Error Analysis for Parameter Estimates from Constrained Inversion

I. G. Enting

DIVISION OF ATMOSPHERIC RESEARCH TECHNICAL PAPER No. 2
COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANIZATION, AUSTRALIA 1983

Error Analysis for Parameter Estimates from Constrained Inversion

By I. G. Enting

Division of Atmospheric Research Technical Paper No. 2

Commonwealth Scientific and Industrial
Research Organization, Australia
1983

National Library of Australia Cataloguing-in-Publication Entry

Enting, I. G.

Error analysis for parameter estimates from constrained inversion.

ISBN 0 643 03523 0.

1. Geochemistry–Mathematical models. I. Commonwealth Scientific and Industrial Research Organization (Australia). Division of Atmospheric Research. II. Title. (Series: CSIRO Division of Atmospheric Research technical paper; no. 2).

551.9'0724

© CSIRO 1983. Printed by CSIRO, Melbourne

83.215-1060

CSIRO Aust. Div. Atmos. Res. Techn. Pap. No. 2, 1-18 (1983)

Error Analysis for Parameter Estimates from Constrained Inversion

I.G. Enting

CSIRO Division of Atmospheric Research

Abstract

Constrained inversion techniques can provide a useful approach to the problem of estimating the parameters of geochemical models. These techniques can be approached in various ways which make differing statistical assumptions and which lead to different expressions for the parameter uncertainties. For models based on well-defined geophysical concepts, Bayesian analysis provides an appropriate way of using additional geophysical information in the model calibration process.

1. Introduction

The difficulties of calibrating and validating geochemical models are one of the major limitations to the use of such models in making plausible predictions of geochemical changes. Frenkiel and Goodall (1978) have commented that 'systematic methods of approaching model estimation and validation have not received much attention in the simulation literature'. This is particularly true when considering attempts to evaluate the uncertainties in the model calibrations. This report is a compilation of a number of different statistical approaches that are relevant in model calibrations. The main emphasis is on statistical techniques for evaluating estimates from constrained inversion calculations, in particular for the case in which a constrained inversion formulation is obtained on the basis of a Bayesian statistical analysis. Constrained inversion formalisms have been used in a number of geophysical contexts, such as remote sensing of atmospheric properties, interpretation of seismic data and ocean transport studies, (see Twomey, 1977 for a review of techniques and some examples of applications). A similar technique called 'ridge regression' has been described by Hoerl and Kennard (1970a,b). There have been a few studies in which such techniques have been obtained on the basis of a Bayesian statistical approach, as for example in the work of Whitney, (1977), Kuczera, (1982) and Bretherton et al. (1983). More recently inversion techniques have been applied to geochemical studies, as in the work of Bolin et al. (1983) in ocean geochemical transport, Fraser et al. (1983) and Enting (1983) in global carbon cycle modelling.

The constrained inversion formalisms are introduced in problems that are under-determined or at least ill-determined. The ways in which these under-determined aspects arise have been described in connection with the various applications listed above, and also to some extent by Twomey, (1977). Section 2 below describes the general mathematical framework within which the estimation problem is expressed. Section 3 describes a number of different statistical interpretations that can be applied to the results of the estimation procedure. Section 4 applies some of the most useful of these approaches to the problem of determining the sensitivity of model predictions to the uncertainties in the model calibration. The notation is defined in an appendix.

2. Parameter Estimation

In abstract terms, the calibration problem for geochemical models can be expressed as follows.

The model is defined in terms of a set of K parameters, x_k, $k = 1,K$ The values of these parameters are not known precisely and have to be estimated by matching the predictions of the geochemical model to measurements of observable geochemical quantities. The various measured quantities are denoted m_j, $j = 1, J$ and the model predictions for these quantities are denoted $y_j(\underset{\sim}{x})$, indicating that the y_j are functions of the whole set of x_k.

The problem of determining the x_k from the m_j is often under-determined or at least exceedingly ill-determined. Constrained inversion formalisms (Twomey, 1977) are designed to obtain estimates by supplying additional assumptions. Examples of the additional requirements that are assumed are

(i) minimal solutions i.e. solutions that minimise the x_k so as not to have the values any larger than what is required by the m_j.

(ii) smoothest solutions. This is applicable when the x_j described the expansion of a function in terms of a set of basis functions. The aim is not to include in the function any 'structure' that is not required by the data.

(iii) smallest deviation from some initial solution. The idea is to take an initial solution and to make only those changes that are necessitated by the m_j.

(iv) minimum magnification of errors - following Backus and Gilbert (1968, 1970).

In each case the fit that is selected is a compromise between fitting the m_j and satisfying the additional assumptions. A simple mathematical form that is commonly used is to estimate the x_k by minimising θ where

$$\theta = \sum_{j=1}^{J} (y_j(\underset{\sim}{x}) - m_j)^2 + \gamma \sum_{k=1}^{K} (x_k - q_k)^2 \qquad (2.1)$$

where the set of q_k, k = 1, K represent the initial solution described in case (iii) above. Case (i) corresponds to all the q_k being zero and case (ii) (i.e. minimum derivatives rather than minimum functions) can be transformed into case (i) by changing the definitions of the parameters. In inversion formalisms the parameter γ is regarded as adjustable and is varied to produce stable estimates of the x_i. Since γ is an arbitrary quantity, a desirable class of solution is that in which the solution is insensitive to the value of γ over some reasonably wide range. If the solution changes abruptly as γ changes it can indicate incompatibility between the two solutions defined by the m_j and q_k respectively in that minimising θ will lead to values near one or other of the limiting solutions except in a narrow unstable region where the solution is unlike either limit. A small variation with γ can indicate the existence of a broad range of compromise solutions all of which fit both sets of information to a similar extent (see Twomey, 1977).

In order to assess the uncertainties in the estimation procedure, some knowledge of the uncertainties in the measurements is needed.

Initially we assume

$$m_i = y_i(\underset{\sim}{x}_{true}) + \varepsilon_j \qquad (2.2)$$

where ε_j represents the known uncertainty in the observations. In the analysis that follows this is taken as being normally distributed although much of the analysis (especially the desirability of using a least square fit) does not depend on this assumption.

A more common situation is to have

$$m_i = y_i(\underset{\sim}{x}_{true}) + \eta_i + \varepsilon_i \qquad (2.3)$$

where the η_i represent errors due to the mismatch between a geographically low-resolution lumped model value of y_i and a value of m_i based on the mean of a small number of samples. In some cases, such as the catchment models considered by Kuczera (1982), the η_i can be reasonably regarded as random effects and absorbed into the ε_i. In geochemical models such as the carbon cycle model described by Enting and Pearman (1982, 1983) the η_i may represent non-random

biases due to covariance effects in non-linear model processes. As pointed out by Enting and Pearman, (1982, §2) this type of effect can sometimes be described by the use of 'effective parameters' that differ from the simple mean values of parameters that show spatial and temporal variation. In any case the initial analysis proceeds on the assumption that the errors are random as indicated by (2.1). When spatial or temporal variations are large, the cautious approach is to follow the suggestion of Bolin et al. (1981) and use the range of variation as a measure of the variation of the m_i.

Once the variances have been estimated, the observations are scaled so that the variance of each of the ε_i is equal to 1. In terms of the least squares fit, this procedure is equivalent to using a weighted least-squares procedure with the weights chosen so as to minimise the variance of the parameter estimates (Deutsch, 1965).

It is also necessary to define some choice of scale for the x_k and q_k in the case $\gamma \neq 0$. In the applications considered here, it is assumed that there is a realistic estimate q_k, for each x_k and also that there is a knowledge of the uncertainties. This approach is particularly applicable in models such as the carbon cycle model described by Enting and Pearman (1982, 1983) which was specifically designed so that each parameter represented a well-defined geophysical quantity.

The Bayesian analysis (Box and Tiao, 1973) takes this approach one step further by treating the parameters as random variables so that the q_k and the variance are obtained from a prior distribution that represents what is known about the parameters before the model is fitted.

Because of the conceptual differences between the Bayesian and the conventional (sampling theory) approaches to estimation, it is important to specify the meanings of the various quantities in each context.

Each approach produces a set of parameter estimates, denoted $\hat{\underset{\sim}{x}}$, which are obtained by minimising the function θ defined in equation (2.1).

Sampling theory treats $\hat{\underset{\sim}{x}}$ as being a vector of random variables since the components are functions of the random variables ε_j. The quantification of uncertainties consists of specifying the way in which the vector $\hat{\underset{\sim}{x}}$ is distributed and this specification is most useful when we specify the distribution of $\hat{\underset{\sim}{x}}$ around the 'true' parameters $\underset{\sim}{x}_{true}$, by specifying the distribution of the difference $\underset{\sim}{d}$ where

$$\underset{\sim}{d} = \hat{\underset{\sim}{x}} - \underset{\sim}{x}_{true} \qquad (2.4)$$

Confidence intervals obtained from the distribution of $\underset{\sim}{d}$, must, strictly speaking, be interpreted as giving probabilities that particular ranges of $\hat{\underset{\sim}{x}}$ values would be obtained, assuming that $\underset{\sim}{x}_{true}$ was actually equal to the computed $\hat{\underset{\sim}{x}}$.

In contrast, the Bayesian analysis takes $\underset{\sim}{x}_{true}$ as being unknown and describes its possible values in terms of probability distributions that

describe the amount of information that we have about $\underset{\sim}{x}_{true}$ both before and after using the m_i. Thus $\underset{\sim}{x}_{true}$ is treated as a vector of random variables but $\hat{\underset{\sim}{x}}$ is not a random vector because the analysis is concerned with the information obtained from a single set of measurements: the m_j and the ε_j are fixed by making the observations.

The uncertainties in the Bayesian analysis can again be described by the vector $\underset{\sim}{d}$ defined by (2.4) but in this case the distribution of $\underset{\sim}{d}$ measures the distribution of the unknown $\underset{\sim}{x}_{true}$ about the estimate $\hat{\underset{\sim}{x}}$.

The analysis in the following section is based on linearising the model, using the sensitivity matrix S where

$$S_{jk} = \frac{\partial y_j}{\partial x_k} \tag{2.5}$$

The sampling theory analysis linearises about the unknown $\underset{\sim}{x}_{true}$ giving

$$y_j(\hat{\underset{\sim}{x}}) = y_j(\underset{\sim}{x}_{true}) + \sum_k S_{jk}\, d_k \tag{2.6}$$

as in equation (3.1).

The Bayesian analysis linearises about $\hat{\underset{\sim}{x}}$ giving

$$y_j(\underset{\sim}{x}_{true}) = y_j(\hat{\underset{\sim}{x}}) - \sum_k S_{jk}\, d_k \tag{2.7}$$

as in equation (3.23).

3. Quantifying the uncertainties in parameter estimates

As indicated in the previous section, there are a number of different points of view from which the constrained inversion formalism can be approached. Although several of the approaches that have been described lead to the same estimates for the parameter values, the different statistical assumptions lead to different expressions for the uncertainties in these estimates and to different statistical tests required to test the validity of the estimation (see Deutsch, 1965).

(a) The use of observations only, assuming known variance

The use of only observations implies $\gamma = 0$ and so

$$\theta = \sum_{j=1}^{J} (y_j(\hat{\underset{\sim}{x}}_j) - m_j)^2$$

$$= \sum_j (\sum_k S_{jk}\, d_k - \varepsilon_j)^2 \quad . \tag{3.1}$$

The minimum of θ is given by d_k which is the solution of

$$\sum_{jk} S_{jk}\, S_{jk'}\, d_k = \sum_j S_{jk} \varepsilon_j \quad .$$

$$= \sum_k V_{kk'}\, d_k = \sum_j S_{jk}\, \varepsilon_j \tag{3.2}$$

whence $$d_k = \sum_{jk'} C_{kk'} S_{jk'} \varepsilon_j \qquad (3.3)$$

where the matrices C and V are defined by

$$V_{kk'} = \sum_{j} S_{jk} S_{jk'} \qquad (3.4)$$

and $$C = V^{-1} \quad . \qquad (3.5)$$

Using this solution, θ can be re-expressed as

$$\theta = \theta_{min} + Q \qquad (3.6)$$

where $$\theta_{min} = \Sigma (y_j(\hat{\underset{\sim}{x}}) - m_j)^2 \qquad (3.7)$$

and $$Q = \sum_{jkk'} d_k S_{jk} S_{jk'} d_{k'} \quad . \qquad (3.8)$$

If var $(\varepsilon_j) = \sigma^2$ for all j

then θ_{min} is distributed as $\sigma^2 \chi_\nu^2$

where

$$\nu = J - K \quad . \qquad (3.9)$$

This means that a χ^2 test (with ν degrees of freedom) can be applied to the value of θ_{min} to see if the model gives a statistically significant fit to the data.

Note that from 3.2,

$$E\{d_k\} = 0 \qquad \text{i.e. } E\{\underset{\sim}{x}\} = \underset{\sim}{x}_{true} \qquad (3.10)$$

and

$$E\{(x_k - \hat{x}_k)(x_m - \hat{x}_m)\}$$

$$= \sum_{k'm'jj'} E\{C_{kk'} S_{jk'} \varepsilon_j C_{mm'} S_{j'm'} \varepsilon_{j'}\}$$

$$= \sum_{m'k'jj} C_{kk'} C_{mm'} S_{jk'} S_{j'm'} \sigma^2 \delta_{jj'}$$

$$= C_{km} \sigma^2 \quad . \qquad (3.11)$$

Thus C, the inverse of $V = S^T S$, is the covariance matrix for the parameter estimates if $\sigma^2 = 1$, i.e. if the scaling assumed in section 2 is correct. If the errors are normally distributed, the d_k will have a multivariate normal distribution.

Confidence intervals on the K parameter distribution of d_k are defined by

$$Q(\underset{\sim}{d}) = \sum_{kk'} d_k V_{kk'} d_{k'} \leq \chi^2 (K, \alpha) \qquad (3.12)$$

with probability $1 - \alpha$.

(b) Only observations used, variance assumed unknown (see Deutsch, 1965)

This case corresponds to the most common situation in regression analysis. It is assumed that the errors ε_j have (after scaling) a common variance σ^2 but that this variance is unknown. The value of θ_{min} can be used to estimate σ^2 or, as below, used directly in the specification of the distribution of parameter estimates. Thus it is not possible to use the value of θ_{min} obtained from a single fit to assess the validity of the estimation procedure. What can be done is to take the values of θ_{min} obtained using different numbers of parameters fitted and apply the 'F-test' to see whether the improved fit that is obtained using the larger parameter set can be regarded as a statistically significant improvement. (Note that the validity of the F-test depends strongly on the normality (and particularly on symmetry) of the error distribution).

Both the joint distribution of the parameter estimates and the various marginal distributions are given by (multivariate) t distributions as defined by Box and Tiao (1973), equations 2.7.20, 2.7.22. Ellipsoidal confidence regions with probability $1-\alpha$ are defined by

$$\frac{(J-K)Q(\underset{\sim}{d})}{K\theta_{min}} = F(K,J-K,\alpha) \qquad (3.13)$$

where $F(K,J-K,\alpha)$ is the upper 100α percentage point of an F distribution with (K,J-K) degrees of freedom.

This form of parameter fitting is intuitively appealing because of its correspondence to conventional regression analysis. In the geochemical modelling context, it suffers from the defect that it throws away the possibility of using θ_{min} to check the model validity and that it relies on the assumption that the defects in the fit can be treated as though they were due to random errors in the data rather than oversimplification in the model.

(c) Constrained inversion with uncertainties determined only by observations

The conventional approach to constrained inversion (Jackson, 1972, Twomey, 1977) treats the parameter determination problem in terms of vector spaces and defines two subspaces within which the parameters are adequately determined and undetermined respectively. The error analysis concerns the uncertainties within the 'adequately-determined' space. The under determined subspace is dealt with by taking a single representative vector. This is chosen on the basis of additional assumptions which are not generally analysed in statistical terms. Since only part of the uncertainty is analysed, this treatment is not particularly suitable for geochemical modelling studies and so only a brief summary is given here.

The assumption is that the additional constraints represented by the q_k represent the 'true' parameter values, so that

$$\theta = \sum_j (\sum_k S_{jk} d_k - \varepsilon_j)^2 + \gamma \sum_k d_k^2 \quad . \qquad (3.14)$$

This assumption is strictly quite unrealistic but it can be regarded as an approximation to the situation in which the q represents a better approximation to the $\underset{\sim}{x}_{true}$ than can be obtained from fitting only the observations as in section 3a, above.

Minimising θ with respect to the d_k gives

$$\sum_{ij} (S_{ij}S_{ij'} + \gamma I_{jj'})d_j = \sum_j S_{ij}\varepsilon_j$$

or
$$\sum_k V_{kk'}(\gamma)d_j = \sum_j S_{ij}\varepsilon_i \qquad (3.15)$$

or
$$d_j = \sum_{ij'} C_{jj'}(\gamma)\, S_{ij}\varepsilon_i \qquad (3.16)$$

where $V(\gamma)$ and $C(\gamma)$ its inverse, are the generalisations of the matrices V,C in the $\gamma = 0$ case above. The matrix $C(\gamma)$ can be regarded as a constrained inverse of $V(0)$. The constraint acts primarily on the subspace of parameter space spanned by those eigenvectors of $V(0)$ that have eigenvalues $\lambda_\mu \leq \gamma$. Since the parameter variability in a particular eigenvector direction is proportional to λ_μ^{-1}, arbitrarily replacing all eigenvalues λ_μ by $\lambda_\mu + \gamma$, i.e. using $C(\gamma)$ rather than $C(0)$, artificially reduces this variability. This approach does not give quite the sharp, distinction between the 'adequately determined' and 'undetermined' subspaces that is implied by the introductory remarks of this section. The sharp distinction is associated with the procedure described by Jackson (1972, section 5) which replaces λ_μ^{-1} by 0 if $\lambda_\mu < t$ for some threshold t, rather than replace λ_μ^{-1} by $1/(\lambda_\mu + \gamma)$ for all μ.

Using equation (3.14) above gives

$$E\{d_k d_{k'}\} = E\{\sum_{\substack{mm'\\ii'}} C_{km}(\gamma)\, C_{m'm'}(\gamma)\, S_{im}\, S_{i'm'}\, \varepsilon_i \varepsilon_{i'}\}$$

$$= \sigma^2 \sum_{mm'} C_{km}(\gamma)\, C_{m'm'}(\gamma)\, V_{mm'}(0) \quad . \qquad (3.17)$$

This will tend to decrease with increasing γ, corresponding to reduced uncertainty associated with reduced resolution in the terminology of Jackson (1972).

As mentioned above, this formalism is not particularly appropriate for geochemical modelling studies as it only considers a restricted class of uncertainties. In particular it takes q as being, for the purposes of error analysis, unbiased estimates of the true x. This aspect is explicitly addressed by Hoerl and Kennard (1970a,b) who described an equivalent formalism (ridge regression) for the purpose of reducing instabilities at the expense of a possible bias.

(d) Direct Bayesian inference (see Box and Tiao, 1973)

The direct form of Bayesian inference takes the ranges used in the scaling as giving a literal interpretation to the probabilities. Once the variables are scaled as defined in section 2 the literal interpretation of the probabilities requires that $\gamma = 1$. This is because the Bayesian formalism begins by treating both the observations m_i and the parameter values x_i as random variables and uses the distribution defined in the appendix.

$$P(\underset{\sim}{m},\underset{\sim}{x}) = P(\underset{\sim}{m}|\underset{\sim}{x})P(\underset{\sim}{x}) = P(\underset{\sim}{x}|\underset{\sim}{m})\, P(\underset{\sim}{m}) \qquad (3.18)$$

whence, for a fixed set of measured values $\underset{\sim}{m}$,

$$P(\underset{\sim}{x}|\underset{\sim}{m}) \propto P(\underset{\sim}{m}|\underset{\sim}{x})\ P(\underset{\sim}{x}) \qquad (3.19)$$

The assumptions are that

$$P(\underset{\sim}{x}) \propto \exp\ (-\tfrac{1}{2} \sum_k (x_k - q_k)^2) \qquad (3.20)$$

and

$$P(\underset{\sim}{m}|\underset{\sim}{x}) \propto \exp\ (-\tfrac{1}{2} \sum_j (y_j(\underset{\sim}{x}) - m_j)^2) \quad . \qquad (3.21)$$

The mode (most probable x) of the posterior density function $P(\underset{\sim}{x}|\underset{\sim}{m})$ is thus given by the $\hat{\underset{\sim}{x}}$ that minimises

$$\theta = \sum_k (x_k - q_k)^2 + \sum_j (y_j(x) - m_j)^2 \quad . \qquad (3.22)$$

Linearising about the minimum gives

$$\theta = \theta_{min} + \tfrac{1}{2} \sum_{kk'} (x_k - \hat{x}_k) \frac{\partial^2 \theta}{\partial x_k \partial x_{k'}} (x_{k'} - \hat{x}_{k'}) \qquad (3.23)$$

so that in this approximation, $P(\underset{\sim}{x}|\underset{\sim}{m})$ has a multivariate normal distribution with the inverse covariance matrix

$$V_{ij}(1) = \frac{\partial^2 \theta}{\partial x_k \partial x_{k'}} = I_{kk'} + \sum_j S_{jk} S_{jk'} \quad . \qquad (3.24)$$

The goodness of the fit i.e. the value of θ_{min}, plays no role in determining the uncertainties at this stage of the analysis. The goodness of fit is however involved in testing the validity of the fit, as in section 3a above. The main tests of validity involve testing the compatibility of the prior information, $P(\underset{\sim}{x})$, with the model results $P(\underset{\sim}{m}|\underset{\sim}{x})$ to see if the assumption that there is a single x that is compatible with both sets of information is statistically acceptable.

Kuczera (1982) uses the parameter estimates obtained from two independent sources as a basis for measuring compatibility. In the present context the two sets of parameter estimates are the prior, for which $\underset{\sim}{x}-\underset{\sim}{q}$ is assumed to be multivariate normal with a unit covariance matrix, and the unconstrained least squares (see section 3a) for which $\underset{\sim}{x}-\hat{\underset{\sim}{x}}_o$ is assumed to be multivariate normal with covariance matrix C(O).

The random vector of differences will be distributed about $\hat{\underset{\sim}{x}}_o - \underset{\sim}{q}$ with a multivariate normal distribution with covariance matrix I + C(O). This is regarded as consistent with a difference of $\underset{\sim}{0}$ i.e. consistent parameter estimates if

$$\tilde{Q} = \sum_{kk'} (\hat{x}_{ok} - q_k) W_{kk'} (\hat{x}_{ok'} - q_{k'}) \leq \chi^2_{k,\alpha} \qquad (3.25)$$

with confidence level 1-α. The matrix W is the inverse of I + C(0), and formally, diagonalising V(0) as

$$V_{ij} = \sum_{\alpha} R_i \lambda_\alpha R_{j\alpha} \qquad (3.26)$$

gives

$$C_{ij} = \sum_{\alpha} R_{i\alpha} \lambda_\alpha^{-1} R_{j\alpha} \qquad (3.27)$$

$$W_{ij} = \sum_{\alpha} R_{i\alpha} (1 + \lambda_\alpha^{-1})^{-1} R_{j\alpha}$$

$$= \sum_{\alpha} R_{i\alpha} (\lambda_\alpha / (1 + \lambda_\alpha)) R_{j\alpha} \qquad (3.28)$$

so that W is well-defined, even if some of the λ_α are effectively zero, by

$$W = [I + V^{-1}]^{-1} = V[I + V]^{-1} = V(0)\, C(1) \quad . \qquad (3.29)$$

Note that if the components of eigenvector β are $R_{j\beta}$ and if $\lambda_\beta = 0$ and $\hat{x}_{ok}$ is replaced by $\hat{x}_{ok} + R_{j\beta}$ then the value of $\hat{Q}$ is unchanged. The compatibility test can be applied using any vector lying in the underdetermined subspace or more generally it can be approximated using any good approximation to $\hat{x}_o$ in the cases for which small λ_β lead to the precise value of $\hat{x}_o$ being difficult to determine.

(e) A Bayesian interpretation of $\gamma > 1$: redundant observations

The Bayesian analysis in the previous section was based on the assumption that the variances of both the observations and the prior estimates were known, and had been scaled to one (and that the covariances were zero). The use of $\gamma \neq 1$ implies that these assumptions are inadequate. It is in principle possible to use a Bayesian analysis to estimate one or both sets of variances but in practice, for non-linear estimation, the computational difficulties are considerable. If however a value of $\gamma \neq 1$ is selected, using one of the criteria described by Twomey, 1977, it may still be possible to interpret the results in Bayesian terms and apply the appropriate generalisations of the results given above.

The main possibility for inappropriate variance estimates arises from the degree of subjective judgement that needs to be applied. Bolin et al (1981, section 1.2.3) suggest that the appropriate variance measure of lumped data describes the variability across the whole geographical range rather than the variance of the mean value. In terms of the notation of section 2, this is an attempt to estimate the uncertainties due to the errors η_i. Such an approach can lead to a requirement for $\gamma > 1$ in at least two ways.

The first possibility is that the estimates of the variances of the prior values have, in order to avoid excessive constraints, been set too high, and should be scaled down by a factor of γ.

Thus (3.20) is replaced by

$$P(\underset{\sim}{x}) \propto \exp\ (-\tfrac{1}{2}\gamma \sum_k (x_k - q_k)^2) \qquad (3.30)$$

and $V(\gamma)$, $C(\gamma)$ replace $V(1)$, $C(1)$ throughout the analysis of the previous section.

Alternatively, values of $\gamma > 1$ may be appropriate if the observations m_i do not have independent errors. With a series of similar measurements such as successive measurements at yearly intervals, the errors arising from experimental technique will usually be independent and, if the spacing is sufficiently large, errors arising from environmental variability will also be independent. However errors arising from effects such as calibration uncertainties, or the fact that the set of observations does not adequately represent a quantity with spatial or temporal variability will not usually be independent. In terms of the notation of section 2, if ε_i includes only measurement errors they may be independent, but if they are required to include the errors η_i associated with spatial variability, this independence may be lost. The use of $\gamma > 1$ would be appropriate when this type of error dominates so that, for the purposes of fitting the model, the m_i represent a number of equivalent measurements, and an improved approximation to $P(\underset{\sim}{m}|\underset{\sim}{x})$ would be

$$P(\underset{\sim}{m}|\underset{\sim}{x}) \propto \exp\ (-\tfrac{1}{2}\gamma^{-1} \sum (y_i(\underset{\sim}{x}) - m_i)^2) \quad . \qquad (3.31)$$

This results in the distribution $P(\underset{\sim}{x}|\underset{\sim}{m})$ being multivariate normal with inverse covariance matrix (or precision matrix)

$$\tilde{V}(\gamma) = I + \gamma^{-1} C(0) \quad . \qquad (3.32)$$

This use of γ to reduce the significance ascribed to the observations is most likely to be appropriate in calculations such as the fluorocarbon inversion described by Fraser et al. (1983) and Enting and Fraser, (1983) in which all the observations are measurements of the same quantity at various places and times. When a number of different types of observation are involved, the adjustment of their significance by changing a single number γ must be regarded as a crude first approximation. In any case, if the results of the analysis appear to be in disagreement with the statistical assumptions that are used in the model calibration, the estimates and residuals should be examined in detail with a view to determining the inadequacies and refining the model and/or the calibration procedure.

In spite of the limitations of this interpretation of values of $\gamma > 1$, the use of $\tilde{V}(\gamma)$ as an inverse covariance matrix for the parameter estimates does appear to agree with the usual interpretation of the inversion calculation. For very large γ, the fit is determined by the prior estimates of the parameters and it is the uncertainties in these estimates that is reflected in the final uncertainties. For $\gamma = 1$, the expression (3.32) agrees with the direct Bayesian analysis that gives an optimal combination of both sets of information.

(f) Weighted least squares

It is of course possible to regard both sets of variances as correct and simply treat the use of $\gamma \neq 1$ as a (non-optimal) weighted least-squares fit. Treating both the m_i and the q_k as independent random variables with variance 1 leads to

$$\text{Cov}\ (d_k, d_m) = \sum_{k'm'} C_{kk'}(\gamma)\ C_{mm'}(\gamma)\ [V_{k'm'}(0) + \gamma^2 I_{k'm'}] \qquad (3.33)$$

which is the multivariate generalisation of the univariate weighted least squares result that

$$\text{Var}\ (d) = (J + \gamma^2 K)/(J + \gamma K)^2 \quad . \qquad (3.34)$$

When the relative weighting of subsets of J and K observations are given the weighting J and K observations are given the weighting 1 : γ, the minimum variance occurring at $\gamma = 1$. While expression (3.33) covers the full range from using purely prior information ($\gamma \to \infty$) to the direct Bayesian ($\gamma = 1$) use of all information through to using only the m_i ($\gamma = 0$) and gives the appropriate covariances in each case, expression (3.33) relies on the assumption that each of the scaled variances is equal to 1. Evaluating (3.33) can be useful to show how $\gamma = 1$ gives estimates of lower variance than other γ values but if a value $\gamma \neq 1$ is used for obtaining final estimates then the implication is that the assumptions on which (3.33) is based are in fact incorrect.

4: Sensitivity Analysis

This section considers the extent to which uncertainties in the parameter estimates are reflected in the uncertainties in predictions made by the model. The analysis will be confined to the case of Bayesian analysis with possible data redundancy (as in section 3e above). The standard Bayesian analysis (section 3d) is just a special ($\gamma = 1$) example of this general analysis while if the Bayesian analysis with over cautious prior variances is encountered, the analysis should be reinterpreted in terms of a standard Bayesian analysis. The analysis for $\gamma = 0$ is a conventional least-square analysis and the sensitivity analysis for a carbon cycle model is described by Enting and Pearman (1982, 1983).

From equation (3.32) the deviations d_k have a multivariate normal distribution with covariance matrix $\tilde{C}(\gamma) = \tilde{V}(\gamma)^{-1}$. Additionally a single d_k deviation has a normal distribution with variance $\tilde{C}_{kk}$. The quantity θ can be approximated by

$$\theta = \theta_{min} + \gamma \sum_{kk} d_k\ \tilde{V}_{kk'}\ d_{k'} \quad . \qquad (4.1)$$

The quantity $\theta^* = \theta - \beta\phi d_j$, (4.2)

in which $\beta\phi$ acts like a Lagrange multiplier, is minimised when

$$\frac{\partial\theta^*}{\partial x_k} = \sum_{k'} 2\gamma \tilde{V}_{kk'} d_{k'} - \beta\phi I_{ji} = 0 \qquad (4.3)$$

or, on multiplying both sides by $\tilde{C}$,

$$d_k = -\frac{\phi\beta}{2\gamma} \tilde{C}_{kj} \qquad (4.4)$$

so that the variance of d_j can be obtained as $2\gamma/\phi\beta$ times the value of d_j^* that minimises θ^*.

If we consider a model prediction Z that depends linearly on one of the parameters as

$$Z = Z(\hat{x}) + \beta d_j \qquad (4.5)$$

Then var (Z) = β^2 var (d_j)

and since

$$\theta^* = \theta - \phi(Z - Z(\hat{x})), \qquad (4.6)$$

$\theta - \phi Z$ is minimised by the value of x that has

$$\begin{aligned} Z^* - Z(\hat{x}) = \beta d_j^* &= -\frac{\phi\beta^2}{2\gamma} \tilde{C}_{jj} \\ &= -\frac{\phi\beta^2}{2\gamma} \text{var } (d_j) \\ &= -\frac{\phi}{2\gamma} \text{var } (Z) \end{aligned} \qquad (4.7)$$

or

$$\text{var } (Z) = -\frac{2\gamma}{\phi} (Z^* - Z(\hat{x})) \qquad (4.8)$$

Since this expression has removed all reference to particular vector directions or matrix elements, it will be independent of the basis used to define the parameter space and will apply to any function of the model parameters that has an approximately linear variation about the best fit point. This type of analysis has been applied to predictions of future atmospheric CO_2 concentrations by Enting (1983).

5. Conclusions

The main results of this report are the various variance/covariance expressions listed in section 3. The range of different expressions shows clearly that even though the different approaches to the inversion calculation can lead to the same parameter estimates, the different statistical assumptions lead to different expressions for the parameter uncertainties. The choice of the most appropriate expression is a matter for the scientific judgement of the modeller. When a sufficient amount of prior information exists, the Bayesian analysis becomes particularly appealing because it formalises some of the ideas that appear intuitively in other approaches. In addition, as emphasised by Box and Tiao (1973) the Bayesian approach can be used in a process of iterative model refinement as new information is obtained.

Acknowledgement A number of helpful discussions with Dr. P. Kabaila and Dr. G. Kuczera are gratefully acknowledged.

Appendix:

Notation

$C, C(\gamma)$	Matrix with elements C_{ij}, or $C_{ij}(\gamma)$ which gives covariance of parameter estimates. Equal to the inverse of matrix V, $V(\gamma)$.
d_k	Deviation of parameter x_k from best fit value $\hat{x}_k$
$E\{\ \}$	Statistical expectation (mean) of quantity in brackets.
F	F statistic
i	integer index
I	Identity matrix, usually K x K
J	Number of observations, m_j, to be fitted
k	integer index, usually range 1 to K
K	number of model parameters x_k
$\underset{\sim}{m}$	Vector of observed values, components m_j
$P(\ )$	Probability. Bayesian analysis uses a number of probability distributions: $P(\underset{\sim}{x})$: This is the prior distribution, describing the probabilities of parameter values as assessed before any observations have been made. $P(\underset{\sim}{m})$: The probability distribution for observing particular values given the error distribution for observations and the probability distribution of the parameters. $P(\underset{\sim}{m},\underset{\sim}{n})$: Joint probability distribution for a set of observed values and parameter values. $P(\underset{\sim}{m}\,|\,\underset{\sim}{x})$: Probability distribution for a set of observed values, conditional on a specified set of parameter values. This is assumed to be given by a multivariate normal distribution about the most probable observations which are calculated by the model from a given set of parameters. $P(\underset{\sim}{x}\,|\,\underset{\sim}{m})$: The probability of the set of parameter values, given that a set of observed values have been measured, i.e. the probability distribution that describes the results of fitting the model.
$\underset{\sim}{q}$	Vector of prior parameter estimates, components q_k
Q	Measure of deviations of parameter vector from best-fit.
R	Matrix used to diagonalise V
S	Sensitivity matrix, elements $S_{jk} = \frac{\partial y_j}{\partial x_k}$
t	Threshold of significance of eigenvalues

$V(\gamma)$	Matrix $\gamma I + S^T S$
W	$[I + V^{-1}]^{-1}$ matrix used in testing compatibility of information
$\underset{\sim}{x}$	Vector of parameter values x_i
$\underset{\sim}{x}_{true}$	Unknown vector of 'true' parameter values
$\hat{\underset{\sim}{x}}$	Estimate of $\underset{\sim}{x}_{true}$, based on $\underset{\sim}{m}$ and $\underset{\sim}{q}$
$\hat{\underset{\sim}{x}}_o$	Estimate of $\underset{\sim}{x}_{true}$ based on m only
$\underset{\sim}{y}$	Vector of model predictions for the observable quantities whose measured values are given by $\underset{\sim}{m}$
Z	Used in sensitivity analysis to denote any quantity predicted by the model
α	Probability level used in confidence levels
β	Scale factor used in calculating model sensitivity
γ	Quantity describing the relative contributions of $\underset{\sim}{m}$ and $\underset{\sim}{q}$ to the parameter estimates
$\underset{\sim}{\varepsilon}$	Errors in $\underset{\sim}{m}$ - especially errors in observations
θ	Function whose minimum defines the parameter estimates
λ_μ	Eigenvalue
ν	Number of degrees of freedom in fit
μ	Eigenvalue index
$\underset{\sim}{\eta}$	Additional errors in $\underset{\sim}{m}$ due to effects such as non-representativeness of measurements
χ^2_ν	Chi-squared random variable with ν degrees of freedom
$\chi^2_{\nu,\alpha}$	Value of χ^2_ν corresponding to probability $1 - \alpha$.

References

Backus, G. and Gilbert, F. (1968). The resolving power of gross earth data. Geophys. J. Astr. Soc. 16:169-205.

Backus, G. and Gilbert, F. (1970). Uniqueness on the inversion of inaccurate gross earth data. Phil. Trans. R. Soc. Lond. 266A 123-192.

Bolin, B., Björkström, A., Holmen, K., and Moore, B. (1983). The simultaneous use of tracers for ocean circulation studies. Tellus (to be published).

Box, G.E.P. and Tiao, G.C. (1973). Bayesian inference in statistical analysis. Addison-Wesley (Reading, Mass.) 588 pp.

Bretherton, F.P., McPhaden, M.J. and Kraus, E.G. (1983). Preprint entitled 'Design studies for climatological measurements of heat storage'.

Deutsh, B. (1965). Estimation theory. Prentice Hall (Englewood Cliffs N.J.) 260 pp.

Enting, I.G. (1983). Constrained inversion in carbon cycle modelling (in preparation).

Enting, I.G. and Pearman, G.I. (1982). Description of a one dimensional carbon cycle model. CSIRO Division of Atmospheric Physics Technical Paper #42. 95 pp.

Enting, I.G. and Fraser, P.J. (1983). Interpretation of observations of chlorofluorocarbons. (in preparation)

Enting, I.G. and Pearman, G.I. (1983). A one-dimensional carbon cycle model with temperature dependence in the air-sea exchange. (in preparation).

Fraser, P.J., Hyson, P., Enting, I.G. and Pearman, G.I. (1983). Atmospheric CCl_3F : Observations and modelling of the global distribution and southern hemisphere trends. Nature 302:692-695.

Frenkiel, F.N. and Goodall, D.W. (eds) (1978). Simulation modelling of environmental problems. SCOPE #9. John Wiley and Sons (Chichester) 112 pp.

Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: Biased estimation for non-orthogonal problems. Techometrics 12:55-67.

Hoerl, A.E. and Kennard, R.W. (1970b). Ridge regression: Applications to non-orthogonal problems. Technometrics 12:69-82.

Jackson, D.D. (1972). Interpretation of inaccurate, insufficient and inconsistent data. Geophys. J. R. Astr. Soc. 78:97-109.

Kuczera, G. (1982). Improved inference of catchment model parameters. Civil engineering research report 1/1982, Monash University (Clayton, Australia) 137 pp.

Twomey, S. (1977). Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements. Elsevier (Amsterdam). 243 pp.

Whitney, C. (1977). Extending radiative transfer models by the use of Bayes' rule. J. Atmos. Sci. 34:766-772.