

## Ability of the Different Inversion Methods to Handle the data Gaps

Ren-Fen Zeng<sup>1,a\*</sup> , Cui-Yun Li<sup>2</sup>

<sup>1</sup>Jingdezhen Ceramics Institute, Jiangxi, Jingdezhen, 333000, China

<sup>2</sup>Jingdezhen Ceramics Institute, Jiangxi, Jingdezhen, 333000, China

<sup>a</sup>rfzeng@163.com

**Keywords:** Inversion problem, regularization, data gaps, smooth way.

**Abstract.** Inversion problem is always represented as different from different conditions, especially in the practical application problem such as geophysics, geodetics, ocean science and so on. One of important theory about inversion problem is Bayesian. We shortly introduce Bayesian formulation, and some analysis has been done to decide the ability of the three methods on handling the data gaps. Each method has their special way to invert the data and give inference for data gaps with zero padding, linearize interpolation or smooth way at different noise level.

### Introduction

Inversion problem is more and more concerned. There are many inversion methods. A bunch of methods have been used to determine the regularization term such as L-curve method [12,3], U-curve [4] and desired chi-square misfit [5]. Mitsuhashi et al. [6] had applied a statistical method to determine the trade-off parameter called Akaike Bayesian Information criterion (ABIC). They solve implementation of this method on the condition when no much data uncertainty is known. They did not discuss how to decide an inversion method when given a set of measured data or when some gaps have been happened.

Bayesian formula, resolving an inverse problem can be represented as updating *a priori* beliefs on the physical model by deploying information acquired from the real observations which highly result in *a posteriori* information on the model sought. Both observation data and unknown environment parameters are considered as random variables, related to each other through the realization probability [2].

In this paper, we first develop some theory for regularization method based on the Bayesian theory. Ability of the three methods to handle the data gaps has been discussed.

### Bayesian Formula

Let  $\mathbf{m}$  represent physical model vector and  $\mathbf{d}$  represent observation data generated from certain physical system. Based on the definition of the conditional probabilities, Bayes' rule may be expressed as:

$$P(\mathbf{m} | \mathbf{d}) \propto P(\mathbf{d} | \mathbf{m})P(\mathbf{m}) \quad (1)$$

Where the Posterior Probability Density (PPD),  $P(\mathbf{m} | \mathbf{d})$  represents the conditional probability for the environment parameters combining both prior information,  $P(\mathbf{m})$ , and data information  $P(\mathbf{d} | \mathbf{m})$ . Interpreting the  $\mathbf{d}$  probability given model  $\mathbf{m}$  for certain observation as a function of

parameter  $\mathbf{m}$  defines the likelihood function as follows:

$$P(\mathbf{d} | \mathbf{m}) \propto \exp[-E(\mathbf{m})] \quad (2)$$

Where  $E(\mathbf{m})$  is the data misfit function.

So following the rule of Dosso' paper (the PPD can be written as:

$$P(\mathbf{m} | \mathbf{d}) = \frac{\exp[-f(\mathbf{m})]}{\int \exp(-f(\mathbf{m}', \mathbf{d})) d\mathbf{m}'}, \quad (3)$$

Where the  $f(\mathbf{m})$  is generalized misfit which combines data and prior information defining as:

$$f(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m}) \quad (4)$$

and the integration domain spans the M-dimensional parameter space.

From the Bayesian theory view, the formula Eq.(3) represents the general solution to the physical inversion problem. How to interpret the PPD for multi-dimension inverse problem had been discussed by many papers. and this is beyond our discussion in this paper. Here we just consider the computation of parameter estimates, especially the maximum *a posteriori* (MAP) estimate (i.e., the most probable model), which are defined as following:

$$\hat{\mathbf{m}}_{MAP} = Arg_{\max} \{P(\mathbf{m} | \mathbf{d})\} \quad (5)$$

Equation (4) can be interpreted as finding an environment model which has a highest probability to give arise the observation data. In real problem we need to make inferences from finite and noisy data. There will usually be an infinity of models that fit the observation data well. Thus it is very useful to look to the probability theory for help.

### The numerical results

If we have some logging, geological or other geophysical information as *a priori* information, we can make a decision whether to use flattest method, smoothest one or even smallest. It is also a good idea to use several of them due to the non-uniqueness. For the sharp discontinuity problem with enough information of the data, it is always preferentially to take the flattest method.

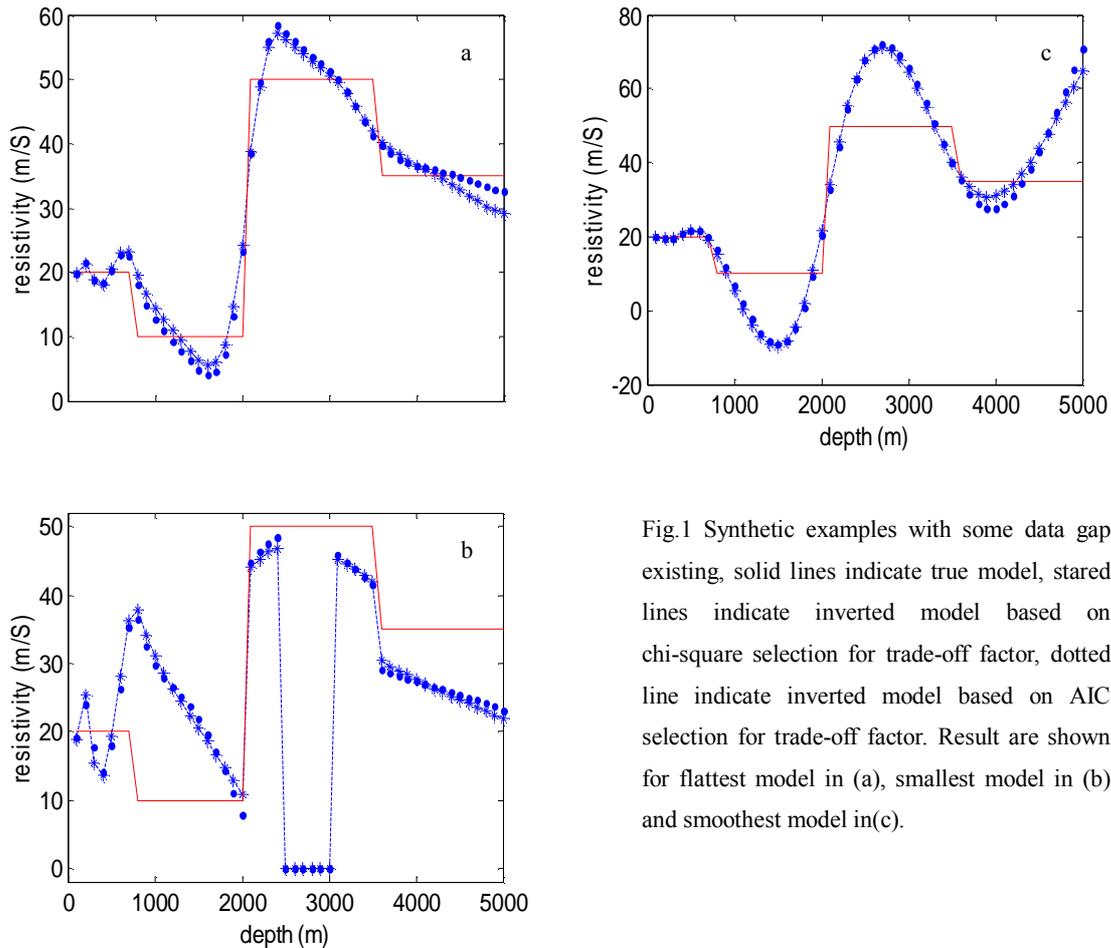


Fig.1 Synthetic examples with some data gap existing, solid lines indicate true model, starred lines indicate inverted model based on chi-square selection for trade-off factor, dotted line indicate inverted model based on AIC selection for trade-off factor. Result are shown for flattest model in (a), smallest model in (b) and smoothest model in(c).

Here we consider some gaps happened in the data which is common situation in our real application. Some comparison has done to analyze the ability and the way to handle this kind of problem for the three regularization methods, assuming data information missing from 2500m to 3000m. Fig.1(a) shows the recovery model from flattest regularization which linearly interpolates the constructed model with gaps. This provides an appropriate way to handle the missing data information. The solution in Fig.1 (c) gives a smooth approximation for which the data can't give information. This is always true and justified under the truth of physical continuity. The resulted resistivity structure is showed in Fig.1 (b) for smallest regularization. It set all the mode parameters which can't be resolved by the data to 0. Obviously it is not a good choice to pad the missing data model parameter in this way.

From the Fig.1 the star line indicate the inversion result on basis of Chi-square fitting selection, and dotted line for AIC selection. Validity can be proved for the AIC selection for the regularization parameter.

## Summary

In the situation of missing some information in some parameters, flattest regularization method can give a linearized interpolate resolution. On the other side the smoothest regularization can give a smooth solution for those parts of parameters.

## References

- [1] D. Krawczyk-Stando ,M.Rudnicki. Regularised synthesis of the magnetic field using the L-curve

approach[J].International Journal of Applied Electromagnetics and Mechanics, Vol.22,No. 3,4, 2005,pp. 233-242.

[2] D. Krawczyk-Stando , M.Rudnicki. The use of L-Curve and U-Curve in inverse electromagnetic modeling[J].Studies in Computational Intelligence, 119, 2008 ,73-82.

[3] P.C. Hansen ,D.P. O'Leary. the use of L-curve in the regularization of discrete ill-posed problem[J].SIAM Journal of Scientific computing, Vol. 14, 1993, pp. 1487-15083.

[4] D. Krawczyk-Stando ,M. Rudnicki. Regularization parameter in discrete ill-posed problems- the use of the U-curve[J]. International Journal of Applied Mathematics and Computer Science, Vol. 17, 2007,pp. 101-108.

[5] S. Dosso , M. R. Fallat.Array element localization for horizontal arrays via Occam's inversion[J]. J. Acoust. Soc. Am. 104, 1998,846-859.

[6] Y. Mitsuhashi,T. Uchida, Y. Murakami , H. Amano.The Fourier transform of controlled-source time-domain electromagnetic data by smooth spectrum inversion[J].Geophys. J. Int., **144**, 2000,123-135.