Advances in Water Resources 110 (2017) 166–181

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

On uncertainty quantification in hydrogeology and hydrogeophysics

Niklas Linde[*,a], David Ginsbourger[b,c], James Irving[a], Fabio Nobile[d], Arnaud Doucet[e]

[a] Environmental Geophysics Group, Institute of Earth Sciences, University of Lausanne, Lausanne 1015, Switzerland
[b] Uncertainty Quantification and Optimal Design group, Idiap Research Institute, Centre du Parc, Rue Marconi 19, PO Box 592, Martigny 1920, Switzerland
[c] Institute of Mathematical Statistics and Actuarial Science, Department of Mathematics and Statistics, University of Bern, Alpeneggstrasse 22, Bern 3012, Switzerland
[d] Calcul Scientifique et Quantification de l'Incertitude, Institute of Mathematics, Ecole polytechnique fédérale de Lausanne, Station 8, CH 1015, Lausanne, Switzerland
[e] Department of Statistics, Oxford University, 24-29 St Giles', Oxford, OX1 3LB, United Kingdom

ARTICLE INFO

Keywords:
Uncertainty quantification
Hydrogeology
Hydrogeophysics
Inversion
Proxy models
Modeling errors
Petrophysics

ABSTRACT

Recent advances in sensor technologies, field methodologies, numerical modeling, and inversion approaches have contributed to unprecedented imaging of hydrogeological properties and detailed predictions at multiple temporal and spatial scales. Nevertheless, imaging results and predictions will always remain imprecise, which calls for appropriate uncertainty quantification (UQ). In this paper, we outline selected methodological developments together with pioneering UQ applications in hydrogeology and hydrogeophysics. The applied mathematics and statistics literature is not easy to penetrate and this review aims at helping hydrogeologists and hydrogeophysicists to identify suitable approaches for UQ that can be applied and further developed to their specific needs. To bypass the tremendous computational costs associated with forward UQ based on full-physics simulations, we discuss proxy-modeling strategies and multi-resolution (Multi-level Monte Carlo) methods. We consider Bayesian inversion for non-linear and non-Gaussian state-space problems and discuss how Sequential Monte Carlo may become a practical alternative. We also describe strategies to account for forward modeling errors in Bayesian inversion. Finally, we consider hydrogeophysical inversion, where petrophysical uncertainty is often ignored leading to overconfident parameter estimation. The high parameter and data dimensions encountered in hydrogeological and geophysical problems make UQ a complicated and important challenge that has only been partially addressed to date.

1. Introduction

The subsurface environment is highly heterogeneous and non-linear coupled processes take place at multiple spatial and temporal scales. Valuable information about subsurface structures and processes can be obtained from borehole measurements, outcrops, laboratory analysis of field samples, and from geophysical and hydrogeological experiments; however, this information is largely incomplete. It is critical that basic scientific studies and management decisions for increasingly complex engineering challenges (e.g., enhanced geothermal systems, carbon capture and storage, nuclear waste repositories, aquifer storage and recovery, remediation of contaminated sites) account for this incompleteness in our system understanding. This enables us to consider the full range of possible future outcomes, to base scientific findings on solid grounds and to target future investigations. Nevertheless, uncertainty quantification (UQ) is highly challenging because it attempts to quantify what we do not know. For example, it is extremely difficult to properly describe prior information about a hydrogeological system, to accurately quantify complex error characteristics in our data, and to quantify model errors caused by incomplete physical, chemical, and biological theories.

Eloquent arguments have been put forward to explain why numerical models in the Earth Sciences cannot be validated (Konikow and Bredehoeft, 1992; Oreskes et al., 1994). These arguments are based on Popperian viewpoints (Tarantola, 2006) and on the recognition that natural subsurface systems are open and inherently under-sampled. This implies that UQ in the Earth Sciences can never be considered to be complete. Instead, it should be viewed as a partial assessment that is valid for a given set of prior assumptions, hypotheses, and simplifications. With this in mind, UQ in terms of probability distributions, often characterized in terms of probability density functions (pdfs), can still greatly help to make informed decisions regarding, for example, strategies for mitigating the effects of climate change, how to best exploit natural resources, how to minimize exposure to environmental pollutants, and how to protect environmental goods such as clean groundwater.

This review focuses on UQ in hydrogeology and hydrogeophysics. Using the term UQ, we refer both to (i) the forward UQ problem,

* Corresponding author.
E-mail address: niklas.linde@unil.ch (N. Linde).

http://dx.doi.org/10.1016/j.advwatres.2017.10.014
Received 4 May 2017; Received in revised form 11 October 2017; Accepted 13 October 2017
Available online 16 October 2017
0309-1708/

namely how to characterize the distribution of output variables of interest (e.g., to determine the risk of contamination in a water supply well) given a distribution of input variables (e.g., subsurface material properties); and (ii) the solution of the Bayesian inverse UQ problem, whereby prior knowledge is merged with (noisy) observational data and numerical modeling in order to obtain a posterior distribution for the input variables. Note that it is beyond the scope of this work to make an exhaustive review of UQ or to present all existing and potential applications in hydrogeology and hydrogeophysics. Rather, we try to connect a number of recent methodological advances in UQ with selected contemporary challenges in hydrogeology and hydrogeophysics. The mathematical development and the description of the methods are kept to a minimum and ample references are provided for further reading. We emphasize general methods that do not necessarily rely upon linearizations or Gaussian assumptions. The price to pay for this generality is a substantial increase in computational cost, which is reflected by the fact that more approximate approaches are presently favored (e.g., Ensemble Kalman filters (Evensen, 2009), quasi-static linear inversion (Kitanidis, 1995)). Clearly, these approximate methods are not only used because they are comparatively fast, but also because they have shown to produce useful and robust results in a wide range of application areas.

After introducing the main concepts and notations (Section 2), we discuss the definition of prior distributions for spatially distributed parameter fields (Section 3.1). This is followed by a discussion on the role of proxy models in forward UQ (Section 3.2), after which we present how Multi-Level Monte Carlo and related techniques can be used within forward UQ to propagate prior uncertainties into quantities of interest (Section 3.3). Next, we consider the Bayesian inverse problem where we examine likelihood functions (Section 4.1) and discuss sampling approaches with an emphasis on particle methods (Section 4.2). This is followed by an outlook towards how to best account for model errors (Section 5.1) and petrophysical-relationship uncertainty in hydrogeophysical inversions (Section 5.2).

2. Main concepts and notations

In hydrogeology, it is often desirable to predict and characterize uncertainties on Quantities of Interest (QoI) given a set of inputs described by a multivariate parameter $\mathbf{u}$. Depending on the problem, $\mathbf{u}$ may refer to a vector, a field, a more general function, or combinations thereof; here, without loss of generality, we use the "field" as a generic term to denote $\mathbf{u}$. As an example, $\mathbf{u}$ may represent a permeability field and a contaminant source region, and the QoI may be the contaminant concentration in a water supply well at some future time. In this case, the forward model that links the two would typically be a numerical solver of the advection-dispersion equation for some set of (possibly uncertain) boundary and initial conditions. Herein, $\mathbf{u}$ is treated either as a discretized (finite-dimensional) or continuous (infinite-dimensional) object. This distinction might seem superfluous at first because discretization is always needed at some stage when dealing with numerical forward models; however, considering an infinite-dimensional formalism can be highly relevant as discussed later.

A given QoI, denoted by $\mathbf{Q}$, is a function of the output from the considered *solution map* (in practice, the output of a numerical simulator), formalized as a deterministic function $\mathscr{R}: \mathbf{u} \mapsto \mathscr{R}(\mathbf{u})$ that is generally non-linear. Here, we use $\mathscr{Q}$ for the function mapping $\mathbf{u}$ to $\mathbf{Q}$. This function can be formulated as $\widetilde{\mathscr{Q}} \circ \mathscr{R}$ for some function $\widetilde{\mathscr{Q}}$ as $\mathbf{Q}$ is assumed to depend on $\mathbf{u}$ solely via $\mathscr{R}(\mathbf{u})$ so that $\mathbf{Q} = \mathscr{Q}(\mathbf{u}) = \widetilde{\mathscr{Q}}(\mathscr{R}(\mathbf{u}))$.

In essence, the probabilistic approach to forward UQ consists of endowing the considered set of $\mathbf{u}$'s with a probability distribution μ_0, and propagating this distribution to $\mathbf{Q}$ by using uncertainty-propagation techniques. The standard means of doing this, referred to as the basic Monte-Carlo method, consists of drawing a sample $\{\mathbf{u}_1, \cdots, \mathbf{u}_N\}$ from μ_0, calculating the corresponding sample $\{\mathscr{Q}(\mathbf{u}_1), \cdots, \mathscr{Q}(\mathbf{u}_N)\}$, and empirically approximating expectations of functions of $\mathbf{Q}$ under the discrete probability distribution $\frac{1}{N}\sum_{i=1}^{N} \delta_{\mathscr{Q}(\mathbf{u}_i)}$.

Practical and theoretical work over the past decade has focused on how to best account for imperfect numerical modeling (see Section 3.2), for instance via error models, and how to take advantage of multiple numerical models with different levels of fidelity and computation times (see Section 3.3). Overall, propagating uncertainties in the inputs, accounting for imperfect numerical modeling, and addressing real-world problems using statistical procedures and numerical models are broadly considered as part of *uncertainty propagation* or *forward UQ*.

Inverse problems have played an important role in applied mathematics for more than a century and are of crucial importance in hydrogeology (e.g., Carrera et al., 2005; McLaughlin and Townley, 1996; Zhou et al., 2014) and geophysics (e.g., Menke, 2012; Parker, 1994; Tarantola, 2005). The starting point when solving an inverse problem is to write the relation linking observed data $\mathbf{y}$ to model parameters $\mathbf{u}$

$$\mathbf{y} = \mathscr{G}(\mathbf{u}) + \epsilon, \tag{1}$$

where the *forward map* $\mathscr{G}$: $\mathbf{u} \mapsto \mathscr{G}(\mathbf{u})$ can be viewed as the combination of a *solution map* $\mathscr{R}$ and an *observation map* $\mathscr{O}$ that returns $n \geq 1$ functionals of $\mathscr{R}(\mathbf{u})$ (typically linear forms, such as point-wise evaluations at specific locations and/or times), and ϵ typically stands for observational noise. In simpler terms, $\mathscr{O}$ extracts from the output of the solution map the information that is needed to calculate the forward responses $\mathscr{G}(\mathbf{u}) = \mathscr{O}(\mathscr{R}(\mathbf{u}))$, that are to be compared with the observed data $\mathbf{y}$.

For example, $\mathbf{u}$ may stand for lithological properties of an aquifer, with $\mathscr{R}$ returning the space-time evolution of contaminant concentration within this aquifer. The corresponding $\mathscr{O}$ could indicate concentrations at specific well locations and times, and the inverse problem would then consist of recovering the unknown lithology from noisy measurements $\mathbf{y}$ at these locations. In practice, $\mathscr{G}$ is the best possible numerical prediction of an experiment, but it is never a perfect map in a strict mathematical sense. This implies that virtually all $\mathscr{G}$'s in the geosciences could be considered as proxy models (see Section 3.2) and we use $\mathscr{G}$ herein when referring to high-fidelity forward simulations. While we do not explicitly consider ϵ terms that incorporate model errors at this stage, the topic is implicitly tackled in forthcoming sections on likelihood functions and error modeling.

The inherent inaccuracies of forward solvers $\mathscr{G}$ have two origins. First, geological and physical heterogeneity are present at all scales, but numerical forward solvers can only handle heterogeneity up to a given spatial (e.g., model cell size) or spectral (e.g., truncation of spherical harmonics) resolution. The impact of limited resolution on simulation results depends strongly on the physics involved. For example, predicted gravimetric or groundwater-level responses will be comparatively insensitive, whereas seismic or ground penetrating radar (GPR) full-waveform modeling or tracer transport simulation results may be highly sensitive (Dentz et al., 2011). Second, considerable simplifications of the underlying physics are often made, even when using the most advanced simulation algorithms. The needed simplifications and their impacts are strongly problem dependent. For instance, gravimetric modeling can be performed using physical descriptions that are highly accurate, whereas GPR forward modeling typically does not account for the well-known frequency-dependence of subsurface electrical properties or the finite sizes of transmitter and receiver antennas (Klotzsche et al., 2013). Furthermore, the accuracy of $\mathscr{G}$ for a given physical description and model domain depends also on the numerical schemes (e.g., in time) and equation solvers (e.g., iterative, direct) employed. Despite these simplifications, evaluating $\mathscr{G}(\mathbf{u})$ (i.e., solving the forward problem) often leads to significant computing times (e.g., Fichtner, 2010; Geiger et al., 2004), which limits the number of forward simulations that can be practically considered.

In hydrogeology and geophysics, $\mathbf{u}$ is generally high-dimensional, $\mathscr{G}$ is costly to evaluate and non-linear, and the size of $\mathbf{y}$ is limited by data acquisition constraints. Bayesian inversion (the inverse UQ problem) provides a framework to make inferences on $\mathbf{u}$ from observations $\mathbf{y}$ by