

Uncertainty quantification for regularized inversion of electromagnetic geophysical data. Part II: Application in 1D and 2D problems.

Daniel Blatter^a, Matthias Morzfeld^a, Kerry Key^b, Steven Constable^a

^a*Scripps Institution of Oceanography, University of California, San Diego,*

La Jolla, CA 92037, USA. E-mail: dblatter@ucsd.edu

^b*Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964 USA.*

SUMMARY

This paper is Part II of a two-part series on a mathematical and computational framework for computing a meaningful uncertainty quantification (UQ) for regularized inversions of electromagnetic data. In Part I, we explained the theory behind a sampling algorithm, which we call RTO-TKO, and in Part II, we showcase RTO-TKO in practice. We individually and jointly invert seafloor magnetotelluric (MT) and surface-towed controlled source electromagnetic field data, collected for imaging offshore freshened groundwater beneath the US Atlantic margin. We also invert seafloor MT data collected for subsalt imaging to produce 2D resistivity models and uncertainty estimates that characterize the salt body geometry and surrounding sediments. We compare the UQ of the RTO-TKO with results from trans-dimensional sampling, and explain the differences arising from different underlying (prior) assumptions of the two algorithms. We also discuss the practical implications of these findings. Most importantly, however, the 2D case study unam-

biguously demonstrates the computational advantages of RTO-TKO and its ability to make use of massive parallelism.

1 INTRODUCTION

Most geophysical inverse problems are nonlinear and, for that reason, their solutions are not unique. In practice, this means that many different models of the subsurface can explain the geophysical data, which are often sparse and limited to measurements taken at the surface. As a result, an understanding of the uncertainty in the inverted model parameters is paramount. This is particularly true when the inverted model values are used to constrain other Earth properties of interest (e.g., Selway et al. 2019; Blatter et al. 2019).

Despite its importance, uncertainty quantification (UQ) in geophysics has proven difficult because of computational limitations. Due to nonlinearity, most numerical methods are either inadequate (because they neglect nonlinearity to achieve computational efficiency) or impractical. Bayesian sampling methods, for example, are in principle capable of computing a nonlinear UQ, but may require weeks or even months to solve problems in 2D or 3D (e.g. Blatter et al. 2021; Zhang et al. 2018; Agostinetti et al. 2015), depending on the computational cost of the forward problem.

In this work, we present the mathematical and computational framework for computing a meaningful UQ for regularized inversions of electromagnetic (EM) data that can overcome the computational bottleneck usually associated with Bayesian sampling. Specifically, Part I of this two-part series describes the mathematics of a sampling algorithm which we call the “RTO-TKO.” RTO-TKO achieves computational efficiency by combining sampling with regularized inversion, leveraging massive parallelism. The former means that RTO-TKO repurposes robust, deterministic inversion codes; the latter means that RTO-TKO scales onto massively parallel computing architectures (see Part I for the mathematics of why). Besides computational efficiency, RTO-TKO has the advantages of requiring little or no tuning, being unaffected by refinement of the parameter grid, and automatically adjusting the regularization strength parameter, similar in spirit to the Occam inversion (Constable et al. 1987).

Here in Part II we focus on practical aspects of RTO-TKO. We first demonstrate the versatility of RTO-TKO by individually and jointly inverting seafloor MT and surface-towed controlled source electromagnetic (CSEM) field data, (Gustafson et al. 2019; Blatter et al. 2019) for 1D seafloor resistivity models. We then study the computational efficiency of RTO-TKO in the context of a 2D MT problem. Specifically, we invert seafloor MT data (Key et al.

2006) for 2D resistivity models of a salt body and surrounding sediments in the northern Gulf of Mexico. We compare the computational cost of RTO-TKO to a recently developed Bayesian sampling technique named ‘TDGP,’ which is a Bayesian sampler that uses a sparse trans-dimensional model parameterization for efficient sampling (Ray & Myer 2019; Blatter et al. 2021). We find that RTO-TKO needs 2 days (or less) to solve this problem, whereas TDGP requires about 10 days.

We argue that RTO-TKO, coupled with a fast regularized inverse solver, could make 3D Bayesian sampling of MT data feasible. As a thought experiment, assume that one regularized inversion takes about 0.5 hr on 100 cores. Then, leveraging parallelism and 1,000 cores, RTO-TKO can obtain about 500 samples in one day, which is likely sufficient for an approximate UQ. A similar estimate of computational resources for TDGP leads to a computing time of about a year (assuming that a single forward response calculation takes about one minute on 100 cores).

In all case studies (1D and 2D), we show and discuss the quantitative uncertainty produced by RTO-TKO, and compare it against the conventional regularized inversion obtained using the Occam method (Constable et al. 1987) and to results of trans-dimensional Markov Chain Monte Carlo (MCMC) samplers. It is also possible to obtain uncertainty estimates from a regularized inversion model via local linearization (Tarantola 2005). We briefly elaborate the theory behind this approach, apply it to the 1D field MT data, and discuss its shortcomings.

We also illustrate, via the various case studies (in 1D and 2D), the crucial point that prior assumptions about model regularization and parameterization have a large effect on the resulting (posterior) uncertainty estimates. We discuss the practical implications of these facts.

Throughout, we only show results using field data, but also provide two synthetic examples in the appendix. One motivation for showing results of the synthetic examples is that we constructed examples in which additional constraints on the regularization penalty weight are needed.

2 USING RTO-TKO TO JOINTLY INVERT MULTIPLE EM DATA SETS

The basic theory behind stochastic optimization in general, and RTO-TKO in particular, is discussed at length in Part I of this series. For the purposes of this discussion, we briefly reproduce the principal points here.

RTO-TKO solves two optimization problems per iteration. The first, and more costly,

solves for the model parameters at constant regularization penalty weight μ ,

$$\min_{\mathbf{m}} f(\mathbf{m}) = \frac{1}{2} \left\| C_d^{-1/2} (\mathbf{F}(\mathbf{m}) - \tilde{\mathbf{d}}) \right\|^2 + \frac{\mu}{2} \left\| L(\mathbf{m} - \tilde{\mathbf{m}}) \right\|^2 \quad (1)$$

where \mathbf{m} is a vector of parameters representing the model; f is the objective function to be minimized; C_d is the data covariance matrix; \mathbf{F} is the forward modeling function that maps the model parameters to predicted data; $\tilde{\mathbf{d}} \sim \mathcal{N}(\mathbf{d}, C_d)$ is the perturbed data (with \mathbf{d} the measured data); L is a regularization operator; and $\tilde{\mathbf{m}}$ is a draw from the prior $\mathcal{N}(0, \frac{1}{\mu}(L^T L)^{-1})$, which is a Gaussian with zero mean and covariance $C_m = \frac{1}{\mu}(L^T L)^{-1}$. We note that typically L (or $L^T L$) is not invertible and discuss this issue in detail below. In a second step, RTO-TKO adjusts the regularization strength parameter μ , by solving a scalar optimization problem in μ , but with a fixed model \mathbf{m}

RTO-TKO is readily adaptable to joint inversion of multiple data sets that all infer on the same physical parameters. Adding additional data simply extends $\tilde{\mathbf{d}}$ and C_d to include the new data and their uncertainty and redefines F to include all relevant forward operators, such that the size and type of the predicted data match that of the measured data. In this case, we inverted collocated surface-towed CSEM and seafloor MT data, part of a survey of low salinity groundwater in the continental shelf offshore New Jersey (Gustafson et al. 2019), for 1D models of resistivity using RTO-TKO. We inverted seafloor MT station NJ05 and the adjacent surface-towed CSEM data. The frequency range of the MT data was 0.625-56.0 Hz, while for the CSEM data we inverted 0.25 and 0.75 Hz at four offsets of 600, 870, 1120, and 1380 m. Further details of the data inverted are provided in Gustafson et al. (2019); Blatter et al. (2018). The geological setting in the region and the depth sensitivity of the data (the long period MT data were not used) are such that 1D models are appropriate. The CSEM forward calculations were done using the Dipole1D code (Key 2009) while the MT forward calculations used the standard recursion relationship (Ward & Hohmann 1987).

When performing the optimization in Eq. 1, the model parameters can be constrained to lie within set bounds. In the case of this joint inversion, we constrained \mathbf{m} to lie between 0.1 ohm-m and 1,000 ohm-m. This decision was made based on knowledge of the local geology, which is dominated by porous, fluid saturated sediments. This constraint also applies to $\tilde{\mathbf{m}}$, which has the same physical units as \mathbf{m} . In practice, this means that drawing a sample $\tilde{\mathbf{m}}$ from C_m involves solving the constrained linear system

$$\sqrt{\mu} L \tilde{\mathbf{m}} = \boldsymbol{\eta} \quad (2)$$

$$m_{lb} \leq m_i \leq m_{ub} \quad \forall i \quad (3)$$

where $\boldsymbol{\eta} \sim \mathcal{N}(0, I)$, with I being the identity matrix of the same size as L , and m_{lb} and m_{ub} are the upper and lower bounds on \mathbf{m} , respectively. Using the constrained least squares solve also enables us to use non-invertible regularization operators. In all inversions performed as part of this study, the initial model is set to a resistive halfspace (100 ohm-m).

The MT and CSEM data contain complementary information about the subsurface (Blatter et al. 2019). The MT data are more sensitive to and therefore better able to constrain the more conductive regions of the subsurface, while the CSEM data have a higher sensitivity to resistive layers. Further, the MT data give sensitivity to greater depths than the CSEM data, whose relatively short offsets give sensitivity that is limited to the shallower parts of the model.

The RTO-TKO and trans-D MCMC posteriors (see Figure 1) are similar in many ways, yet different in others. Both MT posteriors, for example, estimate the conductive portions of the model similarly well, while in the resistive portions where the data are relatively uninformative their divergence reflects differences in their prior assumptions. RTO-TKO assumes the subsurface resistivity is smoothly varying and composed of a fixed, relatively fine grid of layers. The trans-D MCMC algorithm, on the other hand, determines the number of layers during the inversion and imposes no smoothness requirement between them. Instead, it relies on Bayesian parsimony (MacKay 2003) to regularize the model, such that simpler models are more probable than complex ones. This amounts to a prior assumption that the subsurface is composed of relatively few, piece-wise constant layers with large jumps in physical properties allowed between them. The RTO-TKO model parametrization and regularization give rise to smoothly varying, meandering models whose minimum and maximum values are constrained by the roughness penalty, in addition to the data. On the other hand, the trans-D MCMC model parametrization and (implicit) regularization produce simple, blocky models with abrupt transitions, resulting in ‘stripes’ of probability density where large, single layers fit the data well (this is especially evident in the CSEM posterior).

With the different ways these two algorithms parametrize and regularize the model in mind, a similar geologic interpretation can be derived from the posteriors in Fig. 1. The MT posteriors for both RTO-TKO and trans-D MCMC highlight conductive zones in the intervals 0-100 m, 200-400 m, and 400-600 m. In the interval 100-200 m (the depth interval coinciding with a freshened aquifer), the 5th percentiles both indicate a minimum resistivity of about 10 ohm-m, while the 95th percentiles disagree on the basis of regularization—smoothing is enforced in the RTO-TKO inversion, while none is applied in the trans-D MCMC inversion, while the MT data themselves contain little to no information about the maximum allowed

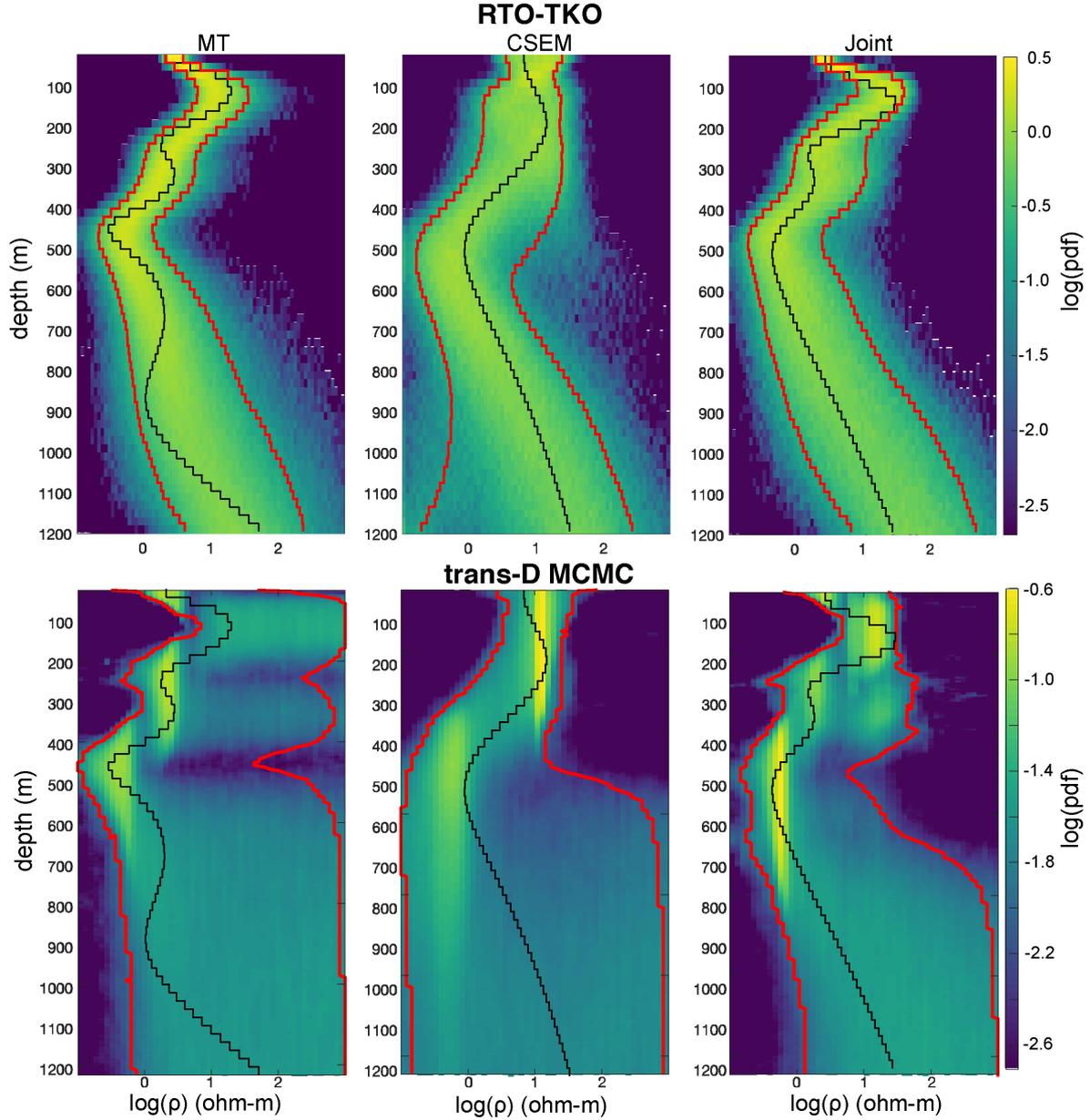


Figure 1. Marginal posterior uncertainty for resistivity as a function of depth for individual and joint inversions of field MT and CSEM data (Gustafson et al. 2019) using RTO-TKO and trans-D MCMC. Warmer colors indicate higher regions of probability. The left and right red lines denote the 5th and 95th percentiles, respectively, while the black dotted line represents the Occam inversion. The complementary information contained in the two data sets combines to produce a higher-resolution image of the freshwater aquifer (300 m depth) and the saline brines (400 m) than in either of the individual inversions. The stripes in the trans-D MCMC posteriors reflects the trans-D preference for models with fewer parameters, while the high 95th percentile in the MT inversion reflects the lack of derivative smoothing in the trans-D MCMC prior.

resistivity in this layer. The CSEM posteriors, meanwhile, tell a consistent story—a resistive layer down to about 400 m above a conductive layer of indeterminate depth.

One could argue that 95th percentile of the RTO-TKO posterior is overly optimistic—for example, at 150 m depth—given MT data’s inability to sense the absolute resistivity of thin resistive layers. The subsurface in this layer could indeed be far more resistive than the RTO-TKO 95th percentile suggests and still fit the data (the proof of this is in the trans-D MCMC models that do exactly that). Such models could *not*, however, simultaneously satisfy the constraint of being *smooth*, which we impose on the model as prior information about the subsurface when we choose to use derivative smoothing as model regularization. As a result, the RTO-TKO 95th percentile is not overly optimistic, it is simply faithfully quantifying the uncertainty related to asserting smoothness as prior information—a choice that is common practice in EM geophysics, for better or for worse.

The joint posteriors contain the information from both data sets and tell a more complex, nuanced story than either of the individual posteriors alone. The 95th percentile of the trans-D MCMC joint posterior is better constrained in the region of the freshened aquifer (100-200 m) by the CSEM data, but also indicates the possibility of freshened water deeper down, from 300-400 m. The joint RTO-TKO posterior, meanwhile, indicates a more resistive 95th percentile generally, but especially in the 100-200 m and 300-400 m intervals.

It may be tempting to view differences in the posterior uncertainties estimated by these two Bayesian sampling algorithms as evidence of a flaw in either algorithm (or both). Instead, it is instructive to view their differences as indicative of the fact that the data do not contain complete information about subsurface resistivity, leaving ample room for their respective priors to leave a strong imprint on their associated posteriors. In essence, through our choices of model parametrization and regularization we fill in the gaps left by the data with our explicit or implicit biases on preferred model structure. This may in some sense be unfortunate, but it is also inevitable. Geophysical data rarely contain sufficiently complete information about the model to render *a priori* assumptions more or less irrelevant. As a result, the prior—including, crucially, how to parametrize the model and what regularization (implicit or explicit) to apply—must be selected with care, mindful that it will likely have a strong impact on the posterior uncertainty.

3 2D MT UNCERTAINTY QUANTIFICATION USING RTO-TKO

The 1D examples illustrate the various desirable features and capabilities of RTO-TKO but the real benefits of this approach to uncertainty quantification (UQ) lie in computational sav-

ings and total run time reduction in 2D (or 3D) problems. To demonstrate the computational efficiency of stochastic optimization, we invert field MT data for 2D models of subsurface resistivity using RTO-TKO. We compare the resulting uncertainty estimates with those obtained from TDGP, a state-of-the-art trans-dimensional MCMC inversion algorithm (Ray & Myer 2019; Blatter et al. 2021). We then discuss the relative computational cost of both algorithms, in terms of total run time and flops.

We modified the MARE2DEM inversion code (Key 2016) to carry out the fixed- μ minimization in Eq. 1, which was straightforward using MARE2DEM’s existing regularized Gauss-Newton minimization framework. The RTO-TKO algorithm itself is coded in Julia, and we use a Julia-Fortran interface to connect the RTO-TKO code to MARE2DEM. The minimization over μ at fixed \mathbf{m} (the TKO step, see Part I) we implement in Julia using a Golden Section search method, though MARE2DEM could easily be modified to conduct this minimization in a more computationally efficient way.

By using the existing MARE2DEM input file format, the 2D models can retain all the benefits of MARE2DEM’s unstructured model parameterization that allows for accurate inclusion of bathymetry and other fixed 2D structures. In addition, using MARE2DEM makes it simple to run a canonical regularized inversion (e.g., Occam inversion) first for reference prior to running RTO-TKO. While we only test MT inversion here, the code is also able to invert CSEM and DC resistivity data, since MARE2DEM already supports these data types.

Here we invert a profile of data from the Gemini MT survey (Key et al. 2006)—the same profile that was inverted in Blatter et al. (2021). These TE mode MT data image a resistive salt body in the Gulf of Mexico, surrounded by conductive, seawater-saturated sediments. The resistive target is challenging to image since thin, spatially constrained resistors are difficult for MT data to resolve.

The Gemini data are shown in Fig. 2, spanning the period band 1-200 sec. Data uncertainties were estimated using standard MT processing workflows. Error floors of 2.5% for apparent resistivity and phase were applied. We assumed the data covariance, C_d , to be diagonal (the data errors are uncorrelated). We inverted these data with RTO-TKO, drawing 8,831 samples using ten independent samplers (each sampler consisted of a single compute node with 24 CPUs). The lower and upper bounds on resistivity were set to 0.1 and 10 ohm-m. We use a constrained, log-normal prior for the regularization penalty weights with lower and upper bounds of 10 and 1,000, and mean of 100 and standard deviation of 0.5 log units (so that most of the prior probability mass is within the assumed, relatively wide bounds).

Fig. 3 shows the RMS misfit, compute time and forward evaluations per RTO-TKO iter-

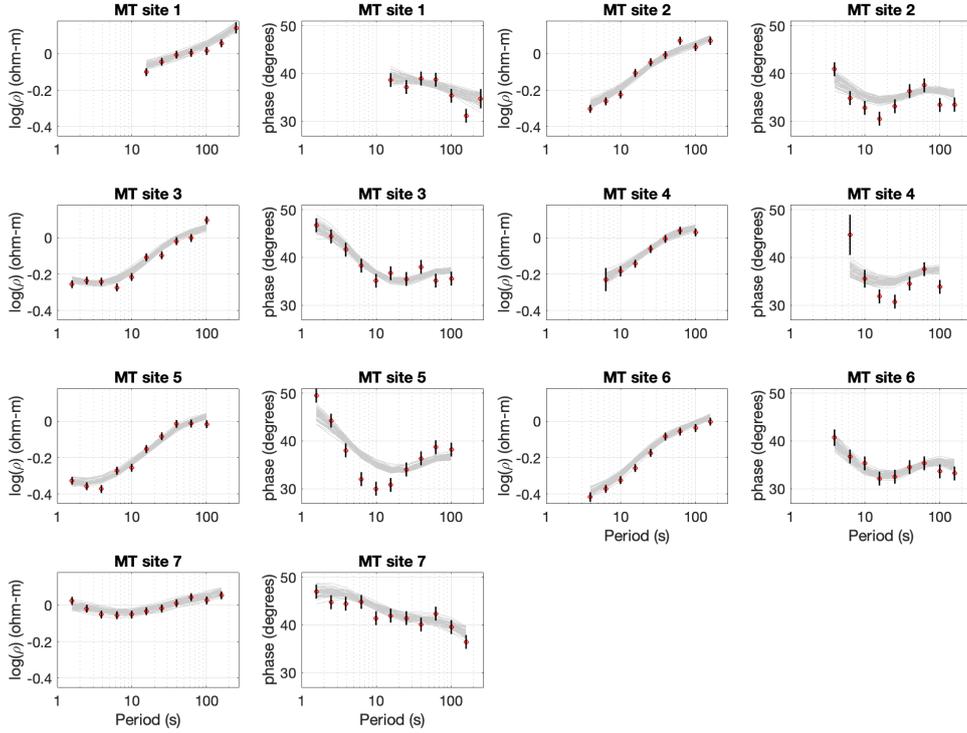


Figure 2. TE mode MT data (apparent resistivity and phase) from seven ocean bottom MT stations (white triangles in Fig. 4) of the Gemini data set (Key et al, 2006). The data span the period range 1-200 sec. Model responses from 50 randomly selected models from an RTO-TKO ensemble are plotted in gray.

ation, and posterior sampling over μ . As Fig. 3d-e indicate, the posterior distribution of the regularization penalty weight μ is stationary with a far lower variance than the prior, indicating that the data and model prior on resistivity contain sufficient information to constrain the regularization penalty weight (with nearly no influence of the wide prior over μ). We refer to the appendix for synthetic numerical experiments where stronger prior constraints on μ are necessary.

The mean of the RTO-TKO distribution is shown in Fig. 4 alongside the Occam inversion result and the mean of a TDGP inversion of the same data (the TDGP model ensemble contained one million models). All three image the conductive, shallow sediments, the resistive salt body, and the deeper, resistive basement rock. The RTO-TKO mean is slightly smoother than the Occam inversion model. The Occam inversion finished with a regularization penalty weight of 6.7 whereas the RTO-TKO inversions converged on values of about 60 – 110 (Fig.

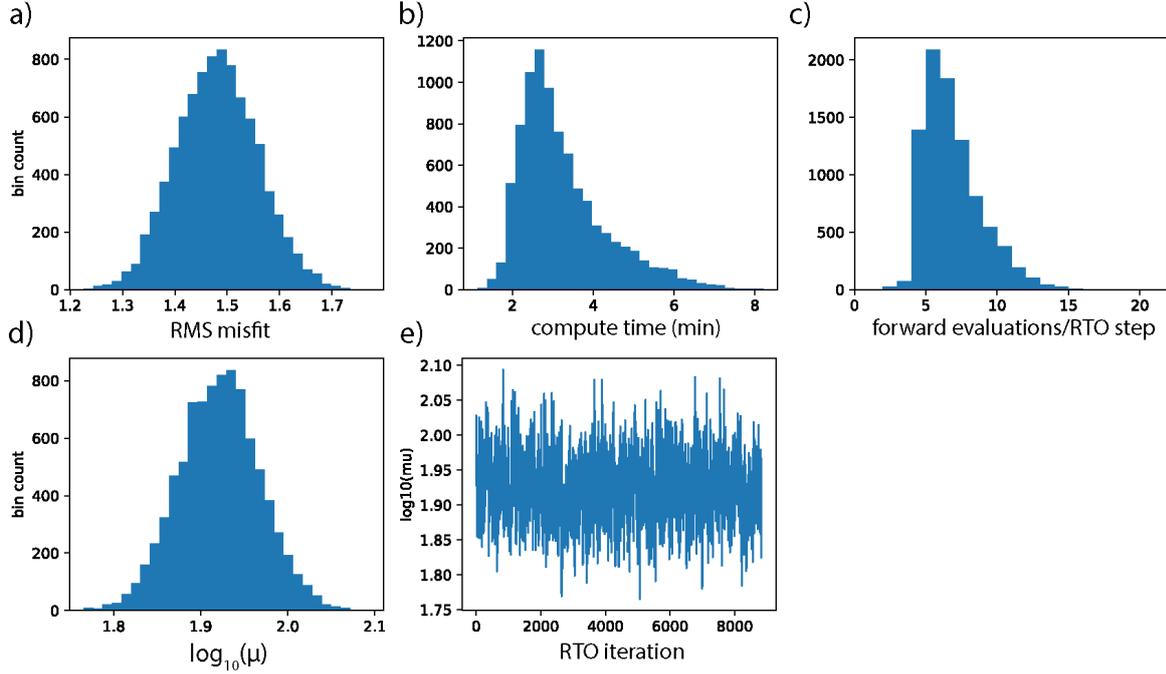


Figure 3. RTO-TKO inversion statistics for the Gemini 2D MT field data. Compare with Fig. A6. Note that even though an informative prior was not specified for μ , its sampling distribution is stationary (d-e). Compute time (b) is run time per RTO-TKO iteration.

A6). The RTO-TKO mean is likewise smoother than the TDGP mean due to the nature of the trans-dimensional Gaussian Process (GP) parametrization, which allows large, local changes in resistivity in the vicinity of the GP interpolation nodes.

The uncertainty in the subsurface resistivity is shown in Fig. 5 and Fig. 6, which contain marginal distributions for horizontal and vertical slices through the model, respectively. The warmer colors in these figures indicate regions of higher probability density, while the red lines bound the 90% credible interval. The white squares in both figures represent the Occam inversion model values within each depth or lateral position slice. The TDGP posterior uncertainty for the same horizontal and vertical model slices are plotted in each figure for reference.

Significant differences between the RTO-TKO and TDGP algorithms are evident in the uncertainties shown Fig. 5 and Fig. 6. The 90% credible interval estimated by RTO-TKO tends to be smaller than that estimated by TDGP. The RTO-TKO distribution of probability density also tends to be more Gaussian (though by no means symmetric about the Occam inversion model) while the TDGP distribution features a strong ‘stripe’ of probability density around 1 ohm-m (the mean of the TDGP model ensemble). Despite this, the general shape of

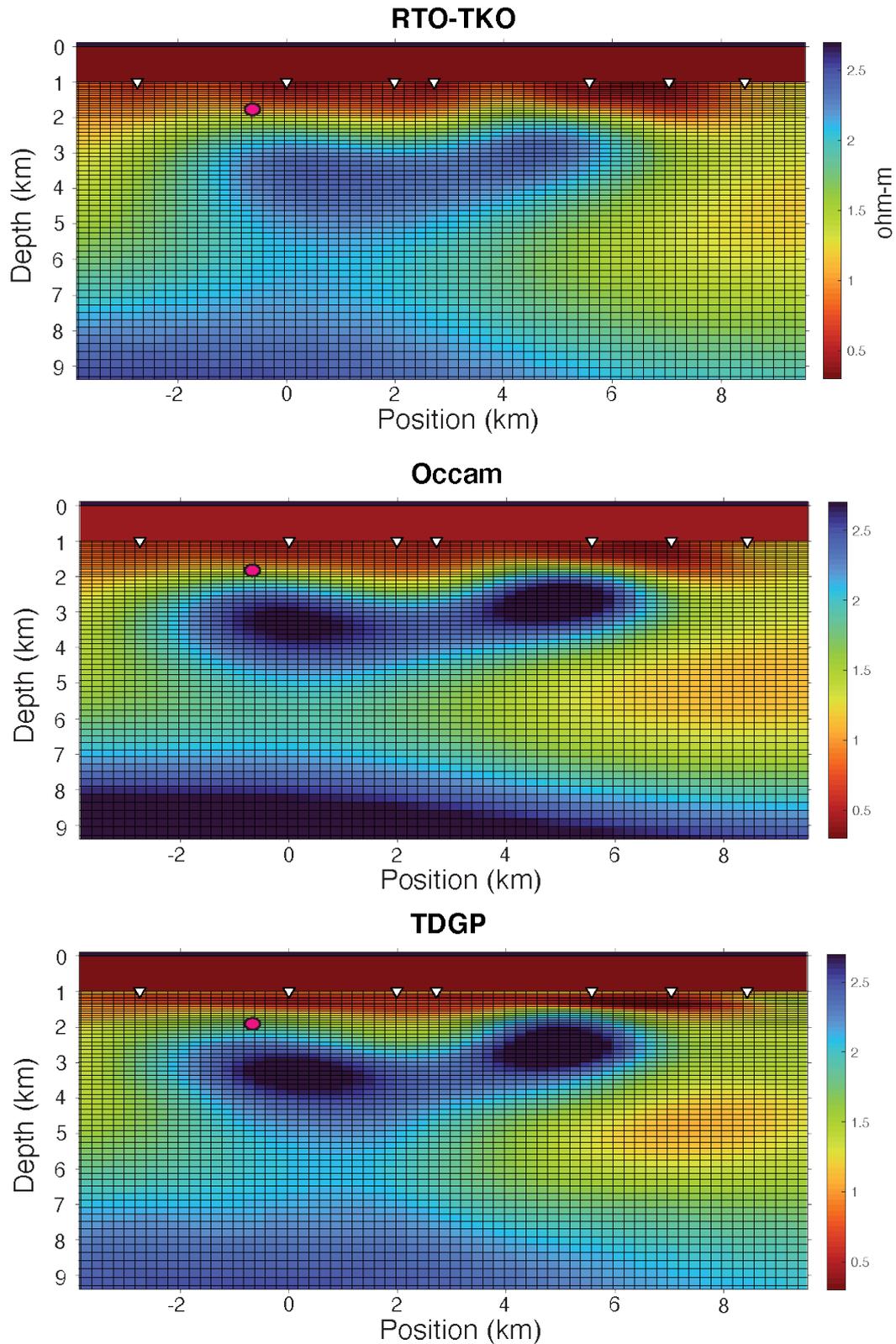


Figure 4. Inversion results of the Gemini field MT data set. The mean of an RTO-TKO ensemble of 8,831 models resembles the Occam inversion model and the mean of the TDGP ensemble of one million models. The MT station locations are shown as white triangles. The pink circle indicates the location of the model parameter discussed in Fig. 10. The color scales are in linear ohm-m (not log units)

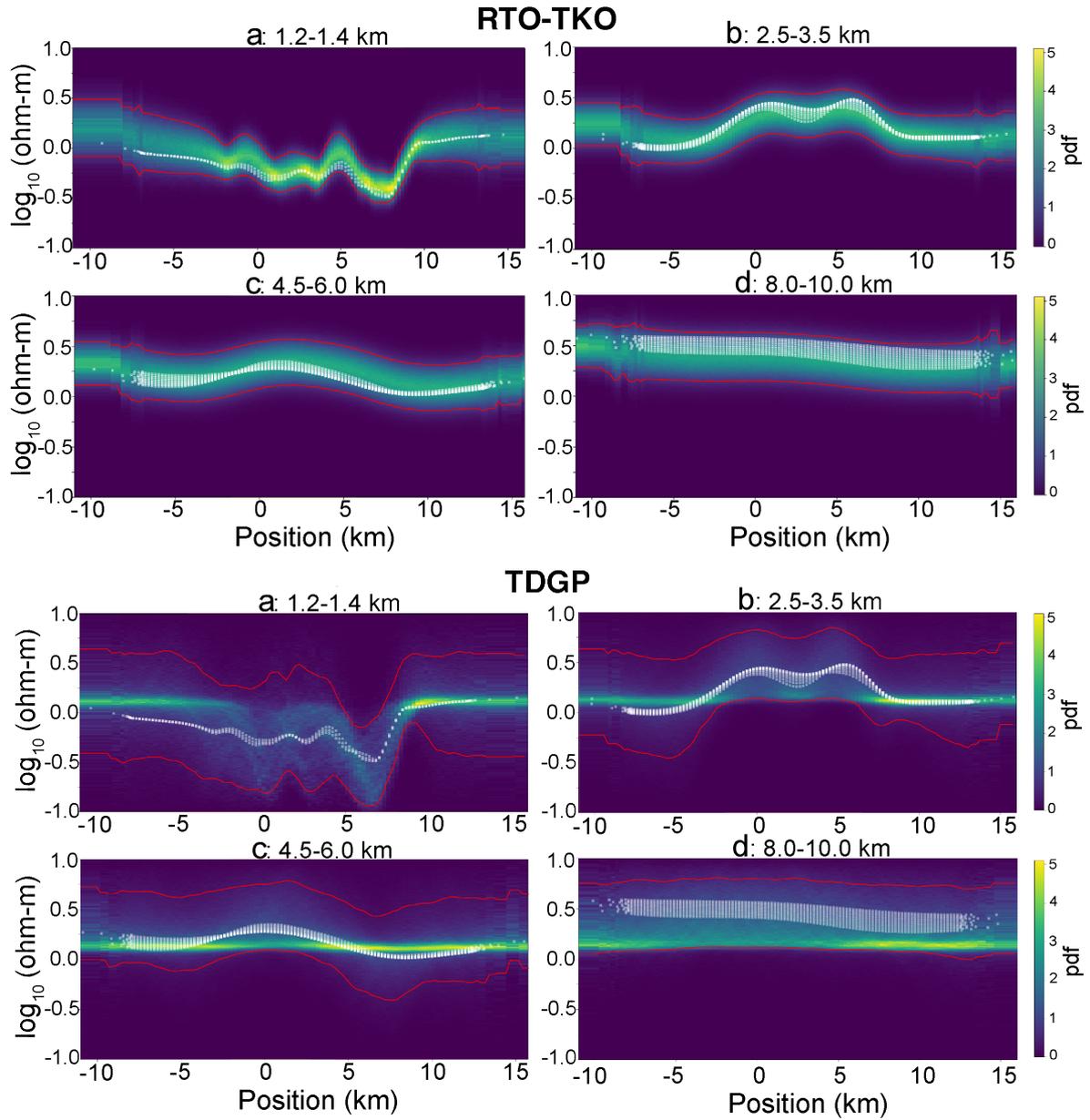


Figure 5. Marginal posterior probability distributions for resistivity as a function of lateral position over specified depth slices using both the RTO-TKO and TDGP algorithms. Warmer colors indicate regions of higher probability. Lower and upper red lines indicate 5th and 95th percentiles, respectively, while white squares represent the Occam inversion model values within the depth slice. The differences in the posteriors, while reflecting the same subsurface structure, are indicative of the differences in their respective priors, especially how the model is parametrized and regularized.

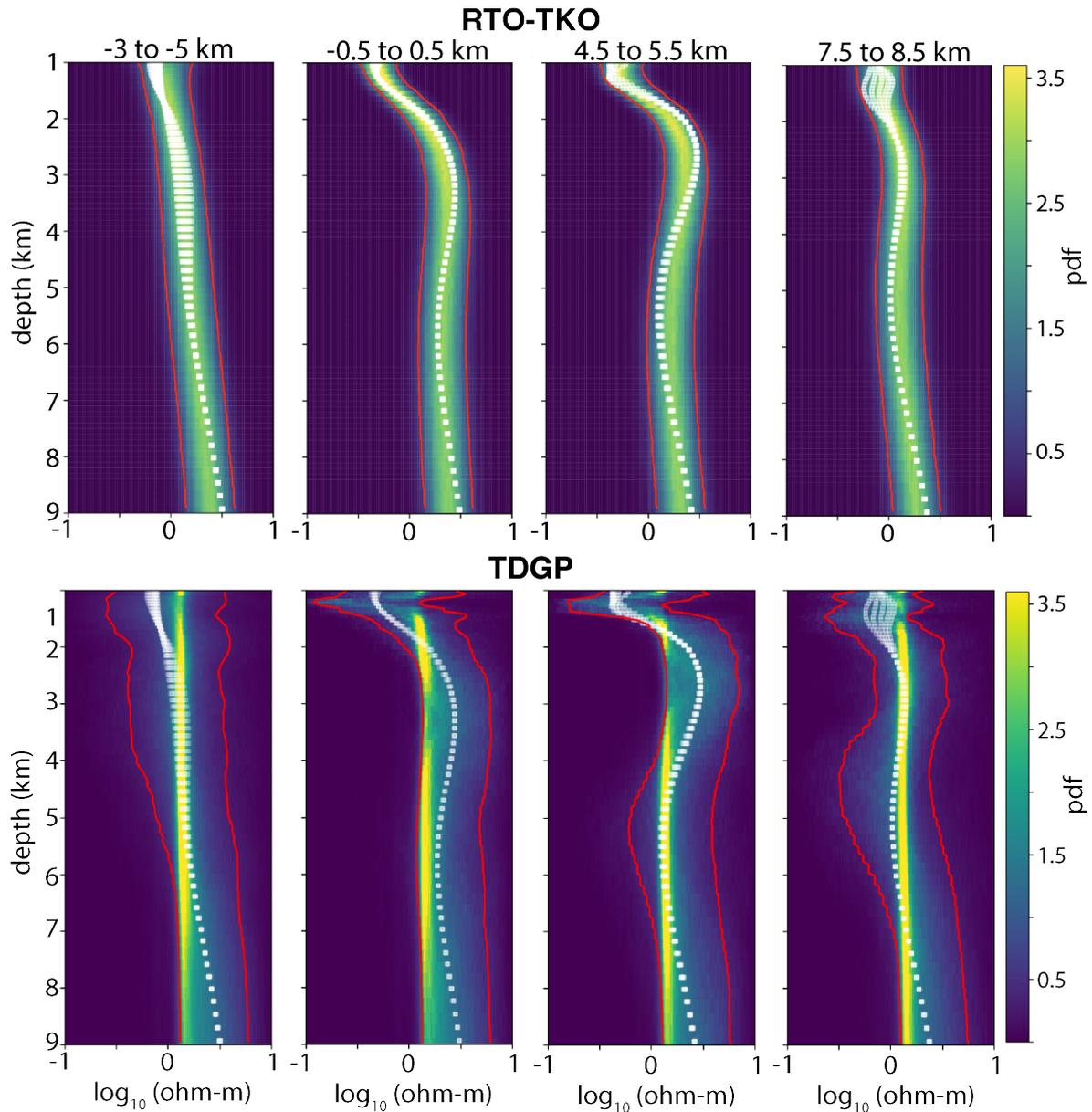


Figure 6. Marginal posterior probability distributions for resistivity as a function of depth over specified lateral position slices for both the RTO-TKO and TDGP model ensembles. Warm colors indicate regions of higher probability density, white squares indicate the Occam inversion model, and red lines delineate the 90% credible interval. Because the data (likelihood) are the same, the dissimilarities in the posterior reflect the different ways the algorithms parametrize and regularize the model.

the 90% credible interval produced by both algorithms is similar, the result of the information from the data common to both inversions. The differences between the RTO-TKO and TDGP posterior uncertainties are due largely to differences in the prior, particularly the very different

ways the two algorithms parametrize and regularize the model, and indicate that the data contain incomplete information about the subsurface.

RTO-TKO uses a fixed grid of resistivity cells (shown by the black grid lines in Fig. 4) and searches for smooth models that fit the data. “Smooth” is to be understood in view of the prior distribution of the resistivity, which is Gaussian with mean zero and covariance $C_m = \frac{1}{\mu}(L^T L)^{-1}$. Here, different values of the regularization penalty weight μ are explored based on a joint Bayesian posterior distribution, so that the resulting models fit the data and are appropriately regularized (but where the type of regularization is fixed by the choice of regularization matrix L).

TDGP on the other hand determines the resistivity value of the grid cells (both RTO-TKO and TDGP used the same model grid for forward calculations and *a posteriori* analysis) using GPs (Williams & Rasmussen 1996), which perform a stochastic interpolation based on a set of nodes whose number, location, and associated resistivity are determined during the inversion. The smoothness and degree of model structure in the TDGP models is determined largely by the correlation length (or “zone of influence”) of the interpolation nodes as well as the shape of the correlation kernel. We chose the correlation length to increase with depth to correspond with the decreasing resolution of MT data with depth and chose a kernel that decays exponentially with distance from the interpolation node. These choices produce models with smoothly varying ‘blobs’ centered around interpolation nodes or node clusters. The structure of the TDGP posterior is also determined by the number of nodes since points in the model at significant distance (in terms of the correlation length) from an interpolation node revert to the mean resistivity of the model, producing the characteristic resistivity stripes seen in Fig. 5 and Fig. 6 wherever the mean is sufficient to fit the data. These stripes loosely correspond to the probability stripes seen in the trans-D MCMC inversions in Fig. 1. In addition, Bayesian parsimony prefers TDGP models with fewer interpolation nodes, all else being equal. The TDGP algorithm thus samples the space of smooth, parsimonious models that fit the data, letting the resistivity revert to the model mean wherever it does not harm the data fit.

These two priors (RTO-TKO’s gradient smoothing against a random, smooth model vs. the smooth, parsimonious GPs) produce very different looking models. Fig. 7 shows four models chosen at random from each of their respective model ensembles. The RTO-TKO models show the imprint of the randomly chosen prior model, against which smoothing is performed. The TDGP models reflect the locations and resistivity of the interpolation nodes, as well as their correlation lengths, and the exponential decay of the influence kernel. What is striking is

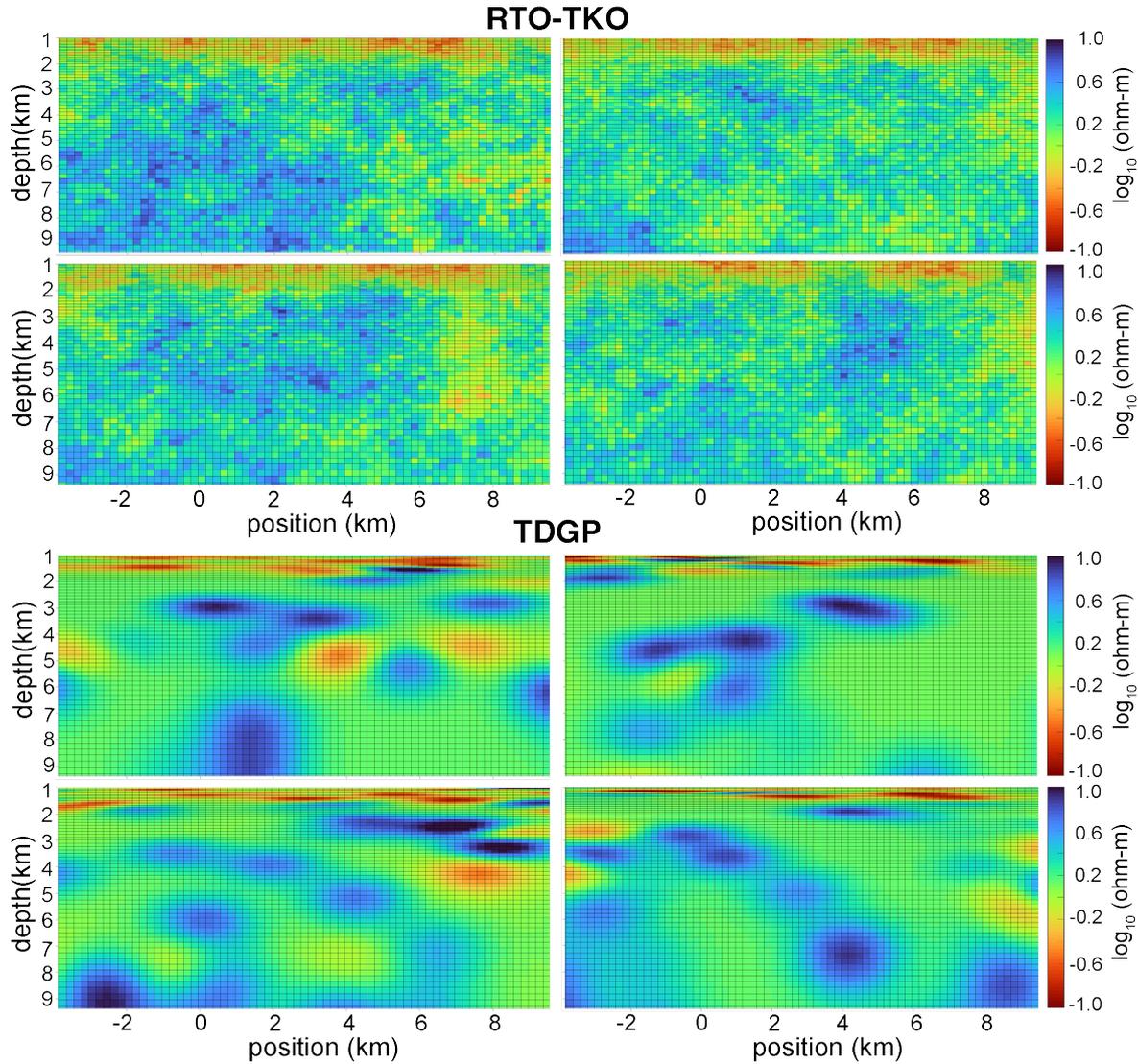


Figure 7. Example Gemini resistivity models chosen from the RTO-TKO and TDGP model ensembles. The RTO-TKO models exhibit the imprint of the randomly chosen prior model against which derivative smoothing is performed. TDGP models reflect the location and resistivity of interpolation nodes as well as the correlation kernel and its length scale. All models fit the data to within the measurement error. Consequently, differences in the model reflect differences in their respective priors.

that each of these models fits the data to within the data uncertainty. Since the information from the likelihood is the same for both RTO-TKO and TDGP, the pronounced differences in the models results entirely from differences in the prior. Yet since the prior range of allowed resistivity was the same in both cases (0.1-10 ohm-m), the differences evident in Fig. 7 are largely the result of the different ways these two algorithms parametrize and regularize the

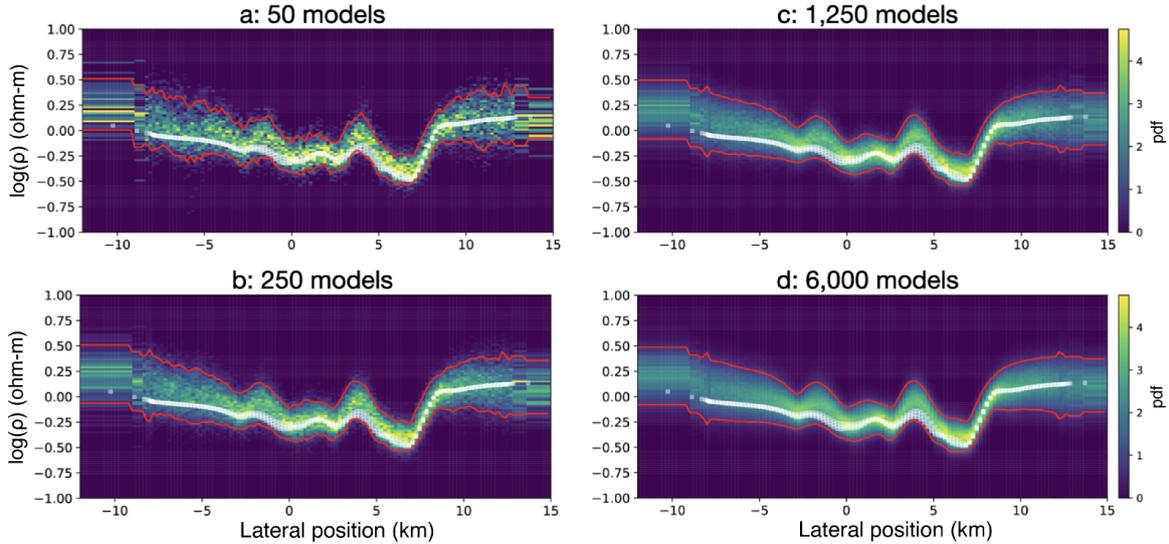


Figure 8. Marginal posterior uncertainty estimates from the RTO-TKO inversion of the Gemini MT data for the 1.2-1.4 km depth slice shown in Fig. 5. When only the first 50 RTO-TKO samples are used to make the estimate, the distribution looks ragged but the location and width of the 90% credible interval is already fairly accurately estimated. The distribution has fully converged by the time 1,250 samples have been drawn.

model. Thus, the RTO-TKO solution does not converge to the TDGP solution, since each is targeting a separate posterior due to their unique priors.

The fact that such disparate models can fit the data equally well is evidence that the MT data in this case do not contain complete information about the subsurface. In the absence of strong constraints from the data, the inversion algorithms fill in the gaps in information with our own prior conceptions of the subsurface. As we noted in Section 2, the heavy imprint of the prior is in some sense regrettable (we would certainly prefer the data be perfectly informative). Yet, it has the added benefit of obligating the practitioner to clearly state their prior assumptions and be mindful of their impact on the posterior uncertainty. Perhaps the greatest risk in UQ is not that the chosen algorithm will get the uncertainty ‘wrong’ in some sense, but that it will be the result of *a priori* choices of which the practitioner is not fully aware.

To assess whether the 8,831 RTO-TKO samples were sufficient to estimate the posterior uncertainty, we estimated the marginal distribution for the horizontal slice over the depth range 1.2-1.4 km shown in Fig. 5 using subsets of the total ensemble (Fig. 8). By the time the first 1,250 models in the RTO-TKO ensemble are used, the estimated posterior is smooth and only very minimal changes occur when the sample size is increased to 6,000 models. On

the other hand, while the probability density and 5th and 95th percentile estimates are rough when only the first 50 or 250 models were used, the shape and location of the 90% credible interval (the region spanned by 90% of the models in the ensemble) is already apparent. This means that a solid estimate of the range of parameter values compatible with the data and prior assumptions can be obtained with only a modest number of samples. We discuss the computational implications of this later in this study.

4 COMPARING LINEARIZED AND NONLINEAR UQ

Uncertainty estimates on inverted model parameters can be obtained without calling on the full machinery Bayesian sampling or the RTO-TKO. This is done by linearizing about a chosen reference model and computing an estimated Gaussian posterior covariance. Such linearized uncertainty estimates are relatively computationally cheap to compute but suffer from limitations. Interested readers are referred to Tarantola (2005). We discuss both the method of linearized UQ and its limitations below and compare to the nonlinear UQ from RTO-TKO.

The first step in obtaining linearized uncertainty estimates is to select a reference model. This is usually chosen to be the solution of deterministic inverse problem. Next, we assume that the forward problem is linear in the vicinity of the reference model. Making this crucial assumption means that the resulting uncertainty estimates are only valid in the vicinity of this model. It also means that the associated posterior distribution is now Gaussian (and hence symmetric) with mean given by the reference model (which we denote by $\hat{\mathbf{m}}$) and covariance given by

$$\hat{C}_m = (I - KG)C_m \quad (4)$$

where G is the linearized forward modeling operator, C_m is the prior model covariance, and $K = C_m G^T (G C_m G^T + C_d)^{-1}$ is called the Kalman gain (Kalman 1960). The meaning of Eq. 4 is that the posterior is equal to the prior minus the information gain from the data, represented by the product KGC_m . Since $\hat{\mathbf{m}}$ is already known (it must be known prior to computing the linearized uncertainty), defining the linearized, Gaussian posterior uncertainty amounts to computing \hat{C}_m , which we emphasize is only an approximation based on linearization of the true, nonlinear forward problem about a particular reference model.

To compute \hat{C}_m , we must first obtain C_m and G . The linearized forward problem Jacobian matrix, G , is an $n_d \times n_m$ matrix of first-order partial derivatives of the model responses with respect to the model parameters. This matrix is typically used to steer non-linear regularized

inversions towards good fitting models and thus can be reused for local linearized UQ. While it can be computed using using finite differences of forward responses about model perturbations, it is often significantly more efficient to either apply the chain-rule to the forward operator equations for layered 1D models (e.g., Key 2009) or to use the adjoint method for multidimensional models (e.g., McGillivray et al. 1994).

The prior covariance C_m is defined by the regularization parameter μ and the matrix L : $C_m = \frac{1}{\mu}(L^T L)^{-1}$. This means that linearized uncertainty estimates depend upon the regularization penalty weight, which (see Part I) has a strong influence on the posterior uncertainty. In our case, we chose μ to be the value selected by the Occam inversion, but this is by no means the only possible choice. This reliance on an *a priori* choice of regularization strength renders linearized uncertainty to some extent a matter of subjective preference: what uncertainty do you want (e.g., Constable et al. 2015)?

Furthermore, since L is typically a first or second derivative operator, L is not invertible by construction. We thus must use a pseudo-inverse of the matrix $\mu L^T L$ to define the prior covariance, C_m . Again, the choice of pseudo-inverse and the details of how the pseudo inverse is computed has a large effect on the resulting linearized uncertainty quantification.

One possible way to compute the pseudo inverse is to decompose $\mu L^T L$ into its orthonormal eigenvectors and their corresponding eigenvalues. For the 1D MT field data problem (see Section 2) with n_m model layers, L is a first order finite differencing scheme applied to adjacent layers such that it has rank equal to $n_m - 1$, since an integral (the inverse of a derivative) can only capture a function up to an unknown constant. As a result, one of the eigenvalues of $\mu L^T L$ is zero, which we omit, instead defining C_m by the $n_m - 1$ non-zero eigenvalues and corresponding eigenvectors of $\mu L^T L$. More specifically, let λ_j , $j = 1, \dots, n_m - 1$, be the non-zero eigenvalues with corresponding eigenvectors u_j . Under these assumptions, the prior covariance is

$$C_m = \frac{1}{\mu} \sum_{j=1}^{n_m-1} \frac{1}{\lambda_j} (u_j u_j^T). \quad (5)$$

Now that the prior covariance is in hand, the posterior covariance can be computed using Eq. 4. For small 1D problems, the required dense matrix and eigen-decomposition operations are trivial, while for large 2D and 3D problems, they may require significant computational resources. However, given that G is already used in the regularized inversion scheme, computing the linearized uncertainty via Eq. 4 should be tractable for 2D and 3D problem that can already be solved using a given inversion code. Another very similar approach, known as ‘‘Hessian UQ,’’ is described in Loose & Heimbach (2021). We implemented both the Kalman

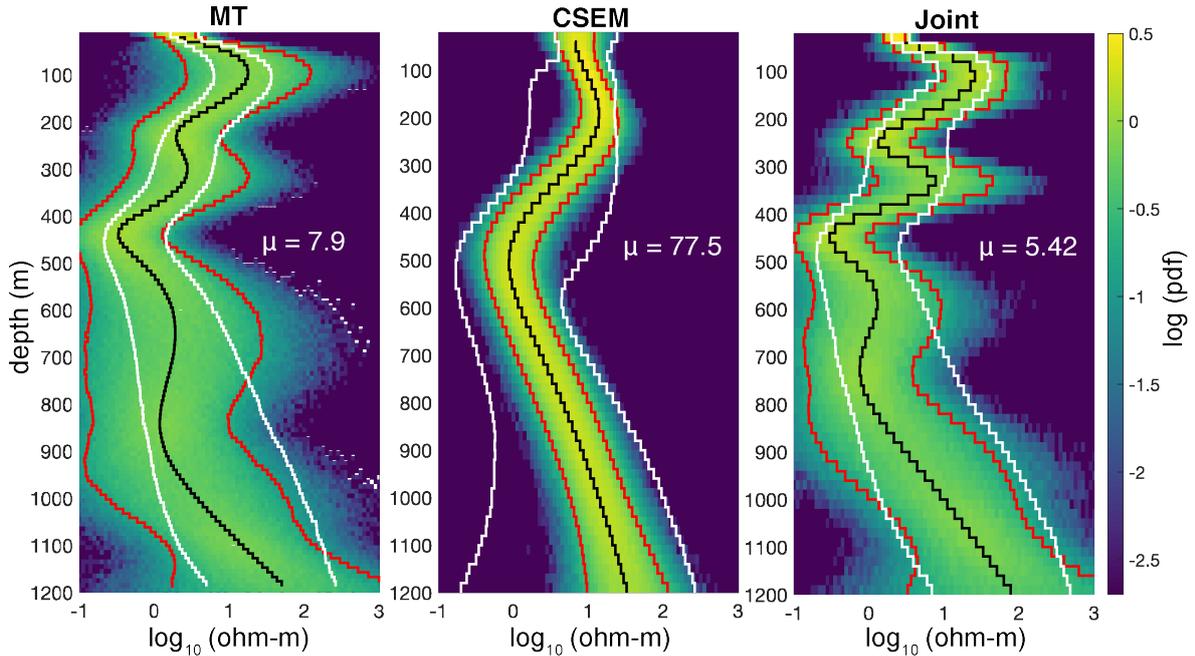


Figure 9. Linearized, Gaussian uncertainty (warmer colors indicate higher probability) estimated about the Occam inversion model (black line) for the MT, CSEM, and joint field data sets. The red lines delineate the 90% credible interval from the linearized estimate, while the white lines indicate the nonlinear RTO-TKO estimate of the 90% credible interval. The impact of the choice of regularization parameter (indicated on each plot) on the linearized uncertainty estimate is evident. The influence of the choice of reference model is equally apparent.

gain method and Hessian UQ and obtained the same results (when using the same recipe for the required pseudo inverse).

Using the Kalman gain method, we computed linearized posterior uncertainties about the Occam inversion model for the 1D MT, CSEM, and joint MT-CSEM field data problems. The regularization penalty strength used in each linearization is that determined by each respective Occam inversion. We used Eq. 4 to generate ten thousand posterior models for each \hat{C}_m . The estimated posteriors are plotted in Fig. 9 where here the warmer colors now indicate regions of higher *linearized* posterior probability and the black line is the Occam inversion result. The 90% credible interval estimated using RTO-TKO is plotted in white for reference.

The two main limitations of linearized UQ are evident in Fig. 9. First, the strong dependence of the posterior variance upon the choice of regularization strength is immediately apparent. In the MT field data and joint inversion problems, a relatively smaller regularization penalty weight results in a commensurately larger posterior variance, while in the CSEM case, larger regularization penalty leads to smaller posterior variance. This is actually counter-

intuitive, since the smaller regularization penalty weight in the MT and joint inversion cases reflects the MT data’s greater sensitivity to model structure (these data were collected in shallow water at relatively high frequencies) relative to the surface-towed CSEM data, which were limited to only two frequencies and four source-receiver offsets. Because the MT data are more sensitive, more model structure is needed to fit the data and the regularization penalty chosen by Occam is commensurately smaller. As a result, the linearized uncertainty method estimates higher uncertainty for the data set with higher resolution than for the data set with lower resolution.

While we chose μ to be the value found by the Occam inversion, this represents a subjective preference. A different choice would yield correspondingly different results. Intriguingly, the linearized posterior variance only depends on the data in so far as the data contributed to determining the reference model. Otherwise, it only depends on the prior and the size of the data uncertainty (see Eq. 4), making these choices all the more influential.

The absence of the data from the expression for the linearized posterior covariance suggests the other primary weakness of linearized uncertainty quantification: the dependence on the mean, or reference model. The only influence the data and nonlinear forward modeling might be allowed to have on linearized uncertainty is in the selection of the reference model, which serves as the mean of the posterior distribution. G is a linear approximation, and therefore valid only in the immediate vicinity of $\hat{\mathbf{m}}$. This means that solutions to the inverse problem (i.e. models that satisfy both the prior and the data) that diverge substantially from $\hat{\mathbf{m}}$ are not reflected in \hat{C}_m , the linearized posterior uncertainty. If there is little structure in the model (e.g. the CSEM posterior in Fig. 9), then perhaps one reference model will serve as well as any other. But it is relatively rare in complex models that all model structures are simultaneously well resolved. For instance, there are often trade-offs between model parameters, such that model structure in one region acts to compensate for additional structure in another. In such cases, linearized UQ, based on a single inversion model (itself a reflection of a choice of regularization strength), risks excluding other equally fit models and suggesting an overly optimistic assessment of the parameter uncertainty.

5 COMPUTATIONALLY EFFICIENT 2D MT UQ

There is always a significantly increased cost to obtaining a UQ for inverted model parameters. For many geophysical inverse problems, however, the cost of Bayesian sampling methods such as Metropolis-Hastings MCMC can be prohibitive. Even MCMC algorithms that go to great lengths to achieve efficiency—for instance, by using gradient information to tune the

proposal distribution (Zhao & Sen 2020) or by describing the model using a parsimonious parameterization (Ray 2021)—can still require run times of weeks, months or longer when applied to large problems (e.g. with 2D or 3D model geometries).

For example, the same Gemini MT field data set inverted in this study was inverted using TDGP, a trans-dimensional MCMC algorithm that achieves efficiency through a parsimonious model space representation through the use of Gaussian Processes. The parsimonious parameterization allows the TDGP algorithm to sample over fewer model parameters than would be necessary for fixed-grid methods such as RTO-TKO. Ideally, this enables quicker exploration of model space and faster, guaranteed convergence to its associated Bayesian posterior distribution.

Even so, and with an average cost per MCMC iteration of only 0.85 sec, inverting the Gemini data set using the TDGP algorithm took ten days running on 168 cores, or roughly 40,320 core hours (Blatter et al. 2021). Crucially, adding more cores would not have decreased the total run time, since in order to ensure convergence to the posterior it was determined that each Markov chain needed to explore the model space for a minimum of one million steps. While each forward computation was done in parallel across multiple cores to reduce the cost of forward modeling as much as possible, each MCMC step had to be taken in series, one after another, such that one million steps at 0.85 sec/step would necessarily cost roughly ten days of compute time.

By contrast, our RTO-TKO inversion of the Gemini data set was accomplished using ten independent samplers—one compute node of 24 cores per sampler—drawing a total of 8,831 samples. The total run time was 48 hours and the compute cost was 11,520 core hours. While the cost reduction in terms of core hours is itself impressive, the real efficiency gain is in the ability to reduce the run time by harnessing more HPC resources. This is due to the parallel nature of RTO-TKO, which draws (nearly) independent samples. For instance, if we had utilized 100 samplers instead of ten to draw the same number of samples (8,831), the total cost in core hours would have remained the same, yet the inversion would have been completed in just under five hours—fast enough to run to completion within a standard work day. All that is required to reduce the run time cost of RTO-TKO is simultaneous access to more processors—something that is becoming more widely available with time. And we emphasize that this does not increase the overall computational cost in terms of core hours.

The run time bottleneck suffered by Metropolis-Hastings MCMC becomes more severe as the problem size increases. For instance, consider a small 3D MT data set such that computing the forward response requires one minute on 100 cores. To draw one million samples using

MCMC would take nearly 700 days of compute time! Frustratingly, the only way to reduce this impractical run time is to reduce the cost of each forward computation. Yet even a 50% reduction in the run time cost of each forward calculation—representing a significant achievement given that most large forward solvers have already been optimized for efficiency—would result in a run time of 350 days, almost an entire year.

MCMC can take advantage of HPC in some ways. Running multiple Markov chains in parallel on separate computers can increase the rate at which samples are drawn, but the effectiveness of this is limited by two factors. First, each chain will have to overcome the burn-in period (the sequence of initial samples with low posterior probability that are ultimately discarded) on its own. For large problems, the burn-in can be hundreds of thousands of samples long (for the 2D MT inversion using TDGP, it was 300,000). Thus, each MCMC chain must be *at least* that long before even beginning to draw posterior samples.

Second, the post burn-in portion of each chain must reach stationarity independently or risk biasing the solution away from the target posterior by an unknown amount. Thus, for instance, 10,000 Markov chains with a post burn-in of length of 100 samples are not equivalent to one Markov chain with a post burn-in length of 1,000,000 samples.

Algorithms do exist that allow multiple Markov chains run in parallel to communicate with one another in order to enhance the speed and robustness of convergence, such as parallel tempering (PT) (Sambridge 2014). However, these algorithms have limitations of their own. Each chain must still overcome a burn-in period and the samples from most of the chains are discarded because they sample from a ‘tempered’ likelihood. Thus, attempts to run MCMC chains in parallel do not alter the fundamentally serial nature of the algorithm.

The difference in the size of the respective model ensembles (8,831 models in the RTO-TKO ensemble compared to one million models in the TDGP ensemble) is due to differing ways the two algorithms explore the model space. MCMC algorithms like TDGP explore the model space slowly, taking incremental, local steps with high degrees of correlation between nearby models in the ensemble (see Part I of the paper series). RTO-TKO, on the other hand, produces (nearly) independent samples. As a result, RTO-TKO does not suffer from a long burn-in period, the models in the chain prior to the algorithm finding high probability regions of model space, and that are ultimately discarded. Nor does it need to draw a prohibitively large number of samples to be certain of convergence.

To demonstrate the power of drawing (nearly) independent samples, we chose a model parameter at random from the same dense resistivity grid used for forward calculations in both the RTO-TKO and TDGP Gemini inversions (see pink dot in Fig. 4). This parameter

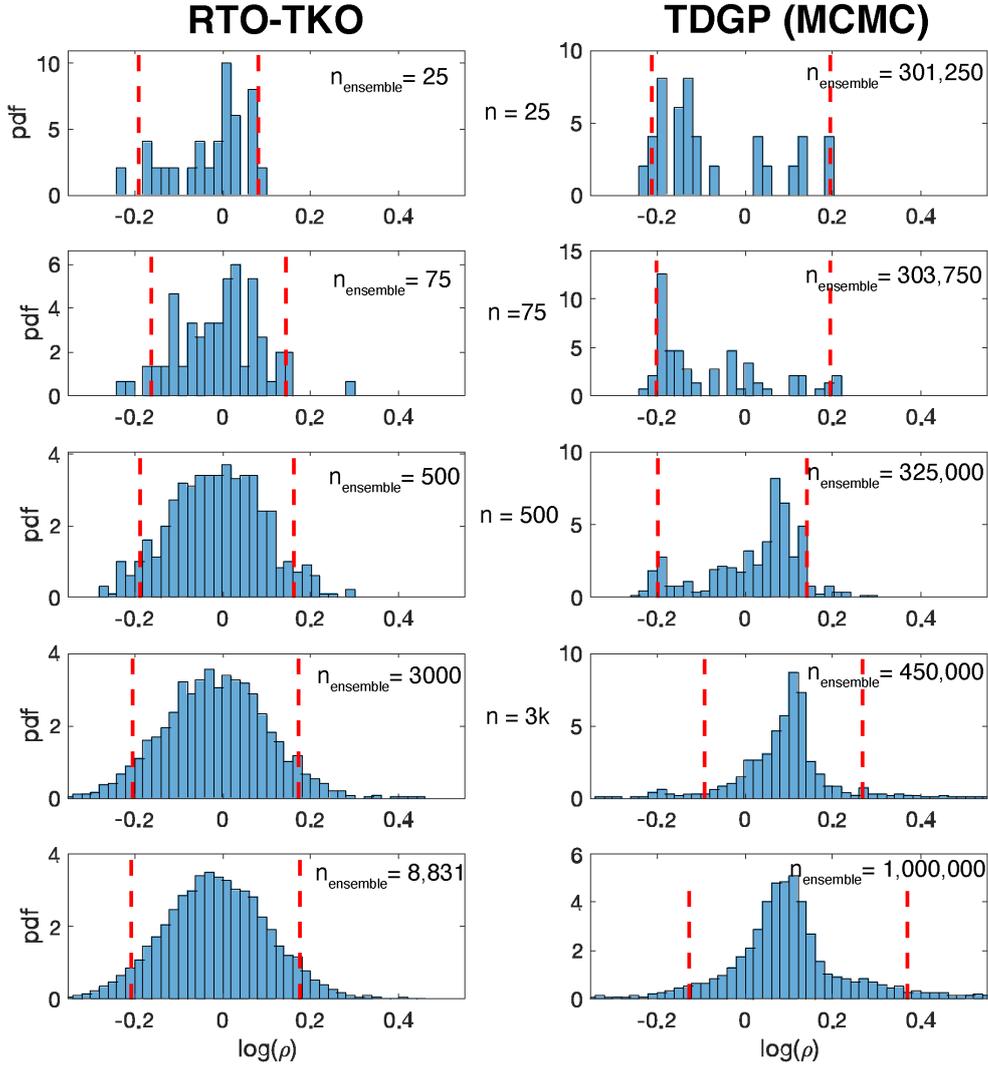


Figure 10. Uncertainty estimates for a model parameter chosen at random from the same dense resistivity grid used for forward calculations in both the RTO and TDGP inversions of the Gemini data set. The location of this parameter is indicated by the pink circle in Fig. 4, is within the MT station array, and is at a depth that should be well resolved by the data. The first n samples (indicated in the center of each row) of each model ensemble were used to make each row of estimates. The total number of samples that had to be drawn by each algorithm to produce each estimate ($n_{ensemble}$) is shown in the upper-right corner of each plot (the TDGP $n_{ensemble}$ includes burn-in and the fact that only every 50th model was saved, for computer memory reasons). The 5th and 95th percentiles of each distribution are indicated by the left and right red dashed lines, respectively. The last row of plots shows the uncertainty estimate made with each respective full model ensemble.

is within the coverage of the MT array and at a depth that should be well resolved by the data. We estimated the posterior uncertainty for this model parameter using the first $n_{ensemble}$ samples of the RTO and TDGP model ensembles (left and right columns of Fig. 10, respectively). The pair of estimates in each row of Fig. 10 were made using the same number of models (indicated by n in the center of each row). The total number of samples required to make each posterior uncertainty estimate ($n_{ensemble}$) is indicated in the upper-right corner of each plot. Because of the burn-in (estimated to be the first 300,000 models of the TDGP ensemble) and the fact that, due to computer memory constraints, only every 50th TDGP model was saved, the total number of samples required is much higher for TDGP than for RTO.

The final row of plots in Fig. 10 contains the posterior estimate for this model parameter made using the full RTO and TDGP model ensembles, respectively. We again emphasize that, because the two algorithms start from different prior assumptions about the subsurface, they are converging to different posterior distributions. Even so, the bulk of the probability density is concentrated between 0.6-2.5 ohm-m, though there is considerable difference in the tails of the distributions.

Two important, computational conclusions can be drawn from Fig. 10. The first is that only a small number of RTO samples is required to obtain a reliable estimate of the 90% credible interval (bounded by the 5th and 95th percentiles, shown as red dashed lines). While the shape of the RTO-TKO distribution looks rough until more than a thousand samples have been drawn, the estimates of the 90% credible interval remain remarkably consistent, even after only 25 samples. This is not true of MCMC methods such as TDGP, where the estimates of the 90% credible interval change considerably as the sampler slowly explores the model space. The second is that far fewer total samples are required for the RTO-TKO ensemble to converge than for the TDGP ensemble, in part due to the burn-in (in this case, the first 300,000 TDGP models) and in part due to the slow, incremental way M-H MCMC algorithms explore the model space.

The efficiency of MCMC sampling can be improved by fine-tuning the proposal distribution to match the posterior as closely as possible. This is challenging, however, since the M-H MCMC proposal is by definition local while the posterior varies spatially. What's more, tuning the proposal can itself be an extremely time- and HPC resource-consuming task. (The TDGP posterior was fine-tuned at the cost of weeks' worth of time and many tens of thousands of core hours of trial runs). The RTO-TKO proposal, however, is by its very design global rather than local and always produces models of high posterior probability, such that it needs to

tuning at all other than what is necessary to make the optimizer work (something the EM geophysics community is already adept at doing).

It bears noting that the computational comparison being made here is fair in the sense that the underlying resistivity grid used for forward computations was identical in both cases. The only difference is that the TDGP algorithm used a parsimonious model representation to explore the model space more efficiently, while the RTO-TKO explored the model space using the same dense resistivity grid as used to compute the forward response of each model.

The unique ability of RTO-TKO to produce reliable estimates of uncertainty (e.g. the 90% credible interval, interquartile range, etc) with only a small number of samples brings uncertainty quantification for 3D MT problems within reach. Consider again the hypothetical 3D MT problem discussed above. If a lightly-parameterized regularized inversion of this data set takes 30 minutes on 100 cores, then 1,000 cores (representing 10 independent samplers) could draw roughly 500 RTO-TKO samples in a single day. While this is likely not enough to generate smooth uncertainty distributions like the ones in Fig. 5, it may well be enough to obtain a solid estimate of the 90% credible interval, as indicated by the convergence test results in Fig. 8 and Fig. 10.

6 CONCLUSIONS

Geophysical inversion algorithms that can provide quantitative, nonlinear uncertainty on the inverted model parameters in a reasonable amount of time are essential to better understanding the subsurface. We described, in a two-part series, a mathematical and computational framework for computing nonlinear UQ for regularized models (which are widely used in geophysics, particularly for EM methods). The key insight behind our approach, which we call RTO-TKO, is that the well-known and robust machinery of regularized inversion can be put to use for nonlinear UQ. More specifically, RTO-TKO is easy to use because it relies on well-understood algorithmic infrastructure (regularized inversion codes), requires little to no tuning, automatically adjusts the regularization in view of the data, and is computationally efficient and easy to parallelize. We described the mathematics and rationale behind RTO-TKO in Part I.

Part II (this paper) focuses on practical aspects of RTO-TKO. We illustrated the applicability and versatility of RTO-TKO by inverting MT and CSEM data separately and jointly for 1D resistivity models. We demonstrated the computational efficiency of RTO-TKO by inverting field MT data for 2D resistivity models. We further compared the computational cost of RTO-TKO to TDGP, a parsimonious trans-dimensional MCMC algorithm and found

that RTO-TKO is more efficient than TDGP, both in total flops and total run time (reducing run-time from ten days to two days). What’s more, the (near) independence of the RTO-TKO samples means that a solid estimate of the range of parameter values compatible with the data and prior assumptions (e.g. the 90% credible interval) can be obtained in even less time, so that nonlinear UQ for 3D MT problems is within reach (and largely dependent on access to a large computer and a robust 3D deterministic inversion code).

We further studied the uncertainty estimates of RTO-TKO and compared the results to those of trans-dimensional MCMC samplers and to a linearized uncertainty quantification (largely relying on the Occam inversion). We found that all three methods give different UQs. In the case of the linearized UQ, the differences are largely due to the linearization of an underlying nonlinear problem, and thus the results are not very useful. In the case of the nonlinear techniques, the differences in uncertainties are due to different sets of prior assumptions, especially in how the models are parametrized and regularized—by interpolation between a sparse set of nodes in the case of TDGP, and by regularization across a dense mesh in the case of RTO-TKO. These prior assumptions have a large effect on the posterior uncertainty estimates because the data do not contain sufficient information to fully constrain all model parameters. In the absence of constraint from the data, prior assumptions are propagated more or less directly into posterior uncertainty estimates.

Thus, we should not expect that one approach to UQ is “right” and others are “wrong”, and we should not view a UQ as providing global bounds on what is allowed by the data. Rather, the UQ should be carefully interpreted with prior assumptions (regularization and model parameterization) in mind. For example, for two of the cases we have considered in this work, the 1D MT and 1D resistivity problems, we know that acceptable models (indeed, the *best* models in a least squares sense) are pathological. These include infinite conductivities in the 1D MT case (Parker & Whaler 1981) and arbitrarily thin layers of extreme conductivity in the 1D resistivity case (Parker 1984). Likely, undiscovered pathologies exist for the higher dimensional versions of these methods. But this is exactly why a nonlinear UQ is useful—these types of models have zero probability of existing in the real world and can be ruled out by employing additional prior assumptions.

While there exists no clear-cut answer to which UQ is “best,” the numerical/computational machinery of the RTO-TKO provides an efficient way of computing nonlinear UQ under a given set of (prior) assumptions. These assumptions are in line with the set of explicitly regularized models. Regularized inversion algorithms, such as Occam, have proved very useful in practice and most geophysical inversions to date rely on them. Yet obtaining a UQ on

such models has remained elusive, especially in view of nonlinearity and the large amount of covariance between model parameters that is implicit in regularization. RTO-TKO provides, perhaps for the first time, a nonlinear UQ for regularized models that is computationally efficient enough to be useful in 2D problems and, with access to a large computer, to 3D geophysical inverse problems.

Finally, we stress that while we have demonstrated RTO-TKO’s usefulness for diffusive EM methods, absent any mathematical theory constraining the size of the RTO-TKO ‘bias’ we cannot guarantee similar success for other geophysical methods. Further study is necessary to determine the breadth of this UQ algorithm’s applicability.

DATA AVAILABILITY

The data and computer codes underlying this article will be shared on request to the corresponding author.

Acknowledgements

DB is supported by the Green Foundation’s John W Miles postdoctoral fellowship in theoretical and computational geophysics. MM is supported by the US Office of Naval Research (ONR) grant N00014-21-1-2309.

REFERENCES

- Agostinetti, N. P., Giacomuzzi, G., & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophysical Journal International*, **201**(3), 1598–1617.
- Blatter, D., Key, K., Ray, A., Foley, N., Tulaczyk, S., & Auken, E., 2018. Trans-dimensional Bayesian inversion of airborne transient EM data from Taylor Glacier, Antarctica, *Geophysical Journal International*, **214**(3), 1919–1936.
- Blatter, D., Key, K., Ray, A., & Gustafson, C., 2019. Bayesian joint inversion of controlled source electromagnetic and magnetotelluric data to image freshwater aquifer offshore New Jersey, *Geophysical Journal International*, **218**(3), 1822–1837.
- Blatter, D., Key, K., & Ray, A., 2021. Two-dimensional Bayesian inversion of magnetotelluric data using trans-dimensional Gaussian processes, *Geophysical Journal International*, **226**(1), 548–563.
- Constable, S., Orange, A., & Key, K., 2015. And the geophysicist replied: “Which model do you want?”, *Geophysics*, **80**(3), E197–E212.

- Constable, S. C., Parker, R. L., & Constable, C. G., 1987. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, **52**(3), 289–300.
- Gustafson, C., Key, K., & Evans, R. L., 2019. Aquifer systems extending far offshore on the U.S. Atlantic margin, *Scientific Reports*, **9**(1), 1–10.
- Kalman, R. E., 1960. A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, **82**(1), 35–45.
- Key, K., 2009. 1D inversion of multicomponent, multifrequency marine CSEM data: Methodology and synthetic studies for resolving thin resistive layers, *Geophysics*, **74**(2), F9–F20.
- Key, K., 2016. MARE2DEM: a 2-D inversion code for controlled-source electromagnetic and magnetotelluric data, *Geophysical Journal International*, **207**(1), 571–588.
- Key, K. W., Constable, S. C., & Weiss, C. J., 2006. Mapping 3D salt using the 2D marine magnetotelluric method: Case study from Gemini Prospect, Gulf of Mexico, *Geophysics*, **71**(1), B17–B27.
- Loose, N. & Heimbach, P., 2021. Leveraging Uncertainty Quantification to Design Ocean Climate Observing Systems, *Journal of Advances in Modeling Earth Systems*, **13**(4), e2020MS002386.
- MacKay, D., 2003. *Information Theory, Inference and Learning Algorithms*, Cambridge University Press.
- McGillivray, P. R., Oldenburg, D. W., Ellis, R. G., & Habashy, T. M., 1994. Calculation of sensitivities for the frequency-domain electromagnetic problem, *Geophysical Journal International*, **116**(1), 1–4.
- Parker, R. L., 1984. The inverse problem of resistivity sounding, *Geophysics*, **49**(12), 2143–2158.
- Parker, R. L. & Whaler, K. A., 1981. Numerical methods for establishing solutions to the inverse problem of electromagnetic induction, *Journal of Geophysical Research: Solid Earth*, **86**(B10), 9574–9584.
- Ray, A., 2021. Bayesian inversion using nested trans-dimensional Gaussian processes, *Geophysical Journal International*, **226**(1), 302–326.
- Ray, A. & Myer, D., 2019. Bayesian geophysical inversion with trans-dimensional Gaussian process machine learning, *Geophysical Journal International*, **217**(3), 1706–1726.
- Sambridge, M., 2014. A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophysical Journal International*, **196**(1), 357–374.
- Selway, K., O'Donnell, J. P., & Özaydin, S., 2019. Upper Mantle Melt Distribution From Petrologically Constrained Magnetotellurics, *Geochemistry Geophysics Geosystems*, **20**(7), 3328–3346.
- Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics.
- Ward, S. H. & Hohmann, G. W., 1987. 4. Electromagnetic Theory for Geophysical Applications, in *Electromagnetic Methods in Applied Geophysics*, pp. 130–311, Society of Exploration Geophysicists.
- Williams, C. K. I. & Rasmussen, C. E., 1996. Gaussian processes for regression, in *Advances in neural information processing systems*, pp. 514–520, eds Touretzky, D. S., Mozer, M. C., & Hasselmo, M. E.
- Zhang, X., Curtis, A., Galetti, E., & de Ridder, S., 2018. 3-D Monte Carlo surface wave tomography,

Geophysical Journal International, **215**(3), 1644–1658.

Zhao, Z. & Sen, M. K., 2020. A gradient-based Markov chain Monte Carlo method for full-waveform inversion and uncertainty analysis, *Geophysics*.

APPENDIX A: SYNTHETIC INVERSIONS IN 1D AND 2D

We present synthetic inversion results in 1D and 2D. Both examples illustrate additional and interesting features of RTO-TKO. In this context, recall that RTO-TKO solves two stochastic optimization problems per iteration. The first, and more costly, solves for the model parameters at constant regularization penalty weight (Eq. 1). The second (smaller) stochastic optimization problem solves for the scalar regularization penalty weight by minimizing

$$\min_{\mu} g(\mu) = \frac{1}{2} \left\| C_d^{-1/2} \left(\mathbf{F} \left(\sqrt{\frac{\mu_{ref}}{\mu}} \mathbf{m} \right) - \tilde{\mathbf{d}} \right) \right\|^2 - \log(p(\mu)) \quad (\text{A.1})$$

where $p(\mu)$ is the prior distribution on μ , μ_{ref} is the regularization penalty weight used in the first stochastic optimization step (Eq. 1) and the model parameters are held constant. In Eq. A.1, smaller values of μ uniformly increase the roughness of the model, while larger values decrease the model roughness. The solution to Eq. A.1 is then used during the minimization of f in Eq. 1 during the next iteration of the algorithm, allowing RTO-TKO to hierarchically sample the regularization penalty weight. This is important, as μ can have a large impact on the posterior uncertainty (for more details, see Part I of this two-part series).

What is left to discuss is the prior on μ and its effects on the joint posterior distribution (and this is the focus of this appendix). Ideally, one wants to use uninformative priors to see all effects of the data. Thus, one may choose a (wide) uniform distribution or a log-normal distribution with a large variance, as we did in the inversions of field data. Uninformative priors, however, can be problematic if the data do not contain enough information about all model and regularization parameters. If this is indeed the case, one should choose a (slightly) more informative prior on μ to stabilize the hierarchical inversion. Below, we show two examples where an instability occurs because the data are not informative and how to fix these issues by defining appropriate priors for μ .

A1 Synthetic 1D MT inversion

We generate a synthetic data set (Fig. A1a-b) by adding 5% Gaussian noise to the MT (apparent resistivity and phase) model responses from a simple, 1D model of a thin, conductive layer in a resistive half-space (Fig. A1c). We then invert this data using RTO-TKO. Fig. A2a shows the hierarchically sampled distribution for μ when an uninformative prior is used. The

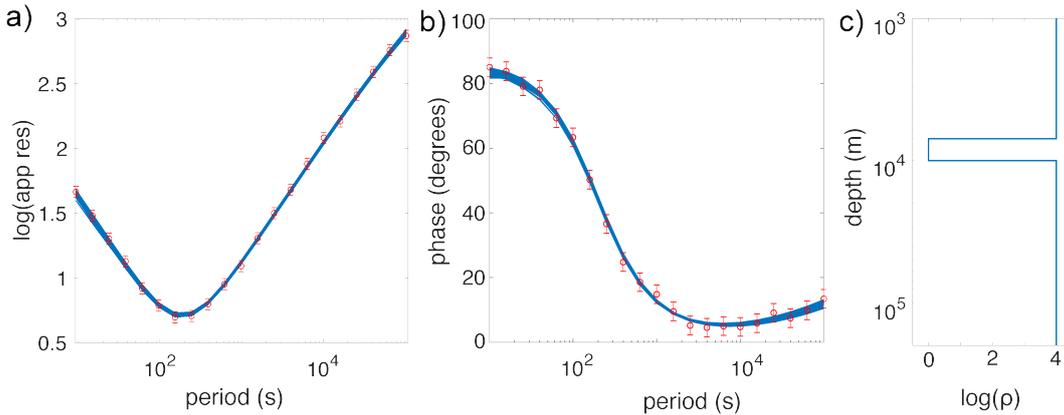


Figure A1. Synthetic MT apparent resistivity (a) and phase (b) data generated from the model on the right (c) of a thin conductor buried at depth within a resistive halfspace. 5% Gaussian noise was added to the model responses. The model responses of 200 randomly selected RTO-TKO inversion models are plotted on top of panels (a) and (b) in blue.

unconstrained regularization penalty weight drops steadily from its initial value until it hits the minimum allowed value (in this case, 10^{-6}). Fig. A3a shows the posterior uncertainty for the model parameters as a function of depth estimated using this model ensemble. This estimate of the posterior exhibits a ragged and uneven distribution of probability density symptomatic of a poorly constrained linearized inverse problem.

If instead we assume a log-Gaussian prior on μ

$$p(\mu) \propto \exp\left(-\frac{\mu - \bar{\mu}}{2\sigma_\mu}\right) \quad (\text{A.2})$$

where all the quantities in Eq. A.2 are expressed in base-ten log units, $\bar{\mu}$ is the log-mean and σ_μ the log-standard deviation, then the ‘TKO’ optimization step over μ becomes

$$\min_{\mu} \hat{g}(\mu) = \frac{1}{2} \left\| C_d^{-1/2} \left(\mathbf{F} \left(10^{0.5(\mu_{ref} - \mu)} \mathbf{m} \right) - \tilde{\mathbf{d}} \right) \right\|^2 + \frac{1}{2\sigma_\mu} \left\| \left(\mu - \bar{\mu} \right) \right\|^2 \quad (\text{A.3})$$

where, again, μ , $\bar{\mu}$, σ_μ , and μ_{ref} are all in base-ten log units, and μ_{ref} is the regularization penalty weight used in the previous, fixed- μ optimization step over the model parameters. In Eq. A.3, the log-Gaussian prior on μ plays the role of model regularization, with σ_μ playing the role of regularization penalty weight. With this regularization term stabilizing the inverse problem, repeated solutions to Eq. A.3 produce a stationary distribution for μ (see Fig. A2b-c) whose mean and variance depend on the choices of $\bar{\mu}$ and σ_μ . The corresponding posterior uncertainties are shown in Fig. A3b-c. As the choice of log-mean and log-standard deviation affect the posterior uncertainty, care must be put into making them.

In selecting $\bar{\mu}$ and σ_μ , it can help to determine upper and lower bounds on μ . A natural

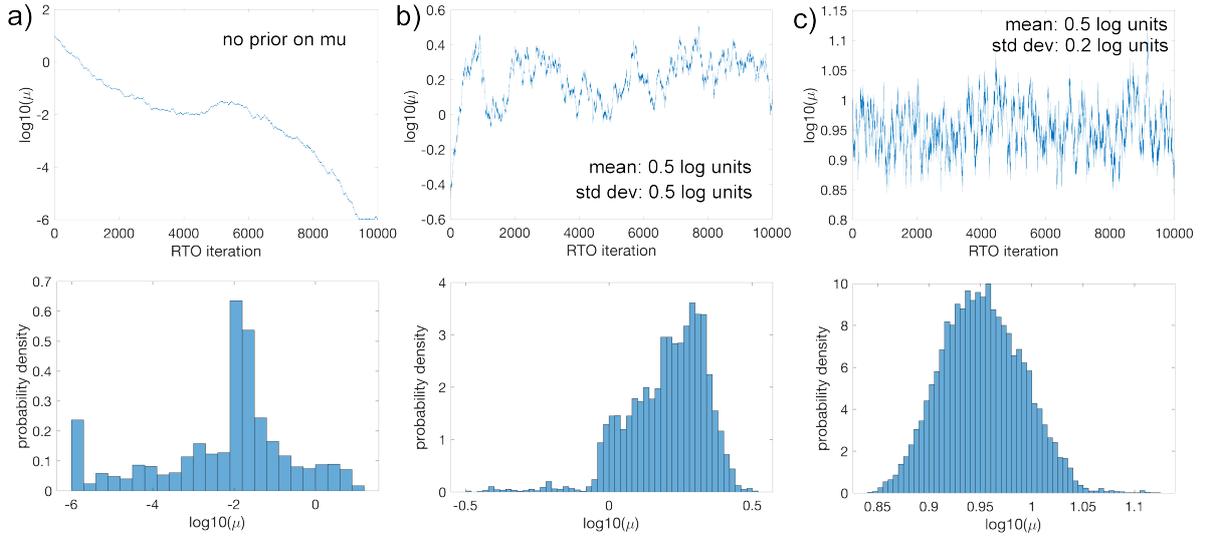


Figure A2. Regularization parameter, μ , for RTO-TKO inversions applying varying degrees of prior information. (a) In the absence of any prior constraint, μ decreases until it hits a minimum allowed value. (b) When a Gaussian prior with mean 0.5 log units and standard deviation 0.5 log units is applied, μ decreases from its starting value until it reaches a stable distribution centered around -0.6 log units. (c) When the standard deviation is decreased to 0.2 log units, the stationary distribution centers around 0.25 log units.

upper bound on μ occurs when roughness is penalized to such an extent that the resulting models no longer fit the data to within the measurement error, or when the data fit starts to deteriorate significantly. A natural choice for the lower bound is suggested by the prior covariance, $C_m = \frac{1}{\mu}(L^T L)^{-1}$, which is a function of μ . As discussed in section 2, the same constraints that apply to the model parameters \mathbf{m} equally apply to the prior model $\tilde{\mathbf{m}}$. Smaller values of μ lead to larger prior model covariance. But since we draw prior models by solving the constrained linear optimization in Eq. 3, it makes little sense to allow μ to drop so low that the bulk of the prior covariance lies outside the prior bounds on the model. At that point, it would be the prior model bounds constraining the roughness of $\tilde{\mathbf{m}}$ rather than μ . Once the lower and upper bounds on μ have been selected, a logical choice for the mean of the log-Gaussian prior is the midpoint (in log units) between them. One choice for the standard deviation (also in log units) is the largest value that produces a stationary distribution for μ that lies well within the upper and lower bounds.

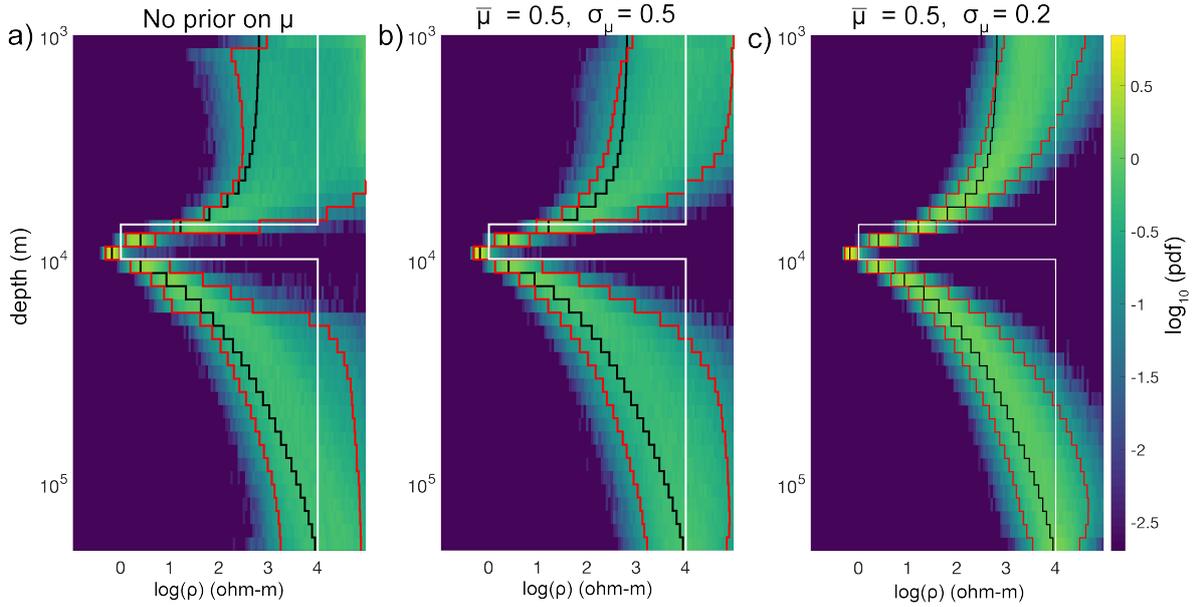


Figure A3. Estimated uncertainty in electrical resistivity as a function of depth. Warmer colors indicate higher probability, cooler colors lower probability. The left and right red lines are the 5th and 95th percentiles, respectively, of the probability distribution at each depth, while the black line is the Occam inversion (Constable, 1987) result. The white line represents the true model. (a) when no prior on μ is specified, the posterior can exhibit behavior typical of overfitting. (b-c) When a Gaussian prior of varying width is imposed, the width partially determines the posterior uncertainty. Compare with Fig. A2.

A2 Synthetic 2D MT inversion

We generate a synthetic MT data set by computing model responses (apparent resistivity and phase) for 5 MT stations spaced 50 km apart above a model consisting of a rectangular, conductive 1 ohm-m anomaly 40 km thick and 40 km wide at a depth of 40 km, buried in a resistive 100 ohm-m half-space. The modeled periods span the range 10-100,000 sec. We add 5% Gaussian noise to the model responses to simulate measurement error. The true model is shown in Fig. A4a. The synthetic data are plotted in Fig. A5.

Inverting this data set with the RTO-TKO algorithm with an uninformative prior on μ (a wide, uniform distribution) is unstable, evidence that the data and the prior bounds on the subsurface resistivity do not contain sufficient information about the appropriate value of this parameter. To stabilize the hierarchical inversion for μ (the TKO step), we specified maximum and minimum values for μ of 0.25 and 1.8 log units, respectively. These were chosen based on the guidelines suggested in Appendix A1. In addition, we specified a Gaussian prior on μ with mean equal to 1.025 log units (the midpoint of the upper and lower bounds) and a standard

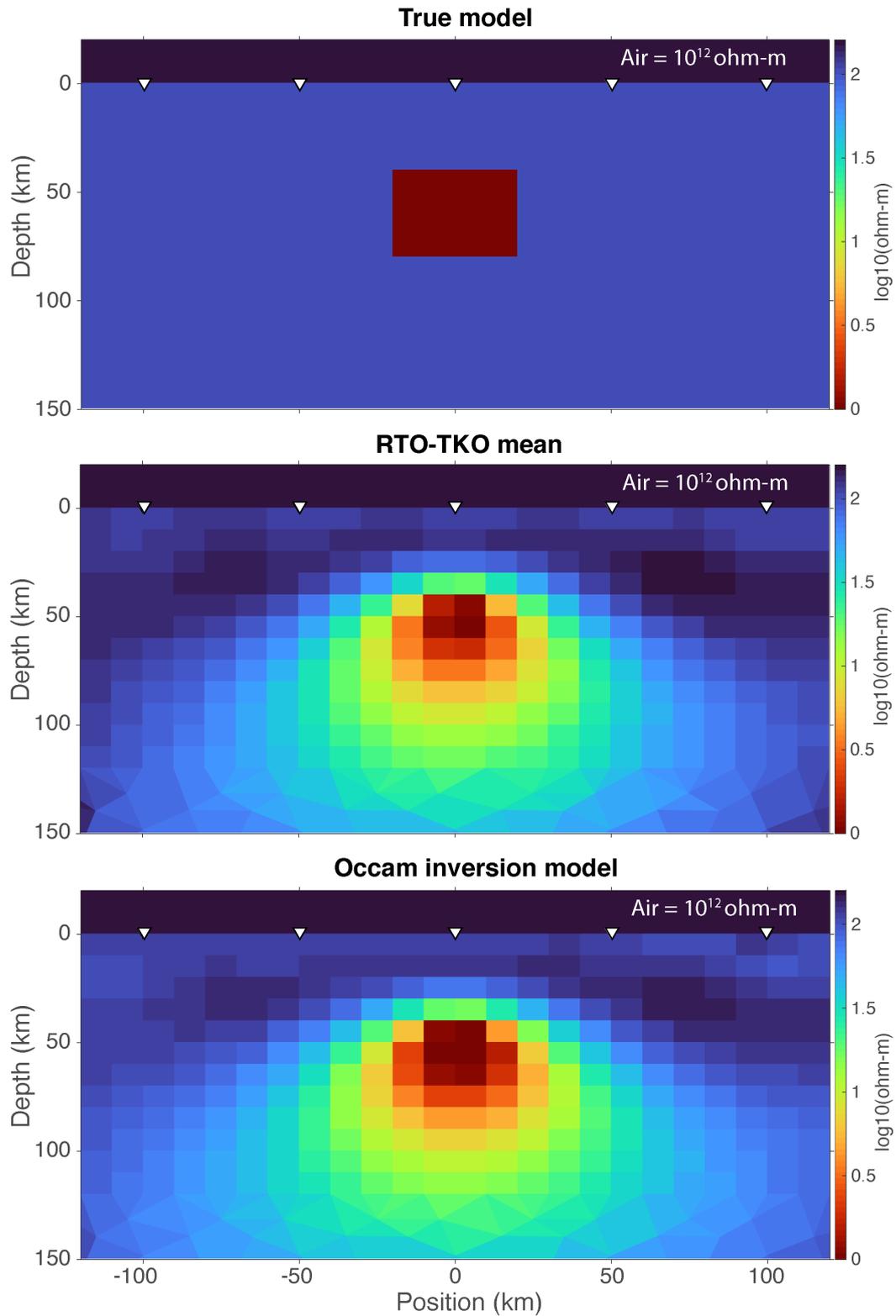


Figure A4. Resistivity model from which synthetic MT data are generated (see Fig. A5) consists of a conductive anomaly in a resistive half-space. MT stations are shown as white triangles. The mean of an RTO-TKO ensemble of 4,277 models recovers a smoothed version of the conductive anomaly as well as the resistive half-space. Only very minor differences exist between the RTO-TKO mean and the Occam inversion result.

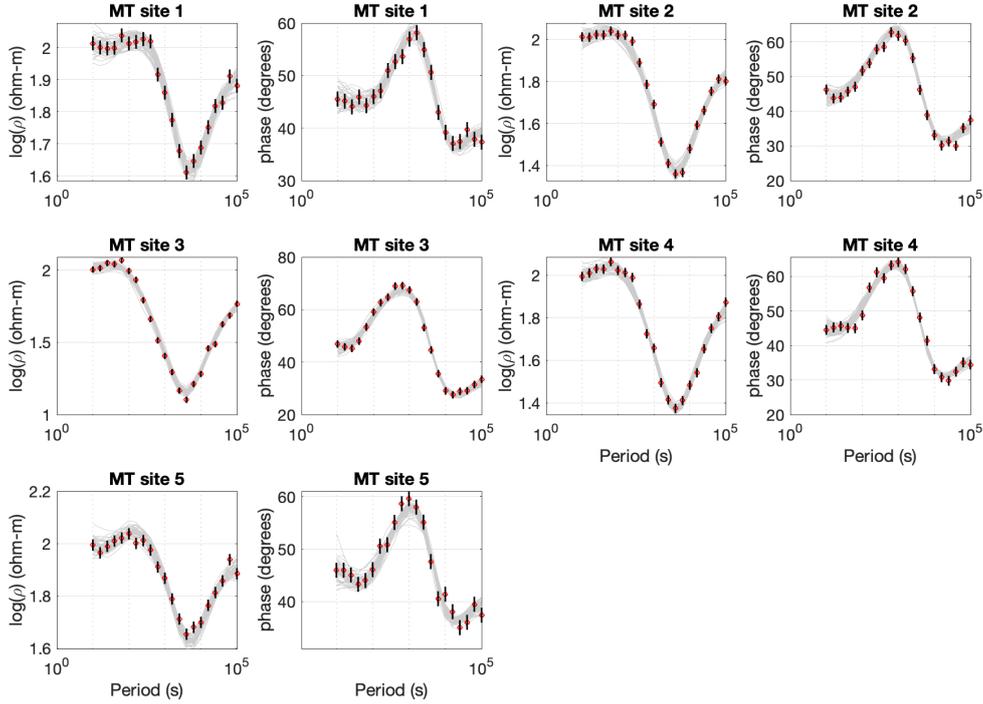


Figure A5. Synthetic MT data (apparent resistivity and phase) generated for 5 MT stations (white triangles in Fig. A4) over the period range 10-100,000 sec. Model responses from 50 randomly selected models from an RTO-TKO ensemble are plotted in gray.

deviation of 0.015. We also specified, in addition to the smoothness constraint, that the model parameters must remain greater than 0.1 ohm-m and less than 100,000 ohm-m—reasonable constraints given a true model representing a sub-lithospheric conductive anomaly.

We inverted this data set using four independent samplers, each drawing around 1,000 samples for a total of 4,277 models. The posterior distribution for μ , root mean square (RMS) misfit, and convergence time and number of iterations for each stochastic optimization step are shown in Fig. A6. The mean of the ensemble, shown in Fig. A4b, captures a smooth version of the conductive anomaly as well as the resistive half-space.

The uncertainty in the model resistivity is captured not in the mean but in the variation across the members of the model ensemble. The model standard deviation at each point in the model is shown in Fig. A7a, where cool colors indicate lower uncertainty and warm colors larger uncertainty. The general trend is an increase in uncertainty with depth, with the smallest uncertainty directly beneath the MT stations (white triangles). The uncertainty is

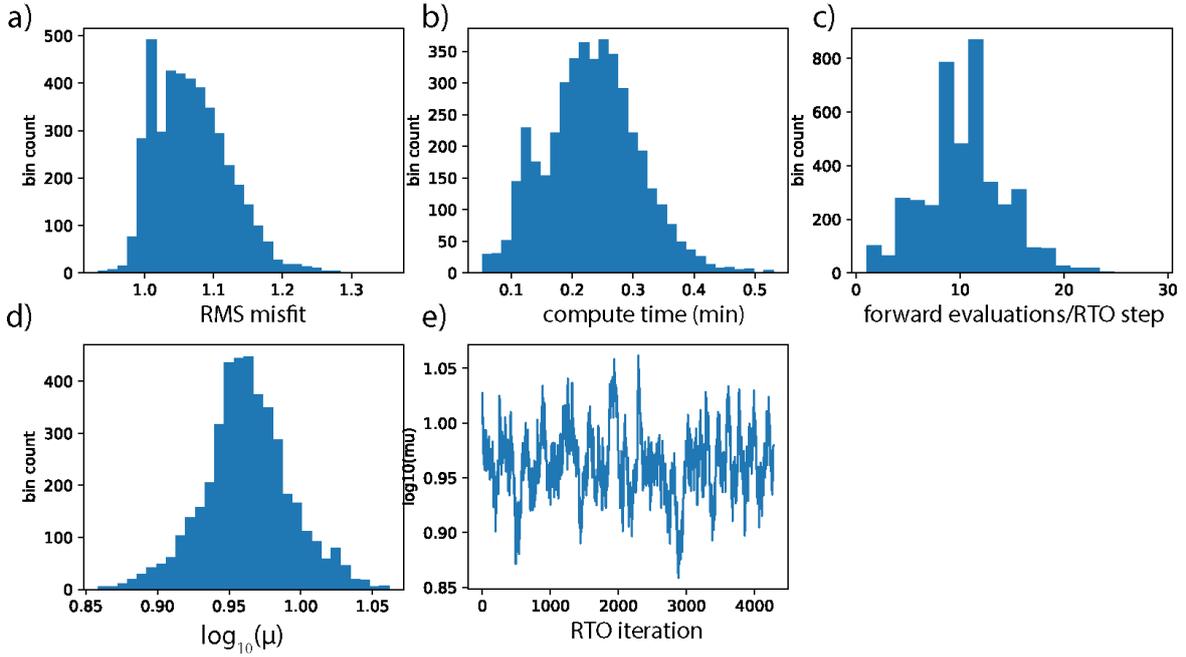


Figure A6. RTO-TKO inversion statistics for the synthetic 2D MT data. (a) Distribution of RMS misfit across 4,277 models in the RTO-TKO ensemble. (b) Total compute time in minutes per RTO-TKO sample. (c) Histogram of number of forward evaluations required per RTO-TKO sample. (d) Distribution of samples for the regularization parameter, μ , as well as (e) the sequence of μ samples.

highest at great depth and at the top of the conductive anomaly where the data sensitivity is high and the drop in resistivity (two orders of magnitude) is large.

Marginal distributions taken from slices through the model are shown in Fig. A7b-c. Here, cooler colors indicate lower probability and warmer colors higher probability. The white squares indicate the true model values while the red lines delineate the 90% credible interval. For the most part, the true model values lie within the 90% credible interval, except where the model jumps abruptly.

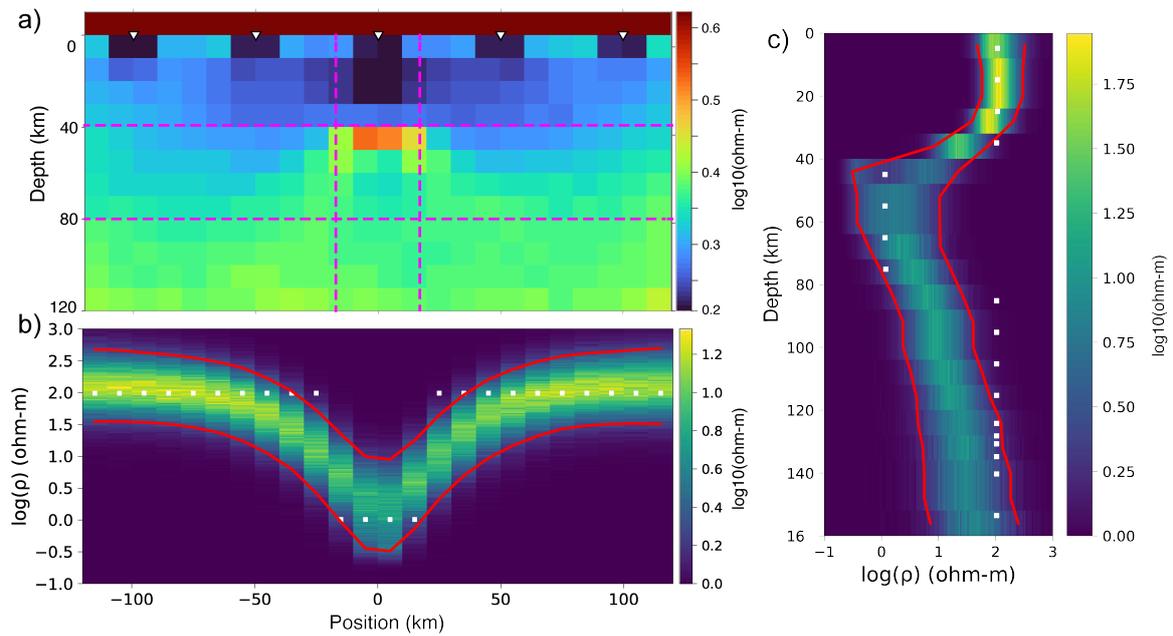


Figure A7. Quantitative uncertainty on subsurface electrical resistivity obtained by inverting the synthetic MT data (Fig. A5) using RTO-TKO. (a) The standard deviation shows increasing uncertainty with depth, as well as uncertainty regarding the depth to the top of the conductor. The lowest uncertainty is obtained beneath the stations. MT stations are indicated by white triangles. (b-c) Marginal distributions of uncertainty for horizontal and vertical slices through the conductor, respectively (dashed lines in (a)). White squares denote the true model values, while the red lines delineate the 90% credible interval.