Advances in Water Resources 86 (2015) 86–101

Contents lists available at ScienceDirect

Advances in Water Resources

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

Geological realism in hydrogeological and geophysical inverse modeling: A review

Niklas Linde [a,*], Philippe Renard [b], Tapan Mukerji [c], Jef Caers [c]

[a] Applied and Environmental Geophysics Group, Institute of Earth Sciences, University of Lausanne, Géopolis - bureau 3779, CH-1015 Lausanne, Switzerland
[b] Stochastic Hydrogeology Group, Centre for Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Switzerland
[c] Stanford Center for Reservoir Forecasting, Department of Energy Resources Engineering, School of Earth Sciences, Stanford University, California.

ARTICLE INFO

Article history:
Received 22 May 2015
Revised 15 September 2015
Accepted 23 September 2015
Available online 5 October 2015

Keyword:
Geostatistics
Inverse problems
Hydrogeology
Geophysics
Hydrogeophysics
Markov chain Monte Carlo

ABSTRACT

Scientific curiosity, exploration of georesources and environmental concerns are pushing the geoscientific research community toward subsurface investigations of ever-increasing complexity. This review explores various approaches to formulate and solve inverse problems in ways that effectively integrate geological concepts with geophysical and hydrogeological data. Modern geostatistical simulation algorithms can produce multiple subsurface realizations that are in agreement with conceptual geological models and statistical rock physics can be used to map these realizations into physical properties that are sensed by the geophysical or hydrogeological data. The inverse problem consists of finding one or an ensemble of such subsurface realizations that are in agreement with the data. The most general inversion frameworks are presently often computationally intractable when applied to large-scale problems and it is necessary to better understand the implications of simplifying (1) the conceptual geological model (e.g., using model compression); (2) the physical forward problem (e.g., using proxy models); and (3) the algorithm used to solve the inverse problem (e.g., Markov chain Monte Carlo or local optimization methods) to reach practical and robust solutions given today's computer resources and knowledge. We also highlight the need to not only use geophysical and hydrogeological data for parameter estimation purposes, but also to use them to falsify or corroborate alternative geological scenarios.

1. Introduction

Geophysical data help to understand geological processes and to test scientific hypotheses throughout the Earth Sciences, while also providing critical information and constraints for forecasting and management of subsurface formations (e.g., oil and gas reservoirs, mineral prospects, aquifers, and the critical zone). The processing of virtually all geophysical surveys involves inversion, a computational process in which measurement responses (e.g., signals in time and space for seismic and electromagnetic data) are translated into multi-dimensional images of physical properties (e.g., seismic wavespeed, density, electrical conductivity) [124,165] or into properties of direct relevance for geological applications (e.g., lithotype, porosity, fluid saturation) [11–13]. Subsurface heterogeneity, signal attenuation, averaging inherent to the underlying physics (e.g., diffusion), incomplete data coverage and noisy data limit the scale at which these properties can be resolved [6].

Solute transport in the subsurface can be highly sensitive to geological features (e.g., fractures [2] or connected high conductivity forms [174]) at scales below the resolution limits offered by geophysical sensing. Resolution-limited geophysical models alone are thus often inadequate for applications related to mass transfer in the subsurface (oil, gas, water). Even if improved geophysical acquisition systems and imaging algorithms allow resolving ever-finer details, fundamental resolution limits persist. At the high resolution necessary for flow- and transport modeling, the geophysical inverse problem has a possibly infinite set of solutions.

This non-uniqueness is traditionally overcome by using an optimization approach with a model regularization term, thereby focusing solely on model features that are necessary to explain the geophysical data [32]. Such a regularization term generally lacks geological justification and results in blurry models that are overly smooth and geologically unrealistic [42]. One step forward is to artificially introduce fine-scale information by adapting multi-Gaussian geostatistical models that describe the correlation between two points in space throughout the volume of investigation (we refer to Chilès and Delfiner [23] for a general introduction to geostatistics). However, similar to the overly smooth models obtained by regularized inversion, the multi-Gaussian framework is often insufficient to

* Corresponding author. Tel.: +41 21 692 4401; fax : +41 21 692 44 05.
E-mail address: Niklas.Linde@unil.ch (N. Linde).

http://dx.doi.org/10.1016/j.advwatres.2015.09.019
0309-1708/

describe realistic geological structure and especially those impacting flow responses [49,57,89,95].

In many cases, the measured hydrogeological or geophysical data can be complemented by ancillary information on the heterogeneity of subsurface formations that is obtained from borehole data, analog outcrops or databases of previously studied sites. Expert knowledge is also important. For example, sedimentologists may provide geological descriptions of the architecture of rock facies, their mutual spatial relationships, geometrical constraints or rules of deposition. In applications where supporting data are sparse and the geological context is unclear, it is perhaps even more important to assimilate and formally test competing conceptual geological models [49,109,134].

This review describes existing approaches to incorporate prior geological understanding in the inversion of geophysical and hydrogeological data to better predict subsurface flow- and transport processes at relevant temporal and spatial scales. This assimilation problem is at the forefront of many exploration, environmental, and research challenges of relevance for the Earth Sciences. Research in the area is very active, but publications are widely spread over various discipline journals with little interaction across disciplines (e.g., oil/gas vs. groundwater). Only a few attempts have been made to bridge these community gaps (e.g., [82]).

The presentation is structured as follows. Section 2 formulates the inverse problem as the integration of the information offered by geophysical and hydrogeological data, their relationship, and an underlying conceptual Earth model. Section 3 describes approaches to create geologically realistic priors and how to generate geologically realistic realizations by sampling this prior. Section 4 introduces approaches on how to parameterize models and propose model updates that are representative samples of a geologically realistic prior. Section 5 reviews how the inverse problem can be solved in the general case using sampling techniques and under more approximate conditions using stochastic search and optimization. Section 6 proposes two alternative strategies for bringing the various pieces (Sections 3–5) together in solving practical field cases. Section 7 provides concluding remarks.

2. The inverse problem

2.1. General formulation

Tarantola and Valette [167] formulated the general nonlinear inverse problem as a combination of the information provided by N data, $\mathbf{d}$, by a priori information about M model parameters, $\mathbf{m}$, and by theories that relate the two $p(\mathbf{m}, \mathbf{d})$. In the following, a slightly less general formulation is considered that is based on a traditional Bayesian framework [86].

The posterior probability density function (pdf) $p(\mathbf{m}|\mathbf{d})$ is

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}, \tag{1}$$

where $L(\mathbf{m}|\mathbf{d}) \equiv p(\mathbf{d}|\mathbf{m})$ is the likelihood function that typically summarizes the statistical properties of the error residuals between observed and simulated data and $p(\mathbf{m})$ is the prior pdf. The evidence $p(\mathbf{d})$ is important for model selection and averaging, but it can be neglected when considering a fixed model parameterization. In this case, the unnormalized density suffices

$$p(\mathbf{m}|\mathbf{d}) \propto L(\mathbf{m}|\mathbf{d})p(\mathbf{m}). \tag{2}$$

The solution to the inverse problem can be represented as a closed-form expression of $p(\mathbf{m}|\mathbf{d})$, an approximation based on samples from this distribution or one representative model obtained by optimization.

2.2. The likelihood

The forward problem consists of simulating the data response $\mathbf{d}^{\mathrm{sim}}$ of a proposed model $\mathbf{m}^{\mathrm{prop}}$

$$\mathbf{d}^{sim} = g(\mathbf{m}^{prop}). \tag{3}$$

The forward simulator $g(\,-\,)$ typically involves numerical simulations based on a physical theory (e.g., the advection–dispersion equation to predict tracer breakthrough curves or the electromagnetic wave equation to simulate ground-penetrating radar responses).

Assuming that measurement and modeling errors follow a Gaussian distribution, the likelihood function is

$$L(\mathbf{m}|\mathbf{d}) = \frac{1}{(2\pi)^{N/2} det(\mathbf{C}_{\mathrm{D}})^{1/2}} \times exp\left(-\frac{1}{2}(g(\mathbf{m}) - \mathbf{d} - \mathbf{b}_{\mathrm{D}})^{T}\mathbf{C}_{\mathrm{D}}^{-1}(g(\mathbf{m}) - \mathbf{d} - \mathbf{b}_{\mathrm{D}})\right), \tag{4}$$

where $\mathbf{C}_{\mathrm{D}}$ is a covariance matrix given by the sum of the covariance matrices describing modeling $\mathbf{C}_{\mathrm{T}}$ and observational errors $\mathbf{C}_{\mathrm{d}}$ (e.g., [165]) and $\mathbf{b}_{\mathrm{D}} = \mathbf{b}_{\mathrm{T}} + \mathbf{b}_{\mathrm{d}}$ describing bias terms associated with modeling and observational error distributions that are not centered on zero [64].

It is common practice to assume that both data and modeling errors are uncorrelated, thus, making $\mathbf{C}_{\mathrm{D}}$ a diagonal matrix. This choice is often made out of convenience and because it is challenging to determine proper error models of field data (observational and geometrical errors) and forward solvers (simplified physics, numerical approximations, effects of parameterization, etc.). Gaussian error models are very sensitive to outliers and alternative distributions, for example, symmetric exponentials may provide more robust results (e.g., [26]). Furthermore, replacing $\mathbf{C}_{\mathrm{D}}$ with a diagonal matrix and ignoring bias terms can lead to important inversion artifacts [64], but determining $\mathbf{C}_{\mathrm{D}}$ and $\mathbf{b}_{\mathrm{D}}$ can be very challenging in practice. One approach is to use a computationally expensive, but physically correct forward simulator, to build an error model that is used in subsequent inversions that rely on simplified forward models [64]. Another approach is to approximate these errors with an assumed functional form, while inferring parameter values (e.g., those in an autoregressive model) during the inversion process [155].

Furthermore, statistical rock physics models can be included in the likelihood function (e.g., [41]) to link physical properties (sensed by geophysical data) and hydrogeological target properties. These relationships are often more straightforward when dealing with time-lapse data (i.e., monitoring of geophysical variables over time). Statistical rock physics is an area of active research. At present, the spatial support and correlation of the scatter in rock physics relationships, their scaling as a function of observational scale, and how parameters vary in space are often largely unknown.

2.3. The prior

In its simplest form, the M model parameters refer to material properties in a regular mesh. In this case, the standard multi-Gaussian description of the prior pdf $p(\mathbf{m})$ takes a similar form as the likelihood function [165]

$$p(\mathbf{m}) = \frac{1}{(2\pi)^{M/2} det(\mathbf{C}_{\mathrm{M}})^{1/2}} exp\left(-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^{T}\mathbf{C}_{\mathrm{M}}^{-1}(\mathbf{m} - \mathbf{m}_0)\right), \tag{5}$$

with $\mathbf{C}_{\mathrm{M}}$ the model covariance matrix describing the spatial correlation between model cells and $\mathbf{m}_0$ the expected value of the model parameters. Assuming a multi-variate distribution of the prior will strongly influence the spatial characteristics of the posterior solutions. A Gaussian prior with a Gaussian likelihood function leads, in the linear case, to an explicit pdf for the posterior which is also Gaussian (e.g., [165]). Similarly, a Gaussian mixture prior with a Gaussian