### Title

RamBO: Randomized blocky Occam, a practical algorithm for generating blocky models and associated uncertainties




### Authors

Huitzil, Eliana Vargas

Morzfeld, Matthias

Constable, Steven



### Publication Date

2025-02-13

### DOI

10.1093/gji/ggaf055

submitted to *Geophys. J. Int.*

# RamBO: Randomized blocky Occam, a practical algorithm for generating blocky models and associated uncertainties.

Eliana Vargas Huitzil, Matthias Morzfeld, Steven Constable

*evargashuitzil@ucsd.edu, mmorzfeld@ucsd.edu, sconstable@ucsd.edu*

*Institute of Geophysics & Planetary Physics, Scripps Institution of Oceanography,*

*University of California, San Diego, La Jolla, CA 92093-0225*

27 February 2025

**SUMMARY**

We present new numerical tools for geophysical inversion and uncertainty quantification (UQ), with an emphasis on blocky (piecewise-constant) layered models that can reproduce sharp contrasts in geophysical or geological properties. The new tools are inspired by an "old" and very successful inversion tool: regularized, nonlinear inversion. We combine Occam's inversion with total variation (TV) regularization and a split Bregman method to obtain an inversion algorithm that we call *blocky Occam*, because it determines the blockiest model that fits the data adequately. To generate a UQ, we use a modified randomize-then-optimize approach (RTO) and call the resulting algorithm *RamBO* (randomized blocky Occam), because it essentially amounts to running blocky Occam in a randomized parallel for-loop. Blocky Occam and RamBO inherit computational advantages and stability from the combination of Occam's inversion, split Bregman and RTO, and, therefore, can be expected to be robustly applicable across geophysics.

**Key words:**

Marine electromagnetics; inverse theory; Bayesian inference

## 1 INTRODUCTION

Some geophysicists are lucky, and maps or images of their data carry meaningful information that is directly interpretable in terms of geological structure. Examples include maps of the gravity or magnetic field and seismic or radar reflection profiles. Those of us who work with electromagnetic (EM) methods are not so lucky, and from the beginning have had to use some sort of inverse method to extract models of electrical resistivity from otherwise obscure data (e.g. Parker (1970); Inman et al. (1973)). Of course, other geophysicists use inverse methods also, particularly those who seek the seismic velocity structure of the mantle, but as Sven Treitel (personal communication) pointed out, the electromagnetic community has made significant contributions to inverse methods because it needs them more than most.

Model space is infinite, even for a one-dimensional resistivity function of depth, yet data are both finite and noisy. This means that the inverse problem is under-determined and ill-posed, and also non-unique; if one solution fits the data then an infinite number will. Early approaches to tackling these problems were to reduce the size of model space by inverting for the resistivities and thicknesses of a small number of layers (Inman et al. 1973) or by solving for averages over some kind of resolving kernel (Parker 1970). These early approaches had three problems: (i) the solutions depend on *a priori* choices; (ii) too many layers lead to instability; and (iii) nonlinear inversions had to be started fairly close to a solution.

The introduction of a smoothing regularization algorithm called Occam's inversion (Constable et al. 1987) solved all of these problems. Occam's inversion collapses the infinite solution space onto a single, useful solution, by searching for the "smoothest model" that fits the data adequately. The result is a remarkably stable algorithm that is largely independent of the number of layers (i.e., using "too many" layers is not a problem) or initial guess (a half-space is sufficient and indeed desirable). Occam's inversion was introduced for one dimensional (1D) problems, but it was readily scaled up to 2D (DeGroot-Hedlin & Constable 1990) and 3D (Siripunvaraporn & Sarakorn 2011) geometries.

Smooth regularized inversion as introduced by the Occam algorithm has become ubiquitous in geophysics, but it has its problems.

(i) If Earth's resistivity structure is not smooth, then Occam's inversion can produce artifacts in the model and a bias in estimated depth of structure. This is not an "academic" problem – sharp resistivity contrasts can occur in the real world, such as edges of sedimentary basins, faults, and many other geological structures. This "Gibbs type" phenomenon (Gibbs 1899) of Occam's inversion has been known since the introduction of the algorithm, but to the best of our knowledge never documented in print (see Section 2.1 for more detail).

(ii) Creating a uniquely smoothest model makes it extremal. The resistivity contrasts are the minimum required to fit the data, not the most likely. A bounded model can be useful in many circumstances, but sometimes the best estimate of the actual rock resistivity is what is wanted, say for a porosity estimate.

(iii) It is difficult to compute and uncertainty quantification (UQ) associated with the inversion. The method currently in vogue, Markov chain Monte Carlo (MCMC), must resort to using sparsely parameterized models in order to force stability and limit computational cost (see, e.g., Malinverno (2002); Blatter et al. (2021) for applications of MCMC in EM geophysics).

We create new computational tools that inherit all benefits of Occam's inversion but that can recover sharp resistivity contrasts, generate a UQ, and give an estimate of the most probable models. We first consider a single inversion (no UQ) and search for a blocky model by swapping the smoothing regularization for a Total Variation (TV) regularization (Rudin et al. 1992). TV regularization had great successes in image deblurring and compressed sensing, and we incorporate it into a nonlinear Occam-style inversion which we call "*blocky Occam*." Blocky Occam follows the tried and true recipe of an Occam's inversion. We linearize around the current model and obtain a *linear* TV-regularized problem. We then adjust the regularization strength to minimize misfit of the *nonlinear* model. These steps are iterated until convergence and once a target RMS is reached, we choose the largest regularization strength that achieves the desired target RMS (searching for the blockiest model that fits the data). Key to success here is our use of the split Bregman method (Goldstein & Osher 2009) to solve the linearized TV-regularized problem at each iteration. Split Bregman is one of the fastest methods to solve linear TV-regularized inverse problems, but it has not been used within iterative, nonlinear inversion.

We equip blocky Occam with a UQ via a modified "*randomize-then-optimize*" (RTO) ap-

proach. RTO generates a UQ by repeatedly solving perturbed inverse problems and RTO has been used for decades under various names in various fields. In short, the RTO implies that we can re-purpose blocky Occam for UQ, by essentially running blocky Occam in a parallel for-loop on perturbed inverse problems. We call the resulting algorithm RamBO (randomized blocky Occam).

Blocky Occam and RamBO are built on the robust framework of Occam's inversion and, for that reason, inherit very desirable numerical characteristics:

(i) the initial guess can be far from the solution and the optimization is stable (all numerical experiments start with a half-space, just as classical Occam's inversion);

(ii) the use of too many layers is inconsequential because the TV regularization suppresses unnecessary features;

(iii) the iteration converges quickly so that computations are manageable (convergence is comparable to classical Occam's inversion);

(iv) the algorithms are largely tuning-free.

Both RamBO and blocky Occam linearize the model and require that the Jacobian of the model is computable, either via adjoints or automatic differentiation. RamBO inherits additional computational efficiency from the RTO approach, so that very few samples (50 or so) are sufficient to obtain a reliable UQ (Blatter et al. 2022a,b), doing away with randomized and expensive searches characteristic of (trans-dimensional) MCMC.

The rest of this paper is organized as follows. Section 2 reviews background materials on Occam's inversion, randomize-then-optimize, the search for blocky models and split Bregman for linear TV regularized inversion. Blocky Occam and RamBO are explained in detail in Sections 3 and 4. The use of blocky Occam and RamBO is illustrated on two EM data sets (Constable et al. 1984; Gustafson et al. 2019) in Section 5, where we also compare the new inversions and UQ to Occam's inversions and UQs obtained via trans-dimensional MCMC (Malinverno 2002; Blatter et al. 2019). We end the paper with a summary of conclusions in Section 6.

## 2 BACKGROUND

Regularized inversion remains the standard method for solving geophysical inverse problems. The basic idea is to define and subsequently optimize a cost function that combines data misfit and model regularization (see, e.g., Parker 1994). To set up the notation, we denote the data by the $n_d$-dimensional vector $d$, the unknown model parameters (e.g. resistivities) of a discretized model are stored in the $n_m$-dimensional vector $m$ and the forward model that predicts the data (usually a sophisticated computer code) is denoted by $\mathcal{F}(m)$. Errors associated with the data are stored in a $n_d \times n_d$ (diagonal) matrix $W$ (reciprocal error weights). A typical cost function can now be written as

$$\mathcal{C}(m) = \|W(\mathcal{F}(m) - d)\|^2 + \mu \|Dm\|^2, \tag{1}$$

where $D$ is a finite differencing matrix and where two vertical bars denote the $\ell_2$-norm of a vector, i.e., $\|x\|_2 = \sqrt{\sum_i x_i^2}$. Throughout, we will refer to the first term of the cost function as the "data-misfit" and the second term as the "regularization." The "strength" of the regularization is controlled by the scalar $\mu > 0$.

### 2.1 Occam's inversion

Occam's inversion (Constable et al. 1987) is an iterative algorithm that has been used for decades for regularized inversion. During the iteration, Occam's inversion adjusts the regularization strength $\mu$ and finds the smoothest model that fits the data – the quadratic regularization term favors smooth models. The iteration of Occam's inversion is as follows. At step $k$, the model is $m_k$ and we approximate the forward model via Taylor expansion:

$$\mathcal{F}(m_{k+1}) \approx \mathcal{F}(m_k) + J_k(m_{k+1} - m_k), \tag{2}$$

where $J_k = \partial\mathcal{F}/\partial m$ is the Jacobian matrix, evaluated at $m_k$. Using the linearization in (1), yields a quadratic cost function for $m_{k+1}$

$$\mathcal{C}(m_{k+1}) = \left\|W(J_k m_{k+1} - \hat{d})\right\|^2 + \mu \|Dm_{k+1}\|^2, \tag{3}$$

where

$$\hat{d} = d - \mathcal{F}(m_k) + J_k m_k, \tag{4}$$

is "a kind of data vector" that accounts for errors due to linearization. We can easily optimize the quadratic function (least squares) to find $m_{k+1}$ and we do so for various regularization strengths $\mu$. Once a regularization $\mu$ is selected, the process repeats until the iteration converged. During the iterations, we choose $\mu$ to minimize root mean squared error (RMS)

$$\text{RMS} = \frac{1}{\sqrt{n_d}} \left\| W(\mathcal{F}(m) - d) \right\|, \tag{5}$$

associated with the *nonlinear* model $\mathcal{F}(\cdot)$. Once the iteration reached a target RMS, we chose the largest $\mu$ that achieves the target RMS. A good choice for a target RMS is one or slightly larger. Some implementations of Occam's inversion, e.g., MARE2DEM (Key 2016), include a "fast Occam" option which dispenses with the line search minimization and accepts any $\mu$ that decreases misfit at a given iteration.

We monitor convergence of Occam's inversion via the "model roughness," i.e., we stop the iteration (and declare convergence), when

$$\Delta R = \frac{\|Dm_{k+1}\| - \|Dm_k\|}{\|Dm_k\|} \leq \text{tol.}, \tag{6}$$

where tol is a small number (usually $10^{-2}$ or so).

### *2.1.1 Gibb's phenomenon in Occam's inversions*

If smooth inversions are carried out for models that have sharp changes in resistivity one observes a Gibbs type phenomenon (Gibbs 1899), in which the regularized inversion overshoots the resistivity jump. We illustrate this Gibbs phenomenon with a simple synthetic model study in which MT data with various error levels are inverted for a jump in resistivity (see Appendix A for details). The resulting models are displayed in Figure 1(a), showing that once the error is below 10% an overshoot develops on both sides of the resistivity jump, but more so on the resistive side (something that persists if the layers are swapped to make the top layer resistive). There is the danger that for more complicated models the spurious peaks in resistivity could be interpreted as real structure. Taking the midpoint of the resistivity change in the regularized models over-estimates the depth of the resistivity jump by about a factor of 2. We can verify that smooth inversions recover smooth models without such artifacts. In Figure 1(b) the step func-

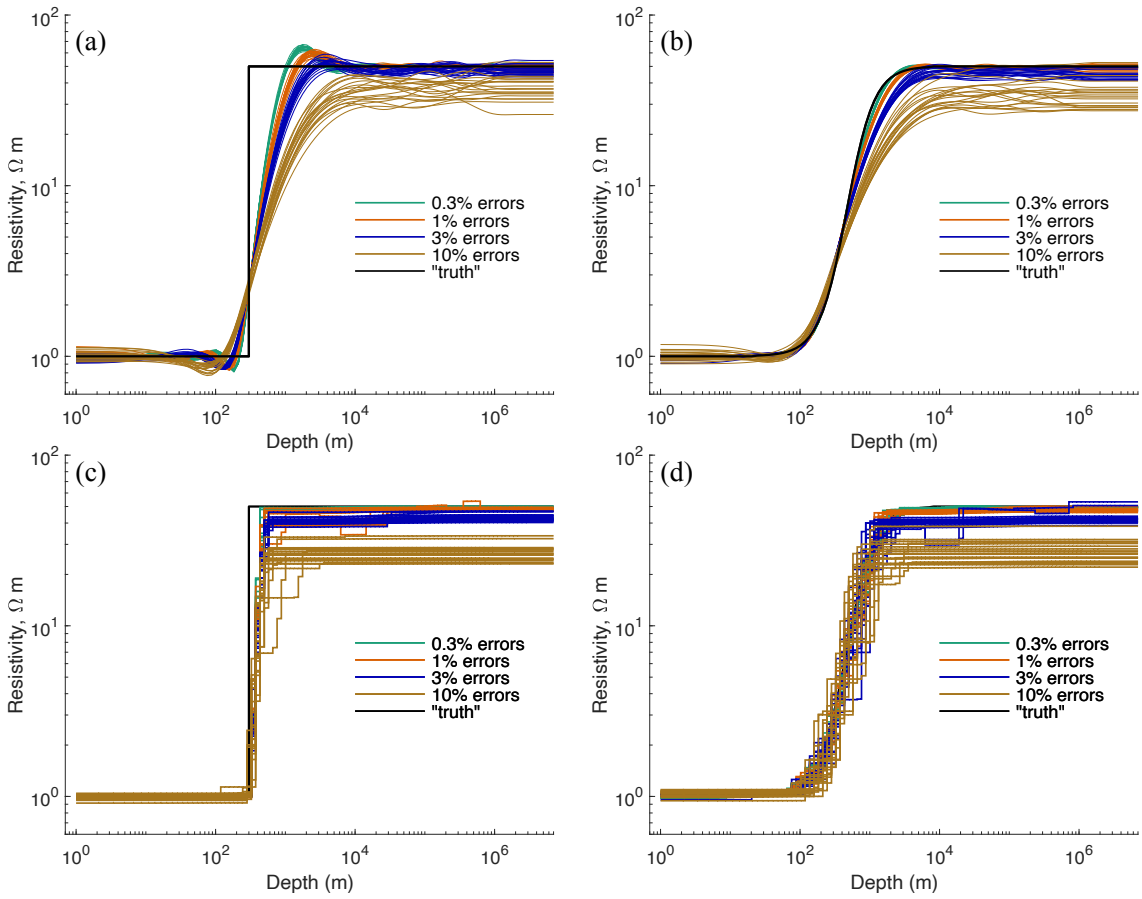



**Figure 1.** Inversion of synthetically generated MT data with various levels of noise added. (a): Occam's inversions of a step increase in resistivity ("truth", black). (b): Occam's inversions of a smooth (sigmoid) increase in resistivity. (c): Blocky Occam's inversions of a step increase in resistivity. (d): Blocky Occam's inversions of a smooth (sigmoid) increase in resistivity.

tion is replaced with a sigmoid function. No overshoot is observed as the error level is reduced, and all except inversions of the most noisy data recover the model faithfully.

We ran the same inversions using "blocky Occam," as described below, with results shown in Figure 1(c)-(d). Blocky Occam improves depth estimation without producing unreliable results for the smooth model and practically eliminates the Gibbs phenomenon. The synthetic tests further illustrate that blocky Occam can capture smooth transitions when needed (see Figure 1(d)).

## 2.2 Uncertainty quantification via randomize-then-optimize

The popular approach to uncertainty quantification (UQ) is via Bayes' theorem, which states that

$$p(m|d) \propto p(d|m)p(m), \tag{7}$$

where $p(m|d)$ is the probability of the model given the data (the posterior probability), $p(m)$ is a prior probability of the model (often taken to be Gaussian), and where $p(d|m)$ is the likelihood, connecting the model $m$ to the data $d$ via the forward model $\mathcal{F}$. The symbol $\propto$ denotes proportionality, i.e., the quantity to the left differs from the quantity to the right by a multiplicative constant. In Bayes' theorem, the missing constant is the probability of the data, $p(d)$, which is called the "evidence." The evidence is not so relevant for UQ, but it can be useful for model selection (Sambridge et al. 2006).

There are many connections between regularized inversion and Bayesian UQ (see, e.g., Blatter et al. 2022a). For example, we can interpret an Occam-style optimization (with a cost function as in equation (1)) as the search for the model that maximizes the posterior probability

$$p(m|y) \propto \exp\left(-\frac{1}{2}\left(\|W(\mathcal{F}(m)-d)\|^2 + \mu\,\|Dm\|^2\right)\right). \tag{8}$$

Connections between a Bayesian posterior distribution and optimization can be exploited to yield efficient and scalable, but approximate sampling methods for UQ. Specifically, one can sample the posterior distribution by solving perturbed optimization problems

$$\arg\min_m \left(\|W\left(\mathcal{F}(m)-(d+\eta)\right)\|^2 + \mu\,\|Dm+\xi\|^2\right), \tag{9}$$

where $\eta$ and $\xi$ are Gaussian random variables that represent perturbations to the data ($\eta$) and to the regularization ($\xi$). 8More specifically, the data perturbations $\eta$ are mean zero Gaussians and their covariance is matrix is $(W^TW)^{-1}$, which is representative of the assumed errors in the data. The perturbations $\xi$ are mean zero Gaussian with covariance matrix $(1/\mu)I$, where $I$ is the $n_m \times n_m$ identity matrix. Both perturbations (data and regularization) are needed or else variances may be underestimated (see Blatter et al. 2022a).

The above optimization-based sampling process has been invented and re-invented in many fields. It is called RTO (randomize-then-optimize, Bardsley et al. (2014); Blatter et al. (2022a))

in the mathematical community, "ensemble of data assimilation" in numerical weather prediction (Isaksen et al. 2010), it goes by the name of "randomized maximum likelihood" in the oil and gas industry (Oliver et al. 1996; Chen & Oliver 2012), and is referred to as "parametric bootstrapping sampling" in hydrology (Kitanidis 1995; Lee & Kitanidis 2013; Gunning et al. 2010). The process is thus well-understood and known to scale to large models and large data sets. RTO is exact only if the forward model is linear, but it has proven to be very useful for solving nonlinear problems in a large number of very different applications (see Blatter et al. (2022a) for more details).

### 2.3 Blocky models

The philosophy behind Occam's inversion is to construct models devoid of features not required by the data, achieved by finding the smoothest model (in some sense). However, many, perhaps even most, geological features of interest are associated with rapid, not smooth, changes in physical properties. Examples include the interface between sedimentary and igneous or volcanic rocks, groundwater tables, edges of magmatic reservoirs, fault structures, and many others. Occam models are useful in such circumstances because the interpreter understands that sharp boundaries will be smoothed by the inversion algorithm, but the actual boundary in question is not localized in space, and the physical property contrast (e.g. electrical resistivity) is smaller than it is in the true Earth (see Section 2.1.1 for a simple illustration).

One way forward is to move from quadratic (Tikhonov) regularization to $\ell_1$-norm regularization, which produces "blocky" (piecewise constant) models. Indeed, smooth and blocky inversions have competed with each other for decades (see, e.g., Portniaguine & Zhdanov 1999; Farquharson & Oldenburg 1998), and variations of the idea have been pondered over for many years, (see, e.g., Guitton & Symes 2003; Theune et al. 2010; Lee & Kitanidis 2013; Sun & Li 2014; Wang et al. 2017; Fournier & Oldenburg 2019; Tang et al. 2021; Wei & Sun 2021). But the methods have never really found their way to mainstream applications. We suspect that the reasons include that some methods are computationally expensive, while others are awkwardly described or unnecessarily complicated. Moreover, some methods do not address the required search over the "nuisance" parameter $\mu$ and a UQ has rarely (if ever) been attempted.

We address these issues and port $\ell_1$ regularization ideas to the well-known, robust and efficient framework of Occam's inversion. We then further equip our inversions with an efficient UQ, implemented via a modified RTO approach.

## 2.4 Split Bregman

Before describing our nonlinear inversion algorithms, we take a short detour and discuss the solution of *linear* inverse problems with total variation (TV) regularization via split Bregman (Goldstein & Osher 2009). Specifically, we wish to minimize

$$\mathcal{C}(x) = \|Jm - d\|^2 + \mu\,|Dm|\,, \tag{10}$$

where $m$ and $d$ are vectors of size $m_n$ and $m_d$, $J$ is a $n_d \times n_m$ matrix, $D$ is a finite difference matrix and $\mu > 0$ is a (given) scalar; here $|\cdot|$ denotes the $\ell_1$-norm, i.e., for a $n_x$-dimensional vector

$$|x| = \sum_{i=1}^{n_x} |x_i|. \tag{11}$$

The regularization $|Dm|$, i.e., the $\ell_1$ norm applied to the derivative of the unknown $m$, is often called total variation (TV) regularization (Rudin et al. 1992).

The split Bregman method, applied to this problem, introduces the auxiliary variable $u = Dm$ and the Bregman variable $b$ to reformulate the cost function as

$$\mathcal{C}_{\text{Breg}}(m, u) = \|Jm - d\|^2 + \mu|u| + \gamma\|u - Dm - b\|^2 \tag{12}$$

where $\gamma$ is a second Lagrange multiplier (but $\gamma = 2\mu$ is a robust choice). The above cost function is optimized by iterating the following three steps, which updates $m$, $u$ as an approximation of $Dm$, and the Bregman variable $b$ sequentially:

(i) For a given $u_k$ and $b_k$, minimize $\mathcal{C}_{\text{Breg}}$ over $m$ by solving the least squares problem

$$m_{k+1} = \arg\min_m \left(\|Jm - d\|^2 + \gamma\|u_k - Dm - b_k\|^2\right) \tag{13}$$

(ii) Given $b_k$ and $m_{k+1}$, minimize $\mathcal{C}_{\text{Breg}}$ over $u$ by solving the optimization problem

$$u_{k+1} = \arg\min_u \left(\mu|u| + \gamma\|u - Dm_{k+1} - b_k\|^2\right). \tag{14}$$

The solution is a soft-thresholding so that

$$u_{k+1} = \text{ST}(Dm_{k+1} + b_k; 2\mu/\gamma), \tag{15}$$

where

$$\text{ST}(x; \alpha) = \text{sign}(x) \max\left(|x| - \alpha, 0\right) \tag{16}$$

is the soft-thresholding function (applied element-wise to the vector in (15)).

(iii) The third step updates the Bregman variable

$$b_{k+1} = b_k + (Dm_{k+1} - u_{k+1}). \tag{17}$$

The above three steps are iterated until we reach convergence. Note that all three steps are easy to implement and scalable: step (i) is a least squares solve; step (ii) is a simple soft-thresholding; and step (iii) is a simple updating (vector addition and matrix-vector multiplication). Indeed, split Bregman is arguably the fastest and most robust method (Goldstein & Osher 2009) for minimizing the TV regularized cost function (10) and, has been very successfully applied to various large scale linear inverse problems.

In the numerical illustrations in Section 5 we set the additional Lagrange multiplier $\gamma = 2\mu$ (as recommended) and use a simple convergence criteria to stop the iteration if

$$\frac{\|m_{k+1} - m_k\|}{\|m_k\|} \leq \text{tol}_{\text{SB}}, \tag{18}$$

with tolerance $\text{tol}_{\text{SB}} = 10^{-4}$, or if a maximum number of iterations ($k_{\max} = 300$) is reached. Split Bregman is summarized as an algorithm in Appendix B.

## 3 BLOCKY OCCAM

We now describe "a kind of" Occam's inversion which we call *blocky Occam*. Blocky Occam discovers the blockiest model that fits the data with the fewest changes in resistivity. To find blocky models, we swap the quadratic regularization in (1) with a total-variation (TV) regularization (Rudin et al. 1992)

$$\mathcal{C}(m) = \|W(\mathcal{F}(m) - d)\|^2 + \mu \left|Dm\right|, \tag{19}$$

where $|\cdot|$ denotes the $\ell_1$-norm. The TV regularization ($\mu\,|Dm|$) enforces sparsity of the *derivative* of the model, by applying the sparsity-promoting $\ell_1$-norm to it. For these reasons, TV regularization promotes piece-wise constant, blocky models as desired.

We mimic Occam's inversion and set up an iteration. Linearizing (see equation (2)) around the current iterate $m_k$ gives

$$\mathcal{C}(m_{k+1}) = \left\| W(J_k m_{k+1} - \hat{d}) \right\|^2 + \mu\,|Dm_{k+1}|\,, \tag{20}$$

where, as in Occam's inversion, $J_k$ is the Jacobian of the forward model and $\hat{d} = d - \mathcal{F}(m_k) + J_k m_k$ (compare the above equation with (3)). In Occam's inversion, one obtains a least squares problem after linearization (which is easy to solve). Linearization in blocky Occam leads to a linear TV-regularized inverse problem. This problem can be solved efficiently with split Bregman (see Section 2.4) for a range of regularization parameters $\mu$. Once we chose a $\mu$, we can proceed with the iteration. During the iterations, we either chose $\mu$ to minimize RMS (of the nonlinear model) or, if RMS is below the target RMS, we use the *largest* $\mu$ that results in the target RMS (generating the blockiest model that fits the data adequately). One may consider adapting ideas of fast Occam (Key 2016) to the TV-regularized problem. Convergence of blocky Occam is assessed via the model roughness as in Occam's inversion (see Section 2.1). We summarize blocky Occam in Algorithm 1.

Blocky Occam inherits the robustness and numerical efficiency from Occam's inversion:

(i) The regularization strength is adjusted automatically during the iteration, which enhances robustness of the iteration and almost always results in quick convergence (rarely divergence). The only tunable parameter in blocky Occam is the desired target RMS and the initial model, which is usually a half space (constant resistivity).

(ii) Just as in Occam's inversion, one does not need to worry about the layer thickness or, more generally, the grid of the forward model. The TV regularization enforces blocky models with few resistivity changes independently of the underlying grid ("too many" layers are not a concern, blocky Occam will find the simplest, blockiest model that fits the data).

(iii) One can create a blocky Occam code with only minor modifications to an existing Occam code. The only difference is that we swap the least squares solves after linearization with a

**Algorithm 1** Blocky Occam

**while** $k \leq k_{\text{max}}$ **do**

Compute the Jacobian $J_k$ and the modified data vector $\hat{d} = d - \mathcal{F}(m_k) + J_k m_k$

**for** $\mu \in [\mu_{\min}, \mu_{\max}]$ **do**

Apply split Bregman to solve the optimization problem

$$\underset{m_{k+1}}{\arg\min} \left\| W(J_k m_{k+1} - \hat{d}) \right\|^2 + \mu \left| D m_{k+1} \right|,$$

Compute RMS of the optimizer using the nonlinear model $\mathcal{F}(\cdot)$

**end for**

**if** RMS $\leq$ RMS$_{\text{target}}$ **then**

Pick largest $\mu$ that leads to an RMS below the target

**else**

Pick $\mu$ to minimize RMS

**end if**

Compute the relative change in roughness:

$$\Delta R = \frac{\|D m_{k+1}\| - \|D m_k\|}{\|D m_k\|},$$

**if** $k > k_{max}$ **or** $\Delta R < 10^{-2}$ **then**

**break**

**end if**

$m_k \leftarrow m_{k+1}$

**end while**

split Bregman method, which is also easy to implement and scalable (almost like least squares). The additional Lagrange multiplier that occurs during split Bregman is adjusted automatically and in accordance with the regularization strength $\mu$.

## 4 RANDOMIZED BLOCKY OCCAM

It is desirable and increasingly important to not only invert for one model, but to equip the inversion with an estimate of associated uncertainties in the model. We use a randomize-then-optimize (RTO) approach (Bardsley et al. 2014), originally proposed by Kitanidis (1995); Oliver

**Algorithm 2** Randomized blocky Occam (RamBO)

**for** $k \leq k_{\max}$ **do**

Draw a sample $\eta_k$ from $\eta \sim \mathcal{N}(0, (W^T W)^{-1})$ and a sample $\nu_k$ from $\nu \sim \mathcal{L}(0, 1/\mu)$.

Use blocky Occam with fixed $\mu$ to solve the perturbed optimization problem

$$\arg\min_m \|W(\mathcal{F}(m) - (d + \eta_k))\|^2 + \mu \, |Dm + \nu_k| \, ,$$

**end for**

et al. (1996) and extended to TV regularized problems by Lee & Kitanidis (2013). The RTO approach entails solving perturbed optimization problems with perturbed cost functions

$$\mathcal{C}(m) = \|W(\mathcal{F}(m) - (d + \eta))\|^2 + \mu \, |Dm + \nu| \, , \tag{21}$$

where, as before, $\eta$ is Gaussian with mean zero and covariance matrix $(W^T W)^{-1}$ and where $\nu \sim \mathcal{L}(0, 1/\mu)$ has a Laplace distribution with scale parameter $1/\mu$ (Lee & Kitanidis 2013). We can optimize the perturbed cost functions using blocky Occam, but with *fixed* regularization strength $\mu$. The implementation is easy and only requires that we replace the data $d$ in the cost function (19) by the perturbed data $(d + \eta)$ and that we account for the perturbation $\nu$ in split Bregman (which we describe in the Appendix B). The resulting procedure, which we call "randomized blocky Occam" (RamBO), is summarized in Algorithm 2 and essentially amounts to running blocky Occam within a (parallel) for-loop. For numerical efficiency, we initialize all optimizations during RamBO with the result of a blocky Occam (but initializing with a half-space gives comparable results at a larger computational cost).

Note that the blocky Occams within RamBO do *not* automatically adjust the regularization strength $\mu$. For that reason, the optimization can be less stable and we introduce a stepsize $\alpha \in (0, 1]$ so that the model in the next iteration is a linear combination of the model we found via split Bregman and the current model, i.e., the "replace" step in Algorithm 1 becomes

$$m_k \leftarrow \alpha m_{k+1} + (1 - \alpha) m_k, \tag{22}$$

where $m_k$ is chosen along with a regularization strength $\mu$ to either minimize RMS, or, if the target RMS is reached, along with the largest $\mu$ that yields the target RMS (blockiest model).

The remaining question is: *If RamBO does not automatically adjust the regularization*

*strength $\mu$, how should $\mu$ be determined*? One way forward is to adopt a hierarchical approach and sample models $m$ and regularization strengths $\mu$ jointly from the posterior distribution $p(m, \mu|d)$. This strategy is used in the RTO-TKO (Blatter et al. 2022a,b), and this technology could be adapted to TV regularized problems. An easier and more efficient way forward is to pick a relatively small value for $\mu$, e.g., we pick $\mu = 0.1$ in the numerical illustrations in Section 5. The reason is that by choosing a small $\mu$, we compute the most uncertain blocky models (another use of Occam's principle). The value $\mu = 0.1$ may not be universal and we recommend to first run a blocky Occam (which one may be tempted to do anyways) and monitor the range of regularization strength encountered during blocky Occam.

Finally, we note that Wang et al. (2017) explored $\ell_1$ regularization in the context of RTO via a clever *invertible* change of variables. The TV regularization we need here for blocky models, however, makes the change of variables not invertible and, hence, not applicable (see also (Lee 2021)).

### 4.1 RamBO and trans-dimensional MCMC

A common approach to UQ in geosciences is trans-dimensional Markov chain Monte Carlo (trans-D MCMC) (see, e.g., Sambridge et al. 2013, 2006; Malinverno 2002). Layered models, as discussed here, are parameterized by layer thicknesses and their resistivity. In trans-D MCMC, the number of layers is an *unknown*, and the trans-dimensional formulation induces a natural parsimony to favor models with a small number of layers over models with a large number of layers (Sambridge et al. 2006). Numerical solution via trans-D MCMC entails a randomized search over the number of layers, their thicknesses, and their resistivities.

RamBO achieves parsimony via TV regularization, which enforces that the number of "blocks" is as small as possible (and if smooth parts of the model are required by the data the inversion will allow that also, see Figure 1(d)). The models used in RamBO are parameterized by a *known* (large) number of layers of fixed thickness, and optimization is used to generate samples of layer resistivity.

We expect that RamBO and trans-D MCMC give somewhat similar results when applied to invert the same data because both algorithms are designed with the same goal in mind – finding

blocky models that fit the data. RamBO and trans-D MCMC only differ in *how* these models are obtained numerically and RamBO has some computational advantages.

(i) Trans-D MCMC is slow to converge and, therefore, requires a large number of forward model evaluations. RamBO can generate a reliable UQ from a few samples (50 or so, see numerical examples below).

(ii) The implementation of trans-D MCMC is cumbersome and problem dependent – there is no "general purpose" trans-D MCMC sampler available that can be straightforwardly applied. RamBO is easy to apply, especially if an Occam-style code is already available.

On the flipside, RamBO relies on the Jacobian for fast convergence since one sample requires the solution of an optimization problem, but derivatives are not needed for trans-D MCMC.

## 5 NUMERICAL ILLUSTRATIONS

We illustrate the use of blocky Occam and RamBO on two data sets. The Schlumberger data set collected at Wauchope station in central Australia (Constable et al. 1984) to study crustal resistivity, and a marine magnetotelluric (MT) data set that was more recently collected offshore New Jersey (Gustafson et al. 2019) to understand low salinities observed in wells in nearby areas. For the marine MT data set, we note that the relatively shallow waters in the region (20–100 m) were insufficient to attenuate high frequencies (1–100 Hz), and this allowed resolving upper subsurface structures. Following Blatter et al. (2021), we consider station N05 and invert for 1D resistivity models.

We first perform blocky Occam inversions in Section 5.1 and compare the blocky models to smooth models obtained via Occam's inversion. In Section 5.2, we compute uncertainties using RamBO, and we compare our results to the trans-D MCMC inversions of Malinverno (2002) (for Schlumberger) and Blatter et al. (2019) (for marine MT). In our inversions and UQ, we use the standard deviations reported as part of the Schlumberger and marine MT data sets to construct the weighting matrix $W$ in the cost functions (1) and (19). The model Jacobians are computed via finite-differences, but a more careful implementation should use adjoints or auto-

matic differentiation to reduce the number of required forward model evaluations – we use finite differences here to keep the code clean and because the 1D forward models are computationally inexpensive.

### 5.1 Blocky Occam inversions

We apply blocky Occam to invert the Schlumberger and marine MT data sets and compare the results to Occam's inversions that generate smooth models. All inversions start with a half-space model and both inversion algorithms are given a range of regularization strengths and a target RMS (which we set to one).

For the Schlumberger data set, Occam's inversion converges in 5 iterations, while blocky Occam requires 10 iterations, leading to RMS values of 0.97 (Occam's inversion) and 0.90 (blocky Occam). The resistivity models obtained by blocky Occam and Occam's inversion are shown in Figure 2(a) and the fits to the data are shown in the supplementary Figure A1(a) in Appendix C. As expected and as desired, the blocky Occam models looks like blocky versions of the smooth models obtained via Occam's inversion. More specifically, we find that Occam's inversion reveals two main features: a conductive zone beneath a 2 m dry surface layer and a deeper resistive zone. Because Occam's inversion generates the smoothest model that fits the data, the transition between the conductive and resistive zones is blurry and not well-defined. In comparison, blocky Occam provides a more distinct separation between the conductive and resistive layers, particularly the base of the conductive layer at approximately 200 m depth. Some model smoothness is still required by the data, however.

For the marine MT data set, Occam's inversion converges in 22 iterations, while blocky Occam requieres 9 iterations. The inversions lead to RMS values of 0.98 for Occam's inversion and 1.01 for blocky Occam. The resistivity models are illustrated in Figure 2(b) and the data fits can be found in supplementary Figures A1(b)-(c). The smooth model obtained via Occam's inversion shows two distinct peaks that correspond to resistive and conductive features. The resistive zone between 40 m to 160 m is associated with sediments hosting low salinity water. The conductive feature at about 400 m suggests sediments hosting seawater. The smooth model shows oscillations around 300 m, where the transition between low and high resistivity zones

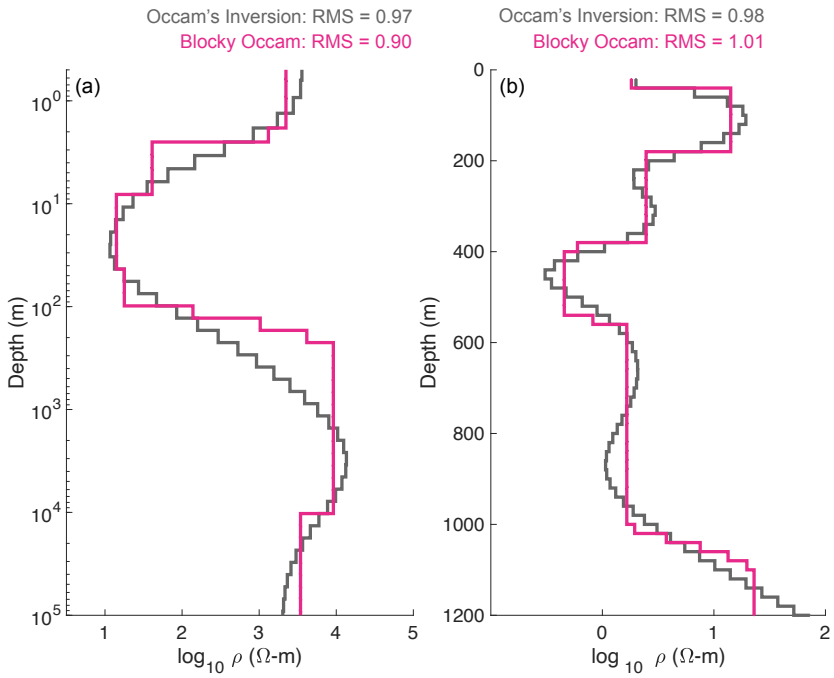



**Figure 2.** Blocky Occam compared to Occam's inversion. Shown are the resistivities as a function of depth for (a) the Schlumberger data set and (b) the marine MT data set. Blocky Occam (pink) and Occam's inversion (gray) lead to nearly identical RMS and the blocky Occam solution looks like a blocky version of the Occam's inversion as desired and as expected.

occurs. These oscillations appear because the smooth inversion should really be blocky or in other words, we have sharp changes in resistivity, and for that reason, we observe a Gibbs-type phenomenon in the transitions in smooth models (see Section 2.1.1). In contrast, the blocky Occam model defines a simpler boundary between the high and low resistivity zones. A basement layer at a depth of about 1,100 m is more clearly defined by blocky Occam than classical Occam.

Finally, we note that blocky Occam applies the split Bregman iteration within the linearizing "outer loop" of an Occam's inversion. The overall computational cost of blocky Occam thus depends on how fast split Bregman converges. Here, convergence of split Bregman is assessed by equation (18) and we chose a small tolerance to obtain very blocky models. The split Bregman iteration converges faster if we use a larger tolerance, but then the resulting models are not really blocky. With our choices, split Bregman converges on average within 181 iterations for the Schlumberger data set and within 268 iterations for the marine MT data set. We acknowledge

that the number of iterations is quite large, which may result in high computational costs in 2D or 3D problems for which the linear algebra of solving least squares problems is more involved than in our 1D test cases (step (i) of split Bregman, see Section 2.4). Our experiments with 1D electromagnetic data thus suggest that split Bregman generates a computational overhead compared to Occam's inversion, but this overhead is needed to obtain truly blocky models. We are unaware of numerical techniques that are more efficient than split Bregman. All other ideas we tried, including approximating $\ell_1$ norms via Eckblom norms or Huber losses, interior point methods for $\ell_1$ convex optimization (see, e.g., Nocedal & Wright 2006), or trans-dimensional MCMC, were computationally more expensive, led to smoother models, or both. The search for blocky models may always be computationally more expensive than searching for smooth models: the TV-regularized inverse problem (19) is inherently more difficult to solve than the nonlinear least squares problem in (1).

### 5.2 UQ with RamBO

We now apply RamBO to the Schlumberger and marine MT data sets to compute an uncertainty quantification. RamBO amounts to running blocky Occam, with a fixed regularization strength $\mu = 0.1$ in a parallel for-loop. We obtain similar results with similar $\mu$, but if we choose $\mu$ to large (e.g., $\mu = 2$), then the uncertainty bounds are very narrow due to the large influence of the Laplacian prior. If $\mu$ is too small (e.g, $\mu = 0.01$), the optimization is unstable. In general, one should adjust $\mu$ for blocky Occam to be as *small* as possible to compute the largest possible uncertainty. A range of possible regularization strength values is often apparent after inspecting the results of blocky Occam or Occam's inversion.

Since the 1D inversions are inexpensive, and since competing trans-D MCMC codes usually require a very large number of forward model evaluations, we draw a large number of samples $(10^4)$ for comparison with trans-D MCMC. For both data-sets, the optimization of RamBO occasionally leads to a large RMS $> 2$ or fails. We filter out these failed attempts and are then left with $8574$ samples for the Schlumberger data set and $9774$ samples for the marine MT data set. We use these samples in Figure 3 to create histograms of resistivity (log-scale) as a function of depth, similar to Figure 12 in Malinverno (2002) and Figure 10(b) in Blatter et al. (2019).

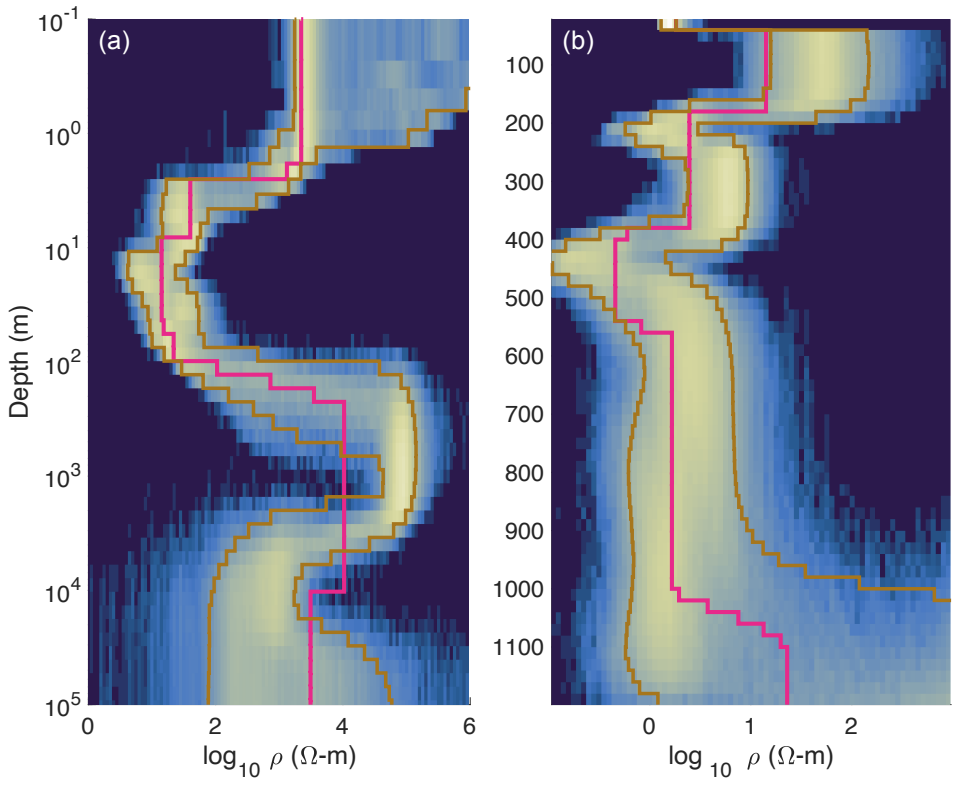



**Figure 3.** Uncertainty quantification for the Schlumberger data set (a) and marine MT data set (b). Shown are histograms of resistivity (log-scale) as a function of depth. Warmer colors (green and yellow) indicate higher probability and cool colors (blues) indicate low or no probability (dark blue). The brown lines indicate 5% and 95% quartiles and the pink lines correspond to the blocky Occam results described above.

For the Schlumberger data set (Figure 3(a)), we find an uncertain but resistive surface layer to a depth of 2 m, followed by three similarly conductive layers (3.5–10 m, 10–30 m and 30–100 m). Between 170 m and 4500 m we detect a resistive layer where the most probable models are nearly an order of magnitude larger than either the smooth or blocky Occam models. These results are in good agreement with the trans-D MCMC results reported by Malinverno (2002) and, to a lesser extent, also with the results of Blatter et al. (2022b), which uses a quadratic regularization (compare Figure 3 with Figure 1 in Blatter et al. (2022b) and Figure 12 in Malinverno (2002)). Both studies result in large posterior uncertainties in the resistivities and interface depths of the deeper layers, and the posterior density estimates obtained via RamBO are consistent with the established results. Moreover, resistivity changes identified by the trans-D MCMC are comparable to those of RamBO (Compare Figure 3a to Figure 12 of Malinverno (2002)), but RamBO obtains these results at a much reduced computational cost.

In addition, we note that uncertainty is *not* symmetric about the blocky Occam model (pink line in Figure 3(a)). This is to be expected because the blocky Occam model is an extreme model – the steps between blocky regions are still the smallest that fit the data.

For the marine MT data set, RamBO defines a resistive layer (40–200 m) and a conductive layer (400–500 m). Below 500 m, the uncertainty is rather large, which is in good agreement with the trans-D results reported by Blatter et al. (2019). These results highlight three major changes in resistivity (0-100 m, 200 – 400 m and 400-600 m) , consistent with results obtained via RamBO (see Figure 3b). Again, the uncertainty is not symmetric around the blocky Occam solution (as expected). The blocky Occam solution rather picks out the least resistive model that is rendered likely by RamBO when the data are informative (above 400 m), which makes sense since MT is more sensitive to thin conductors than thin resistors (e.g. Key et al. (2006)).

The data fits of models generated by RamBO for the Schlumberger and marine MT data sets are shown in Figures A2(a,c,d) in Appendix C. Histograms of RMS of models generated by RamBO are shown in Figures A2(b,e). RamBO explores many models that fit the data well and the distribution of RMS is near one for both data sets.

In summary, RamBO generates a UQ that is comparable to what other methods have produced. Compared to trans-D MCMC, however, RamBO has two advantages:

(i) The UQ can be computed at a reduced computational cost.

(ii) RamBO relies on optimization and can be implemented with only minor modifications of an existing Occam's inversion code. Trans-D MCMC, on the other hand, is usually tailor-made for each problem and trans-D MCMC codes are not easily portable from one inversion to another.

The computational advantage of RamBO compared to trans-D MCMC is more apparent if we constrain the number of samples. With RamBO, about 50 models may be sufficient to get an idea of the uncertainty of the inversion, provided the posterior distribution is unimodal (Blatter et al. 2022b). We illustrate this idea in Figure 4, where we show a "spaghetti plot" of 50 samples of RamBO. The 50 samples are sufficient to eyeball regions of large or small uncertainty and the 5% and 95% quartiles are already comparable to those obtained from $O(10^4)$

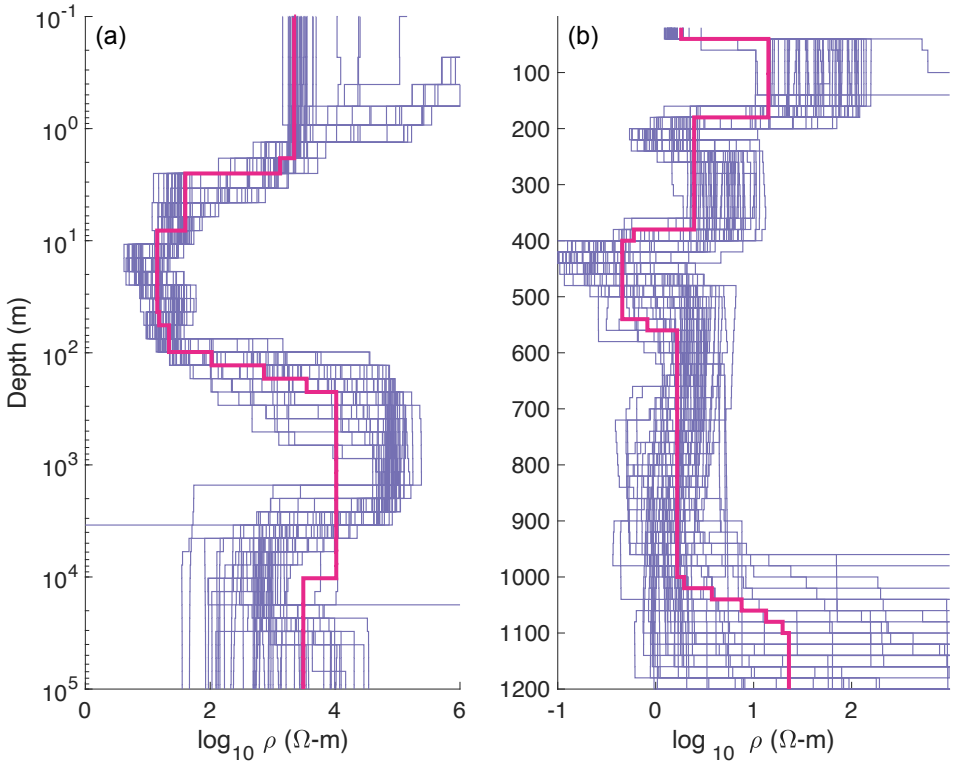



**Figure 4.** Spaghetti plots of 50 samples obtained by RamBO (purple) for the Schlumberger data set (a) and the marine MT data set (b). Shown in pink is the blocky Occam model

samples. RamBO inherits the computational efficiency for UQ from RTO, which was already reported and discussed at length in the context of inverting EM data by Blatter et al. (2022a,b). MCMC in general, and trans-D MCMC in particular, routinely require thousands or millions of forward solves due to slow convergence (and the convergence becomes slower with dimension/the number of layers). As an illustration, trans-D MCMC routinely requires $10^6$ (or more) samples, each one requiring one forward solve. One sample of RamBO requires an optimization, which requires about 10 iterations (conservatively speaking), each requiring one forward and one adjoint solve. Assuming forward and adjoint solves are comparable, we can estimate the cost of one RamBO sample to about 20 forward solves, so that 50 RamBO samples require $10^3$ forward solves, which is reduces the cost by three orders of magnitude compared to trans-D MCMC. Thus, RamBO may be a computationally efficient and more robustly applicable alternative to trans-D MCMC.

## 6 SUMMARY AND CONCLUSIONS

We present new computational tools for geophysical inversion that can recover sharp (resistivity) contrasts and generate an uncertainty quantification (UQ). Specifically, we incorporate total variation (TV) regularization within an Occam's inversion and we call the resulting inversion algorithm *blocky Occam*. Blocky Occam determines the blockiest model that fits the data adequately. A modified randomize-then-optimize (RTO) approach allows us obtain a UQ by essentially running blocky Occam in a parallel for-loop on perturbed optimization problems. The resulting UQ algorithm is called *RamBO* (randomized blocky Occam).

Blocky Occam and RamBO are built upon the robust foundation of an Occam's inversion, and, for that reason, they are:

(i) largely tuning-free;

(ii) insensitive to the number of layers or starting guess;

(iii) fast to converge.

Moreover, blocky Occam can be obtained from an existing Occam code with only minor modifications and RamBO is easy to implement once a blocky Occam code is available. Key to the efficiency of blocky Occam is that we use split Bregman, one of the fastest methods for solving *linear* TV regularized problems, within an Occam-style *nonlinear* inversion. Crucial for the efficiency of RamBO is the RTO approach, which implies that a reliable UQ can be obtained from a small number of samples (50 or so are sufficient).

We demonstrated the use of blocky Occam and RamBO on 1D electromagnetic (EM) data sets. Our blocky models display the same structures found using Occam's inversion, but with sharper transitions and clearer distinctions between resistivity contrasts. A UQ generated by RamBO is comparable to one obtained by trans-dimensional MCMC, but RamBO is easier to implement and requires fewer forward model evaluations. In the future, we expect we or others will apply blocky Occam and RamBO to 2D and perhaps 3D problems, but our 1D formulation may still prove useful in some aspects of geophysics. For example, the hugely popular SkyTEM system (Sorensen & Auken 2004) uses hundreds of stitched 1D inversions as an interpretation

product, and might benefit from the combination of better depth resolution and minimal tuning of blocky Occam.

As explained in the introduction, we are motivated by the desire to interpret electromagnetic data, but neither blocky Occam or RamBO know nothing of the physics in the forward problem and can be deployed widely in geophysics. RamBO in particular may turn out to be a computationally inexpensive and easy-to-implement alternative to trans-D MCMC (for layered models), doing away with expensive, randomized models searches and custom-codes, and likely extended to 2D and even 3D problems.

## 7 DATA AVAILABILITY STATEMENT

The data and code used in this paper are available on Zenodo: https://doi.org/10.5281/zenodo.14769229. The code is maintained on Github at https://github.com/evargashuitzil.

## 8 ACKNOWLEDGMENTS

EVH, MM and SC are supported by NSF grant EAR-2433476. EVH and MM are supported by ONR grant N00014-21-1-2309.

### 8.1 Author Contribution

EVH and MM equally contributed to the project's conception and execution. SC took the lead on interpreting the results and deciphering geophysical accuracy and relevance of the new methodology. EVH took the lead on writing the code. All three authors wrote the paper together.

## REFERENCES

Bardsley, J. M., Solonen, A., Haario, H., & Laine, M., 2014. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems, *SIAM Journal on Scientific Computing*, **36**(4), A1895–A1910.

Blatter, D., Key, K., Ray, A., Gustafson, C., & Evans, R., 2019. Bayesian joint inversion of controlled source electromagnetic and magnetotelluric data to image freshwater aquifer offshore New Jersey, *Geophys. J. Int.*, **218**(3), 1822–1837.

Blatter, D., Ray, A., & Key, K., 2021. Two-dimensional Bayesian inversion of magnetotelluric data using trans-dimensional Gaussian processes, *Geophys. J. Int.*, **226**(1), 548–563.

Blatter, D., Morzfeld, M., Key, K., & Constable, S., 2022a. Uncertainty quantification for regularized inversion of electromagnetic geophysical data – Part I: motivation and theory, *Geophys. J. Int.*, **231(2)**, 1057–1074.

Blatter, D., Morzfeld, M., Key, K., & Constable, S., 2022b. Uncertainty quantification for regularized inversion of electromagnetic geophysical data – Part II: application in 1-D and 2-D problems, *Geophys. J. Int.*, **231(2)**, 1075–1095.

Chen, Y. & Oliver, D., 2012. Ensemble randomized maximum likelihood method as an iterative ensemble smoother, *Math Geosci.*, **44**, 1–26.

Constable, S., Parker, R., & Constable, C., 1987. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, **52**(3), 289–300.

Constable, S. C., McElhinny, M. W., & McFadden, P. L., 1984. Deep Schlumberger sounding and the crustal resistivity structure of central Australia, *Geophys. J. Int.*, **79**(3), 893–910.

DeGroot-Hedlin, C. & Constable, S., 1990. Occam inversion to generate smooth, 2-dimensional models from magnetotelluric data, *Geophysics*, **55**(12), 1613–1624.

Farquharson, C. G. & Oldenburg, D. W., 1998. Non-linear inversion using general measures of data misfit and model structure, *Geophys. J. Int.*, **134**(1), 213–227.

Fournier, D. & Oldenburg, D. W., 2019. Inversion using spatially variable mixed lp norms, *Geophys. J. Int.*, **218**(1), 268–282.

Gibbs, J., 1899. Fourier's series, *Nature*, **9**, 606.

Goldstein, T. & Osher, S., 2009. The split Bregman method for L1-regularized problems, *SIAM journal on imaging sciences*, **2**(2), 323–343.

Guitton, A. & Symes, W. W., 2003. Robust inversion of seismic data using the Huber norm, *Geophysics*, **68**(4), 1310–1319.

Gunning, J., Glinsky, M. E., & Hedditch, J., 2010. Resolution and uncertainty in 1d csem inversion: A bayesian approach and open-source implementation, *Geophysics*, **75**(6), F151–F171.

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.

Inman, J., Ryu, J., & Ward, S., 1973. Resistivity inversion, *Geophysics*, **38**, 1088–1108.

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., & Raynaud, L., 2010. Ensemble of data assimilations at ECMWF.

Key, K., 2016. MARE2DEM: a 2-D inversion code for controlled-source electromagnetic and magnetotelluric data, *Geophys. J. Int.*, **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.

Kitanidis, P. K., 1995. Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, **10**(31),

564 2411–2419.

565 Lee, J. & Kitanidis, P. K., 2013. Bayesian inversion with total variation prior for discrete geologic
566 structure identification, *Water Resources Research*, **49**(11), 7658–7669.

567 Lee, Y., 2021. $l_p$ regularization for ensemble kalman inversion, *SIAM Journal on Scientific Com-*
568 *puting*, **43**(5), A3417–A3437.

569 Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear
570 geophysical problem, *Geophysical Journal International*, **151**(3), 675–688.

571 Nocedal, J. & Wright, S. J., 2006. *Numerical Optimization*, Springer, New York, NY, USA, 2nd edn.

572 Oliver, D. S., He, N., & Reynolds, A. C., 1996. Conditioning permeability fields to pressure data.

573 Parker, R. L., 1970. The inverse problem of electrical conductivity in the mantle, *Geophysical Journal*
574 *of the Royal Astronomcial Society*, **22**, 121–138.

575 Parker, R. L., 1994. *Geophysical Inverse Theory*, Princeton, New Jersey.

576 Portniaguine, O. & Zhdanov, M. S., 1999. Focusing geophysical inversion images, *GEOPHYSICS*,
577 **64**(3), 874–887.

578 Rudin, L. I., Osher, S., & Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms,
579 *Physica D: Nonlinear Phenomena*, **60**(1), 259–268.

580 Sambridge, M., Gallagher, K., Jackson, A., & Rickwood, P., 2006. Trans-dimensional inverse problems,
581 model comparison and the evidence, *Geophys. J. Int.*, **167**(2), 528–542.

582 Sambridge, M., Bodin, T., Gallagher, K., & Tkalčić, H., 2013. Transdimensional inference in the geo-
583 sciences, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering*
584 *Sciences*, **371**(1984), 20110547.

585 Siripunvaraporn, W. & Sarakorn, W., 2011. An efficient data-space conjugate gradient Occam's method
586 for three-dimensional magnetotelluric inversion, *Geophys. J. Int.*, **186**(2), 567–579.

587 Sorensen, K. I. & Auken, E., 2004. Skytem-a new high-resolution helicopter transient electromagnetic
588 system, *Exploration Geophysics*, **35**(3), 194–202.

589 Sun, J. & Li, Y., 2014. Adaptive Lp inversion for simultaneous recovery of both blocky and smooth
590 features in a geophysical model, *Geophys. J. Int.*, **197**(2), 882–899.

591 Tang, W., Li, J., Zhang, W., Zhang, J., Geng, W., & Li, Y., 2021. Time-lapse difference inversion based
592 on the modified reflectivity method with differentiable hyper-Laplacian blocky constraint, *Geophysics*,
593 **86**(6), R865–R878.

594 Theune, U., Jensas, I. O., & Eidsvik, J., 2010. Analysis of prior models for a blocky inversion of seismic
595 AVA data, *Geophysics*, **75**(3), C25–C35.

596 Wang, Z., Bardsley, J. M., Solonen, A., Cui, T., & Marzouk, Y. M., 2017. Bayesian inverse problems
597 with l1 priors: a randomize-then-optimize approach, *SIAM Journal on Scientific Computing*, **39**(5),
598 S140–S166.

599 Wei, X. & Sun, J., 2021. *3D probabilistic geology differentiation using mixed Lp norm joint inversion*
600 *constrained by petrophysical information*, pp. 1231–1235.

## APPENDIX A: PARAMETERS USED TO CREATE FIGURE 1

A total of 50 MT amplitudes and phases logarithmically spaced between 100 Hz and 100,000 s were computed for a simple one dimensional model of a 300 m thick 1 $\Omega$m layer underlain by a 50 $\Omega$m half-space and also a model replacing the step with a sigmoid function centered on 500 m depth. The data were perturbed with normally distributed noise and inverted using a standard Occam approach. The inverted model consisted of 100 layers increasing exponentially in thickness from 1 m to 1,000 km. Regularization was a first difference between each layer, unweighted by layer thickness or depth. Noise was set to 0.3%, 1%, 3%, and 10% of linear apparent resistivity and propagated into $\log_{10}$(apparent resistivity) and linear phase, which were the inverted data. We carried out 20 inversions for each noise level to capture variations associated with the noise statistics, all converging to a root-mean-square misfit of 1.0.

## APPENDIX B: SPLIT BREGMAN WITH OR WITHOUT PERTURBATIONS

We wish to minimize the cost function

$$\mathcal{C}(x) = \|Jm - d\|^2 + \mu \left|Dm + \nu\right|, \tag{B.1}$$

with split Bregman. The perturbation $\nu$ is needed for RamBO. For blocky Occam, we simply set $\nu = 0$.

The auxiliary and Bregman variables are as in Section 2.4 and the reformulated optimization problem becomes:

$$\mathcal{C}_{\text{Breg}}(m, u) = \|Jm - d\|^2 + \mu|u + \nu| + \gamma\|u - Dm - b\|^2. \tag{B.2}$$

The reformulated optimization problem is solved by iterating the following three steps.

(i) For a given $u_k$ and $b_k$, minimize $\mathcal{C}_{\text{Breg}}$ over $m$ by solving the least squares problem

$$m_{k+1} = \arg\min_m \|Jm - d\|^2 + \gamma\|u_k - Dm - b_k\|^2 \tag{B.3}$$

(ii) Given $b_k$ and $m_{k+1}$, minimize $\mathcal{C}_{\text{Breg}}$ over $u$ by solving the optimization problem

$$u_{k+1} = \arg\min_u \mu|u + \nu| + \gamma\|u - Dm_{k+1} - b_k\|^2. \tag{B.4}$$

via soft-thresholding:

$$s_{k+1} = \text{ST}(\nu + Dm_{k+1} + b_k; 2\mu/\gamma), \tag{B.5}$$

$$u_{k+1} = s_{k+1} - \nu \tag{B.6}$$

(iii) Update the Bregman variable

$$b_{k+1} = b_k + (Dm_{k+1} - u_{k+1} - \nu). \tag{B.7}$$

We summarize split Bregman with perturbations $\nu$ in Algorithm 3, where we set the Lagrange multiplyer $\gamma = 2\mu$, as recommended by Goldstein & Osher (2009). The algorithm for split Bregman *without* perturbations, as used in the blocky Occam of Section 3, can be obtained by setting $\nu = 0$.

## APPENDIX C: ADDITIONAL FIGURES

**Algorithm 3** Split Bregman

Initialize: $u = 0, b = 0$

**while** $k \leq k_{\text{max}}$ **do**

Solve the least squares problem

$$m_{k+1} = \arg\min_{m} \|Jm - d\|^2 + \gamma \|u_k - Dm - b_k\|^2$$

Use soft-thresholding to find $u_{k+1}$

$$s_{k+1} = \text{ST}(\nu + Dm_{k+1} + b_k; 2\mu/\gamma),$$

$$u_{k+1} = s_{k+1} - \nu$$

Update the Bregman variable

$$b_{k+1} = b_k + (Dm_{k+1} - u_{k+1} - \nu).$$

**if** convergence **then**

**break**

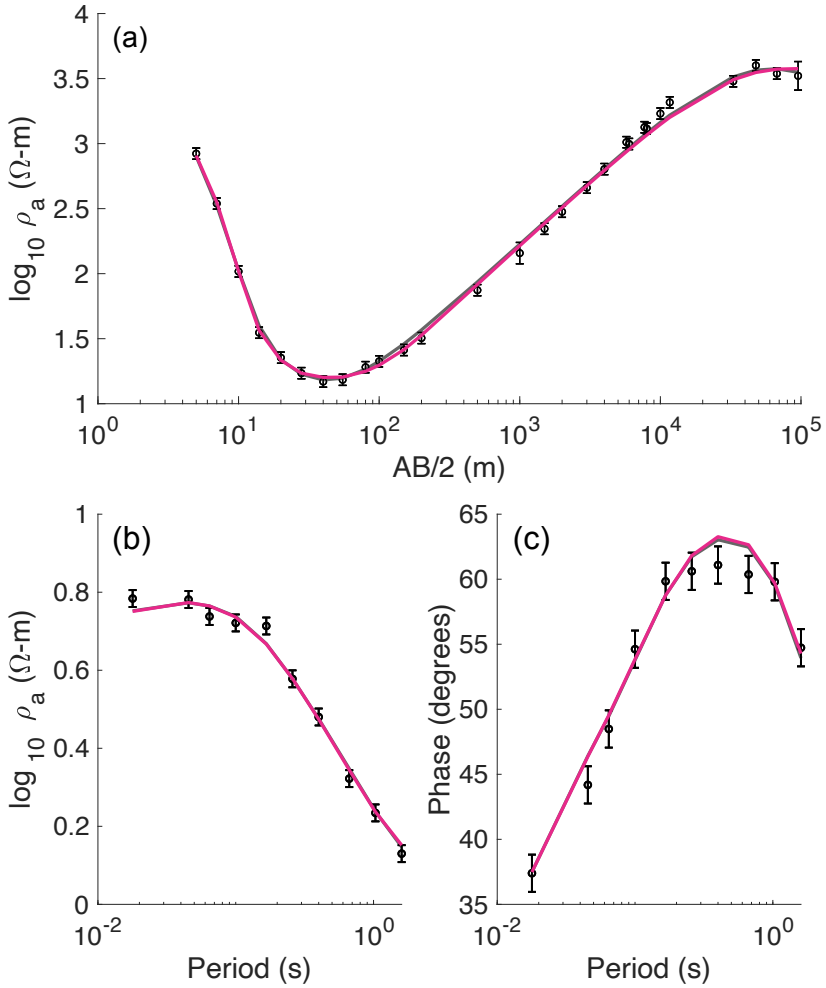
**end if**

$m_k \leftarrow m_{k+1}$

$u_k \leftarrow u_{k+1}$

$b_k \leftarrow b_{k+1}$

**end while**




[tb]

**Figure A1.** Blocky Occam compared to Occam's inversion. Panel (a) shows apparent resistivity (logspace) as a function of electrode spacing (AB/2) for the Schlumberger data set, along with error bars and the data fits of blocky Occam (pink) and Occam's inversion (gray). Panels (b) and (c) show apparent resistivity (logspace) and phase as a function of period, along with error bars. The data fits for blocky Occam and Occam's inversion are shown in pink and gray. In all panels, Occam's inversion is partially hidden by blocky Occam.

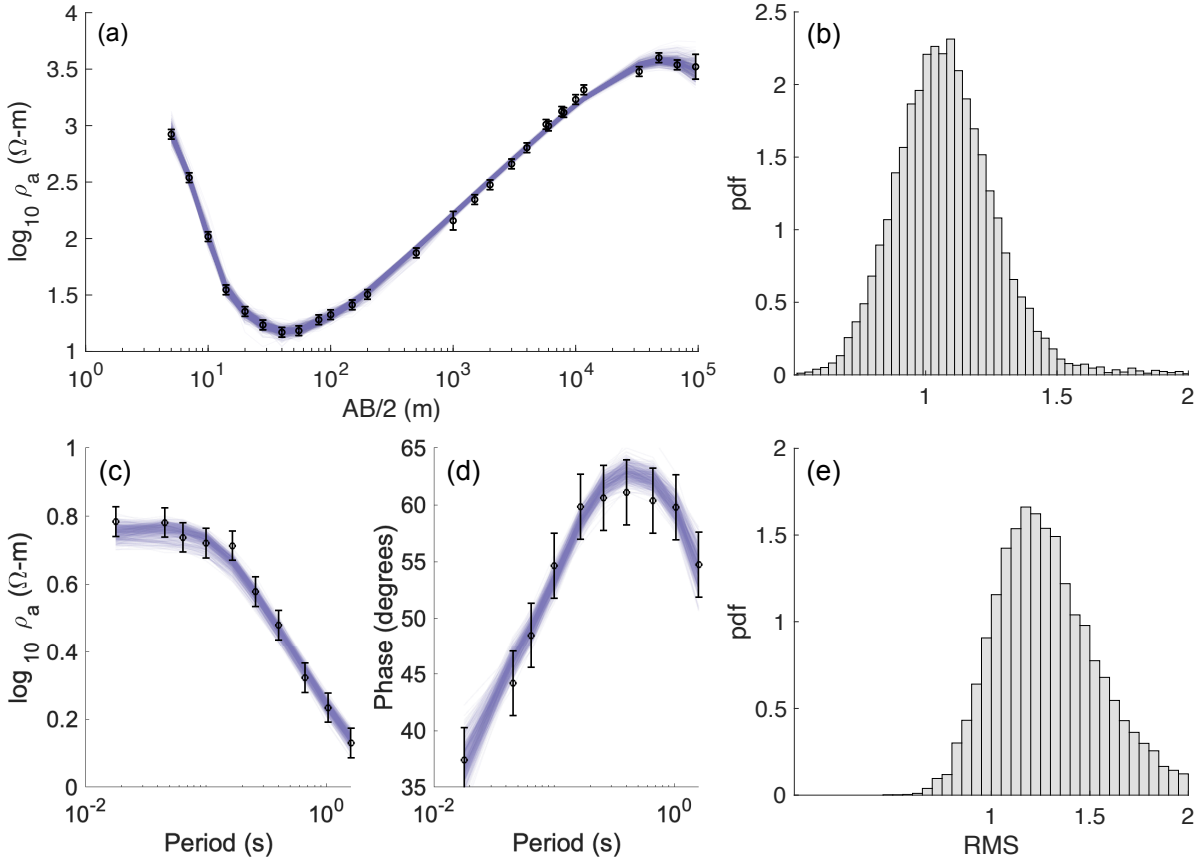



**Figure A2.** (a) Data fits of 500 models generated by RamBO for the Schlumberger data set. (b) Histogram of RMS corresponding to the models generated by RamBO (Schlumberger). (c,d) Data fits of 500 models generated by RamBO for the marine MT data set. (e) Histogram of RMS corresponding to the models generated by RamBO (marine MT).