Journal of Hydrology 514 (2014) 271–280

Contents lists available at ScienceDirect

Journal of Hydrology

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

Integration of hydrological and geophysical data beyond the local scale: Application of Bayesian sequential simulation to field data from the Saint-Lambert-de-Lauzon site, Québec, Canada

Paolo Ruggeri [a,*], Erwan Gloaguen [b], René Lefebvre [b], James Irving [a], Klaus Holliger [a]

[a] Applied Geophysics Group, Center for Research of the Terrestrial Environment, University of Lausanne, CH-1015 Lausanne, Switzerland
[b] Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Quebec City, Quebec, Canada

ARTICLE INFO

Article history:
Received 26 August 2013
Received in revised form 1 April 2014
Accepted 14 April 2014
Available online 20 April 2014
This manuscript was handled by Peter K. Kitanidis, Editor-in-Chief, with the assistance of Niklas Linde, Associate Editor

Keywords:
Hydrogeophysics
Downhole methods
Tomography
Probabilistic forecasting

SUMMARY

Adequate characterization of aquifer heterogeneity is critically important for the sustainable use, protection, and remediation of groundwater resources. The combined use of hydrological and geophysical measurements is arguably the most effective means of achieving this objective. In this regard, significant progress has been made on the quantitative integration of geophysical and hydrological data at the local scale. However, the extension of such approaches to larger, more regional scales remains a major research challenge. In this paper, we demonstrate the application of a recently developed regional-scale hydrogeophysical data integration approach, which is based on Bayesian sequential simulation, to a field database from Quebec, Canada consisting of low-resolution, surface-based geoelectrical measurements as well as high-resolution direct-push and borehole-based measurements of the electrical and hydraulic conductivities. The results of our study, which involved the integration of data along an approximately 250-m-long survey line, confirm that this novel methodology, with suitable adaptation, is fully applicable to field data and has the potential of providing realistic estimates of the spatial distribution of hydraulic target parameters at the regional-scale. Equally importantly, through the generation of multiple stochastic realizations, the methodology allows for quantitative assessment of the uncertainty associated with the inferred subsurface models, which in turn is essential for interpreting subsequent predictions of the flow and transport characteristics of the studied region.

1. Introduction

The protection, remediation, and sustainable management of the world's increasingly fragile groundwater resources require adequate models of the spatial distribution of hydraulic conductivity as prerequisites for realistic predictions of groundwater flow and contaminant transport (e.g., Delleur, 1999; Chen et al., 2001). The hydraulic conductivity is an inherently challenging material property to estimate because it varies over many orders-of-magnitude, typically exhibits a pronounced degree of spatial heterogeneity, and can in general only be measured through dedicated experiments (e.g., Domenico and Schwartz, 1998; Ezzedine et al., 1999; Rubin, 2003; Schön, 2004; Butler, 2005).

Traditionally, the hydrological characterization of aquifers has been based on evidence from drill cores, hydraulic borehole logs, and tracer and pumping experiments. Core- and borehole-based measurements can provide detailed local information, but such information is inherently 1D and spatially sparse in nature, while tracer and pumping experiments tend to capture only the gross average properties of the probed subsurface region. Correspondingly, there is a large gap in terms of spatial coverage and resolution between these conventional hydrological techniques and hence they are, without complementary information, often inadequate for characterizing heterogeneous aquifers (e.g., Sudicky, 1986; McKenna and Poeter, 1995; Schreibe and Chien, 2003; de Marsily et al., 2005). While geophysical methods have the potential of bridging this gap in resolution and coverage associated with traditional hydrological measurements, they do in general not exhibit any direct sensitivity to the hydraulic conductivity. Moreover, any potential rock physical relationships between geophysical parameters and the hydraulic conductivity tend to be site- and scale-specific (Purvance and Andricevic, 2000; e.g., Schön, 2004; Lesmes and Friedman, 2005; Hyndman and Tronicke, 2005; Linde, 2006).

To overcome these complications, a number of strategies have been proposed for local-scale aquifer characterization, that is, at

* Corresponding author. Tel.: +41 767688562.
E-mail address: paolo.ruggeri@unil.ch (P. Ruggeri).

http://dx.doi.org/10.1016/j.jhydrol.2014.04.031
0022-1694/

lateral distances ranging from approximately 5 to 50 m, typical involving a combination of core- and/or borehole-based hydraulic measurements and high-resolution crosshole tomographic geophysical surveys (e.g., Hyndman et al., 1994; Hyndman and Harris, 1996; Chen et al., 2001; Singha and Gorelick, 2005; Paasche et al., 2006; Dafflon et al., 2009b, 2009a, 2010; Dubreuil-Boisclair et al., 2011; Lochbühler et al., 2013) Most of these local-scale data integration approaches are based on geostatistical methods, which are not only well suited for assimilating diverse sources of information of varying resolution and hardness, but also offer the possibility of constraining the uncertainty of the inferred models. These local-scale data integration approaches are reaching a certain degree of maturity. Due to the lack of closely spaced boreholes for effective crosshole tomographic imaging, the extension of these local-scale techniques to larger scales does, however, represent a major and until recently essentially unresolved challenge. This is unfortunate since in many, if not most, cases it is at these larger scales that the greatest benefits of improved flow and transport predictions can be reaped (Domenico and Schwartz, 1998).

To address this problem, Ruggeri et al. (2013a) recently proposed a novel method for the quantitative integration of larger-scale geophysical and hydrological data based on a geostatistical technique known as Bayesian sequential simulation or BSS (Doyen and Boer, 1996). This approach showed significant promise when applied to realistic synthetic data for heterogeneous larger-scale aquifer models, but its practical viability remained unproven. Some promising, albeit highly preliminary, initial results of the application of this novel data integration techniques to real data were recently presented by Ruggeri et al. (2013b) in the context of a recent broad-public review.

The objective of this paper is to extend and complement the previous work by Ruggeri et al. (2013a, 2013b) by rigorously testing this novel data integration approach on a typical sub-regional-scale geophysical and hydrological field database. More specifically, we wish to explore the method's capacity and robustness for generating, in a computationally efficient manner, realistic conditional stochastic realizations of the larger-scale hydraulic conductivity field as well as for assessing the uncertainties of the thus inferred stochastic aquifer models.

2. Methodological background

In the following, we briefly outline the methodological foundations of the BSS-based data integration approach of Ruggeri et al. (2013a) before proceeding to assess its practical potential by applying it to field measurements. The BSS method (Doyen and Boer, 1996) allows for the generation of multiple, spatially correlated realizations of some variable of interest, referred to as the primary variable, conditioned to (i) spatially extensive measurements of a related secondary variable, as provided for example by geophysical surveying; and (ii) sparsely distributed measurements of the primary variable, as provided for example by borehole data. The following simplified version of Bayes' theorem forms the basis for the technique:

$$p(A_n|B_n, A_1, \ldots, A_{n-1}) = c \cdot p(B_n|A_n) \cdot p(A_n|A_1, \ldots, A_{n-1}), \tag{1}$$

where A and B denote the primary and secondary variables, respectively, $p(\cdot)$ denotes a probability distribution, and c is a normalization constant. Eq. (1) is valid under the assumption of conditional independence of B_n with respect to $A_1, A_2, \ldots, A_{n-1}$ when given A_n. That is, we assume that $p(B_n|A_1, \ldots, A_n) = p(B_n|A_n)$.

As with all sequential simulation procedures, the generation of each stochastic realization using BSS is accomplished iteratively, whereby previously simulated values for the primary variable at points along a randomly chosen path through the model space are treated as known "data" when simulating this variable at subsequent points (e.g., Goovaerts, 1997). Before the simulation begins, the covariance matrix for the primary variable is defined based on the horizontal and vertical variograms computed from existing values. In each iteration of the procedure, a value for the primary variable at cell n is then randomly drawn from the posterior distribution $p(A_n|B_n, A_1, \ldots, A_{n-1})$, which is obtained by multiplying the prior distribution $p(A_n|A_1, \ldots, A_{n-1})$ with the likelihood function $p(B_n|A_n)$. The prior distribution is conditional to the measured and previously simulated values of the primary variable in cells 1 through $n-1$, and is obtained by simple kriging of those values to obtain a Gaussian distribution having the kriging mean and variance. The likelihood function, which expresses the range of values for the primary variable in cell n that is consistent with a particular measured value of the secondary variable at the same location, is determined from the joint probability density $p(A,B)$, which in turn is computed from collocated measurements using a non-parametric kernel-based smoothing approach (Silverman, 1986; Wand and Jones, 1995). The posterior distribution can be viewed as an updated state of information that accounts for the prior and likelihood information at the chosen location. One realization of the primary variable is generated when all unknown cells in the model space have been simulated. Quite importantly, multiple stochastic realizations can be readily obtained by changing the order of the visited cells and repeating the simulation procedure. The latter allows for an assessment of the posterior ensemble uncertainty.

It should be noted that the BSS method is highly flexible with regard to the relationship that exists between the primary and secondary variables, in the sense that the likelihood is estimated empirically from collocated measurements of these variables. The quality of the relationship between A and B is thus reflected in the variability of the output stochastic realizations. Also note that, unlike cokriging-based simulation methods, the BSS approach does not rely on a generalized linear regression model, which is not an appropriate choice when the relationship between the primary and secondary variables strongly deviates from being linear and multi-Gaussian. On the other hand, when these conditions are satisfied, approaches such as collocated cokriging could be equally well implemented.

The hydrogeophysical data integration approach of Ruggeri et al. (2013a), which again showed significant potential in the course of its initial testing on synthetic data, consists of two key steps, both of which are based on the general BSS methodology outlined above. Fig. 1 summarizes the overall procedure. In the first step, high- and low-resolution geophysical parameter estimates (primary and secondary variable, respectively) are used to generate fine-scale realizations of the underlying geophysical property. The aim of this step is to effectively downscale the low-resolution geophysical parameter estimates and quantify the corresponding uncertainty with regard to the fine-scale grid. In this procedure, the high-resolution data are considered to be measurements of the geophysical parameter at a small number of sparsely distributed borehole locations throughout the aquifer volume. The low-resolution data, on the other hand, are considered to be a tomographic image obtained through the inversion of geophysical survey data, which can be regarded as set of uncertain spatially averaged measurements of the "true" subsurface geophysical parameter field. The likelihood function is thus estimated from the joint probability inferred from collocated or quasi-collocated high- and low-resolution geophysical parameter estimates at the borehole locations.

In the second step of the data integration approach of Ruggeri et al. (2013a), borehole measurements of the hydraulic conductivity (primary variable) and point-by-point statistics of the high-resolution geophysical parameter field derived from the realizations