BEFORE INDEPENDENT HEARING COMMISSIONERS

AT NAPIER & WAIPAWA

I MUA NGĀ KAIKŌMIHANA WHAKAWĀ MOTUHAKE

KI AHURIRI & WAIPAWA

IN THE MATTER	of the Resource Management Act 1991
AND	
IN THE MATTER	of the hearing of submissions on applications for the take and use of water from the Ruataniwha Basin.

STATEMENT OF EVIDENCE OF DR NICK DUDLEY WARD
ON BEHALF OF APPLICANTS

31 OCTOBER 2022

RICHMOND
CHAMBERS

Counsel Instructed
B J Matheson
Richmond Chambers
PO Box 1008
Shortland Street
Auckland 1140
E. matheson@richmondchambers.co.nz

1. INTRODUCTION

1.1 My name is Nick Dudley Ward.

Qualifications and experience

1.2 I have a BSc and DPhil in Mathematics and an MS in Civil Engineering. I am also a chartered professional engineer (CPEng) and a practice area assessor for Engineering New Zealand.

1.3 Between 2015-2020 I was a senior lecturer in civil engineering at the University of Canterbury where I taught and carried out research on groundwater, with particular emphasis on the underlying physics, mathematics, and uncertainties.

1.4 Since 1991 I have had academic/research roles at a number of universities and institutes around the world including the Universities of Auckland, Otago, Leeds, the University of Pennsylvania, and the University of Eastern Finland, and at the French national research institute, INRIA.

1.5 In particular, between 2003-2006 I lectured courses in statistics and probability at the undergraduate and postgraduate levels at the Wharton School, University of Pennsylvania. Techniques from statistics and probability feature prominently in my research.

1.6 Between 1998-2001 I worked on the modelling and assessment of catastrophic risk in the Lloyd's market, and between 2007-2010 I worked as an engineer for PDP.

1.7 Since 2020 I have been resident in Australia and between 2018-2021, I was an academic visitor in the Research School of Engineering at the Australian National University. In 2021 I established a Tasmanian based company, SEPCO Technologies, a spin-off of my university research. The company specialises in improved characterisation of groundwater resources using geophysical measurements and advanced algorithms, http://www.prospectinganz.com.

1.8 Since 2007 I have published extensively on groundwater hydraulics, and novel techniques for groundwater characterisation and uncertainty quantification. I am familiar with, and am responsible for some of, the tools that are commonly used to quantify the impacts of groundwater development in New Zealand.

Code of Conduct

1.9 I have read the Environment Court's Code of Conduct and agree to comply with it. My qualifications and relevant experience as an expert are set out above. I confirm that the issues addressed in this statement of evidence are within my area of expertise, except where I rely on the opinion or evidence of others where noted.

Scope of evidence

1.10 The aim of my evidence is to provide some clarification on the value of groundwater models, their uncertainties, and their use in making decisions.

1.11 My evidence is divided into the following sections:

(a) General comments on numerical groundwater models and their uncertainties.

(b) Comments on the Aqualinc numerical model.

(c) Assessment of the key uncertainties.

1.12 I also participated in joint witness conferencing with Mr Thomas and Mr Weir, and I signed the record of that conference.

2. GENERAL COMMENTS ON GROUNDWATER MODELS AND THEIR UNCERTAINTIES

2.1 The purpose of a catchment scale model is to simulate large-scale effects in space and time.

2.2 It is important to set reasonable expectations and recognise that there are limits to what can be achieved. Predictive uncertainty cannot be eliminated

from any model simply by making it more and more sophisticated. As with COVID-19 and Climate Change models there will always be an envelope of predictive uncertainty and scope for error, but this does not detract from their critical role in decision-making.

2.3 There are three aspects of a model that are important for decision-making and require transparency:

(a) how the model is built;

(b) how it is calibrated; and

(c) its predictive reliability.

2.4 Generally, a model will have value for decision making if it is built on reasonable physical assumptions. Simplicity is usually the best guide – trying to over-model is usually uninformative and counter-productive. The modeller needs to set clear objectives and find the right balance.

2.5 The calibration and estimation of predictive reliability of a groundwater model are complex and mathematically challenging tasks.

2.6 Calibration means being able to reproduce historical data to a certain tolerance. It is very common that data sets are small, have gaps or are inaccurate. There are many well-founded statistical techniques for producing data in order to calibrate a model. In other cases, data may not be available, and the modeller may choose to use synthetic data either as a model input or calibration data.

2.7 Calibration of a model requires a starting point. For a groundwater model that aims to simulate groundwater flow over time a very sensible way to begin the calibration is to calibrate a steady-state version. These parameters are used to initialise the calibration of the transient model and run a mathematical optimiser.

2.8 It is important to recognise that model calibration is never unique – there are usually many different calibrated models that explain data.

2.9 The estimation of the predictive reliability of a model is the most challenging task of all. Again, one has to set reasonable expectations bearing in mind that the predictive uncertainty cannot be reduced to zero. The task of the modeller is to interrogate the calibrated model and estimate uncertainties.

2.10 Obviously, however well-calibrated a model, uncertainty will remain – the predictions may not be correct. We will never know exactly what will happen until a scheme is implemented and appropriately monitored. This is the case with all models. The actual effects will need to be adaptively managed by monitoring and controls. Indeed, the purpose of a groundwater model from a decision-making perspective is to provide a baseline for adaptive management. I have considered this matter in detail in my joint paper - *Uncertainty, decision, and control: issues and solutions*, New Zealand Journal of Hydrology, pages 53-91, 2014. (A copy of this paper is attached to this evidence.)

3. COMMENTS ON THE AQUALINC NUMERICAL MODEL

3.1 I have read the latest version of the Aqualinc report 'RUATANIWHA BASIN Tranche 2 Groundwater Modelling (Revised 2)' together with the modelling review by Patrick Durney (Lincoln Agritech) and the evidence of Neil Thomas (PDP) dated August 2022. Both these latter documents have led to some useful refinements in the model interrogation. Mr Weir has replied in detail to the points raised in these two documents. I will confine myself to some general comments.

3.2 I find that Mr Weir has been clear with explaining his intent and objectives with his numerical model. In particular the model has been built at an appropriate resolution. Mr Weir has chosen to calibrate his model to measured groundwater levels and low river flows. This matches his modelling objectives.

3.3 The most important thing to comment on is the model calibration: Weir first calibrates a steady state model to averaged data. The aims of doing this are to get a baseline set of parameters and, more importantly, to

initialise the calibration process for his transient model. Given the computational resources available to him this is a perfectly sensible thing to do.

3.4 This point seems to have been misunderstood by Mr Thomas in his evidence referenced in item 20, who suggests that the calibration is 'unusual'. This is not the case.

3.5 With regards to the quality of model fit I make the following comments. With all statistical fitting it is always the case that some data is over-predicted, and some data is under-predicted, unless both the model and data are perfect. There are generally accepted benchmarks on what is an acceptable calibration, usually given as summary statistics. This is the result of a great deal of collective experience over many years. The US Army Corps of Engineers provides guidelines as do many other organisations. Based on the reported statistics in Mr Weir's report I would say that the model is well-calibrated relative to his modelling objectives.

3.6 With regard to the model resolution, it is simply not feasible or informative to include every single stream in the model. This would end up with a highly uncalibrated model and would be counter productive. In my view, given the available data, Mr Weir has made an appropriate choice relative to his modelling objectives. As I stated above the purpose of a model is to capture the salient physical features of interest – and not to duplicate reality at a minuscule scale. Estimates of stream depletion on the modelled streams and rivers may be taken as an index for unmodelled streams.

3.7 Both Mr Durney and Mr Thomas have discussed the potential adverse effects on existing bores. Well interference assessments using simple analytic models, as suggested by Mr Durney, and commonly carried out are usually highly subjective and do little to reduce the uncertainty of the interference that is the focus of interest. The process of estimating worst case effects then becomes arbitrary. It is probably more informative to estimate well interference using a globally calibrated model as Mr Weir has done and provide an estimate of the uncertainty of the potential

interference using the calibrated model. Regardless, I note that Mr Weir has also used a simple analytic model to estimate well interference and reports the interference is broadly consistent with his global model. This is a useful check.

3.8 One way to verify a catchment scale model is to consider the large-scale effects and carry out simple scale analysis. For example, the reduction in the volume of water in the aquifer due to the proposed scheme is a quantity of critical interest. It is very easy to get a handle on this based on simple assumptions and avoid clouded narratives about the uncertainties of large-scale models. The central question is: are the predictions of the model consistent with estimated large-scale effects? Mr Weir considers this specific question in section 3.14 of his report and concludes that his estimate of global drawdown is broadly consistent with the volume of groundwater extracted.

3.9 In conclusion, I think the Ruataniwha groundwater model is appropriately assembled and calibrated relative to the modelling objectives. I think the model has revealed as much as can be reasonably asked of it, or indeed any groundwater model, and attention should be focused on mitigating and controlling the outstanding uncertainties that are a universal feature of almost all models. The interrogation of the predictive uncertainty by sensitivity testing and Monte Carlo sampling provides envelopes of uncertainty on predicted groundwater level changes and estimates of what may happen. These have been commented on in the JWS on Groundwater Modelling and in Mr Weir's evidence.

4. ASSESSMENT OF THE KEY UNCERTAINTIES

4.1 I will now confine myself to the two key uncertainties associated with the proposal, namely

(a) the efficacy of the surface flow augmentation; and

(b) effects of the proposed groundwater development on other groundwater users.

4.2 The idea of augmenting surface flows potentially depleted by groundwater abstraction is a novel aspect of the current proposal.

4.3 The key uncertainty here is establishing its efficacy. In Mr Weir's report it is mentioned that most applicants propose to discharge augmented water directly into streams while two propose injecting groundwater into shallow wells directly hydraulically connected to the relevant stream. This is less likely to be effective. Mr Thomas also refers to this in his evidence. In any case it should be required of applicants to establish the efficacy of the particular scheme they propose to implement and show that the augmentation of surface flows is proving effective through appropriate monitoring and controls.

4.4 The primary intent of the monitoring is to ensure that the augmentation is working. This also has a secondary benefit in providing valuable data that would inform future schemes proposing a similar approach.

4.5 The second key uncertainty are the potentially adverse effects of the proposed development on other groundwater users. Mr Weir includes an assessment of this in his report in which he concludes that the effects will be small. Again, this uncertainty can be managed through appropriate monitoring and controls.

Nick Dudley Ward

31 October 2022

Journal of Hydrology (NZ) 53 (1): 53-91 2014

Uncertainty, decision and control: issues and solutions

Nicholas Dudley Ward[1] and Jari Kaipio[2]

[1] *Otago Computational Modelling Group Ltd, PO 8098, Gardens, Dunedin, New Zealand. Corresponding author: nick@ocmo.co.nz*

[2] *Department of Mathematics, University of Auckland, Private Bag 92019, Auckland, New Zealand.*

Abstract

In the light of the L'Aquila verdict in which seven scientists were found guilty of manslaughter resulting from poor communication of risk and uncertainty, scientific modelling and general assessment of risk and uncertainty is being carefully scrutinised. While the majority of scientists recognise the importance of uncertainty, there is in general much less clarity on what is a good uncertainty assessment. For this reason we give a broad survey of uncertainty assessment drawing on examples from many areas, including earthquake forecasting, liquefaction assessment, process engineering, biomedical imaging and hydrology. We review least-squares methodology and show that uncertainty assessment based on least squares can be misleading. We show how problems associated with least squares can be avoided by working in the Bayesian framework. The aim of the article is to contribute to the general debate, and show how explicit modelling of uncertainties can provide a sounder basis for decision making.

Introduction

Uncertainty is one of those words that means all things to all people. Consequently discussions about uncertainty have a tendency to descend into confusion. Even the mathematical study of uncertainty, the theory of probability, which provides the most coherent framework for dealing with uncertainty, has a reputation for confounding the expert. In this paper our interest is in quantitative uncertainty; in short, that part of uncertainty that can be effectively modelled using probability theory.

In an abstract setting we say that the uncertainty of an event has been quantified if we can attach a probability to that event. As far as the mathematical theory of probability is concerned no prescription is given how to assign probabilities. It is sufficient that probabilities are attached to events so that some self-evident axioms are satisfied (as self-evident as the axioms of arithmetic). The axioms of probability are a distillation of our basic notions about randomness and chance. These notions are themselves wrapped up in the idea of repeatability: suppose we can repeat an experiment indefinitely with n different outcomes (or events). Then our basic intuition about the proportion of occurrences of a particular event is that it will approach the probability of that event. This intuition forms the basis of the law(s) of large numbers which are themselves subtle distillations of the protean 'law of averages'. Of course, had probability been simply confined to computing odds for games of chance, a simple interpretation of probability would be sufficient. However, the contemporary scope of probability covers many fields in science, economics, and engineering, because it has proved to be an extremely powerful

tool. Views about probability range from the frequentist interpretation to a more subjective or Bayesian interpretation, where some proponents speak of 'degree of belief'. While the axioms of probability might be self-evident, there is much less agreement on how one should interpret the probability of an event. Both the frequentist and Bayesian interpretations may be regarded as extreme. For this reason it is preferable to contain prejudice and keep a flexible view towards probability because this will preserve perspicacity.

We are interested here in mathematical modelling and how probability can inform the modelling process. Traditionally, mathematical modelling has been based on solving systems of equations which (ideally) encapsulate the essential physics of a process. Usually models express conservation of some quantity like mass or momentum (e.g., groundwater flow, shallow water equations, Navier-Stokes equations). There are two directions that modelling can go. We can simply take our distilled model and study its properties in order to extend or clarify our understanding of a process. Or we can attempt to simulate the physical process by choosing physically realistic parameters (from experimental or observational data), and therefore obtain predictions about the behavior of the system. Both approaches are examples of 'forward modelling'.

Forward modelling ignores explicit consideration of uncertainty. In general it is unreasonable to expect that a model will predict exactly what happens. It is clear that we need to get to grips with predictive uncertainty and quantify the 'envelope of predictive uncertainty'.[1] Indeed, quantified uncertainty is a crucial part of the modelling process if our aim is accurately simulate physical processes in the sense that what actually happens is captured in the envelope of uncertainty. The aim of computer simulation is prediction – by mimicking a physical process we hope to understand better how the real process behaves. Our primary interest therefore is in 'inverse modelling', that is fitting the forward model to data. By explicitly modelling uncertainties and building them into the overall model architecture we can more fully explore the relationship between models and data. Crucially we can meaningfully quantify predictive uncertainty and therefore develop control measures to manage the risks.

Because interest in uncertainty assessment is currently widespread we begin with a general discussion of modelling and uncertainty. Our aim is to clarify what we mean by quantitative uncertainty and to avoid the obfuscation that is frequently associated with discussions about uncertainty. Also because uncertainty is extremely challenging, we first present an example from finance because it shows in a simple setting the essential elements of quantitative uncertainty. Next we present three examples of uncertainty assessment, which are both 'hot topics' in New Zealand and typify the broad spectrum of uncertainty assessment: engineering assessment of liquefaction, earthquake forecasting and the management of catastrophic risks (e.g., earthquakes and floods). In particular, we want to make a clear demarcation between qualitative and quantitative uncertainty. We remark that quantitative uncertainty is not the same as sensitivity testing in which model parameters are varied using, say, Monte Carlo random sampling.

Given the prevalence of least squares methodology in model calibration, we then review the least-squares method. We show by means of some simple 'classroom examples' that uncertainty assessment based on least

[1] That is, specify an interval such that there is (say) a 90 or 95% chance that what happens is contained in the interval.

squares can be misleading. We then discuss the Bayesian framework in some detail and show how problems associated with least squares can be resolved and how uncertainty can be more meaningly quantified.

In the final part of the article we present published examples from different applications areas in which uncertainty quantification carried out in the Bayesian framework has led to demonstrable advances – medical imaging, process engineering and groundwater hydrology, respectively. In our view some of the biggest technical advances in 'Bayesian inversion' have been in the first two areas, and we think it is instructive to present examples that are outside hydrology. Our broad aim is to encourage informed debate on the difficult topic of how we can develop a risk-based management framework for the management of natural resources and hazards.

Modelling

Most readers will be familiar with the much quoted comment attributed to G.E.P. Box that 'all models are wrong, but some are useful'. Unfortunately, this comment has often been given as a defence of (frequently poor) modelling. Probably, Box meant this in the context of statistical modelling, but it now seems to have universal application. However, it plainly ignores the subtlety of, and skill required for, good mathematical modelling.

Mathematical modelling is a compromise between physical insight, good sense and practicality. The Nobel-prize-winning physicist P.A.M. Dirac (better known in hydrology for the Dirac δ function) observed that 'almost all the physical processes that arise in applied sciences and engineering can be modeled accurately using the principles of quantum mechanics' (Weinan, 2011). However, such an approach is intractable because the resulting system has a *very* large number of variables. Therefore we tend to opt for systems of equations that are actually solvable on a reasonable timescale. There are two simplifying processes of mathematical modelling which stand out, and form the basis of much hydrological modelling, namely linearisation and homogenization. In a hydrological setting a good example of linearisation is Darcy's law, which states that the rate of groundwater flow is proportional to the hydraulic gradient. The constant of proportionality, the hydraulic conductivity, is a lumped parameter which depends on the solid matrix (the permeability), and fluid properties (the density and the dynamic viscosity) encapsulates some small-scale physics. Furthermore, the groundwater flow equation is a macroscopic description of flow in porous media and can be deduced from the general equations of fluid flow, the Navier-Stokes equations, by averaging out the small scale effects of flow. While the Navier- Stokes equations are themselves macroscopic models (they ignore molecular interactions, for example), the success of fluid dynamics models is indisputable.

Therefore, it is not so much the case whether a mathematical model is wrong or not. The issue is whether the correct scale has been coded into the model that is actually solved.

Averaging over small-scale processes is called homogenisation. An extreme example of homogenisation in a hydrological setting is evident (and second nature) with the groundwater flow models with few parameters which are used to fit pump test data, e.g., the Theis or Boulton flow model. In effect, all small-scale heterogeneities are smoothed out so that one parameter represents hydraulic conductivity, storativity or leakage. It is plain that a pump test is unlikely to identify an unseemly large boulder in an aquifer, but it may identify obstructions to flow, preferential flow paths, or external water sources.

To have a realistic simulation of a physical process we need to calibrate to data, and solve the *inverse problem* of fitting a forward model to data. In short, a good forward model is one that captures the right scales (i.e., the ones that are observable in data), and is solvable on a time scale that permits solution of the corresponding inverse problem on a time scale that is useful.

Why do uncertainty?

To motivate consideration of uncertainty it is instructive to consider the following (unrealistic) example of an investor portfolio.

Consider a portfolio consisting of 100 randomly selected shares. We can safely assume that a sensible investor would be unwilling to invest in this portfolio armed with this simple knowledge. Suppose then that the investor is told the annual average returns of the individual shares. We can assume (perhaps with less confidence) that the savvy investor would still be unwilling to invest. Probably, our investor would want to know the uncertainty (usually known as the 'volatility') of the portfolio. However we frequently make decisions in the absence of uncertainty. There are many reasons for this. In our opinion two stand out: the technical problem of quantifying uncertainty for realistic models and the absence of user-friendly tools. Second is the challenge of fitting uncertainty into a decision-making framework. However, it is hard to solve the latter without solving the former. Instead of getting bogged down in a discussion about uncertainty, we prefer here to present several examples from different application areas to highlight some of the recent successes of 'uncertainty quantification'.

To continue with the investor example, to quantify the uncertainty of the portfolio we need to posit a model for the share prices. Let X^i denote the value of share i at time t, and let $dX^i = X^i_t - X^i_{t-dt}$ be the change in value between t and $t + dt$.

If the value were purely deterministic we might posit a fixed return, say μ^i, which gives the following model for the return:

$$\frac{dX^i_t}{X^i_t} = \mu^i dt \tag{1}$$

However, share prices are of course never deterministic and normally we model the nondeterministic part by a stochastic process like Brownian motion:

$$\frac{dX^i_t}{X^i_t} \quad \mu^i dt + \sigma^i dW_t \tag{2}$$

(A simple way to realize a Brownian motion is to flip a coin and record 1 for heads and –1 for tails and take the cumulative sum). Equation 2 defines geometric Brownian motion and constitutes a naive model of a share price used in option pricing. Whether or not this is a useful model depends on how realistic the underlying assumptions are, and clearly the only way to test these is to fit the model to share price data and test the assumptions. Therefore we need to solve the inverse problem of estimating the model parameters, namely the mean return μ^i per unit time and the volatility σ^i, for $i = 1, \ldots,$ 100 which is a problem of statistical estimation. Figure 1 shows some realizations of geometric Brownian motion for an imaginary share. The broken curve represents the long-term trend given by $e^{\mu t}$. The shaded region marks the envelope between the 2.5th and 97.5th percentile as a function of time; thus for a given time there is a 95% probability that the share price will be between the lower and upper limits. For this example the inverse problem for an individual share consists of estimating the mean return μ and the volatility σ from *one* realization (share price history). It is clear that simply providing a single estimate for each parameter is unhelpful, because each realization will give a different estimate, possibly some distance from the true value. The best we can hope

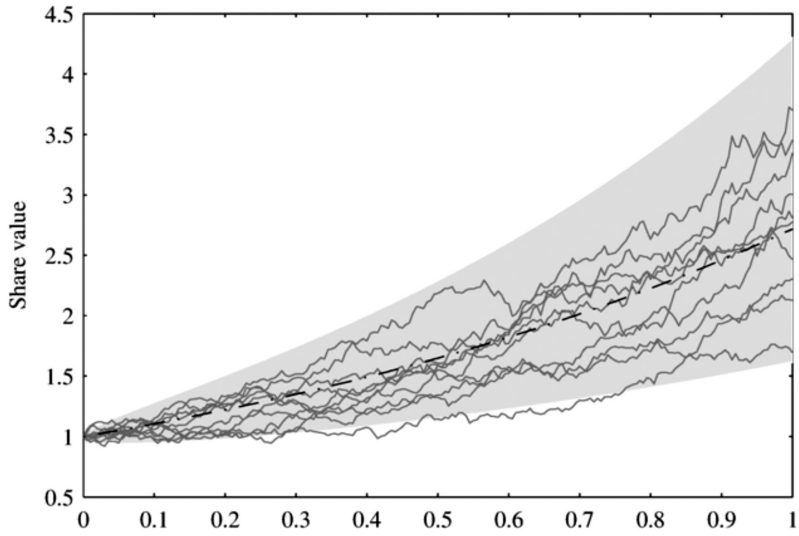

Figure 1 – Ten realizations of geometric Brownian motion. The broken line shows the mean return. Here the envelope of uncertainty is the region between the 2.5 and 97.5th percentile of returns as a function of time.

for is that the uncertainty estimate *includes* the true value. This example is complicated by the fact that share prices are likely to have some correlation, so the inverse problem for the portfolio is more challenging since we need to simultaneously estimate 5,050 parameters (taking into account all the possible correlations between different shares) to estimate the volatility of the portfolio. We further remark that, for a given share history, a poor estimate of return, for example, creates an offset which will not be well modeled by the random part of the model in Eq. (1), which will be exposed by the statistical diagnostics.

For an investor, quantified uncertainty is important because uncertainty is risk, and an experienced investor will make risk- based decisions. The investor will have ideas about risk and reward. Clearly, share price movements are not random in the sense of flipping coins, and randomness is used as a proxy for ignorance. The relevant observation is that it turns out to be a very good assumption: we can make useful models that fit well to data, and we can test the underlying modelling assumptions. In the same manner we can build uncertainty into mathematical models of physical processes, typically evolution equations. As we have seen models are generally approximations to the true physical processes, so there is always a degree of model uncertainty, which we may model in turn by stochastic processes. Furthermore, measurement data always contains errors. Again it is often reasonable to model measurement error as a stochastic process. These may be regarded as ansatz which may or may not prove useful for a particular problem. Part of the modelling consists in testing the assumptions, and establishing that for a given problem they are reasonable.

Why is this interesting? It is plain that models are interesting precisely because of their predictive power. However, even a small model or measurement error can result in large and unacceptable predictive uncertainty. Large uncertainty can, of course, be a fact of life: its sets a limitation

on what we can deduce about a process relative to model and data. Small (predictive) uncertainty is clearly preferable: we would like to accurately quantify groundwater/ surface water interaction or the motion of a contaminant plume, for example. The better we can predict, the easier it is to manage and achieve cooperation with stakeholders. At the extreme end of the uncertainty scale, a large envelope of uncertainty with a substantial downside is clearly undesirable because we probably need to impose more precautionary management. Furthermore, cooperation with stakeholders can prove more challenging. However, large uncertainties can sometimes be controlled; unacceptable uncertainties with unacceptable consequences can be reduced by investigating the measurement process and asking what improvement in data acquisition and assimilation or model is required to reduce the uncertainties to an acceptable level for effective management without adverse economic consequences, for example. We remark that an over optimistic (i.e., small) assessment of uncertainty is clearly a bad thing; models should be ground-truthed.

From qualitive to quantitative uncertainty

Earthquake forecasting and Monte Carlo simulation

One approach to qualitative (not quantitative) uncertainty and model fitting is to vary model parameters, a procedure often referred to as sensitivity testing. This is justified by the assumption that giving a range of parameter estimates goes some way to quantify uncertainty. This is often done using Monte Carlo simulation. As an example we consider the United States Geological Survey earthquake forecast in the San Francisco Bay Area (USGS, 1999), and Freedman and Stark's (2003) analysis of the methodology.

This forecast was constructed in two stages. In the first stage the USGS scientists built a collection of 2,000 models, each of which was used to predict the regional rate of tectonic deformation. This was done by constructing a probability distribution using data and expert opinion. Freedman and Stark pointed out that they had problems understanding the details but reckoned that the models differed in the geometry and dimensions of the fault segments, the fraction of slip released aseismically on each fault segment, the relative frequencies with which different combinations of fault segments rupture together, the relationship between fault area and earthquake size, and so forth. The models were then selected by Monte Carlo simulation; if the predicted deformation was not close enough to the measured rate of deformation it was rejected.

In the next stage three generic stochastic models for fault segment ruptures were created using the long-term recurrence rates from the first stage to estimate the model parameters. These stochastic models were then used to estimate the probability that there will be at least one earthquake of magnitude 6.7 or greater by 2030. This is shown schematically in Figure 2.

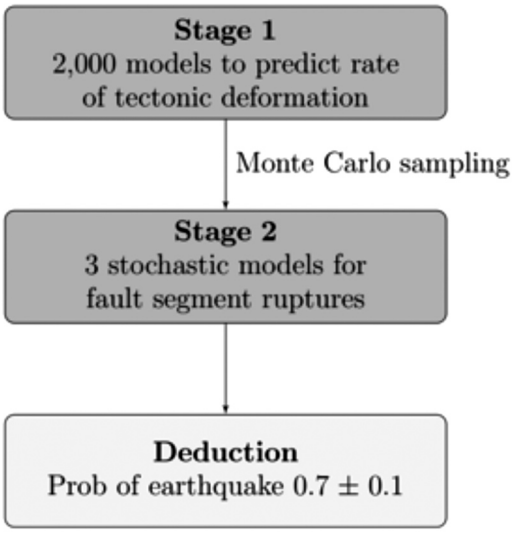

Figure 2 – Flow diagram showing main stages of USGS forecast.

Earthquake magnitude is related empirically to the area of a fault. Particularly noteworthy in the USGS analysis is the Gaussian modelling of unknown quantities such as fault lengths and depths, as well as slip rates, where parameters are given on the basis of expert opinion. These probability distributions were then sampled using Monte Carlo, the resulting models were then rejected if a regional slip constraint was not satisfied. Rather than discuss in detail Freedman and Starks' dissection of the USGS model, we simply note that they observe that 'many steps involve models that are largely untestable; modeling choices often seemed arbitrary. The USGS forecast is 0.7 ± 0.1, where 0.1 is an uncertainty estimate. They go onto say that 'by a process we do not understand, those uncertainties (estimated in stage 1) were propagated through stage 2 to estimate the uncertainty of the estimated probability of a large earthquake. If this view is correct, 0.1 is a gross underestimate of the uncertainty'. They then list ten sources of error which had been overlooked.

This approach to model building and uncertainty analysis lacks scientific rigor and can be both uninformative and unhelpful. It is a procedure that is comparable to a statistician's exploratory analysis of data, useful as an initial analysis at most. In particular, we emphasise that while over-confident estimates of uncertainty (i.e., unrealistically small) as obtained in the earthquake example are appealing, they can be extremely misleading.

We introduce the expression *envelope of uncertainty* to describe the quantified range of uncertainty for any particular quantity of interest. This could relate to the uncertainty of a probability, as in the USGS earthquake forecast, model parameter uncertainty, or most importantly the envelope of predictive uncertainty. Rather than over-confident estimates of uncertainty, it is clearly much better to accept that the envelope of uncertainty may indeed be very large.

In our opinion the USGS example is generic; uncertainty assessment is often carried out by randomising unknown quantities. Unless the unknown quantities can be sensibly estimated from data or the uncertainty in the unknown quantity itself modelled by a stochastic model, Monte Carlo sampling can give extremely misleading conclusions. It is very easy to get lost in number: just because one has generated several thousand samples in no way entails we have a better grip on the problem of interest.

Assessment of liquefaction

Of great contemporary relevance in New Zealand is the engineering assessment of liquefaction. Here we examine the 'simplified approach' to liquefaction assessment that is carried out in consulting. For further details see Seed and Idris (2006) and Idris and Boulanger (2006). Stress is transmitted through a saturated medium through the effective stress (transmitted through the granular particles of the solid matrix), and the pore water pressure that makes up the saturated part. Cyclic loading induced by an earthquake causes the pore water pressure to increase and a consequent reduction in effective stress (and hence strength of the solid matrix). As a result, the matrix loses strength and this may lead to liquefaction as the effective stress tends to zero. This is shown in Figure 3.

Figure 4 shows the methodology that makes up the simplified method. In brief, if the soil's propensity to withstand liquefaction is less than the earthquake loading then liquefaction may occur.

To define the cyclic shear stress ratio at a depth z, the ratio of the total vertical stress σ_{vo} and the effective vertical stress σ'_{vo} is scaled by the maximum horizontal acceleration at the ground surface $a_{\max}$, where g is the acceleration due to gravity, and r_d is a stress reduction coefficient. The

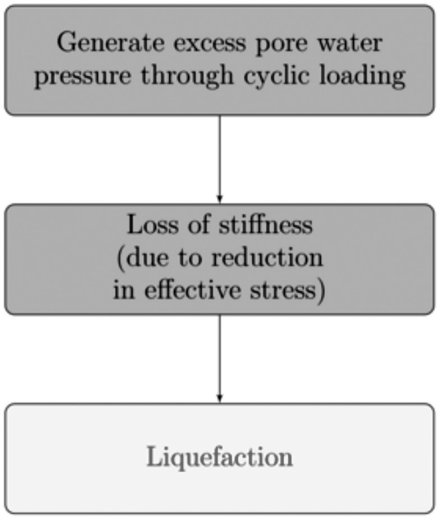

Figure 3 – Flow diagram showing liquefaction process.

0.65 'is used to convert the peak cyclic stress ratio to a cyclic stress ratio that is representative of the most significant cycles over the full duration of loading' (Idris and Boulanger, 2006). The cyclic resistance ratio estimates the soils propensity to resist liquefaction and is generally determined using in-situ index tests (e.g., standard or cone penetration tests); empirical formulae relate tip resistance to the cyclic resistance ratio (CRR) for an earthquake of magnitude 7.5. This is then scaled by a magnitude scaling factor *MSF*, and a further overburden correction factor K_σ.

Note that the cyclic stress ratio (CSR) is given in terms of the peak ground acceleration, whereas one might expect frequency content to play a major role, since laboratory experiments show that the pore

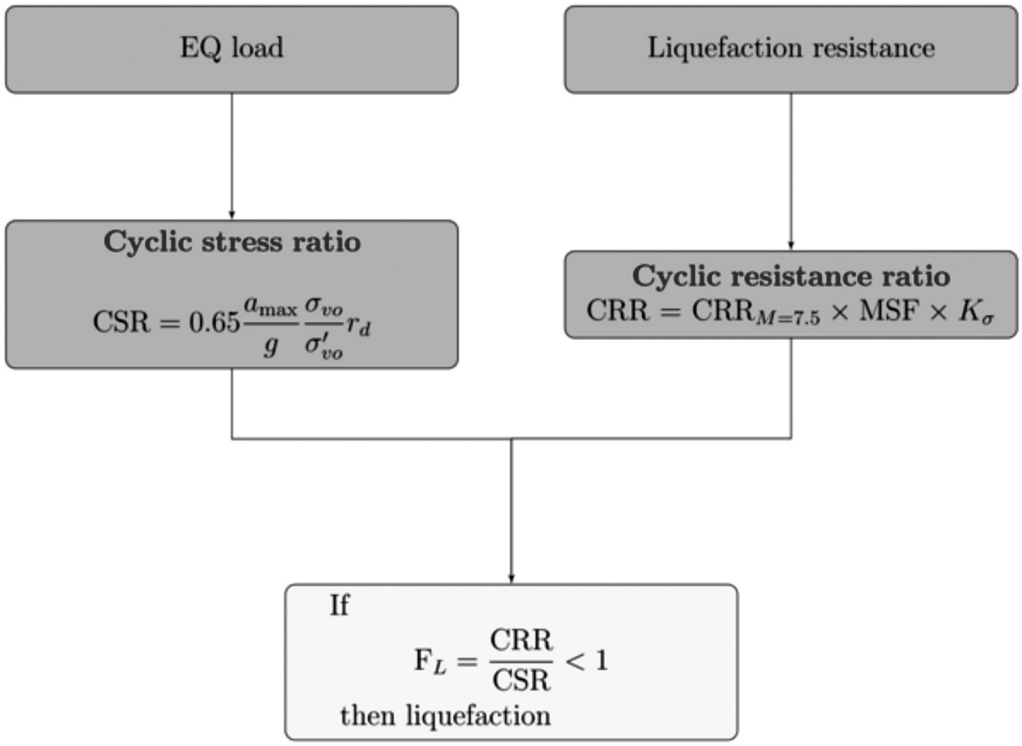

Figure 4 – Flow diagram showing main stages of liquefaction assessment.

water pressure is cranked up following cyclic shaking. In the simplified procedure, an irregular time history (which would happen in practice) is converted to an equivalent number of uniform cycles (which is much easier to simulate in a laboratory), which in turn is related to earthquake magnitude.

Compounding inexactitudes

Even if one accepts the procedure described above, there are numerous obvious uncertainties with the scalings:

- 0.65
- Stress reduction factor, r_d (many empirical formulae)
- Magnitude scaling factor, MSF
- Effects of overburden stress K_σ

Furthermore, if the cyclic resistance ratio is estimated from laboratory tests, there are five further adjustments:

$$\mathrm{CRR}_{\mathrm{field}} = C_1 C_2 C_3 C_4 C_5 \mathrm{CRR}_{\mathrm{triaxial}}$$

There are further corrections for soil type, sloping ground, etc. Finally, we have not considered the all-important interpretation of data, and the empirical relationships that relate earthquake magnitude to number of cycles.

The simplified method gives a single prediction of liquefaction risk for a given earthquake. Is it a reliable prediction – i.e., does it correspond to what actually happens for the same magnitude earthquake? Based on the compounding inexactitudes that make up the cyclic resistance and cyclic stress ratios, it would be exceedingly hard to justify the reliability. Common sense suggests even with the best methods and field tests there are going to be significant uncertainties and, based on the semi-empirical simplified method, these are likely to result in very large predictive uncertainty.

The assessment of liquefaction risk is clearly a very challenging problem, and the simplified approach is an attempt to encapsulate the essential physical mechanisms while retaining some practicability. However, in our view, the method has a very large envelope of uncertainty, and it is surely inappropriate not to recognise this in any engineering assessment.

An approach to uncertainty assessment that is often carried out is to run different models which generally show a range of predictions. Again this is qualitative uncertainty, because no meaning is given to the range. All such assessments show is that the methods are inaccurate, and little confidence can be placed in the assessments.

In our view, the simplified procedure of liquefaction assessment is an extreme example of a style of risk assessment in which semi-empirical relationships provide the intellectual backbone that is fairly common in engineering practice. The problem is that one tends to get lost in a labyrinth of detail, with a concomitant loss of clarity. Of course, if we could improve the modelling of the physical mechanisms and capture the relevant processes and scales in a mathematical model, *and* interpret test data somewhat more precisely (or figure out what test data is necessary to make a sufficiently accurate assessment), we could develop a more quantitative basis for liquefaction assessment.

Lloyd's underwriting

Another example that is pertinent in a general discussion of uncertainty is the underwriting of catastrophic risks such as earthquakes and floods. Typically a Lloyd's underwriter will assign probabilities to risks and think in terms of 'return periods', e.g., a 1-in-100 year return period for an earthquake. These are based on collective experience, allowing a classification into 1-in-10 year, 1-in-20 year, 1-in-50 year, . . ., events. There is generally enough data on market losses to have a fair degree of confidence in the lower return periods, although the large return periods (1-in-200, 300 etc.) are more qualitative statements summing up the collective ignorance about

events rather than the collective state of knowledge.

In the same way, hydrologists assign probabilities to rare flood events and speak of a 1-in-100 year flood event, for example. It is hard to validate small probabilities using data. What is important for the underwriter of a portfolio of catastrophic risks and the hydrologist designing flood protection measures is that assigning return periods to rare risks provides a *consistent* framework for management of risks. What should be avoided is over-interpreting a return period (in particular, supposing it causal); it is simply an informal expression of the law of averages. Therefore we can think of 'designer' earthquakes or floods, for which the attached probabilities are a combination of historical evidence and opinion.

Discussion

Assigning probabilities to events may be appealing on the grounds that it can give reassurance that we know more than we do. However, far more important than the point estimate of the probability is the envelope of uncertainty, because that will (or should) determine any decision we make. In cases in which it is not possible to sensibly model the uncertainties, the envelope of uncertainty will frequently be much larger than we would like. However, that is reality, and it is plainly wiser to accept this rather than cover the uncertainty in a cloak of obfuscation. There are of course many instances in which qualitative uncertainty assessment is the most that can be done, in which case it is always best to appeal to simplicity, and take particular care with communication. In the remainder of this paper we consider examples in which uncertainties are explicitly and meaningful modelled as probability distributions.

Least squares and the related estimates for uncertainty

The least-squares approach to estimate parameters and compute model predictions is probably the most commonly used technique. It sometimes works, and sometimes does not work, as we will see in the following simple examples.

In this section, we consider the archetypical least-squares estimation problem: linear (affine) model fitting and the related assessment of the estimation error. In particular, we will show when the least-squares estimate can turn out to be poor. The particular point here is to show how model reduction (i.e., using a simpler model) and uncertainties can destroy the reliability of both the parameter estimates and model predictions.

As noted above, the aim of assimilating (fitting) a model may be either parameter estimation or prediction based on the model. In the following, we address both topics.

In Figure 5, a typical physics lab situation is shown in which a set of points $(y_k\ ,\ t_k)$, $k = 1, \ldots 20$, seem to fit a straight line model. Here, the correct model for the observations is

$$y(t) = \alpha_0 + \alpha_1 t + e(t) \qquad (3)$$

and the actual values are $\alpha_0 = 1$ and $\alpha_1 = 0.5$. Furthermore, here we use normal independent identically distributed noise $e \sim N\ (0,\ \sigma_e^2 I\ )$ with $\sigma_e = 0.02$ to generate the data. In this case, when we fit an affine model (straight line) to the data using the least-squares approach, we are using the *actual correct model* (i.e., the one that generated the data) but we do not assume an explicit model for the distribution of the error e when computing the least-squares fit.

We write equation (3) in vector form

$$y = Ax + e \in \Re^{20}$$

where $x = (\alpha_0,\ \alpha_1)$ is the vector of the unknown coefficients, and obtain the least

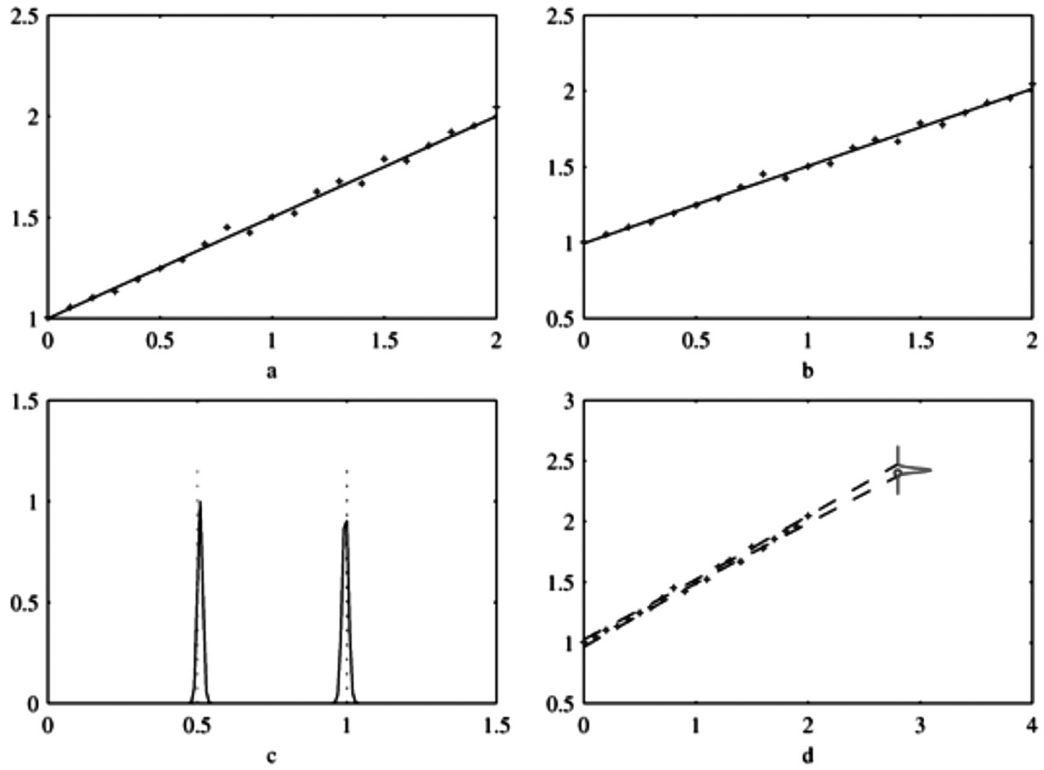

Figure 5 – a) A straight line and noisy observations, b) least-squares fit, c) least-squares parameter error estimates and the actual intercept and slope parameters ($\alpha_0 = 1$ and $\alpha_1 = 0.5$), d) actual value (circle) and predicted envelope of uncertainty (three standard deviations from mean predicted value), at $t = 2.8$.

squares estimate

$$x_{\mathrm{LS}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}y$$

which is a (vector-valued) random variable since it is a function of y which contains the noise e. The estimate gives us the residuals

$$r = y - Ax_{\mathrm{LS}}$$

The fitted straight line is shown in Fig. 5b and it seems to describe the data well.

To obtain an uncertainty estimate using the least-squares approach, we now assume that the original errors e are independent identically distributed (homoscedastic). This suggests estimating the covariance of x_{LS} by

$$\Gamma^{x} = \hat{\sigma}^{2}_{r}\left(A^{\mathrm{T}}A\right)^{-1}$$

where

$$\hat{\sigma}^{2}_{r} = \frac{1}{N-p}\sum_{k=1}^{N} r_k^2$$

and where $p = 2$ (in this case) is the number of estimated parameters. The uncertainty in the parameter estimates can now be estimated (assuming normal residuals, which can, of course, be tested) from

$$x \sim N(x_{\mathrm{LS}}, \Gamma_x)$$

The marginal distributions, together with the actual parameters are shown in Fig. 5c. The uncertainty estimates are clearly feasible in the sense that the actual values are captured by the indicated uncertainty envelope.

The predictions given by the model are of the form

$$\hat{y}(t) = \alpha_0 + \alpha_1 t$$

Since the parameters $x = (\alpha_0, \alpha_1)$ are not known but estimated random variables, we obtain for the prediction

$$\hat{y}(t) = \alpha_0 + \alpha_1 t = V x_{\mathrm{LS}}$$

where $V = [1\ t]$, and further for the variance

$$\mathrm{var}\, y(t) = V \Gamma_x V^{\mathrm{T}}$$

The observations, the least-squares estimate for $y(t)$, and the envelope of predictive uncertainty ($\pm 3\sigma_{y(t)}$ from mean predicted value $\hat{y}(t)$ – which depend on t) are shown in Fig. 5d. Furthermore, we also show the predictions of the model beyond the extent of the observations up to $t = 2.8$ (original observations covered $t \in (0, 2)$). Again, the uncertainty estimates are clearly good in the sense that the actual values are supported by (contained in) the indicated uncertainty envelope, in particular, the estimate for $\hat{y}(t = 2.8)$.

Next, consider the data shown in Fig. 6. The actual data has been generated by a third-order polynomial

$$y(t) = \alpha_0 + \sum_{k=1}^{3} \alpha_k t^k + e(t)$$

where (α_0, α_1) are as above, $\alpha_2 = -0.22$ and $\alpha_3 = 0.06$, and the noise statistics are as above. Considering the data only (the '+' in plot (a)) suggests using an affine model, as with the data in Figure 5. We therefore estimate the affine model exactly as above, the only difference being the new values taken for $y(t_k)$.

Again, judging by the fit shown in Figure 6b, we appear to be doing fine. The norm (variance) of the residuals is slightly larger than in the first example, but we cannot employ this fact in any way: in the least-squares approach, we do not assume a model for the error statistics. As for the parameter estimates, the intercept estimate $\hat{\alpha}_0$ is *almost* reasonable in the sense that the estimated value $\hat{\alpha}_0 = 1.042$ is near the correct value $\alpha_0 = 1$, but the error estimate $\sigma_{\alpha 0} = 0.011$ is so small that the actual value is around 4 standard deviations from the estimated value:

$$|\alpha_0 - \hat{\alpha}| \approx 4\sigma_{\alpha 0}\,.$$

Much worse, for the slope we have $\alpha_1 = 0.5$, $\hat{\alpha}_1 = 0.282$, and $\sigma_{\alpha_1} = 0.009$, so that the actual slope is around 30 standard deviations from the estimated value:

$$|\alpha_1 - \hat{\alpha}_1| \approx 30\sigma_{\alpha 1}$$

which is very poor (see Fig. 6c). However, the prediction $\hat{y}(t)$ and the indicated uncertainty intervals shown in Fig. 6d are feasible over the *interpolation interval* $t \in (0, 2)$, but *not* beyond the interval $t \in (0, 2)$. For example, the true value at $t = 2.8$ is 9 standard deviations from the estimated value:

$$|y(2.8) - \hat{y}(2.8)| \approx 9\sigma_{y(2.8)}$$

Such behaviour is typical for parameter estimation problems when approximate models are used. The gravity of the misinterpretation of the results is variable and depends on whether we consider the parameter estimates or the predictions, and whether the predictions are de facto interpolations (noise removal) or extrapolations. The fitting of a line is numerically a stable problem. The assimilation of well measurements, for example, into a groundwater model, is an unstable inverse problem, and the related feasibility problems can be significantly worse. For example, fitting the model to pump test data, would possibly give feasible predictions at the *same* well in another pump test that is carried out *in exactly the same way*. But predicting the extraction behaviour of the aquifer (non-interpolative prediction) is potentially subject to a similar estimate infeasibility as the above (numerically stable) polynomial fitting example. The model reduction in the above classroom line-fitting example was by no means necessary, since the data could, of course, have been fitted by a third-order polynomial too. However, for

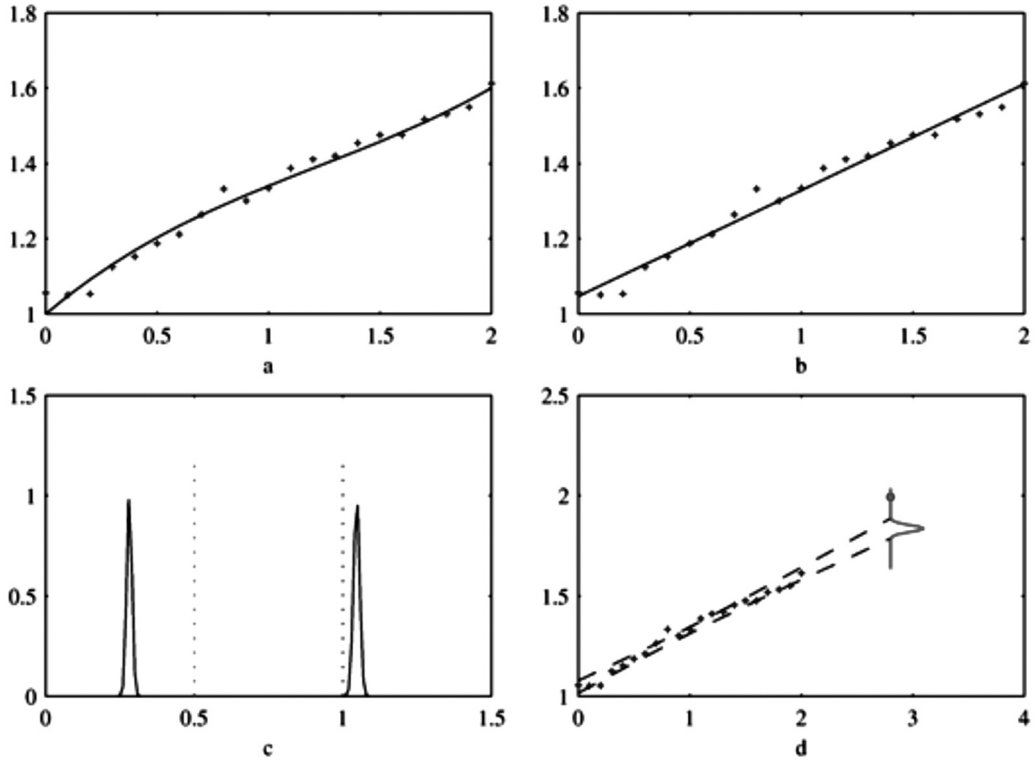

Figure 6 – a) A third-order polynomial and noisy observations, b) least-squares fit of a first-order polynomial, c) least-squares error estimates and the actual intercept and slope parameters ($\alpha_0 = 1$ and $\alpha_1 = 0.5$), d) actual value (circle) and predicted uncertainty at $t = 2.8$.

data assimilation when we encounter models with distributed (possibly spatially varying) parameters, such as groundwater models, model reduction is an unavoidable task.

Good and poor uncertainty quantification

For clarity we draw a distinction between 'good' and 'poor' uncertainty quantification. The first example above is *good* uncertainty quantification, since the envelopes of parameter uncertainty and predictive uncertainty capture the true parameter and predictive values respectively. The second example is *poor* uncertainty quantification since the envelopes of uncertainty neither capture the true parameter values nor, more importantly, the true predictive values. Therefore the predictive reliability of the second model is poor.

We note that minimisation of (posterior) uncertainty is not a primary objective. Good uncertainty quantification in the sense defined above is clearly preferable to poor uncertainty quantification. It can turn out, of course, that uncertainty estimates can have little value in decision making because the envelope of uncertainty is too large to be of practical value. However such estimates *are by definition not misleading* since they include the true values, i.e., what actually happens. If this is the case, the correct interpretation is simply that more data is needed to reduce, and hence *control* the uncertainty.

In a later section on the least-squares problem, we revisit this problem of fitting an affine model to the above noisy third-order polynomial data. We build an explicit model for the uncertainty in the model and take

this into account. This leads to a model that may yield significantly higher uncertainty estimates, but are good in the sense that the true values are captured in the envelope of uncertainty.

Uncertainty quantification in the Bayesian framework

Our discussion now focuses exclusively on the Bayesian framework in which all uncertainties, including those that are related to the models themselves, are explicitly modelled as probability distributions. Easiest to consider is measurement noise, which is often modelled as correlated or uncorrelated Gaussian white noise, while more challenging uncertainties like domain boundaries, sources or sinks and other model uncertainties can be incorporated into the Bayesian framework (see below). Why is this important? Very simply, measurement and model uncertainty (even very small) can lead to significant parameter uncertainty, which in turn can lead to significant predictive uncertainty. We have seen this in the case of the least-squares examples above. Traditional modelling is inherently deterministic; model parameters are tuned in order to give a forward mapping that approximates the process under consideration. (Furthermore parameters are often lumped values, determined from experiments or published values.) The problem is that this often results in poor predictive reliability. Since we can never reasonably expect to predict exactly how a process will evolve, the best we can do is to quantify the 'envelope of uncertainty', and hope that what actually happens is captured in the envelope. Ultimately we can only test the efficacy of a model by using synthetic examples, or calibrating to a subset of field or test data and testing the predictive reliability of the model. In the inverse problems literature, it is a standard procedure to begin an investigation with synthetic data so that the model can be ground-truthed. Calibrating models in the first instance with real data should, in general, be avoided.

Bayes formula

To formulate an inverse problem in the Bayesian framework, we use Bayes' formula, which is a simple algorithm for updating the probabilities of events, in our case the probabilities of model parameters relative to data. Somewhat figuratively, we speak of the *prior uncertainty* of parameters, in which we posit a stochastic model of the parameter uncertainty *a priori*, i.e. prior to any measurements or experiments. The prior model is therefore completely *un*-informed by data. Measurement and model uncertainties are then built into the likelihood, which gives the relative probability of observing the data for any given set of model parameters. Bayes' formula then links the *posterior probability* of parameters relative to data as shown in the flow diagram in Figure 7.

For inverse problems, the stochastic modelling of uncertainties can be viewed as part of a larger model building process. The end game is the posterior distribution of parameters, which measures the probability of parameters relative to data.

The solution to the inverse problem is then given as summary statistics over the posterior distribution, which necessarily includes a measure of the uncertainty. A typical estimator of a parameter would be the mean (average) of a parameter or the mode or *maximum a posteriori* estimate. A measure of the uncertainty would be the standard deviation or, in Bayesian vocabulary,

Figure 7 – Bayes' formula

credibility intervals and their multi-dimensional counterparts. Even for models with few parameters, one cannot in general compute these quantities analytically, usually the posterior distribution is 'explored' using a random sampling technique such as Markov Chain Monte Carlo.

It is helpful to draw distinctions and speak of low- and high-dimensional inverse problems. In a low-dimensional inverse problem, the model may have up to 10-15 parameters, while a high-dimensional problem may have several thousand unknown parameters (a geothermal field model may have around 10,000 parameters). In the low dimensional case the domain is generally homogenised so that bulk parameters describe quantities like storativity or transmissivity in a groundwater setting, while in the high case parameters are assumed to be spatially (and sometimes temporally) distributed. Not surprisingly, they present different orders of computational challenge, and much of the current research effort in inverse problems is focused on reducing the computational bottlenecks that characterise large-scale problems. We emphasise that quantifying uncertainty is challenging at the best of times, and many problems are rendered infeasible by 'direct attack' because it takes too long to compute the required solution. Both classes have their uses; simple problems are clearly appropriate when measurements are very sparse, and in combination with dimensional analysis can be used, for example, to study the efficacy of a new procedure before lengthy and expensive experiments are carried out. One of the clear merits of the Bayesian framework and the explicit modelling of uncertainties is the scope for establishing *uncertainty reduction*. For example, we might ask to what extent measurements of heat exchange between surface water and groundwater reduce the uncertainty of groundwater/surface water interaction.

Statistical inverse problems

We give a more formal derivation of the Bayesian approach to inverse problems in this subsection. Further details can be found in Kaipio and Somersalo (2007), Tarantola (2004) and Calvetti and Somersalo (2007). In particular, we detail the model building process, which is shown in the flow diagram in Figure 8.The first task is to build a model of the physical process. Usually this is expressed as a partial differential equation, together with the appropriate boundary and initial values. In most cases the model is solved numerically using a finite difference, volume or element scheme.[2] Examples might be models of the movement of ground or surface water and temperature, or models of seismic wave propagation. The latter could be either active earthquakes (high frequency signals) or background passive seismic waves (low frequency signals) caused by ocean waves. The unknown parameters in these models range from hydraulic quantities like fluxes and heat diffusivity, to elastic properties of the media. We denote model parameters by x and measurements by y. The next task is to build stochastic models of measurement and model errors. Measurement error, of course, is somewhat more straightforward. In general, we assume that the measurements are corrupted by 'noise', the simplest model being an additive noise model

$$y = \bar{y} + n \tag{4}$$

where $\bar{y}$ is the unobserved 'true' signal, and n denotes Gaussian white noise. Thus the

[2] We note that we often also need the "operator" forms of the numerical solvers, especially when uncertainty models are incorporated. Thus software that simply computes the solution, given all parameters, is not always adequate.

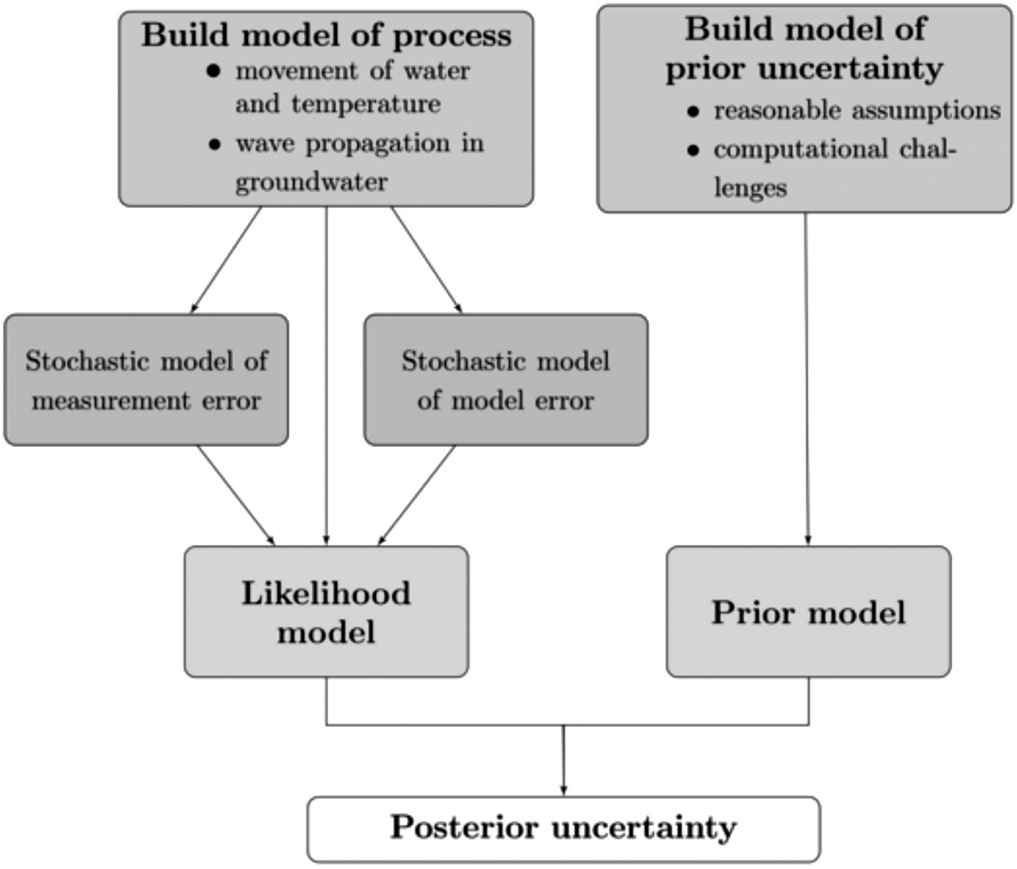

Figure 8 – Flow diagram showing main stages of model-building process.

(observed) corrupted signal is manufactured from the true signal perturbed by uncorrelated random samples from a normal distribution with mean 0 and standard deviation σ_n. We note that while this is a common assumption, measurements may, in practice, be correlated (or the Gaussian may be inappropriate), in which case it is necessary to build this into the model for measurement error.

Dealing with model uncertainty is much more challenging. A simple way to deal with this is to lump the model uncertainty into the additive noise model for measurement uncertainty and assume a stochastic model for the standard deviation. If there is a systematic bias, another approach would be to take a statistical modelling approach and extract the bias. A third way is to use the *Bayesian approximation error approach* described in the next section, so that unmodelled structure is explicitly accounted for. The mapping f between the model parameters x and noise-free observations $\bar{y}$ is called the *forward mapping* between the parameter

$$f : x \mapsto \bar{y} \tag{5}$$

space and the noise-free observations. Assuming, for simplicity, the additive noise model Eq. (4) and that x and n are mutually independent, and ignoring model uncertainty gives the *likelihood*

$$x(y \mid x) = \pi_n(y - f(x)) \tag{6}$$

where π_n is the probability distribution of measurement noise n in Eq. (4). This measures the relative probability of observing

data y, given any choice of parameters x. The likelihood can be thought of as an offset function over parameters between data and model predictions.

Next, we build a stochastic model of the prior probability distribution $\pi(x)$ of model parameters. For low-dimensional problems this is generally quite easy and it is sufficient to assume a 'flat prior' (i.e., equiprobability over the feasible range of parameters) or a normal distribution with a sufficiently large standard deviation. For large-scale problems, considerably more care needs to be taken in designing an appropriate prior; often this is a compromise between computational efficiency and physical reasonableness. For examples of small- and large-scale problems in a groundwater setting see Cui and Ward (2013) and Cui *et al.* (2013) respectively.

The *posterior distribution* $\pi(x \mid y)$ is the object of interest, and measures the relative probability of parameters given data. Bayes formula links the posterior distribution to the likelihood and prior distribution:

$$\pi(x \mid y) \propto \pi(y \mid x)\pi(x) \tag{7}$$

which is a formal expression of the updating formula given in the flow diagram above. As discussed in the previous section, a solution to the inverse problem is given as summary statistics over the posterior distribution, which includes an estimate of the uncertainty.

Modelling errors and the Bayesian approximation error approach

Here, we give a brief review on the Bayesian approximation error approach. We list typical modelling errors and approximations that occur when we consider real world cases, and cite work in which the approach has been successfully employed to recover from these errors and uncertainties.

Distributed parameter estimation problems induced by partial differential equations and the related initial-boundary value problems constitute perhaps the largest class of inverse problems. Such models always provide a more-or-less simplified approximation for the physical reality. In particular, the following modelling problems are common:

- Geometry is unknown (biomedical imaging, geophysics),
- Measurement sensor locations are only approximately known (geophysics, some biomedical imaging),
- Measurement noise statistics are poorly known (most applications),
- Approximate physical measurement models are used (biomedical, geophysical and industrial applications),
- Significant uncertainty in models for the unknown variables (geophysics, industrial flows), and
- Boundary conditions are partially unknown (biomedical, geophysics and industrial applications).

The approximation error approach was introduced in Kaipio and Somersalo (2005, 2007) originally to handle pure model reduction errors. For example, in electrical impedance and electrical resistance tomography and deconvolution problems, it was shown that significant model reduction is possible without essentially sacrificing the feasibility of estimates. With electrical impedance tomography, for example, this means that very low-dimensional finite element approximations can be used. Later, the approach has also been applied to handle other kinds of approximation and modelling errors as well as other inverse problems: model reduction, domain truncation and unknown anisotropy structures in diffuse optical tomography were treated in Arridge *et al.* (2006), Kolehmainen *et al.* (2009), Heino and Somersalo (2004) and Heino *et al.* (2005). Missing boundary data in the case of image processing and geophysical

electrical impedance and electrical resistance tomography were considered in Calvetti *et al.* (2006) and Lehikoinen *et al.* (2007), respectively. Furthermore, in Nissinen *et al.* (2008) and Nissinen *et al.* (2009) the problem of recovery from simultaneous domain truncation and model reduction was found to be possible, and in Nissinen *et al.* (2011a, 2011b) the recovery from the errors related to inaccurately known body shape was shown feasible.

The approximation error approach was extended to nonstationary inverse problems in Huttunen and Kaipio (2007b), in which linear nonstationary (heat transfer) problems were considered, and in Huttunen and Kaipio (2007a) ,and Huttunen and Kaipio (2009), in which nonlinear problems and state space identification problems were considered, respectively. The earliest similar but partial treatment in the framework on nonstationary inverse problems was considered in Seppänen *et al.* (2001a), in which the boundary data that is related to stochastic convection diffusion models was partially unknown. A modification in which the approximation error statistics can be updated with accumulating information was proposed in Huttunen *et al.* (2010) and an application to hydrogeophysical monitoring in Lehikoinen *et al.* (2010).

From pure model reduction and unknown parameters or boundary data, a step forwards was recently considered in Tarvainen *et al.* (2010) in which the physical forward model itself was replaced with a (computationally) much simpler model. In their model, the radiative transfer model (Boltzmann transfer equation) which is considered to be the most accurate partial differential equation model for light transfer in (turbid) media, was replaced with a diffusion approximation. It was found that in this kind of a case, the statistical structure of the approximation errors enabled the use of a significantly less complex model, again simultaneously with significant model reduction for the diffusion approximation. But also here, both the absorption and scattering coefficients were estimated simultaneously, while in Kolehmainen *et al.* (2011), the scattering coefficient was modelled as a normal Markov random field and was not estimated.

The approximation error approach relies on the Bayesian framework for inverse problems described in the previous section. In the Bayesian framework, all unknowns are subject to inference (i.e., estimation) simultaneously, which often results in excessively heavy computational loads. Generally, Markov Chain Monte Carlo algorithms have to be used to obtain a representative set of samples from the posterior distribution. Then, after a set of samples has been computed, marginalization over the uninteresting unknowns is straightforward Only in few special but important cases, such as the additive error model, some of the uninteresting unknowns can be eliminated before inference.

The most comprehensive treatment of the Bayesian approximation error approach was given in Kaipio and Kolehmainen (2013), which also deals with numerical and computational considerations that are related to high-dimensional observations. We now give a simplified introduction to the idea behind the approximation error method in which only the model reduction is considered; for more details and a formulation that is relevant with general model uncertainties, we refer to the appendix.

For simplicity we consider an accurate forward model $A(\bar{x})$ with spatially distributed parameters $\bar{x}$:

$$\bar{x} \mapsto \bar{A}(\bar{x})$$

Usually the model is nonlinear in the unknown parameters. We assume an additive noise model e for the relation between measurements and unknown parameters $\bar{x}$:

$$y = \bar{A}(\bar{x}) + e \in \Re^m$$

Next we approximate $\bar{A}(\bar{x})$ by a similar model $A(x)$. In the example above, the accurate model is the cubic while the reduced model is the straight line. Then we can write

$$y = \bar{A}(\bar{x}) + e \tag{8}$$

$$= A(x) + \left(\bar{A}(\bar{x}) - A(\bar{x})\right) + e \tag{9}$$

$$= A(x) + \varepsilon + e \tag{10}$$

where we define the *approximation error* $\varepsilon = \phi(\bar{x}) = \bar{A}(\bar{x}) - A(x)$. Thus, the approximation error is the discrepancy of predictions of the measurements (given the unknowns) when using the accurate model $\bar{A}(\bar{x})$ and the approximate model $A(x)$. The mean of the approximation error can be interpreted as a systematic correction factor when carrying out computations with the approximate model.

Furthermore, the (second-order) statistics of the approximation error convey information about the effects (variability) of the predictions of the forward model under model errors and uncertainties. Using the approximation error one can obtain a modified likelihood model where the bulk of computations are carried out using the approximate model, which are corrected using the mean and covariance of the approximation error. Importantly, we note that the mean and covariance of the approximation error do not depend on data, and can be carried out off-line, whereas the relatively cheaper calculations using the approximate model and data are carried out as on-line computations. We now show this in the simple case of the cubic data example of the section on least squares and the related estimates for uncertainty.

The least-squares problem revisited

In the section on least squares and the related estimates for uncertainty, we considered the classroom example of fitting a straight line (first-order polynomial) to data that visually suggested such a model. The conclusion was that if the model is correct (and the problem is a stable one), the (least-squares) estimates can be feasible. On the other hand, if the model is not correct or is highly approximate, which is typically the case when model reduction is employed, the estimates can be grossly infeasible.

In the following, we use the Bayesian approximation error approach to build an extended likelihood model. Thus, we write

$$y = \bar{A}\bar{x} + e = Ax + e + \varepsilon(x)$$

where the approximation error

$$\varepsilon(x) = \left(A - AP\right)\mathbf{x},\ \mathbf{x} = (\alpha_0, ..., \alpha_3),\ x = (\alpha_0, \alpha_1)$$

and $P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$

As a prior model, we used a normal model for the coefficients $\bar{\alpha}$ with means $\mathrm{E}\alpha = \bar{x}_* =$ (0.7, 0.3, –0.18, 0.055) and variances $\sigma_\alpha =$ (0.1, 0.1, 0.04, 0.005) and modelling the coefficients as mutually independent, we have a diagonal covariance Γ_x. These were chosen so that the actual coefficients are 2-4 standard deviations away from the respective means. We note also that the modelled uncertainty is quite significant, which is expected to yield significant error estimates. But, again, we are after feasibility (i.e., good uncertainty quantification) rather than apparently small uncertainties. Draws from this distribution, together with the actual (errorless) polynomial, are shown in Figure 9.

As noted in the previous section, we calculate the posterior probability distribution of parameters $x = (\alpha_0, \alpha_1)$ for the reduced linear model, using the correction given the approximation error, which is a function of the accurate model, namely the cubic polynomial.

In the case of least-squares estimates, the 'fit' is simply equal to the predictions. In the case of approximation error estimates, this

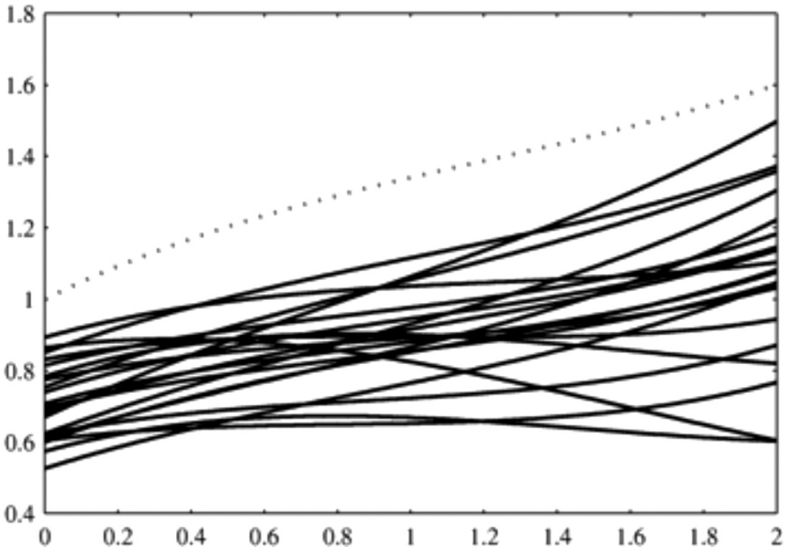

Figure 9 – Draws from an uncertainty model for the polynomials (expecting high uncertainty of the affine model) and the actual polynomial (dotted line). Note that the actual polynomial is nowhere near the modelled mean behaviour.

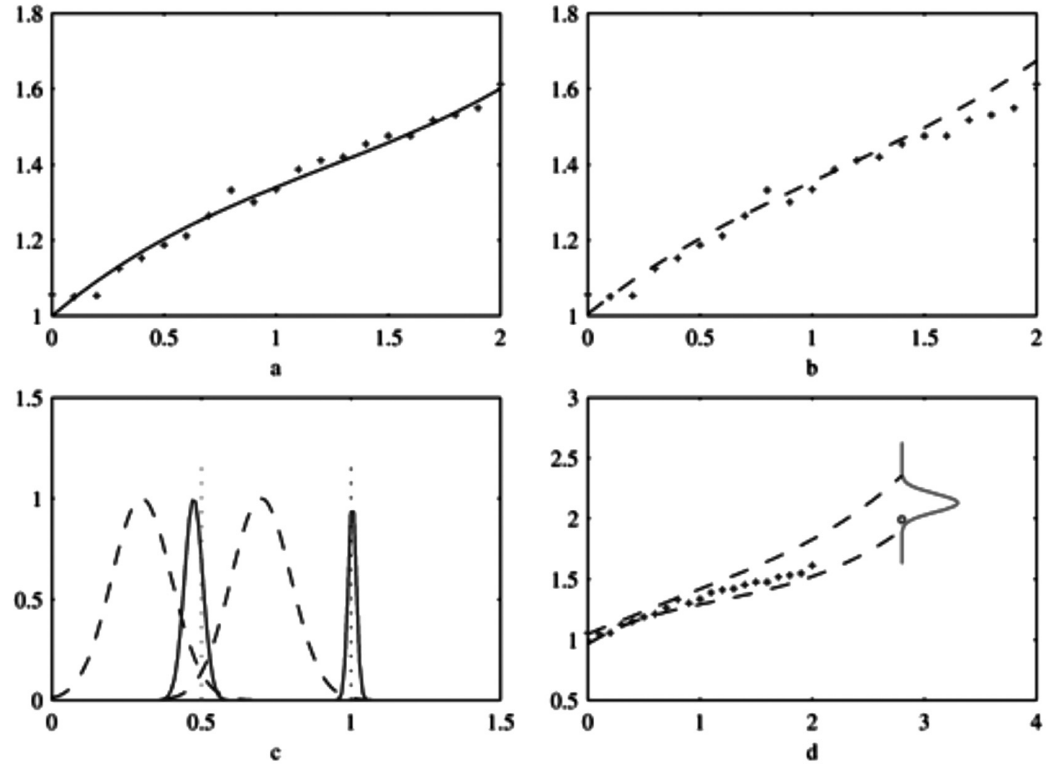

Figure 10 – a) A third-order polynomial and noisy observations, b) predictions given by an approximation error approach, c) the prior (dashed lines) and posterior (solid lines) densities, and the actual intercept and slope parameters ($\alpha_0 = 1$ and $\alpha_1 = 0.5$), d) the actual value (circle) and predicted envelope of uncertainty (three standard deviations from mean predicted value) at $t = 2.8$.

is not necessarily the case, since the mean of the approximation errors does not usually vanish.

The results are shown in Figure 10. As noted above, the 'fit' in Figure 10b is not a straight line. The parameter estimates in Figure 10c are clearly feasible, although the prior probability of the actual values was quite low. This means that the parameter estimates are still 'driven by the data' rather than the prior model. This is an essential requirement when numerical experiments are carried out. Furthermore, the envelopes of parameter uncertainty are clearly larger than the envelopes computed in the least squares section. However, this time *they actually capture the true values*. The same applies to the estimated predictive uncertainty, and we have achieved now good uncertainty quantification.

Examples from biomedical imaging and process tomography

In this section, we consider some previously published examples that illustrate the importance of modelling uncertainties properly. The examples are distributed parameter estimation type inverse problems, in which the primary unknowns are spatially varying coefficients. In all these cases, we consider tomographic data, that is, boundary measurement data only. Furthermore, in all the cases, we employ feasible prior models that exhibit relatively large prior uncertainty. In all cases, the primary unknowns are modelled as normal Markov random fields.

Different prior models are used in different cases, but in the comparison between using a conventional observation model (modelling errors not modelled) and the observation model that is constructed based on modelling and marginalising over the approximation errors, the same prior model is employed. In all cases, the measurement error statistics are estimated by analysing the behavior of the (electrical) measurement system and by conducting a large number of repeated measurements. Since in most real world applications the computational resources can be very limited, we carry out varying approximations in the computational models, such as (finite element) model reduction and approximation of all probability distributions as normal. Together with the unknown boundary conditions, unknown boundary geometry, unmodelled parts of domains (the unknowns that affect the measurements but are not being estimated) and such, all the uncertainties are treated simultaneously.

Local X-ray tomography

In X-ray tomography, beams of high energy photons are directed into the target volume. Depending on the spatially varying absorbtion coefficient, the transmitted beam intensity is attenuated. Irradiating the entire target over an angular aperture of minimum 180°, the attenuation data is usually adequately rich and low-noise to lead to a mildly ill-posed inverse problem. This technology was established in the 1960s and most people have seen X-ray (computerized) tomography images of different cross-sections of the human body.

Let us repeat the main points that lead to mildly ill-posed problems: i) use relatively high intensity beams to provide low relative noise levels; ii) irradiate the entire body over iii) angular aperture of minimum of 180°. All the three topics lead to high radiation exposure and thus X-ray tomography cannot be carried out on a person very often. Imaging of the head and spine are particularly delicate targets. In addition, irradiating the entire body over 180° cannot always be realized, for example, in dental imaging. Furthermore, if one is only interested in the target in a region of interest, it is natural to try to limit the radiation exposure of uninteresting subdomains.

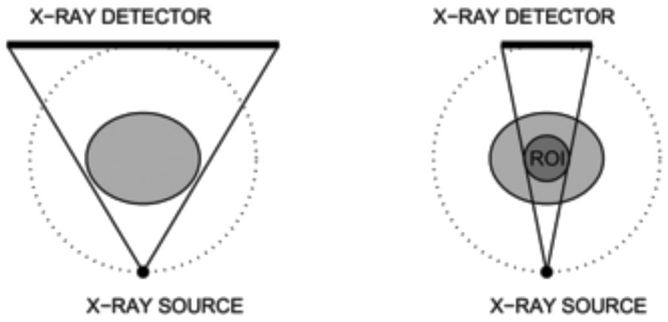

Figure 11 – Global (left) and local (right) tomography. The source and detector panel rotate around the object along the dotted circle. In the right hand image, the dark gray patch represents the region of interest subdomain that is present in all of the projection images in the local tomography case. In local tomography, only the region of interest is to be reconstructed, with a minimal number of projection data.

In such cases, one is dealing with *local tomography* (entire target is not irradiated) and/or *limited-angle* tomography (see Figure 11 for the geometry of local X-ray *tomography*).

In contrast to the full-angle non-local tomography, limited-angle tomography is problematic, and standard approaches can only yield more-or-less meaningless estimates. At best, the estimates (reconstructed cross-sections) are typically only qualitatively feasible. For Bayesian analysis of the limited angel X-ray tomography, see Siltanen *et al.* (2003) and Kolehmainen *et al.* (2003). For the analysis of the local X-ray tomography in the Bayesian framework, see Kaipio and Kolehmainen (2013).

We study a full-angle local dental X-ray imaging case which was originally considered in Kaipio and Kolehmainen (2013). The global tomography estimate which serves as the ground truth is shown in Figure 12, together with the local tomography estimates. From Figure 12, it is clear that the local tomography estimate without a feasible uncertainty model can usually provide at best a hazy qualitative estimate or, as in this case, a completely useless estimate. It is to be noted that the lower middle image in Figure 12 *fits the measurement data within the statistical accuracy of the measurement errors*. On the other hand, when the uncertainties are modelled, the estimates are *quantitatively feasible*, which can be seen in Figure 13, in which related marginal densities of a pixel together with the actual value are shown. As in the section on least squares and the related estimates for uncertainty, the parameter

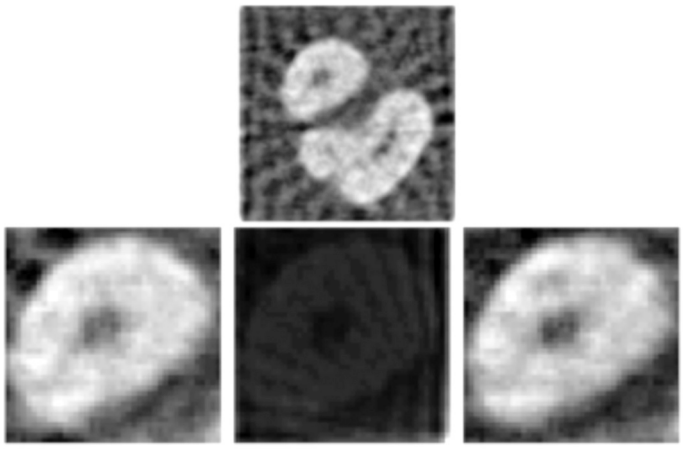

Figure 12 – Estimates from X-ray projection data from a tooth specimen (23 projections from a view angle of 187 degrees: full angle data). *Top:* global tomography, entire target. *Left:* global tomography (zoom to region of interest), *middle:* local tomography, uncertainties are not modelled, and *right:* local tomography, uncertainties are modelled.

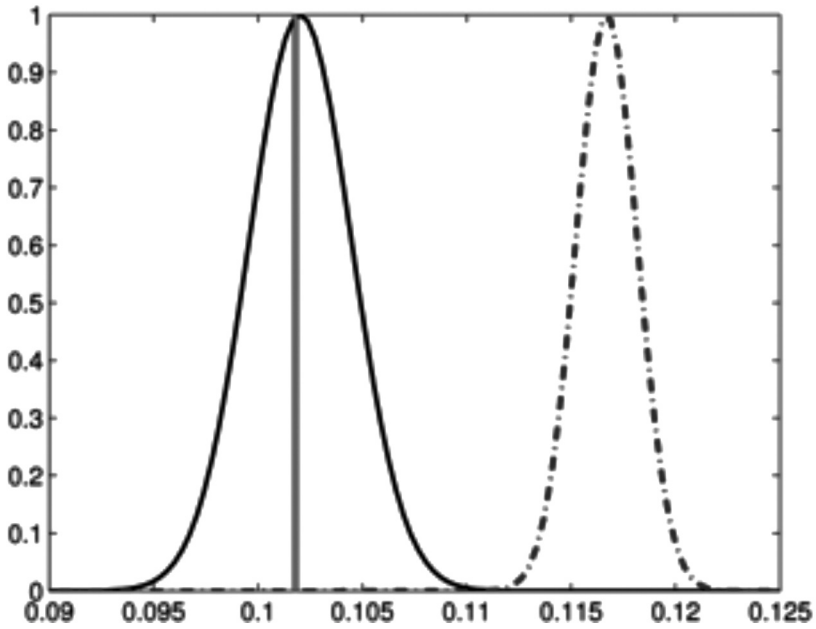

Figure 13 – A posterior marginal density of a pixel (local X-ray tomography) when the approximation and modelling errors have been modelled (solid line), have not been modelled (dash-dot line); and the actual value (vertical line) of a pixel in a local tomography problem.

estimates without modelled approximation errors are infeasible in that the actual value is practically claimed to be impossible, i.e., well outside the envelope of uncertainty. On the other hand, the modelled approximation errors widen and shift the posterior distribution somewhat, so that the envelope of uncertainty increases but now captures the actual value. We note that the actual medical application of X-ray imaging is most commonly used to detect small deviations of the coefficient values from the nominal ones, in particular, related to the detection of tumors. In decision theory, the detection are based on the posterior distribution, if the distribution estimates can be assumed to be reliable (feasible).

Biomedical electrical resistance tomography

Electrical resistance (or impedance) tomography is used in biomedical imaging as a functional modality. The anatomical accuracy of the resistance tomography images is much poorer than with X-ray tomography, but resistance tomography is fast and portable, and could be used in ambulances to detect, for example, lung collapse, pneumothorax or internal haemorrhage. The reason why electrical resistance tomography has not yet been adopted in such ambulatory practice is that it tolerates geometric uncertainty very poorly. In the thoracical imaging case, for example, the mathematical model should be based on the actual outer geometry of the patient's torso. This is clearly an impossible requirement.

But again, the uncertainty in the geometry can be modelled using a probability distribution. A very large number of thorax cross-sections are readily available, and these can be used as a so-called anatomical atlas. This atlas, in turn, can be used to build the stochastic forward mapping. In Nissinen *et al.* (2011a, 2011b), this was carried out; see Figure 14 for draws from the distribution model for the domain boundary.

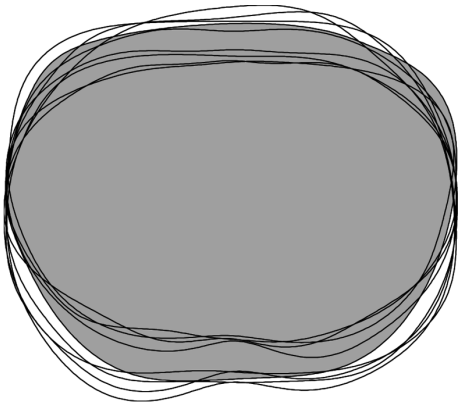

Figure 14 – Samples from an anatomical atlas of a thorax. The actual geometry used in the test is shown as a grey domain.

To test the approach, a test tank with non-circular geometry was constructed, and the ERT (electrical resistance tomography) data was used to estimate the spatial conductivity distribution. The actual target (tank), a reconstruction in the actual domain (this serves as a reference: in real life this geometric data is not available), together with reconstructions in a standard circular geometry are shown in Figure 15. Again, the uncertainty that is related to the boundary geometry can be handled. Since the approximation error model also incorporates severe model reduction, the AE (approximation error) estimate is fast to compute. Furthermore, the AE estimates can be used to provide an estimate for the actual geometry, which is a secondary task here. Nevertheless, the actual geometry, the estimate and approximate error estimates are shown in Figure 16. This shows again that with proper modeling of the primary uncertainty, the error estimates can be feasible.

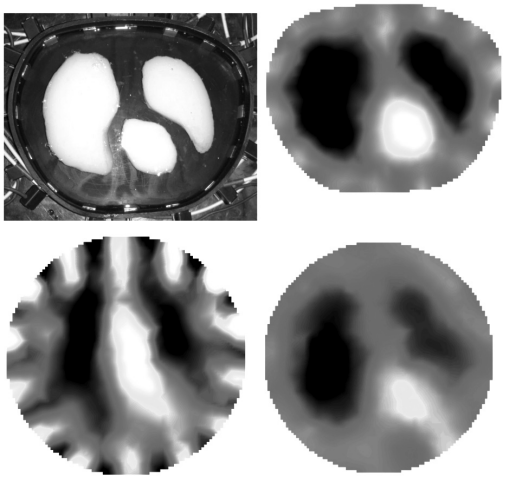

Figure 15 – *Top left:* The thorax model phantom with realistic conductivity characteristics. *Top right:* reconstruction using a conventional measurement error model and the correct domain (this information is unavailable in actual situations). This serves as an idealized reference. *Bottom left:* reconstruction using the conventional measurement error model and conventional circular model domain *Bottom right:* reconstruction using the approximation error model and the circular domain.

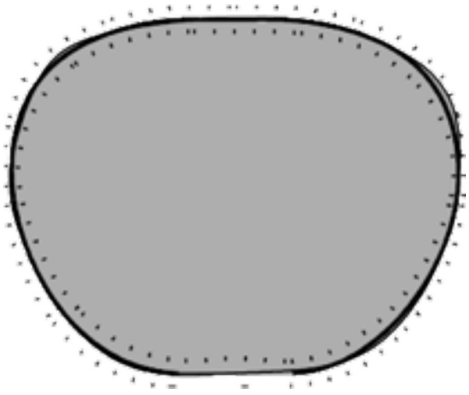

Figure 16 – Estimation of the geometry (sequentially after the estimation of the internal conductivity distribution). The cross-section of the actual domain is shown as a gray patch. The estimated boundary with the augmented approximation error model is shown with a solid line, and two posterior (approximate) standard deviation limits for the approximate posterior with dashed lines. Note that the difference in the actual domain and the estimated domain is visible only in a couple of patches on the boundary!

Process tomography: detecting inhomogeneities in pipeline flows

In process tomography, the dynamics of the (multiphase) flows are estimated based on measurements that are carried out on the boundary. The most common measurement modalities are the electromagnetic ones such as electrical impedance and capacitance tomography. In addition to the challenges posed by these diffuse tomographic modalities (such as the example in the previous paragaph), the primary unknowns (such as the conductivity) are now time varying. Furthermore, only a few measurements (real numbers) can be obtained from the target before the target has changed (due to the typically fast flow along a pipeline). Also, the time to compute the reconstructions is typically of the order of one millisecond, which means that highly approximate reduced order models have to be used. Mathematically, process tomography is classified as a nonstationary inverse problem.

Nonstationary inverse problems are a special class of Bayesian inverse problems in which the primary unknown can be modelled with a stochastic evolution model, typically a stochastic partial differential equation (Kaipio and Somersalo, 1999; 2005). These problems can be tackled with sequential Bayesian estimation methods Doucet *et al.* (2001), and, in special cases, recursive algorithms such as the Kalman (1960) filter and its variants (Anderson and Moore, 1979; Kaipio and Somersalo, 2005).

The most common class of problems that induces nonstationary inverse problems is the one related to transport phenomena such as diffusion, advection and convection in fluids, solids and porous materials, as well as chemical reactions. In particular, problems that are governed by stochastic convection-diffusion models in fluids have served as standard examples for sources of nonstationary inverse problems (C; Seppänen *et al.*, 2007 Seppänen *et al.*, 2008; Voutilainen and Kaipio, 2009). Here, the temporally and spatially varying concentration of some substance is typically the physical entity of primary interest. In these problems, the stochasticity of the models is typically due to unknown boundary data for the convection-diffusion model or the fluid dynamics model (Seppänen *et al.*, 2001b; Voutilainen *et al.*, 2007), unknown velocity field (Seppänen *et al.*, 2001c; Lipponen *et al.*, 2010), and highly approximate reduced order models for the primary unknowns (Huttenen and Kaipio, 2007a; 2007b; Liponnen *et al.*, 2010).

The uncertainties include unknown boundary conditions for both the electromagnetic problem and the evolution model for the flow, unknown electrode contact impedances, and errors due to very severe model reduction. Furthermore, the actual flow field is nonstationary and unknown, while in the computational model a fixed nominal stationary flow field has to be used.

An example of a multiphase pipeline flow with a single solid object flowing past the measurement electrodes, is shown in Figures 17 and 18. The relatively good estimates shown in Figure 17 can be obtained using model that is practically infeasible (computational complexity is so high that the time to compute the estimates is orders of magnitude higher than the actually available time). With a practically feasible (with respect to the computational complexity) model without modelling the approximation errors, the estimates are completely useless, and are not shown here. On the other hand, the estimates in Figure 18 clearly show the passing of the inhomogeneity in the flow. Here, the approximation errors were modelled (and employed), and the computational model was practically feasible.

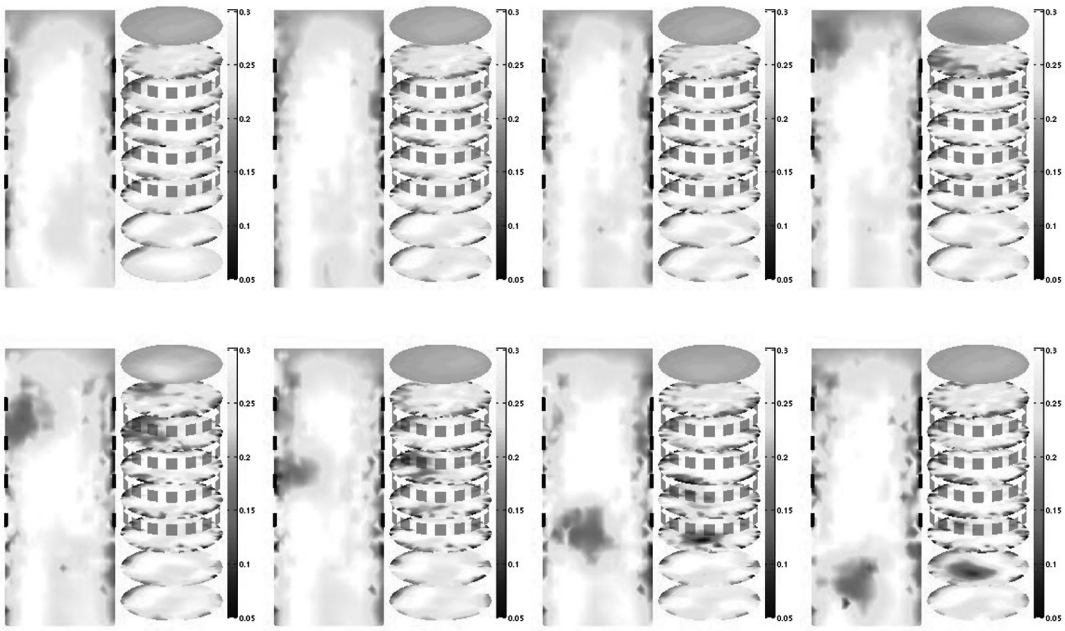

Figure 17 – Reconstruction of a multiphase flow (a single solid object moving fast in a pipeline). Uncertainties are not modeled but a very accurate and computationally excessively complex, and practically infeasible, model is used.

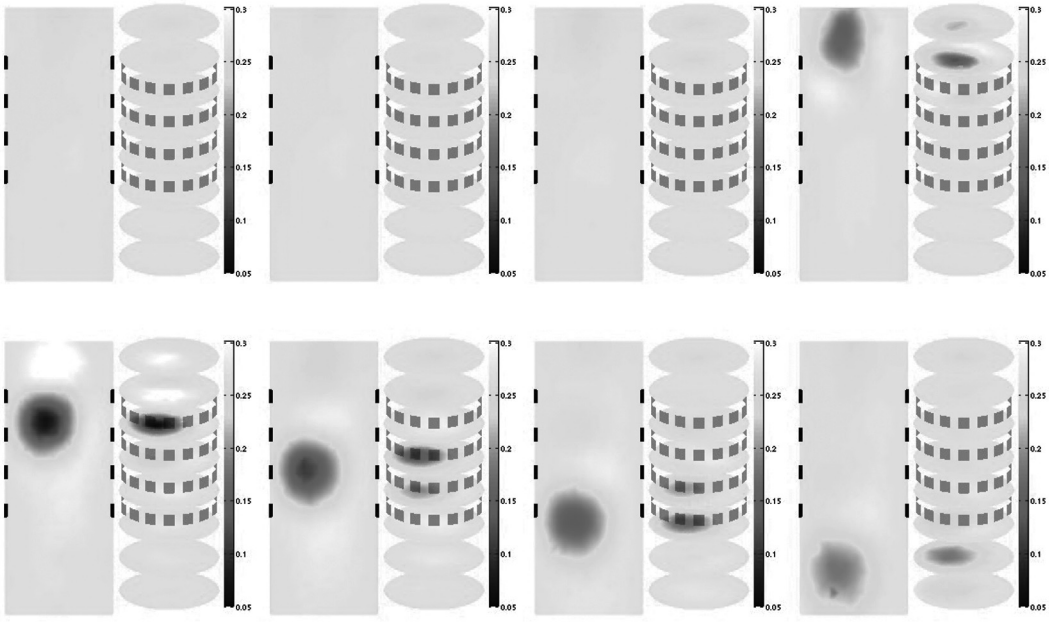

Figure 18 – Reconstruction of a multiphase flow (a single solid object moving fast in a pipeline). Uncertainties are modeled and the computational model is practically feasible.

Example: Hydrology

Finally we consider an application of uncertainty quantification to hydrology which exhibits one of the merits of working within the Bayesian framework. A challenging problem in the management of water resources is quantifying surface water depletion due to groundwater abstractions. As an example, we consider the problem of estimating the rate of stream depletion from a pumping test for the following scenario given in Figure 19. Here we consider the effects of pumping a well screened in a semi-confined aquifer which is hydraulically connected to a phreatic aquifer, which in turn is connected to a stream. For simplicity we suppose that the stream is rectilinear, in which case the model permits an analytic solution given in Dudley Ward and Lough (2011). The following synthetic case study is taken from Cui and Dudley Ward (2013), where the underlying true parameters are given in Table 1. Drawdown data over a five-day pump test from two observation wells screened in each aquifer, with $x = 50$ and $y = 100$, are shown in Figure 20; d_2 denotes the measured data in the pumped aquifer and d_1 denotes the measured data in the upper table aquifer. (The 'true' unobserved signal has been corrupted with Gaussian white noise with a standard deviation $s = 0.01$m).

Table 1 – True parameters for layered aquifer model in Figure 19

Physical Parameter	Value
Transmissivity (water table aquifer) T_1	500m^2/day
Transmissivity (semi-confined aquifer) T_2	3,000m^2/day
Storativity (water table aquifer) S_1	1×10^{-1}
Storativity (water table aquifer) S_2	1×10^{-4}
Aquitard leakage $\frac{K'}{B'}$	3×10^{-2}1/day
Stream bed conductance λ	1m/day

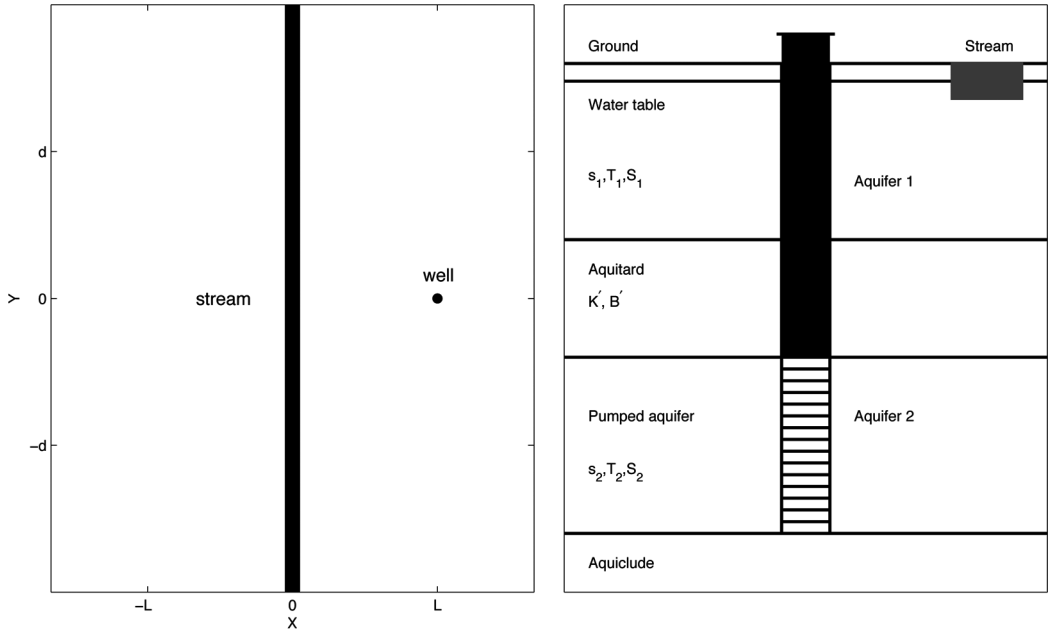

Figure 19 – Plan and cross-sectional views of Dudley Ward and Lough (2011) model.

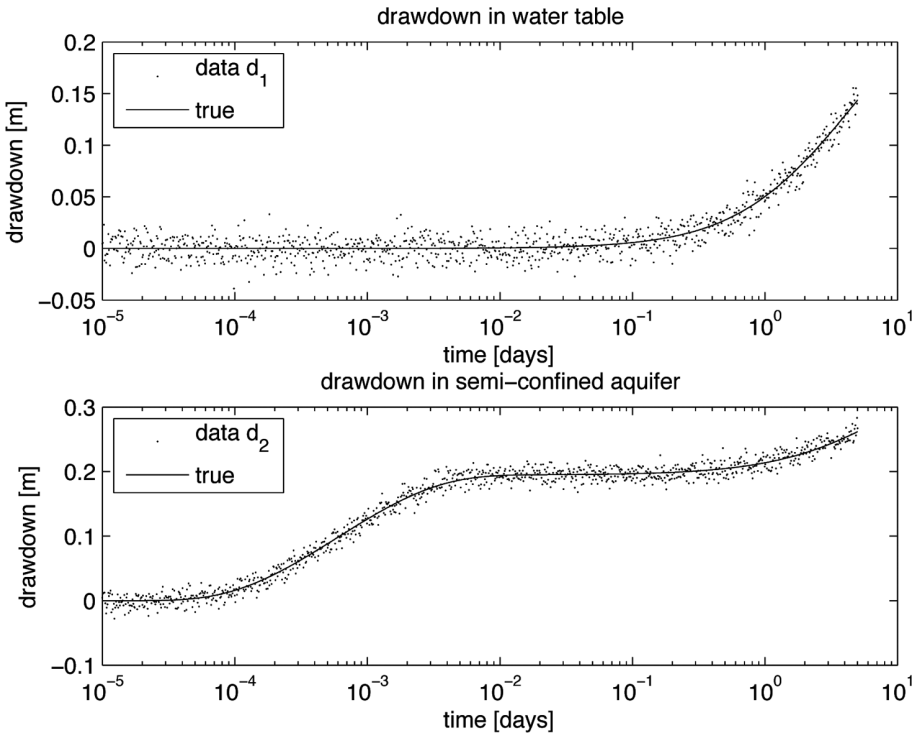

Figure 20 – Synthetic drawdown data for the aquifer model in Figure 19.

The assumed white noise model of measurement data is a reasonable model of pressure transducer data (Dudley Ward and Fox, 2012). We remark that the transmissivity T_1 of the upper aquifer plays an important role in the analysis of data, since relatively small values of T_1 will permit large values of the streambed conductance, since vertical flow to the semi-confined aquifer will be dominated by the aquitard leakage. Figure 21 shows the natural logarithm of the likelihood as a function of T_1 and λ for the fixed (true) values $T_2 = 500$ m/day, $S_1 = 1 \times 10^{-1}$, $S_2 = 1 \times 10^{-4}$, and $\frac{K'}{B'} = 3 \times 10^{-2}$ 1/day. Figure 21 (a) shows the log-likelihood when just the measurement data from the pumped aquifer, d_2, are used, while Figure 21 (b) uses measurement data in both aquifers (d_1 and d_2).

The pump test data tells us absolutely nothing about the stream bed conductance over the time scale of the pumping test. The maximum log-likelihood is given as ridges in each plot, so that arbitrarily large values of the stream bed conductance λ are permitted without a significant reduction in the likelihood; the test data thus does not provide a useful constraint on the conductance. In simple terms, we have unbounded uncertainty of stream depletion, and the data d_2 does very little to constrain the uncertainty. This is, of course, because the signal-to-noise ratio is rather small, and it would be necessary to carry out a pump test over an impractically long period to actually capture the quantity of interest in the measurement data.

Left with this situation there is little that can be done except to collect more test data. Given the relative importance of T_1 and λ, it makes sense to carry out a further pump test in the phreatic aquifer, which on grounds of economy is carried out over 24 hours. The

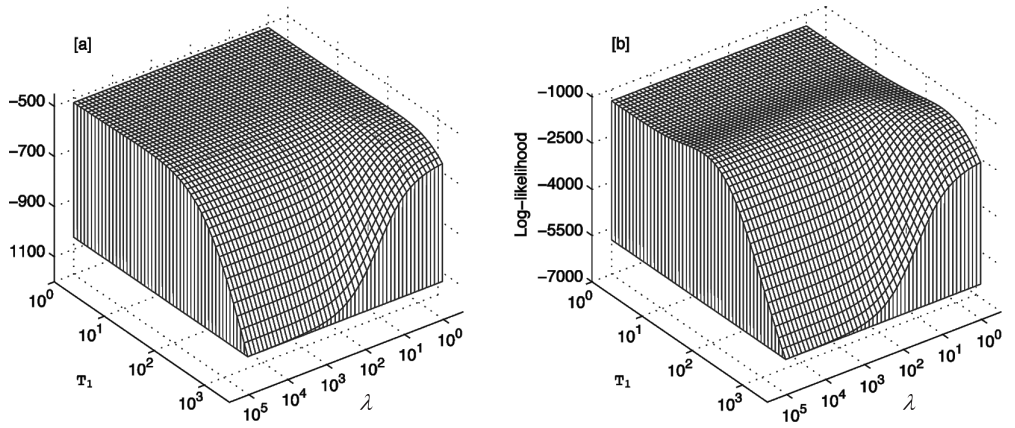

Figure 21 – Logarithm of the conditional likelihood of T_1 and λ for T_2 = 3,000 m^2/day, $S_1 = 0.1$, $S_2 = 1 \times 10^{-4}$, $\frac{K'}{b'} = 3 \times 10^{-2}$1/day, with respect to $\lambda_{\text{true}} = 1$ m/day. Plot (a) uses data d_2, while plot (b) uses d_1 and d_2.

observed data d_3 is plotted in Figure 22. It proves very instructive to investigate to what extent the new data reduces the posterior uncertainty of the stream bed conductance and the predictive uncertainty of stream depletion. Since d_1 had little effect on reducing the uncertainty, we first consider (on grounds of computational economy) only data d_2 and d_3. Figure 23 shows the posterior probability distribution of parameters obtained from Markov Chain Monte Carlo sampling. Of particular note is the posterior distribution of the streambed conductance λ shown in the bottom right hand corner of

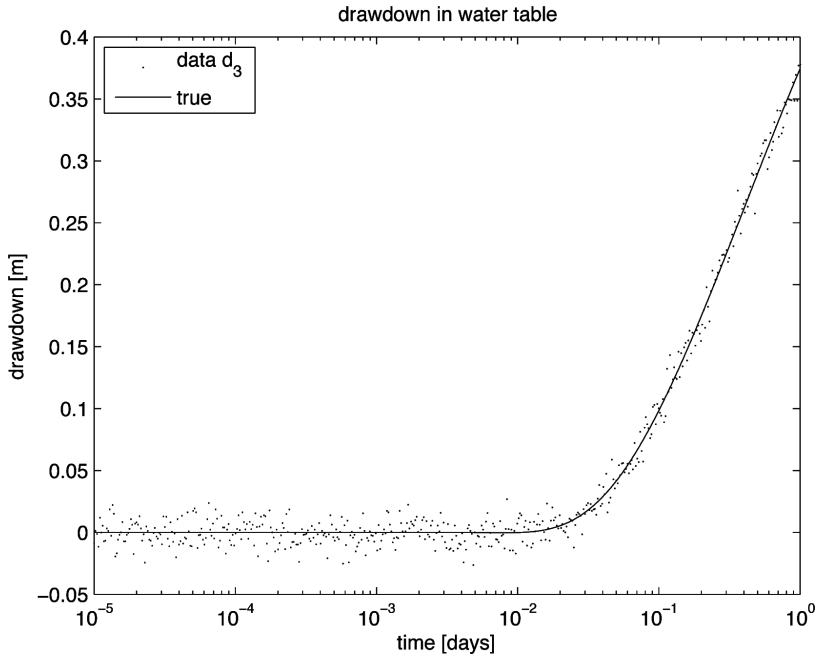

Figure 22 – Drawdown data for shallow pumping test

Figure 23. While λ is now constrained, finite large values of λ are still permitted. Figure 24 shows the predicted probability distribution of stream depletion as a function of time. While the 'envelope of uncertainty' captures the actual depletion, the mean rate of stream depletion which would be a sensible estimator of the actual depletion is a gross overestimate. Finally, for completeness we consider all data d_1, d_2 and d_3, Figure 25 shows the posterior distribution of parameters. This time the uncertainty in λ is constrained to physically realistic values and the predictive uncertainty of stream depletion in Figure 26 is now sensible.

The moral here is that a seemingly innocuous bit of information while in isolation may have an insignificant role in reducing uncertainty, may play a leading part when combined with other sources of information.

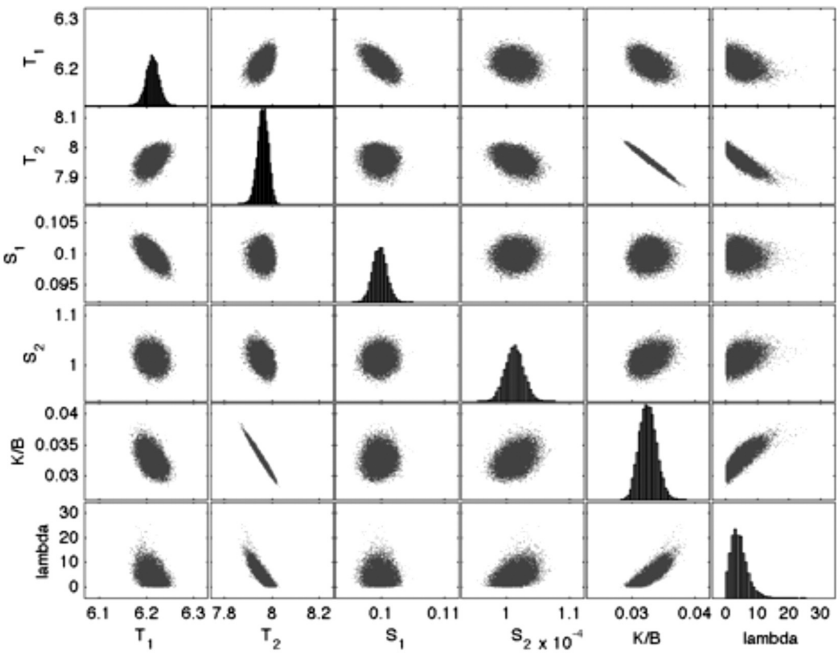

Figure 23 – Scatter plot showing the posterior marginal and joint distributions of parameters using the deep pumping test data d_1 from Figure 20 and the shallow pumping test data d_3 from Figure 22.

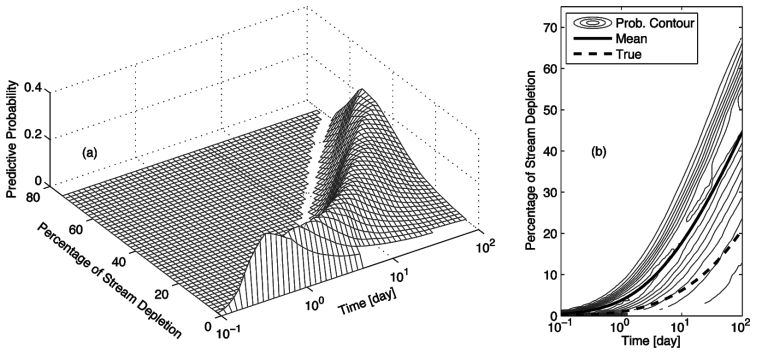

Figure 24 – Posterior probability distribution of percentage of stream depletion as a function of time using the posterior distribution in Figure 23. Plot (b) is a contour plot of (a), while the bold line is the mean predicted stream depletion, and the dotted line is the underlying 'true' stream depletion.

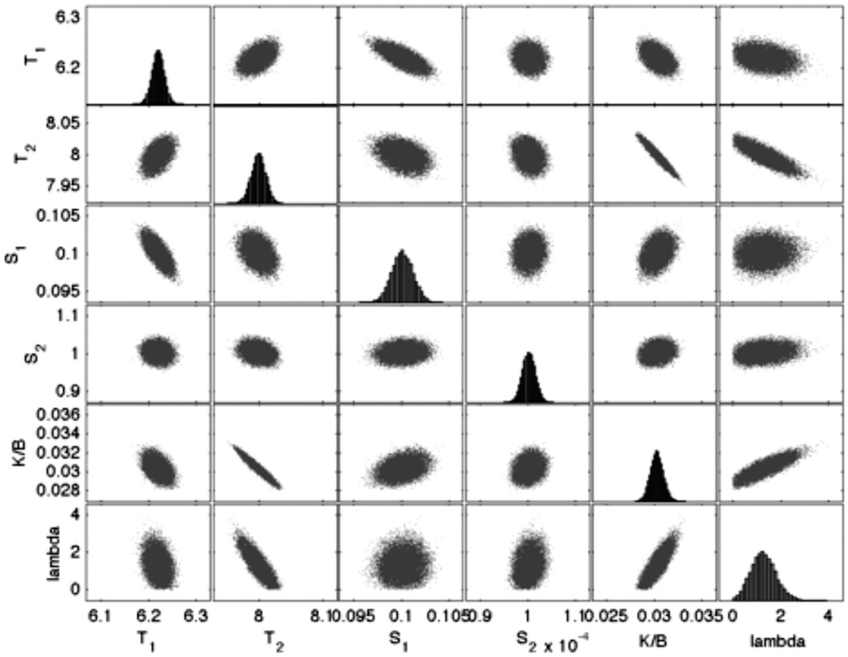

Figure 25 – Scatter plot showing the posterior marginal and joint distributions of parameters using the deep pumping test data d_1, d_2 from Figure 20 and the shallow pumping test data d_3 from Figure 22.

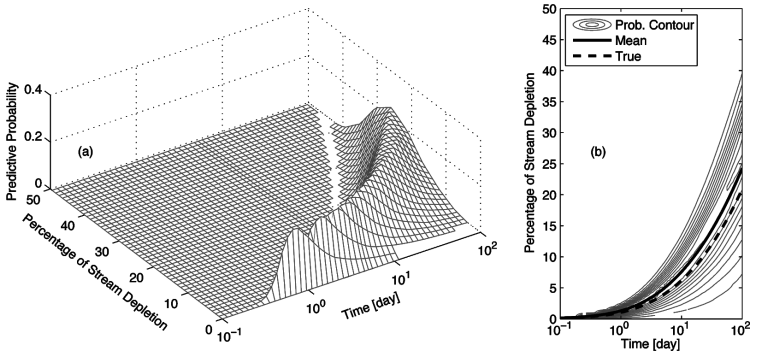

Figure 26 – Posterior probability distribution of percentage of stream depletion as a function of time using the posterior distribution in Figure 25. Plot (b) is a contour plot of (a), while the bold line is the mean predicted stream depletion, and the dotted line is the underlying 'true' stream depletion.

Regularisation approaches

The classical approach to computing estimates for inverse problems is called *regularisation*. Regularisation approaches were originally based on the mathematical theory of ill-posed problems and focused on *exactly known models* with *very small, or asymptotically vanishing* measurement errors. These methods are not suitable for analysing modelling errors, and especially are not suitable for incorporating the associated statistical error models. Essentially all regularization approaches are equipped with a tuning parameter, that is fixed to obtain model fitting residuals that are consistent with the *assumed magnitude of measurement errors*.

The most commonly used regularization approach, Tikhonov regularization, is (in its most commonly used form), a stabilized least-squares approach. This approach, as

well as the other regulariation approaches, inherit the problems of least squares that we encountered with the classroom examples above: outside the statistical regime, the error estimates are unavailable, or meaningless, with real physical errors. Furthermore, the estimates can usually be shown to converge to the true (minimum norm) estimates only when there are no modelling errors and even the measurement errors vanish completely.

In practise, if (the ubiquitous) modelling errors are to be taken into account, one almost invariably chooses the Bayesian framework for analysis and model building.

Discussion

The relevance of uncertainty to decision-making is incontrovertible. However, as we have seen, the spectrum of uncertainty assessments is extremely broad and ranges from the broadly qualitative to quantitative. While qualitative assessments undoubtably have their place, it is extremely important not to assign more value to them than they merit. The broad aim of uncertainty assessment in a scientific or engineering context is to describe the 'envelope of predictive uncertainty'; sometimes this can only be done qualitatively, in which case it is important to recognise that. If, from a policy point of view, it is sufficient to have a bench mark, a consistent framework for classifying risk may be satisfactory – however this should not be dressed up as science. However, it makes sense that the mechanism for assigning risk is as simple as possible.

In our view the examples of earthquake forecasting and liquefaction assessment typify some of the less desirable features of uncertainty assessment. In particular it is extremely important to draw a clear demarcation between sensitivity analysis and uncertainty quantification. In the former it may be acceptable to vary parameters to get an idea of their relative importance, but arbitrarily assigning probability distributions to parameters which are not grounded in data can be misleading, particularly if the assessment of uncertainty is evidently vastly over-confident, as in the earthquake example.

In the case that uncertainties can be effectively modelled and quantified, the envelope of predictive uncertainty is typically assumed to be plus or minus 2 or 3 standard deviations from the estimated value. In this situation the distinction between a 'good' and 'poor' calibrated model is clear; for the former the 'truth', i.e., what actually happens, should be captured by the envelope of predictive uncertainty. Usually this entails that the true parameter values are captured in their respective envelopes of uncertainty (in the case of synthetic examples, and in special measurement campaigns in which the underlying 'truth' is known).

The examples from biomedical imaging and process engineering show how properly modelled uncertainties can significantly reduce uncertainty and lead to much greater clarity (in a demonstrable sense). Hence they can directly inform decision making. The downside of making a poor diagnosis, or poor control of an industrial process, can be catastrophic. Thus the incentive to do better is compelling. Methods based on regularisation, when applied to large scale inverse problems, can usually at best be described as qualitative. Furthermore, there is the problem of interpreting the regularisation parameter, as discussed above.

Most importantly in our view, we have shown how the least squares approach to uncertainty quantification can fail, even for the most trivial examples. This is plainly of immense consequence, because one can be lulled into a false sense of security, since a model can be apparently well-calibrated while at the same time making predictions outside the envelope of uncertainty that are misleading by orders of magnitude. This, at

the very least, should make practitioners sit up and take stock of their tool kit.

In our view one of the great merits of the Bayesian approach to model calibration and uncertainty quantification is that it provides a meaningful framework for exploring the relationship between models and data. In particular, it allows us to answer the question of which measurements are necessary in order that the uncertainty in the quantity of interest is sufficiently small. Thus it provides a framework for uncertainty reduction that goes well beyond sensitivity testing.

In our experience, hydrological modelling presents particular challenges since, more often than not, data is quite poor and sparse. In addition, the computational domains are large and the models may be over-resolved compared to the available data. We note that uncertainty quantification requires, at the least, an estimate of the measurement uncertainty. Spot measurements (which are frequently the case in practical hydrology) do not provide a reliable estimate of measurement uncertainty. Made-up estimates of uncertainty, while they might lead to a solution (i.e., a calibrated model), are not objective. As a result, uncertainties are larger than is usually assumed. We note that one way to deal with a dearth of knowledge about measurement uncertainties is to model these uncertainties as unknowns and *infer* them. Thus all uncertainties (parameter, measurement, and predictive) get bundled up in the same problem.

Large envelopes of uncertainty are to be welcomed (and certainly not ignored), since theysetlimitsonourcurrentstateofknowledge, and present a challenge to reduce the gap in uncertainty. This may be done by improving measurements (e.g., better instrumentation) or extending the measurement process and using new measurements (and models) to reduce uncertainties. For example, analysis of groundwater responses to the Canterbury earthquakes suggested that measurements of spontaneous seismic activity may be used to identify aquifer features of interest such as depth to basement and water table, see Gulley *et al.*, 2013. Initial studies of this problem have been carried out in Lähivaara *et al.* (2013a) and Lähivaara *et al.* (2013b), where significant computational reductions were achieved using the approximation error method. In this situation it is necessary to simulate seismic wave propagation in coupled elastic and poroelastic media. Here, the bulk of the computations were carried out using an elastic wave approximation to the poroelastic waves, rendering the problem both computationally viable and accurate. In this way quantified uncertainties can be better characterised and controlled. However, the downside (almost always) is that quantifying uncertainty presents very significant computational challenges. Ingenuity, as well as advances in computational power, are required to render a general programme of uncertainty quantification feasible.

The other challenge with quantified uncertainties is how they fit into a decision-making framework. It is much easier to make a decision on a point estimate, and provided the consequences are not too serious this may appear reasonable. The danger is that while that may work now with no obvious detrimental consequences, the longer-term effect may be catastrophic, particularly if we have seriously under- or over-estimated a quantity. In situations where economic consequences are significant, as in oil prospecting or in countries with serious water resource problems, the incentive to do better work is compelling. However, while New Zealand does not have a real problem with water quantity – at the moment – it is not inconceivable that in the future the tables will be turned. The honest thing to do is to present all the information (estimate plus uncertainty), and the onus is then on policy makers to incorporate this into rules and regulations.

Acknowledgements

In conclusion we express our sincere gratitude to Leanne Kirk of the Mathematics Department of the University of Otago for converting the original manuscript from Latex to Word. We also thank Chris Palmer for his expert advice in the preparation of the Word version of this article. This work has been funded by the New Zealand Ministry of Science and Innovation as part of the SMART Aquifer Project.

References

Anderson, B.D.O.; Moore, J.B. 1979: *Optimal filtering*. Prentice-Hall Englewood Cliffs, NJ.

Arridge, S.R.; Kaipio, J.P.; Kolehmainen, V.; Schweiger, M.; Somersalo, E.; Tarvainen, T.; Vauhkonen, M. 2006: Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems 22* : 175-195.

Calvetti, D.; Kaipio, J.P.; Somersalo, E. 2006: Aristotelian prior boundary conditions. *International Journal of Mathematics 1*: 63-81.

Calvetti, D.; Somersalo, E. 2007: *An Introduction to Bayesian Scientific Computing Ten Lectures on Subjective Computing*. Springer.

Cui, T.; Dudley Ward, N.F. 2013: Uncertainty quantification for stream depletion tests. *Journal of Hydrologic Engineering 18(12)* : 1581-1590.

Cui, T.; Ward, N.F.D.; Kaipio, J.P. 2013: Characterisation of parameters for a spatially heterogenous aquifer from pumping test data. *Journal of Hydrologic Engineering 18(12)* : 1581-1590. .

Doucet, A.; de Freitas, N.; Gordon, N. 2001: *Sequential Monte Carlo Methods in Practice*. Springer.

Freedman, D.A.; Stark, P.B. 2003: What is the probability of an earthquake? In F. Mulargia and R.J. Geller (*eds.*), *Earthquake Science and Seismic Risk Reduction. NATO Science Series IV: Earth and Environmental Sciences, vol. 32* : 201-213, Kluwer, Dordrecht, The Netherlands.

Gulley, A.K.; Dudley Ward, N.F.; Cox, S.C.; Kaipio, J. 2013: Groundwater responses to the recent Canterbury earthquakes: a comparison. *Journal of Hydrology 504* : 171-181.

Heino, J.; Somersalo, E. 2004: A modelling error approach for the estimation of optical absorption in the presence of anisotropies. *Physics in Medicine and Biology 49* : 4785-4798.

Heino, J.; Somersalo, E.; Kaipio, J.P. 2005: Compensation for geometric mismodelling by anisotropies in optical tomography. *Optics Express 13(1)* : 296-308.

Huttunen, J.M.J.; Kaipio, J.P. 2007a: Approximation error analysis in nonlinear state estimation with an application to state- space identification. *Inverse Problems 23* : 2141-2157.

Huttunen, J.M.J.; Kaipio, J.P. 2007b: Approximation errors in nonstationary inverse problems. *Inverse Problem and Imaging 1(1)*: 77-93.

Huttunen, J.M.J.; Kaipio, J.P. 2009: Model reduction in state identification problems with an application to determination of thermal parameters. *Applied Numerical Mathematics 59* : 877-890.

Huttunen, J.M.J.; Lehikoinen, A.; Hämaläinen, J.; Kaipio, J.P. 2010: Importance filtering approach for the nonstationary approximation error method. *Inverse Problems 26(12)* : 125003.

Idris, I.M.; Boulanger, R.W. 2006: Semi-empirical procedures for evaluating liquefaction potential during earthquakes. *Soil Dynamics and Earthquake Engineering 26* : 115-130.

Kaipio, J.P.; Somersalo, E. 1999: Nonstationary inverse problems and state estimation. *Journal of Inverse and Ill-posed Problems 7* : 273-282.

Kaipio, JP.; Somersalo, E 2005: *Statistical and Computational Inverse Problems*. Springer- Verlag.

Kaipio, J.P.; Somersalo, E. 2007: Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics 198(2)86* : 493-504.

Kaipio, J.P.; Kolehmainen, V. 2013: Approximate marginalization over modeling errors and uncertainties in inverse problems. *In* P. Damien, N. Polson, and D. Stephens (eds.), *Bayesian Theory and Applications*. Oxford University Press.

Kalman, R.E. 1960: A new approach to linear filtering and prediction problems. *Transactions of the ASME. Series D, Journal of Basic Engineerin 82D(1)* : 35-45.

Kolehmainen, V.; Schweiger, M.; Nissilä, I.; Tarvainen, T.; Arridge, S.R.; Kaipio, J.P. 2009: Approximation errors and model reduction in three-dimensional diffuse optical tomography. *Journal of the Optical Society of America 26(10)* : 2257-2268.

Kolehmainen, V. Siltanen, S.; Järvenpää, S.; Kaipio, J.P.; Koistinen, P.; Lassas, M.; Pirttilä, J.; Somersalo, E. 2003: Statistical inversion for medical x-ray tomography with few radiographs ii: Application to dental radiology. *Physics in Medicine and Biology 48(10)* : 1465-1490.

Kolehmainen, V.; Tarvainen, T.; Arridge, S.R.; Kaipio, J.P. 2011: Marginalization of uninteresting distributed parameters in inverse problems – application to optical tomography. *International Journal for Uncertainty Quantification 1(1)* : 1-17.

Lähivaara, T.; Dudley Ward, N.; Huttunen, T.; Kaipio, J.P.; Koponen, J. 2013a:. Estimation of aquifer dimensions from passive seismic signals with approximate wave propagation models. *Inverse Problems*,

T. Lähivaara, T.; Kaipio, J.P.; Dudley Ward, N.; Huttunen, T. 2013b:Simulation study on seismic monitoring of aquifers. *Proceedings of Meetings on Acoustics 19* : 045057.

Lehikoinen, A. Finsterle, S. Voutilainen, A. Heikkinen, L.M.; Vauhkonen, M.; Kaipio, J.P. 2007: Approximation errors and truncation of computational domains with application to geophysical tomography. *Inverse Problems and Imaging 1*: 371-389.

Lehikoinen, A.; Huttunen, J.M.J.; Voutilainen, A.; Finsterle, S.; Kowalsky, M.B.; Kaipio, J.P. 2010: Dynamic inversion for hydrological process monitoring with electrical resistance tomography under model uncertainties. *Water Resources Research 46(4)* : W04513.

Lipponen, A.; Seppänen, A.; Kaipio, J.P. 2010: Nonstationary inversion of convection-diffusion problems – recovery from unknown nonstationary velocity fields. *Inverse Problems*: 463-468.

Nissinen, A.; Heikkinen, L.M.;. Kaipio, J.P. 2008: The Bayesian approximation error approach for electrical impedance tomography – experimental results. *Measurement Science and Technology 19* : 015501.

Nissinen, A.; Heikkinen, L.M.; Kolehmainen,V.; Kaipio, J.P. 2009: Compensation of errors due to discretization, domain truncation and unknown contact impedances in electrical impedance tomography. *Measurement Science and Technology 20* : 105504.

Nissinen, A. Kolehmainen, V.; Kaipio, J.P. 2011a: Compensation of modelling errors due to unknown domain boundary in electrical impedance tomography. *IEEE Transactions on Medical Imaging 30(2)* : 231-242.

Nissinen, A.; Kolehmainen, V.; Kaipio, J.P. 2011b: Reconstruction of domain boundary and conductivity in electrical impedance tomography using the approximation error approach. *International Journal for Uncertainty Quantification 1(3)* : 203-222.

Seed, H.B.; Idris, I.M. 2006: A simplified procedure for evaluating liquefaction potential. *Earthquake Research Center, University of California Berkeley* 26: 115-130,.

Seppänen, A.; Heikkinen, L.; Savolainen, T.; Voutilainen, A.; Somersalo, E.; Kaipio, J.P. 2007: An experimental evaluation of state estimation with fluid dynamical models in process tomography. *Chemical Engineering Journal 127* : 23-30.

Seppänen, A.; Vauhkonen, M.; Vauhkonen, P.J.; Somersalo,E.; Kaipio, J.P. 2001a: State estimation with fluid dynamical evolution models in process tomography – an application to impedance tomography. *Inverse Problems 17* : 467-484.

Seppänen, A.; Vauhkonen, M.; Vauhkonen, P.J.; Somersalo,E.; Kaipio, J.P. . 2001b: State estimation with fluid dynamical evolution models in process tomography – an application to impedance tomography. *Inverse Problems*, *17* : 467-483.

Seppänen, A.; Vauhkonen, M.; Vauhkonen, P.J.; Somersalo,E.; Kaipio, J.P. 2001c: State estimation with fluid dynamical evolution models in process tomography – an application to impedance tomography. *Inverse Problems*, *17* : 467-484.

Seppänen, A.; Vauhkonen, M.; Vauhkonen, P.J.; Voutilainen, A.; Kaipio, P.J. 2008:. State estimation in process tomography – Three-dimensional impedance imaging of moving fluids. *International Journal for Numerical Methods in Engineering 73(11)* : 1651-1670.

Siltanen, V. Kolehmainen, V.; Järvenpää, S.; Kaipio, J.P.; Koistinen, P.; Lassas, M.; Pirttilä, J,; Somersalo, E 2003: Statistical inversion for medical x-ray tomography with few radiographs i: General theory. *Physics in Medicine and Biology 48* : 1437-1463.

Tarantola, A. 2004: *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia.

Tarvainen, T.; Kolehmainen, V.; Pulkkinen, A.; Vauhkonen, M.; Schweiger, M.; Arridge, S.R.; Kaipio, J.P. 2010: Approximation error approach for compensating modelling errors between the radiative transfer equation and the diffusion approximation in diffuse optical tomography. *Inverse Problems 26* : 015005.

USGS 1999: Earthquake probabilities in the San Francisco Bay Region: 2000-203 – a summary of findings. Working Group on California Earthquake Probabilities Technical report, United States Geological Survey.

Voutilainen, A.; Kaipio, J.P. 2009: Model reduction and pollution source identification from remote sensing data. *Inverse Problems and Imaging 3(4)*: 711-730.

Voutilainen, A.; Pyhälahti, T.; Kallio, K.Y.; Pulliainen, J.; Haario, H.; Kaipio, J.P. 2007: A filtering approach for estimating lake water quality from remote sensing data. *International Journal of Applied Earth Observations and Geoinformation 9(1)*: 50-64.

Ward, N.F.D.; Fox, C. 2012: Identification of aquifer parameters from pumping test data with regard for uncertainty. *Journal of Hydrologic Engineering 17(7)* : 769-781. doi: 10.1061/(ASCE)HE.1943-5584.0000521.

Ward, N.F.D.; Lough, H. 2011: Stream depletion from pumping a semi-confined aquifer in a two-layer leaky aquifer system. *Journal of Hydrologic Engineering 16(11)* : 955-959. doi:10.1061/(ASCE)HE.1943-5584.0000382.

Weinan, E. 2011: *Principles of Multiscale Modeling*. Princeton University Press.

Appendix

Approximation Error Method

We detail the approximation error method which is based on Kaipio and Kolehmainen (2013).

Let the unknowns be (x, z, ξ, e), where e represents additive errors while ξ represents auxiliary uncertainties such as unknown boundary data, and (x, z) are two distributed parameters of which only x is of interest. The accurate forward model

$$(x,z,\xi) \quad \bar{A}(x,z,\xi) \tag{11}$$

is usually a nonlinear one. The uncertainties ξ must sometimes be modelled as mutually dependent with (x, z), especially when ξ represents boundary data on the computational domain boundary and (x, z) are modelled as random fields. On the other hand, if ξ represents unknown boundary shape, ξ might be modelled as mutually independent with (x, z). In the following, we consider the case in which the noise e is additive and the unknowns (x, z, ξ) are not necessarily mutually independent.

Let

$$y = \bar{A}(\bar{x},z,\xi) + e \in \Re^m$$

denote an accurate model for the relation between the measurements and the unknowns and let e be mutually independent with (x, z, ξ).

Below, we approximate the accurate representation of the primary unknown $\bar{x}$ by $x = P\bar{x}$, where P is typically a projection operator. Let $\pi(x, z, \pi, \xi, e)$ be a feasible model for the joint distribution of the unknowns. We identify $x = P\bar{x}$ with its coordinates in the associated basis.

In the approximation error approach, we proceed as follows. Instead of using the accurate forward model $(\bar{x},z,\xi) \mapsto \bar{A}(\bar{x},z,\xi)$ with $(\bar{x},z,\xi)$ as the unknowns, we fix the random variables (z, ξ) $\leftarrow$ (z_0, ξ_0) and use the computationally (possibly drastically reduced) approximate model

$$x \mapsto A(x,z_0,\xi_0)$$

Here, the predictions of the two models $\bar{A}(\bar{x},z,\xi)$ and $A(x, z_0,\xi_0)$ may be drastically different. Thus, we write the measurement model in the form

$$y = \bar{A}(\bar{x},z,\xi) + e \tag{12}$$

$$= A(x,z_0,\xi_0) + \left(\bar{A}(\bar{x},z,\xi) - A(x,z_0,\xi_0)\right) + e \tag{13}$$

$$= x \mapsto A(x,z_0,\xi_0) + \varepsilon + e \tag{14}$$

where we define the *approximation error* $\varepsilon = \phi(x,z,\xi) = \bar{A}(\bar{x},z,\xi) - A(x,z_0,\xi_0)$. Thus, the approximation error is the discrepancy of predictions of the measurements (given the unknowns) when using the accurate model $\bar{A}(x,z,\xi)$ and the approximate model $A(x, z_0, \xi_0)$. Note that Eq. (14) is exact.

Formally, after the models $\bar{A}$ and A are fixed, we have $\pi\left(\varepsilon \mid \bar{x}, z, \xi\right) = \delta\left(\varepsilon - \phi\left(\bar{x}, z, \xi\right)\right)$. We will later, however, employ approximate joint distributions and therefore consider $\pi\left(\varepsilon, \bar{x}, z, \xi\right)$ without any special structure. As the first approximation, we approximate $\phi\left(\bar{x}, z, \xi\right) \approx \phi\left(Px, z, \xi\right)$ and thus $\pi\left(\varepsilon \mid \bar{x}, z, \xi\right) \approx \pi\left(\varepsilon \mid Px, z, \xi\right)$. This means that we assume that the model predictions and thus the approximation error is essentially the same for $\bar{x}$ as $x = P\bar{x}$. This assumption holds for inverse problems in most cases.

We use the Bayes' formula repeatedly

$$\begin{aligned}
\pi(y, x, z, \xi, e, \varepsilon) &= \pi(y \mid x, z, \xi, e, \varepsilon)\, \pi(x, z, \xi, e, \varepsilon) \\
&= \delta(y - A(x, z_0, \xi_0) - e - \varepsilon) \\
&\qquad \pi(e, \varepsilon \mid x, z, \xi)\, \pi(z, \xi \mid x)\, \pi(x) \\
&= \pi(y, z, \xi, e, \varepsilon \mid x)\, \pi(x)
\end{aligned}$$

Hence

$$\begin{aligned}
\pi(y \mid x) &= \iiiint \pi(y, z, \xi, e, \varepsilon \mid x)\, de\, d\varepsilon\, dz\, d\xi \\
&= \iint \delta\left(y - A(x, z_0, \xi_0) - e - \varepsilon\right) \\
&\qquad \cdot \iint \pi\left(e, \varepsilon \mid x, z, \xi\right) \pi(z, \xi \mid x)\, dz\, d\xi\, de\, d\varepsilon \\
&= \iint \delta\left(y - A(x, z_0, \xi_0) - e - \varepsilon\right) \pi(e, \varepsilon \mid x)\, de\, d\varepsilon \\
&= \int \pi_e\left(y - A(x, z_0, \xi_0) - \varepsilon\right) \pi_{\varepsilon|x}(\varepsilon \mid x)\, d\varepsilon \qquad (15)
\end{aligned}$$

since e and x are mutually independent, and Eq. (15) is a convolution integral with respect to ε. In particular, since e is mutually independent with (x, z, ξ), e and ε are also mutually independent.

At this stage, in the approximation error approach, both π_e and $\pi_{\varepsilon|x}$ are approximated with normal distributions. Let the normal approximation for the joint density $\pi(\varepsilon, x)$ be

$$\pi(\varepsilon, x) \alpha \exp\left\{ -\frac{1}{2} \begin{pmatrix} \varepsilon - \varepsilon_* \\ x - \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \Gamma_\varepsilon & \Gamma_{\varepsilon x} \\ \Gamma_{x\varepsilon} & \Gamma_x \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon - \varepsilon_* \\ x - \end{pmatrix} \right\} \qquad (16)$$

Thus we write $\quad e \sim N\left(e_*, \Gamma_e\right), \quad \varepsilon \mid x \sim N\left(\varepsilon_{*,x}, \Gamma_{\varepsilon|x}\right)$

where

$$\varepsilon_{*,x} = \varepsilon_* + \Gamma_{\varepsilon\, x}\, \Gamma_{x\, x}^{-1} \left(x - x_*\right) \qquad (17)$$

$$\Gamma_{\varepsilon|x} = \Gamma_e - \Gamma_{\varepsilon x} \Gamma_x^{-1} \Gamma_{x\varepsilon} \qquad (18)$$

Define the normal random variable ν so that $\nu \mid x = e + \varepsilon \mid x$

$$\nu \mid x \sim N(\nu_{*|x}, \Gamma_{\nu|x})$$

where

$$\nu_{*|x} = e_* + \varepsilon_* + \Gamma_{\varepsilon x} \Gamma_x^{-1} \left(x - x_*\right) \qquad (19)$$

$$\Gamma_{v|x} = \Gamma_e + \Gamma_\varepsilon - \Gamma_{\varepsilon x}\Gamma_x^{-1}\Gamma \tag{20}$$

Thus, we obtain the approximate likelihood model:

$$y|x \sim N\left(y - A(x, z_0, \xi_0) - v_{*|x}, \Gamma_{v|x}\right)$$

Since we are after computational efficiency, a normal approximation for the prior model would also typically be employed in the construction of the posterior model

$$x \sim N(x_*, \Gamma_x)$$

Thus, we obtain the approximation for the posterior distribution

$$\pi(x \mid y) \propto \pi(y \mid x)\pi(x) \propto \exp\left(-\frac{1}{2}V(x)\right)$$

where $V(x)$ is the posterior potential

$$\begin{aligned} V(x) &= (y - A(x, z_0, \xi_0) - v_{*|x})^T \Gamma_{v|x}^{-1} (y - A(x, z_0, \xi_0) - v_{*|x}) \\ &\quad + (x - x_*)^T \Gamma_x^{-1} (x - x_*) \end{aligned} \tag{21}$$

$$= \left\| L_{v|x}(y - A(x, z_0, \xi_0) - V_{*|x}) \right\|^2 + \left\| L_x (x - x_*) \right\|^2 \tag{22}$$

and where $\Gamma_{v|x}^{-1} = L_{v|x}^T L_{v|x}$, $\Gamma_x^{-1} = L_x^T L_x$ and $v_{*|x} = v_{*|x}(x)$.

The MAP estimate of x with the approximation error model is obtained by

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\{ \left\| L_{v|} \left(y - A\left(x, z_0, \xi_0\right) - v_{*|}\right) \right\|^2 + \left\| L_x \left(x - x_*\right) \right\|^2 \right\} \tag{23}$$

Then, the approximate posterior covariance would be computed by linearizing $A(x, z_0, \xi_0)$ at $x = \hat{x}$

$$\hat{\Gamma}_{x|d} \approx \left(J^T \Gamma_{v|x}^{-1} J + \Gamma^{-1}\right)^{-1} \tag{24}$$

where $J = J + \Gamma_{\varepsilon x}\Gamma_x^{-1}$ and J is the Jacobian of $A(x, z_0, \xi_0)$ evaluated at $x = \hat{x}$.

The associated means and covariances are then (almost always) estimated with the respective sample averages.

Manuscript received 8 April 2013; accepted for publication 16 December 2013