Modeling of uncertainties in statistical inverse problems

Jari Kaipio¹,²
¹ Department of Mathematics, University of Auckland, New Zealand
² Department of Physics, University of Kuopio, Finland
E-mail: jari@math.auckland.ac.nz, jpkaipio@physics.uku.fi

Abstract. In all real world problems, the models that tie the measurements to the unknowns of interest, are at best only approximations for reality. While moderate modeling and approximation errors can be tolerated with stable problems, inverse problems are a notorious exception. Typical modeling errors include inaccurate geometry, unknown boundary and initial data, properties of noise and other disturbances, and simply the numerical approximations of the physical models. In principle, the Bayesian approach to inverse problems, in which all uncertainties are modeled as random variables, is capable of handling these uncertainties. Depending on the type of uncertainties, however, different strategies may be adopted. In this paper we give an overview of typical modeling errors and related strategies within the Bayesian framework.

1. Introduction
The mathematical definition of ill-posed problems is based on the continuity and stability of the problem [1]. On a practical level, inverse problems can be defined as problems that tolerate measurement and modelling errors poorly.

In the context of Bayesian inference, or (in our case) the statistical inversion paradigm, the poor tolerance has to be interpreted in a quite different manner since the aim is to extract the uncertainty of the unknowns given the measurements and the model. In this context, large measurement and modelling errors lead to large uncertainty of the unknowns.

Consider a linear ill-posed problem with additive measurement errors

\[ y = A_\sigma x + e \]

where the operator \( A_\sigma \) depends on the parameters \( \sigma \in \mathbb{R}^p \) which are unknown\(^1\). Here, we would interpret \( x \) as the primary (interesting) unknown and \( \sigma \) as an uninteresting but unknown parameter.

In a case in which \( \sigma \in \mathbb{R} \), known bounds \( \sigma \in (\sigma_1, \sigma_2) \) and knowledge that \( \sigma \rightarrow A_\sigma \) is smooth, we can of course use any regularization method to get an estimate for \( x \) using different parameters \( \sigma \) and thus get an estimate for the bounds for \( x \).

\(^1\) In this paper, when we refer to some variables as unknowns, we don’t necessarily mean that they are completely unknown, that is, there would be no known bounds on the variables or that their marginal distributions would be uninformative.
In a case in which $\sigma$ does not allow for a small dimensional approximate representation, however, this approach becomes infeasible. Furthermore, if we try to obtain bounds on the solution using a model $y = Ax + e$ based on bounds on the operator norm $\| A - A_{\sigma} \|$, the resulting bounds may be enormous.

In most practical cases, the approach has been to guess a "good" value for $\sigma$, use this value to obtain a single estimate for the unknown, and hope for the best. We refer to the issue of using an operator $A$ that is not (in some sense) the correct one, as introduction of modelling or approximation errors, depending on the nature of the adopted $A_{\sigma}$.

As for the models for the (measurement) noise, the typical implicit assumption has been that the noise is Gaussian with independent identical distributed (iid) uncorrelated components. While the Gaussian distribution as such can be supported in several physical measurement setting due to the central limit theorem, it is often hard to support the uncorrelatedness, especially when some of the measurement are carried out simultaneously. It is also to be noted that even under the uncorrelated iid model, the error norm $\delta = \| e \|$ is actually a random variable rather that a known number\(^2\).

The experience of the author is that in most practical cases that can be described as inverse problems, the measurements can be carried so that the modelling errors are the prevailing ones. In such cases, if the modelling errors are not handled properly, the inherent accuracy of the measurements, and thus also the unknowns, is lost.

Our aim is to stress that any reliable inference of the primary interesting unknowns necessitates feasible modelling of all uncertainties and the exploitation of these models. The bad news is that the feasibility is crucial: modelling the uncertainties too small will give misleading estimates, and modelling the uncertainties too large (to be on the safe side) will lead to losing the inherent accuracy of the information. General references on Bayesian statistics and Markov chain Monte Carlo modelling are [2, 3, 4] while [5, 6, 7] include treatises from the inverse problems point of view.

2. Inverse problems in the Bayesian framework

What makes Bayesian inverse problems any different from Bayesian inference in general? The answer is reminiscent of the classical definition of ill-posed problems: with stable, or well-posed, problems, small measurement and modelling errors have small impact on the estimates. In the Bayesian framework, loosely speaking, this means that the actual (correct) unknown lies within the posterior uncertainty estimates with high probability. With ill-posed problems, on the other hand, the correct unknown can easily be interpreted as an impossible solution to the problem.

In the following, we use the following conventions: $y$ denotes the measurement, $x$ the primary (typically the interesting) unknown, $\sigma$ the secondary (typically uninteresting) unknowns and $e$ the (conventional) measurement errors. In the case of structurally different competing models, we denote the models with $M_{\ell}$, where $\ell = 1, 2, \ldots$.

2.1. Formulation

All information of the random variables $\chi$ is decoded in the joint density $\pi(\chi)$, where in our case $\chi = (y, x, \sigma, e)$. According to the Bayes’ theorem, the joint density can be written in terms of the conditional densities as

$$
\pi(\chi) = \pi(\chi_1|\chi_2)\pi(\chi_2) = \pi(\chi_2|\chi_1)\pi(\chi_1)
$$

where $\chi_1, \chi_2$ are any two subsets of $\chi$ so that $\chi = \chi_1 \cup \chi_2$.

\(^2\) Let $e \in \mathbb{R}^N$ be an Gaussian iid random vector. Then $\delta = \left( \sum_k e_k^2 \right)^{1/2}$ is a function of random variables and thus a random variable itself with some probability distribution.
The information and uncertainty of the random variables $\chi_1$ given $\chi_2$ (fixed/known/measured values for $\chi_2$) is coded in the conditional distribution $\pi(\chi_1|\chi_2)$. Assume next that one is interested in the RV’s $\chi_1$, the RV’s $\chi_2$ have been measured and the RV’s $\chi_3$ are uninteresting. Then the formal task is to compute the conditional distribution

$$
\pi(\chi_1|\chi_2) = \int \pi(\chi_1, \chi_3|\chi_2) \, d\chi_3
$$

where the integration may be anything from a straightforward to a formidable task.

Let $\chi = (y, x, e)$ and consider the common task, that is, determine the conditional distribution of the primary unknown given the measurements under the additive noise model

$$
y = Ax + e
$$

and let $\pi_e(e)$ be the marginal distribution of the noise term. Then, if $x$ and $e$ are mutually independent, we can write $\pi(x, e) = \pi(x)\pi_e(e)$ and

$$
\pi(y|x, e) = \delta(y - Ax - e)
$$

which leads to

$$
\pi(x|y) = \int \pi(x, e|y) \, de = c\pi_e(y - Ax)\pi(x)
$$

where $c = \pi(y)$ is a norming coefficient, $\pi(x)$ is called the prior density and $\pi(y|x) = \pi_e(y - Ax)$ is called the likelihood density. Although we refer to $\pi(x)$ and $\pi(y|x)$ as the prior and likelihood densities (distributions), respectively, it is, however advisable to keep in mind that in real life these are always models.

### 2.2. Inference

It is sometimes declared that the (formal) solution to an inverse problems in the Bayesian framework is the posterior distribution itself. This definition could be supported in the sense that all (statistical) questions could in principle be answered given the posterior distribution only. In practice, the posterior distribution can be a function of thousands or millions of variables and no understanding of the posterior uncertainty is acquired just by looking at the functional form of the distribution. Moreover, it is possible that the posterior distribution does not even assume an explicit (closed) form. Thus, after the construction of the posterior model, one is always left with the task of exploration, that is, computing of point estimates, interval estimates and marginal distributions.

The most common point estimates are the maximum a posteriori $x_{MAP}$ and the conditional mean $x_{CM}$ estimates. These are

$$
x_{MAP} = \text{sol max}_x \pi(x|y)
$$

$$
x_{CM} = E(x|y)
$$

The computation of the MAP estimate is an optimization problem while the computation of the CM estimate is an integration problem.

If the posterior distribution is smooth (absolutely continuous with respect to all variables), standard optimization approaches such as Newton-Raphson are commonly used. In smooth posterior cases the MAP estimate is almost always computed.

For the computation of the CM estimate the only feasible strategy in almost all cases is to resort to Markov chain Monte Carlo (MCMC) sampling. The idea in using the computationally burdensome MCMC sampling is to try to obtain a comprehensive set of samples that are
distributed according to the posterior distribution. Denote a set of samples by \( \{x^{(\ell)}, \ell = 1,\ldots,L\} \sim \pi(x|y) \). Then, for example, we have

\[
E(x) \approx \frac{1}{L} \sum_{\ell=1}^{L} x^{(\ell)}
\]

\[
P(0 < x^{(10)} \leq 5) \approx \frac{N(0 < x^{(10)} \leq 5)}{L}
\]

where \( N(0 < x^{(10)} \leq 5) \) is the number of samples for which this condition is fulfilled and \( P \) refers to the probability of an event.

It is very popular, however, to try to construct feasible Gaussian approximations for the prior and likelihood models especially in the case of linear (and mildly nonlinear) forward problems. For the linear Gaussian case, no minimization or MCMC sampling is needed, since the maximum equals the conditional mean which has a closed form solution. In addition, all marginal distributions and credibility intervals (corresponding to constant posterior densities) can be obtained in closed form.

3. Uncertainties

3.1. Random variables vs. error bounds

Error bounds are usually a feasible way to describe uncertainties only when there is only a single (real valued) unknown variable or at most two variables. To be on the safe side, the uncertainties should not be underestimated. On the other hand, a feasible model for the uncertainties is seldom an uniform distribution and, moreover, the error bounds for the estimate would not correspond to the end points defining an uniform distribution. Employing realistic lower and upper bounds for the uncertain variable almost exclusively leads to infeasibly large uncertainty estimates for the primary unknowns.

Treating the secondary unknowns as random variables enables one to achieve much more consistent and feasible (error/uncertainty) estimates for the primary unknowns. It has to be noted that, however, that statistical modelling of the secondary unknowns can turn out to be a major task both in statistics and (physical) modelling. A good example is the problem of having partially unknown boundary data in a case in which the forward model is induced by a boundary value problem, say, a Dirichlet problem, and a finite element scheme is used for the computational implementation. In this case a statistical model should be constructed for the nodal values on the unknown boundary. In most problems, the adoption of a Gaussian iid model would be completely infeasible.

Also, in many cases the primary and secondary unknowns can be heavily correlated, such as in the case of modelling errors (see Section 5.4). Implementing such information and models into error bounds can turn out to be impossible or at least infeasible.

3.2. Primary and secondary unknowns

The categorization of unknowns into primary (interesting) and secondary (uninteresting) ones is usually clear. Posterior information for the former group is sought while the second group represents “nuisance” parameters.

There are, however, situations in which the variables that are usually treated as secondary, may actually be the primary ones. For example, in a blind deconvolution problem that is related to an imaging device, the width of the point spread function might be considered as a nuisance parameter, but might actually be the variable that is sought to analyze the performance of the overall imaging system or to estimate in which state the imaging system is.
3.3. Inherent vs. realization related uncertainties and errors

We can differentiate between three levels of modelling related errors and uncertainties.

On the first level there are inherent uncertainties in the (forward) model, such as unknown boundary conditions and uninteresting secondary parameters. For these, a prior model has to be designed apart from the prior model for the primary unknowns.

On the second level, there are cases in which the “natural” forward problem is known up to all details, but it is too cumbersome to implement as the forward model for the likelihood.

On the third level, the forward model for the likelihood is well known and as such implementable. In some cases, however, especially in those that are related to on-line operation in industrial settings, there is limited time to compute the estimates and the forward problem. Then, especially in problems related to PDE’s, a (too) small-dimensional approximation for the forward problems has to constructed.

The dividing line between the first level and the two latter level cases the errors due to the approximations (either with respect to model approximation or numerical approximation) are in fact functions of the unknown. This topic is discussed further in Section 5.4.

3.4. Risks: underestimation and overestimation of uncertainties

The case which absolutely has to be avoided, is to underestimate the errors, be they due to traditional measurement errors or modelling based errors. The outcome of underestimation is usually that the estimates are incompatible with reality altogether. With this we refer to a case in which the actual unknown \( x^* \) is such that we have \( \pi(x^*|y,M) \approx 0 \), where we denote the adopted (computational) model with \( M \).

On the other hand, overestimation of uncertainties and underestimation of, say, measurement accuracy means that we lose information unnecessarily. In other words, the information in measurements (and models) is not exploited properly. The key is that the uncertainties can have structure that can be exploited. We shall return to this topic in Section ??.

4. Typical uncertainties in inverse problems

4.1. Forward and likelihood models

Consider only the simplest possible likelihood model, linear observations with additive noise model

\[
y = Ax + e
\]

If the model \( A \) can be considered exact and deterministic, we usually refer to the mapping \( x \to Ax \) as the forward model. Then, given a model for the additive noise and assuming that \( e \) and \( x \) are independent, we get the likelihood model

\[
y|x \sim \pi_e(y - Ax)
\]

In a typical measurement setting, however, a realistic (feasible) model is seldom such that the mapping \( A \) would be accurate and exactly known. Below, we list typical uncertainties. We note, again, that ill-posed inverse problems by definition are problems that tolerate measurement and modelling errors poorly.

4.1.1. Partially unknown boundary data

In practically all problems in which the forward model is related to a partial differential equation (PDE), at least some of the boundary data or conditions are unknown.

As a typical case, consider the geophysical electrical resistance tomography setting in which a feasible boundary value problem, the complete electrode model, is well known. The measurements are carried out by injecting currents into the soil via electrodes that are placed...
on the ground or in boreholes. The resulting voltages on other electrodes are the measured. Given the electrode contact impedances and boundary conditions, the forward model is thus specified.

For the computational realization, however, we have to truncate the computational domain so that all electrodes and the overall region of interest is within this (computational) domain. In such a situation, we know the boundary conditions only on the surface of the ground but not on the truncation boundaries. Furthermore, the related boundary operator, the Diriclet to Neumann mapping depends on the resistivity outside the computational domain, which we should know to be able to specify the proper boundary conditions.

The situation is similar in essentially all real world PDE induced problems.

4.1.2. Uncertain geometry Problems in which the overall geometry and/or sensor locations are only approximatively known, are quite common. This is typical especially in problems that are related to biomedical engineering. For example, in electroencephalography source location problem, the measurement electrodes are commonly attached to a rubber web that accommodates to the shape of the head. The forward model, however, is still today often based on a standard head model with standard locations of the electrodes. In other words, both the geometry of the domain and the electrode locations are only approximate.

4.1.3. Gross model As an example, the commonly agreed best model for light propagation in turbid media is the radiative transfer equation. When diffuse optical tomography is modelled, however, the so-called diffusion approximation is almost always employed, see, for example, the article by Simon Arridge in this volume. While in cases in which the scattering coefficient is large enough, the diffusion approximation can be taken to be an adequate approximation for the radiative transfer model, many interesting practical situation are on the verge of this condition.

4.1.4. Model reduction Especially with nonlinear inverse problems with accurate measurements, which would require a very accurate forward solver, one could find that the computational requirements surpass the available resources. This is especially typical in industrial problems. To accommodate to the available resources, one might be forced to resort to numerical and computational approximations that would introduce significant errors (larger than the measurement errors) to the model predictions. If these errors could not be modelled, the resulting estimates would most probably be worthless.

The dual problem to this is that given fixed computational resources, if the actual measurement accuracy would be increased, one would not be able to exploit this increased accuracy. These kind of problems can be considered by employing the approximation error approach, see Section 5.4.

4.1.5. Noise model Additive measurement error, or noise, models arise naturally. The most commonly employed additive error model is the zero mean Gaussian iid model, that is, we write (again) \( y = Ax + e \) and set \( e \sim N(0, \lambda I) \). But even in cases in which the additive error model can be supported, the choices of i) zero mean, ii) independence (or uncorrelatedness) and iii) normality (Gaussianity) of the error variables can often be questioned. Furthermore, the (additive) errors in the likelihood model are almost exclusively taken to be mutually independent with the primary unknowns.

Consider an industrial measurement setting in which a number of measurements are carried out simultaneously and the measurement information is carried via analogue transfer lines to the data loggers. For example, in a thermomechanical pulp plant, there can be a total of 20 MW of electrical motor power dissipated within a 20m radius around the measurement setting. The electromagnetic field strength in the region that the measurement leads occupy...
are easy to estimate. Furthermore, the associated frequencies are typically small and thus the wavelengths are very long. Thus, the mutual (cross) correlations of the errors induced to the measurement leads can be very high instead vanishing altogether. Although this is an extreme example, significant correlations with additive errors are quite common. Most remarkably, these correlations (as well as the error mean) can in most cases be measured directly.

4.2. Prior model
We stress that in most cases prior distributions have to be understood as models whose characteristics could (and should) be evaluated. By this we mean that draws form this distribution should have the expectable characteristics of the unknowns. As with likelihood models, using a too narrow prior distribution is likely to lead to posterior models with vanishing density for the actual unknowns.

A note concerning smoothness priors: it is advisable to use proper smoothness prior model, that is, models whose covariance exists, rather than improper models whose inverse covariance is of the form $\Gamma^{-1} = D^T D$ where $D$ is a differential operator with a nonvanishing null space.

As with likelihood models, the underestimation of uncertainty (variance and covariance) of the prior model is to be avoided. In the case of prior model, however, the overestimation of the uncertainty is often not as serious as in the case of the likelihood model.

Equivalently as with likelihood (additive noise) models, it is very common to use an iid Gaussian (white noise) model for the unknowns. Although white noise models could be argued to be safe in the sense that they are not prone to introduce non-existent structures (artefacts) to the estimates, they are very often completely infeasible for the problem.

4.3. Competing models
Consider a case in which there are two or more competing models either for the likelihood or the prior. For example, in a medical imaging problem, the image might be from a healthy tissue or a tumor. In such a situation one is primarily interested whether the tissue sample is healthy or not.

Let us have two competing, say, prior models $\mathcal{M}_1$ and $\mathcal{M}_2$. We denote these two prior models with $\pi(x|\mathcal{M}_1)$ and $\pi(x|\mathcal{M}_2)$ and set the prior probabilities for the models as so that $P(\mathcal{M}_1) + P(\mathcal{M}_2) = 1$. Thus we have $\pi(x, \mathcal{M}) = \pi(x|\mathcal{M})P(\mathcal{M})$ which is a formal notation only.

Formally, we have thus the posterior model

$$\pi(x, \mathcal{M}|y) = \pi(y|x, \mathcal{M})\pi(x, \mathcal{M})$$
$$= \pi(y|x, \mathcal{M})\pi(x|\mathcal{M})P(\mathcal{M})$$

It is to be noted that even when $\pi(y|x, \mathcal{M})$ and $\pi(x, \mathcal{M})$ were Gaussian for both $\mathcal{M}_1$ and $\mathcal{M}_2$, the posterior is not; in fact, it is not even continuous and thus, for example, a MAP estimate cannot be computed with a minimization type algorithm. Markov chain Monte Carlo approaches, however, can tackle these kind of problems. See [8] for the reversible jump MCMC algorithm, which is especially suitable to this kind of problems and can handle problems with nonconstant number of variables.

5. Approaches
In the following we assume that the posterior model is feasible, that is, the modelled uncertainties are compatible with the actual uncertainties.
5.1. Markov chain Monte Carlo

If a feasible sampler can be constructed and definitive answers are sought, sampling is the way to go. The two most common schemes are the Metropolis-Hastings algorithm and the Gibbs sampler and their variants. We do not go into the details of these schemes and only refer to [2, 3, 4] as general texts for how to implement MCMC samplers. The references [5, 6, 7] include treatises from the inverse problems point of view.

Regarding sampling, the main challenge with inverse problems is related to the usually relatively accurate measurements and the resulting narrow likelihoods which usually also render the posteriors at least as narrow, which in high dimensional problems makes the problem hard. Of course, one could artificially widen the likelihood model simply by multiplying the (additive) noise variances with a largish number, but this will destroy the inherent accuracy of the measurements themselves.

The sampling approach facilitates the exploration of nonsmooth posterior distributions, models with varying dimension (reversible jump MCMC) as well as evaluation of models themselves, and thus also model selection: computing posterior probabilities for the competing models.

5.2. Hypermodels

If the uncertainty in models can be modelled with a small number of parameters, the hypermodel approach is often a good approach. As a typical example, assume that we know the structure of the additive noise but not its overall strenght. Such a situation occurs, for example, when the disturbance source is an external electromagnetic field and the location of the measurements varies so that the overall field strength is not known exactly. Then we could write, for example,

\[ y = Ax + e, \quad e \sim N(0, \lambda \Gamma) \]

where \( \Gamma \) contains the structure of the errors and \( \lambda \) is a real-valued random variable, whose distribution could often easily be modelled by carrying out dummy measurements in the field. Analogously, if the prior model is a smoothness prior model but the scale of the unknown is not known, a prior model of the form \( \pi(x) = N(x_0, \kappa \Gamma_x) \) can be adopted.

Here, an important practical remark must be made. Especially in the case of Gaussian models, the normalizing constant is usually dropped. In the above cases, the normalizing constant is not a constant with respect to the scaling variable. Thus we have to write

\[
\pi(x) \propto \kappa^{-N/2} \exp \left( ||\kappa^{-1/2} L_x (x - x_0)||^2 \right)
\]

where \( \Gamma_x^{-1} = L_x^T L_x \) and \( x \in \mathbb{R}^N \).

5.3. Analytic marginalization

In block and single variable variants of the sampling algorithms, analytic marginalization (computation of the full conditional densities) with respect to some variables or subsets of variables can often be carried out. If we are able to draw from these full conditional densities, for example, in block Metropolis-Hastings type algorithms the acceptance probability becomes unity and all proposals will be accepted.

5.4. Approximation errors

The approximation error approach may be a feasible way to handle modelling errors and numerical approximations in cases in which the number of uncertain variables is large. This approach can be applied to a wide variety of problems in which a reasonably accurate forward solver for the problem is accessible.
The formulation of the approach in a case in which there are uncertain parameters and, simultaneously, a crude numerical approximation for the forward problem is employed, is as follows.

Let \( y = \bar{A}_\sigma(\bar{x}) + \epsilon \), where the statistics of \( \epsilon \) is known and there is a proper statistical (prior) model for both \((x, \sigma)\). For the likelihood, however, we are to use the model \( y = A(x) + \epsilon \), where the operator \( A \) is realized typically in a sparse mesh and a fixed value for the random variable \( \sigma \) has been used.

We then write

\[
\begin{align*}
y &= \bar{A}_\sigma(\bar{x}) + \epsilon \\
&= A(x) + e + (\bar{A}_\sigma(\bar{x}) - A(x)) \\
&= A(x) + e + \varepsilon
\end{align*}
\]

where \( \varepsilon = \bar{A}_\sigma(\bar{x}) - A(x) \). Now since we have a model \( \pi(x, \sigma) \) and \( \varepsilon \) is a function of the two random variables, \( \varepsilon \) is also a random variable and we can construct an estimate for its (at least second order) statistics. Since we have realizations for both \( \bar{A}_\sigma \) and \( A \), we can simply draw from \( \pi(x, \sigma) \) and compute the approximation errors \( \varepsilon \), and an approximation for the joint statistics \( \pi(x, \sigma, \varepsilon) \). See [5, 9, 10] for details.

6. Conclusions

To facilitate reliable inference for any given application, it is of primary importance to get the models and especially the related uncertainties right. The adoption of means that facilitate accurate inference, such as MCMC methods, could be said to be justified only when the respective models are correspondingly accurate or feasible. The problem is that in most cases it is impossible to assess the feasibility of the models. On the other hand, the task of developing efficient samplers for large dimensional problems is of utmost importance in itself.

References

[1] Tikhonov A N and Arsenin V Y 1977 Solution of ill-posed problems (Washington, DC: Winston & Sons)
[2] Gilks W, Richardson S and Spiegelhalter D 1996 Markov Chain Monte Carlo in Practice (Chapmann & Hall)
[3] Liu J 2005 Monte Carlo Strategies in Scientific Computing (Springer Series in Statistics no ISBN 0-387-95230-6) (Springer)
[4] Robert C and Casella G 2004 Monte Carlo Statistical Methods (Springer Texts in Statistics no ISBN 0-387-21239-6) (Springer)
[5] Kaipio J and Somersalo E 2005 Statistical and Computational Inverse Problems (New York: Springer)
[6] Calvetti D and Somersalo E 2007 An Introduction to Bayesian Scientific Computing · Ten Lectures on Subjective Computing ISBN 978-0-387-73393-7 (Springer)
[7] Tarantola A 2004 Inverse Problem Theory and Methods for Model Parameter Estimation (Philadelphia: SIAM)
[8] Green P 1995 Biometrika 82 711–732
[9] Kaipio J and Somersalo E 2006 J Comput Appl Math In Press
[10] Arridge S, Kaipio J, Kolehmainen V, Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 22 175–195