Geophys. J. Int. (1991) **106**, 387–388

Comment on 'Magnetic appraisal using simulated annealing' by S. E. Dosso and D. W. Oldenburg

Steven Constable

Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, La Jolla, CA 92093-0225, USA

Accepted 1991 February 21. Received 1991 February 21; in original form 1991 January 4

During the last few years the electromagnetic community has enthusiastically adopted extremal inversion as a tool for constructing models from data. The concept is simple: one minimizes or maximizes some penalty on the model, $\mathbf{m}$, whilst demanding that the model fit the observed data to within some reasonable level of misfit, χ_*. A Lagrange multiplier formulation is usually employed to realize this concept, so that an unconstrained functional U

$$U = R(\mathbf{m}) + \mu^{-1}\chi(\mathbf{m}), \tag{1}$$

is minimized. Dosso & Oldenburg (1991, this issue) use this approach, as can be seen by inspection of their equation (9). The first term, $R(\mathbf{m})$ is a functional of the model which returns a property which one wishes to penalize; Dosso & Oldenburg choose a boxcar average of the model. The term $\chi(\mathbf{m})$ measures the misfit obtained by the model. The relative importance of the model penalty and the misfit is controlled by the Lagrange multiplier μ^{-1}; when μ is zero minimizing U simply minimizes the misfit at whatever cost to model penalty; when μ is infinite minimizing U minimizes the penalty on the model no matter how badly the penalized model fits the data. μ is chosen so that $\chi(\mathbf{m}) = \chi_*$.

Dosso & Oldenburg (1989) solved (1) for a non-linear problem by linearizing and solving the linear problem iteratively. The purpose of their present paper is to demonstrate that the linearization obtained a global minimum by solving the same problem using simulated annealing. I think that they have achieved their purpose, and the demonstration of the use of simulated annealing in the electromagnetic context is useful. However, in both papers the authors use synthetically generated data to demonstrate the methods, but fail to include synthetic noise. This is wrong. In a classic least-squares approach failure to add noise to synthetic data usually makes little difference to the outcome but, as I will demonstrate, in an extremal inversion the result will always be biased. Even if it is argued that the effect is minimal in this circumstance, a bad example is being set. Indeed, this mistake is becoming very common.

Why, then, is it so heinous to generate accurate synthetic data and *pretend* that it includes noise (in this case 1.8 per cent) instead of actually adding noise?

Firstly, data without noise can be fit exactly by the algorithm in question, which is rarely the case for a realistic data set. Some inversion algorithms perform very much better when the data are accurate than when they are noisy, or even fail completely for noisy data.

Secondly, if misfit *is* allowed in an extremal inversion, as it should be, then the misfit budget will be spent selectively to increase the penalty on the model. This selective partitioning of the misfit has two undesirable consequences: the model will be penalized more than would be the case if the misfit budget was spent accommodating the variations in data, and the data residuals will be serially correlated, violating assumptions about properties of data noise, including the use of the expected value for $\chi(\mathbf{m})$. I will illustrate these points with an extremely simple example. Let us consider the underlying forward problem to be a line:

$$f(x, \mathbf{m}) = a_1 x + a_2, \tag{2}$$

where the model vector is $\mathbf{m} = (a_1, a_2)^{\mathrm{T}}$. We will seek an extremal model in which the magnitude of the line's slope is minimized; the penalty term becomes $\|\mathbf{Rm}\|^2$ where the 2×2 matrix $\mathbf{R}$ is

$$\mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Let us take 20 data $\mathbf{d} = (d_1, d_2, \ldots, d_{20})^{\mathrm{T}}$ and construct a diagonal weighting matrix $\mathbf{W} = \mathrm{diag}\,(1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_{20})$ from the data errors σ_i. The usual two-norm misfit measure is then $\chi(\mathbf{m}) = \|\mathbf{Wd} - \mathbf{WGm}\|^2$ where $\mathbf{G}$ is a 20×2 matrix describing the forward problem; the first column is $(x_1, x_2, \ldots, x_{20})^{\mathrm{T}}$ and the second column is filled with ones. Our unconstrained functional is

$$U = \|\mathbf{Rm}\|^2 + \mu^{-1}\|\mathbf{Wd} - \mathbf{WGm}\|^2. \tag{3}$$

Following Constable, Parker & Constable (1987), U is minimized by differentiation with respect to $\mathbf{m}$ and rearrangement to yield an expression for the model:

$$\mathbf{m} = [\mu \mathbf{R}^T\mathbf{R} + (\mathbf{WG})^{\mathrm{T}}\mathbf{WG}]^{-1}(\mathbf{WG})^{\mathrm{T}}\mathbf{Wd}. \tag{4}$$

Application of (4) will produce the models with smallest slopes that fit the data to various levels of misfit, depending on the choice of μ (since the problem is linear no iteration is required). With Gaussian noise, μ is usually chosen so that the misfit is the expected value of the chi-squared distribution, or a root mean square (rms) residual of one standard error.

Now we examine a synthetic model study in which the data are generated from a known model, $f(x) = 1.0x + 1.0$. Fig. 1(A) shows 20 data generated from this model and assumed to have errors of 0.75, without actually adding any noise. If we fit a minimum slope model to rms 1.0 the entire

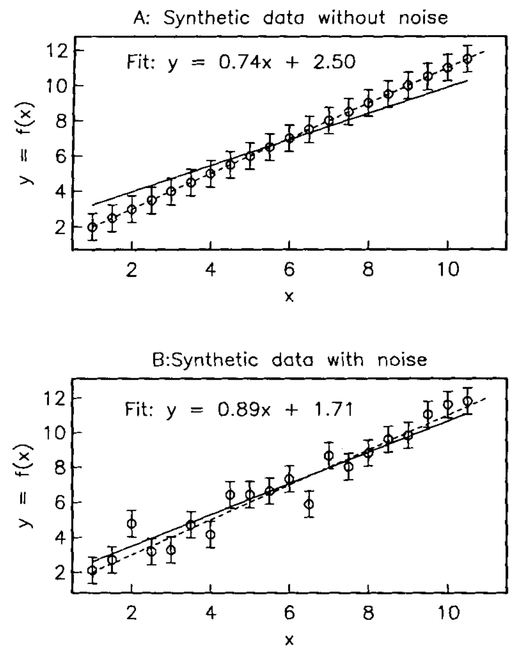

Figure 1. Two sets of synthetic data generated from the model $f(x) = 1.0x + 1.0$. (A) No noise has been added to the data, but errors of 0.75 are assumed. (B) Gaussian noise with zero mean and standard error 0.75 has been added to the data. The broken line in both plots shows the model used to generate the data.

misfit budget is used to lower the slope to 0.74. Also, from inspection it will be clear that the data residuals are serially correlated.

We can generate noise which is approximately Gaussian using the algorithm $\varepsilon = \sigma\sqrt{-2\ln(r_1)}\cos(2\pi r_2) + \bar{x}$, where r_1 and r_2 are random numbers uniformly distributed in [0, 1] and σ and $\bar{x}$ are the desired standard deviation and mean. When we add noise (zero mean, 0.75 standard deviation) to the synthetic data and fit a minimum slope line (Fig. 1B) most of the misfit budget must be used to fit the noise in the data, and we achieve a larger (less penalized) slope. Note that our objective is *not* to recover the starting model; that is always impossible with noisy (realistic) data, although in this case simple least squares would provide the most efficient attempt. This simple example illustrates more practical and complicated situations where the extremum sought has some useful interpretation, such as in smooth modelling (Constable *et al.* 1987) or, as in Dosso & Oldenburg (1989, 1991), extremal averages. A smooth model with more structure or a smaller extremal average places tighter constraints on Earth structure than otherwise.

While the minimum slope for the no-noise case will always be 0.74, the slope obtained when noise is added depends on the particular realization of noise used; in the example it is 0.89. For 200 independent realizations of noise, the mean minimum slope was 0.88 and its standard deviation 0.08. The distribution of minimum slopes was indistinguishable from Gaussian and so a slope smaller than 0.74 will only appear 2 per cent of the time. There is no right answer; the variability in extremal solutions is part of the problem resolution being examined by synthetic modelling. Of course, if two different inversion techniques are being compared, as in Dosso & Oldenburg, then the same realization of noise must be used for both.

In conclusion, failure to add noise to computed data in synthetic model studies not only fails to provide a realistic test of inversion algorithms, but in extremal inversions will bias the result significantly. The bias is in a direction which reduces the apparent resolving power of the experiment being modelled.

REFERENCES

Constable, S. C., Parker, R. L. & Constable, C. G., 1987. Occam's Inversion: a practical algorithm for generating smooth models from EM sounding data, *Geophysics*, **52**, 289–300.

Dosso, S. E. & Oldenburg, D. W., 1989. Linear and non-linear appraisal using extremal models of bounded variation, *Geophys. J. Int.*, **99**, 483–495.

Dosso, S. E. & Oldenburg, D. W., 1991. Magnetotelluric appraisal using simulated annealing, *Geophys. J. Int.*, **106**, 379–385.