WIREs WATER

Article type: Overview

Article title: Falsification and Corroboration of Conceptual Hydrological Models Using Geophysical Data

Authors:

First author
Niklas Linde, Applied and Environmental Geophysics Group, Faculty of Geosciences and the Environment, University of Lausanne; niklas.linde@unil.ch

Abstract

Geophysical data may provide crucial information about hydrological properties, states and processes that are difficult to obtain by other means. Large data sets can be acquired over widely different scales in a minimally invasive manner and at comparatively low costs, but their effective use in hydrology makes it necessary to understand the fidelity of geophysical models, the assumptions made in their construction and the links between geophysical and hydrological properties. Geophysics has been applied for groundwater prospecting for almost a century, but it is only in the last twenty years that it is regularly used together with classical hydrological data to build predictive hydrological models. A largely unexplored venue for future work is to use geophysical data to falsify or rank competing conceptual hydrological models. A promising cornerstone for such a model selection strategy is the Bayes factor, but it can only be calculated reliably when considering the main sources of uncertainty throughout the hydrogeophysical parameter estimation process. Most classical geophysical imaging tools tend to favor models with smoothly varying property fields that are at odds with most conceptual hydrological models of interest. It is thus necessary to account for this bias or use alternative approaches in which proposed conceptual models are honored at all steps in the model building process.

INTRODUCTION

Geophysical data are increasingly used to monitor hydrological processes and to image controlling sub-surface structure over various spatial and temporal scales.[1-3] Geophysical models describing the distribution of physical properties in space and time can sometimes be translated into distributed estimates of relevant hydrological properties (e.g., porosity, tortuosity) and states (e.g., water content, salinity) or used to image controlling subsurface features (e.g., bedrock topography, faults). Geophysics not only enables comparatively cheap and less invasive site investigations compared to extensive drilling programs, but it may also provide information at scales that are difficult to sample with classical hydrological methods.[4] One of the main challenges in using geophysical techniques in hydrology is that the geophysical data are primarily sensitive to physical properties that are of little direct use in hydrology (e.g., electrical permittivity or conductivity). Petrophysical relationships are thus needed to translate these properties into the hydrological properties and state variables of interest. The distinction between geophysical and hydrological measurement techniques is often rather arbitrary as similar calibration steps with underlying transfer functions are often used to produce hydrological data (e.g., time-domain reflectometry to infer water content or water conductivity to provide tracer concentration, etc.).[5] Hydrological investigations that use geophysical data are often grouped in the maturing multidisciplinary field of hydrogeophysics, with the first attempts to integrate geophysics in quantitative stochastic hydrogeology dating back to the early 1990s[6]. Recent reviews describe hydrological applications that can benefit from geophysics, the most relevant petrophysical models that link hydrological and geophysical properties and successful case-studies.[1-3,7-10] Geophysics offers excellent means to complement the information gained from classical hydrological investigations, but it should not be seen as a competitor nor a replacement of traditional data (e.g., pumping or tracer tests) that often are more sensitive to the underlying hydrological system behavior. The theoretical framework for deterministic[11] and stochastic[12] geophysical parameter estimation (i.e., geophysical inverse theory) is well developed, which is also the case in hydrogeology[13-14]. Several textbooks provide up-to-date introductions to the most common near-surface geophysical methods.[15] See Ref 16 for the terminology associated with the hydrological modeling process.

Model comparison studies in the Earth Sciences reveal that model predictions are often more dependent on the underlying conceptual model than on the parameter estimation procedure employed to determine the most suitable parameter values[17-19]. Ref 19 analyzed 412 geological interpretations of the same synthetic high-resolution seismic section by as many trained individuals. Only 21% of the participants interpreted the correct tectonic setting. More importantly, the interpretations of the conceptual model were strongly affected by the interpreters' past training and the interpretational tools used. This highlights the subjectivity associated with interpretations of geophysical data or models and the obvious risk of promoting a false sense of certainty if one interpretation is taken as the sole basis for further analysis. Furthermore, it has been argued that conceptual Earth models can never be validated.[20] Field data are noise-contaminated and data coverage is always incomplete. Under such circumstances, it is well known that if one model is found that agrees with the data, then there exists an infinite set of models that explain these data equally well.[21] It is therefore an illusion to consider that a model that explains the available data is validated regardless of the types, numbers, or quality of the data used. It is better to investigate different conceptual models and consider the ones that perform well to be corroborated[22] or behavioural[23]. In

hydrogeophysics, insufficient emphasis has been placed on using formal theory to define and test alternative conceptual models. Geophysical data are often used to parameterize hydrological models[6,24], but less frequently to test underlying assumptions about parameter structure, initial or boundary conditions. A shift towards falsification and ranking of competing conceptual models would help to place the geophysical data at the centre of scientific investigations and stimulate close collaboration across disciplines. Non-uniqueness will always remain a central characteristic of hydrological and geophysical investigations regardless of developments in terms of more flexible modeling and inversion codes, more powerful computers, better sensors, more data and new methods.[23] Emphasis should be placed on what the resulting images tells us about the scientific, management or engineering questions being asked.[25] This can only be achieved if the impacts of all choices in the hydrogeophysical parameter estimation problem are understood and considered when interpreting the data and making recommendations. The focus of this overview is to describe the main steps and challenges associated with using geophysical data to falsify and rank conceptual hydrological models. It is my hope that this text can serve as a starting point to stimulate research on hydrological model selection using geophysical data and to help non-geophysicists to avoid some of the most common pitfalls associated with the use of geophysical data or models in hydrology.

NON-UNIQUENESS

When first confronted with geophysical data, it is easy to be impressed by the large data volumes. The information content of tens or hundreds of thousands of geophysical data might appear immense (e.g., compared to piezometric observations in a few boreholes), but it is limited. This implies that only a finite number of model parameters can be independently inferred from the data.[21] Even if modern geophysical surveys based on acquisitions in boreholes, on the ground, within water bodies, and from airborne platforms (helicopters, airplanes, satellites, drones) can result in vast data sets, the possible acquisition geometry is fundamentally limited. The data are noise-contaminated and the different physical phenomena underlying the geophysical methods (e.g., diffusion, wave propagation) limit the resolution at which physical property variations can be sensed.

Experimental design allows increasing the information content by optimizing geophysical campaigns.[26] For a given budget and equipment pool, it allows estimating the most suitable measurement sequence. Experimental design is challenging due to the non-linear response of most geophysical methods, which implies that the optimal design depends on the unknown physical property distribution of the subsurface. Instead of relying on algorithms, the chosen experimental design is, therefore, in practice often a result of the geophysicist's past experience and the problem at hand. An in-depth understanding of geophysical methods, their imaging capabilities and link to target properties are therefore key components in defining an appropriate experimental design. Synthetic modeling experiments that consider different subsurface scenarios can often be useful to assess how well a proposed layout is likely to perform.

With modern computer-based inversion algorithms, it is relatively easy to obtain three-dimensional distributed models that explain the available data within their estimated errors, but finding such models is no guarantee that they are useful. For instance, for the 1-D (the subsurface is represented by horizontal layers) inverse problem of resistivity sounding it has been shown that the best fitting model explaining an arbitrary layered electrical resistivity distribution have layers of near-zero

thickness followed by a perfect electrical conductor.[27] This implies that there is no inherent depth information in these data and that any mapping of geological structure at depth is also a consequence of model regularization (e.g., user-defined constraints on the magnitude and patterns of the inferred model). In a 2-D setting (i.e., physical properties are assumed to be constant in one spatial direction), a relatively complex test model was used to demonstrate that electrical data could be equally well explained by an isotropic smooth model as models consisting of sequences of either horizontal or vertical stripes.[28] None of these inversion models has a morphology that is close to the true test model. A similar situation is expected in 3-D where physical properties vary throughout the investigation volume.

In geophysics, the classical solution to the non-uniqueness problem described above is to artificially impose uniqueness of the inverse problem by searching for the model that shows the least spatial variations under the constraint that the model fits the data[29]. Such a smoothness-constrained or minimum-structure approach is useful when very limited prior information is available or when locating anomalies (e.g., mineral or oil prospection), but imposing smoothness in the resulting models is likely to result in a subsurface model with a morphology that is rather different from the true subsurface structure.[30] Figure 1a shows a relatively simple test model representing a gravel aquifer (300 Ohm-m) with a few embedded clay lenses (30 Ohm-m) that overlays bedrock (3000 Ohm-m). The upper meter of the gravel formation is only partly saturated (1000 Ohm-m). An electrical resistivity tomography (ERT) survey was simulated using 50 electrodes placed on the ground surface with an electrode spacing of 2 m. The simulated data were contaminated with uncorrelated Gaussian errors with a standard deviation of 3% of the simulated values. The model from a smoothness-constrained ERT inversion with the same data misfit as the test model is shown in Figure 1b. The gravel, clay, bedrock and the vadose zone are all outlined, but the detailed geometry of the clay lenses and the topography of the bedrock-gravel aquifer are smoothed out and the true resistivity values of the bedrock is not recovered.

A more satisfying approach is to use Bayes theorem to combine prior information and the available data to derive a closed-form expression or samples from a posterior probability density function[12]. The data are combined in a data likelihood term that describes how likely it is that a given model is responsible for the observed data under an assumed error model. The prior probability density function describes how likely a model is before considering its agreement with the data. Bayesian methods can be very advanced and allow for introducing uncertain prior information and poorly known error statistics. Nevertheless, it has been argued that the true reason for the popularity of such Bayesian inverse methods does not necessarily lie in the elegance at which prior information and data are combined, but rather in that the choice of the prior probability density function (pdf) allows tuning the properties of the proposed models.[31] Such a tweaking of the “prior information” is at odds with underlying Bayesian theory and may lead to results that simply confirm the preconceptions of the modeler. Such pragmatic inversions can be useful for many practical applications, but it is important to explicitly state and understand that the resulting models and uncertainty estimates are strongly influenced by possibly unrealistic assumptions and that they are not pure results of irrefutable prior information and observed data. This argument applies also to classical geostatistical inversion that is based on the assumption of an underlying multiGaussian random field.[32] The resulting models are presented within a probabilistic framework, but they all have minimal correlation of extreme values, which makes them poor at describing realistic geological structures.[33] Indeed, significant issues may arise when calibrating field data from a very

connected subsurface setting (e.g., buried channels) with an underlying assumption of multiGaussian fields.[34] Nevertheless, simplifying assumptions are often needed in practical situations and the information gained by classical geostatistical inversion algorithms[35] can often be very high.[36] For twenty years, the geostatistical community has developed methods that rely on higher-order spatial statistics (so-called multiple-point statistics) to obtain geologically realistic models by using training images describing the expected subsurface structure.[37] These training images are nothing else than another type of model regularization that constrain the model space of the final solutions, even if they may be more adequate than standard regularization choices if appropriate training images are available.

Model regularization can be thought of as any constraints that are imposed on the permissible model structure (model parameterization, presupposed structure, damping or constraints on variability, etc.) that doesn't come from the observed data. When using geophysical models to test a conceptual hydrological model (e.g., the subsurface is characterized by a continuous multiGaussian field or a discontinuous field of lithofacies), it is important that the model regularization used in the geophysical inverse problem is in agreement with the underlying conceptual model to be tested (e.g., a classical minimum structure or multiGaussian geostatistical inversion will never produce a model with discontinuous property variations; adding known mass constraints in the geophysical inversion will avoid proposing physically inconsistent models that don't conserve mass) or that the resulting bias is taken into account. This is seldom possible when using standard geophysical inversion tools and the bias introduced by rather inflexible model regularizations may make the resulting geophysical models unsuitable for model testing.

SOURCES OF UNCERTAINTY AND THEIR REPRESENTATION

This section describes some of the key components of the hydrogeophysical parameter estimation problem and the associated sources of discrepancies between measured data and the simulated model response. Let **x** represent the spatial vector, for example, $\mathbf{x} = (x, y, z)$ for a 3-D Cartesian coordinate system.

At any time *t*, the hydrological system will be in a state that can be described by three-dimensional fields of state variables $S_i(\mathbf{x}, t)$, where *i* may indicate water content, chemical composition of the water phase, pore pressure, temperature, etc. The evolution of the hydrological state depends on physical laws, the distribution of system properties $P_j(\mathbf{x}, t)$, where *j* may indicate permeability, porosity, specific storage, etc., and *k* types of boundary conditions $B_k(\mathbf{x}, t)$ including any forcing to the system (e.g., precipitation or pumping in a borehole).

The most suitable conceptual model describing the key aspects of a hydrological system under study is application-dependent. It may be a simple four-parameter model used in rainfall-runoff modeling[38], a refined physically based and distributed model representing a nuclear waste repository[39], or something in-between for modeling a hillslope[40]. The conceptual model is designed to describe the main aspects of $S_i(\mathbf{x}, t = t_0)$, $P_j(\mathbf{x}, t)$ and $B_k(\mathbf{x}, t)$ to allow predicting (in a qualitative or quantitative sense) the evolution of $S_i(\mathbf{x}, t)$ and any derived quantities (e.g., arrival

time of a tracer). The parameters describing this conceptualization can be assumed to be known or unknown. In both cases, it is crucial to choose a model parameterization that allows describing the main features of the system with a limited number of parameters. A general parameterization of a property field $p(\mathbf{x})$ is[41]

$$p(\mathbf{x}) \approx p^{\mathrm{m}}(\mathbf{x}) = \sum_{l=1}^{L} c_l \psi_l(\mathbf{x};\mathbf{a}), \tag{1}$$

where $\boldsymbol{\psi} = (\psi_1,\ldots,\psi_L)$ denotes a set of *L* basis functions, and $\mathbf{m} = \left[\mathbf{c}^T\ \mathbf{a}^T\right]$ represents model parameters. The superscript m is used to represent a model of the true unknown fields. The exterior coefficients **c** govern the magnitude of the model parameters and the interior coefficients **a** describe the parameter structure through the basis functions. Some model parameters might be perfectly known or assumed known as they only have a limited importance. The most common approach is to use pre-defined basis functions (e.g., a pixel-based model; a lithological model of known geometry, but unknown aquifer properties; a uniform property field, a multiGaussian model, etc.). In some situations, it might be useful to assume that the basis functions are unknown (e.g., the depths to bedrock are allowed to vary within specified ranges). Equation 1 highlights that even if the optimal model parameters are found for a given model dimension, it might still be a rather poor description of the underlying system. As an example, Figure 2a shows a porosity field and Figure 2b the "best" upscaled representation at a given spatial resolution. The differences (Figure 2c) are correlated in space.

Geophysical measurements are directly sensitive to the distribution of physical properties $G_g(\mathbf{x},t)$, where *g* might indicate electrical conductivity, density, etc. They are likely to have some relation to $P_j(\mathbf{x},t)$ and $S_i(\mathbf{x},t)$ through an unknown petrophysical relationship

$$G_g(\mathbf{x},t) = R_g\left(S_i(\mathbf{x},t), P_j(\mathbf{x},t), N_n(\mathbf{x},t)\right), \tag{2}$$

with $N_n(\mathbf{x},t)$ indicating "nuisance" parameters that influence the geophysical property, but not the hydrological system (e.g., the relative permittivity of the soil matrix). This petrophysical mapping is often only approximately known and is a major source of uncertainty in hydrogeophysical studies.

To enable geophysical forward modeling (i.e., to predict the geophysical response for a proposed model), the hydrological model needs to be mapped into a distributed geophysical field using a petrophysical model

$$G_g^{\mathrm{m}}(\mathbf{x},t) = R_g^{\mathrm{m}}\left(S_i^{\mathrm{m}}(\mathbf{x},t), P_j^{\mathrm{m}}(\mathbf{x},t), N_n^{\mathrm{m}}(\mathbf{x},t)\right). \tag{3}$$

This relationship can take many forms and its predictive value can vary strongly. The discrepancy between the true and modeled geophysical property fields is given by

$$\Delta G_g(\mathbf{x}) = R_g\left(S_i(\mathbf{x},t), P_j(\mathbf{x},t), N_n(\mathbf{x},t)\right) - R_g^{\mathrm{m}}\left(S_i^{\mathrm{m}}(\mathbf{x},t), P_j^{\mathrm{m}}(\mathbf{x},t), N_n^{\mathrm{m}}(\mathbf{x},t)\right). \tag{4}$$

Figure 2d shows how the true porosity field (Figure 2a) is mapped into radar wave speed using a petrophysical transform, while Figure 2e displays the radar wave speed field obtained by applying an

approximately known petrophysical relationship to the upscaled porosity field (Figure 2b). The differences between the two fields (Figure 2f) display a spatial correlation.

A geophysicist performs diverse experiments to acquire data that are characterized by a certain precision and accuracy, with possible interference from parasitic signals (e.g., power lines for electromagnetic (EM) data; diffractions from trees for ground penetrating radar (GPR) data). The geophysical response is given by

$$\mathbf{d}(t) = f\left(G_g(\mathbf{x},t), Exp\right) + \boldsymbol{\varepsilon}(t), \tag{5}$$

where *f*(-) is the physical response of an experiment *Exp* given $G_g(\mathbf{x},t)$, and $\boldsymbol{\varepsilon}(t)$ is a vector of data errors associated with the experiment.

Geophysical forward modeling $f_g^{\mathrm{m}}(-)$ maps the geophysical model into a simulated instrument response. This is typically achieved using numerical methods (e.g., finite element or finite difference methods) that solve the governing ordinary or partial differential equations. The modeling is generally based on a number of simplifying assumptions that decrease computing times (e.g., antennas or electrodes are assumed infinitely small, boreholes are not explicitly parameterized in the modeling, topography is ignored), which leads to modeling errors. The simulated geophysical model response is given by

$$\mathbf{d}^{sim}(t) = f_g^m\left(G_g^m(\mathbf{x},t), Exp^m\right) \tag{6}$$

with corresponding data residuals

$$\Delta\mathbf{r}_g(t) = \left[f\left(G_g(\mathbf{x},t), Exp\right) + \boldsymbol{\varepsilon}(t)\right] - \left[f_g^{\mathrm{m}}\left(G^{\mathrm{m}}(\mathbf{x},t), Exp^{\mathrm{m}}\right)\right] \text{ or} \tag{7}$$

$$\begin{aligned}\Delta\mathbf{r}_g(t) = &\left[f\left(R_g\left(S_i(\mathbf{x},t), P_j(\mathbf{x},t), N_n(\mathbf{x},t)\right), Exp\right) + \boldsymbol{\varepsilon}(t)\right] + \\ &-\left[f_g^{\mathrm{m}}\left(S_i^{\mathrm{m}}(\mathbf{x},t), P_j^{\mathrm{m}}(\mathbf{x},t), N_n^{\mathrm{m}}(\mathbf{x},t), Exp^{\mathrm{m}}\right)\right]\end{aligned}. \tag{8}$$

A geophysical model response with observational errors is shown for the real wave speed field (Figure 2d). The responses of the upscaled model (Figure 2e) were calculated assuming incorrect borehole geometry. The data residuals (Figure 2i) from this illustrative example indicate a distribution that is neither symmetric nor centered at zero. Any attempts to invert these data assuming zero-mean uncorrelated Gaussian errors will fail in recovering a model that is close to the one shown in Figure 2b.

The data residuals between the actual subsurface response and the one obtained by the “best possible” upscaled model are thus a function of (1) errors in the petrophysical model, (2) errors in the geophysical forward modeling, including geometrical errors in representing the geophysical experiments, (3) observational errors of the actual experiments and (4) simplifications and assumptions made to parameterize and assign values to the conceptual hydrological model. Any meaningful procedure that aims to distinguish between different conceptual models using geophysical data must therefore also consider error sources (1-3) that are briefly discussed below:

1. Petrophysical relationships have been under intense study in geophysics with primary focus on reservoir rocks and seismic properties[42-43], but a rich (albeit incomplete) literature exists for conditions of relevance to hydrology[44]. The parametric forms of the petrophysical models are often obtained from theoretical considerations (e.g., differential effective media theory; volume averaging) or from statistical relations between collocated geophysical and hydrological estimates. The petrophysical relationships are likely to be scale dependent and non-stationary, which implies that the discrepancy between predicted and actual properties is correlated in space. Scale-dependence, distribution of optimal parameters and error correlation are difficult to assess and are ignored in most hydrogeophysical investigations without further considering the consequences on the resulting predictions.

2. Numerical modeling tools in hydrology and geophysics are constantly evolving and modeling errors are often investigated thoroughly for cases where analytical solutions are available.[45] Significant modeling errors will continue to prevail as geophysicists are pushed to look on ever-finer features of the data to obtain ever-higher resolution models that are increasingly sensitive to small imperfections in the forward models. In addition to errors in the numerical modeling, geometrical errors with respect to source and receiver positions and topography are known to have important effects. To a certain extent, it is possible and beneficial to invert for the uncertain experimental geometry[46] or to build a correlated noise model[47]. Generally, most quantitative studies in high-resolution hydrogeophysics necessitate cm to dm scale positioning accuracy, which implies the routine use of modern surveying methods (e.g., differential global positioning system, tacheometers, borehole deviation tools, etc.).

3. Observational errors are comparatively easy to handle. Statistics can be derived by analyzing the accuracy of the sensor in capturing the signal under interest by considering calibration tests, the sampling rate of the data acquisition device, or frequency-analysis of time-series from which signal-to-noise ratios can be derived. The underlying pdf is often assumed to be a zero-mean uncorrelated Gaussian process with a known standard deviation, even if it is possible to infer the standard deviation as a part of the inverse problem.[48] When noise statistics are unknown, it is often preferable to assume heavy-tailed distributions, such as symmetric exponentials, as the resulting models are then less sensitive to individual data outliers.[49]

It is only for linear theory and for Gaussian noise that it is theoretically justified to combine the three errors described above into one combined error (to be estimated or not), but this is by far the prevailing approach in practice even when dealing with nonlinear problems. In this case, the corresponding likelihood function is[12]

$$L(\mathbf{m}|\mathbf{d}) = \frac{1}{(2\pi)^{N/2}\det(\mathbf{C}_\mathrm{d})^{1/2}}\exp\left(-\frac{1}{2}\left(g(\mathbf{m})-\mathbf{d}\right)^T\mathbf{C}_\mathrm{d}\left(g(\mathbf{m})-\mathbf{d}\right)\right), \tag{9}$$

where $\mathbf{C}_\mathrm{d}$ is the data covariance matrix that contains the error variances and covariances of *N* data and **d** is the data vector representing the geophysical data acquired in the field. The prior model distribution (or regularization term) is most often assumed to have the form[12]

$$p(\mathbf{m}) = \frac{1}{(2\pi)^{M/2}\det(\mathbf{C}_{\mathrm{m}})^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{\mathrm{ref}})^T\mathbf{C}_{\mathrm{m}}(\mathbf{m}-\mathbf{m}_{\mathrm{ref}})\right), \tag{10}$$

where $\mathbf{C}_{\mathrm{m}}$ is the model covariance matrix, M is the number of model parameters and $\mathbf{m}_{\mathrm{ref}}$ is the reference model. If $\mathbf{C}_{\mathrm{d}}$ and $\mathbf{C}_{\mathrm{m}}$ are assumed known, it is possible to express the posterior distribution as[12]

$$p(\mathbf{m}|\mathbf{d}) = k\exp(-\phi(\mathbf{m})), \tag{11}$$

with k a constant and

$$\phi(\mathbf{m}) = \frac{1}{2}\left((g(\mathbf{m})-\mathbf{d})^T\mathbf{C}_{\mathrm{d}}(g(\mathbf{m})-\mathbf{d}) + (\mathbf{m}-\mathbf{m}_{\mathrm{ref}})^T\mathbf{C}_{\mathrm{m}}(\mathbf{m}-\mathbf{m}_{prior})\right). \tag{12}$$

For linear inverse problems, it is possible to write $g(\mathbf{m}) = \mathbf{G}\mathbf{m}$, and minimize the misfit function $\phi(\mathbf{m})$ by taking its derivative to be zero. This yields the following least-squares solution[12]

$$\hat{\mathbf{m}} = \left(\mathbf{G}^T\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G} + \mathbf{C}_{\mathrm{m}}^{-1}\right)^{-1}\left(\mathbf{G}^T\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d} + \mathbf{C}_{\mathbf{m}}^{-1}\mathbf{m}_{ref}\right). \tag{13}$$

It is common to assume that $\mathbf{C}_{\mathrm{d}}$ is given by a diagonal matrix of expected error variances, which implies that the data errors are uncorrelated. The prior or model regularization is often further simplified by replacing $\mathbf{C}_{\mathrm{m}}$ with a matrix that quantifies the first derivative of the proposed model (flatness), the second derivative (roughness) or its deviation from $\mathbf{m}_{\mathrm{ref}}$ (damping) multiplied by a model regularization weight λ.

A data misfit function or distance function (e.g., the first term in parenthesis in equation (12)) is used to assess if the simulated model response of the proposed model is in agreement with the assumed error distribution. Under the assumptions made above, it should follow a chi-square distribution with an expected value of N and the misfit measure (e.g., equation (12)) can be used to obtain an optimal or probabilistic description of the uncertain model parameters. The misfit measure should ideally be sensitive to the most important parameters in the conceptual model, while being robust with respect to nuisance parameters, assumptions on the statistics of the observational noise, to modeling and geometrical errors. The definition of an appropriate distance measure necessitates assumptions regarding the observational noise, the appropriateness of the conceptual model, the underlying petrophysics, and modeling errors. Field data may display important outliers that may severely affect the resulting results when using least-squares methods and it is often beneficial to use more robust measures and norms.[12,49]

MODEL-SELECTION

Motivation for model selection

In hillslope hydrology, a hydrologist might want to know if a soil-bedrock interface exists within, say, the ten top meters, and if so, with what topography. Assuming impermeable bedrock, it is clear that the absence or presence of such an interface plays a key role in determining the conceptual model of the hillslope. However, will the sharp soil-bedrock interface show up when applying standard

geophysical inversion tools? The test case in Figure 1 showed that the presence of a bedrock interface was clearly resolved by the data, but that it was not possible to image its actual topography using this inversion method and data set.

Geophysical data are often analyzed with interpretational software developed for general-purpose applications (e.g., Figure 1b). Limited emphasis is given to estimate or evaluate the impact of the choice of noise statistics or the model regularization, even if they have important impacts on the resulting models and uncertainty estimates.[50] One reason for this is that commercial and academic geophysical software tools provide few options to define noise characteristics and model regularization is often restricted to smoothness[29] or damping constraints. The translation of the resulting geophysical model into properties relevant to hydrologists is often made through a generic and simplified petrophysical model, which might lead to significant bias.[51] The model resolution and uncertainty of the resulting models are (if ever considered) often based on linear inverse theory[11] that treats the model regularization and noise statistics as solid prior information. The typical outputs of the geophysical inverse process may thus provide an unsuitable basis for constructing or testing conceptual hydrological models or to infer parameter values.[52-53]

A trained geophysicist might have a good understanding about model features that are well resolved, but this information is not always communicated to the final hydrological user. This leads to the obvious risk that the hydrologist over-interprets the geophysical information by treating each model pixel as an independent well-known truth or "data point" that is either perfectly[6] or approximately known[54]. An opposing risk is that well-resolved features are ignored or that the models are disregarded altogether if they are presented such that they have little apparent value to the hydrologist.[55] Over- and under-interpretation of geophysical models, simplified uncertainty assessments and the choices made in translating geophysical models into hydrological properties often affect the utility of geophysics in hydrology. Close collaboration between geophysicists and hydrologists might partly reduce these problems, but only if the geophysicist communicates realistic expectations prior to the investigation, understands the underlying hydrological problems, and makes an effort in communicating uncertainty in either objective or subjective terms. There are fundamental limits in our ability to capture model and prediction uncertainty. Most hydrogeophysical studies focus on parameter estimation under the (implicit or explicit) assumption of rather simple generic conceptual model types (e.g., multiGaussian geostatistics, a layered model), while hydrological system response may be strongly affected by features that only make up a small fraction of the subsurface, such as macro pores in soils[56] or thin high-permeability channels[57] that are likely to be unresolved when applying a standard inversion approach. Ignoring such features, for example, by only estimating multiGaussian fields, leads to unrealistically low uncertainty estimates and incorrect predictions. Explicit testing of conceptual models is perhaps the most suitable way to assess some of these deeper sources of uncertainty and to explore our "border with ignorance"[58] (e.g., even if there is no field evidence of macro pores or channels at a site, this is possibly only a consequence of the type and number of data available together with the interpretational framework used). A deeper and potentially dominating uncertainty are features and system properties that we simply ignore, but that may have an important impact on system behavior.[59] This ignorance is difficult to integrate in an uncertainty analysis and it is often necessary to resort to subjective choices involving an "engineering factor" that inflates the estimated uncertainty range, which is particularly important for complex systems (non-linear and heterogeneous) when the stakes associated with the modeling outcome are high.[60] It is often computationally prohibitive to

investigate large subsets of models that agree with available data and prior knowledge, which implies that it is important to focus on the aspects of the system that are expected to have the largest impact on the spread of the final modeling predictions.[25] This often implies analyzing and comparing conceptual subsurface models.

A method to compare competing conceptual hydrological models using geophysical data should ideally have the following characteristics:

- It should allow incorporating prior preferences for the different conceptual models based on the modeler's experience and site understanding;
- It should allow comparisons of models with different degrees of complexity;
- It should consider that the parameter values are uncertain;
- It should allow for nuisance parameters that affect the geophysical data, but provide limited or no relevant information about the hydrology;
- It should be computationally feasible in typical field applications.

Bayes factor

The Bayes factor[61] and approximations thereof respond well to the characteristics outlined above[59]. The Bayes factor B_{ij} for two competing conceptual models H_i and H_j is

$$B_{ij} = \frac{p\left(\mathbf{d}\middle|H_i\right)}{p\left(\mathbf{d}\middle|H_j\right)} = \frac{\int p\left(\mathbf{d}\middle|\mathbf{m}_i, H_i\right) p\left(\mathbf{m}_i\middle|H_i\right) d\mathbf{m}_i}{\int p\left(\mathbf{d}\middle|\mathbf{m}_j, H_j\right) p\left(\mathbf{m}_j\middle|H_j\right) d\mathbf{m}_j}, \tag{14}$$

where $p\left(\mathbf{d}\middle|H_i\right)$ is the evidence (or marginal probability) of the data given H_i obtained by integrating the joint density of $\left(\mathbf{d}, \mathbf{m}_i\right)$ given **d** over $\mathbf{m}_i$. The evidence is large for parameterizations in which the posterior model space of likely models takes up an important part of the prior model space. This is more likely to happen in lower model dimensions or for strongly constrained solutions, and the evidence offers thereby a natural counter-balance to the lower data misfits typically associated with models that have many freely varying parameters. The Bayes factor can be interpreted as a summary of the evidence provided by the data in favor of one conceptual model compared to another. The Bayes factor can also be viewed as the success of H_i relative to H_j in predicting the data. If one hypothesis is favored a priori, it is possible to combine Bayes factor with prior probabilities $p\left(H_i\right)$ as follows

$$\frac{p\left(H_i|\mathbf{d}\right)}{p\left(H_j|\mathbf{d}\right)} = \frac{p\left(\mathbf{d}\middle|H_i\right)}{p\left(\mathbf{d}\middle|H_j\right)} \frac{p\left(H_i\right)}{p\left(H_j\right)}. \tag{15}$$

Calculating Bayes factor and its simplifications

To calculate Bayes factors, the integrations in equation (14) must often be carried out using computationally expensive numerical methods. Fast asymptotic solutions exist, such as the Laplace

method that uses a quadratic expansion around the maximum likelihood estimate[62]. A more general (but computationally more expensive approach) is Markov chain Monte Carlo (MCMC) sampling with equation (14) being evaluated by importance sampling (i.e., regions of high likelihood are preferentially sampled, but with the sampling bias being accounted for). Unfortunately, MCMC is computationally expensive in high parameter dimensions[63].

The Schwarz criterion $\left(SBC_{ij}\right)$ or the related Bayes Information Criterion $\left(BIC = -2SBC_{ij}\right)$ gives a rough approximation to the logarithm of Bayes factor. This asymptotic criterion ignores the effect of the prior pdf by assuming that the information content in the data is much larger and that the likelihood function is well represented by its maximum likelihood. The Schwarz criterion is given by

$$SBC_{ij} = \log p\left(\mathbf{d}\middle|\hat{\mathbf{m}}_i, H_i\right) - \log p\left(\mathbf{d}\middle|\hat{\mathbf{m}}_j, H_j\right) - \frac{1}{2}\left(M_i - M_j\right)\log N, \qquad (16)$$

where $\hat{\mathbf{m}}_i$ indicates the maximum likelihood solution of model type *i* and M_i is the corresponding number of model parameters. Significant care must be taken in defining the effective number of model parameters in presence of model regularization. In the statistical literature on ridge regression, it is common to determine the effective number of model parameters by the trace of the smoother[64], which is referred to as the data resolution matrix in the geophysical literature[10].

Consider 100 hypothetical measurements of bulk resistivity ρ_b using samples with known porosity *n* (see Figure 3). The data were obtained by applying the petrophysical relationship, $\rho_b = \rho_w n^{-m}$ for a pore water resistivity ρ_w of 100 Ohm-m, a cementation exponent *m* of 1.8, and contaminating the resulting data with Gaussian noise with a standard deviation of 20%. The generated electrical resistivity data have a chi-square data misfit of 99. Least-squares inversion was carried out (equation 13) to infer a relation between porosity and bulk resistivity with a discretization of 250 different water content intervals (i.e., 250 model parameters). To stabilize the inversion, a roughness matrix penalizes the second derivative of this relation with respect to neighboring water contents. By varying the regularization weight, it is possible to vary the data misfit and the effective number of model parameters (calculated by the trace of the smoother as discussed above). The first model (underfitted; Figure 3) has a Schwartz criterion with respect to the true model of 48 (i.e., it is approximately 10^{42} times less likely than the actual petrophysical relationship). It does not explain the data very well (chi-square data misfit of 195), but uses only two effective model parameters. The second model (fitted; Figure 3) has a Schwartz criterion of 6.4. It has the same chi-square as the actual petrophysical relationship, but is penalized because it has 4.7 effective model parameters. Finally, a third model is presented (overfitted; Figure 3) with a Schwartz criterion of 65. The model fits the data very well (chi-square data misfit of 50), but it uses 41 effective model parameters to do so. This simple example clearly indicates that the underlying petrophysical relationship is by far the most likely, and that the second model (fitted; appropriate fit to the data and not too many effective model parameters) is preferred compared with the other two models considered. Finally, it shows that default parameterization of inverse problems in terms of finely discretized models and classical regularization constraints can be problematic. Indeed, a careful parameterization might allow for a suitable data misfit with fewer effective model parameters.

To demonstrate the use of Bayes factors for model selection, consider a hypothetical crosshole GPR data set acquired between two boreholes. Four conceptual models of the porosity distribution are

compared: (1) the true geostatistical model (exponential model with known mean and horizontal and vertical integral scales of 10 and 2 m); (2) same as in (1), but with integral scales of 6 m in both directions; (3) the same as in (1), but with the horizontal integral scale being 2 m and the vertical being 10 m; (4) porosity is constant throughout the aquifer assuming the same univariate statistics as for the porosity distribution in (1-3). A total of 400 travel times were simulated using a station and receiver spacing of 0.4 m. Assuming linear theory, the data covariance matrix consists of contributions due to observational errors (standard deviation of 1 ns), errors in the petrophysical relationship (random errors in water content of 0.012 at each grid cell of 0.1 m), and geometrical errors as the horizontal and vertical positions of each borehole in this example are only assumed known with an accuracy of 5 cm. The petrophysical and geometrical errors and the associated covariances are obtained by linear error propagation. The Bayes factor is here evaluated with respect to model type (1) by integrating the likelihoods over the prior model space by drawing realizations from the prior distribution. The data response was calculated for 20,000 random samples of each model type. Figure 4a shows the porosity distribution used to generate the data with a chi-square data misfit of 402, while Figures 4b-e show the best fitting realizations for each of the model types (corresponding chi-square data misfits are 412, 446, 488, and 421). Based on the calculated Bayes factors, model type (1) is 2.7×10^2 times more likely than model type (2), 1.7×10^{17} times more likely than model type (3), and 8.2×10^4 times more likely than model type (4).

Computational limitations have so far somewhat limited the use of Bayes factors in geophysics. Among published studies, it has been applied to data sensitive to the lunar interior to show that the hypothesis of a molten lunar core is more likely than a solid core[65] and to compare alternative parameterizations of sediment shear-wave velocity profiles inferred from interface-wave dispersion data[66]. A closely related methodology is transdimensional (or reversible jump) MCMC[67], which is of growing interest in geophysics[68-69]. The main attraction of this method is that the number of model parameters is treated as an unknown in the inversion, which decreases the impact of subjective choices made when deciding upon the number of model parameters to include. The posterior odds of any model parameterization are given by the frequency at which they are visited. Transdimensional MCMC is naturally parsimonious (i.e., it favors models with few parameters) as it incorporates the evidence in its calculations. It can be very useful when a large number of alternative model parameterizations are to be evaluated, but is less efficient when only a few different model parameterizations are to be compared, for which separate MCMC inversions followed by estimates of Bayes factors appear more efficient.[70] Bayes factors allow for a ranking of competing conceptual models, but it would be a mistake to only retain the most likely model for future uncertainty assessment if other competing models have similar Bayes factors. Bayesian model averaging (BMA)[71] allows for combining the posterior pdf derived from different conceptual models to better account for model uncertainty. Ref. 72 used BMA to combine predictions from two conceptual models that link geophysical data (self-potential) to the position of the groundwater table.

Hydrogeophysical developments relevant for model selection

To be useful for model selection, it is important that the hydrogeophysical parameter estimation framework is largely in agreement with the conceptual model to be tested or that the bias

introduced is considered. The first option is clearly preferable, but often implies considerable code development. One attractive approach is so-called fully–coupled inversion, in which an underlying hydrological model gives rise to hydrological state variables that are subsequently transformed into geophysical properties. Forward modeling is then performed to calculate the corresponding data misfits. The hydrological and petrophysical models are then varied within pre-defined ranges to find many models or one optimal model that adequately explain the observed data.[24,73-75] One challenge with the fully-coupled approach is the need to integrate the geophysical and hydrological modeling within the same modeling framework. Other more loosely coupled approaches have been proposed, in which the geophysical data are inverted using a flexible model regularization framework.

Classical inversion methods that imposes smoothness will, by construction, provide the smoothest model that is in agreement with the geophysical data.[29] It is possible to use alternative formulations to produce multiple realizations that agree with the data and, for example, uncertain semi-variogram information[76], possibly with different semi-variograms in different parts of the model region with sharp boundaries between pre-defined lithological units[77]. Such a regularization decoupling can be most effective and is easy to include in existing inversion algorithms. Furthermore, powerful freely available academic softwares are increasingly available to respond to this type of modeling needs (e.g., www.resistivity.net).

To illustrate the value of regularization decoupling across known boundaries, consider Ref 78. A gravel bar with a length of 250 m formed following the restoration of the Thur River in Switzerland was investigated using 3-D GPR and ERT. The processed GPR data clearly outline three distinct regions (see Figure 5a with unit boundaries highlighted): an upper region of the gravel aquifer predominantly consisting of gravel sheets (high signal coherency and layering), a lower region in the gravel aquifer with more heterogeneous deposits (e.g., foresets), and the underlying lacustrine clays. A standard 3-D ERT inversion produced a model (Figure 5b) showing decreasing resistivities with depth, but it is difficult to use this model to accurately locate the units that are so clearly seen in the GPR data. An alternative ERT inversion was performed, in which the model regularization was disconnected across the two GPR-defined boundaries and the known water table. The resulting model provides representative resistivities for the two gravel regions. The lower layer has a lower resistivity, which here indicates the presence of fines. It is thus likely that the upper part of the aquifer is more permeable than the lower part. This approach was used here to impose known information into the final model, but such an approach could also be used to test different assumptions about interfaces within a model selection framework.

User-defined regularization decoupling is effective if reliable information about the locations of the most important interfaces is available (e.g., from borehole logs, other types of geophysical data, or if it is part of the conceptual model to be tested). When the actual locations are unknown, it is possible to invert directly for sharp interfaces between more homogeneous units. For example, hydrogeological mapping in sedimentary environments has proven more efficient when smooth horizontally varying properties are sought together with the geometry of sharp layer interfaces.[79] Layered model parameterizations are unsuitable when lithological zones have lateral extents that are smaller than the model domain and other inversion strategies should then be used. For example, inversion based on level sets allows inverting for both the geometry of lithological units and their properties[80]. It is also possible to use clustering[81], data-driven zonation with Kalman filtering[82] or other zonation-promoting strategies[83] to derive models describing lithological units. Other promising

approaches involve the direct inference of geostatistical parameters from the geophysical data. Recently, an open-source inversion toolkit has been developed that not only allows obtaining multiple realizations that are in agreement with a given geostatistical model, but also to infer these parameters as part of the inversion procedure.[84-85] It is also possible to use GPR or seismic images to infer the aspect ratio of correlation lengths.[86]

Many hydrogeophysical studies use time-lapse data, for example, by repeating a sequence of measurements during a tracer test. It is well known that models obtained from classical smoothness-constrained geophysical inversions do not conserve injected mass and provide biased plume geometries.[87] Inversion methods have thus been developed that explicitly conserve the known injected mass and minimize plume distortions using model parameterizations related to geometrical moments of the plume.[88-89] Another approach is to not only invert for parameter values, but also for the most appropriate basis functions to represent the plume.[41]

If only a standard smoothness-based inversion model is available, it is necessary to account for the bias introduced by the regularization. The regularization can be seen as a spatially varying filtering process and its effects on parameter estimates has been investigated in many studies.[28,51,90-91] In a recent contribution[92], it was demonstrated how to produce multiple point statistics realizations that are in agreement with conceptual models in terms of geological facies models and smoothness-constrained tomograms. For a given section of a hydrogeological training image (Figure 6a), it is possible to apply a petrophysical transform to obtain a possible distribution of a geophysical property (Figure 6b). A synthetic experiment that mimics the acquisition of field data is performed and the noise-contaminated data are inverted to obtain a geophysical tomogram (Figure 6c). As discussed above, an underlying hydrological field (Figure 6d) is somehow related to a geophysical field (Figure 6e) from which the geophysicist can obtain a geophysical tomogram (Figure 6f). By repeating the transform and inversions of different sections of the training image, it is possible to obtain a large set of models that relate the high-resolution training images to the corresponding lower-resolution geophysical images. Under the assumption that the training image, the petrophysical model, and the simulation of the geophysical data acquisition and inversion process are realistic, it is possible, by using direct sampling[93] to find multiple realizations (Figure 6g) that have the same geological characteristics as the training image while being coherent with the field-based tomogram. This method can be adapted to many types of geophysical data and allows comparing models obtained under different assumptions of the geological setting.

To decrease non-uniqueness, it also helps to combine multiple types of geophysical data in a joint inversion[94-96] and to explore the full information content of the geophysical data[97]. There are also geophysical tools that are under-utilized, despite having the potential of being sensitive to key hydrological properties. Such methods include seismoelectrics[98], self-potential[99], time-lapse gravity[100], magnetic resonance sounding[101] and spectral induced polarization[102].

There are instances when the conceptual hydrological model imposes probabilistic approaches. A large set of hydrological (tracer, flowmeter, and televiewer) and geophysical (GPR imaging and monitoring during saline tracer injections) data acquired in two boreholes within a granitic formation were considered.[103] To understand transport pathways in the connected fracture networks that dominate flow and transport behavior at the site, it was necessary to parameterize the subsurface with a 3-D discrete fracture network model and derive large sets of fracture network realizations

conditioned to these data. The acquisition geometry (two boreholes only) does not allow constraining the exact positions of the fractures, thereby making it necessary to rely on a Monte Carlo method. Figure 7a illustrates how the fractures are connected and Figure 7b displays the simulated hydraulic head for one injection-withdrawal experiment. Obtaining only one such model is clearly limited as no meaningful uncertainty assessment can be made, while the analysis of large sets of conditioned models drawn from a formal posterior pdf allows capturing uncertainty. Different variations of this conceptual model (number of fractures, matrix properties, small-scale unresolved features, etc.) could be investigated using Bayes factors.

Advanced geophysical imaging and modeling should be employed in hydrogeophysics, not only to maximize the geophysical information content, but also to decrease the risk of biased estimates. One of the most commonly used methods in hydrogeophysics is borehole ERT, but it is only in the last few years that borehole effects have been considered in the inversion. The importance of doing this has been demonstrated in a saturated gravel aquifer, in which the boreholes were explicitly parameterized and the model regularization was disconnected between the aquifer and the boreholes.[104] The resulting model (Figure 8a) highlights a more resistive region (i.e., lower porosity) in the central part of the aquifer that is clearly outlined by seismic and GPR data.[96] If the boreholes are ignored, the resulting models are characterized by more resistive structures in the vicinity of the boreholes (Figure 8b) as a direct consequence of ignoring current channeling in the conductive water-filled boreholes. Indeed, joint inversion of multiple types of geophysical data could only provide coherent subsurface models when the borehole effects were considered.[96] It is thus clear that careful experimental design, field investigations and data interpretations using suitable geophysical modeling and interpretational tools are pre-requisites for consistently successful use of geophysical methods in hydrology.

CHALLENGES IN HYDROGEOPHYSICAL MODEL SELECTION

A Bayesian framework to model selection has the advantage that all choices in the modeling process must be carefully considered and quantified in probabilistic terms. Hierarchical[48] and transdimensional[68] Bayesian methods allow for including uncertain error sources, statistical properties about, say, correlation lengths, and unknown number of model parameters. In any probabilistic formulation, there will always be simplifications or assumptions made at some level (concerning modeling errors, truncation errors, the exact form of the prior distribution, data correlation, etc.) that often have important effects on the resulting models, especially when large data sets are considered[23]. Some authors[23] have drawn the radical conclusion to say that even a fully probabilistic framework will often provide too restrictive estimates of the acceptable model space. This problem of infinite regress (i.e., how to define the uncertainty of uncertainty estimates of uncertain parameters, and so on) makes it impossible to fully assess uncertainty within a formalistic framework.[55] It might happen that apparently mild assumptions or negligence of certain poorly understood error sources lead to situations where the parameter estimates do not capture the true underlying model. Using crosshole GPR to infer upscaled lower-dimensional representations of tracer plumes, it was found that the results were so adversely affected by unaccounted small-scale features that the most appropriate models (morphologically-speaking) were not part of the posterior estimates.[89,105] It is possible to parameterize this error and include it in the inversion, but such a parameterization will be imperfect and a model describing this imperfection will also be imperfect, etc. Many error sources will only be approximately known, so it is therefore important to derive pragmatic and robust approaches to model selection.

CONCLUSIONS

The Bayes factor or transdimensional inversion offers a theoretically satisfying, but largely unexplored, approach to compare alternative conceptual hydrological models using geophysical and hydrological data. Its routine use is presently hindered by computational costs and multiple error sources that are only approximately known. The extent to which geophysical data can be used to falsify or corroborate conceptual models depends on their complexity, the underlying subsurface heterogeneity, the choice of geophysical data and survey design, as well as the knowledge about petrophysical relationships linking geophysical estimates to the hydrological properties of interest. Simplifying assumptions at any level lead (if unaccounted) to biased assessments and possibly misleading conclusions. The use of Bayes factors has the added advantage that it demands clear statements about the state of knowledge concerning the different steps in the hydrogeophysical parameter estimation process. In the future, it is necessary to gain experience with the Bayes factor and its asymptotic solutions in field-based hydrogeophysical studies that cover a wide range of application types. Model selection and ranking of conceptual hydrological models is perhaps the best way for hydrologists to gain the most from geophysics and for geophysicists to have a larger impact on hydrology. There are many “easy”, but important problems that geophysics can help to solve, but it is important to accommodate any geophysically inferred hydrological results with clear assessments about uncertainty, assumptions and the “prior information” used to assure that hydrogeophysics remains credible in the long run.

ACKNOWLEDGEMENTS

This study was partly funded by the Swiss National Science Foundation and is a contribution to the ENSEMBLE project (grant no. CRSI22_132249). I thank the members of ENSEMBLE and my students and post-doc for discussions that helped to form this manuscript. I am grateful to Thomas Günther who provided the software needed to produce Figure 1. Comments from two anonymous reviewers and Associate Editor Stuart Lane helped to improve the manuscript.

REFERENCES

1. Rubin Y, Hubbard SS, editors. Hydrogeophysics. Springer, Dordrecht, 2005, 523.

2. Binley A, Cassiani G, Deiana R. Hydrogeophysics: opportunities and challenges. *B Geofis Teor Appl* 2010, 51:267-284.

3. Hubbard SS, Linde N. Hydrogeophysics. In: Uhlenbrook S, editor. Treatise on Water Science. Vol. 2. Elsevier, Amsterdam, 2011, 401-434.

4. Hubbard SS, Rubin Y. Hydrogeological parameter estimation using geophysical data: a review of selected techniques. *J Contam Hydrol* 2000, 45:3-34.

5. Odoni NA, Lane SN. Knowledge-theoretic models in hydrology. *Prog Phys Geog* 2010, 34:151-171.

6. Rubin Y, Mavko G, Harris J. Mapping permeability in heterogeneous aquifers using hydrological and seismic data. *Water Resour Res* 1992, 28:1809-1816.

7. Linde N, Chen J, Kowalsky MB, Hubbard S. Hydrogeophysical parameter estimation approaches for field scale characterization. In: Vereecken H, Binley A, Cassiani G, Revil A, Titov K, editors. Applied Hydrogeophysics. Springer, Dordrecht, 2006, 9-44.

8. Slater L. Near surface electrical characterization of hydraulic conductivity: From petrophysical properties to aquifer geometries—A Review. *Surv Geophys* 2007, 28:169-197.

9. Robinson DA, Binley A, Crook N, Day-Lewis FD, Ferre TPA, Grauch VJS, Knight R, Knoll M, Lakshmi V, Miller R, et al. Advancing process-based watershed hydrological research using near-surface geophysics: a vision for, and review of, electrical and magnetic methods. *Hydrol Process* 2008, 22:3604-3635.

10. Revil A, Karoulis M, Johnson T, Kemna A. Review: Some low-frequency electrical methods for subsurface characterization and monitoring in hydrogeology. *Hydrogeol J* 2012, 20:617-658.

11. Menke W. Geophysical data analysis: Discrete inverse theory, revised edition. Academic Press, San Diego, 1989, 289.

12. Tarantola A. Inverse problem theory and methods for model parameter estimation. SIAM, Philadelphia, 2005, 342.

13. McLaughlin D, Townley LR. A reassessment of the groundwater inverse problem. *Water Resour. Res.* 1996, 32:1131-1161.

14. Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ. Inverse problem in hydrogeology. *Hydrogeol J* 2005, 13:206-222.

15. Everett ME, Near-surface applied geophysics. Cambridge University Press, Cambridge, 2013, 403.

16. Gupta HV, Clark MP, Vrugt JA, Abramowitz G. Towards a comprehensive assessment of model structural adequacy. *Water Resour Res* 2012, 48:W08301.

17. Zimmerman DA, de Marsily G, Gotway CA, Marietta MG, Axness CL, Beauheim RL, Bras RL, Carrera J, Dagen G, Davies PB, et al. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. *Water Resour Res* 1998, 34:1373-1413.

18. Refsgaard JC, van der Sluijs JP, Brown J, van der Keur P. A framework for dealing with uncertainty due to model structure error. *Adv Water Resour* 2006, 29:1586-1597.

19. Bond CE, Gibbs AD, Shipton ZK, Jones S. What do you think this is? “Conceptual uncertainty” in geoscience interpretation. *GSA Today* 2007, 17(11):4-10.

20. Oreskes N, Shrader-Frechette K, Belitz K. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 1994, 263:641-646.

21. Backus G, Gilbert F. Uniqueness in the inversion of inaccurate gross earth data. *Philos Tr R Soc S-A* 1970, 266:123-192.

22. Konikow LF, Bredehoeft JD. Groundwater models cannot be validated. *Adv Water Resour* 1992, 15:75-83.

23. Beven K. A manifesto for the equifinality thesis. *J Hydrol* 2006, 320:18-36.

24. Kowalsky MB, Finsterle S, Peterson J, Hubbard S, Rubin Y, Majer E, Ward A, Gee G. Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data. *Water Resour Res* 2005, 41:W11425.

25. Caers J. Modeling uncertainty in the earth sciences. Wiley-Blackwell, Chichester, 2011, 246.

26. Maurer H, Curtis A, Boerner DE. Recent advances in optimized geophysical survey design. *Geophysics* 2010, 75:A177-A194.

27. Parker RL. The inverse problem of resistivity sounding. *Geophysics* 1984, 49:2143-2158.

28. Oldenburg DW, Li YG. Estimating depth of investigation in dc resistivity and IP surveys. *Geophysics* 1999, 64:403-416.

29. Constable SC, Parker RL, Constable CG. Occam’s inversion—A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics* 1987, 52:289-300.

30. Ellis RG, Oldenburg DW. Applied geophysical inversion. *Geophys J Int* 1994, 116:5-11.

31. Scales JA, Snieder R. To Bayes or not to Bayes? *Geophysics* 1997, 62:1045-1046.

32. Feyen L, Caers J. Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations. *Adv Water Resour* 2006, 29:912-919.

33. Gómez-Hernández JJ, Wen X-H. To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Adv Water Resour* 1998, 21:47-61.

34. Kerrou J, Renard P, Franssen HJH, Lunati I. Issues in characterizing heterogeneity and connectivity in non multiGaussian media. *Adv Water Resour* 2008, 31:147-159.

35. Kitanidis PK. Quasi-linear geostatistical theory for inversing. *Water Resour Res* 1995, 31:2411-2419.

36. Cardiff M, Barrash W, Kitanidis PK. A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily-emplaced equipment. *Water Resour Res* 2012, 48:W05531.

37. Hu LY, Chugunova T. Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. Water Resour Res, 44:W11413.

38. Jakeman AJ, Hornberger GM. How much complexity is warranted in a rainfall-runoff model. Water Resour Res 1993, 29:2637-2649.

39. Wu YS, Haukwa C, Bodvarsson GS. A site-scale model for fluid and heat flow in the unsaturated zone of Yucca Mountain, Nevada. *J Contam Hydrol* 1999, 38:185-215.

40. Freer J, McDonnell JJ, Beven KJ, Peters NE, Burns DA, Hooper RP, Aulenbach B, Kendall C. The role of bedrock topography on subsurface storm flow. *Water Resour Res* 2002, 38:1269.

41. Lien M. Simultaneous joint inversion of amplitude-versus-offset and controlled-source electromagnetic data by implicit representation of common parameter structure. *Geophysics* 2013, 78:ID15-ID27.

42. Schön JH. Physical properties of rocks: Fundamentals and principles of petrophysics. Elsevier, Oxford, 2004, 584.

43. Mavko G, Mukerji T, Dvorkin J. The rock physics handbook: tools for seismic analysis of porous media, 2nd edition. Cambridge University Press, Cambridge, 2009, 524.

44. Lesmes DP, Friedman SP. Relationships between the electrical and hydrogeological properties of rocks and soils. In Rubin Y, Hubbard SS, editors. Hydrogeophysics. Springer, Dordrecht, 2005, 523.

45. Rücker C, Günther T, Spitzer K. Three-dimensional modeling and inversion of dc resistivity data incorporating topography –I. Modeling. *Geophys J Int* 2006, 166:495-505.

46. Maurer H, Green AG. Potential coordinate mislocations in crosshole tomography: Results from the Grimsel test site, Switzerland. *Geophysics* 1997, 62:1696-1709.

47. Cordua KS, Nielsen L, Looms MC, Hansen TM, Binley A. Quantifying the influence of static-like errors in least-squares-based inversion and sequential simulation of cross-borehole ground penetrating radar data. *J Appl Geophys* 2009, 68:71-84.

48. Malinverno A, Briggs VA. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics* 2004, 69:1005-1016.

49. Claerbout JF, Muir F. Robust modeling with erratic data. *Geophysics* 1973, 38:826-844.

50. Scales JA, and Tenorio L. Prior information and uncertainty in inverse problems. *Geophysics* 2001, 66:389-397.

51. Day-Lewis FD, Singha K, Binley AM. Applying petrophysical models to radar travel time and electrical resistivity tomograms: Resolution-dependent limitations. *J Geophys Res* 2005, 110:B08206.

52. Looms MC, Binley A, Jensen KH, Nielsen L, and Hansen TM. Identifying unsaturated hydraulic parameters using an integrated data fusion approach on cross-borehole geophysical data. *Vadose Zone J* 2008, 7:238-248.

53. Binley A, Beven K. Vadose zone flow uncertainty as conditioned on geophysical data. *Ground Water* 2003, 41:119-127.

54. Farmani MB, Kitterod NO, Keers H. Inverse modeling of unsaturated flow parameters using dynamic geological structure conditioned by GPR tomography. *Water Resour Res* 2008, 44:W08401.

55. Carrera J, Martinez-Landa L, Perez-Estaun P, Vasquez-Sune E. Geophysics and hydrogeology: will they ever marry? Abstract H42A-02 presented at the 2012 Fall Meeting, AGU, San Francisco, Calif., 3-7 Dec. 2012.

56. Jarvis NJ, Bergström L, Dik PE. Modeling water and solute transport in macroporous soil. 2. Chloride breakthrough under nonsteady flow. *J Soil Sci* 1991, 42:71-81.

57. Ronayne MJ, Gorelick AM, Zheng CM. Geological modeling of submeter scale heterogeneity and its influence on tracer transport in a fluvial aquifer. *Water Resour Res* 2010, 46:W10519.

58. Funtowicz SO, Ravetz JR. Uncertainty and quality in science for policy. Kluwer Academic Publishers, Dordrecht, 1990, 229.

59. Taleb NN. The black swan: The impact of the highly improbable. Random House, New York, 2007, 366.

60. Funtowicz SO, Ravetz JR. Science for the post-normal age. *Futures* 1993, 25:739-755.

61. Jeffreys H. Some tests of significance, treated by the theory of probability. *Math Proc Cambridge* 1935, 31:203-222.

62. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995, 90:773-795.

63. Laloy E, Vrugt JA. High-dimensional posterior exploration of hydrological models using multiple-try DREAM$_{(ZS)}$ and high-performance computing. *Water Resour Res* 2012, 48:W01526.

64. Berk RA, Statistical learning from a regression perspective. Springer, New York, 2010, 358.

65. Khan A, Mosegaard K. Further constraints on the deep lunar interior. *Geophys Res Lett* 2005, 32:L22203.

66. Dettmer J, Dosso SE, Osler JC. Bayesian evidence computation for model selection in non-linear geoacoustic inference problems. *J Acoust Soc Am* 2010, 128:3406-3415.

67. Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995, 82:711-732.

68. Malinverno A. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophys J Int* 2002, 151:675-688.

69. Bodin T, Sambridge M, Rawlinson N, Arroucau P. Transdimensional tomography with unknown data noise. *Geophys J Int* 2012, 189:1536-1556.

70. Chib S, Jeliazkov I. Marginal likelihood from the Metropolis-Hastings output. *J Am Stat Assoc* 2001, 96:270-281.

71. Draper D. Assessment and propagation of model uncertainty. *J Roy Stat Soc Ser B* 1995, 57:45-97.

72. Linde N, Revil A, Bolève A, Dagès C, Castermant J, Suski B, Voltz M. Estimation of the water table throughout a catchment using self-potential and piezometric data in a Bayesian framework. *J Hydrol* 2006, 334:88-98.

73. Irving J, Singha K. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. *Water Resour Res* 2010, 46:W11514.

74. Hinnell AC, Ferre TPA, Vrugt JA, Huisman JA, Moysey S, Rings J, Kowalsky MB. Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. *Water Resour Res* 2010, 46:WOOD40.

75. Pollock D, Cirpka OA. Fully coupled hydrogeophysical inversion of a laboratory salt tracer experiment monitored by electrical resistivity tomography. *Water Resour Res* 2011, 48:W01505.

76. Johnson T, Routh PS, Clemo T, Barrash W, and Clement WP. Incorporating geostatistical constraints in nonlinear inversion problems. *Water Resour Res* 2007, 43:W10422.

77. Johnson TC, Versteeg RJ, Rockhold M, Slater LD, Ntargliannis D, Greenwood WJ, Zachara J. Characterization of a contaminated wellfield using 3D electrical resistivity tomography implemented with geostatistical, discontinuous boundary, and known conductivity constraints. *Geophysics* 2012, 77:EN85-EN96.

78. Doetsch J, Linde N, Pessognelli M, Green AG, Günther T. Constraining 3-D electrical resistance tomography with GPR reflection data for improved aquifer characterization. *J Appl Geophys* 2012, 78:68-76.

79. Auken E, Christiansen AV. Layered and laterally constrained 2D inversion of resistivity data. *Geophysics* 2004, 69:752-764.

80. Cardiff M, Kitanidis PK, Bayesian inversion for facies detection: An extensible level set framework. *Water Resour Res* 2009, 45:W10416.

81. Paasche H, Tronicke J. Cooperative inversion of 2D geophysical data sets: A zonal approach based on fuzzy c-means cluster analysis. *Geophysics* 2007, 72:A35-A39.

82. Eppstein MJ, Dougherty DE. Efficient three-dimensional data inversion: Soil characterization and moisture monitoring from cross-well ground-penetrating radar at a Vermont test site. *Water Resour Res* 1998, 34:1889-1900.

83. Hyndman DW, Harris JM, Gorelick SM. Coupled seismic and tracer test inversion for aquifer property characterization. *Water Resour Resear* 1994, 30:1965-1977.

84. Hansen TM, Cordua KS, Looms MJ, Mosegaard K. SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 1–Methodology. *Comput Geosci* 2013, 52:470-480.

85. Hansen TM, Cordua KS, Looms MJ, Mosegaard K. SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 2–Application to crosshole GPR tomography. *Comput Geosci* 2013, 52:481-492.

86. Irving J, Holliger K. Geostatistical inversion of seismic and ground-penetrating radar reflection images: What can we actually resolve? *Geophys Res Lett* 2010, 37:L21306.

87. Day-Lewis FD, Chen Y, Singha K. Moment inference from tomograms. *Geophys Res Lett* 2007, 34:L22404.

88. Pidlisecky A, Singha K, Day-Lewis FD. A distribution-based parameterization for improved tomographic imaging of solute plumes. *Geophys J Int* 2011, 187:214-224.

89. Laloy E, Linde N, Vrugt JA. Mass conservative three-dimensional water tracer distribution from MCMC inversion of time-lapse GPR data. *Water Resour Res* 2012, 48:W01526.

90. Moysey S, Singha K, Knight R. A framework for inferring field-scale rock physics relationships through numerical simulation. *Geophys Res Lett* 2005, 32:L08304.

91. Singha K, Moysey S. Accounting for spatially variable resolution in electrical resistivity tomography through field-scale rock-physics relations. *Geophysics* 2006, 71:A25-A28.

92. Lochbühler T, Pirot G, Straubhaar J, Linde N. Conditioning of multiple-point statistics facies simulations to tomographic images. *Math Geosci*. 2013 (in press).

93. Mariethoz G, Renard P, Straubhaar J. The Direct Sampling method to perform multiple-point geostatistical simulations. *Water Resour Res* 2010, 46:W11536.

94. Lochbühler T, Doetsch J, Brauchler R, Linde N. Structure-coupled joint inversion of geophysical and hydrological data. *Geophysics* 2013, 78:ID1–ID14.

95. Gallardo LA, Meju MA. Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradient constraints. *J Geophys Res* 2004, 109:B03311.

96. Doetsch J, Linde N, Coscia I, Greenhalgh SA, Green AG. Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data. *Geophysics* 2010 75:G53-G64.

97. Ernst JR, Green AG, Maurer H, Holliger K. Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data. *Geophysics* 2007, 72:J53-J64.

98. Dupuis JC, Butler KE, Kepic AW. Seismoelectric imaging of the vadose zone of a sand aquifer. *Geophysics* 2007, 72:A81A85.

99. Revil A, Jardani A. The self-potential method: Theory and applications in environmental geosciences. Cambridge University Press, Cambridge, 2013, 369.

100. Davis K, Li YG, Batzle M. Time-lapse gravity monitoring: A systematic 4D approach with application to aquifer storage and recovery. *Geophysics* 2008, 73:WA61-WA69.

101. Lehmann-Horn JA, Walbrecker JO, Hertrich M, Langston G, McClymont AF, Green AG. Imaging groundwater beneath a rugged proglacial moraine. *Geophysics* 2011, 76:B165-B172.

102. Borner FD, Schopper JR, Weller A. Evaluation of transport and storage properties in the soil and groundwater zone from induced polarization measurements. *Geophys Prospect* 1996, 44:583-601.

103. Dorn C, Linde N, Le Borgne T, Bour O, De Dreuzy, J-R, Conditioning of stochastic 3-D fracture networks to hydrological and geophysical data. *Adv Water Resour*. 2013 (in press).

104. Doetsch JA, Coscia I, Greenhalgh S, Linde N, Green A, Günther T. The borehole–fluid effect in electrical resistivity imaging. *Geophysics* 2010, 75:F107-F114.

105. Linde N, Vrugt JA. Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion. *Vadose Zone J* 2013, 12.

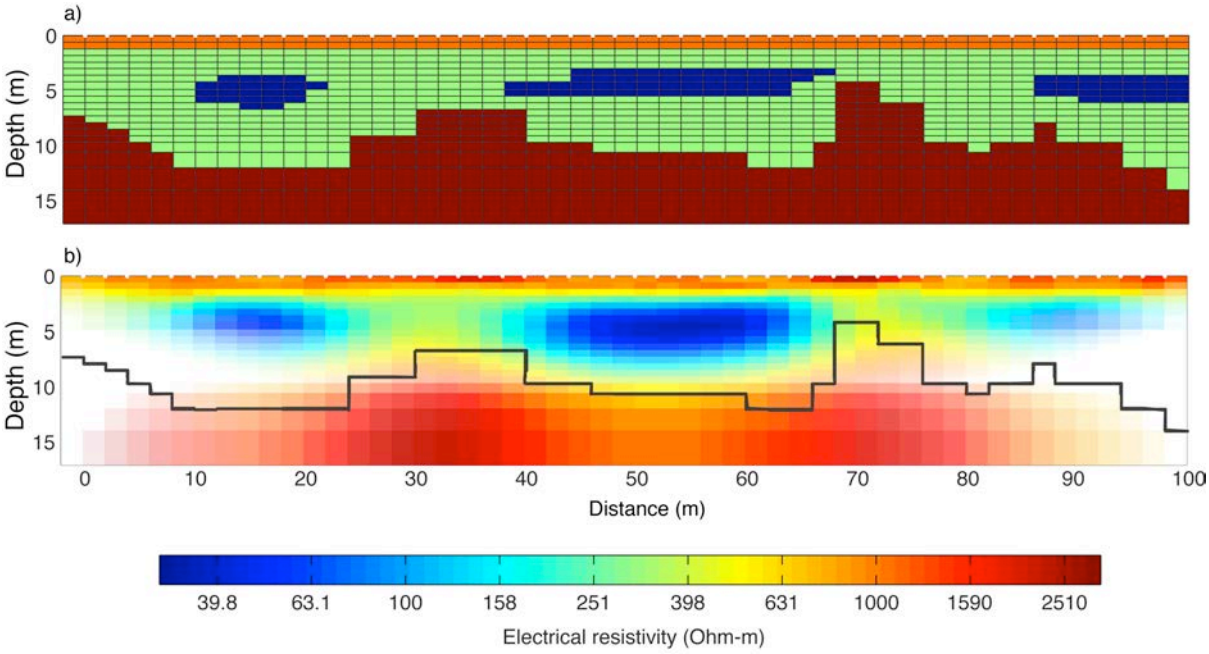

Figure 1: (a) Test case representing a gravel aquifer with embedded clay lenses overlying bedrock. (b) The result of a standard smoothness-constrained inversion of ERT data acquired on the ground surface clearly indicates the presence of bedrock, but is unable to resolve detailed variations in the bedrock topography or its actual resistivity.

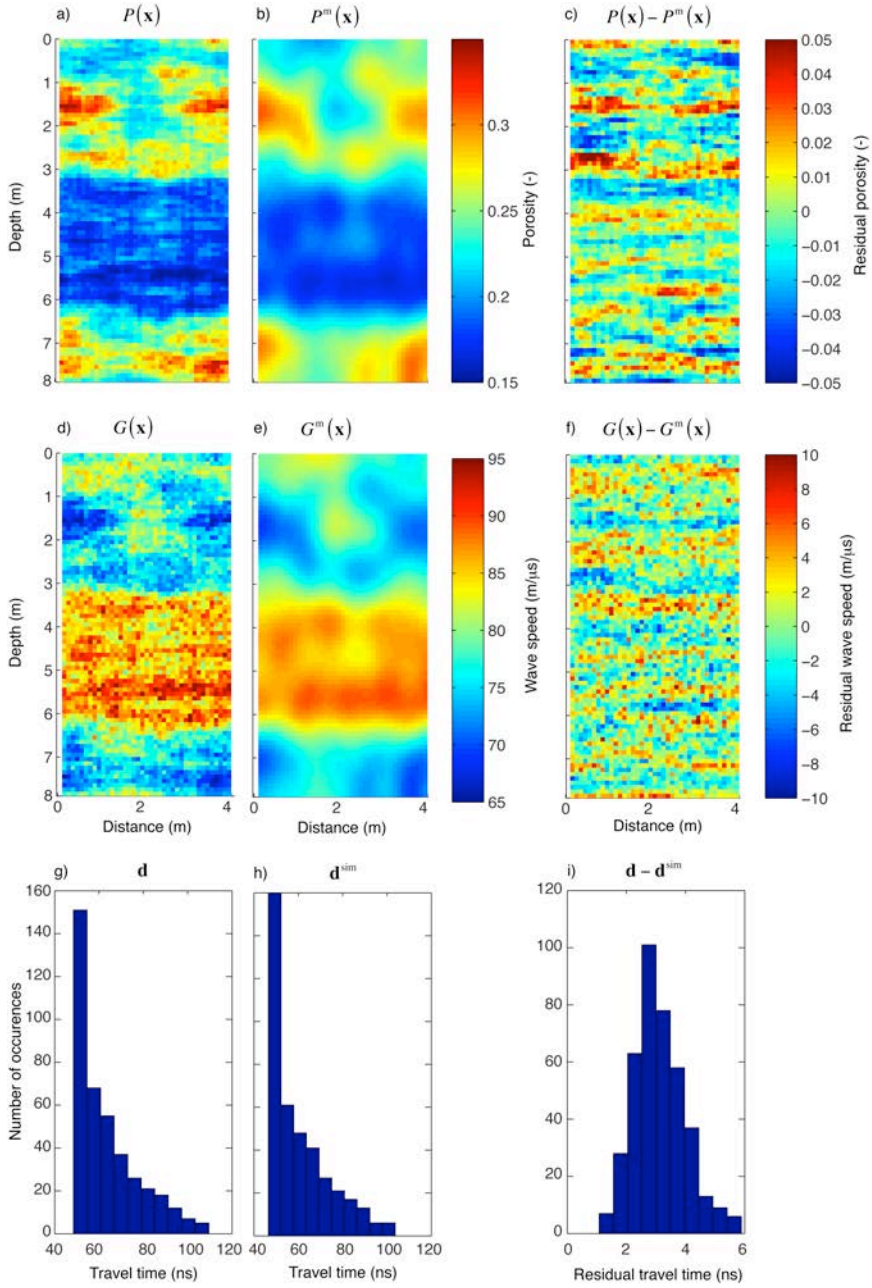

Figure 2: Example illustrating errors associated with the model parameterization, uncertain petrophysical relationships, geometrical, and observational errors. The (a) actual porosity field cannot be fully described by (b) the model parameterization used, which leads to (c) upscaling errors. The actual porosity field is related to a (d) geophysical field (radar wave speed) through a petrophysical relationship. This relationship is never perfectly known, so additional uncertainty and bias occur when using a petrophysical model to represent the (e) expected wave speed as shown in the (f) model residuals. The (g) observed geophysical data are contaminated with observational errors, while the (h) calculated model response of (e) is affected by errors in the forward model or in the geometry, which contribute to (i) data residuals that are biased. These types of errors can be minimized by careful investigations, but not fully removed. Their expected distributions should be carefully considered when inverting (h) in order to retrieve a representative model similar to (b).

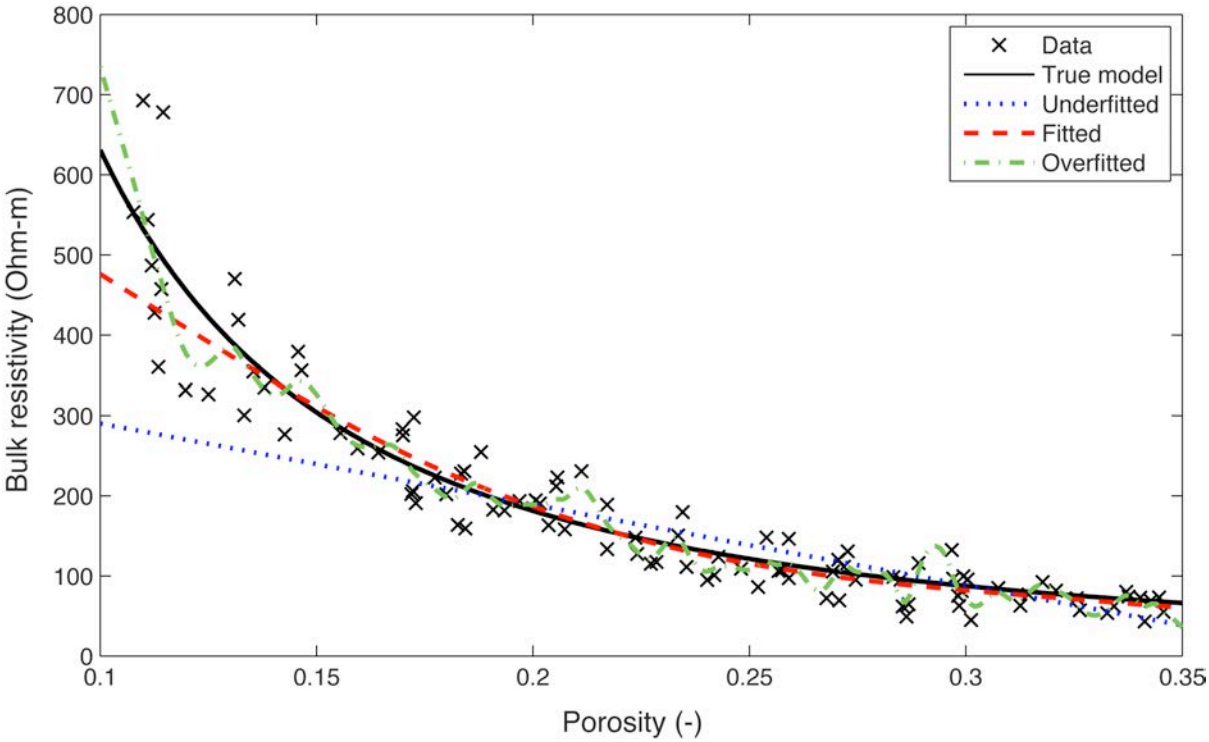

Figure 3: A hypothetical set of 100 noisy data describing the observed relationships between porosity and bulk resistivity. Three models obtained by regularized smoothness-constrained inversion display decreasing data misfits (underfitted; fitted; and overfitted), but also increasing model complexity (i.e., more oscillations). The most likely model based on the Schwartz criterion is the fitted model, but it is far from being as satisfactory as the two-parameter model used to generate the data. A careful choice of model parameterization is thus most important in hydrogeophysical studies.

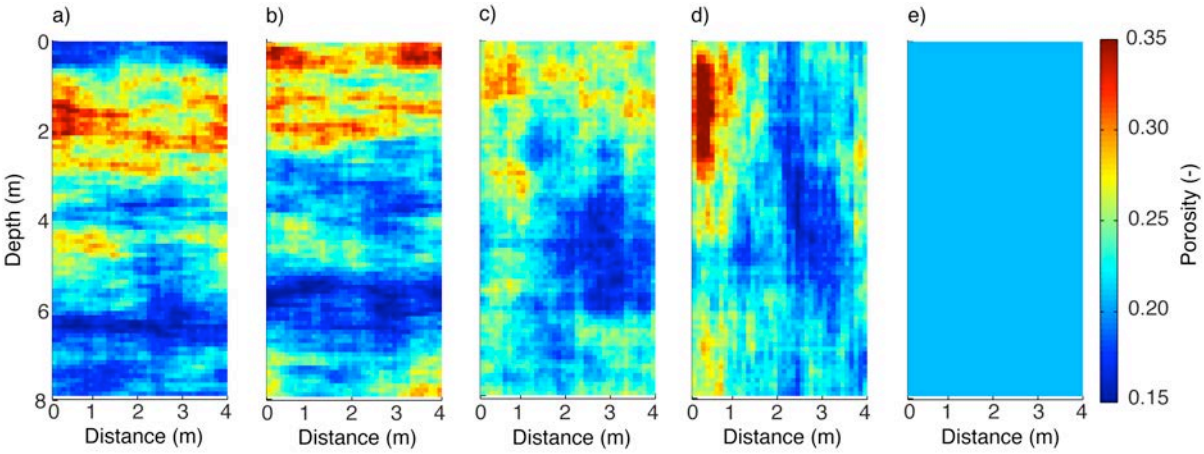

Figure 4. (a) The actual porosity field used to generate noise-contaminated data. From 20,000 random draws, (b-e) show for different conceptual models the realizations that best explain the noise-contaminated crosshole GPR data generated by (a). The true geostatistical model is known in case (b), isotropy is assumed in (c), vertical anisotropy is assumed in (d) and uniform porosity is assumed in (e).

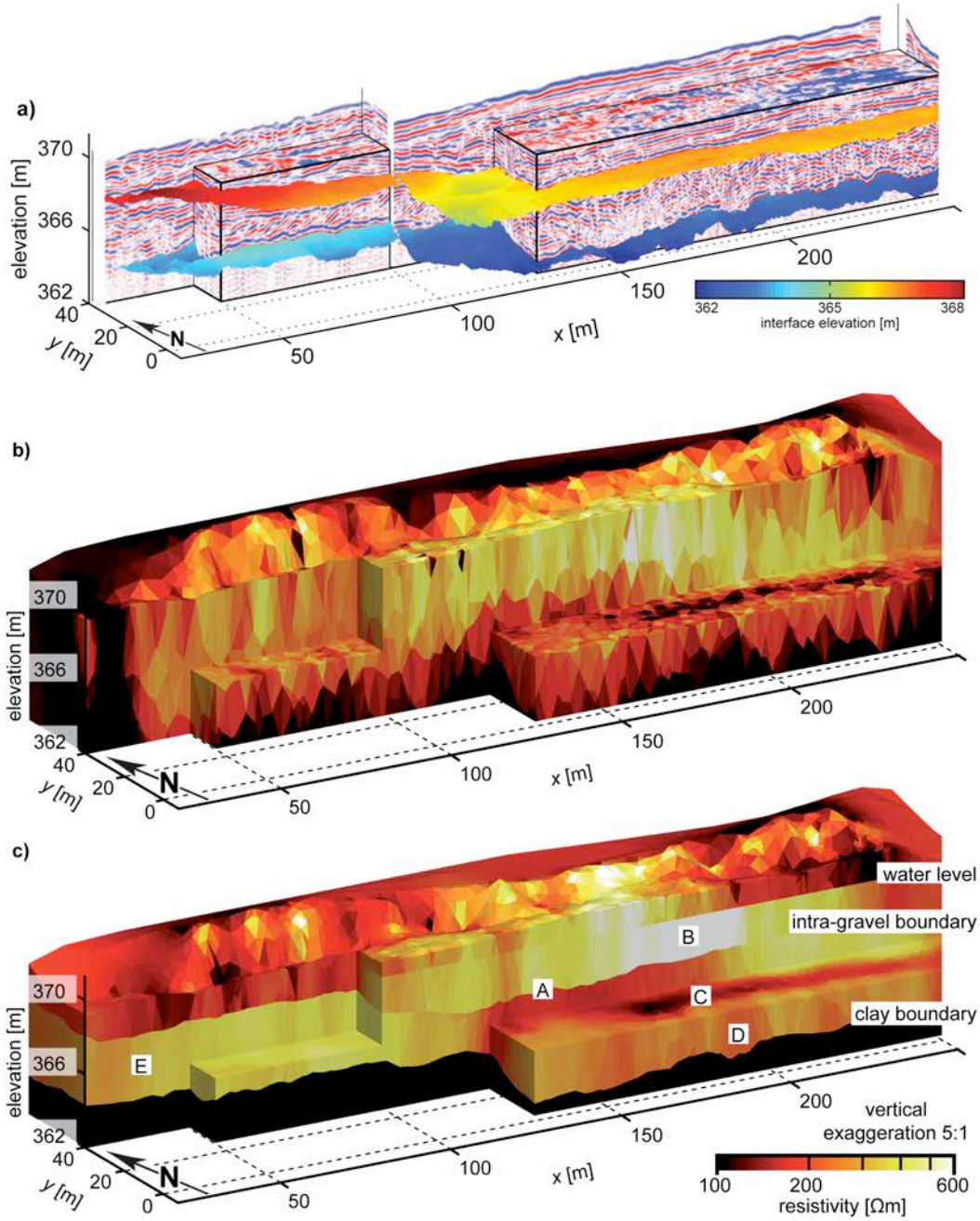

Figure 5: (a) Chair plot of a GPR data volume acquired across a gravel bar in the vicinity of the Thur River in Switzerland. (b) ERT inversion results using standard smoothness-constrained model regularization and (c) for the case, in which no model regularization is imposed across GPR-defined boundaries outlined in (a) and across the water table. The regularization decoupling enforces similarity between the GPR and ERT results, thus facilitating interpretation. Furthermore, the constraints from high-resolution GPR (very sensitive to lithological boundaries) improve the estimates of bulk electrical properties. Adapted from Figures 5 and 8 in Ref. 78.

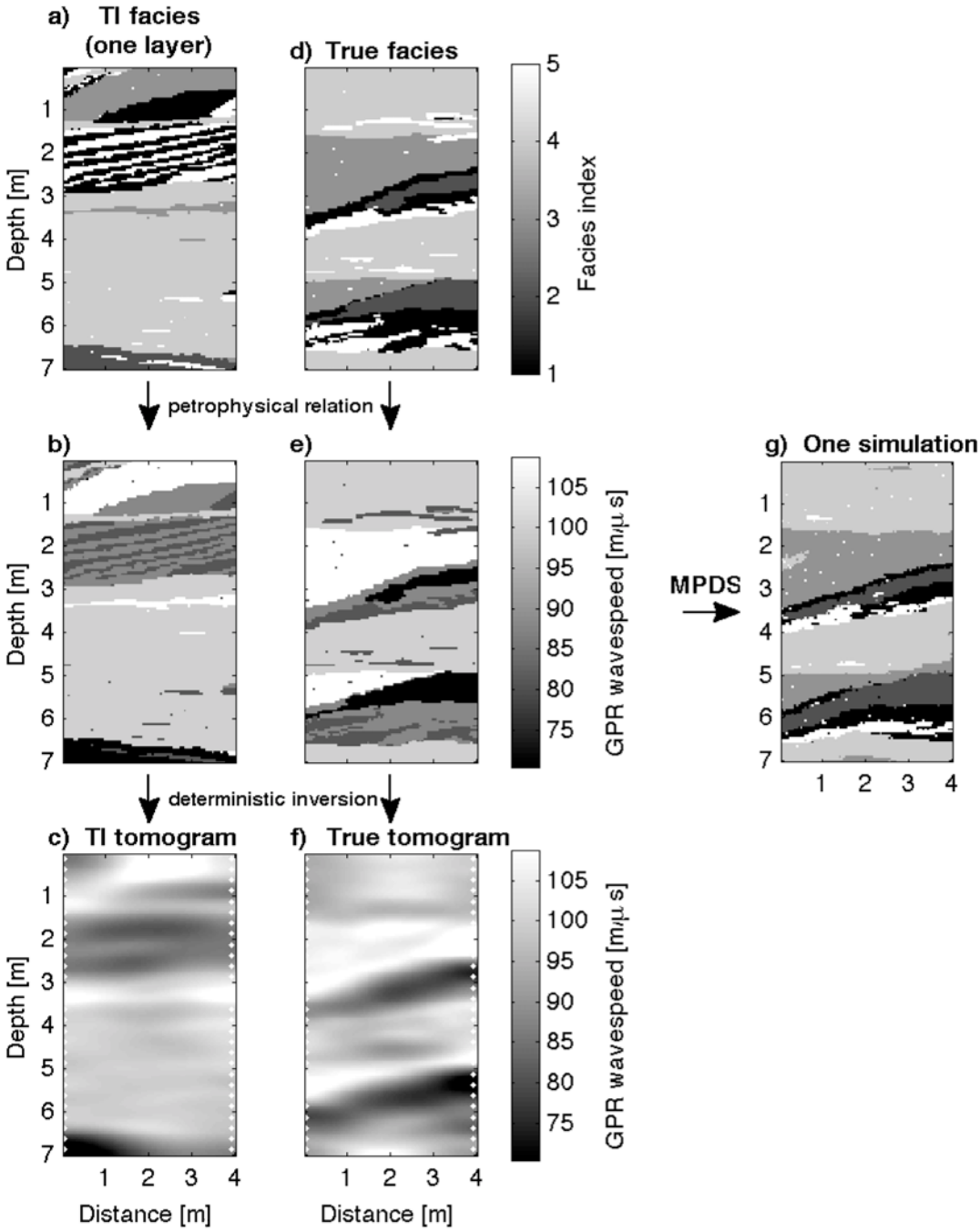

Figure 6: Conditioning of multiple-point statistics facies simulations to geophysical tomograms. Many sections of a (a) training image (TI) that is representative of the expected depositional setting is generated and transformed using a petrophysical relation into (b) geophysical fields on which realistic geophysical data are generated and inverted to obtain a (c) tomogram. The real (d) underlying facies distribution is related to an underlying (e) geophysical property field from which a geophysicist can derive a lower resolution (f) tomogram. Using direct sampling[93], it is possible to generate (g) multiple realizations of subsurface models that are coherent with both the geophysical tomogram and the training image.

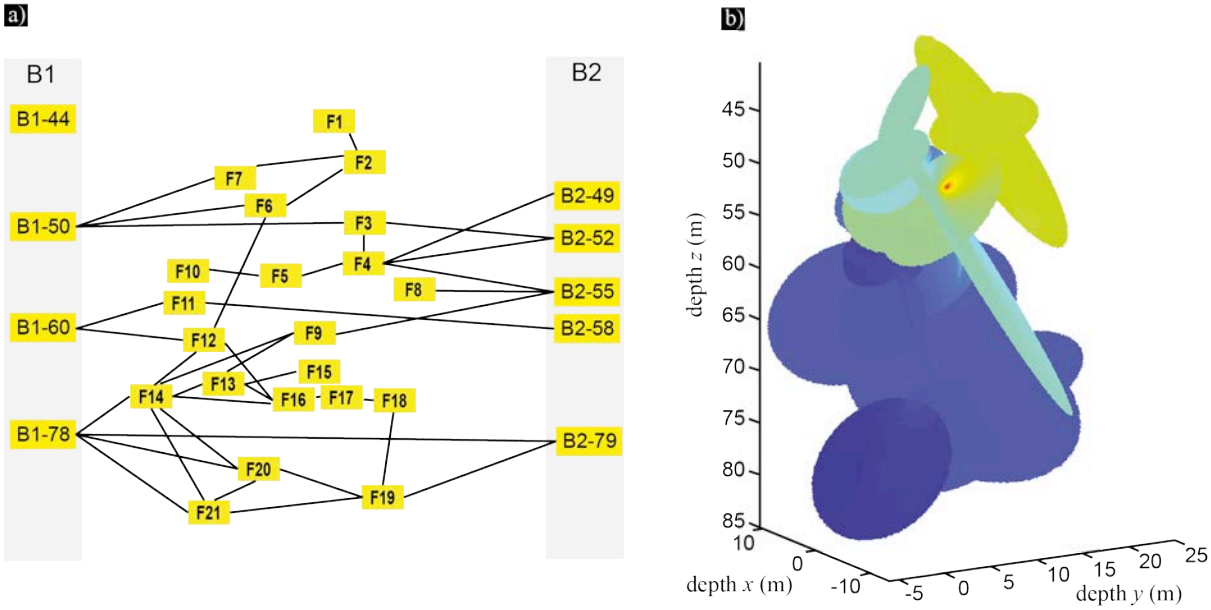

Figure 7: One realization of a discrete fracture network model that is conditioned to a large set of geophysical and hydrological data acquired at the Ploemeur research site in France. (a) Graph representation of the network and (b) the simulated hydraulic head distribution for an injection-withdrawal experiment.

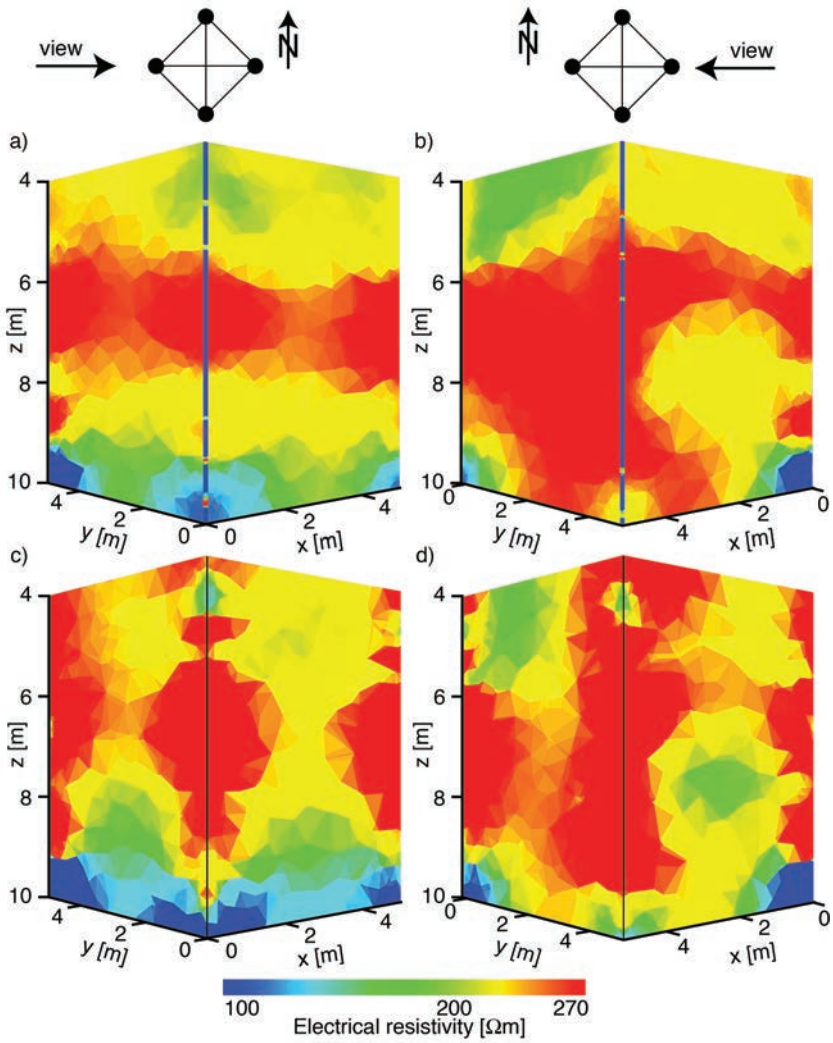

Figure 8: (a-b) Different views of a crosshole 3-D ERT inversion model for which borehole effects were explicitly included in the modeling. The high resistive central region (low porosity) is in agreement with models obtained by other geophysical models. (c-d) Corresponding views for an inversion model obtained without considering borehole effects. These latter results are of limited value in building a predictive aquifer model. Figure 6 from Ref. 103.

Related Articles

DOI	Article title
10.1002/wics.199	The Bayesian information criterion: background, derivation, and applications
10.1002/wics.194	Monte Carlo methods
10.1002/wcc.220	Ensemble modeling, uncertainty and robust predictions