Uncertainty Quantification in Bayesian Inversion

Andrew M Stuart*

Abstract. Probabilistic thinking is of growing importance in many areas of mathematics. This paper highlights the beautiful mathematical framework, coupled with practical algorithms, which results from thinking probabilistically about inverse problems arising in partial differential equations.

Many inverse problems in the physical sciences require the determination of an unknown field from a finite set of indirect measurements. Examples include oceanography, oil recovery, water resource management and weather forecasting. In the Bayesian approach to these problems, the unknown and the data are modelled as a jointly varying random variable, typically linked through solution of a partial differential equation, and the solution of the inverse problem is the distribution of the unknown given the data.

This approach provides a natural way to provide estimates of the unknown field, together with a quantification of the uncertainty associated with the estimate. It is hence a useful practical modelling tool. However it also provides a very elegant mathematical framework for inverse problems: whilst the classical approach to inverse problems leads to ill-posedness, the Bayesian approach leads to a natural well-posedness and stability theory. Furthermore this framework provides a way of deriving and developing algorithms which are well-suited to the formidable computational challenges which arise from the conjunction of approximations arising from the numerical analysis of partial differential equations, together with approximations of central limit theorem type arising from sampling of measures.

Mathematics Subject Classification (2010). Primary 35R30; Secondary 62C10.

Keywords. Inverse problems, Bayesian inversion, Uncertainty quantification, Monte Carlo methods, Stochastic partial differential equations.

1. Introduction

Let X, R be Banach spaces and $G : X \to R$. For example G might represent the *forward* map which takes the input data $u \in X$ for a partial differential equation (PDE) into the solution $r \in R$. **Uncertainty quantification** is concerned with

*The author is grateful to EPSRC, ERC and ONR for financial support which led to the work described in this lecture. He is grateful to Marco Iglesias for help in preparing the figures and to Yuan-Xiang Zhang for careful proofreading.

determining the propagation of randomness in the input u into randomness in some *quantity of interest* $q \in Q$, with Q again a Banach space, found by applying operator $\mathcal{Q} : R \to Q$ to $G(u)$; thus $q = (\mathcal{Q} \circ G)(u)$. The situation is illustrated in Figure 1.

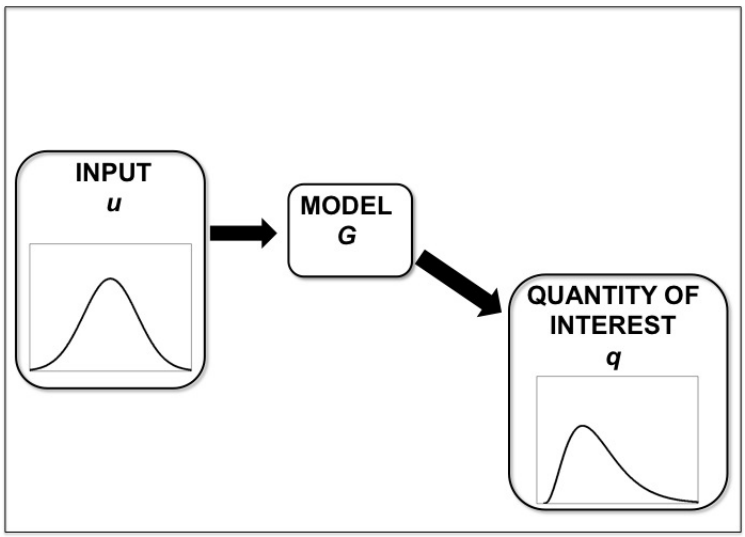

Figure 1. Uncertainty Quantification

Inverse problems are concerned with the related problem of determining the input u when given noisy *observed data* y found from $G(u)$. Let Y be the Banach space where the observations lie, let $\mathcal{O} : R \to Y$ denote the *observation operator*, define $\mathcal{G} = \mathcal{O} \circ G$, and consider the equation

$$y = \mathcal{G}(u) + \eta \tag{1.1}$$

viewed as an equation for $u \in X$ given $y \in Y$. The element $\eta \in Y$ represents *noise*, and typically something about the size of η is assumed known, often only in a statistical sense, but the actual instance of η entering the data y is not known. The aim is to reconstruct u from y. The **Bayesian inverse problem** is to find the the conditional probability distribution on $u|y$ from the joint distribution of the random variable (u, y); the latter is determined by specifying the distributions on u and η and, for example, assuming that u and η are independent. This situation is illustrated in Figure 2.

To formulate the inverse problem probabilistically it is natural to work with separable Banach spaces as this allows for development of an integration theory (Bochner) as well as avoiding a variety of pathologies that might otherwise arise; we assume separability from now on. The probability measure on u is termed the *prior*, and will be denoted by μ_0, and that on $u|y$ the *posterior*, and will be denoted by μ^y. Once the Bayesian inverse problems has been solved, the uncertainty in q can be quantified with respect to input distributed according to the posterior on

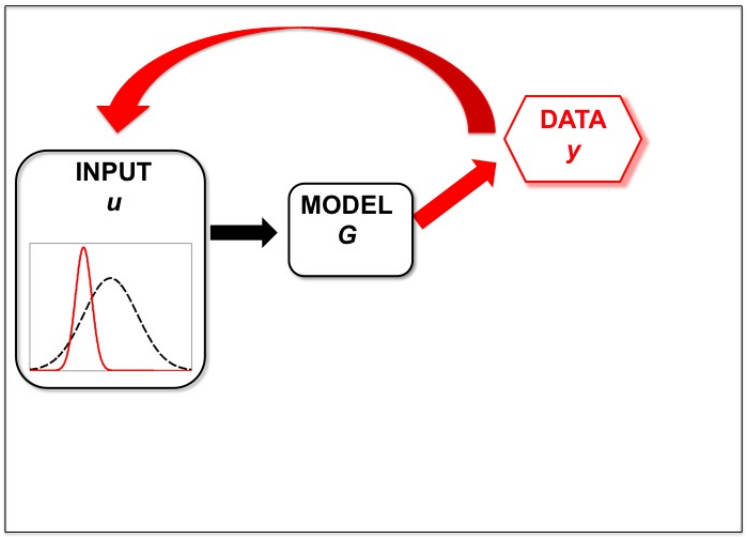

Figure 2. Bayesian Inverse Problem

$u|y$, resulting in improved quantification of uncertainty in comparison with simply using input distributed according to the prior on u. The situation is illustrated in Figure 3. The black dotted lines demonstrate uncertainty quantification prior to incorporating the data, the red curves demonstrate uncertainty quantification after the data has been incoprorated by means of Bayesian inversion.

Carrying out the program illusrated in Figure 3 can have enormous benefits within a wide-range of important problems arising in science and technology. This is illustrated in Figure 4. The top two panels show representative draws from the prior (left) and posterior (right) probability distribution on the geological properties of a subsurface oil field, whilst the bottom two panels show predictions of future oil production, with uncertainty represented via the spread of the ensemble of outcomes shown, again under the prior on the left and under the posterior on the right. The unknown u here is the log permeability of the subsurface, the data y comprises measurements at oil wells and the quantity of interest q is future oil production. The map G is the solution of a system of partial differential equations (PDEs) describing the two-phase flow of oil-water in a porous medium, in which u enters as an unknown coefficient. The figure demonstrates that the use of data significantly reduces the uncertainty in the predictions.

The reader is hopefully persuaded, then, of the power of combining a mathematical model with data. Furthermore it should also be apparent that the set-up described applies to an enormous range of applications; it is also robust to changes, such as allowing for correlation between the noise η and the element $u \in X$. However, producing Figure 4, and similar in other application areas, is a demanding computational task: it requires the full power of numerical analysis, to approximate the forward map G, and the full power of computational statistics, to probe

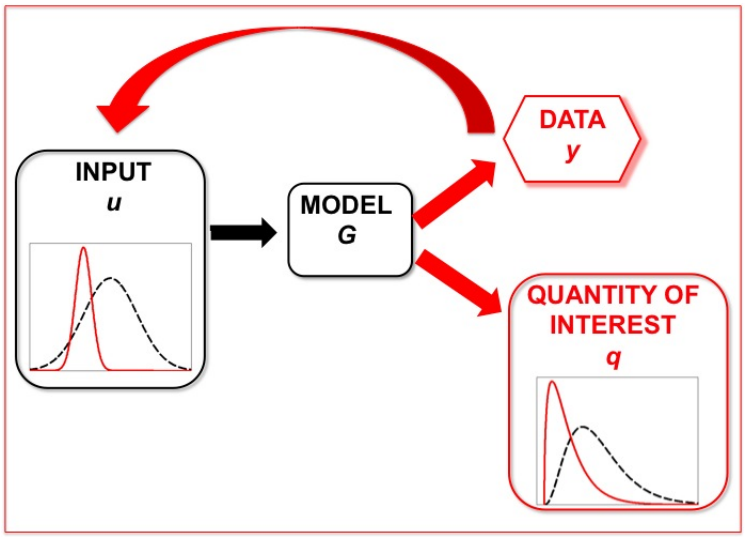

Figure 3. Uncertainty Quantification in Bayesian Inversion.

the posterior distribution. The central thrust of the mathematical research which underlies this talk is concerned with how to undertake such tasks efficiently. The key idea underlying all of the work is to conceive of Bayesian inversion in the separable Banach space X, to conceive of algorithms for probing the measure μ^y on X and, only once this has been done, to then apply discretization of the unknown field u, to a finite dimensional space $\mathbb{R}^N$, and discretization of the forward PDE solver. This differs from a great deal of applied work which discretizes the space X at the very start to obtain a measure $\mu^{y,N}$ on $\mathbb{R}^N$, and then employs standard statistical techniques on $\mathbb{R}^N$. The idea is illustrated in Figure 5. Of course algorithms derived by the black route and the red route *can* lead to algorithms which coincide; however many of the algorithms derived via the the black route do not behave well under refinement of the approximation, $N \to \infty$, whilst those derived via the red route do since they are designed to work on X where $N = \infty$. Conceptual problem formulation and algorithm development via the red route is thus advocated.

This may all seem rather discursive, but a great deal of mathematical meat has gone into making precise theories which back-up the philosophy. The short space provided here is not enough to do justice to the mathematics and the reader is directed to [73] for details. Here we confine ourselves to a brief description of the historical context for the subject, given in section 2, and a summary of some of the novel mathematical and algorithmic ideas which have emerged to support the philisophy encapsulated in Figure 5, in sections 4 and 5. Section 3 contains some examples of inverse problems which motivated the theoretical work highlighted in sections 4 and 5, and may also serve to help the reader who prefers concrete settings. Section 6 contains some concluding remarks.

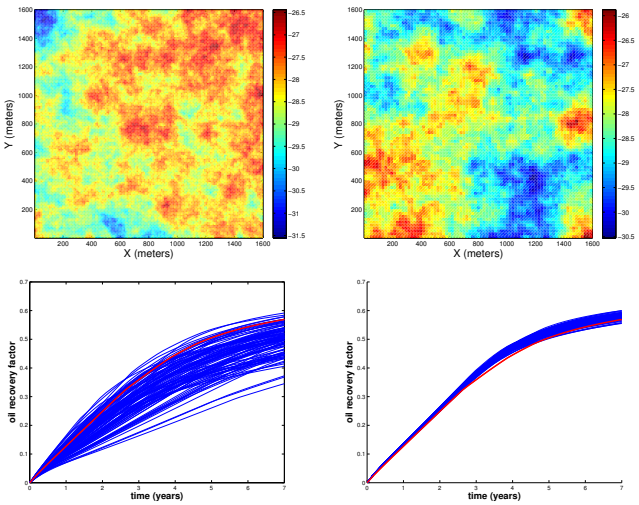

Figure 4. Upper panels: typical draws from the prior (left) and posterior (right). Lower panels: uncertainty in oil production under the prior (left) and posterior (right).

2. Historical Context

A cornerstone in the mathematical development of uncertainty quantification is the book [28] which unified and galvanized a growing engineering community interested in problems with random (uncertain) parameters. The next two and a half decades saw remarkable developments in this field, on both the applied and theoretical sides; in particular a systematic numerical analysis evolved which may be traced through the series of papers [77, 5, 6, 7, 59, 60, 15, 16, 17, 69, 62] and the referenes therein. Inverse problems have a long history and arise in an enormous range of applications and mathematical formulations. The 1976 article of Keller [38] is widely cited as foundational in the classical approach to inverse problems, and the modern classical theory, especially in relation to PDE and integral equations, is overviewed in a variety of texts: see [25, 39], for example.

The classical theory of inverse problems does not quantify uncertainty: typically it employs knowledge of the size of η but not its statistical distribution. However as long ago as 1970 the possibility of formulating PDE inverse problems in terms of Bayes' formula on the space X was recognized by Franklin [27] who studied classical linear inverse problems, such as inverting the heat kernel, from this perspective. That paper focussed on the rational basis for deriving a regularization using the Bayesian approach, rather than on quantifying uncertainty, but the posterior (Gaussian in this case) distribution did indeed provide a quantification of uncertainty. However it is arguable that the work of Franklin was so far ahead of its time that it made little impact when it appeared, primarily because the computational power needed to approach practical problems from this perspec-

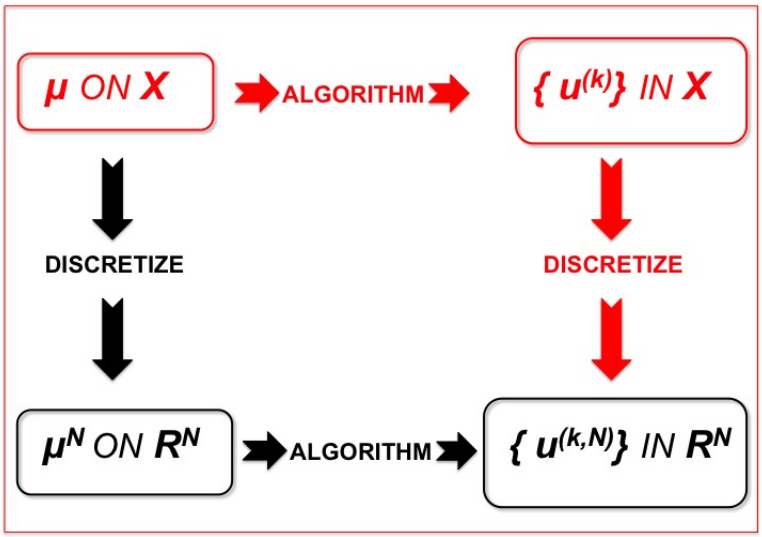

Figure 5. The red route is conceptually benefical in comparison with the black route.

tive was not available. The book of Kaipio and Somersalo [40] in 2005, however, had immediate impact, laying out a Bayesian methodology for inverse problems, and demonstrating its applicability to a range of important applications; computer power was ripe for the exploitation of fully Bayesian analyses when the book was published. However the perspective in [40] corresponded essentially to the black route outlined in Figure 5 ($N < \infty$) and did not take an infinite dimensional perspective in X.

In the interim between 1970 and 2005 there had been significant development of the theory of Bayesian inversion in X for linear problems, building on the work of Franklin [54, 50], and working directly in the infinite dimensional space X. Lasanen then developed this into a fully nonlinear theory [45, 46, 48, 49], also working on X. This theoretical work was not concerned directly with the development of practical algorithms and the need to interface computational Bayesian practice with numerical analysis; in particular the need to deal with limits $N \to \infty$ in order to represent elements of X was not addressed. However others within the Bayesian school of inverse problems were interested in this question; see, for example, the paper [51]. Furthermore, in contrast to classical inversion, which is (often by definition [25]) ill-posed, Bayesian inversion comes with a desirable well-posedness theory on X which, itself, underpins approximation theories [72]; we will survey some of the developments which come from this perspective in what follows. Cousins of this well-posedness theory on X may be found in the papers [55, 58] both of which consider issues relating to perturbation of the posterior, in the finite dimensional setting $N < \infty$.

The primary applications which drive the theoretical and algorithmic developments highlighted in this article are in subsurface geophysics and in the atmosphere-

ocean sciences. In the subsurface two major forces for the adoption of the Bayesian approach to inversion have been the work of Tarantola and co-workers and of Oliver and co-workers; see the books [76, 61] for further references. In the ocean-atmosphere sciences the Bayesian perspective has been less popular, but the book of Bennett [9] makes a strong case for it, primarily in the oceanographic context, whilst the work of Lorenc [53] has been a powerful force for Bayesian thinking in numerical weather prediction.

3. Examples

We provide in this section three examples to aid the reader who prefers concrete applications, and to highlight the type of problems which have motivated the theoretical develepments overviewed in the following sections. All of the examples can be placed in the general framework of (1.1).

3.1. Linear Inverse Problem. Consider the bounded linear map $K : X \to Y$, with X, Y separable Banach spaces, and the problem of finding $u \in X$ from noisy observations y of the image of u under K, given by

$$y = Ku + \eta.$$

For example if u is the initial condition of the heat equation on bounded open set $D \subset \mathbb{R}^d$, $X = L^2(D)$ and K denotes the solution operator for the heat equation over time interval T, then this is a widely used example of a classically ill-posed inverse problem. Ill-posedness arises beause of the smoothing property of the heat kernel and the fact that the noise η may take y out of the range space of K. Further ill-posedness can arise, for example, if K is found from the composition of the solution operator for the heat equation over time interval T with an operator comprising a finite set of point evaluations; the need to find a function u from a finite set of observations then leads to the problem being under-determined, further compounding ill-posedness. Linear inverse problems were the subject of the foundational paper [27], and developed further in [54, 50]. Natural applications include image processing.

3.2. Data Assimilation in Fluid Mechanics. A natural nonlinear generalization of the inverse problem for the heat equation, and one which is prototypical of the inverse problems arising in oceanography and weather forecasting, is the following. Consider the Navier-Stokes equation written as an ordinary differential equation in the Hilbert space $X := L^2_{\mathrm{div}}(\mathbb{T}^2)$ of square-integrable divergence-

free functions on the two-dimensional torus:

$$\frac{dv}{dt} + \nu Av + B(v,v) = f, \quad v(0) = u \in X.$$

This describes the velocity field $v(x,t)$ for a model of incompressible Newtonian flow [74] on a two-dimensional periodic domain. An inverse problem protoypical of weather forecasting in particular is to find $u \in X$ given noisy *Eulerian observations*

$$y_{j,k} = v(x_j, t_k) + \eta_{j,k}.$$

Like the heat equation the forward solution operator is smoothing, and the fact that the observations are finite in number further compounds ill-posedness. In addition the nonlinearity adds further complications, such as sensitive dependence on initial conditions arising from the chaotic character of the equations for $\nu \ll 1$. There are many interesting variants on this problem; one is to consider *Lagrangian observations* derived from tracers moving according to the velocity field v itself, and the problem is prototypical of inverse problems which arise in oceanography. Determining the initial condition of models from fluid mechanics on the basis of observations at later times is termed *data assimilation*. Both Eulerian and Lagrangian data assimilation are formulated as Bayesain inverse problems in [13].

3.3. Groundwater Flow. The following is prototypical of inverse problems arising in hydrology and in oil reservoir modelling. Consider the Darcy Flow, with log permeability $u \in X = L^\infty(D)$, described by the equation

$$\begin{aligned} -\nabla \cdot \big(\exp(u)\nabla p\big) &= 0, \quad x \in D, \\ u &= g, \quad x \in \partial D. \end{aligned}$$

Here the aim is to find $u \in X$ given noisy observations

$$y_j = p(x_j) + \eta_j.$$

The pressure p is a surrogate for the height of the water table and measurements of this height are made by hydrologists seeking to understand the earths subsurface. The resulting classical inverse problem is studied in [67] and Bayesian formulations are given in [21, 22]. The space $L^\infty(D)$ is not separable, but this difficulty can be circumvented by working in separable Banach spaces found as the closure of the linear span of an infinite set of functions in $L^\infty(D)$, with respect to the $L^\infty(D)$-norm.

4. Mathematical Foundations

In this section we briefly outline some of the issues involved in the rigorous formulation of Bayesian inversion on a separable Banach space X. We start by discussing

various prior models on X, and then discuss how Bayes' formula may be used to incorporate data and update these prior distributions on u into posterior distributions on $u|y$.

4.1. Priors: Random Functions. Perhaps the simplest way to construct random priors on a function space X is as follows. Let $\{\varphi_j\}_{j=1}^{\infty}$ denotes an infinite sequence in the Banach space X, normalized so that $\|\varphi_j\|_X = 1$. Define the deterministic random sequence $\gamma = \{\gamma_j\}_{j=1}^{\infty} \in \ell_w^p(\mathbb{R})$, where $\ell_w^p(\mathbb{R})$ denotes the sequence of $p^{th}-$power summable sequences, when weighted by the sequence $w = \{w_j\}_{j=1}^{\infty}$. Then let $\xi = \{\xi_j\}_{j=1}^{\infty}$ denote the i.i.d sequence of centred random variables in $\mathbb{R}$, normalized to that $\mathbb{E}\xi_1^2 = 1$. We define $u_j = \gamma_j \xi_j$ and pick a *mean* element $m \in X$ and then consider the random function

$$u = m + \sum_{j=1}^{\infty} u_j \varphi_j. \tag{4.1}$$

The probability measure on the random sequence implies, via its pushforward under the construction (4.1) a probability measure on the function u; we denote this measure by μ_0. Of course the fact that the φ_j are elements of X does not imply that μ_0 is a measure on X: assumptions must be made on the decay of the sequence γ. For example, using the fact that the random sequence $u = \{u_j\}_{j=1}^{\infty}$ comprises independent centred random variables we find that

$$\mathbb{E}^{\mu_0}\|u - m\|_X^2 = \sum_{j=1}^{\infty} \gamma_j^2.$$

This demonstrates that assuming $\gamma = \{\gamma_j\}_{j=1}^{\infty} \in \ell^2(\mathbb{R})$ is sufficient to ensure that the random function is almost surely an element of X. If the space X itself is not separable, this difficulty can be circumvented by working in a separable Banach space X' found as the closure of the linear span of the φ_j with respect to the norm in X.

Expansions of the form (4.1) go by the name Karhunen-Loeve in the Gaussian case [1] arising when ξ_1 is a Gaussian random variable. The so-called Besov case was introduced in [51] and concerns the case where ξ_1 is distributed according to Lebesgue density proportional to a power of $\exp(-|\cdot|^q)$, subsuming the Gaussian situation as the special case $q = 2$. Schwab has been a leading proponent of random functions constructed using compactly supported random variables ξ_1 – see [69, 71] and the references therein; although not so natural from an applications viewpoint, the simplicity that follows from this assumption allows the study of key issues in uncertainty quantification and Bayesian inversion without the need to deal with a variety of substantial technical issues which arise when ξ_1 is not compactly supported; in particular integrability of the tails becomes a key technical issue for non-compactly supported ξ_1, and there is a need for a Fernique theorem [26] or its analogue [51, 22]. For a general treatment of random functions constructed as in (4.1) see the book Kahane [37].

4.2. Priors: Hierarchical. There are many parameters required in the prior constructions of the previous subsection, and in many applications these may not be known. In such situations these parameters can be inferred from the data, along with u. Rather than giving a general discussion we consider the example of Gaussian priors when X is a Hilbert space. A draw u from a Gaussian is written as $u \sim N(m, C)$ where $N(m, C)$ denotes a Gaussian with mean m and covariance C. Here the covariance operator C is defined by

$$\begin{aligned} C &= \mathbb{E}^{\mu_0}(u-m) \otimes (u-m) \\ &= \sum_{j=1}^{\infty} \gamma_j^2 \varphi_j \otimes \varphi_j. \end{aligned}$$

Note that then

$$C\varphi_j = \gamma_j^2 \varphi_j.$$

An example hierarchical prior may be constructed by introducing an unknown parameter δ, which scales the covariance, and positing that

$$\begin{aligned} u|\delta &\sim N(0, \delta^{-1}C). \\ \delta &\sim \mathrm{Ga}(\alpha, \beta). \end{aligned}$$

Here Ga denotes the Gamma distribution, and of course other prior assumptions on δ are possible. The potential for the use of hierarchical priors in linear inverse problems has been highlighted in several recent papers, see [10, 11, 8] for example, all in the finite dimensional context; such models have been studied in the large dimension and infinite dimensional limit in [2].

4.3. Priors: Geometric. The probability measures constructed through random functions are inherently infinite dimensional, being built on an infinite sequence of random coefficients. In the previous subsection we showed how these could be extended to priors which included an extra unknown parameter δ specifying the scale of the prior; there are numerous generalizations of this basic concept. Here we describe one of them that is particularly useful in the study of subsurface inverse problems where the geometry imposed by faults, old fluival structures and so forth is a major determining fact in underground porous medium fluid flow.

Examples of problems to which our theory applies may be found in Figure 6. In the top left we show a layered structure in which a piecewise constant function is constructed; this maybe generalized to include faults, as in the bottom left. The top right shows a generalization of the layered structured to allow a different Gaussian random field realization in each layer, and the bottom right shows a generalization to allow for a channel-like structure, typical of fluvial deposition.

The development of layered prior models was pioneered in [12]. The chanellized structure as prior was developed in [44] and [79]. All of this work was finite dimensional, but a theoretical frameowork subsuming these particular cases, and set in infinite dimensions, is developed in [36].

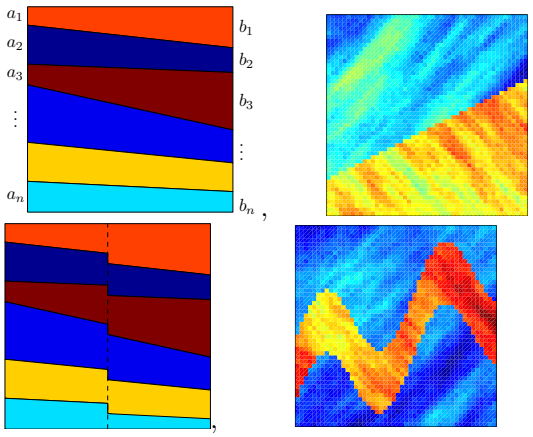

Figure 6. Uncertainty quantification under the prior and the posterior

.

4.4. Posterior. Recall that the Bayesuah solutiuon to the inverse problem of funding u from data y given by (1.1) is to determine the probability distribution on $u|y$, which lives on the space X, from the probability distribution of the joint random variable (u, y) which lives on $X \times Y$. In order to do this we specify to the situation where $Y = \mathbb{R}^J$, so that the number of observations is finite, and assume that $\eta \sim N(0, \Gamma)$, with Γ an invertible covariance matrix on $\mathbb{R}^J$. Many generalizations of this are possible, to both infinite dimensions or to non-Gaussian noise η, but the setting with fnite dimensional data allows us to expose the main ideas.

We define the *model-data mismatch functional*, or *least squares functional*, given by

$$\Phi(u; y) := \frac{1}{2}\left|\Gamma^{-\frac{1}{2}}\big(y - \mathcal{G}(u)\big)\right|^2$$

where $|\cdot|$ denotes the Euclidean norm. Classical Bayesian inversion is concerned with minimizing $\Phi(\cdot; y)$, typically with incoporation of regularization through addition of a penalty term (Tikhonov regularization) or through specification of seeking minimizers within a compact subset of X [25]. It is natural to ask how a Bayesian approach relates to such classical approaches.

Bayes' formula is typically stated as

$$\frac{\mathbb{P}(u|y)}{\mathbb{P}(u)} \propto \mathbb{P}(y|u)$$

and our wish is to formulate this precisely in the infinite dimensional context where u lives in a separable Banach space. Given a prior measure μ_0 on u and a posterior measure μ^y on $u|y$ a typical infinite dimensional version of Bayes' formula

is a statement that μ^y is absolutely continuous with respect to μ_0 and that

$$\frac{d\mu^y}{d\mu_0}(u) \propto \exp\Big(-\Phi(u;y)\Big). \tag{4.2}$$

Note that the righ-hand side is indeed proportional to $\mathbb{P}(y|u)$ whilst the left-hand side is an infinite dimensional analogue of $\frac{\mathbb{P}(u|y)}{\mathbb{P}(u)}$. The formula (4.2) implies that the posterior measure is large (resp. small), relative to the prior measure, on sets where $\Phi(\cdot;y)$ is small (resp. large). As such we see a clear link between classical inversion, which aims to choose elements of X which make $\Phi(\cdot;y)$ small, and the Bayesian approach.

There is a particular structure which occurs in the linear inverse problem of subsection 3.1, namely that if η is distributed according to a Gaussian, then the posterior on $u|y$ is Gaussian if the prior on u is Gaussian; the prior and posterior are termed *conjugate* in this situation, coming from the same class. See [42, 3] for a discussion of this Gaussian conjugacy for linear inverse problems in infinite dimensions.

4.5. Well-Posed Posterior. For a wide range of the priors and examples given previously there is a well-posedness theory which accompanies the Bayesian perspective. This theory is developed, for example, in the papers [72, 13, 21, 22, 36]. This theory shows that the posterior μ^y is Hölder in the Hellinger metric with respect to changes in the data y. The Hölder exponent depends on the prior, and is one (the Lipschitz case) for many applications. However it is important to strike a note of caution concerning the robustness of the Bayesian approach: see [63].

4.6. Recovery of Truth. Consider data y given from truth $\mathrm{u}^\dagger$ by

$$y = \mathcal{G}(\mathrm{u}^\dagger) + \epsilon\,\eta_0, \quad \eta_0 \sim N(0,\Gamma_0).$$

Thus we have assumed that the data is generated from the model used to construct the posterior. It is then natural to ask how close is the posterior measure μ^y to the truth $\mathrm{u}^\dagger$? For many of the preceding problems we have (refinements of) results of the type:

$$\text{For any } \delta > 0,\ \mathbb{P}^{\mu^y}\big(|u - \mathrm{u}^\dagger| > \delta\big) \to 0 \text{ as } \epsilon \to 0.$$

Examples of theories of this type may be found for linear problems of subsection 3.1 in [3, 4, 42, 43, 47, 66], for the Eulerian Navier-Stokes inverse problems of subsection 3.2 in [68], and for the groundwater flow problem of subsection 3.3 in [78].

5. Algorithms

The preceding chapter describes a range of theoretical developments which allow for precise characterizations of, and study of the properties of, the posterior distribution μ^y. These are interesting in their own right, but they also underpin algorithmic approaches which aim to be efficient with respect to increase of N in the approximation of μ^y by a measure $\mu^{y,N}$ on $\mathbb{R}^N$. Here we outline research in this direction.

5.1. Forward Error = Inverse Error. Imagine that we have approximated the space X by $\mathbb{R}^N$; for example we might truncate the expansion (4.1) at N terms and consider the inverse problem for the N unknown coefficients in the representation of u. We then approximate the forward map $\mathcal{G}$ by a numerical method to obtain $\mathcal{G}^N$ satisfying, for u in X,

$$|\mathcal{G}(u) - \mathcal{G}^N(u)| \leq \psi(N) \to 0$$

as $N \to \infty$. Such results are in the domain of classical numerical analysis. It is interesting to understand their implications for the Bayesian inverse problem.

The approximation of the forward map leads to an approximate posterior measure $\mu^{y,N}$ and it is natural to ask how expectations under μ^y, the ideal expectations to be computed, and under $\mu^{y,N}$, expectations under which we may approximate by, for example statistical sampling techniques, compare. Under quite general conditions it is possible to prove [18] that, for an appropriate class of test functions $f : X \to \mathsf{S}$, with S a Banach space,

$$\|\mathbb{E}^{\mu^y} f(u) - \mathbb{E}^{\mu^{y,N}} f(u)\|_{\mathsf{S}} \leq C\psi(N).$$

The method used is to employ the stability in the Hellinger metric implied by the well-posedness theory to show that μ^y and $\mu^{y,N}$ are $\psi(\mathcal{N})$ close in the Hellinger metric and then use properties of that metric to bound perturbations in expectations.

5.2. Faster MCMC. The preceding subsection demonstrates how to control errors arising from the numerical analysis component of any approximation of a Bayesian inverse problem. Here we turn to statistical sampling error, and in particular to Markov Chain-Monte Carlo (MCMC) methods. These methods were developed in the statistical physics community in [57] and then generalized to a flexible tool for statistical sampling in [34]. The paper [75] demonstrated an abstract framework for such methods on infinite dimensional spaces.

The full power of using MCMC methodology for inverse problems was highlighted in [40] and used for interesting applications in the subsurface in, for example, [24]. However for a wide range of priors/model problems it is possible to show that standard MCMC algorithms, derived by the black route in Figure 5, mix in

$\mathcal{O}(N^a)$ steps, for some $a > 0$ implying undesirable slowing down as N increases. By following the red route in Figure 5, however, it is possible to create new MCMC algorithms which mix in $\mathcal{O}(1)$ steps.

The slowing down of standard MCMC methods in high dimensions is demonstrated by means of diffusion limits in [56] for Gaussian priors and in [2] for hierarchical Gaussian priors. Diffusion limits where then used to demonstrate the effectiveness of the new method, derived via the red route in Figure 5, in [64] and a review explaining the derivation of such new methods maybe found in [19]. The paper [32] uses spectral gaps to both quantify the benefits of the method studied in [64] ($\mathcal{O}(1)$ lower bounds on the spectral gap) compared with the drawbacks of traditional methods, such as that studied in [56] ($\mathcal{O}(N^{-\frac{1}{2}})$ upper bounds on the spectral gap.)

These new MCMC methods are starting to find their way into use within large-scale engineering inverse problems and to be extended and modified to make them more efficient in large data sets, or small noise data sets scenarios; see for examples [29, 14, 20].

5.3. Other Directions. The previous subsection concentrated on a particular class of methods for exploring the posterior distribution, namely MCMC methods. These are by no means the only class of methods available for probing the posterior and here we give a brief overview of some other approaches that may be used.

The determinsitic approximation of posterior expectations, by means of sparse approximation of high dimensional integrals, is one approach with great potential. The mathematical theory behind this subject is overviewed in [69] in the context of standard uncertainty quantification, and the approach is extended to Bayesian inverse problems and uncertainty quantification in [71], with recent computational and theoretical progress contained in [70].

It is also possible to combine sparse approximation techniques with MCMC and the computational complexity of this approach is analyzed in [33], and references to the engineering literature, where this approach was pioneered, are given. The idea of multilevel Monte Carlo [30] has recently been generalized to MCMC methods; see the paper [33] which analyzes the computational complexity of such methods, the paper [41] in which a variant on such methods was introduced and implemented for the groundwater flow problem and the thesis [31] which introduced the idea of multilevel MCMC within the context of sampling conditioned diffusion processes.

Another computational approach, widely used in machine learning when complex probability measures need to be probed, is to look for the best approximation of μ^y within some simple class of measures. If the class comprises Dirac measures then such an approach is known as *maximum a posterior estimation* and corresponds in finite dimensions, when the posterior has a Lebesgue density, to finding the location of the peak of that density [40]. This idea is extended to the infinite dimensional setting in [23]. In the context of uncertainty quantification the MAP estimator itself is not of direct use as it contains no information about fluctuations.

However linearization about the MAP can be used to compute a Gaussian approximation at that point. A more sophisticated approach is to directly seek the best Gaussian approximation $\nu = N(m, C)$ wrt relative entropy. Analysis of this in the infinite dimensional setting, viewed as a problem in the calculus of variations, is undertaken in [65].

6. Conclusions

Combining uncertainty quantification with Bayesian inversion provides formidable computational challenges relating to the need to control, and optimally balance, errors arising from the numerical analysis, and approximation of the forward operator, with errors arising from computational statistical probing of the posterior distribution. The approach to this problem outlined here has been to adopt a way of deriving and analyzing algorithms based on thinking about them in infinite dimensional spaces, and only then discretizing to obtain implementable algorithms in $\mathbb{R}^N$ with $N < \infty$. This requires formulation and analysis of the Bayesian inverse problem in infinite dimensions. We have overviewed the mathematical theory that goes into this formulation and analysis, in section 3, and overviewed the algorithmic developments which follow from it, in section 4.

In some applications it is starting to be feasible to compute accurate approximations of the Bayesian posterior distribution, and it is to be expected that there will be great strides in this area over the next decade, both in terms of range of applications and algorithmic innovation, with the latter based on the infinite dimensional perspective given here, but making more careful exploitation of data and structure of the likelihood. Even where the fully Bayesian approach is out of the question for the forseeable future, for example in weather forecasting, the Bayesian approach described here can be important as it may be used as a gold standard against which to benchmark algorithms which are useable in practice. This approach is employed in [52, 35] in the context of model problems of the type shown in sections 3.2 and 3.3, and variants on them.

Finally the reader is reminded that this article is in essay form and contains no mathematical details. For an overview of the subject in which mathematical details are given the reader is refered to [73].

References

[1] R. Adler. *The Geometry Of Random Fields.* SIAM, 1981.

[2] S.Agapiou, J. Bardsley, O. Papaspiliopoulis and AM.Stuart Analysis of the Gibbs sampler for hierarchical inverse problems. `arxiv.org/abs/1311.1138.pdf`

[3] S. Agapiou, S. Larsson, and A. M. Stuart. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. Stochastic Processes and their Applications **123**(2013), 3828–3860.

[4] S. Agapiou, A. M. Stuart, and Y. X. Zhang. Bayesian posterior consistency for linear severely ill-posed inverse problems. To appear Journal of Inverse and Ill-posed Problems. `arxiv.org/abs/1210.1563`

[5] I. Babuska, R. Tempone and G. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. SIAM J. Num. Anal. **42**(2004), 800–825.

[6] I. Babuska, R. Tempone and G. Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. Applied Mechanics and Engineering, **194**(2005), 1251-1294.

[7] I. Babuska, F. Nobile and R. Tempone. A Stochastic Collocation method for elliptic Partial Differential Equations with Random Input Data. SIAM J. Num. Anal. **45**(2007), 1005–1034.

[8] J Bardsley. MCMC-Based image reconstruction with uncertainty quantification. SISC **34**(2012), 1316–1332.

[9] AF Bennett. Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press, 2002.

[10] D.Calvetti, H.Hakula, S.Pursiainen and E. Somersalo. Conditionally Gaussian Hypermodels for Cerebral Source Localization. SIAM J. Imaging Sciences **2**2009, 879909.

[11] D.Calvetti and E. Somersalo. Hypermodels in the Bayesian imaging framework. Inverse Problems **24**2008, 034013.

[12] J. Carter and D.White. History matching on the Imperial College fault model using parallel tempering. Computational Geosciences **17**(2013), 43–65.

[13] S. Cotter, M. Dashti, J. Robinson, and A. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems* **25**, (2009), doi:10.1088/0266–5611/25/11/115008.

[14] A.Cliffe, O.Ernst and B.Sprungk. In preparation, 2014.

[15] A.Cohen, R.DeVore and S.Schwab. Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. Foundations of Computational Mathematics **10**(2010), 615–646.

[16] A.Cohen, R.DeVore and S.Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs. Analysis and Applications **9**(2011), 11–47.

[17] A.Chkifa, A.Cohen, R.DeVore and S.Schwab. Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs. ESAIM: Mathematical Modelling and Numerical Analysis **47**(2013), 253–280.

[18] S. Cotter, M. Dashti, and A. Stuart. Approximation of Bayesian inverse problems. *SIAM Journal of Numerical Analysis* **48**(2010), 322–345.

[19] S. Cotter, G. Roberts, A. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science, to appear; arXiv:1202.0709* .

[20] T. Cui, KJH Law and Y.Marzouk. In preparation, 2014.

[21] M. Dashti and A. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem *SIAM J. Num. Anal.* **49**(2012), 2524-2542.

[22] M. Dashti, S. Harris, and A. Stuart. Besov priors for Bayesian inverse problems. *Inverse Problems and Imaging* **6**(2012), 183–200.

[23] M. Dashti, KJH.Law, AM Stuart and J.Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems Inverse Problems **29**(2013), 095017.

[24] P Dostert, Y Efendiev, TY Hou and W Luo Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification *Journal of Computational Physics* **217**(2006), 123–142.

[25] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer, 1996.

[26] X. Fernique. Intégrabilité des vecteurs Gaussiens. *C. R. Acad. Sci. Paris Sér. A-B* **270**, (1970), A1698–A1699.

[27] J. Franklin. Well-posed stochastic extensions of ill-posed linear problems. *J. Math. Anal. Appl.* **31**, (1970), 682–716.

[28] RG Ghanem and PD Spanos. *Stochastic Finite Elements: a Spectral Approach.* newblock Springer, 1991.

[29] O.Ghattas and T.Bui-Thanh An Analysis of Infinite Dimensional Bayesian Inverse Shape Acoustic Scattering and its Numerical Approximation, SIAM Journal on Uncertainty Quantification , Submitted, 2012.

[30] M.Giles Multilevel Monte Carlo path simulation. Operations Research **56**(2008), 607–617.

[31] D.Gruhlke. Convergence of Multilevel MCMC methods on path spaces. PhD Thesis, University of Bonn, 2013.

[32] M. Hairer, AM Stuart and S.Vollmer. Spectral Gaps for a Metropolis-Hastings Algorithm in Infinite Dimensions. To appear, Ann. Appl. Prob. 2014. `arxiv.org/abs/1112.1392`

[33] V.H Hoang, C. Schwab and AM Stuart Complexity analysis of accelerated MCMC methods for Bayesian inversion. Inverse Problems **29**(2013), 085010.

[34] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, no. 1, (1970), 97–109.

[35] M Iglesias, KJH Law and AM Stuart Evaluation of Gaussian approximations for data assimilation in reservoir models. Computational Geosciences **17**(2013), 851-885.

[36] M.Iglesias, K.Lin and AM Stuart Well-Posed Bayesian Geometric Inverse Problems Arising in Subsurface Flow. `arxiv.org/abs/1401.5571`.

[37] J.-P. Kahane. *Some random series of functions*, vol. 5 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1985.

[38] J.B. Keller. Inverse problems. *Am. Math. Mon.* **83**(1976), 107–118.

[39] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems.* Springer, 1996.

[40] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, vol. 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005.

[41] C.Ketelsen, R.Scheichl and Teckentrup. A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. `arxiv.org/abs/1303.7343`

[42] B. Knapik, A. van Der Vaart, and J. van Zanten. Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39**, no. 5, (2011), 2626–2657.

[43] B. Knapik, A. van der Vaart, and J. van Zanten. Bayesian recovery of the initial condition for the heat equation `arxiv.org/abs/1111.5876`.

[44] JL Landa and RN Horne A procedure to integrate well test data, reservoir performance history and 4-D seismic information into a reservoir description. SPE Annual Technical Conference 1997, 99–114.

[45] S. Lasanen. Discretizations of generalized random variables with applications to inverse problems. *Ann. Acad. Sci. Fenn. Math. Diss., University of Oulu* **130**.

[46] S. Lasanen. Measurements and infinite-dimensional statistical inverse theory. *PAMM* **7**, (2007), 1080101–1080102.

[47] S. Lasanen. Posterior convergence for approximated unknowns in non-Gaussian statistical inverse problems. *Arxiv preprint arXiv:1112.0906* .

[48] S. Lasanen. Non-Gaussian statistical inverse problems. Part I: Posterior distributions. *Inverse Problems and Imaging* **6**, no. 2, (2012), 215–266.

[49] S. Lasanen. Non-Gaussian statistical inverse problems. Part II: Posterior distributions. *Inverse Problems and Imaging* **6**, no. 2, (2012), 267–287.

[50] M. S. Lehtinen, L. Päivärinta, and E. Somersalo. Linear inverse problems for generalised random variables. *Inverse Problems* **5**, no. 4, (1989), 599–612. http://stacks.iop.org/0266-5611/5/599.

[51] M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Problems and Imaging* **3**, (2009), 87–122.

[52] KJH Law and AM Stuart Evaluating data assimilation algorithms. Monthly Weather Review **140**(2012), 3757-3782.

[53] AC Lorenc and O Hammon Objective quality control of observations using Bayesian methods. Theory, and a practical implementation. *Quarterly Journal of the Royal Meteorological Society*, **114**(1988), 515–543.

[54] A. Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrsch. Verw. Gebiete* **65**, no. 3, (1984), 385–397. http://dx.doi.org/10.1007/BF00533743.

[55] Y. Marzouk and D Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. *Communications in Computational Physicss* **6**(2009), 826-847.

[56] J. Mattingly, N. Pillai, and A. Stuart. Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann. Appl. Prob* **22**(2012), 881–930.

[57] N. Metropolis, R. Rosenbluth, M. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, (1953), 1087–1092.

[58] A. Mondal, Y. Efendiev, B. Mallick and A. Datta-Gupta, Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump Markov chain Monte Carlo methods. *Advances in Water Resources*, **3**(2010), 241–256.

[59] F. Nobile, R. Tempone and CG Webester, A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, **46**(2008), 2309–2345.

[60] F. Nobile, R. Tempone and CG Webester, An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, **46**(2008), 2441–2442.

[61] DS Oliver, AC Reynolds and N Liu Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press 2008.

[62] H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns and M. Ortiz. Optimal Uncertainty Quantification. SIAM Review **55**(2013), 271345.

[63] H. Owhadi, C. Scovel and T. J. Sullivan. When Bayesian Inference Shatters. `arxiv.org/abs/1308.6306`

[64] N. Pillai, A. Stuart, and A. Thiery. Gradient flow from a random walk in Hilbert space. To appear, Stochastic Partial Differential Equations. `arxiv.org/abs/1108.1494`.

[65] F.Pinski, G. Simpson, AM Stuart and H Weber. Kullback-Leibler Approximation for Probability Measures on Infinite Dimensional Spaces. `arxiv.org/abs/1310.7845`

[66] K.Ray Bayesian inverse problems with non-conjugate priors Electronic Journal of Statistics, **7**(2013), 1-3169.

[67] G. Richter. An inverse problem for the steady state diffusion equation. *SIAM Journal on Applied Mathematics* **41**, no. 2, (1981), 210–221.

[68] D. Sanz-Alonso and A.M. Stuart In preparation, 2014.

[69] C. Schwab and CJ Gittelson. *Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. Acta Numer.* **20**, (2011).

[70] C.Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems Inverse Problems **29**(2013), 065011.

[71] C. Schwab and A. Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems* **28**, (2012), 045003.

[72] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.* **19**, (2010), 451–559.

[73] A. M. Stuart. The Bayesian approach to inverse problems. `arxiv.org/abs/1302.6989`

[74] R.Temam. Navier-Stokes equations. AMS Chelsea Publishing, Providence, RI, 2001.

[75] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8**, no. 1, (1998), 1–9.

[76] A. Tarantola *Inverse Problem Theory.* SIAM, 2005.

[77] R.Tempone Numerical Complexity Analysis of Weak Approximation of Stochastic Differential Equations. PhD Thesis, KTH Stockholm, Sweden, 2002. `http://www.nada.kth.se/utbildning/forsk.utb/avhandlingar/dokt/Tempone.pdf`

[78] S.Vollmer Posterior consistency for Bayesian inverse problems through stability and regression results. Inverse Problems **29**(2013), 125011.

[79] J.Xie, Y.Efendiev and Datta-Gupta. Uncertainty quantification in history matching of channelized reservoirs using Markov chain level set approaches. SPE Reservoir Simulation Symposium, 2011.

Mathematics Institute, Warwick University, Coventry CV4 7AL, UK
E-mail: a.m.stuart@warwick.ac.uk