-RESEARCH PAPER-

International Journal of Mining and Geo-Engineering

IJMGE

IJMGE 59-1 (2025) 61-67

DOI: 10.22059/IJMGE.2025.382369.595194

A non-parametric resampling method for uncertainty analysis of geophysical inverse problems

Ashkan Rahmati Shad [a], Reza Ghanati [a, *] and Mahdi Fallahsafari [a]

[a] *Institute of Geophysics, University of Tehran, Tehran, Iran.*

Article History:

Received: 14 September 2024.

Revised: 3 December 2024.

Accepted: 2 February 2025.

ABSTRACT

Due to non-uniqueness of geophysical inverse problems and measurement errors, the inversion uncertainties within the model parameters are one of the most significant necessities imposed on any modern inverse theory. Uncertainty analysis consists of finding equivalent models which sufficiently fit the observed data within the same error bound and are consistent with the prior information. In this paper, we present a non-parametric block-wise bootstrap resampling method called moving block bootstrapping (MBB) for uncertainty analysis of geophysical inverse solutions. In contrast to conventional bootstrap in which the dependence structure of data is ignored, the block bootstrap considers the dependency and correlation among the observed data by resampling not individual observations, but blocks of observations. The application of the proposed strategy to different synthetic inverse problems as well as to synthetic and real datasets of geo-electrical sounding inversion is presented. The results demonstrated that through the block bootstrap, it is possible to effectively sample the equivalence regions for a given error bound.

Keywords: *Geophysical inverse problem; Moving block-wise bootstrap resampling; Uncertainty analysis..*

1. Introduction

From deterministic and probabilistic points of views, the aim of geophysical inversion is to infer the subsurface physical properties from limited and noisy data. An integral portion of any parameter estimation is the model appraisal or uncertainty analysis of the inverse solution. The uncertainty has a statistical component associated with the data errors, and a deterministic component that accounts for the finite resolution that is attained in the model estimation as well as systematic errors in the problem. However, more reasons may exist for uncertainty in inversion results, such as linearization, incorrect assumptions (e.g., inappropriate starting model, isotropy, anisotropy, and homogeneity), and numerical approximation. Hence, it is essential to quantify the error between the estimated model parameters and the true model parameters.

The uncertainty analysis (i.e., searching equivalent models) based on the deterministic inversion approaches are quite well understood in linear inverse problems [1-3]. However, these strategies are not often sufficient for non-linear inverse problems since linear inverse theory can only allow a local estimation of the equivalent region. The deterministic-based uncertainty estimation methods depend upon the understanding of the topography of the objective function measuring the data misfit in linear and non-linear problems. An alternative strategy is to cast the problem in a stochastic framework using Bayesian inference [4], [5] and guided random search procedures, including genetic algorithm and simulated annealing [6-9]. Bayesian inference provides a systematic framework for incorporating data uncertainties, forward models, and a priori information to quantify uncertainty of the model parameters through sampling the posterior probability density function of each model parameter by means of sampling techniques, such as Markov Chain Monte Carlo methods. However, the solution of inverse problems in the context of stochastic approaches is prevented, in practice, by the curse of dimensionality and by the high computational cost needed to solve the corresponding forward problems. Therefore, a question that arises here is how to have a reliable uncertainty analysis in a relatively computationally efficient deterministic framework. Uncertainty study of geophysical models has been the subject of many research in geophysical inverse modelling in the context of single and multi-solution methods. Uncertainty analysis based on a single solution can only reflect model uncertainty due to data noise, but will not account for the inherent solution non-uniqueness, while multi-solution methods provide a set of models which sufficiently fit the observed data, and consequently, result in statistical distributions of model parameters. Indeed, in forming single model solutions, one cannot fully exploit the information content of our geophysical data [10].

In this paper, we present a non-parametric block bootstrap resampling method to quantify linear and non-linear geophysical inverse solutions uncertainty. The strategy is based on creating a number of data realizations that generate ensembles of models that fit the data within a certain tolerance in the context of deterministic inversion. The main advantage of this strategy that determines a population of models together with information on how well each model explains the observed data is that this ensemble can be used to infer the statistical properties about the model. A further preference is that sampling the solution space through the bootstrap strategy does not require any particular complicated parameter tuning [11]. The idea of the standard bootstrap resampling was introduced by Efron as a strategy to estimate confidence intervals for model parameters [12]. The conventional bootstrap method assumes that data are independent and identically distributed. However, in some cases correlations between the

* Corresponding author. *E-mail address:* rghanati@ut.ac.ir (R. Ghanati).

data can be observed, where the block bootstrap method partially maintains the underlying dependence structure and creates more realistic pseudo-samples. The bootstrap resampling algorithm has found numerous practical applications in the mathematical literature for statistical inference, such as estimating distribution functions, conducting hypothesis tests, and identifying models. Despite its proven effectiveness, it has been seldom applied for quantitatively assessing uncertainty intervals in geophysical models. For instance, McLaughlin showed that the bootstrap method can be applied to maximum-likelihood magnitude estimation for computation of the seismic event magnitude uncertainty [13]. Tichelaar and Ruff [14] and Shearer [15] proposed an uncertainty estimation of the earthquake parameters using bootstrap resampling. Parsekian and Grombacher [16] and Hertrich [17] took advantage of the conventional bootstrap statistics for estimating uncertainty in the hydro-geophysical parameters (i.e., water content and relaxation time). Schnaidt and Heinson [18] utilized a bootstrap-based model appraisal approach in magneto-telluric modelling. Campanya et al. [19] and Neukirch and Garcia [20] applied the bootstrap resampling method to estimate Magneto-telluric transfer functions. In addition to geophysical applications, Ebtehaj et al. [21] developed a block bootstrap method for hydrological parameter estimation.

This study is intended to deal with the functionality of a variant of block bootstrapping (i.e., moving block bootstrap) for assessing model parameter uncertainty in a known controlled situation using synthetic experiments and a practical situation through application to a 1D resistivity field data. In other words, the presented procedure is aimed at sampling the uncertainty space in linear and non-linear inverse modelling based on the framework of deterministic inversion of multiple data realizations. The moving block bootstrapping-based model appraisal, to the best of our knowledge, has never been applied in geophysical experiments.

Following the introduction section, this paper continues with a brief description of the methodology in section 2, giving a basic knowledge of the block bootstrap procedure with emphasis on moving block bootstrapping. Section 3 deals with the capability of the proposed resampling method through synthetic and real experiments as our numerical experiments. Finally, this paper ends by a short concluding remark and summary.

2. Methodology

The conventional bootstrap method generates samples by resampling the observed data randomly with or without replacement assuming that the observations are independently and identically distributed, and then constructs the respective empirical distribution function. The mathematical symbols applied in the formulas are as follows. We use italics for scalar quantities, boldface lowercase letters for vectors, and boldface capital letters for matrices. Suppose $\mathbf{d} = (d_1 \dots d_n)$ to be a sequence of independent and identically distributed observations. A bootstrap realization $\mathbf{d}^* = (d_1^* d_2^* \dots d_n^*)$ is generated by making n random draws with or without replacement from $\boldsymbol{d}$ for each bootstrapped data set. The bootstrap resampling is repeated until a total of k bootstrapped data sets are obtained $\mathbf{D} = \{\mathbf{d}_1^* \mathbf{d}_2^* \dots \mathbf{d}_k^*\}$. However, the performance of the original bootstrap method can be affected by dependency and correlation of observations. A remedy to this problem is to use the block-wise bootstrap to enhance the accuracy of bootstrap resampling through dividing the data samples into multiple blocks, where the blocks may be made up of non-overlapping or overlapping subsets from the original observed data. In this paper, we follow the strategy of the overlapping blocks known as moving block bootstrapping (MBB) applied by [22, 23], to a large class of weakly dependent random observations. The details of the MBB resampling strategy is summarized by Algorithm 1, as adopted for geophysical applications.

Note that an important concept in the generation of ensembles of models is the randomness in generation of each data realization that one employs, hence, the block length ζ and the number of blocks ℓ are randomly selected at each iteration of random data creation. As a result, the bootstrapping process, at each iteration, provides a variable number of the observed data, depending on the block length and the number of blocks. The MBB, as a model-free procedure represents a general non-parametric approach that can be used when no distributional assumptions are available and so it shows a behavior close to the conventional bootstrap resampling. In addition, the edge effect of the MBB can be removed using a modification of the MBB method, that is, circular block bootstrap, where by wrapping the data around in a circle before blocking ensures that all observations have the same drawing probability. As was mentioned earlier, using the MBB procedure the uncertainty space is sampled in linear and non-linear inverse problems in implementing inversion of random subsamples (i.e., bootstrapped data sets) derived from Algorithm 1. In the context of mathematics, uncertainty analysis in discrete inverse problems consists of finding an ensemble of models that adequately fit the data $\boldsymbol{d} \in \mathbb{R}^{n\times 1}$ within the same error bounds δ and that are consistent with the prior information:

$$\|\mathbf{d} - \mathbf{K}(\mathbf{m})\|_{l_2} \leq \delta \tag{1}$$

Where $\mathbf{K} \in \mathbb{R}^{n\times m}$ is a linear or non-linear forward operator depending on whether the forward problem relies linearly or non-linearly on the unknown parameters $\mathbf{m} \in \mathbb{R}^{m\times 1}$. Referring to Eq. 1, the uncertainty space is defined based on the regions that include the equivalent models satisfying the above condition for a given noise level. According to [24], the equivalent region in linear inverse problems is the portion of the model space inside the hyper-quadric surface of equivalence, whose axes are dependent on the noise level as well as on the ill-posedness of the forward matrix. This hyper-quadric surface is a very oblong ellipsoid and degenerates to an elliptical cylinder in the case of a rank-deficient forward matrix. In the case of non-linear problems, the valley has a croissant shape and can include different disconnected basins. They also showed that the geometry of the region of equivalence is affected by the noise level in both linear and non-linear inverse problems through shifting the solution found by optimization methods and deforming the topography of the cost function. To perform the uncertainty analysis, the original data set was resampled using the MBB algorithm to create different bootstrap realizations of the data. Then, all bootstrapped data realizations were inverted in the context of deterministic inversion using the same inversion parameters and the same initial models resulted in multiple solutions sufficiently fitting the resampled data. It should also be noted that the number of rows in the kernel or Jacobian matrix, respectively, for linear and non-linear problems must be the same size as the number of resampled observations at each iteration of the bootstrapping process to enable multiplication with the resampled data vector. Following the inversion of each data realization, we checked the condition presented in Eq. 1, if the criterion is met, the corresponding model can be taken into account as a member of equivalent models.

Algorithm 1. Moving block bootstrap resampling algorithm

Step 0. Given: noisy data $\mathbf{d} \in \mathbb{R}^{n\times 1} = (d_1 \dots d_n)$, ζ_{min} and ζ_{max}, where $\zeta_{min} > 1$ and $\zeta_{max} < n$.

Step 1. Define the block length ζ randomly in the range of $(\zeta_{min}\ \zeta_{max})$.

Step 2. Construct κ blocks with $\kappa = n - \zeta + 1$ from $\boldsymbol{d}$, where the ith block B_i with starting point d_i consists of ζ elements, i.e., $\mathrm{B}_i = (d_i\ d_{i+1}\ d_{i+2} \dots d_{i+\zeta-1})$ with $1 \leq i \leq \kappa$.

Step 3. The κ overlapping blocks form a matrix with κ rows and ζ columns

$$\mathbf{B}^* = \begin{pmatrix} d_1 & \cdots & d_{1+\zeta-1} \\ \vdots & \ddots & \vdots \\ d_\kappa & \cdots & d_{\kappa+\zeta-1} \end{pmatrix}_{\kappa\times\zeta}.$$

Step 4. Select randomly ℓ blocks (rows) without replacement from $\mathbf{B}^*$.

Step 5. Sort the ℓ blocks (rows) in the ascending order in terms of the number of each row, and then concatenate them together without the repeated samples to form the bootstrapped data $\mathbf{d}^* \in \mathbb{R}^{(\ell\times\zeta)-q}$, with q being the length of the overlap.

Step 1 to **5** is repeated to generate a variety of data realizations.

It is noted that similar to probabilistic approaches relying on sampling of a posteriori probability density function, the proposed method created multiple solutions with higher computational efficiency. Furthermore, we provided *a posterior* analysis using discrete cumulative density functions (see Appendix A) of the retrieved model parameters inferred from the entire bootstrapped data. The variability of the model parameters based on the upper and lower percentiles (i.e., the 1st quartile and the 3rd quartile, respectively) of the inverted models was presented to predict new data sampled and to detect the *a posteriori* probability distribution of the predicted target [11].

3. Numerical experiments

Based on the proposed algorithm for finding the equivalent models that are sampled by the MBB resampling method in the previous section, in this section, we presented a set of experiments using synthetic and real case studies to illustrate the performance of our procedure. Note that the numerical examples merely include non-linear examples due to complexity of uncertainty analysis in non-linear inverse problems; however, the proposed strategy can be easily applied and extended to the case of linear problems.

3.1. Synthetic example

The first example consists of a non-linear inversion for the near-surface application of gravity measurements along a profile, aiming at the retrieval of the depth to center z and radius r of a horizontal cylinder. The gravitational potential $g_z(rz)$ in the vertical direction resulting from an infinitely long cylinder stretched out into the plane, in the Cartesian coordinate system, can be given as:

$$g_z(rz) = 2\pi \mathrm{P}\Delta\rho(r^2 z/(x_i^2 + z^2)) \;\; i = 1 \; \ldots \; 100 \qquad (2)$$

Where $\mathrm{P} = 6.673\mathrm{e}-11\; m^3 kg^{-1} s^{-2}$ is the universal gravitational constant, and $\Delta\rho$ is the anomalous density. In this case $\Delta\rho = -1500\; kgm^{-3}$, and the depth z and the radius r were set to 10 m and 5 m, respectively. The synthetic gravity data were simulated along a profile with a length of 100 at a station interval of 1 m, then the generated data were contaminated with uncorrelated Gaussian noise with a standard deviation equal to 5% of the synthetic data mean. Fig. 1 indicates the non-linear region of equivalence for 300 resampled data using the moving block bootstrapping with the assumption that the anomalous density was known prior and thus the problem is to be solved for the unknown z and r. It should be noted that the points depicted in Fig. 1 are the results of inversion which fulfilled the criterion defined in Eq. 1. The optimization problem is non-linear and requires an iterative solution, for which a Gauss-Newton approach was used. We also showed the contour (red curve) of the equivalence region corresponding to the noise level added to the data. In addition, an approximate posterior analysis using discrete cumulative density functions of both the retrieved parameters inferred from the region of equivalence fulfilling the criterion defined in Eq. 1 is shown in Fig. 2. For the second example, we considered the inversion of a synthetic three-layered earth model with the geo-electrical parameters, including $\rho_1 = 100\; \Omega m\; \rho_2 = 150\; \Omega m\; \rho_3 = 200\; \Omega m\; h_1 = 20\; m \; and \; h_2 = 10\; m$. To create the synthetic data, a Schlumberger configuration with 25 AB/2 spreads ranging from 1.25 to 1000 m was conducted. We assumed that the resistivity and thickness of the second layer and the resistivity of the third layer were known prior; thus, the problem was to be solved for the unknown ρ_1 and h_1. The forward response of the assumed parameters was contaminated with uncorrelated Gaussian noise with a standard deviation equal to 5% of the synthetic data mean. Appendix B provides details of the theoretical formulation of the 1D resistivity forward problem for the input parameters (i.e., resistivity and thickness of each layer). In this case, we built an ensemble of resampled data, including 300 bootstrapped data realizations with a variable number of samples. The inversion of the entire bootstrapped data was implemented with the same starting models and the same starting parameters, to then assess the differences and similarities between the bootstrap models, which are indicators of model uncertainty.

The contour plots of the non-linear misfit function are displayed in Fig. 3 along with the equivalent models surrounded by the region of equivalence (black curve). Likewise, Fig. 4 shows the probability distribution of the model parameters in the region of equivalence defined based on the additive noise level.

The third example used a synthetic earth model, including a five-layer earth with a shallow and deep unconfined aquifer which were separated by a low resistive layer acting as an aquiclude basement to the shallow aquifer. The geo-electrical parameters used to simulate the forward response of the third experiment consist of $\rho_1 = 4\; \Omega m\; \rho_2 = 5\; \Omega m \rho_3 = 5\; \Omega m \rho_4 = 7\; \Omega m \rho_5 = 5\; \Omega m;\; h_1 = 2\; m h_2 = 10\; m h_3 = 10\; m - and\; h_4 = 10\; m$ with a homogenous half-space with infinite thickness. To create the synthetic data, a Schlumberger configuration with 15 AB/2 spreads ranging from 1.25 to 1000 m was conducted. The synthetic data were corrupted by uncorrelated Gaussian noise with a standard deviation equal to 3% of the synthetic data mean. The MBB algorithm was implemented to select 300 bootstrap replicated from the data set. As outlined earlier, the bootstrap replicates include fewer observations than the original data, so that using partial data information introduced ambiguity in the inversion results, made it possible sampling the equivalent model domain efficiently [11]. Two strategies for solving the geo-electrical sounding inverse problem were commonly used in terms of the layer boundaries. These methods were defined based on models with variant and invariant geometry.

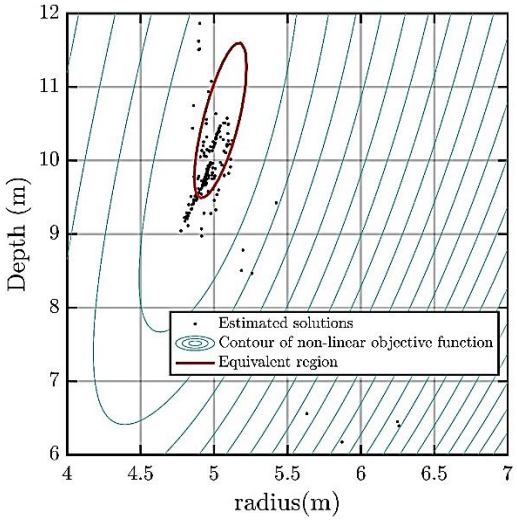

Fig. 1. The contour plot of the non-linear misfit function associated with the first synthetic experiment indicating the non-linear equivalence region (red curve) corresponding to the noise level added to the data and the results of non-linear inversion of the bootstrapped data sets surrounded by the equivalence domain fulfilling the criterion defined in Eq 1.

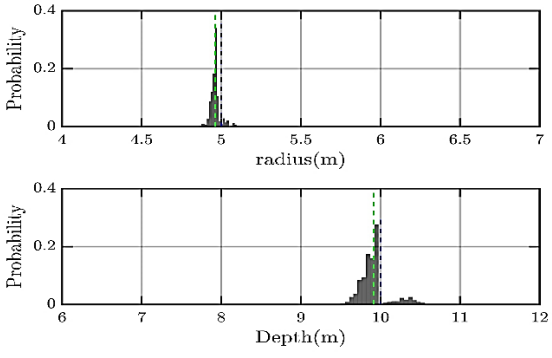

Fig. 2. Approximate posterior analysis of the sampled model parameters, including radius and depth inferred from the equivalent models shown in Fig. 1 using discrete cumulative density functions. The dashed blue and green lines indicate the true and mean values of the model parameters, respectively.

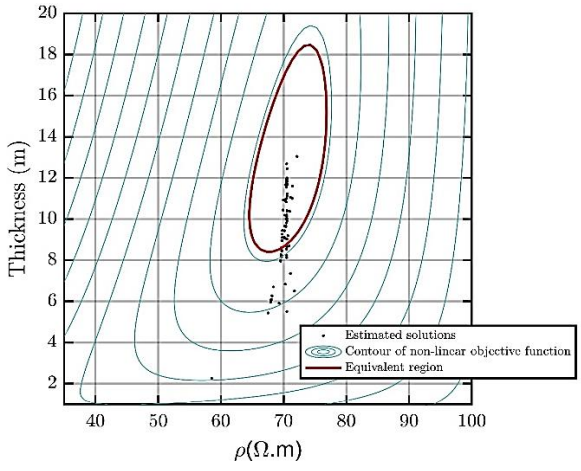

Fig. 3. The contour plot of the non-linear misfit function associated with the second synthetic experiment indicating the non-linear equivalence region (red curve) corresponding to the noise level added to the data and the results of non-linear inversion of the bootstrapped data sets surrounded by the equivalence domain fulfilling the criterion defined in Eq. 1.

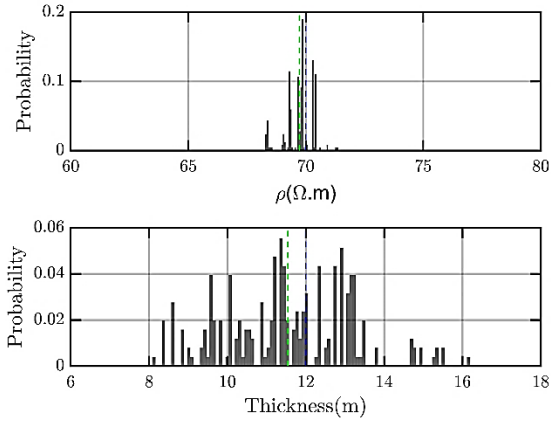

Fig. 4. Approximate posterior analysis of the sampled model parameters, including ρ and h inferred from the equivalent domain shown in Fig. 3 using discrete cumulative density functions. The dashed blue and green lines indicate the true and mean values of the model parameters, respectively.

Regarding variant geometry inversion, the models are divided into a few layers with variable boundaries where both the values of the resistivity and thickness in each layer are allowed to vary. The invariant geometry inversion is based on the assumption that the Earth is divided into many layers with fixed boundaries so that only the resistivity in each layer is allowed to vary. For the variant geometry point of view, an over-determined problem should be solved leading to a more stable solution, but the number of geo-electrical layer remains an additional unknown in the geo-electrical sounding invers problem. In other words, in spite of further stability of the inversion process due to a lower number of parameters than the observed data, the geo-electrical sounding problem is still challenging owing to its ill-posed character (i.e., equivalence problems). Aside from the role of noise, the major reason for the ill-posedness of the geo-electrical sounding inversion is the lack of an appropriate spatial coverage of the field measurements. All bootstrap data sets are inverted in the context of variant geometry strategy with the same initial models. According to Eq. 1, solutions that fulfill the condition are regarded as equivalent models. Fig. 5 indicates the resulting models sampled in the equivalence region for a given error level and the median model (red), and the lower and upper bounds using the first and third quartiles of the model parameters. In addition to a visual comparison of the geo-electrical profiles uncertainty, histograms of the recovered parameters depicting the range of the model parameters variations are shown in Fig. 6.

The last synthetic example consists of the inversion of electromagnetic induction (EMI) loop-loop data for 1D imaging of subsurface soil electrical conductivity. The forward EMI response for a given layered earth model is represented in Appendix C. The subsurface electrical conductivity model consists of a four-layer Earth with the geoelectrical parameters, including $\sigma_1 = 250$ ms. m^{-1} $\sigma_2 = 50$ ms. m^{-1} $\sigma_3 = 100$ ms. m^{-1} $\sigma_4 = 20$ ms. m^{-1} $h_1 = 1.14$ m $h_2 = 2.07$ m *and* $h_3 = 2.2$ m. Data simulation was operated using two circle loops (coplanar) as receiver and transmitter at the distance of 5 m in the horizontal and vertical dipole orientations for 16 heights ranging from 0 to 3 m above the surface. To bring the synthetic EMI data closer to field conditions, Gaussian additive noise with standard deviation equal to 5% of the magnitude of observations' mean was added to give the data set to be inverted. The four-layered Earth model for which the data were generated is shown by the blue line in Fig. 7. The synthetic data set was inverted using multi-layer inversion modelling algorithm where the subsurface model was discretized into many layers of logarithmically increasing thickness with uniform conductivity resulting in a smooth inversion. This strategy is often referred to as Occam's inversion [25], and in contrast to the blocky inversion (i.e., Levenberg-Marquardt algorithm), it includes a larger model space. We followed the strategy described for the previous synthetic experiments to create 300 replicates of the original data set. Fig.7 displays the distribution of the estimated conductivity profiles (light green lines) sampled in the equivalence region with respect to the additive noise level each with the same initial inversion parameters. Furthermore, the median model (green line), and the lower and upper bounds (red line) derived from the first and third quartiles of the inverted profiles are depicted in Fig. 7.

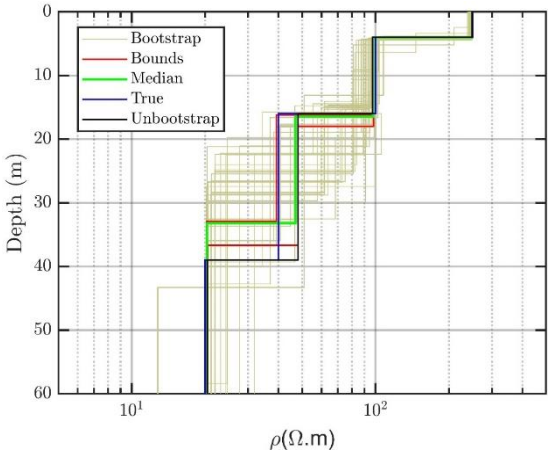

Fig. 5. Synthetic geoelectrical sounding inversion. The resulting models sampled (light green) in the equivalence region for a given error level and the median model (green) as well as the lower and the upper bounds (red) using the first and third quartiles of the model parameters, and the inverted model of unbootstrapped data (black).

3.2. Real example

As a real field example, we refer in this subsection to a real geo-electrical sounding non-linear inverse problem with the purpose of detecting hydrogeological conditions. The resistivity sounding data were acquired at the rural district of Shahid Abad in Alborz province, Northwest Iran through the Schlumberger array consisting of 23 apparent resistivity records with current electrode spacing ranging from 3 to 47 m. At this site, the borehole data and soil classification confirmed a vadose zone with a thickness of 3 m, followed by a fresh water-bearing layer composed of fine sand and clayey sand until a depth of 10 m, underlined by a 9 m mudstone followed by mudstone mixed with siltstone layer until the bottom of the borehole. Estimate of data error level is a significant factor in non-linear inversion of geophysical data, since the point at which convergence is met and also the amount of damping enforced on the data are all influenced by the level of measurement errors. Wrongly characterizing measurement errors can

bring about over- or under-interpreted structures during inversion process. In an electrical resistivity survey, the procedure of reciprocal resistance measurements was used to provide us with a rough estimation of the noise level in the observed data. The proposed bootstrap resampling, which is well adopted for geophysical applications, was implemented on the original data to generate a set of bootstrap replicates. The block bootstrap estimates were calculated by defining the block length ζ randomly in the range of ($\zeta_{min} = 4$ $\zeta_{max} = 10$), and then randomly selecting $\ell = 3$ blocks (rows) without replacement from the matrix $\mathbf{B}^*$. The resampling process was repeated to create 200 bootstrap representations of the original data. All bootstrapped data realizations were inverted using a regularized Levenberg–Marquardt in the context of variant geometry inversion (i.e., block discretization). The starting parameters for each bootstrapped data have to be identical to make them comparable and the uncertainty analysis will to some degree be dependent upon the selection of starting parameters. To control the balance between residual and stabilization terms, the regularization parameter decayed at each iteration of the inversion. In other words, the inversion process began with an adequately large value of the regularization parameter ensuring stronger stability. Then the regularization parameter was continuously reduced by a fixed factor until the convergence criteria were met, where the chi-squared value ($\chi^2 = m^{-1/2}\|\mathbf{W_d}(\mathbf{d} - \mathbf{K}(\mathbf{m}))\|_{l_2}$, where $\mathbf{W_d}$ is the diagonal matrix which includes the noise level in each datum) moved close to one. In addition to the χ^2 criteria, we also considered a convergence test based on the gradient of $\mathbf{K}(\mathbf{m})$. To ensure that the gradient of $\mathbf{K}(\mathbf{m})$ was approximately zero, assuming that the values of $\mathbf{K}(\mathbf{m})$ could be calculated with an accuracy of ε, it was required that [26].

$$\|\nabla \mathbf{K}(\mathbf{m}^k)\|_{l_2} \leq \sqrt{\varepsilon}(1 + |\mathbf{K}(\mathbf{m}^k)|) \quad (3)$$

The uncertainty analysis, including the equivalent models (light green lines) based on the solutions of the resampled data sets along with the median model (green line), and the lower and upper bounds (red lines) using the first and third quartiles of the model parameters is shown in Fig. 8. From this, we saw variable uncertainty in the resistivity profile with depth. The black solid line represents in the inversion of the original data. It is also observed that the median model obtained from the suite of models resolved the aquifer structure and was relatively compatible with the results from the original data. However, there was a slight difference in the resistivity values of the layers. Furthermore, the histograms of the inverted geoelectric parameters (i.e., resistivities and thicknesses) derived from the equivalent geoelectrical profiles shown in Fig. 8 are represented in Fig. 9.

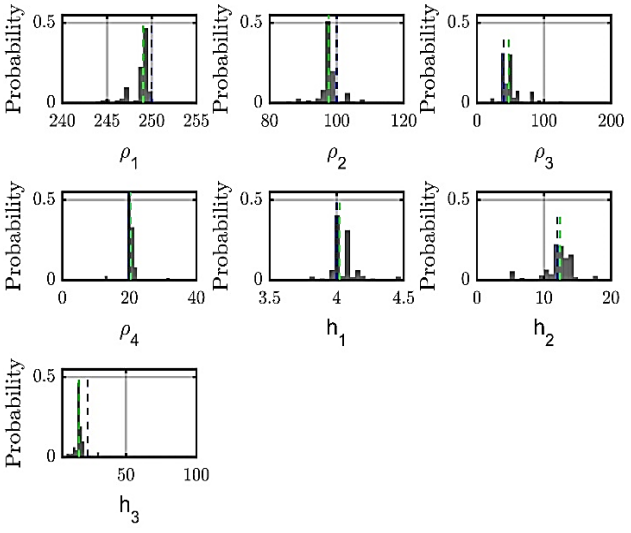

Fig. 6. Approximate posterior analysis of the sampled geoelectrical parameters inferred from the equivalent models shown in Fig. 5 using discrete cumulative density functions. The dashed blue and green lines indicate the true and mean values of the model parameters, respectively.

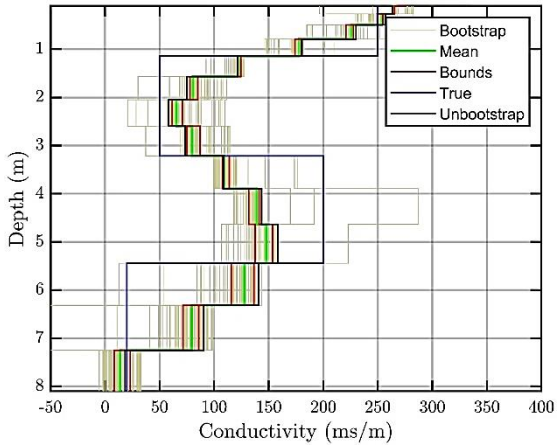

Fig. 7. Synthetic electromagnetic induction inversion. The resulting models sampled (light green) in the equivalence region for a given error level and the median model (green) as well as the lower and the upper bounds (red) using the first and third quartiles of the model parameters, and the inverted model of unbootstrapped data (black).

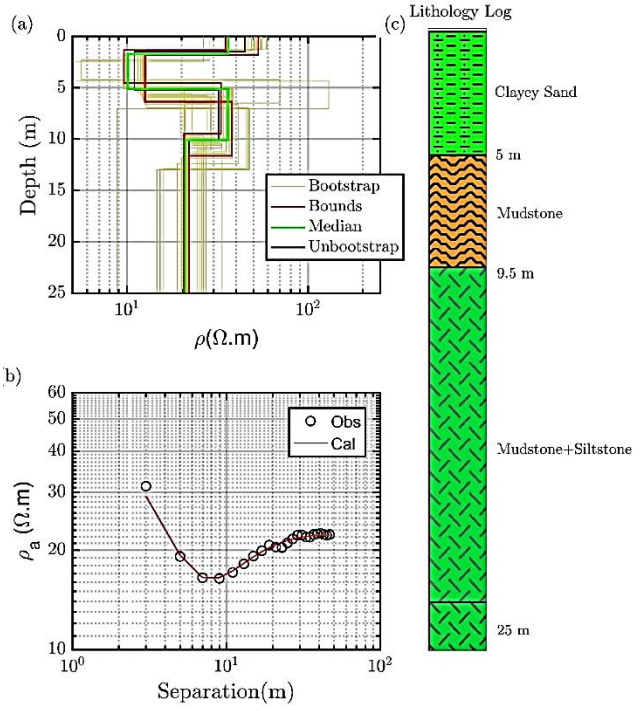

Fig. 8. Real geoelectrical sounding data inversion. a) The resulting models sampled (light green) in the equivalence region for a given error level estimated using the procedure of reciprocal resistance measurements and the median model (green), and the lower and upper bounds (red) using the first and third quartiles of the model parameters. b) Unbootstrapped field data (black circles) and the fitted model response (red line). c) Information on/about a borehole lithology in the vicinity of the survey.

4. Conclusions

The MBB method was developed and adapted as a non-parametric tool for uncertainty quantification of geophysical inverse solutions in the context of linear and non-linear problems. The MBB is an overlapping variant of the block bootstrap resampling method that can surmount the limitations of the conventional bootstrap procedure and can generally be used with any available inversion algorithm.

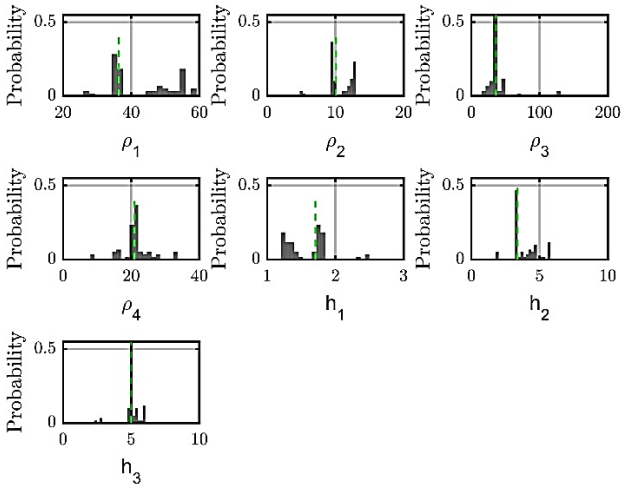

Fig. 9. Approximate posterior analysis of the sampled geoelectrical parameters inferred from the equivalent models shown in Fig. 8 using discrete cumulative density functions.

The strength of the proposed method is that it is easy to implement and accelerates the model space search for uncertainty purposes. In contrast to Bayesian-based model appraisal approaches (e.g., Markov Chain Monte Carlo methods) relying on the highly time- consuming process of sampling of a posteriori probability density function, the proposed method creates multiple solutions with higher computational efficiency, in particular, for large-scale inverse problems. Besides, the proposed method not only provides model uncertainty owing to the presence of noise but also accounts for the inherent solution non-uniqueness. We numerically show the efficiency of the presented technique through its application to different non-linear synthetic experiments in gravimetric, loop-loop electromagnetic induction, and geoelectrical sounding inversion, and to the inversion of a real geoelectrical sounding data set. The numerical results demonstrated the satisfactory application of the algorithm in sampling the different models of the equivalent region leading to a quantitative evaluation of the reliability of the inverse problem solutions. This applicable and straightforward model appraisal strategy can be extended to geophysical imaging applications and the joint inversion of multi-physics data sets.

Appendix A: Discrete cumulative density function

Give a discrete random variable M, and its probability density function $\Pr(M = m) = f(m)$, the discreet cumulative density function is defined as:

$$F(m) = \Pr(M \leq \ell) \tag{A1}$$

$$\Pr(M \leq m) = \sum_{k=m_{min}}^{m} [\Pr(M = k)] \tag{A2}$$

The cumulative density function $F(m)$: $\mathbb{R} \rightarrow [0\ 1]$ describes the probability that the random variable M with the probability distribution $f(m)$, will be found at a value less than or equal to m.

Appendix B: 1D resistivity forward modelling

The forward response of DC resistivity sounding on the surface of a horizontally stratified Earth acquired using Schlumberger configuration is written as [27]:

$$\rho_a^s = \rho_1\left[1 + 2\omega^2 \int_0^\infty G(\lambda) J_1(\lambda\omega)\lambda d\lambda\right] \tag{B1}$$

Where ρ_1 is the resistivity of the first layer, ω is the electrode spacing, $G(\lambda)$ indicates the kernel function depending on the integration variable $\lambda = {}^1/_\omega$ as well as the values of resistivity ($\rho\ (\Omega \cdot m)$) and thickness (h (m)) of all the layers and of the resistivity of the infinite substratum, and , J_1 is the first order Bessel function of the first kind.

By replacing the resistivity transform function $T(\lambda)(\Omega \cdot \mathrm{m}) = \rho_1[1 + 2G(\lambda)]$, which is calculated using the recurrence relationships from bottom to surface:

$$T_i(\lambda) = [T_{i+1}(\lambda) + \rho_i tanh(\lambda h_i)]/[1 + T_{i+1}(\lambda)\, tanh(\lambda h_i)/\rho_i] \tag{B2}$$

$$i = [M - 1\ \ldots\ 1] \tag{B3}$$

into Eq. (B-1), we derived:

$$\rho_a^s(\omega) = \omega^2\left[\int_0^\infty T(\lambda) J_1(\lambda\omega)\lambda d\lambda\right] \tag{B4}$$

Where M stands for the number of layers. The apparent resistivity values were derived from computing integral Eq. (B-3) through the application of the Ghosh's filter [27].

Appendix C: 1D electromagnetic forward response of a multi-layered earth

Forward response of electromagnetic induction for a horizontally stratified Earth acquired using vertical and horizontal dipole coil configurations is presented by the [28], which is created by the commutative electrical conductivity distribution over a certain depth range, and valid under condition of low induction number (i.e., the ratio of the coil separation divided by the plane-wave EM skin depth to be much less than unity), is given by:

$$(\sigma_a)_H = \left(\frac{-4r}{\omega\mu_0}\right) Im\left[\int_0^\infty q_0 J_0(\lambda r)\lambda^2 d\lambda\right] \tag{C1}$$

$$(\sigma_a)_V = \left(\frac{-4r}{\omega\mu_0}\right) Im\left[\int_0^\infty q_0 J_1(\lambda r)\lambda^2 d\lambda\right] \tag{C2}$$

Where $(\sigma_a)_H$ and $(\sigma_a)_V$ stands for the apparent electrical conductivity measured in horizontal and vertical dipole coil configurations, respectively, r is the spacing between the coils, J_0 and J_1 represent the zero-order and first-order Bessel functions, μ_0 is the free space permeability, ω indicates the angular frequency, λ is the radial wave number, and h_c is the height of the coils above the Earth surface, as well as q_0 corresponds to the reflection factor which was calculated recursively through discretizing the subsurface model into N layers with fixed boundaries as follows:

$$q_n(h_n, \sigma_n) = \frac{\frac{\Sigma_n - \Sigma_{n+1}}{\Sigma_n + \Sigma_{n+1}} + q_{n+1}\exp(-2\Sigma_{n+1}h_{n+1})}{1 + \frac{\Sigma_n - \Sigma_{n+1}}{\Sigma_n + \Sigma_{n+1}} q_{n+1}\exp(-2\Sigma_{n+1}h_{n+1})}, \tag{C3}$$

Where $\Sigma_n = (\lambda_n + \omega\mu_0 j\sigma_n)^{1/2}$

References

[1] Backus, G. E., Gilbert, F., (1967). Numerical Applications of a Formalism for Geophysical Inverse Problems. Geophysical Journal of the Royal Astronomical Society, 13, 1-3, 247–276.

[2] Backus, G.E., Gilbert, F., (1968). The Resolving power of Gross Earth Data, Geophysical Journal of the Royal Astronomical Society, 16, 169–205.

[3] Backus, G.E., Gilbert, F., (1970). Uniqueness in the Inversion of inaccurate Gross Earth Data, Philosophical Transactions of the Royal Society of London A, 266, 123-192.

[4] Mosegaard, K., Tarantola, A., (1995). Monte Carlo sampling of solutions to inverse problems. J Geophys Res Solid Earth 100(B7),12431–12447

[5] Gouveia, W. P., Scales, J. A., (1997). Resolution of seismic waveform inversion: Bayes versus Occam, Inverse Problems 13, 323–349.

[6] Moorkamp M, Jones AG, Eaton DW (2007) Joint inversion of teleseismic receiver functions and magnetotelluric data using a genetic algorithm: are seismic velocities and electrical

conductivities compatible? Geophysical Research Letters, 34(16):L16, 311

[7] Akca I, Basokur AT (2010) Extraction of structure-based geoelectric models by hybrid genetic algorithms. Geophysics 75(1):F15–F22

[8] Roy L, Sen MK, Blankenship DD, Stoffa PL, Richter TG (2005) Inversion and uncertainty estimation of gravity data using simulated annealing: an application over Lake Vostok, East Antarctica. Geophysics 70(1):J1–J12

[9] Wang R, Yin C, Wang M, Wang G (2012) Simulated annealing for controlled-source audio-frequency magnetotelluric data inversion. Geophysics 77(2):E127–E133.

[10] Reading, A. M., Cracknell, M. J., and Sambridge, M., (2011). Turning geophysical data into geological information or why a broader range of mathematical strategies is needed to better enable discovery. Preview, 151,24–29. https://doi.org/10.1071/PVv2011n151p24.

[11] Fernández-Muñiz , Z., Khaniani, H., and , J. L. (2019). Data kit inversion and uncertainty analysis. Journal of Applied Geophysics, 161, 228–238.

[12] Efron, B., (1979). Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1-26.

[13] McLaughlin, K.L., (1988). Maximum-likelihood event magnitude estima- tion with bootstrapping for uncertainty estimation, Bull. seism. Soc. Am., 78(2), 855–862.

[14] Tichelaar, B.W. and Ruff, L.J., (1989). How good are our best models? Jack- knifing, bootstrapping, and earthquake depth, EOS, Trans. Am. geophys. Un., 55(12), 1613–1624.

[15] Shearer, P.M., (1997). Improving local earthquake locations using the l1 norm and waveform cross correlation: application to the Whittier Narrows, California, aftershock sequence, Journal of Geophysical Research, 102(B4), 8269–8283.

[16] Parsekian, A. D., and Grombacher, D. (2015). Uncertainty estimates for surface nuclear magnetic resonance water content and relaxation time profiles from bootstrap statistics. Journal of Applied Geophysics, 119, 61–70. https://doi.org/10.1016/j.jappgeo.2015.05.005.

[17] Hertrich, M., (2008). Imaging of groundwater with nuclear magnetic resonance: Progress in Nuclear Magnetic Resonance Spectroscopy, 53, 227–248.

[18] Schnaidt, S., and Heinson, G. (2015). Bootstrap resampling as a tool for uncertainty analysis in 2-D magnetotelluric inversion modelling. Geophysical Journal International, 203(1), 92–106. https://doi.org/10.1093/gji/ggv264.

[19] Campanya, J., Ledo, J., Queralt, P., Marcuello, A. and Jones, A.G., (2014). A new methodology to estimate magnetotelluric (MT) tensor relationships: estimation of local transfer-functions by combining interstation transfer- functions (ELICIT), Geophysical Journal International, 198(1), 484–494.

[20] Neukirch, M. and Garcia, X., 2014. Nonstationary magneto-telluric data processing with instantaneous parameter, Journal of Geophysical Research, 119, 1634–1654.

[21] Ebtehaj, M., Moradkhani, H., and Gupta, H. V. (2010). Improving robust-ness of hydrologic parameter estimation by the use of moving block bootstrap resampling, Water Resources Res., 46, W07515,

[22] Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. Ann.Statist. 17 1217–1261.

[23] Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In Exploring the Limits of Bootstrap (R. Lepage and L. Billard, eds.) 225–248. Wiley, New York.

[24] Fernández Martínez, J. L., Zulima Fernández Muñiz, M and Tompkins, M. J., (2012). On the topography of the cost functional in linear and nonlinear inverse problems, Geophysics 77: W1-W15.

[25] Constable, S. C., R. L. Parker, and C. G. Constable, (1987). Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data: Geophysics, 52,289–300, doi: 10.1190/1.1442303.

[26] Aster, R., Borchers, B., and Thurber, C. (2005). Parameter estimation and inverse problems. Elsevier.

[27] Ghosh, D.P., (1971). The application of linear filter theory to the direct interpretation of geoelectrical resistivity sounding measurements Geophysical Prospecting, 19, 192-217.

[28] McNeill, J.D., (1980). Electromagnetic Terrain Conductivity Measurement at Low Induction Numbers. Tech note TN-6. Geonics Ltd, Mississauga, ON, Canada.