Advances in Water Resources 111 (2018) 346–359

Contents lists available at ScienceDirect

Advances in Water Resources

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

Original Article

Impact of petrophysical uncertainty on Bayesian hydrogeophysical inversion and model selection

Carlotta Brunetti*, Niklas Linde

Applied and Environmental Geophysics Group, Institute of Earth Sciences, University of Lausanne, Lausanne 1015, Switzerland

ARTICLE INFO

Keywords:
Petrophysical uncertainty
Hydrogeophysics
Bayesian model selection
Bayesian inversion
Evidence
Conceptual model

ABSTRACT

Quantitative hydrogeophysical studies rely heavily on petrophysical relationships that link geophysical properties to hydrogeological properties and state variables. Coupled inversion studies are frequently based on the questionable assumption that these relationships are perfect (i.e., no scatter). Using synthetic examples and crosshole ground-penetrating radar (GPR) data from the South Oyster Bacterial Transport Site in Virginia, USA, we investigate the impact of spatially-correlated petrophysical uncertainty on inferred posterior porosity and hydraulic conductivity distributions and on Bayes factors used in Bayesian model selection. Our study shows that accounting for petrophysical uncertainty in the inversion (I) decreases bias of the inferred variance of hydrogeological subsurface properties, (II) provides more realistic uncertainty assessment and (III) reduces the overconfidence in the ability of geophysical data to falsify conceptual hydrogeological models.

1. Introduction

A primary goal in hydrogeophysical studies is often to infer quantitative hydrogeological models from geophysical and any available hydrogeological data. Unfortunately, petrophysical relationships describing links between geophysical properties and hydrogeological parameters and state variables are uncertain and the information content of hydrogeophysically-inferred estimates is significantly affected by their predictive power. We distinguish here between three types of uncertainty in petrophysical (also called rock physics) models: (1) *petrophysical model uncertainty* refers to uncertainty about the most appropriate parametric form (e.g., Archie's law, time propagation model, Wyllie's formula), (2) *petrophysical parameter uncertainty* relates to uncertainty about the most appropriate parameter values (e.g., cementation index, saturation exponent), and (3) *petrophysical prediction uncertainty* describes the scatter and bias around the calibrated petrophysical model (e.g., dispersion around predictions based on Archie's law). These three types of uncertainty are clearly not independent of each other. For instance, petrophysical prediction uncertainty is described by the residuals between the actual prediction quantity (e.g., porosity, hydraulic conductivity) and the predictions for a given petrophysical model and parameter values.

To date, most focus in hydrogeophysical inversion has been on petrophysical parameter uncertainty (e.g., Kowalsky et al., 2005; Lochbühler et al., 2014) with the petrophysical parameter values being inferred (deterministically or probabilistically) as a part of the inversion process. However, ignoring the other two types of uncertainty may lead to biased estimates and unrealistically low uncertainty estimates. For instance, Brunetti et al. (2017) suggest that ignoring petrophysical prediction uncertainty when using Bayesian model selection to discriminate among conceptual hydrogeological models will likely lead to over confidence in the ability of geophysical data to falsify and discriminate between alternative conceptual hydrogeological models (Linde, 2014). Furthermore, it also implies that ad hoc data weighting schemes are needed when jointly inverting geophysical and hydrogeological data (e.g., Lochbühler et al., 2013 in which each data type was given an equal weight in the objective function).

One approach to partly circumvent these issues is to avoid the use of explicit petrophysical relationships altogether. For instance, this can be achieved using structural approaches to joint inversion (Haber and Oldenburg, 1997). The cross-gradient method of Gallardo and Meju (2003) is a widely employed approach to penalize structural dissimilarity between any two parameter fields (defined as the cross-product of the spatial gradients of two parameter fields). Hydrogeophysical adaptations and applications of this method can be found in Doetsch et al. (2010), Linde et al. (2006, 2008), Lochbühler et al. (2013). Unfortunately, minimizing the cross-gradient function is an inappropriate approach when both hydrogeological properties and state variables vary (e.g., Doetsch et al., 2010; Linde et al., 2006). Among a multitude of cluster-based approaches, we highlight the works by Sun and Li (2016, 2017) who develop a multi-domain joint clustering inversion method that uses the fuzzy c-means

* Corresponding author.
E-mail addresses: Carlotta.Brunetti@unil.ch (C. Brunetti), Niklas.Linde@unil.ch (N. Linde).

https://doi.org/10.1016/j.advwatres.2017.11.028
Received 10 July 2017; Received in revised form 1 November 2017; Accepted 29 November 2017

clustering technique to constrain the statistical behaviour of inverted physical property values in the parameter domain. This approach overcomes the problem of determining a priori the appropriate petrophysical model as it is allowed to exhibit different forms in different regions of the model domain. For time-lapse applications, Vasco et al. (2014) circumvent the use of an explicit petrophysical model by relating the time at which a significant change in geophysical data occurs to the time of a saturation and/or pressure change within a reservoir or aquifer. Alternative approaches are presented by Hermans et al. (2016) and Oware et al. (2013). They link geophysical properties to hydrogeological parameters by physically-based regularization operators or direct multivariate statistical models but, unlike other methods, they adopt an explicit petrophysical relationship to create a prior set of subsurface model realizations or training images. This is done to ensure geologically realistic results.

Explicit petrophysical relationships can be integrated in hydrogeophysical inversions using two types of work flows: two-step (or sequential) inversion approaches (Chen et al., 2001; Copty et al., 1993; Doyen, 1988, 2007; Rubin et al., 1992) and coupled inversion approaches (Hinnell et al., 2010; Kowalsky et al., 2005).

The two-step inversion approach consists of two sequential steps: first, the geophysical properties (e.g., electrical permittivity) are inferred from geophysical data (e.g., first-arrival ground-penetrating radar (GPR) travel times) through deterministic or stochastic inversions; second, petrophysical relationships are used to classify and map the inferred geophysical properties into probability density functions (Mukerji et al., 2001) or deterministic estimates of hydrogeological or reservoir properties. This is achieved by different statistical techniques, such as, co-kriging, discriminant analysis, neural networks and Bayesian classification/estimation. In reservoir geophysics, the two-step inversion approach has been favored in conjunction with sophisticated statistical rock physics models. For instance, (Shahraeeni and Curtis, 2011; Shahraeeni et al., 2012) use neural networks to map inferred seismic wave impedances into posterior distributions of porosity, clay content, and water saturation. Grana and Della Rossa (2010), Grana et al. (2012) sample the posterior distribution of reservoir properties using the Monte Carlo method for a given seismic model. They conceptualize petrophysical prediction uncertainty as Gaussian random fields with zero mean and a covariance matrix estimated by comparing predictions with well-log data. In hydrogeophysics, the Bayesian two-step approaches are also used, for instance, by Chen et al. (2001, 2004) to estimate hydraulic conductivity conditioned to GPR velocity, GPR attenuation, and seismic velocity tomograms. In hydrogeophysics, the two-step approach has been criticized as it can lead to inconsistent estimates (apparent mass loss) and spatially-dependent bias (Day-Lewis et al., 2005).

The coupled inversion approach is often formulated within a Bayesian framework in which hydrogeological properties are estimated by inversion of geophysical and, possibly, hydrogeological data. A pioneering work on coupled inversion is (Bosch, 1999) who develops a formal Bayesian procedure, referred to as lithological tomography or lithological inversion. In this approach, Markov chain Monte Carlo (MCMC) is used to integrate geophysical data, geological concepts and uncertain petrophysical relationships. The coupled inversion approach is well suited to integrate multiple geophysical datasets and arbitrary petrophysical relationships. Also, when confronted with non-linear physics and non-linear petrophysical relationships, the coupled inversion approach is preferable to a two-step inversion approach (Bosch, 2004). Most hydrogeophysical works based on coupled inversion approaches assume that the petrophysical relationship is perfect with known or unknown parameter values (Chen et al., 2006; Kowalsky et al., 2005; Lochbühler et al., 2015). When petrophysical parameter values are unknown, they are inverted for simultaneously with the hydrogeological properties of interest. Petrophysical prediction uncertainty has received less attention in coupled inversion. In the rare circumstances it is included at all, it is commonly conceptualized with a multivariate Gaussian distribution with known mean and covariance matrix (Bosch, 2004; 2016; Bosch et al., 2009; Chen and Dickens, 2009). The petrophysical prediction uncertainty is then typically sampled using the brute force Monte Carlo method by adding random multivariate Gaussian realizations to the petrophysical model outputs at each iteration of the MCMC inversion.

In this study, we address the following research questions using a coupled Bayesian hydrogeophysical inversion approach:

1. How can we efficiently incorporate petrophysical prediction uncertainty in MCMC inversions?
2. What are the consequences of ignoring or making incorrect assumptions on petrophysical prediction uncertainty (including its correlation structure) on inferred posterior distributions of interest?
3. Can we reliably infer a geostatistical model of petrophysical prediction uncertainty within the inversion?
4. What are the impacts of petrophysical uncertainty on Bayesian model selection results?

After introducing the theory and method (Section 2), we start out by exploring the above-mentioned research questions by means of porosity estimation using synthetic crosshole GPR travel time data and an explicit well-known petrophysical relationship with known parameters (Section 3). We then present a field case-study (Section 4) aiming at hydraulic conductivity estimation from GPR travel time and hydraulic conductivity (flowmeter) data measured at the South Oyster Bacterial Transport site in Virginia, USA (Chen et al., 2001; Hubbard et al., 2001; Scheibe et al., 2011). Here, we solely assume to know the parametric form of the petrophysical relationship and we infer for its petrophysical parameters (i.e., the petrophysical parameter uncertainty is considered in addition to petrophysical prediction uncertainty).

2. Theory and method

2.1. Bayesian inference and model selection

We present below a short summary of Bayesian inference and model selection.

Given n measurements, $\widetilde{\mathbf{Y}} = \{\widetilde{y}_1, ..., \widetilde{y}_n\}$, and a d-dimensional vector of model parameters, $\theta = \{\theta_1, \cdots, \theta_d\}$, Bayes' theorem defines the posterior probability density function (pdf) of the model parameters, $p(\theta|\widetilde{\mathbf{Y}})$, as

$$p(\theta|\widetilde{\mathbf{Y}}) = \frac{p(\theta)L(\theta|\widetilde{\mathbf{Y}})}{p(\widetilde{\mathbf{Y}})}. \tag{1}$$

The posterior pdf describes the state of knowledge about the model parameters given the observed data and prior knowledge. The prior pdf, $p(\boldsymbol{\theta})$, quantifies the initial state of knowledge about the model parameters before considering the observed data. We consider a likelihood function, $L(\theta|\widetilde{\mathbf{Y}})$, that is Gaussian in shape by imposing uncorrelated and normally distributed measurement errors with constant standard deviation, $\sigma_{\widetilde{\mathbf{Y}}}$,

$$L(\theta|\widetilde{\mathbf{Y}}) = \left(\sqrt{2\pi\sigma^2_{\widetilde{\mathbf{Y}}}}\right)^{-n} \exp\left[-\frac{1}{2}\sum_{h=1}^{n}\left(\frac{\mathscr{F}_h(\theta) - \widetilde{y}_h}{\sigma_{\widetilde{\mathbf{Y}}}}\right)^2\right]. \tag{2}$$

The larger the likelihood, the lower is the data misfit between the simulated forward responses, $\mathscr{F}(\theta)$, and the data, $\widetilde{\mathbf{Y}}$. The evidence, $p(\widetilde{\mathbf{Y}})$, evaluates the support provided by the observed data to a given model parametrization and prior pdf (conceptual model), η, and it is defined as the (multidimensional) integral of the likelihood function over the prior distribution,

$$p(\widetilde{\mathbf{Y}}|\eta) = \int L(\theta, \eta|\widetilde{\mathbf{Y}})p(\theta|\eta)d\theta. \tag{3}$$

Computing the evidence is challenging as, in general, the integral in Eq. (3) cannot be evaluated analytically and it must be approximated by