Geophys. J. Int. (2004) **157,** 583–588 doi: 10.1111/j.1365-246X.2004.02272.x

FAST TRACK PAPER

Consistency regions in non-linear inversion

B. L. N. Kennett

Research School of Earth Sciences, Australian National University, Canberra ACT 0200, *Australia. E-mail: brian@rses.anu.edu.au*

Accepted 2004 February 20. Received 2004 February 3; in original form 2003 December 22

SUMMARY

A disadvantage of fully non-linear methods of inversion, through exploitation of the properties of model space, is the absence of a well-developed framework for error assessment. To rectify this problem an auxiliary weighting function for ensemble properties is introduced that can be used with suitable thresholds to define consistency regions of suitable models. This approach requires neither a detailed knowledge of the misfit distribution nor an underlying probabilistic model. The use of a polyhedral representation of such consistency regions is illustrated with an example from non-linear seismic event location, showing the effect of different choices for the misfit measure.

The auxiliary weight function can be used directly with the composite misfit measure used to drive the exploration of model space in the non-linear inversion. Such composite measures usually combine a data misfit and regularization term. However, considerable benefit can be obtained by storing the data misfit and associated model characteristics for each investigated model, as well as the composite measure. The properties of the model ensemble can then be used retrospectively to define preferred models by the intersection of a consistency region in data misfit with zones constrained by desirable model properties.

Key words: event location, model uncertainty, non-linear inversion.

1 INTRODUCTION

Non-linear inversion methods involving exploration of parameter space such as genetic algorithms (e.g. Gallagher *et al.* 1991), simulated annealing (e.g. Sen & Stoffa 1991) or the neighbourhood algorithm (Sambridge 1999a,b) can be very effective at finding models with a suitable fit to data. These methods do not require the construction of numerical derivatives and can use any convenient measure of consistency between the observations and the prediction from a proposed model. A recent review is provided by Sambridge & Mosegaard (2002).

However, unlike conventional Gauss–Newton linearization schemes, the fully non-linear methods do not provide an immediate product that can be used to characterize the likely uncertainties in model parameters. For low-dimension systems, such as the seismic event location problem with four hypocentral parameters, a systematic search around the preferred model can map out the pattern of misfits (e.g. Billings *et al.* 1994). With an appropriate probability distribution for the misfit, the contours of misfit can be interpreted as confidence levels.

Sambridge (1999b) has shown how to exploit the characteristics of model space in a probabilistic interpretation of uncertainty and resolution analysis. The ensemble of models generated in the exploration of parameter space is augmented by resampling using a neighbourhood algorithm, so that suitable probability integrals can be calculated by Monte Carlo integration. This approach has been used recently by Resovsky & Trampert (2002) in the assessment of the properties of Earth models consistent with free oscillation and surface wave data.

A common approach to the delineation of uncertainty in non-linear inversion is to seek the region in parameter space where the misfit to the observed data is less than some prescribed threshold. This can be attempted by direct mapping of the properties of parameter space or by adapting the non-linear inversion approach to preferentially map acceptable models. Thus, Lomax & Snieder (1994, 1995) have modified a genetic algorithm to encourage the extraction of acceptable models, rather than perform a global optimization. Sambridge (2001) has shown how different classes of modified data fit functions can be used with a neighbourhood algorithm to concentrate attention on acceptable models. This approach can work well even where there are multiple, disjoint regions of parameter space with a comparable fit to data.

The approach developed here is somewhat different in that the properties of the ensemble of models tested in the exploration of parameter space are used to determine a consistency region within which there is comparable representation of the observations. It is convenient to use an auxiliary weighting function to specify the appropriate threshold, because this avoids dependence on the particular form of misfit function employed.

2 ENSEMBLE PROPERTIES IN INVERSION

The process of exploration of parameter space in the non-linear inversion schemes is directed toward minimizing some measure of model suitability E, which will always contain some measure Φ of the misfit between observations and model predictions, but frequently will include a regularization term Ψ to secure, for example, adequate smoothness in a model. Thus,

$$E(\mathbf{m}) = \Phi(\mathbf{m}) + \Psi(\mathbf{m}) \tag{1}$$

for model parameters $\mathbf{m}$. Many different choices can be made for the misfit $\Phi(\mathbf{m})$, such as an L_p norm of residuals. The regularization term $\Psi(\mathbf{m})$ may include a measure of deviations from a reference model $\mathbf{m}_{\mathbf{ref}}$, constraints on gradients etc.

Each of the sampled models $\mathbf{m}_i$ will therefore be represented through the composite measure E_i, the data misfit Φ_i and a regularization term Ψ_i. As we shall see, it may be advantageous to keep track of a range of different properties of the model.

We introduce an auxiliary weighting function $w(E)$ so that we can estimate ensemble properties via, for example,

$$\langle p \rangle = \sum_i w(E_i) p(\mathbf{m}_i) \Big/ \sum_i w(E_i), \tag{2}$$

where the summation is taken over the full ensemble of models traversed in the inversion, or some subset. The weighting function $w(E)$ should be some monotonically decreasing function of the composite measure E so that the ensemble estimate $\langle p \rangle$ emphasizes the properties of those models that are most suitable, i.e. with smaller E. Such a weighting scheme by inverse misfit [$w(E) \propto 1/E$] was used by Shibutani *et al.* (1996) in inversion of receiver function waveforms using a genetic algorithm to derive stable estimates of shear velocity profiles from the best 1000 models sampled.

Ensemble results can be applied directly to the model parameters or to secondary quantities constructed from the model parameters. In the location problem for seismic events, the hypocentre representing the point of initiation of seismic energy in space and time can be represented by a four-vector $\mathbf{h}$ whose elements are the origin time, latitude, longitude and depth of the event. The ensemble location estimate

$$\langle \mathbf{h} \rangle = \sum_i w(E_i) \mathbf{h}_i \Big/ \sum_i w(E_i) \tag{3}$$

will be distinct from the best-fitting estimate $\hat{\mathbf{h}}$ and should be less susceptible to noise in the observations.

We can envisage two different approaches to the choice of the weighting function $w(E)$. The first would be related to the expected probability distribution for E and is thus strongly dependent on the choices made for the data misfit Φ and the regularization Ψ. Even when the behaviour of the residual distribution is well understood, the influence of the regularization term will be difficult to assess. With this style of weighting it will also be difficult to compare results for different measures of fit. The alternative is to adopt a standard form for the weighting function with some tunable parameters, that can be adapted to the problem at hand. The same scheme can then be used for different criteria of fit for the same data set.

From a range of trials of suitable monotonically decreasing functions $w(E)$, we have found that a suitable form is based on the Fermi–Dirac distribution from statistical physics (see, e.g. Schrödinger 1952):

$$w(E) = [\exp\{\beta(E - E_r)/E_0\} + 1]^{-1}, \tag{4}$$

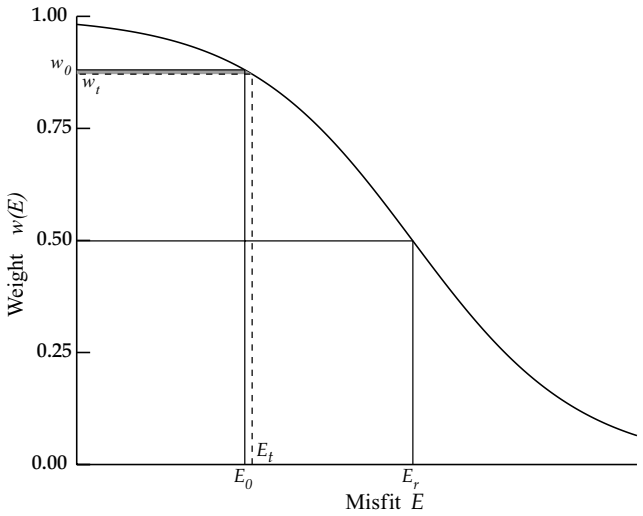

Figure 1. Illustration of the Fermi–Dirac weighting function $w(E)$ and the imposition of a threshold on the weights (w_t) that defines a band of consistent models indicated by the shaded region.

where β controls the rate of change of the weighting function and hence the emphasis to be placed on different aspects of the ensemble. E_r is a reference point such that $w(E_r) = 0.5$. We normalize the misfits with a base composite measure E_0, taken to be slightly less than the smallest value encountered in the ensemble $E_{\min}$; we have used $E_0 = 0.999 E_{\min}$.

The Fermi–Dirac distribution $w(E)$ (eq. 4) is illustrated in Fig. 1 using the parameters ($\beta = 2.0, E_R = 2.0\, E_0$) employed in the seismic location example below. Commonly in statistical physics, β, representing inverse temperature, is large and then $w(E)$ is close to unity until the vicinity of E_r, when there is a rapid transition to near zero values. However, for the current application it is preferable to use a smaller value of β, as shown in Fig. 1, when there is a gentle but steady reduction in $w(E)$ for small values of E with maximum gradient at the reference value E_r. Large values of the composite misfit measure E receive little weighting, because the tail is essentially a negative exponential with respect to misfit.

3 CONSISTENCY REGIONS

The use of the auxiliary weighting function allows a means of extracting a subset of the ensemble of investigated models with similar composite measure properties to the best found in the inversion.

The weights $\{w(E_i)\}$ are evaluated for all the models and we then define a consistency region for which $w > w_t$. For the Fermi–Dirac distribution we take

$$w_t = \frac{1}{2} + t\left[w(E_0) - \frac{1}{2}\right], \tag{5}$$

where the threshold value t is chosen to include a suitable zone near the misfit minimum, e.g. $t = 0.979$ has been employed in the example illustrated in Section 4. For the choice $E_r = 2\ E_0$ the threshold weight can also be written as

$$w_t = \frac{1}{2}\left(1 + t \tanh\frac{\beta}{2}\right). \tag{6}$$

The specification of the threshold on w defines a band of models with consistent misfit properties, as indicated by the shading of the weights in Fig. 1. This subset of the ensemble may then be used

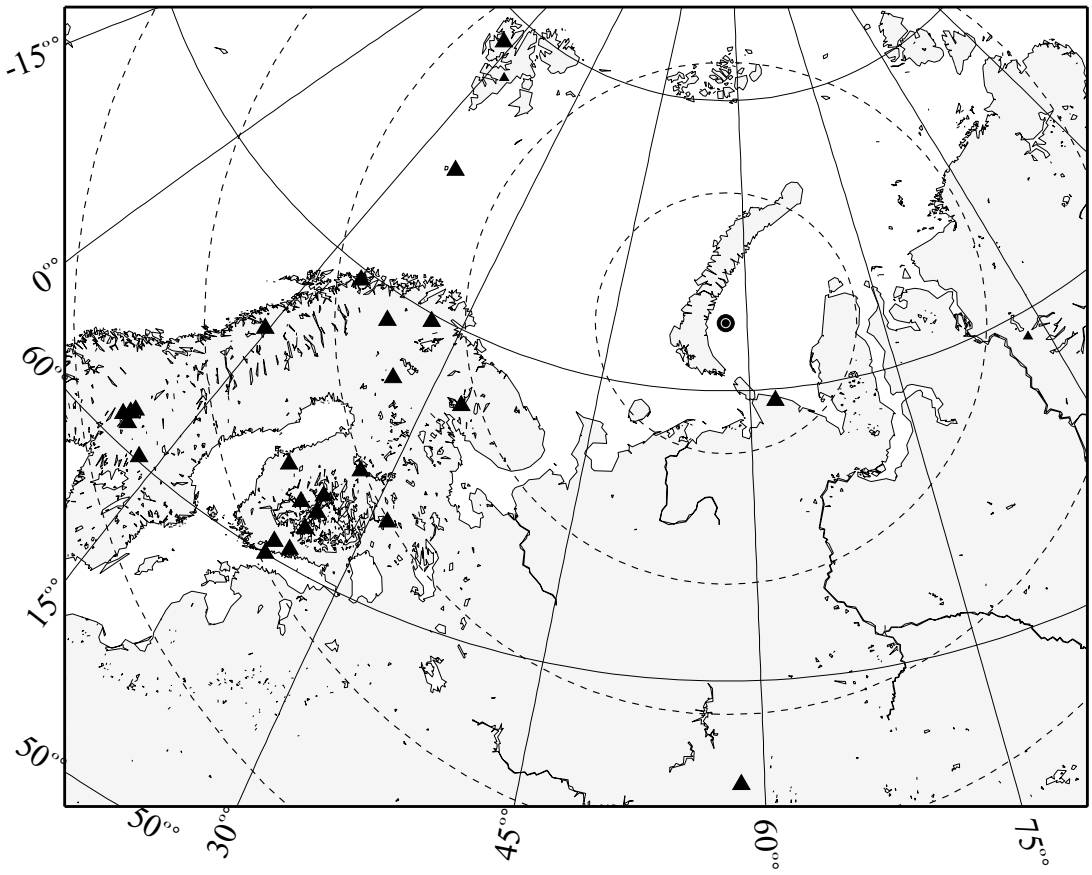

Figure 2. Location of the Kara sea event of 1997 August 16 (circle) and the seismic stations (triangles) with phase readings. The dashed circles show distances from the event in increments of 500 km.

to give a stable estimate of model properties using a sub-ensemble average as in eq. (1). Alternatively, the band may be viewed as defining a consistency region in model parameter space. For simple problems, this region can be adequately defined using a polyhedral representation.

The choice of the threshold w_t for the Fermi–Dirac distribution is equivalent to specifying a composite misfit measure E_t:

$$E_t = E_r + E_0 \frac{1}{\beta} \ln\left(\frac{1}{w_t} - 1\right), \tag{7}$$

as indicated in Fig. 1. However, there are considerable merits in working with the properties of the auxiliary function $w(E)$ rather than making direct use of the misfit values E:

(i) A consistency region can be defined, based purely on the properties in parameter space, for choices of composite measure E that do not correspond to standard probability density functions with known confidence levels.

(ii) We have a standard weighting scheme, which can be applied for different styles of misfit measure in a comparable way. If needed, the shape of the distribution can be tuned to the circumstances by the choice of β, E_r. A larger value of β will, for example, give a tighter band of misfit for the same weighting threshold.

(iii) The details of the misfit distribution are not used, just the relative weight: hence, the process is akin to the use of model rank as a driver for the exploration of parameter space, as in the neighbourhood algorithm (Sambridge 1999a, 2001).

(iv) It is easy to explore the consequences of choices of threshold on the nature of the consistency region. Such choices can be related to probabilistic models for the misfit distribution, but are not forced by specific assumptions.

The consistency region approach can be used directly with the original ensemble generated during an inversion and does not need subsequent resampling. However, it can be readily combined with the augmented ensemble approach of Sambridge (1999b) to provide probabilistic estimates.

Because the consistency region is defined retrospectively, it can be used with auxiliary information on model properties, e.g. smoothness, to single out those regions in parameter space with certain characteristics. Indeed multiple criteria on model properties can be applied provided that the requisite information is collected in the course of the inversion.

4 CONSISTENCY REGIONS IN SEISMIC EVENT LOCATION

As an easily visualized example of the use of consistency regions, we consider the location of a seismic event at far regional distances. We use an event in the Kara Sea in 1997 August, near the former USSR nuclear test site on Novaya Zemlya, whose nature has been the subject of some debate (see, e.g. Schweitzer & Kennett 2002, 2004). As can be seen from Fig. 2, the distribution of available stations is rather uneven. The best results are obtained with the use of regional models for the Barents sea region, but there is sufficient lateral heterogeneity that it is difficult to represent the pattern of observed times well with a single model.

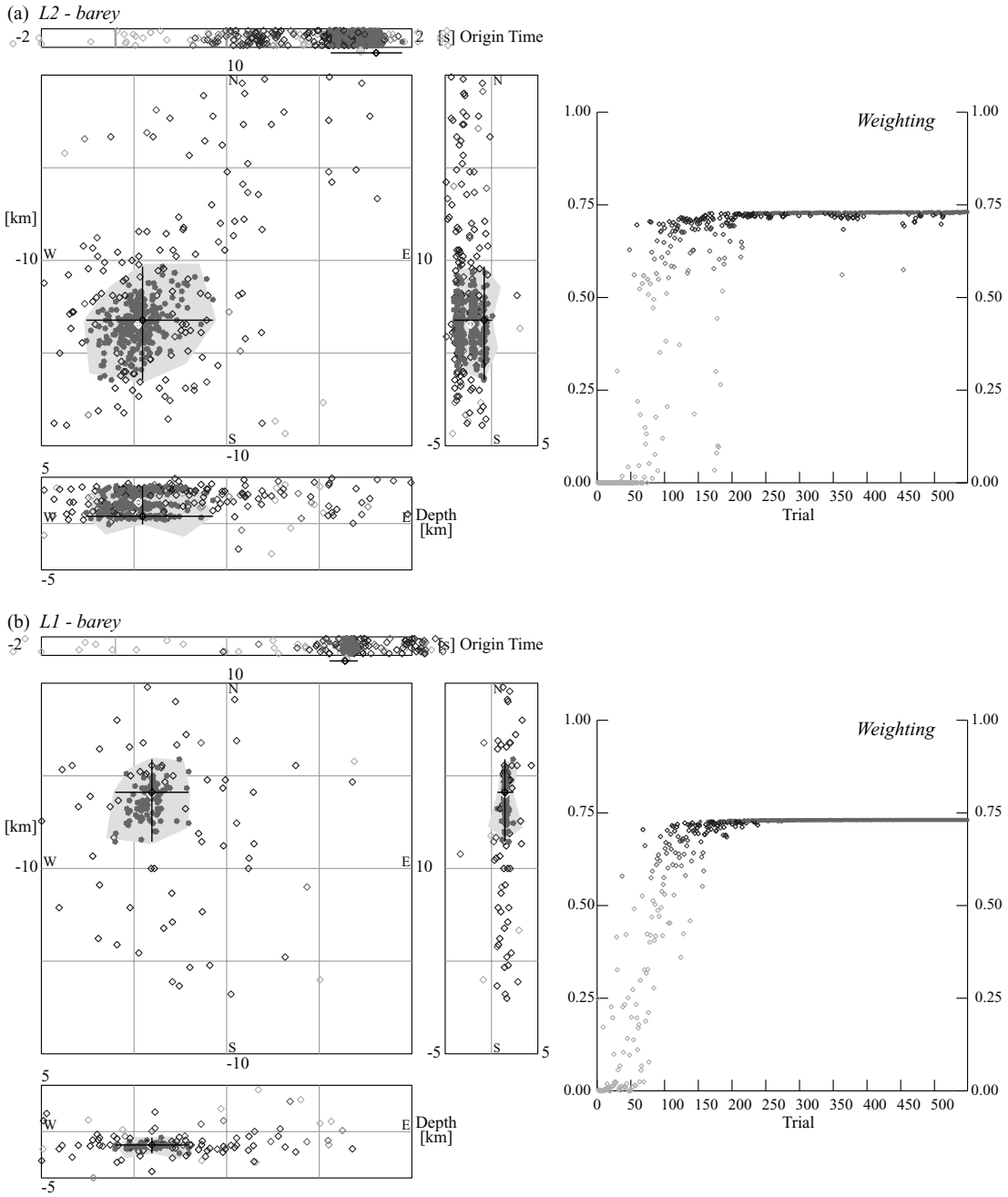

Figure 3. Representation of the progress of the neighbourhood algorithm scheme towards convergence on a location estimate, with definition of the consistency region of data fit from the properties of the weighting function. Only a small portion of the model space about a reference point (72.4°N, 57.9°E, 30 km depth) is shown. The open symbols get darker as data fit improves. Models within the consistency region are indicated by solid pentagons; a polyhedral representation of the region is indicated by a grey tone. The best-fitting location is marked by a black diamond with a grey centre and the ensemble estimate by a grey open diamond. (a) L_2 data misfit function with the *barey* model; (b) L_1 data misfit function with the *barey* model.

We therefore consider the effect of using different measures of misfit with the regional model, *barey* (Schweitzer & Kennett 2004). The inversions for the hypocentral location are carried out using the neighbourhood algorithm (Sambridge & Kennett 2001), with an extended number of iterations to provide good sampling of parameter space in the region of good fit to data. In Fig. 3 we compare the behaviour of the results for (a) the conventional L_2 norm of residuals and (b) the more robust L_1 norm, which is less sensitive to the deficiencies of the simple 1-D regional model. In each case we have used a weighting function directly on data misfit Φ with $\beta = 2.0$, $\Phi_0 = 0.999\Phi_{\min}$, $\Phi_r = 2\Phi_0$: where $\Phi_{\min}$ is the least misfit in the location ensemble. The weighting threshold is specified by $t = 0.979$.

Fig. 3 shows the progress of the neighbourhood algorithm for the location of this Kara sea event, through the evolution of the weighting function and projections of the location estimates for a small region near the best fitting location (±20 km in horizontal position, ±10 km in depth and ±2 s in origin time). The parameter space searched comprised an 800 × 800 km^2 region in horizontal position, ±40 km depth in depth and ±10 s in origin time centered on the

position 72.5°N, 57.5°E and 40 km in depth. The open symbols, for estimates whose weights lie below the threshold, get darker as the data misfit is reduced. Estimates whose weights lie above the threshold are shown as solid pentagons. The polyhedral representation of the consistency regions is indicated by the grey tone.

The choice of threshold $t = 0.979$ was made after a few trials to give compact consistency regions. Because this assessment is retrospective, it can be carried out rapidly using the stored properties of the trial models.

The same form of threshold weighting was applied to the L_2 and L_1 location results. Both Figs 3(a) and (b) are drawn for a small region around the same reference point (displaced from the centre of the original search space). The distinct change in the location estimates arises because of the strong lateral heterogeneity across the region. The locations with the L_2 norm for data misfit have problems with poor fits to some phase picks and this enlarges the zone of comparable misfit, particularly in depth and origin time. The use of the L_1 norm compensates for the limitations of the 1-D velocity model (*barey*) by reducing the influence of the difficulty of fitting some stations and phases. With the same weight threshold, it is clear that the consistency region for the L_1 norm is tighter, notably in time and depth. The convergence of the location estimates was more rapid for case (b) and there is closer agreement between the best-fitting model found and that determined as an ensemble average over the models within the consistency region.

Through the use of the auxiliary weighting scheme, we are able to make a direct comparison of the consistency regions for the two different choices of data misfit function. A polyhedral representation of the consistency zone simply provides a summary of the collection of models whose weights lie above the threshold for the given misfit function. In this simple example, the projection of the hypocentres is sufficient to provide a visual assessment, but this will be more complex for problems with more parameters.

5 DISCUSSION: CONSISTENCY ZONES COMBINING DATA FIT AND MODEL PROPERTIES

The exploration of model parameter space for suitable models to explain observations is most easily performed using a single measure to represent the behaviour of the model. This is why composite measures incorporating both data misfit and regularization are commonly employed. The search in parameter space is then directed towards models that have both good fit to data and desirable properties.

However, although the progress of the inversion may be conveniently controlled by the minimization of a single parameter, it is little extra effort to also store the ensemble of data misfit values $\{\Phi_i\}$ and the set of measures of model properties $\{\psi_i\}$, which can be combined to produce the regularization term $\{\Psi_i\}$.

In the assessment stage of the inversion we can then make use of the ensemble of properties $\{\Phi_i\}$, $\{\psi_i\}$ across the samples in model space to find consistency regions that make direct use of the data misfit and model characteristics (*cf.* Kennett 1978). There are a number of different ways in which such regions can be defined depending on the relative emphasis placed on data fit and regularization.

Rather than develop a consistency region for the composite measure E via $w(E)$, a data consistency zone can be constructed by applying the threshold test to the data misfit using $w(\Phi)$. Within this region with suitable data fit properties, we can then map out the associated properties of the models through the summary measures $\{\psi\}$. The subset of the consistency region for data fit with appropriate model properties can then be regarded as the desired consistency zone of suitable models.

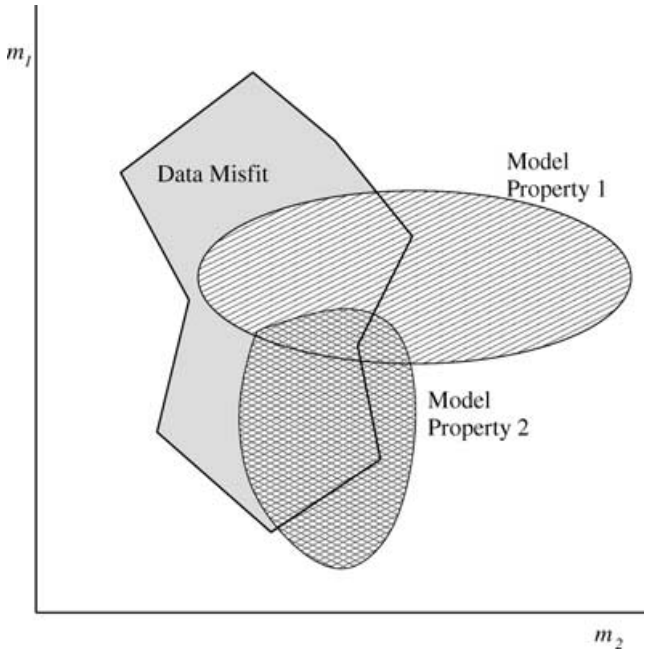

Figure 4. The intersection of the consistency region for data misfit, determined using $w(\Phi)$, with the regions defined by constraints on model properties $\{\psi\}$ defining a consistency region for suitable models.

The separation of the data fit and regularization terms gives considerable flexibility in the specification of the outcomes of the inversion. We can readily place *a priori* specifications on the model properties, whilst the consistency region for data fit retains the benefits discussed above.

The consistency zone is then the intersection of the regions defined by the different criteria, as indicated schematically in Fig. 4. This approach provides an unfettered framework for *a posteriori* analysis. Additional model characteristics can be investigated, such as the use of separate constraints on model norm, gradients and curvature as opposed to a composite Sobolev norm.

In some circumstances, data depends on multiple sets of physical parameters, as in the relation of the frequencies of the normal modes of the Earth to the *P*- and *S*-wave-speed distribution and the density (see, e.g. Dahlen & Tromp 1998). The properties of the subsets of model parameters can then be used independently and collectively with the data fit in the characterization of suitable models.

In the situation where multiple classes of data are generated from a single model description, as for the traveltimes of seismic phases for a 1-D global Earth model (e.g. Kennett *et al.* 1995), multiple data fit zones for the different data classes can be defined in parameter space. The inter-relation of these various zones controlled by different aspects of the data with the model constraints provides a measure of the resolution achieved in the model.

ACKNOWLEDGMENTS

Part of this work was undertaken during a visit to NORSAR supported by the Norwegian Research Council. Discussions with Johannes Schweitzer on the assessment of error in non-linear location schemes helped to stimulate these results.

REFERENCES

Billings, S., Sambridge, M.S. & Kennett, B.L.N., 1994. Errors in hypocentre location: picking, model and magnitude dependence, *Bull. seism. Soc. Am.,* **84,** 1978–1990.

Dahlen, F.A. & Tromp, J., 1998. Theoretical Global Seismology, Princeton University Press, Princeton, NJ.

Gallagher, K., Sambridge, M.S. & Drijkoningen, G.G., 1991. Genetic algorithms: an evolution on Monte Carlo methods in strongly non-linear geophysical optimization problems, *Geophys. Res. Lett.,* **18,** 2177–2180.

Kennett, B.L.N., 1978. Some aspects of non-linearity in inversion, *Geophys. J. R. astr. Soc.,* **55,** 373–391.

Kennett, B.L.N. & Sambridge, M.S., 1992. Earthquake location: Genetic algorithms for teleseisms., *Phys. Earth. planet. Int.,* **75,** 103–110.

Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.,* **122,** 108–124.

Lomax, A. & Snieder, R., 1994. Finding sets of acceptable solutions with a genetic algorithm with application to surface wave group dispersion in Europe, *Geophys. Res. Lett.,* **21,** 2617–2620.

Lomax, A. & Snieder, R., 1995. Identifying sets of acceptable solutions to nonlinear geophysical inverse problems which have complicated misfit functions., *Nonlinear. Proc. Geophys.,* **2,** 222–227.

Sambridge, M.S., 1999a. Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space, *Geophys. J. Int.,* **17,** 479–494.

Sambridge, M.S., 1999b. Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble, *Geophys. J. Int.,* **138,** 727–746.

Sambridge, M.S., 2001. Finding acceptable models in nonlinear inverse problems using a neighbourhood algorithm, *Inverse Problems,* **17,** 387–403.

Sambridge, M.S. & Kennett, B.L.N., 2001. Seismic event location: nonlinear inversion using a neighbourhood algorithm, *Pure appl. Geophys.,* **158,** 241–257.

Sambridge, M.S. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), 1009, doi:10.1029/2000RG000089.

Schrödinger, E., 1952. Statistical Thermodynamics, Cambridge University Press, Cambridge.

Schweitzer, J. & Kennett, B.L.N., 2002. Comparison of location procedures—the Kara Sea event of 16 August 1997, *NORSAR Sci. Report 1-2002*, 97–114.

Schweitzer, J. & Kennett, B.L.N., 2004. The Kara Sea event of 16 August 1997—Comparison of location procedures, *Bull. seism. Soc. Am.*, submitted.

Sen, M.K. & Stoffa, P.L., 1991. Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics,* **56,** 1624–1638.

Shibutani, T., Sambridge, M. & Kennett, B., 1996. Genetic algorithm inversion for receiver functions with applications to crust and uppermost mantle structure beneath eastern Australia, *Geophys. Res. Lett.,* **23,** 1829–1832.

Resovsky, J.S. & Trampert, J., 2002. Reliable mantle density error bars: an application of the neighbourhood algorithm to normal-mode and surface wave data, *Geophys. J. Int.,* **150,** 665–672.