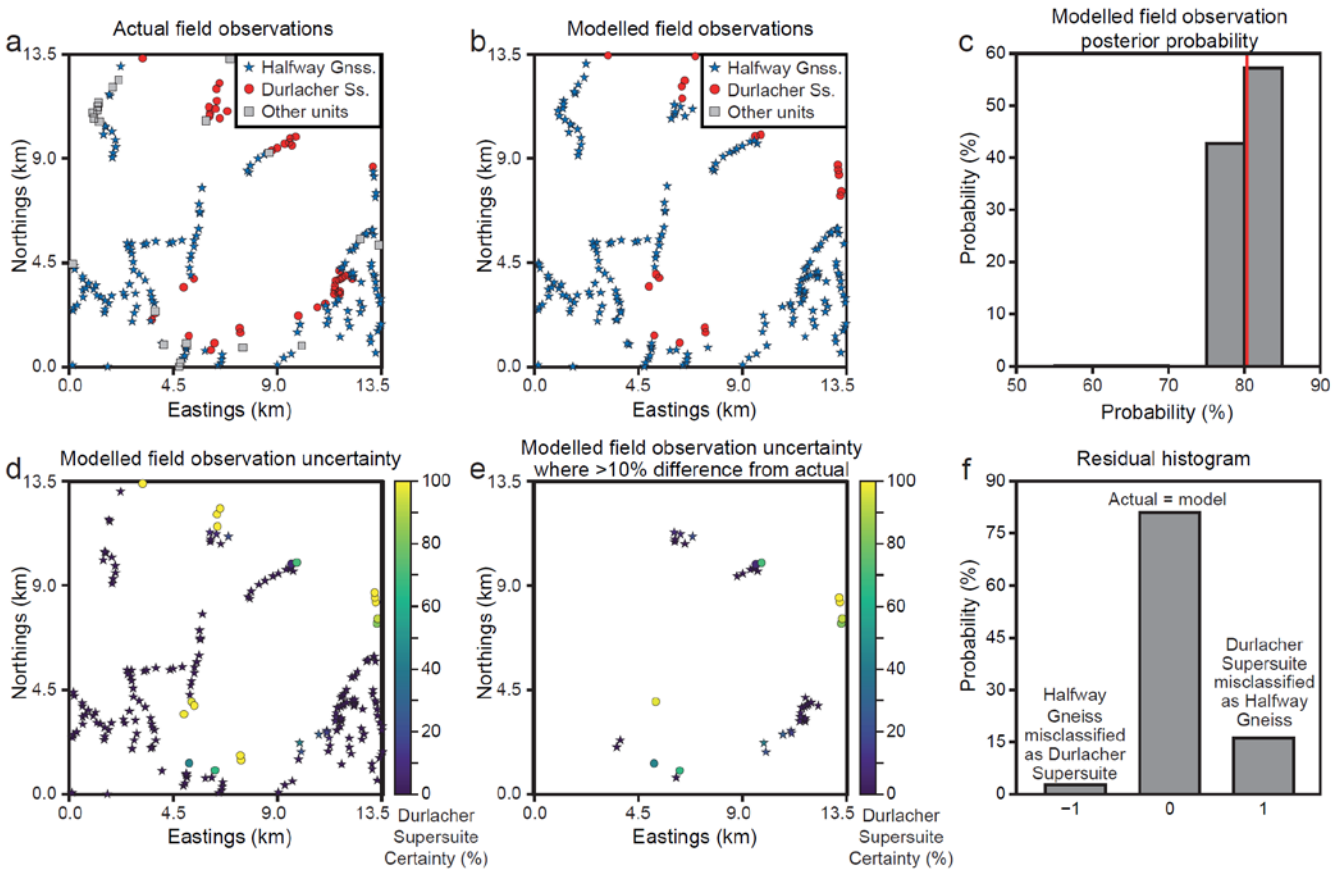

Fig. 9: Actual vs. modelled field observations. (a) Actual field observations. (b) Modelled field observations, showing highest probability geological unit. (c) Modelled field observation posterior probability. (d) Modelled field observations with uncertainty. (e) Modelled field observations with uncertainty, where >10% different from actual field observations. (f) Residual histogram, where –1 is Durlacher Supersuite misclassified as Halfway Gneiss, 0 is correctly classified and 1 is Halfway Gneiss misclassified as Durlacher Supersuite.

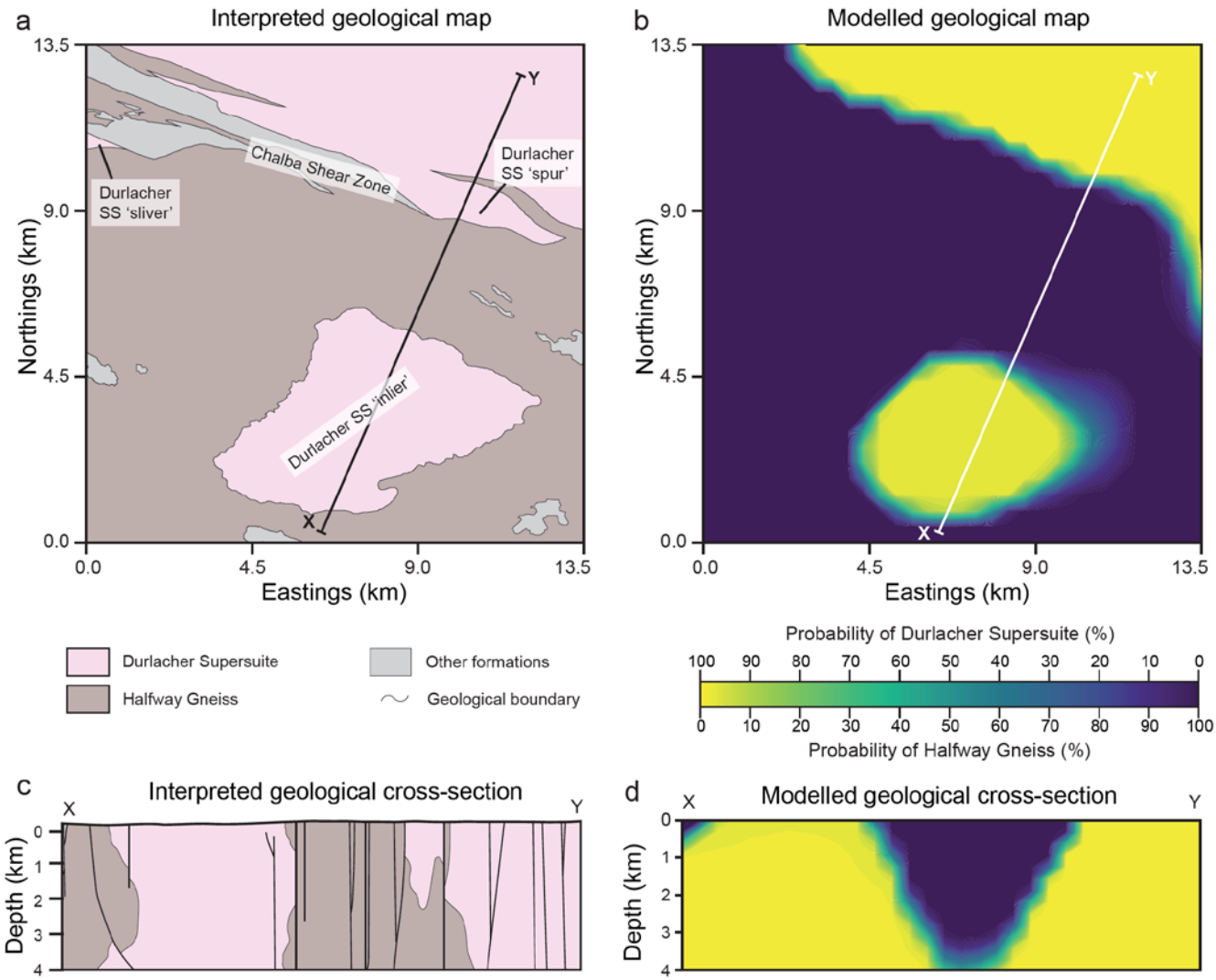

Fig. 10: Voxelized posterior distributions of the 3D geological model compared to simplified GSWA maps and cross-sections. (a) Simplified geological map from Figure 2, showing interpreted boundaries for the Durlacher Supersuite, Halfway Gneiss and other formations (undifferentiated). (b) Model of the surface. (c) Interpreted geological cross-section through X–Y in (a). (d) Modelled cross-section through X–Y in (b). Both cross-sections have 1:1 horizontal to vertical aspect ratios.

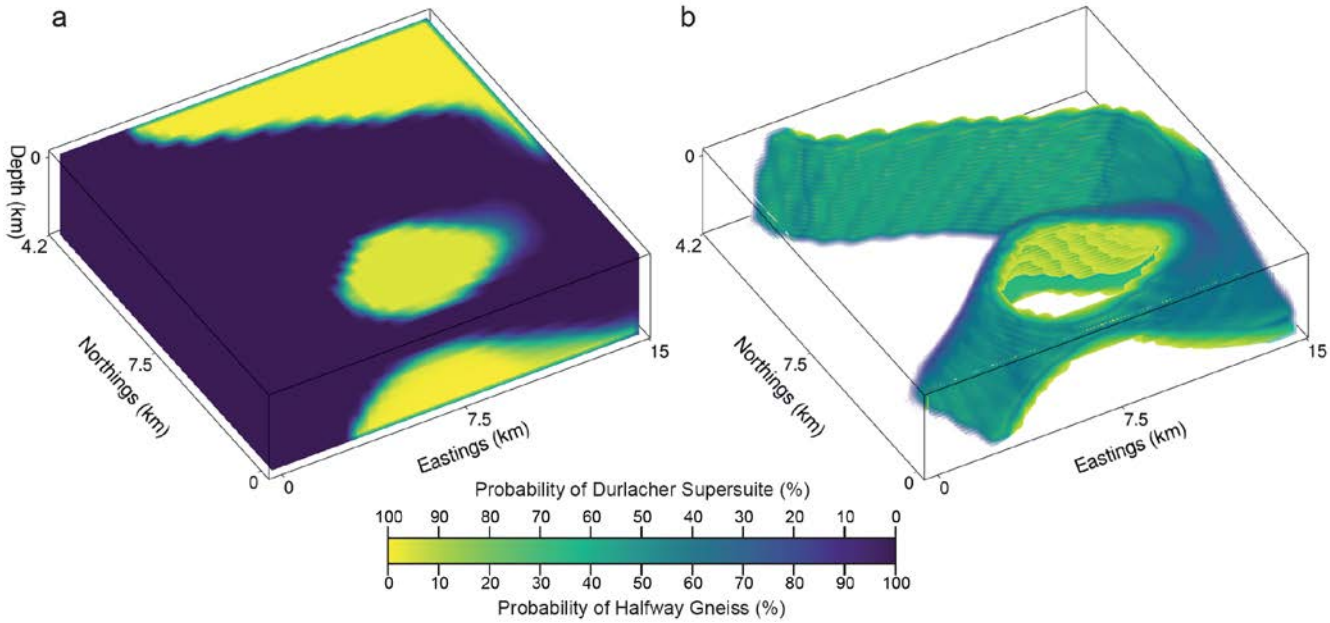

Fig. 11: Voxelized posterior distributions in 3D, looking towards the northeast. (a) Probability of Halfway Gneiss and Durlacher Supersuite. (b) Same as (a) but only showing regions between 5 and 95% probability (i.e., the interface between Durlacher Supersuite.

Formation, petrophysical property	Mean	–2σ	+2σ	Autocorrelation time	Effective n	R̂
Halfway Gneiss, Rock Density (mGal)	2.72	0.13	0.11	11.61	129.1967	1.02
Halfway Gneiss, Log Susceptibility	–3.65	0.53	0.57	14.26	105.2075	1.03
Durlacher Supersuite, Rock Density (mGal)	2.67	0.13	0.11	11.59	129.4099	1.02
Durlacher Supersuite, Log Susceptibility	–2.61	0.05	0.09	13.32	112.6397	1.02

Table 1: Rock property diagnostics for density and magnetic susceptibility for both the Halfway Gneiss and Durlacher Supersuite.

APPENDIX A – Derivation of likelihood functions

Our likelihood assumes that the potential-field data are distributed independently around each forward model as a Gaussian with some variance σ^2. In the case where σ^2 is known exactly a priori, this results in a Gaussian likelihood $P(y \mid \theta, \sigma^2)$, or in a log-likelihood with the familiar weighted-mean-square form. If σ^2 is not known exactly, we can capture any pre-existing expectations about its distribution in the form of a prior $P(\sigma^2)$. We then integrate $P(y \mid \theta, \sigma^2)\, P(\sigma^2)$ over σ^2 to derive the underlying likelihood $P(y \mid \theta)$ that accounts for our uncertainty about the noise level.

Choosing an inverse-gamma prior, $P(\sigma^2) = P(\sigma^2 \mid \alpha, \beta) = IG(\sigma^2 \mid \alpha, \beta)$, allows us to do this integral analytically:

$$P(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(y-\mu)^2}{2\sigma^2}} \qquad \text{- Eqn. 1}$$

$$P(\sigma^2|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, \sigma^{-2(\alpha+1)} e^{\frac{-\beta}{\sigma^2}} \qquad \text{- Eqn. 2}$$

$$P(y|\mu,\alpha,\beta) = \int P(y|\mu,\sigma^2) P(\sigma^2|\alpha,\beta)\, d\sigma^2 \qquad \text{- Eqn. 3.1}$$

$$= \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-\mu)^2}{2\sigma^2}} \times \frac{\beta^\alpha}{\Gamma(\alpha)}\, \sigma^{-2(\alpha+1)} e^{\frac{-\beta}{\sigma^2}} d\sigma^2 \qquad \text{- Eqn. 3.2}$$

$$\nwarrow u = \frac{1}{\sigma^2},\ \sigma^2 = \frac{1}{u},\ d\sigma^2 = \frac{1}{u}\, du \qquad \text{- Eqn. 3.2.1}$$

$$= \frac{1}{\sqrt{2\pi}} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \int u^{\frac{1}{2}+(\alpha+1)-2} e^{-\left[\frac{1}{2}(y-\mu)^2+\beta\right]u}\, du \qquad \text{- Eqn 3.3}$$

$$\nwarrow v = \left[\frac{1}{2}(y-\mu)^2+\beta\right]u,\ du = \left[\frac{1}{2}(y-\mu)^2+\beta\right]^{-1} dv \qquad \text{- Eqn. 3.3.1}$$

$$= \frac{1}{\sqrt{2\pi}} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \times \left[\frac{(y-\mu)^2}{2}+\beta\right]^{-\left(\alpha-\frac{1}{2}\right)-1} \int v^{\alpha-\frac{1}{2}-v} e^{-v}\, dv \qquad \text{- Eqn. 3.4}$$

$$\nwarrow \Gamma(\alpha) = \int v^{\alpha+1} e^{-v}\, dv \qquad \text{- Eqn. 3.4.1}$$

$$= \frac{1}{\sqrt{2\pi}} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \times \left[\frac{(y-\mu)^2}{2}+\beta\right]^{-\left(\alpha-\frac{1}{2}\right)-1} \Gamma\left(\alpha+\frac{1}{2}\right) \qquad \text{- Eqn. 3.5}$$

$$= \frac{1}{\sqrt{2\pi}} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \times \beta^{-\left(\alpha-\frac{1}{2}\right)} \times \left[\frac{(y-\mu)^2}{2\beta}+1\right]^{-\left(\alpha+\frac{1}{2}\right)} \Gamma\left(\alpha+\frac{1}{2}\right) \qquad \text{- Eqn. 3.6}$$

$$\nwarrow \alpha = \frac{\nu}{2} \qquad \text{- Eqn. 3.6.1}$$

$$= \frac{1}{\sqrt{2\pi\beta}} \times \frac{\Gamma\frac{\nu+1}{2}}{\Gamma\frac{\nu}{2}} \times \left[\frac{(y-\mu)^2}{2\beta}+1\right]^{-\left(\frac{\nu+1}{2}\right)} \qquad \text{- Eqn. 3.7}$$

$$\nwarrow \xi = \sqrt{\frac{\nu}{2\beta}}(y-\mu) \qquad \text{- Eqn. 3.7.1}$$

$$= \sqrt{\frac{\nu}{2\beta}} \times \frac{1}{\sqrt{\pi\nu}} \times \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \times \left[\frac{\xi^2}{\nu}+1\right]^{-\left(\frac{\nu+1}{2}\right)} \qquad \text{- Eqn. 3.8}$$

The resulting probability density is thus a t-distribution in the transformed variable ξ with $\nu = 2\alpha$ degrees of freedom and scale $(\beta/\alpha)^{1/2}$, with the last factor of $(\beta/\alpha)^{-1/2}$ introduced in the variable change from y to ξ.

The parameters α and β can then be elicited from experts to produce a distribution that captures reasonable expectations about noise. For the inverse-gamma prior, α describes the thickness of the tail towards large variances, while β/α gives a scale for the mean variance in units of the sample variance of the data.

Similarly, the likelihood for the lithostratigraphic data is assumed to be a binomial distribution for k correct observations out of n total observations, with some probability p that each underlying observation is correct. In practice p is not known exactly, but as with the potential-field data, a prior over p can be specified and integrated out. Choosing a beta-function prior, P(p) = B(p | α, β), allows this integral to be done analytically:

$$P(k|n,\rho) = \frac{n!}{k!(n-k)!}\,\rho^k(1-\rho)^{1-k} \qquad \text{- Eqn. 4}$$

$$P(k|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\,\rho^{\alpha-1}(1-\rho)^{\beta-1} \qquad \text{- Eqn. 5}$$

$$\nwarrow B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\rho-1}dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \text{- Eqn. 5.1}$$

$$P(k|n,\alpha,\beta) = \int_0^1 P(k|n,\rho)P(\rho|\alpha,\beta)dp \qquad \text{- Eqn. 6.1}$$

$$= \int_0^1 \frac{n!}{k!(n-k)!} \times \frac{1}{B(\alpha,\beta)}\rho^{k+\alpha-1}(1-\rho)^{n-k+\beta-1}dp \qquad \text{- Eqn. 6.2}$$

$$= \frac{n!}{k!(n-k)!} \times \frac{1}{B(\alpha,\beta)}\int_0^1 \rho^{k+\alpha-1}(1-\rho)^{n-k+\beta-1}dp \qquad \text{- Eqn. 6.3}$$

$$= \frac{n!}{k!(n-k)!} \times \frac{1}{B(\alpha,\beta)}B(k+\alpha, n-k+\beta) \qquad \text{- Eqn. 6.4}$$

$$= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(k+\alpha)\Gamma(n-k+\beta)}{\Gamma(n+\alpha+\beta)} \qquad \text{- Eqn. 6.5}$$

where now α and β have different interpretations: α can be interpreted as the number of correct answers, and β as the number of incorrect answers, in a hypothetical training dataset used to compose a prior before any of the actual current training data are considered.