B. H. Jacobsen.

"A priori information: Why, which and how?"

This note is a short review of some problems and solutions from inverse theory and from electromagnetics relevant to the interpretation situations discussed at the colloquium in Neustadt 1982.
After a general introduction we shall discuss the questions of the title and finally indicate a simple way of estimating equivalence through the 'ideal body' concept.

The general interpretation situation may be given the following formulation (Jacobsen, 1982):
We discribe the earth by some simplified model, which may be parameterized by M real numbers, $x_1, .. , x_M$.
These parameters may be considered as a vector $\underline{x} = (x_1,..,x_M)$.
During some measurement sequence we collect the new field data, which may be represented as N real coordinates, $\underline{\hat{y}} = (\hat{y}_1,..,\hat{y}_N)$.
From mathematical physics a theoretical noisefree datavector

$$\underline{y}(\underline{x}) = (\underline{y}_1(\underline{x}),..,\underline{y}_N(\underline{x}))$$

may be computed corresponding to any parameter vector $\underline{x}$.
The inverse problem is then to investigate, which parameter vectors are compatible with our information about the earth.
Part of our information is the new field data, $\underline{\hat{y}}$.
The rest of our information about the earth is 'known beforehand' and is therefore named 'a priori information'.

'Pure' inverse theory deals with the case where 'new field data' are our only information.
If the statistical properties of the measurement error is known then the set of 'data compatible' earth models may be discribed by the concept of confidence regions.
If $\underline{y}(\underline{x})$ is a linear function, and the measurement error is Gaussian distributed, then classical linear regression analysis is appropriate.
If further the errors are independent with equal distributions then data compatibility is simply related to the sum of squared residuals, $Q(\underline{x}) = \|\underline{\hat{y}}-\underline{y}(\underline{x})\|^2$.
The singular value decomposition of $\underline{y}(\underline{x})$ may in this latter case help us to understand the set of solutions in which we have confidence (Pedersen, 1979).
If however errors are not Gaussian distributed or $\underline{y}(\underline{x})$ is nonlinear then the actual computation of confidence limits is still an area of research in applied statistics.

Stabilized linearised iteration for finding models with minimum residuals has been in common use since the days of Newton (Marquardt, 1963), and some conditions have been found for the validity of the well known linear confidence analysis of the resulting least squares model (Bates & Watts, 1980).
For linear problems one should not forget linear programming as another powerful tool for finding models which are compatible with the data (Cuer & Bayer, 1980).
From a statistical point of view however, this algorithm is less natural.

We may summarize that search for a model, which describes given data with given error statistics, and analysis of the limits of variability of the model are well studied though not completely solved issues.

We learn from experience, that the set of data compatible earth models often contain models, which from a geologically point of view are very different. This problem called 'equivalence' was illuminated for magnetotelluric data and layered earth by G. Fischer at the colloquium, (cf. Fischer, preprint).
Let us review this example to illustrate the general introduction.

$\underline{\tilde{y}}$ consists of the measured $\tilde{\rho}_a(f_i)$ and $\tilde{\phi}(f_i)$ at the frequencies f_i.
The measurement errors are (hopefully) Gaussian and independent with standard deviations $\Delta\rho_a(f_i)$ and $\Delta\phi(f_i)$.
A 'homogeneous n-layer model' is parameterized in a 2n-1 dimensional vectorspace: $\underline{x} = (\rho_1, h_1, \,..\, , \rho_{n-1}, h_{n-1}, \rho_n)$.
The theoretical response, $\underline{y}(\underline{x})$, is computed from the familiar layer by layer recursions. This vector function is strongly nonlinear.
The standard measure of misfit is

$$Q(\underline{x}) = \sum_1^N (\tilde{y}_i - y_i(\underline{x}))^2/\Delta y_i^2$$

(G. Fischer (preprint) uses ε^2 for a similar measure of misfit).
If the true earth is really described by n homogeneous layers, $\underline{x}_{true}$, then $Q(\underline{x}_{true})$ has the expectation value N:

$$E(Q(\underline{x}_{true})) = N.$$

The best fitting model, $\underline{x}_{best\ fit}$, by definition makes $Q(\underline{x}_{best\ fit})$ a minimum. If the theoretical response, $\underline{y}(\underline{x})$, is a linear function, then regression analysis tells us, that

$$E(Q(\underline{x}_{best\ fit})) = N-M .$$

and therefore

$$E(Q(\underline{x}_{true}) - Q(\underline{x}_{best\ fit})) = M$$

where N is the number of data and M is the number of parameters. Models satisfying

$$Q(\underline{x}) - Q(\underline{x}_{best\ fit}) < K \cdot M$$

make up a confidence region, where K is close to 1 and depending on the decired confidence level and N and M.
These results are approximately true for nonlinear $\underline{y}(\underline{x})$, provided $\underline{y}(\underline{x})$ is not too nonlinear (Bates & Watts, 1980).
For a typical MT-sounding covering two decades of frequencies we may have 40 real data. Such data may almost always be modelled by 4 layers, i.e. 7 parameters. Putting K=1 this gives

$$Q(\underline{x}) < Q(\underline{x}_{best\ fit}) + M \approx 1.21 \cdot Q(\underline{x}_{best\ fit})$$

or using $\varepsilon = \sqrt{Q}$

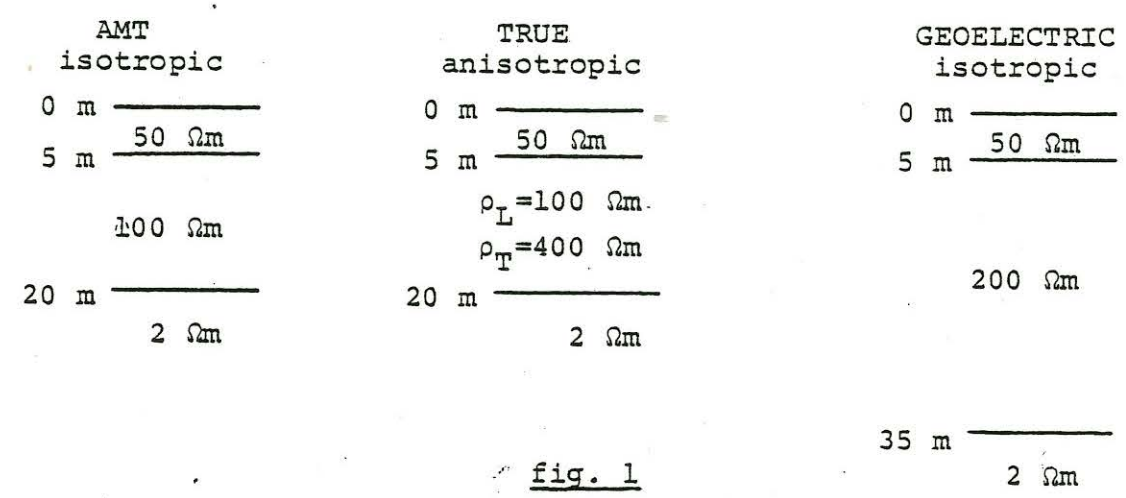
$$\varepsilon < 1.1 \cdot \varepsilon_{best\ fit}$$

as also stated on empirical ground by G. Fischer (ibid.).

We will now discuss the 3 questions proposed in the title.
We asked, why a priori information should be included when interpreting new field data.
The reason is, that a priori information improve the choise of model space and reduce equivalence.
Thus it is well known, that only the total conductivity, $\sigma \cdot h$, of a conductive bed is determined by MT data. If however the conductive layer corresponds to a certain geological formation with known specific conductivity, then the thickness, h, will also be well determined. This illustrates, how a combination of new field data and one 'key piece' of a priori information may elliminate equivalence.

In general however, it is not a trivial question which a priori information to use. The selection must be founded on a sound intuition for both the geology of the region and the physics of the measuring method.
We will give some examples relevant to MT interpretation.
For the estimation of the impedance tensor, physics offers a priori constraints. Weidelt (1972) developed several such constraints, primarily connected with the causality and minimum-phase property of the earths impulse response, and according to Rokityansky (1982,p89f) similar properties are obeyed by the tensor over any 3-D earth.
Such constraints may help to decide the data quality, for instance by discovering bias effects.
For a stratified region a 1-D model is appropriate. In this case

layer conductivities, depths and thicknesses may be known from seismics or drillings.
For regions with more complex geology the location of faults and the shape of near surface structures (for instance oceans) will constitute important a priori information.
This kind of information may at first sight seem more harmful than helpful, as it poses more questions than it answers.
But to ignore such information in order to stay with a 1-D model would of course be 'to wipe the dust under the carpet'. Instead we should view a priori information as an aid in the choice of model space.
A simple example will show, that even in the 1-D case this choice of model space is not merely a matter of selecting the number of homogeneous layers. Assume, that the earth contains an anisotropic layer as illustrated in fig. 1. This anisotropic layer could be sands above the ground water level resting on water saturated clay. Audiomagnetotelluric data will be perfectly interpretable by the left 3-layer isotropic model, while geoelectric souding data will fit the right 3-layer isotropic model.
To fit both datasets with the same homgeneous n-layer model we will need 4 or more layers and their parameters will be very ill determined.



<u>fig. 1</u>

If however we know, that unsaturated sands rest on saturated clays of great thickness in the region, then this information will lead to the formulation of a 3-layer model allowing anisotropy in the second layer. The parameters of this model will be well determined and geologically meaningful.

These examples of a priori information have at the same time shown one side of <u>how</u> to include a priori information in the interpretation: A priori information form the basis of the more or less philosophical choice of model space.

Now assume, that a model space has been selected, and parameterized by some vector $\underline{x}$. Then some algorithm is required to investigate the set of data compatible models in this space.
This algorithm, which is often implemented on a computer, will be named 'the interpretation program'. We assume, that this program is capable of selecting models which satisfy our data:

$$y_i(\underline{x}) = \hat{y}_i \pm \Delta y_i \qquad i = 1 \,..\, N$$

subject to some misfit criterion.

<u>The second way</u> of including a priori information is now to add it to this interpretation program as 'a priori data' about some scalar earth property.
In a 1-D MT model a piece of a priori information may sound
"The resistivity of Postzechstein is probably around 2 Ohmm,
" may be 4 Ohmm, may be 1 Ohmm."
If Postzechstein is the first layer in the model, then this information may be translated like this:
"$\ln(\rho_1) = \ln(2 \text{ Ohmm}) \pm \ln(2)$ ".
This relation is now formulated as a measurement and may be included in the interpretation program on a pragmatic basis.
Any scalar measure, $F(\underline{x})$, on which we have information of a 'most probable' value, c, with an 'uncertainty', Δc, may be entered as an a priori datum:

$$F(\underline{x}) = c \pm \Delta c \; .$$

Notice, that we are not forced to assume the property known exactly and 'frozen'. Instead the a priori datum acts as a 'rubber band' with a 'length' Δc, tapering the variability of $\underline{x}$. The decreased variability of the model parameters will then indeed reflect the information content of the extra datum. It is therefore essential that c is reasonable and that Δc has not been chosen too optimistically.
Some times a more natural formulation of our a priori information is in the form of inequality constraints:

$$F(\underline{x}) < c \; .$$

If $F(\underline{x})$ is linear and $\underline{y}(\underline{x})$ is linear then this kind of constraints are elegantly treated by linear programming.
Weidelt (1972) suggested to use this kind of constraints to impove estimates of the transfer function in magnetotellurics, and Parker (1980) has shown how to treat the problem of 1-D MT modelling in

the framework of linear programming using such a priori constraints.

The word 'a priori data' is presented by Jackson(1979), but the concepts have been used for a long time in statistics under the label 'Bays statistics'.

We have now indicated why a priori information should be included in any interpretation, given examples of such relevant information and devised methods to implement a priori information in an interpretation program.
But no matter how careful we are, we will probably still be left with a large set of earth structures compatible with field data as well as all available a priori information.
It is essential to notice, that this set of solutions is often difficult to 'understand'. Even simple multidimensional hyper ellipsoids (for linear problems and Gaussian noise) demand heavy abstraction.
However, in certain interpretation situation we often have special interest in some particular property of the earth structure.
Assume, that this property may be expressed as a scalar quantity.
It could be "depth to the crystaline basement", "integrated conductivity of the graben structure" or "lateral coordinate of the deep-seated conductivity structure". It would then be interesting to know the extreme values of such scalar measures compatible with the field data and all available a priori information.
A simple trick is now to include the interesting scalar as an a priori datum with small variance in the interpretation program. Fixing this scalar at still larger (smaller) values until compatibility is lost will give upper (lower) bounds to the interesting earth property (Jacobsen,1982,Chap.8).
Notice, that these intervals of variability are not based on any linear approximation close to some optimum model.

An earth model for which a scalar measure is extreme is called an 'ideal body'.
Inspection of the ideal bodies corresponding to extreme values of the scalar measures may give valuable insight into the physics of the measurements and the nature of the entered a priori information.
This technique may also be used to analyze what the resolving power of a planned measurement sequence would be on the expected earth structure.

The name 'ideal bodies' is presented by Parker (1974). Jackson (1976,1979) name a similar approach the 'most squares' method, but the concept of 'extreme solutions' is old.

The problems and solutions reviewed in this note are obvious and trivial from a theoretical point of view.
The actual translation of a priori information to a priori data may pose problems from time to time, partly because the translation forces us to consider the reliability of a priori information. But the applicability of the concept was impressively demonstrated by G. Fischer at the colloquium for the case of MT data and an n-layer earth.
For the much more complex problems of 2-D and 3-D modelling using MT profiles and grids the described techniques should be even more helpful in selecting the modelspace, reducing equivalence and estimating the variability of key properties.

Acknowledgements
The author is greatly indebted to F. H. Andersen, J. Nissen and T. Rasmussen (University of Aarhus) for many valuable discussions on the application of inverse theory to electromagnetic interpretation.

References

Bates, D. M., Watts, D. G., 1980, Relative curvature measures of nonlinearity, J. R. Statist. Soc. B 42, pp. 1-25.
Cuer, M., Bayer, R., 1980, Fortran routines for linear inverse problems, Geophys. vol 45, 1706-1719.
Fischer, G., Le Quang, B. V., preprint, Parameter trade-off in one-dimensional magnetotelluric modelling, Observatoire Cantonal, CH-2000 Neuchatel.
Jacobsen, B. H., 1982, Inverse theory - foundations, techniques and applications, (in danish), GeoKompendier 19, University of Aarhus.
Jackson, D. D., 1976, Most squares inversion, J. Geophys. Res. 81, 1027-1030.
Jackson, D. D., 1979, The use of a priori data to resolve non-uniqueness in linear inversion, Geophys. J. R. Astron. Soc. 57, 137-157.
Marquardt, D.W., 1963, An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Indust. Appl. Math. 11, 431-441.

Parker, R. L., 1974, Best bounds on density and depth from gravity data, Geophysics 39, 644-649.

Parker, R. L., 1980, The inverse problem of electromagnetic induction: Existence and construction of solutions based on incomplete data, J. Geophys. Res. 85, 4421-4428.

Pedersen, L. B., 1979, Constrained inversion of potential field data, Geophys. Prosp. 27, 726-748.

Rokityansky, I. I., 1982, Geomagnetic investigation of the earth's crust and mantle, Springer-Verlag, Berlin Heidelberg New York.

Weidelt, P., 1972, The inverse problem of geomagnetic induction, Z. Geophys. 38, 257-289.