# Tutorial: tomographic Bayesian uncertainty estimation

Ian F. Jones[1*], Rodrigo Felicio[1], Angeliki Vlassopoulou[1] and Claudia Hagen[1]

**Abstract**

Whether we build a subsurface parameter model or deliver a subsurface image, our industry has been sadly lacking in attempting to assign 'error bars' to any of the products created. It transpires that this is an extremely difficult task to undertake in a quantitative manner. Assuming that we have an acceptable migration algorithm that honours the physics of the problem, then there are certain minimum acceptance criteria, which tell us that at least the derived model explains the observed data to within some acceptance threshold. Namely: image gathers that are 'flat' following migration with the obtained model, and which also match available well data — but these criteria do not tell us how accurate or precise the model or image is. Bayesian analysis of tomographic model error offers one approach to quantifying image positioning uncertainty, and here we give an overview of the elements involved in this procedure.

## Introduction

Given that we can never obtain a 'correct' model based on measured data (e.g. Jackson 1972), we should ideally assess how suitable the derived approximate model or resultant image is. However, putting error bars on images turns-out to be an extremely difficult task to undertake in a quantitative manner. Many workers have attempted this over the years, and there are two broad approaches to accomplishing this task.

First, we can assess the measured residual moveout in the final migrated image gathers, after a comprehensive model-building exercise, which usually would involve several iterations of ray or waveform tomography. Then, using an estimate of the inherent measurement uncertainty in the residual moveout (RMO) measurements (e.g., Ashton et al., 1994; Al-Chalabi, 1997; Chen and Schuster, 1999; Tom Armstrong, personal communication, 2008; Jones, 2010), many dozens of slight perturbations are introduced into the RMO values, with a distribution bounded by the estimated RMO uncertainty (and the distribution of uncertainty might be further bounded by picked horizons and/or rock physics constraints). For each of these slight RMO perturbations, a tomographic inversion is run so as to update the velocity model. This yields an ensemble of possible realizations of the velocity model, each of which is consistent with the observed data, to within the uncertainties associated with our measured RMO. Then for a specified target horizon, map demigration is performed (just once), followed by successive map migrations for each of these perturbed, equally valid representative velocity models, giving rise to a spatial distribution of possible positions for each specified picked reflector. Map migration is an approximate migration technique whereby a time horizon is picked from, say, a stacked section, and then, in conjunction with an interval velocity field, Snell's law is used to reposition this horizon to its equivalent depth location. This in effect delivers a low-cost emulation of a full depth-migration. Map demigration is the reverse procedure, converting a depth horizon picked from a depth-migrated image to the corresponding (perhaps multivalued) time horizon positions. Such methods of assessing structural uncertainty have been described by Cognot et al. (1995); Thore and Hass (1996); Thore and Juliard, (1999); and Thore et al. (2002). Rather than using map migrations, we could also run many full-volume migrations (e.g. Bell el al., 2017) which would be more costly. However, in order to assess positioning uncertainty using these many migrated volumes, we would still need to pick specific image elements in order to statistically quantify their volumetric displacement distribution.

The second approach is more esoteric, involving the mathematics of inversion used within the tomographic solvers, dealing with what is called the model-resolution matrix. This can be used to assess the uncertainty associated with each of the inverted parameter values, but it still needs to be used in conjunction with something such as a map migration to assess the effect of model variation on image position (e.g., Jackson, 1972; Menke, 1989; Berryman, 1997, 2001; Chiţu et al., 2008; Etgen, 2008; Osypov et al., 2008; 2011, 2013, Letki et al., 2013; Jones, 2010, 2018; Raffle et al., 2017; Schuster, 2017). Bachrach (2010) discusses similar methods for anisotropic parameter uncertainty estimation.

[1] ION Geophysical
[*] Corresponding author, E-mail: ian.jones@iongeo.com
DOI: 10.3997/1365-2397.2019022

## Intrinsic errors

Whenever we attempt to pick or compare moveout observed on a gather, either before or after migration, there will always be some intrinsic measurement error in what we observe. This can be assessed by considering the difference between moveout applied to a CMP gather with a correct velocity and moveout applied with a velocity slightly in error. For some particularly small error, we will no longer be able to notice the difference between the two moveout corrected gathers, due to the offset and band-limitation of the data.

If we perturb the NMO velocity for an event from $V_{rms}$ to ($V_{rms}$+$\Delta V_{rms}$), then the respective arrival times at offset x are:

$$T^2_{(x,V_{rms})} = T_0^2 + \frac{x^2}{V^2_{rms}} \quad (1)$$

and

$$T^2_{(x,V_{rms}+\Delta V_{rms})} = T_0^2 + \frac{x^2}{(V_{rms}+\Delta V_{rms})^2} \quad (2)$$

and assess the time difference on the far trace, $\Delta T_{nmo}$ resulting from this velocity change, then for $V_{rms}$>>$\Delta V_{rms}$ we obtain to 2nd order in $x$:

$$\Delta T_{nmo} = \frac{\Delta V_{rms}\, x^2}{T_0 V^3_{rms}} \quad (3)$$

Where:

$T_0$ is the zero offset arrival time of the moveout trajectory being analysed

$x$ is the maximum offset, for the event at a given time $T_0$

$V_{rms}$ is the approximation for stacking velocity

For a wavelet of time duration $\tau$, if we adopt a resolution criterion, say the thin bed Rayleigh approximation using $\tau$/4 as the discernible time shift at a maximum offset $x$, and given that $\tau \sim 1/f_d$ then we obtain the following expression for intrinsic RMS velocity error to 2nd order in $x$:

$$\Delta V_{rms} = \frac{T_0 V^3_{rms}}{4 f_d x^2} \quad (4)$$

where:

$f_d$ is the dominant frequency of the reflection event's wavelet under analysis,

and $\Delta V_{rms}$ is the velocity difference being resolved.

## Tomographic inversion

Once residual moveout has been picked (usually following migration), we can invert this information using a tomographic technique, so as to update the velocity model (Stork, 1992; Lo and Inderwiesen, 1994; Luo et al., 2014). For reflection travel-time tomography this procedure can be described as follows. Consider a simple subsurface model divided into nine cells each with its own constant velocity, for a source at location 'A' and a receiver at location 'C', for the ray path reflecting off the dipping surface at 'B' (Figure 1a). The arrival time equation for the ray-path for the third offset shown in Figure 1b comprises contributions from the small elemental travel paths $d_{ij}$, through each of the cells traversed (where the index $i$ pertains to the offset and the index $j$ pertains to the model cell), such that:

$$\textbf{Total time for this offset, } t_3 = d_{31}/v_1 + d_{34}/v_4 + d_{35}/v_5 + d_{38}/v_8 + d_{39}/v_9 + d_{36}/v_6 + d_{33}/v_3. \quad (5)$$

For a CMP gather, we have many arrival time measurements for a given subsurface reflector element, and each source-to-receiver travel-path can be decomposed into its elemental contributions from each cell (note that element $d_{38}$ in Figure 1a is 'v' shaped, as it includes the reflection point). In general, each raypath will not actually be a straight line, but may refract at each grid-cell boundary, and may also be curved. Travel times autopicked on the real data gathers will be compared to the ray-trace computed travel times at each iteration. To determine the velocity distribution along this ray path, tomography tries to solve a set of simultaneous equations, which can be accomplished as we often have more equations than unknowns (i.e., a given cell in Figure 1c is traversed by more than one raypath).

The expression for the travel time from Equation 5 can be written as:

$$\mathbf{t_i = \Sigma_j\, d_{ij}/v_j = \Sigma_j\, d_{ij}\, s_j} \quad (6)$$

where $t_i$ is the total travel time along the $i_{th}$ ray-path

$d_{ij}$ is the path length of the $i_{th}$ ray in the $j_{th}$ cell

$v_j$ is the velocity in the $j_{th}$ cell

$s_j$ is the slowness in the $j_{th}$ cell

Or, in matrix notation: **T = DS.** This is shown as a matrix cartoon in Figure 2.

Contrary to what we have shown in Figure 2, for a realistic problem we would have many more travel-time measurements

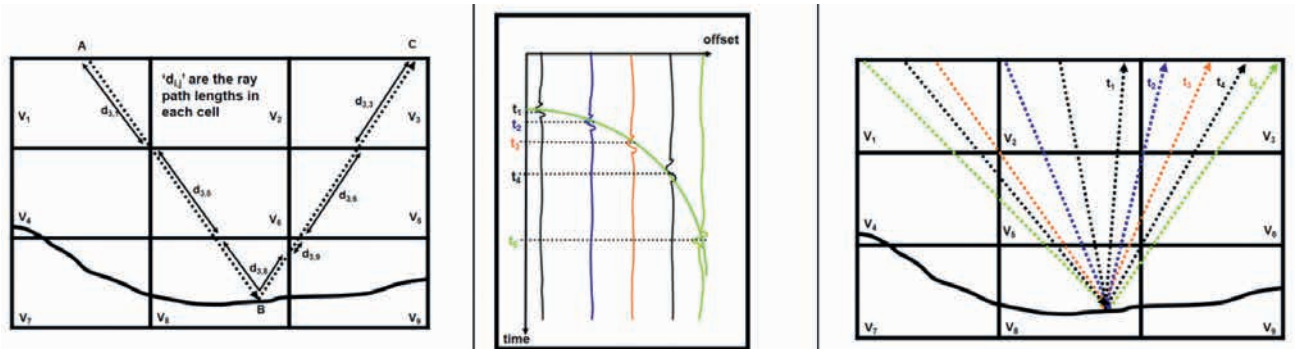


**Figure 1** a) A nine cell model with the raypath for a single source-receiver pair. b) A CMP gather showing moveout for a single reflector for five offsets. c) Raypaths corresponding to the offsets shown in b).

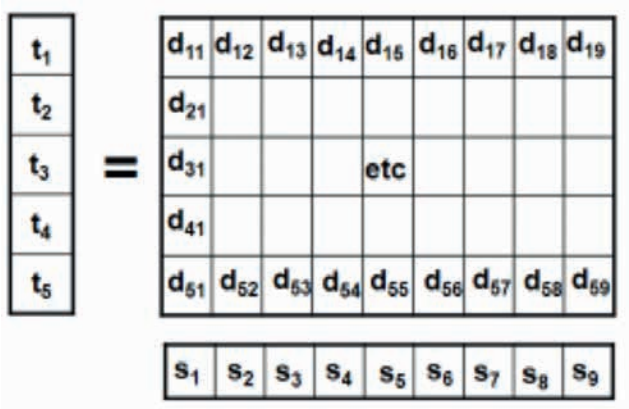


**Figure 2** Matrix cartoon. For model cells with no ray coverage, we just leave a zero in the **D** matrix for that value.

than model cells. For example, consider a survey of 1000km$^2$ characterized by a subsurface parameter model extending to a depth of 10 km, with tomography cell size 100m*100m*100m: in this case we have 10 million model cells. If this survey was covered by 100,000 shots, with 10 cables and 100 receivers per cable, we would have 100 million traces, and if we utilized observed data for 10 reflectors then we would have about one billion ray-equations. Although we would typically have many more ray-equations than model cells (an overdetermined matrix system), due to measurement errors and the fact that certain cells may not have been crossed by any ray (an unconstrained matrix system), the solution is highly non-unique, and we have to solve the simultaneous equations approximately, perhaps with a least-squares approach, and perhaps introducing constraints that represent geologically plausible solutions. Measurement errors can arise, for example, from ambiguity in travel time picking due to errors in wavelet phase or approximated spatial receiver location resulting from binning.

We want to estimate the value of velocity in each cell of our model: to accomplish this we must solve a set of simultaneous equations. We use an auto-picker to measure the observed travel time for an event at a given offset, and we obtain estimates of the distance a ray travelled ($d_{ij}$) in each cell by using a ray-tracing algorithm. Now we want to invert the matrix equation to yield **S**:

$$\mathbf{T} = \mathbf{DS} \tag{7}$$

$$\mathbf{D^{-1}T} = \mathbf{D^{-1}DS} = \mathbf{S} \tag{8}$$

Unfortunately, in most cases, the matrix **D** is not invertible, because it is not square (i.e., in general the number of travel time measurements is not the same as the number of velocity cells in the model) and there is no exact solution due to measurement error and other approximations. So, instead, we have to derive an approximate solution that best fits the observed inconsistent data according to criteria of our choosing (e.g. Jackson 1972).

For example, to obtain the best solution using the least-squares criteria, one pre-multiplies both sides of the equation **T = DS** by the transpose of **D**, $\mathbf{D}^T$, to form the symmetric and invertible square covariance matrix $\mathbf{D}^T\mathbf{D}$, and then the least-squares solution is given by :

$$\mathbf{S} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{T} \tag{9}$$

The quantity $\mathbf{D}^T\mathbf{D}$ is referred to as the Hessian matrix.

Alternatively, we could rearrange equation 7 and consider the vector quantity **(T-DS)** — the residual modelling error**,** which we ideally want to be zero, and then recast the problem as a minimization of a 'cost' or 'objective' function, $\mathbf{F(S) = (T-DS)^T (T-DS)}$ , i.e., the sum of the squares of the residual error quantity, over all time samples, for all traces, in all shots.

Unfortunately, there will always be many slightly different models that could yield forward modelling that matches the observed field data equally well, and also, the inversion may converge on a local minimum (e.g. Jones 2018) which is not truly representative of the real earth structure. Such non-uniqueness in the solution is a direct consequence of the all the corners that we have to cut to solve our problem, such as the linear approximation described in Appendix 1. Other important sources of uncertainty include, for instance, the noise in the input data, errors in the matrix **D**, and limited or no illumination of portions of the discretized velocity model, due to the survey geometry. Hence, one important conclusion to draw from inverse theory and practice is that there is never a correct answer, and never can be: at best we can only obtain models that adequately explain the observed data. This uncertainty is a manifestation of the principle of non-uniqueness in inverse theory.

*So if you ask 'is the model right' then you're asking the wrong question!*

## The structure of the D matrix, and estimating the data and model covariance

Note that the **D** matrix has dimensions of: [n_ measurements X n_model-cells] = [n_rays X n_parameters], and a single column vector in **D** represents the small path length segments of all the rays that pass through a single model cell for a given reflector (Figure 3).

Additionally, we can expand the analysis to also assess behaviour of the measurements themselves, by assessing the (co)

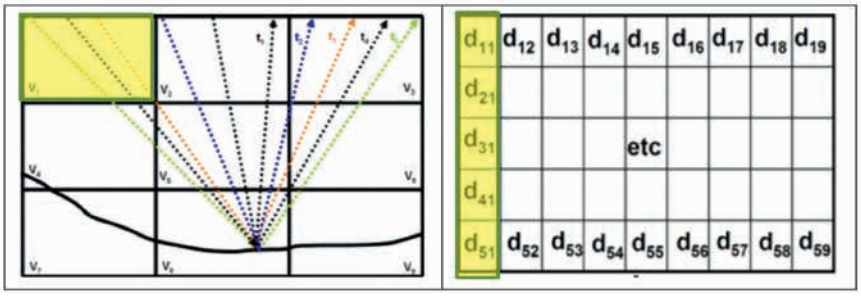


**Figure 3** Ray paths and elements of the D matrix associated with all the small ray-path segments associated with a single model cell.

variance matrix of the input data (under the assumption that they have zero-mean value), $\mathbf{TT}^{\mathbf{T}} = \mathbf{C}_d$. If we assume the data errors (e.g., the errors in the picks made by the auto-picker) are independent and identically distributed about the true value, then the data covariance matrix is the data variance $\sigma^2$ multiplied by the identity matrix, which will have dimensions [n_rays X n_rays], (recall that the identity matrix is a square matrix with ones on the diagonal and zeroes elsewhere). In practice this is never really the case as the autopicker is typically constrained to follow the same phase across offset and CRP location, and any phase error remaining after pre-processing and migration will give rise to small, but non-zero, off-diagonal terms in $\mathbf{C}_d$.

These measurements come from the residual error estimated from the RMO picking made on the CRP gathers: each raypath has an error. For a given reflector event, we can either give all rays in the CRP a related error as from parametric picking or allocate separate errors as from non-parametric picking. The data errors will propagate throughout the inversion process leading to estimated model parameters with model covariance $\mathbf{C}_m$

Remember that a covariance measures how two variables correlate – in this case it tells us how much uncertainty is in the inverted model given the uncertainty in the observed data. Recall that the travel time data (**T**), is related to the model parameters (**S**), via:

$$\mathbf{T} = \mathbf{DS}, \text{ hence simplistically: } \mathbf{S} = \mathbf{D}^{-1}\mathbf{T} \qquad (10)$$

Thus model covariance matrix $\mathbf{SS}^{\mathrm{T}}$ is given by:

$$\mathbf{C}_m = \mathbf{D}^{-1}\,\mathbf{C}_d\,(\mathbf{D}^{-1})^{\mathrm{T}} = (\mathbf{D}^{\mathrm{T}}\,\mathbf{C}_d{}^{-1}\,\mathbf{D})^{-1} = \sigma_d{}^2(\mathbf{D}^{\mathrm{T}}\,\mathbf{D})^{-1} \qquad (11)$$

Where $\sigma_d^2$ is the variance of the input data.

Although here we have shown the derivation of $\mathbf{C}_m$ derived from the overly simplistic expression in equation (10), it can also be derived from the more useful expression in equation (9).

Here we have assumed that the measured data errors are independent of each other, hence their covariance matrix ($\mathbf{TT}^{\mathrm{T}} = \mathbf{C}_d$) has non-zero terms only along its diagonal, and thus we can replace the data covariance matrix $\mathbf{C}_d$ with the individual data variances, as expressed in $\sigma_d$ (these individual data variances could also be identical, but do not have to be).

The model covariance matrix $\mathbf{C}_m$ has dimensions n_parameters x n_parameters and is independent of the model values themselves. It is a function of the matrix **D** and the data errors. It is also worth noting that it is proportional to the inverse of the Hessian ($\mathbf{D}^{\mathrm{T}}\,\mathbf{D}$).

Each individual **row** of the **D** matrix refers to a **ray**, and its contributions to all possible cells in the model. We have '*fold*' traces in a CRP and '*nr*' ray paths in total. Each **column** pertains to a model **cell**, and the small path length segments of all the rays contributing to it (Figure 4). We have '*np*' cells in total.

In contemporary industrial implementation, we perform ray tomography in the depth-migrated domain, measuring residual moveout on CRP gathers. However, the actual inverse problem for such depth-domain tomography is usually solved by ray tracing back from the subsurface image point where this error was picked so as to compute the corresponding travel time back to the surface source and receiver locations, also keeping track of the associated elemental travel path lengths $d_{ij}$ through the gridded parameter model. Hence the descriptions offered here are still those actually used.

## Prior and posterior error distribution

We now need to describe the distribution of possible model errors and relate this to our estimates of initial error (the 'prior' estimates) so as to derive some final errors (the 'posterior' estimates).

So far we have espoused a deterministic point of view and concentrated on finding *one best possible solution*. By using model regularization and introducing weights to the residuals (as described in Appendix 1) we reduce the range of acceptable solutions. However, we have not yet made any inroads into assessing how tomography redistributed and reduced the initial error estimates.

The Bayesian approach, on the other hand, provides us with a way of tackling both issues with a refreshing change of attitude: instead of fighting uncertainty and looking for a single 'best-fit' model, Bayes' theorem lets us embrace it (Bayes, 1763; Duijndam 1988a, 1988b). Now we take **S** and **T** to be random variables and we use Bayes's theorem to estimate how the probability distribution of **S**, $P(\mathbf{S})$, changes given the information contained on **T** and its probability distribution, $P(\mathbf{T})$.

Central to this estimation are the concepts of joint distribution of two random variables and of the conditional probabilities between them. The *joint distribution*, $P(\mathbf{S} \cap \mathbf{T})$ provides the chance of observing simultaneous realizations of both variables. If the two random variables were independent of each other, then

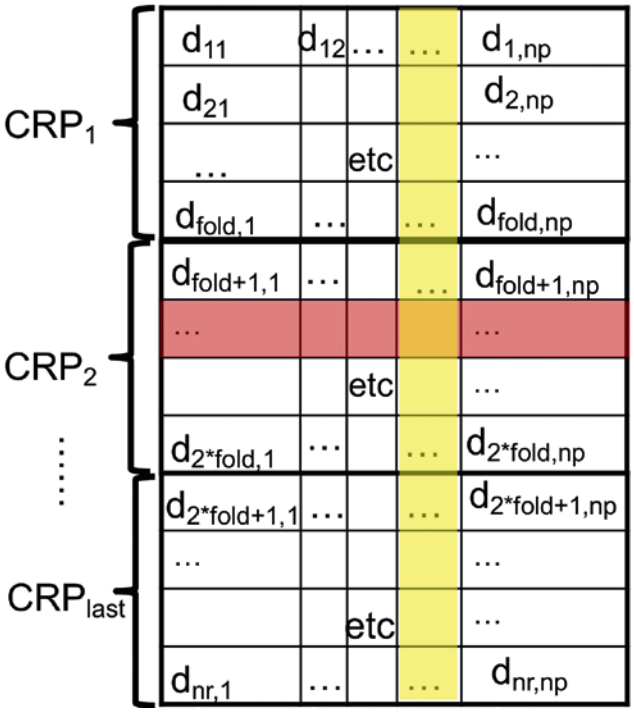


**Figure 4** Structure of the **D** matrix: each **row** of the **D** matrix refers to a **ray** (shaded in pink), and its contributions to all possible cells in the model. There are '*fold*' traces in a CRP and '*nr*' ray paths in total. Each **column** pertains to a model **cell** (shaded in yellow), and the small path length segments of all the rays contributing to it.

this chance is simply the product of each of the variables' individual occurrence probabilities, $P(\mathbf{S}\cap\mathbf{T}) = P(\mathbf{S})\,P(\mathbf{T})$. However, when the variables are not independent, the joint distribution needs to be generalized to incorporate the precedent that we either observe **T** with knowledge of **S** or vice-versa. Bayes's theorem guarantees that the joint distribution is symmetric in relation to both situations, that is:

$$\boldsymbol{P}(\mathbf{T})\,\boldsymbol{P}(\mathbf{S}|\mathbf{T}) = \boldsymbol{P}(\mathbf{S}\cap\mathbf{T}) = \boldsymbol{P}(\mathbf{S})\,\boldsymbol{P}(\mathbf{T}|\mathbf{S}), \tag{12}$$

where $P(\mathbf{T}|\mathbf{S})$ is the *conditional* probability of observing **T** given **S**, whereas $P(\mathbf{S}|\mathbf{T})$ represents the conditional probability of the reverse case. Bayes's Theorem, then, is the key to assess how errors in the input data, and/or in the model description and/or in the physics used to predict travel times are mapped by tomography into our final estimations of slowness. Indeed, by exploiting the equalities shown on Equation 12, we note that:

$$\boldsymbol{P}(\mathbf{S}|\mathbf{T}) = \frac{\boldsymbol{P}(\mathbf{S})\,\boldsymbol{P}(\mathbf{T}|\mathbf{S})}{\boldsymbol{P}(\mathbf{T})}. \tag{13}$$

Hence, the updated (the *posterior*) distribution of our model parameters, $P(\mathbf{S}|\mathbf{T})$, results from the modification of a prior model distribution $P(\mathbf{S})$ by the observed input data via the conditional probability $P(\mathbf{T}|\mathbf{S})$, with $P(\mathbf{T})$ acting as a normalization factor.

Equation 13 can also be viewed as a new cost function at the core of the Bayesian inference (inversion) problem — used to estimate the average model of the posterior distribution — in which the realization $\mathbf{S}_{map}$ connected to the maximum of $P(\mathbf{S}|\mathbf{T})$ is be obtained by maximizing $P(\mathbf{S})\,P(\mathbf{T}|\mathbf{S})$.

Appendix 2 outlines how we may estimate the posterior model covariance matrix $\mathbf{C}_{\mathbf{m\text{-}post}}$, based on the tomographic inverse problem derived from Equation 13, and additionally, how to construct a suite of equi-probable parameter models, **m**, each of which is consistent with the observed data, and satisfies the tomographic solution:

$$\mathbf{m} = \mathbf{m}_{map} + \mathbf{C}^{1/2}_{m\text{-}post}\ \mathbf{r}$$

$$\mathbf{m} = \mathbf{m}_{map} + \mathbf{C}^{1/2}_{m\text{-}prior}\ \mathbf{K}\,\mathbf{r} \tag{14}$$

This is the expression (Equation A2.12) derived in Appendix 2, where

- $\mathbf{m}_{map}$ represents an average model solution,
- $\mathbf{C}_{m\text{-}prior}$ is the prior model covariance
- $\mathbf{C}_{m\text{-}post}$ is the final model covariance after tomographic update
- **r** is a random vector drawn from the Normal distribution
- **K** is a proportionality matrix relating $\mathbf{C}_{\text{m-prior}}$ to $\mathbf{C}_{\text{m-post}}$.

$\mathbf{C}_{m\text{-}prior}$ plays a key role in finding the estimated posterior model distribution, and incorporates influences from the range of model perturbations, the spatial correlations of perturbations, any theoretical/empirical relationships between different types of parameters (e.g. velocity and anisotropic parameters δ, ε), accounts for different types of input used in the model building process (e.g., well logs), and acts as the pre-conditioner for the approximate solution of the Hessian.

Equation 14 relates the known (measurable) initial model covariance $\mathbf{C}_{m\text{-}prior}$, and quantities derived from the tomographic inversion that updated this model, to the final post-tomography model error $\mathbf{C}_{m\text{-}post}$. We then proceed to form a large suite of model perturbations which are used to perform many map migrations, so as to build an ensemble of equally possible images, in order to estimate the associated standard deviation and image-position error bars corresponding to this suite of solutions. Examples of this technique were first outlined by Osypov et al. (2013) and can also be found in Raffle et al. (2017) and Vlassopoulou (2017).

It should be noted that the workflow described above pertains to the final model parameters as they linearly relate to the data residuals. If the final tomography model happens to be trapped at a local minimum, which is occasionally undoubtedly the case, then there is no guarantee that the final model bounded by the posteriori model covariance will incorporate the true model.

## Workflow summary

The preceding steps in preparing a suite of velocity models can be summarised in the flow chart in Table 1. The final production velocity model, associated depth image and gathers are the starting point. The remnant velocity error, as manifested in residual moveout on the CRP gathers, is picked for use in the tomography. The tomographic inversion algorithm is then re-run, but this time not to update the velocity model, but rather to output the various matrices associated with the remnant moveout error and final model.

## Case study example

The example shown here comes from the Ivory discovery (Raffle et al., 2017), which lies within the Nyk High, located in the north-eastern part of the Vøring Basin in Norwegian waters. The main challenge in mapping the extent of the discovery has been seismic imaging at the crest of structures bound by major faults (e.g. fault shadow effects), together with depth conversion uncertainty and a poor well-to-seismic tie. After a comprehensive reprocessing and imaging project, structural uncertainty was then estimated using the Bayesian statistical analysis of the tomographic resolution matrices in conjunction with prior uncertainty estimates, as described earlier.

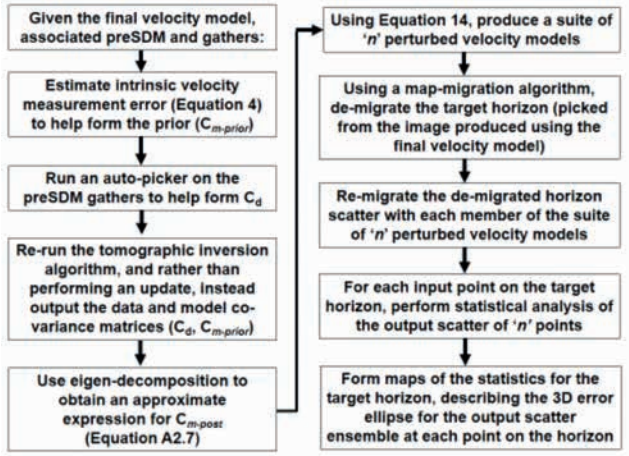


**Table 1** Flowchart outlining the overall workflow for uncertainty estimation.

In this case, we obtained the model prior to error distribution, $\mathbf{C}_{m\text{-}prior}$, by scaling the final velocity model with a lateral average of the intrinsic error derived from Equation 4. It is noted that for Equation 13, the statistical distribution of observed data, $P(\mathbf{T})$, acts as a normalization factor. This can be determined post-facto by calibrating the final result of the statistical analysis against a well database, so as to ensure that the relative distribution of image errors, obtained from the statistics derived from Equation 14, resembles the actual observed well mis-ties.

Figure 5 shows a typical inline section, with the final tomographic interval velocity model superimposed, and Figure 6 shows a subset of the 200 interval velocity model perturbations obtained via the decomposition and random selection described in Equations 14 and A2.5. These 200 model realizations are then used to compute the standard deviation of this distribution (Figure 7), while Figure 8 shows the prior error distribution, the posterior error distribution (following tomographic update), and the percentage change in the standard deviation (an overall reduction in the error): indicating that tomographic inversion results in a significant reduction in error in the shallower section as compared to the deeper section. If we had a very large number of perturbations, rather than just the 200 used here, then the standard deviation of Figure 7 would be the

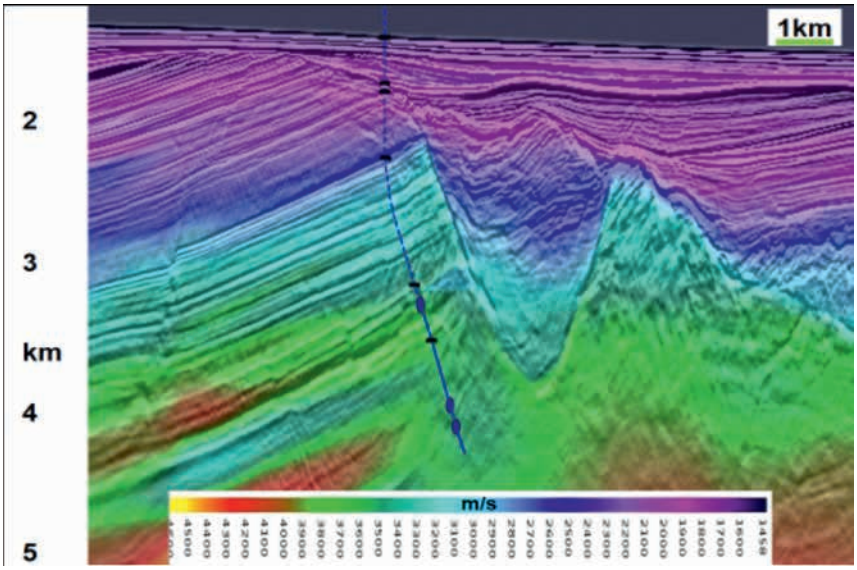


**Figure 5** Typical inline structure of data in the study area with the final tomographic interval velocity model superimposed. Line length displayed ~15 km.

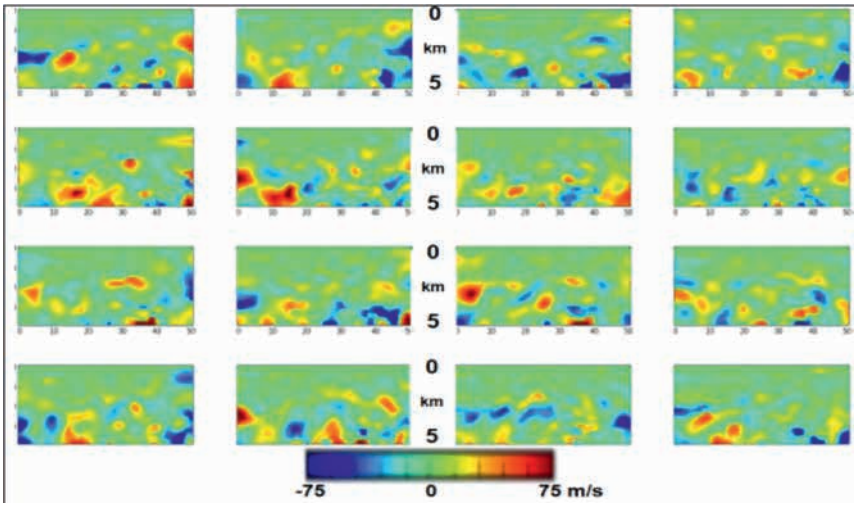


**Figure 6** A few examples of the 200 model realization perturbations for a single inline vertical section (corresponding to the inline shown in Figure 5). The colour bar indicates the velocity variations that could still explain the data to within noise.

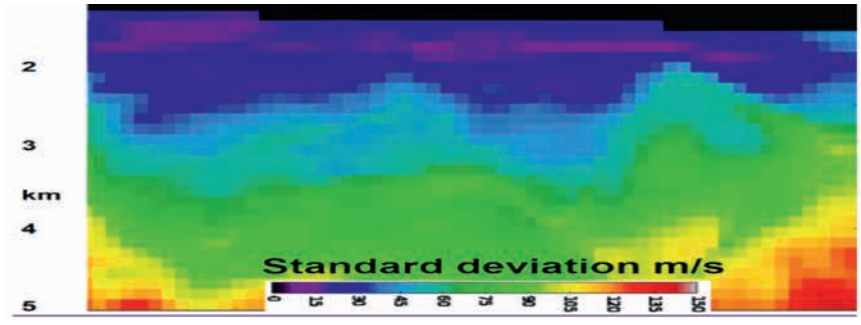


**Figure 7** Standard deviation from 200 model realizations, as exemplified by the subset shown in Figure 6.

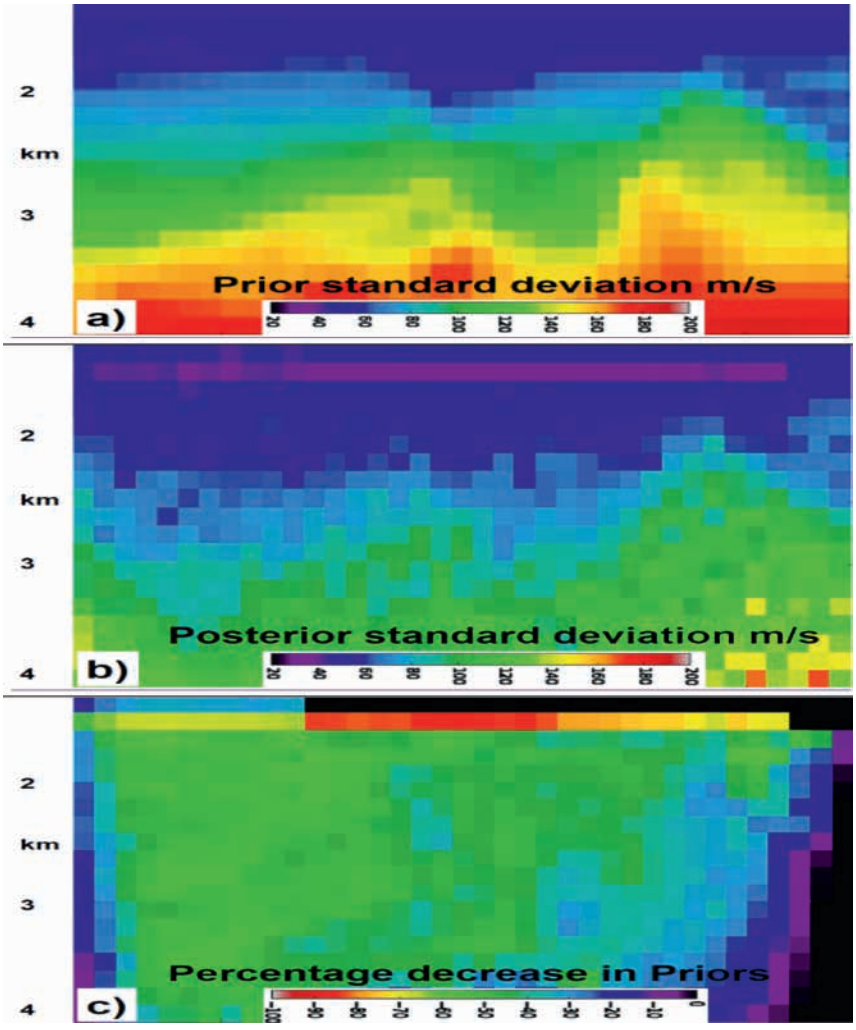


**Figure 8** a) Prior model standard deviation, b) post-tomography final posterior model standard deviation, c) percentage change in the standard deviation ($\sqrt{[C_{m\text{-}post}/C_{m\text{-}prior}]}$-1).

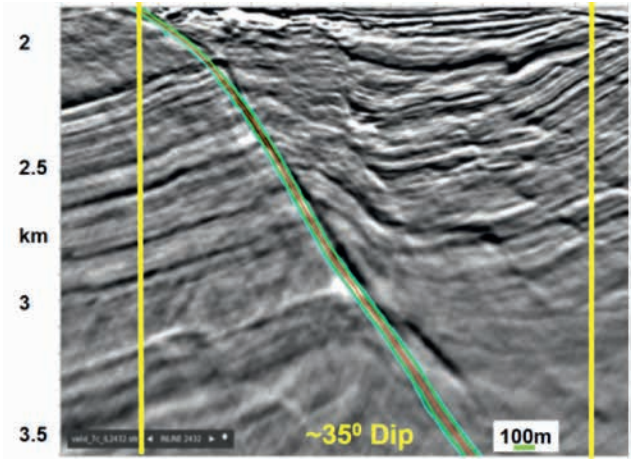


**Figure 9** PreSDM image showing fault plane and uncertainty corridors based on 200 model realizations. Red: fault average position, Green: 2*Sigma confidence interval, Blue: outer bounds. The yellow vertical line indicates the region of the fault were the uncertainty results are considered reliable (with dip approximately 35⁰). Beyond these bounds, the fault imaging is poor or the ray-coverage is sparse (from Raffle et al., 2017, courtesy of Spirit Energy, Norge).

same as the posterior error distribution shown in Figure 8b (and also assuming that all other parameters were consistent).

Figure 9 shows the corridor of 200 model realisations of the main fault plane, indicating in red the mean fault location as well as the ±2 standard-deviation bounds for the fault — shown in green, and in blue the maximal bounds (i.e., the largest actual value of lateral shift for all 200 perturbations with respect to the input sample location). Figure 10 shows the 200-point scatter cloud for a single input point, and also a representation of the inline, crossline, and vertical depth errors for this fault plane, here represented by the simple depth, inline, and crossline thicknesses of the 200-point scatters. This information can also be plotted after an eigenvector rotation so as to follow the major and minor axes of the scatter ellipsoids.

## Discussion and conclusions

Putting error bars on images is not trivial: the mathematics is complicated and the costs of inverting the associated matrices is high. However, if we want to fulfil the requirements for prospect de-risking, then the errors associated with an interpretation based on a seismic image must be somehow quantified. The technique described here can offer one such route to relative error quantification.

Perhaps not surprisingly, if the 'final' model used as the basis for the uncertainty estimation is not very good, then the estimated errors will be larger. For example, Vlassopoulou (2017) and Vlassopoulou et al. (2019) show that the uncertainty estimated after the third iteration of tomographic model update is larger

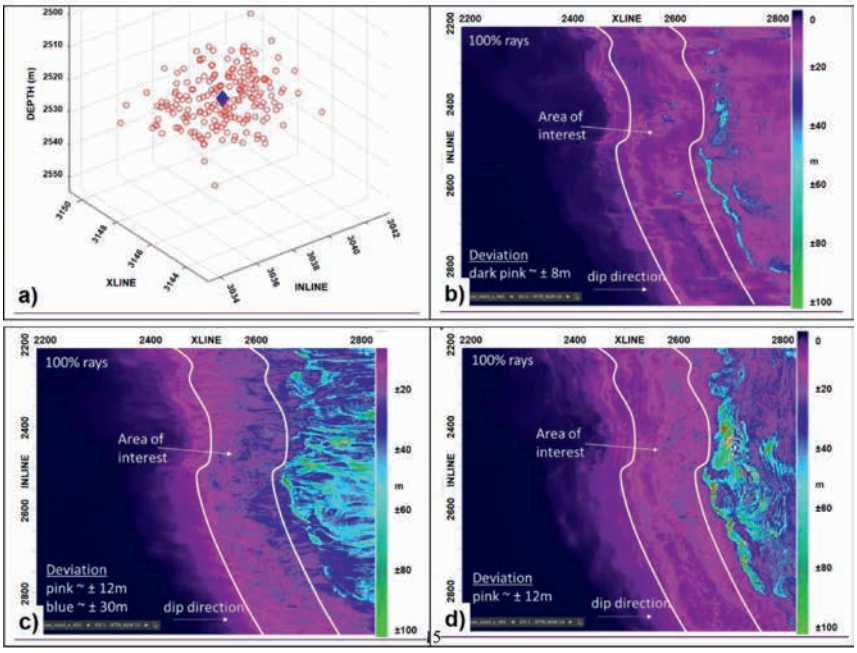


**Figure 10** a) 3D scatter of 200 realizations for a single input point on the fault plane. This behaviour can be characterized in terms of the inline, crossline, and vertical thicknesses of the scatter, expressed in terms of the initial location plus or minus a displacement. b) Vertical thickness, c) inline-thickness, d) crossline-thickness of the scatter-cloud. The wavy white lines indicate the regions of the fault plane where the results are considered reliable (as indicated on Figure 9 by the $35^0$ dip bounds). (From Vlassopoulou, 2017, courtesy of Spirit Energy, Norge).

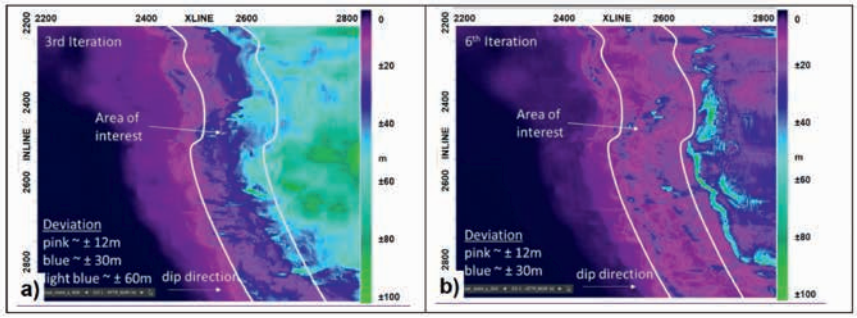


**Figure 11** a) Vertical (depth) scatter corridor thickness after only three iterations of tomography: here the scatter corridor thickness in the zone of interest approaches ±30 m. b) Scatter corridor thickness after the 6th and final iteration of tomographic model update is reduced to approximately ±12 m (from Vlassopoulou, 2017).

than that estimated after the 6th and final iteration of tomographic update (Figure 11).

Are these errors quantitatively absolute? Not really: it can be noted from Equation 14, that the final error distribution is proportional to the specified starting distribution, such that if we doubled the initial prior term $\mathbf{C}^{1/2}_{\text{m-prior}}$ then we would also double the associated velocity perturbation. However, what we do gain from this analysis is the relative spatial distribution of errors, and as long as we calibrate the results to some 'ground truth' measurements (e.g. observed well mis-ties), then the error distribution will be meaningful both qualitatively and quantitatively.

In the example shown here, we selected a relatively simple approach to estimating the prior model uncertainty, namely, looking at the intrinsic measurement errors associated with residual moveout picking, and using this as a scalar applied to the final tomographic velocity model. The posterior error was then calibrated against well measurements, and subsequently drilled wells fell within the predicted error bounds.

## Acknowledgements

We would like to thank Svein Idar Forsund and Mark Ackers Spirit Energy Norge, PL528 partners and ION for permission to publish these results, and especially to ION colleagues Zhijiang Luo for his work on the tomography aspects, and to J.J. Heyd, Doug Sassen and Ben Lasscock for their work on the decomposition code, and to Jacques Leveille, Nick Bernitsas, John Brittan, and Peter Rowbotham for their help in improving the paper.

## References

Al-Chalabi, M. [1997]. Parameter nonuniqueness in velocity versus depth functions. *Geophysics*, **62** (3), 970-979.

Ashton, C.P., Bacon, B., Deplante, C., Sinclair, T. and Redekop, G. [1994]. 3D seismic survey design. *Oilfield Review*, 19-32.

Bachrach, R., [2010]. Applications of deterministic and stochastic rock physics modelling to anisotropic velocity model building. *80th Annual International Meeting, SEG*, Expanded Abstracts, 2436-2440.

Bayes T. [1763]. An Essay towards solving a Problem in the Doctrine of Chances. *Phil. Trans.*, **53**, 370–418. doi:10.1098/rstl.1763.0053.

Bell, A.,. Russo, L., Martin, T. and van der Burg, D., [2017]. Model uncertainty analysis for de-risking seismic image accuracy. *79th EAGE Conference and Exhibition*, WS06.

Berryman, J.G. [1997]. Resolution for Lanczos and Paige-Saunders inverses in tomography, *SEP*, report 77, 1-185.

Berryman, J.G. [2001]. Computing tomographic resolution matrices using Arnoldi's iterative inversion algorith, *SEP*, report 82, 1-176.

Bui-Thanh, T., Ghattas, O., Martin, J. and Stadler, G. [2013]. A computational framework for infinite-dimensional Bayesian inverse problems, part I: the linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, **35**(6), A2494-A2523.

Chen, J. and Schuster, G.T. [1999]. Resolution limits of migrated images. *Geophysics*, **64** (4), 1046-1053.

Chiţu, D.A., Al-Ali, M.N. and Verschuur, D.J. [2008]. Assessing estimated velocity-depth models: Finding error bars in tomographic inversion. *Geophysics,* **73** (5), supp. VE223-VE233.

Cognot, R., Thore, P. and Haas, A. [1995]. Seismic depth estimation. In: Harlan, W.S. (Ed.), *Spring Symp. Geophys. Soc. Tulsa*, SEG, 18–26.

Duijndam, A.J.W. [1988a]. Bayesian estimation in seismic inversion. Part I: Principles. *Geophysical Prospecting,* **36**, 878-898.

Duijndam, A., J.W. [1988b]. Bayesian estimation in seismic inversion. Part 11: Uncertainty analysis. *Geophysical Prospecting,* **36**, 899-918.

Etgen, J.T. [2008]. Applications of the model resolution matrix. *78th SEG Annual Meeting*, Expanded Abstracts, 3702-3703.

Jackson, D.D. [1972]. Interpretation of inaccurate, insufficient, and inconsistent data. *Geophysical Journal of the Royal Astronomical Society*, **28**, 97-109.

Jones, I.F. [2010]. *An introduction to velocity model building*. EAGE, The Netherlands.

Jones, I.F., [2018]. *Velocities, Imaging, and Waveform Inversion (The evolution of characterizing the Earth's subsurface)*. EET 13, EAGE, The Netherlands.

Letki, L.P. Ben-Hadj-Ali, H. Desegaulx, P. [2013]. Quantifying Uncertainty in Final Seismic Depth Image Using Structural Uncertainty Analysis - Case Study Offshore Nigeria. *75th EAGE Conference & Exhibition*, Extended Abstracts.

Lo, T.W. and Inderwiesen, P. [1994]. *Fundamentals of Seismic Tomography*. SEG, USA.

Luo Z., Brittan J., Fan D., Mecham, B., Farmer P. and Martin, G. [2014]. Imaging complexity in the earth. Case studies with optimized ray tomography. *The Leading Edge*, **33** (9), 1016-1022.

Menke, W. [1989]. Geophysical Data Analysis: Discrete Inverse Theory, Revised Edition. *International Geophysics Series*, **45**. Academic Press.

Osypov, K., Nichols, D., Woodward, M., Zdraveva O. and Yarman, C.E., [2008]. Uncertainty and resolution analysis for anisotropic tomography using iterative eigendecomposition, *78th SEG Annual Meeting*, Expanded Abstracts.

Osypov, K., O'Briain, M., Whitfield, P., Nichols, D., Douillard, A., Sexton, P., Jousselin, P. [2011]. Quantifying Structural Uncertainty in Anisotropic Model Building and Depth Imaging - Hild Case Study. *73rd EAGE Conference & Exhibition*, Extended Abstracts.

Osypov, K., Yang, Y., Fournier, A., Ivanova, N., Bachrach, R., Yarman, C.E., You, Y., Nichols, D., Woodward, M., [2013]. Model-uncertainty quantification in seismic tomography: method and Applications. *Geophysical Prospecting,* **61**, 1114–1134

Raffle, J., Earnshaw, T., Fruehn, J.K., Greenwood, S., Singh, J., Hagen, C., Jones, I.F., Felicio, R., Sassen, D., Luo, Z., Forsund, S.I., Ackers, M. and Aamodt, L., [2017]. Risk reduction on the Ivory prospect via geologically constrained non-parametric inversion and Bayesian uncertainty estimation on a fault-bounded reservoir. *79th EAGE Conference & Exhibition*, Extended Abstracts.

Schuster, G.T., [2017]. *Seismic Inversion. Investigations in Geophysics #20*. SEG, USA.

Stork, C. [1992]. Reflection tomography in the postmigrated domain. *Geophysics*, **57**, (5), 680–692.

Tarantola, A., [2005]. *Inverse Problem Theory and Model Parameter Estimation*. SIAM.

Thore, P. and Haas, A. [1996]. A practical formulation of migration error due to velocity uncertainties. *58th EAGE Conference & Exhibition*, Extended Abstracts, X016.

Thore, P. and Juliard, C. [1999]. Fresnel zone effect on seismic velocity resolution. *Geophysics*, **64** (2), 593–603.

Thore, P., Shtuka, A., Lecour, M., Ait-Ettajer, T. and Cognot, R. [2002]. Structural uncertainties: Determination, management, and applications. *Geophysics*, 67, (3), 840–852.

Vlassopoulou, A., Felicio, R., Hagen, C., Jones, I.F., Ackers, M.A., Schjelderup, S. [2019]. Bayesian uncertainty estimation in the presence of tomographic model error. *81st EAGE Conference & Exhibition*, Extended Abstracts

Vlassopoulou, A., [2017]. *Estimation of position uncertainty in subsurface images*. MSc thesis, University of Leeds, UK.

## Appendix 1: least-squares, weighting, and regularization

Recall from Equation 9 that we had:

$$\mathbf{S} = (\mathbf{D}^T\,\mathbf{D})^{-1}\,\mathbf{D}^T\,\mathbf{T} \qquad \text{(A1.1)}$$

The least-squares solution **S** in Equation A1.1 is in reality a *unique* and *exact* solution, but to an *approximated* inversion (optimization) problem. The hidden assumption here is **T=DS**, which approximates the true, but non-linear, relationship between observed data (travel-times) and model parameters (slowness). Thus, we are actually solving a linear inversion problem in which we seek the minimum of a function that provides us the sum of the squares of the differences between modelled and observed travel-times:

$$\mathbf{F(S)} = (\mathbf{T} - \mathbf{DS})^T\,(\mathbf{T}\text{-}\,\mathbf{DS}) \qquad \text{(A1.2)}$$

Where **F(S)** is referred to as the 'cost function' or 'objective function' and is the quantity to be minimized, on average, over all the measurements.

The cost function is then a quadratic function of model parameters **S**; from calculus the minimum (or maximum) of a quadratic function of a single variable is found by setting its first-derivative to zero and solving for the aforementioned variable. Likewise (and ignoring a factor of two, which cancels),

$$\partial/\partial\mathbf{S}\;[\mathbf{F(S)}] = -\mathbf{D}^T(\mathbf{T}\text{-}\mathbf{DS}) = \mathbf{0} \qquad \text{(A1.3)}$$

This is exactly what we have done when writing $\mathbf{D}^T(\mathbf{T}\text{-}\mathbf{DS}) = 0$ above and then finding **S** in Equation A1.1; the only difference is that $\mathbf{D}^T(\mathbf{T}\text{-}\mathbf{DS})$ represents a multi-dimensional generalization of the derivative, called the *gradient* of $F(\mathbf{S})$.

Starting with some initial guess of parameters $\mathbf{S}_0$, the minimum of the cost function **F(S)** will occur when we perturb $\mathbf{S}_0$ by **ΔS** so as to achieve a minimum:

$$\partial/\partial\mathbf{S}\;[\mathbf{F}(\mathbf{S}_0+\Delta\mathbf{S})] = -\mathbf{D}^T(\mathbf{T}\text{-}\mathbf{D}(\mathbf{S}_0+\Delta\mathbf{S})) = \mathbf{0} \qquad \text{(A1.4)}$$

This can then be solved for **ΔS** by expanding $\mathbf{F}(\mathbf{S}_0+\Delta\mathbf{S})$ as a Taylor series and truncating to first order so as to evaluate **ΔS**

using the derivative of the cost function evaluated with the starting model guess $\mathbf{S}_0$,

$$\partial/\partial \mathbf{S}\,[\mathbf{F}(\mathbf{S}_0+\Delta\mathbf{S})] = \partial\mathbf{F}(\mathbf{S}_0)/\partial\mathbf{S} + \partial^2\mathbf{F}(\mathbf{S}_0)/\partial\mathbf{S}^2\ \Delta\mathbf{S} = \mathbf{0}$$

$$= \mathbf{D}^T(\mathbf{T}-\mathbf{D}\mathbf{S}_0) + \mathbf{D}^T\mathbf{D}\,\Delta\mathbf{S} = \mathbf{0} \qquad \text{(A1.5)}$$

And rearranging to obtain $\Delta\mathbf{S}$ gives:

$$\Delta\mathbf{S} = -\,[\partial^2\mathbf{F}(\mathbf{S}_0)/\partial\mathbf{S}^2]^{-1}\,.\,\partial\mathbf{F}(\mathbf{S}_0)/\partial\mathbf{S}$$

$$= -\,[\mathbf{D}^T\,\mathbf{D}]^{-1}\,\mathbf{D}^T(\mathbf{T}-\mathbf{D}\mathbf{S}_0) \qquad \text{(A1.6)}$$

In the expressions for $\Delta\mathbf{S}$ above, the term in the inverse bracket is referred to as the Hessian and the second term as the gradient.

The solution of Equation A1.1, $\mathbf{S} = (\mathbf{D}^T\,\mathbf{D})^{-1}\,\mathbf{D}^T\,\mathbf{T}$ works for overdetermined systems of equations, i.e., when there are more rays (equations) than model cells (unknowns). On other occasions, (as seen in Figure 1b, where some cells have no ray coverage) the converse may happen (there are more unknowns than equations, an underdetermined system) and thus one would like to get **S** such that it represents the simplest solution in some sense. This might be, for example, the solution with smallest norm value of $\mathbf{S}^T\mathbf{S}$ (i.e., a model with small covariance, such that we avoid rapid variation between model cells), while still obeying **T=DS**. Using a weighting vector, referred to as a Lagrangian multiplier $\boldsymbol{\gamma}$, one may write the corresponding cost function as:

$$\mathbf{F}(\mathbf{S}) = \mathbf{S}^T\mathbf{S} + \boldsymbol{\gamma}^T(\mathbf{T}-\mathbf{D}\mathbf{S}) \qquad \text{(A1.7)}$$

By taking the gradient of the cost function **F** (with respect to the variables S and $\boldsymbol{\gamma}$) and setting it to zero and then solving the resulting expression for $\boldsymbol{\gamma}$, one finds the minimum-length solution to be (Menke, 1989):

$$\mathbf{S} = \mathbf{D}^T\left(\mathbf{D}\mathbf{D}^T\right)^{-1}\mathbf{T} \qquad \text{(A1.8)}$$

In general, the tomography problem is of a mixed nature rather than being either purely under- or overdetermined, because there are model cells that are traversed by many rays (they are overdetermined), while others are not probed at all (underdetermined). The most common approach is to combine these two end-member solutions into a single optimization problem leading to a *damped least-squares solution* that attempts a compromise between reducing fitting errors and providing a simple (however simple is defined) solution. The damped qualifier comes from the intrinsic structure of the problem: it is posed like a least-squares problem, but with the addition of a penalty term (or damper) that rewards models that represent simpler solutions (e.g. smoother models or with smaller norms). A cost function for a damped least-squares problem can be written

$$\mathbf{F}(\mathbf{S}) = \varepsilon^2\mathbf{S}^T\mathbf{S} + (\mathbf{T}-\mathbf{D}\mathbf{S})^T(\mathbf{T}-\mathbf{D}\mathbf{S}) \qquad \text{(A1.9)}$$

where ε is a positive real number used to control the size of the penalty term.

As will be shown later in this appendix, the solution of this can be determined as being:

$$\mathbf{S}_{est} = (\mathbf{D}^T\mathbf{D} + \varepsilon^2\mathbf{I})^{-1}\mathbf{D}^T\mathbf{T} \qquad \text{(A1.10)}$$

Given the huge size and sparse nature of the matrix **D**, the tomography algorithm then solves Equation A1.10 iteratively to find **S**. This might be done for all locations on all layers simultaneously, but given the linear approximation, better results are obtained by adding another level of *outer* iterations, updating the model in a top-down manner. In production-type tomography workflows, several such outer iterations are typical, each involving a new forward modelling exercise (e.g. Lo and Inderwiesen, 1994; Jones 2010).

## Regularization

Solving Equation A1.1 directly will yield a possible solution to the inverse problem using the inconsistent noisy data with which we are working. However, the data inconsistencies, along with various approximations we make, means that there will in fact be a large family of possible solutions that all equally well explain the observed data. The volume of model space that contains these possible solutions is referred to as the 'null space' of the solution.

This null space can be reduced by adding various constraints to the solutions we derive. For example, we could require a solution that not only minimized (**T-DS**), but that also gave a model that was spatially smooth (i.e., the model spatial derivatives were small). Such constraints are usually referred to as 'regularization' terms.

***Model Regularization*** of the least-squares solution is one of the techniques commonly used to mitigate the local-minima and non-uniqueness pitfalls. Regularization reduces the range of estimated models by adding penalty terms to the cost function to help steer the solutions away from models that are deemed (however defined) less acceptable or less likely to be true. Regularization terms usually comprise some norm of the obtained model (e.g., be smooth, and/or have a limited range of variation) and may be written in a generalized form as

$$\mathbf{m}^T\mathbf{m} = \left[\mathbf{R}(\mathbf{S}-\mathbf{S}_o)\right]^T\left[\mathbf{R}(\mathbf{S}-\mathbf{S}_o)\right] \qquad \text{(A1.11)}$$

where **R** is a conveniently chosen matrix that encodes further ways to penalize a predefined set of models. Equation A1.11 allows for a variety of choices: one may choose **R=I** (an identity matrix) and $\mathbf{S}_o = 0$, thus penalizing models with larger $\mathbf{S}^T\mathbf{S}$; or, we may take **R** as matrix representation of a derivative, which will favour solutions describing smooth models; using a non-zero reference model means that we seek solutions that are similar to $\mathbf{S}_o$.

Another way of reducing and coping with uncertainty is to introduce weights in the residual vector **(T-DS),** via a linear transformation**, W** (**T – DS**), which will help in dealing with noise in the input data and with errors in the matrix **D**. If the errors are independent, then the matrix **W** can be diagonal.

With these modifications in hand we can write a more sophisticated cost function that now contains two terms, one related to the data fitting objective and another to the model norm penalty from Equation A1.11:

$$\mathbf{F(S) = W(T-DS)^T\, W(T-DS) + \varepsilon^2 m^T m}$$

$$\mathbf{F(S) = W(T-DS)^T\, W(T-DS) + \varepsilon^2 [R(S-S_0)]^T [R(S-S_0)]} \qquad \text{(A1.12)}$$

where $\boldsymbol{\varepsilon}$ (an arbitrarily small positive number) governs the trade-off between those two terms. Despite the addition of the model norm term and the new weights in the residual vector, the minimization of Equation A1.12 is also a linear, least-squares problem, which can be solved as follows. Firstly, noting that we wish to find a minimum in the cost function **F(S),** then as was done with Equations A1.3 and A1.4, take its partial derivative with respect to **S:**

$$\partial/\partial \mathbf{S}\ [\mathbf{F(S)}] = \mathbf{-W^2(T-DS)^T D + \varepsilon^2 [R(S-S_0)]^T R = 0} \qquad \text{(A1.13)}$$

We then evaluate this for a value of **S** that will be at a minimum, namely when we are at $\mathbf{S_0+\Delta S}$.

But also recall that the Taylor series expansion of a function of the form $\mathbf{Q(x_0+\Delta x)}$ can be truncated to first order to yield: $\mathbf{Q(X_0+\Delta X) = Q(X_0) + \Delta X\ \partial/\partial X\ [Q(X_0)]}$.

Then applying such an expansion to Equation A1.13, which already has a first derivative, yields:

$$\partial/\partial \mathbf{S}\ [\mathbf{F(S_0+\Delta S)}] = \partial/\partial \mathbf{S}\ [\mathbf{F(S_0)}] + \mathbf{\Delta S}\ \partial^2/\partial \mathbf{S}^2\ [\mathbf{F(S_0)}] \qquad \text{(A1.14)}$$

Given that the second derivative of **F(S)** is: $\partial^2/\partial \mathbf{S}^2[\mathbf{F(S)}] = \mathbf{W^2 D^T D + \varepsilon^2 R^T R,}$ then Equation A1.14 can be written as:

$$\partial/\partial \mathbf{S}\ [\mathbf{F(S_0+\Delta S)}] = \mathbf{-W^2(T-DS_0)^T D + \varepsilon^2 [R(S_0-S_0)]^T R + \Delta S\ [W^2 D^T D + \varepsilon^2 R^T R]}$$

$$\mathbf{= -W^2(T-DS_0)^T D + \Delta S\ [W^2 D^T D + \varepsilon^2 R^T R] = 0} \qquad \text{(A1.15)}$$

And rearranging gives:

$$\mathbf{W^2(T-DS_0)^T D = \Delta S\ [W^2 D^T D + \varepsilon^2 R^T R]} \qquad \text{(A1.16)}$$

However, noting that $\mathbf{\Delta S = S_{est} - S_0}$, where $\mathbf{S_{est}}$ is the final estimated (ideally 'true') model, then:

$$\mathbf{S_{est} = S_0 + W^2(T-DS_0)^T D\ [W^2 D^T D + \varepsilon^2 R^T R]^{-1}} \qquad \text{(A1.17)}$$

Or alternatively

$$\mathbf{S_{est} = S_0 + [D^T W^2 D + \varepsilon^2 R^T R]^{-1}\ D^T W^2 (T-DS_0)} \qquad \text{(A1.18)}$$

If we do not specify a preferred model (so that $\mathbf{S_0}$ will be 0**),** and set the weights **W** to be unity **(W=I),** and set the regularization also to be **R=I,** then Equation A1.18 reduces to Equation A1.10.

Ultimately we wish to be able to assess how errors and uncertainties in the input data and initial models are mapped by the tomographic operator **D** into our model estimations. With this goal in mind we may define the weights found on Equation A1.18 in terms of both the input data and the model parameters errors. This is justified because we want to attribute more importance to data points that we deem more representative or just less noisy, while also focusing on the estimation of parameters that can actually be resolved by the geometry of our seismic experiment and the physics used to set up **D**.

For example, using a weight proportional to the inverse of the initial data and model covariances ($C_d$ and $C_{m\text{-}prior}$ respectively), and noting that $\mathbf{S_{est}}$ is the posterior model estimate and $\mathbf{S_0}$ is the prior model estimate, we obtain:

$$\mathbf{S_{post} = S_{prior} + [D^T C_d^{-1} D + C^{-1}_{m\text{-}prior}]^{-1}\ D^T C_d^{-1} (T-DS_{prior})} \qquad \text{(A1.19)}$$

This formulation leads-in to the next appendix wherein we reformulate the problem in terms of conditional probability and a more pragmatic and tractable solution for the inverse of the Hessian $\mathbf{[D^T C_d^{-1} D + C^{-1}_{m\text{-}prior}]}$.

## Appendix 2: determining a posterior $C_m$ via Bayesian inference

It was noted earlier that Equation 13 could be used to recast a new minimization problem, obtained by maximizing $P(\mathbf{S})\ P\ (\mathbf{T}|\mathbf{S})$. If we now assume that $P(\mathbf{S})$ and $P(\mathbf{T}|\mathbf{S})$ are described by Gaussian functions, then $P(\mathbf{S}|\mathbf{T})$ is also a Gaussian and its maximum $\mathbf{S_{map}}$ is indeed the sought average model of the posterior distribution. Also, we can now rewrite the cost function **F(S)**, by minimizing the negative of the logarithms of $\boldsymbol{P}(\mathbf{S})$ and $\boldsymbol{P}(\mathbf{T}|\mathbf{S})$ as (Tarantola, 2005):

$$\mathbf{F(S) = -[\log P(S) + \log P(T|S)]} \qquad \text{(A2.1)}$$

Recall from Equation A1.12 that:

$$\mathbf{F(S) = W(T-DS)^T\, W(T-DS) + \varepsilon^2 [R(S-S_0)]^T [R(S-S_0)],}$$

and that in Equation A1.19 we used the covariance matrices as the weighting functions:
$\mathbf{C^{-1}_{m\text{-}prior} = \varepsilon^2 R^T R}$ and $\mathbf{C^{-1}_d = W^2}$; then with such changes:

$$\mathbf{F(S) = (T-DS)^T\, C^{-1}_d\, (T-DS) + (S-S_0)^T\, C^{-1}_{m\text{-}prior}\, (S-S_0)} \qquad \text{(A2.2)}$$

and $\mathbf{S_{map}}$ takes the form of the solution found in Equation A1.19.

To complete our description of the posterior distribution $\mathbf{P(S|T)}$, we need 'only' to evaluate its covariance matrix, which is also given in Equation A1.17. As demonstrated in Tarantola (2005), this covariance matrix is the normalization factor of the gradient in Equation A1.19 (i.e., the Hessian of the cost function) namely:

$$\mathbf{C_{m\text{-}post} = [D^T C_d^{-1} D + C^{-1}_{m\text{-}prior}]^{-1}} \qquad \text{(A2.3)}$$

However, this is computationally impractical to evaluate. Indeed the algorithms used to actually solve Equation A2.3 do that without explicitly forming or inverting this huge matrix. As a work-around, we can form an approximate representation of the most dominant parts of it using a truncated eigenvector/eigenvalue decomposition. Once this truncated eigen-decomposition is done we are then finally able to answer queries such as 'what are the most important (i.e. dominant) aspects of the model error?'

To assess which are the dominant components of **D**, using eigen decomposition we note that:

$$\begin{aligned}\mathbf{C}_{\text{m-post}} &= (\mathbf{D}^{\mathbf{T}}\,\mathbf{C}_{\mathbf{d}}{}^{-1}\,\mathbf{D} + \mathbf{C}^{-1}{}_{\text{m-prior}})^{-1} \\ &= \mathbf{C}_{\text{m-prior}}\,(\mathbf{C}_{\text{m-prior}}\,\mathbf{D}^{\mathbf{T}}\,\mathbf{C}_{\mathbf{d}}{}^{-1}\,\mathbf{D} + \mathbf{I})^{-1} \\ &\approx \mathbf{C}_{\text{m-prior}}\,(\mathbf{U}\,\boldsymbol{\Lambda}\,\mathbf{U}^{\mathbf{T}} + \mathbf{I})^{-1} \qquad \text{(A2.4)}\end{aligned}$$

Recall that the **D** matrix has dimensions of : [n_ measurements X n_model-cells] = [n_rays X n_parameters], and given that we have written $\mathbf{U}\,\boldsymbol{\Lambda}\,\mathbf{U}^{\mathrm{T}} = \mathbf{C}_{\text{m-prior}}\,\mathbf{D}^{\mathbf{T}}\,\mathbf{C}_{\mathbf{d}}{}^{-1}\,\mathbf{D}$, the eigen-vector matrix U has dimensions [n- parameters X n- parameters].

Given the Woodbury matrix identity, which notes that in general for matrices **A, B, U, V**:

$$(\mathbf{UBV+A})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\,\mathbf{U}(\mathbf{B}^{-1} + \mathbf{VA}^{-1}\,\mathbf{U})^{-1}\,\mathbf{VA}^{-1}$$

Then, Equation A2.4 can be written:

$$\begin{aligned}\mathbf{C}_{\text{m-post}} &= \mathbf{C}_{\text{m-prior}}\,[\mathbf{I} - \mathbf{IU}(\boldsymbol{\Lambda}^{-1} + \mathbf{U}^{\mathbf{T}}\mathbf{IU})^{-1}\,\mathbf{U}^{\mathbf{T}}\mathbf{I}] \\ &= \mathbf{C}_{\text{m-prior}}\,[\mathbf{I} - \mathbf{U}(\boldsymbol{\Lambda}^{-1} + \mathbf{I})^{-1}\,\mathbf{U}^{\mathbf{T}}] \qquad \text{(A2.5)}\end{aligned}$$

And noting that: $(\boldsymbol{\Lambda}^{-1} + \mathbf{I})^{-1} = \boldsymbol{\Lambda}/(\mathbf{I} + \boldsymbol{\Lambda})$, this becomes:

$$\mathbf{C}_{\text{m-post}} = \mathbf{C}_{\text{m-prior}}\,[\mathbf{I} - \mathbf{U}\,\{\boldsymbol{\Lambda}/(\mathbf{I} + \boldsymbol{\Lambda})\}\,\mathbf{U}^{\mathbf{T}}] \qquad \text{(A2.6)}$$

Putting $\mathbf{A} = \boldsymbol{\Lambda}/(\mathbf{I} + \boldsymbol{\Lambda})$ this becomes:

$$\mathbf{C}_{\text{m-post}} = \mathbf{C}_{\text{m-prior}}\,(\mathbf{I} - \mathbf{UAU}^{\mathbf{T}}) \qquad \text{(A2.7)}$$

This new form of **C**m-post is more convenient to evaluate, as it removes the explicit requirement to evaluate the inverse term in Equation A2.4.

Once we have found the dominant sources of error, we can select a meaningful subset of the, say *k*, eigenvectors to draw a representative random suite of perturbed velocity models, all of which explain the observed data equally well. The range of perturbations are then bounded by our estimates of possible error.

Bayesian inference provides a quick way to compute the model realisations that equally fit the data:

$$\mathbf{m} = \mathbf{m}_{\mathbf{map}} + \mathbf{C}^{\mathbf{1/2}}{}_{\text{m-post}}\,\mathbf{r} \qquad \text{(A2.8)}$$

Which using Equation A2.7 can then be written as:

$$\mathbf{m} = \mathbf{m}_{\mathbf{map}} + \mathbf{C}^{\mathbf{1/2}}{}_{\text{m-prior}}\,(\mathbf{I} - \mathbf{UAU}^{\mathbf{T}})^{1/2}\,\mathbf{r} \qquad \text{(A2.9)}$$

This can be further simplified by approximating the square-root term, via a truncation of the eigen-decomposition of the of the matrix of the form $\mathbf{B} = (\mathbf{I} - \mathbf{UAU}^{\mathbf{T}})$. Following Bui-Thanh et al. (2013), and noting that the identity matrix can be written as $\mathbf{I} = \sum_{i=1}^{n} \boldsymbol{u}_i \boldsymbol{u}_i^T$, then the spectral representation of **B** is

$$\mathbf{B} = \mathbf{I} - \mathbf{UAU}^{T}$$

$$\mathbf{B} = \sum_{i=1}^{n}\left[1 - \frac{\lambda_i}{(1+\lambda_i)}\right]\boldsymbol{u}_i\boldsymbol{u}_i^T$$

$$\mathbf{B} = \sum_{i=1}^{n}\frac{1}{(1+\lambda_i)}\,\boldsymbol{u}_i\boldsymbol{u}_i^T, \qquad \text{(A2.10)}$$

and its square root is

$$\mathbf{B}^{\frac{1}{2}} = \sum_{i=1}^{n}\frac{1}{\sqrt{(1+\lambda_i)}}\,\boldsymbol{u}_i\boldsymbol{u}_i^T$$

A low-rank approximation of this square root may be obtained by considering the k largest eigenvalues:

$$\mathbf{B}^{\frac{1}{2}} = \sum_{i=1}^{k}\frac{1}{\sqrt{(1+\lambda_i)}}\,\boldsymbol{u}_i\boldsymbol{u}_i^T + \sum_{i=k+1}^{n}\frac{1}{\sqrt{(1+\lambda_i)}}\,\boldsymbol{u}_i\boldsymbol{u}_i^T,$$

Then for sufficiently large k, where the eigenvalues are smaller than one, it is possible to approximate the second term above like so:

$$\mathbf{B}^{\frac{1}{2}} \approx \sum_{i=1}^{k}\frac{1}{\sqrt{(1+\lambda_i)}}\,\boldsymbol{u}_i\boldsymbol{u}_i^T + \sum_{i=k+1}^{n}\boldsymbol{u}_i\boldsymbol{u}_i^T.$$

With further algebraic manipulation, one may simplify the above result to a more suitable form

$$\mathbf{B}^{\frac{1}{2}} \approx \sum_{i=1}^{k}\frac{1}{\sqrt{(1+\lambda_i)}}\,\boldsymbol{u}_i\boldsymbol{u}_i^T - \sum_{i=1}^{k}\boldsymbol{u}_i\boldsymbol{u}_i^T + \sum_{i=1}^{k}\boldsymbol{u}_i\boldsymbol{u}_i^T + \sum_{i=k+1}^{n}\boldsymbol{u}_i\boldsymbol{u}_i^T,$$

$$\mathbf{B}^{\frac{1}{2}} \approx \sum_{i=1}^{k}\frac{1}{\sqrt{(1+\lambda_i)}}\,\boldsymbol{u}_i\boldsymbol{u}_i^T - \sum_{i=1}^{k}\boldsymbol{u}_i\boldsymbol{u}_i^T + \sum_{i=1}^{n}\boldsymbol{u}_i\boldsymbol{u}_i^T,$$

$$\mathbf{B}^{\frac{1}{2}} \approx \sum_{i=1}^{n}\boldsymbol{u}_i\boldsymbol{u}_i^T + \sum_{i=1}^{k}\left(\frac{1}{\sqrt{1+\lambda_i}} - 1\right)\boldsymbol{u}_i\boldsymbol{u}_i^T, \qquad \text{(A2.11)}$$

$$\mathbf{B}^{\frac{1}{2}} \approx \left(\boldsymbol{I} + \mathbf{U}_k\,\mathbf{L}_k\,\mathbf{U}_k^T\right),$$

And hence Equation A2.9 becomes:

$$\mathbf{m} = \mathbf{m}_{\mathbf{map}} + \mathbf{C}^{\mathbf{1/2}}{}_{\text{m-prior}}\,(\mathbf{I} + \mathbf{U}_{\mathbf{k}}\,\mathbf{L}_{\mathbf{k}}\,\mathbf{U}_{\mathbf{k}}{}^{\mathbf{T}})\,\mathbf{r}$$

$$\mathbf{m} = \mathbf{m}_{\mathbf{map}} + \mathbf{C}^{\mathbf{1/2}}{}_{\text{m-prior}}\,\mathbf{K}\,\mathbf{r} \qquad \text{(A2.12)}$$

Where

$\mathbf{C}_{m\text{-}prior}$ is the prior model covariance

$\mathbf{K} = (\mathbf{I} + \mathbf{U}_{\mathbf{k}}\,\mathbf{L}_{\mathbf{k}}\,\mathbf{U}_{\mathbf{k}}{}^{\mathbf{T}})$

**r** is a random vector drawn from the Normal distribution

$$\boldsymbol{L}_k = diag\left(\frac{1}{\sqrt{\lambda_1+1}} - 1, \frac{1}{\sqrt{\lambda_2+1}} - 1, \ldots, \frac{1}{\sqrt{\lambda_k+1}} - 1\right)$$

$\lambda_i$ is the i-th eigenvalue found on the diagonal matrix $\boldsymbol{\Lambda}$