Water Resources Research

RESEARCH ARTICLE
10.1029/2023WR034992

# A Novel Methodology for the Stochastic Integration of Geophysical and Hydrogeological Data in Geologically Consistent Models

**Alexis Neven**[1] **and Philippe Renard**[1,2]

[1]Centre for Hydrogeology and Geothermics, University of Neuchâtel, Neuchâtel, Switzerland, [2]Department of Geoscience, University of Oslo, Oslo, Norway

**Key Points:**

- A methodology is proposed to assimilate geological, hydrogeological, and geophysical data in consistent stochastic models
- The methodology combines a hierarchical geological modeling technique with the ensemble smoother with multiple data assimilation
- The applicability of the methodology is demonstrated using actual field data from the upper Aare valley aquifer in Switzerland

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

A. Neven,
alexis.neven@unine.ch

**Citation:**

Neven, A., & Renard, P. (2023). A novel methodology for the stochastic integration of geophysical and hydrogeological data in geologically consistent models. *Water Resources Research*, *59*, e2023WR034992. https://doi.org/10.1029/2023WR034992

Received 3 APR 2023
Accepted 23 JUN 2023

**Author Contributions:**

**Conceptualization:** Alexis Neven, Philippe Renard
**Funding acquisition:** Philippe Renard
**Methodology:** Alexis Neven, Philippe Renard
**Software:** Alexis Neven
**Supervision:** Philippe Renard
**Visualization:** Alexis Neven
**Writing – original draft:** Alexis Neven, Philippe Renard
**Writing – review & editing:** Alexis Neven, Philippe Renard



**Abstract** To address groundwater issues, it is often necessary to develop geological and hydrogeological models. Combining geological, geophysical and hydrogeological data available on a site to build such models is often a challenge. This paper presents a methodology to integrate such data within a geologically consistent model with robust error estimation. The methodology combines the Ensemble Smoother with Multiple Data Assimilation (ESMDA) algorithm with a hierarchical geological modeling approach (ArchPy). Geophysical and hydrogeological field data are jointly assimilated in a stochastic ESMDA framework. To speed up the inversion process, forward responses are computed in lower-dimensional spaces relevant to each physical problem. By doing so, the final models take into account multiple data sources and regional conceptual geological knowledge. This study illustrates the applicability of this novel approach using actual data from the upper Aare Valley, Switzerland. The results of cross-validation show that the combination of different data types, each sensitive to different spatial dimensions, enhances the quality of the model within a reasonable computing time. The proposed methodology allows the automatic generation of groundwater models with robust uncertainty estimation and could be applied to a wide variety of hydrogeological issues.

**Plain Language Summary** When dealing with groundwater, it is necessary to develop underground models. However, taking into account all the different data types on a site is time-consuming, and there is a lack of uncertainty quantification. In this study, we develop an approach that automatically combines different types of data, including boreholes, geophysical data, and hydrogeological measurements. All data are combined using a stochastic algorithm and produce an ensemble of plausible and data-compatible models. These models can be used, for example, to forecast groundwater availability, pollutant distribution, or the effect of climate change on groundwater.

## 1. Introduction

Groundwater resources represent 99% of the total liquid fresh water on Earth and provides approximately 50% of the total volume of water withdrawn for domestic use in the world (United Nations Educational Scientific and Cultural Organization, 2022). However, because the resource is at depth below the ground surface, the interest of the general public and policy makers tends to wane due to its invisible aspect (United Nations Educational Scientific and Cultural Organization, 2022). In a changing climate, being able to understand, predict and model underground flows and resources is crucial. These questions can only be properly addressed by local to regional groundwater modeling.

Another issue owing to the underground position of groundwater is the difficulty in collecting data. Data are usually sparse and often represent only a tiny fraction of the total volume of the aquifer. The data can therefore miss most of the spatial heterogeneity. Conjectures based on such groundwater models could lead to inadequate decisions. A recent paper by Adams et al. (2022) reviews the different existing remote sensing techniques (including geophysical techniques) and possible future approaches that could help fill this gap. They argue that satellite or airborne remote sensing could be key methods for understanding groundwater dynamics at various scales, which would have been impossible with classical spare in situ observations. One method mentioned in their paper is the electromagnetic induction techniques, and especially the Time Domain Electromagnetic (TEM). The TEM method is relatively inexpensive, fast and has a depth of investigation ranging from a few meters to a few hundred meters (Christiansen et al., 2009; Fitterman, 1987). Due to its inductive principle, the Electromagnetic Magnetic (EM) method will be mainly sensitive to underground electrical resistivity. Consequently, it will also be sensitive

to all parameters that will affect this property, such as porosity, water saturation, clay content, and water electrical conductivity. EM was successfully applied to various hydrogeological situations, usually in combination with sparse borehole data (e.g., Barfod et al., 2018; Christensen et al., 2017; Dumont et al., 2018; Paine, 2003).

However, the use of electromagnetic (EM) data for hydrogeological interpretation also poses some challenges. First, the method will provide only limited information regarding the water storage capacity or the hydraulic conductivity. Second, it is also sensitive to the geological composition of the underground. The amount of conductive clay will strongly affect the measured field and, therefore, the resulting inversion. Third, due to the nature of the inverse problem and its high nonlinearity in the case of EM, the inverted resistivity models are nonunique (Tarantola, 2005). As underlined by Adams et al. (2022), therefore a good understanding of the geological properties of the aquifer and their spatial variations is necessary in order to interpret geophysical data robustly. So far, the most popular workflow has been to independently invert EM data and then relate them to boreholes or other types of data (Christensen et al., 2017; Dumont et al., 2018; Jørgensen et al., 2013, 2015; S. Kang et al., 2021; Knight et al., 2018; Neven, Christiansen, & Renard, 2022; Ringrose & Bentley, 2015). Often, the inversion and some steps in the workflow are considered deterministic. Therefore, the final structural model, the so-called cognitive model, is the one that, according to the modeler, fits most of the comprehensive knowledge available on site (Henriksen et al., 2003; Høyer et al., 2015; Kollet & Maxwell, 2006; Lemieux et al., 2008). The calibration of the groundwater model parameter is then carried out while keeping the cognitive model fixed. By doing so, there is no way to account for additional information about the geometry coming from the hydrogeological data, while the transmissivity and therefore the aquifer thickness clearly influence the head gradient or solute transport. If the geometry was fixed before inverting for hydrogeological parameters, and not correctly estimated, the hydrogeological parameters will likely have to be incorrectly identified during inversion in order to compensate for these initial errors. Furthermore, the use of a single cognitive model derived from geophysical data completely neglects structural uncertainty and, consequently, overall uncertainty in the aquifer characterization procedure.

To ensure consistency and the propagation of errors throughout the workflow, we propose to reverse the methodology described in the previous paragraph. Instead of going from the geophysical data to the structural model, we start by generating a set of prior plausible geological models and then updating these 3D models iteratively to fit the observed data. We adjusted the models jointly on the EM data acquired at the surface and the hydraulic heads measured in the aquifer. Since prior data are generated from the boreholes, we ensure that the final ensemble of models (posterior) agrees with them, even after parameter adjustment.

In this study, we generate geological models using the ArchPy hierarchical modeling approach (Schorpp et al., 2022). This method offers a formal description of a geological concept, the stratigraphic pile, that encompasses the succession of units, lithologies, and properties that must be modeled within a given domain. It also integrate knowledge from wells and conceptual understanding of the geological setup to create geostatistical realizations of potential aquifer geometries and properties. ArchPy utilizes a hierarchical approach that first simulates the main stratigraphic units, then the litho-facies, and finally the petrophysical parameters. The result is a complex and accurate representation of the aquifer structure. However, ArchPy cannot integrate other knowledge, such as geophysical or hydrological. The prior set of models needs then to be updated, in order to account for all the other types of data available on-site.

When updating models, the Markov chain Monte Carlo (MCMC) has been shown to be capable of producing consistent models and properly quantifying the final uncertainty (Irving & Singha, 2010; Jardani et al., 2013; Mosegaard & Tarantola, 1995; Reuschen et al., 2021). One downside of this method is the large number of forward calls required to converge and the associated computational cost (Linde & Doetsch, 2016). Ensemble Smoother algorithms have proven to be a reasonable alternative to MCMC for high-dimensional problems. It managed to identify solutions for complex nonlinear inverse problems, with fewer computational resources and time than MCMC (Emerick & Reynolds, 2013; Juda et al., 2022). Ensemble Smoother with Multiple Data Assimilation (ESMDA) (Emerick & Reynolds, 2013) is a variant of the Ensemble Smoother (ES) algorithm proposed by van Leeuwen and Evensen (1996). It approximates the relationship between the parameters and the data using their covariance computed using a finite prior ensemble of models. For large or extremely non-linear problems, the data can be assimilated multiple times iteratively (Chen & Oliver, 2013). ESMDA was used successfully in various groundwater studies (X. Kang et al., 2019; Lam et al., 2020; Li et al., 2015; Xu et al., 2021). Similar algorithms (ENKF, e.g.,) were also successfully applied to jointly integrate different types of data, such as geophysical and hydrological (Bouzaglou et al., 2018). However, this application was adapted only on a relatively small-scale lab experiment.

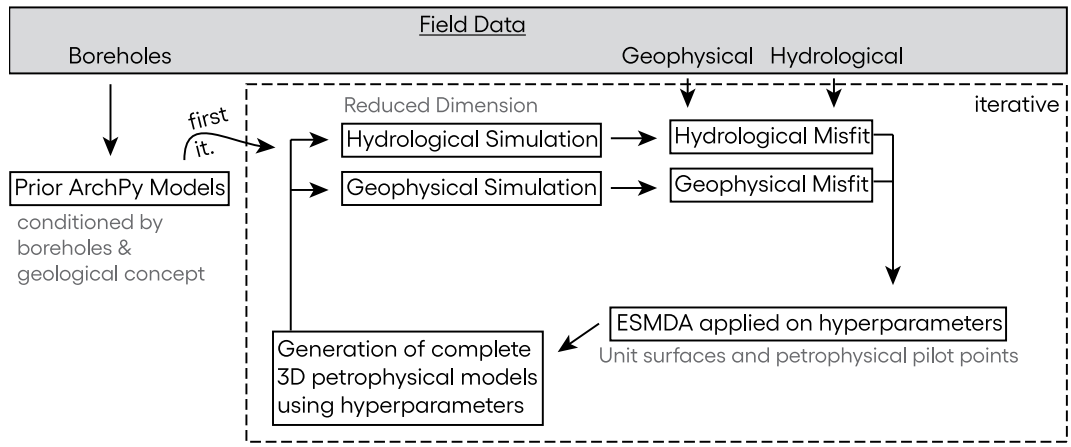


**Figure 1.** General sketch of the proposed data integration workflow.

In a previous study, we used a synthetic 2D data set and showed that joint hierarchical inversion combined with ESMDA can improve the identification of model parameters and the reliability of a prediction and its uncertainty, while integrating complex geological prior knowledge (Neven, Schorpp, & Renard, 2022). In the present study, we extend the method and show that this novel approach can be applied to real field data to integrate geophysical and hydrogeological data in 3D geologically consistent models. A new and important aspect of this paper is the development of a simple and consistent approach to identify the 3D parameter fields while computing the forward simulations in lower-dimensional spaces that can be different for each physical problem. To demonstrate its applicability, the new methodology is illustrated on a part (~15 km$^2$) of the upper Aare Valley, Switzerland. The area is densely sampled, with a towed Transient Electromagnetic survey (Neven et al., 2021), several hundred boreholes, and 25 piezometers.

## 2. Methodology

The proposed methodology can be divided into three steps (Figure 1): (a) the generation of prior geological models, (b) the forward simulations in a reduced dimensionality, and (c) the iterative update of the ensemble of models using ESMDA.

### 2.1. Generation of Prior ArchPy Models

We generate the prior ensemble of models using the open source stochastic hierarchical code, ArchPy, (http://www.github.com/randlab/ArchPy) (Schorpp et al., 2022). This Python-based tool allows for a detailed and hierarchical depiction of the geological environment, starting from the simulation of the contacts of large syndepositional layers, using 2D geostatistical methods, to the distribution of categorical lithofacies within these layers, and ultimately the petrophysical properties associated to the lithologies. ArchPy relies on the concept of stratigraphic pile, which is a formal description of the known regional stratigraphy of a given area. A large ensemble of plausible geological models compatible with regional knowledge and borehole data can be easily generated. For an in-depth description of the ArchPy methodology and examples of its capabilities, the reader is referred to Schorpp et al. (2022).

In addition to the existing ArchPy approach, to address possible nonstationarity in the petrophysical parameter simulations, we added the possibility to employ petrophysical pilot points. The underlying assumption is that large areas of the same aquifer can be significantly more (or less) hydraulically permeable than others. The conductive clay content can also change and cause variations in terms of electrical resistivity within the same unit. Consequently, the local average used for the simulation cannot be assumed to be uniform in space in the whole domain. To account for this, pilot points are randomly placed in each layer using a stratified sampling strategy. This ensures that the points are drawn randomly through the entire domain. The values of the pilot point are then drawn in a distribution of plausible range. Then, a linear variogram is used to krigge between the pilot points within each unit. This smooth map is then used as the local mean for the 3D petrophysical

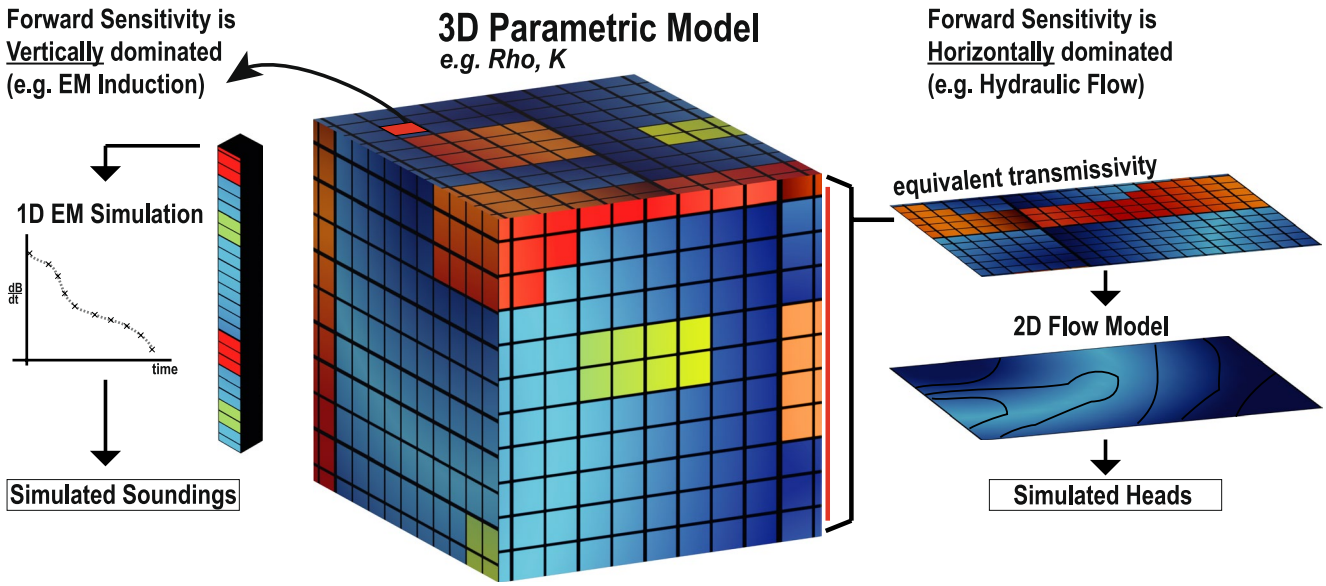


**Figure 2.** General sketch of the Forward in reduced dimensionality principle.

geostatistical simulation. The petrophysical parameters will vary following the variogram model provided around a spatially variable mean. A set of pilot points is defined for each layer and for each petrophysical parameter.

### 2.2. Forward Simulations in Reduced Dimensionality

As pointed out, for example by Linde and Doetsch (2016), the inversion of high-dimensional models leads to a large number of iterations and a large number of forward model calls for solving the physical, possibly nonlinear, spatial and/or time-dependent equations. Forward calculations are therefore often the most computationally expensive part of an inversion. Consequently, it is common to reduce the number of parameters (Linde & Doetsch, 2016), or to use surrogate modeling (Asher et al., 2015), for example.

In this study, we propose applying the inverse procedure by correcting the 3D geological and parameter models, but before running the forward flow and geophysical calculations, we reduce the dimensionality of the model to accelerate the forward simulations. Computing time is drastically reduced with a marginal loss of information. Figure 2 illustrates the principle of this approach. The idea is to exploit the fact that different physical processes are sensitive to different subsets of the 3D parameter fields. For example, the propagation of the EM signal requires us to solve the Maxwell equations to simulate the secondary magnetic field decay. For a horizontal coil, the main sensitivity is oriented downward. For most applications, solving the complete 3D equations will only slightly affect the final results but will result in a high increase in computational complexity (Engebretsen et al., 2022). For this reason, the EM geophysical forward is only computed using multiple 1D models extracted along the $Z$ axis. Each geophysical sounding is then computed in parallel using a fraction of the original model. The simulation time is greatly reduced, from several orders of magnitude. In this first example, the dimension is reduced by omitting some part of the model that is expected to have a limited effect on the predicted data.

However, flow within an aquifer is an example of horizontally dominated sensitivity. Instead of extracting some parts of the model, we use another approach here. Dimensional reduction is achieved by integrating spatial properties along the vertical direction. Because the overall thickness is small compared to the lateral extension and because we only have one hydraulic head measurement per piezometer, it is reasonable to solve the groundwater flow only in 2D. The complete 3D hydraulic conductivity model is then vertically integrated, and the horizontal transmissivity of the total aquifer is estimated at any location. The equation is the following:

$$T_{x,y} = \sum_{iz=1}^{nz} K_{x,y,iz} \cdot d_{iz} \quad (1)$$

where $T_{x,y}$ is the transmissivity in a cell $x$, $y$, $nz$ is the number of cells in the model along the vertical axis, $K$ is the hydraulic conductivity and $d$ is the thickness of the cell. The equivalent transmissivity is used to solve the 2D equivalent flow.

Of course, the simplified 1D and 2D forward simulations will not completely represent the actual 3D processes. The gain in terms of computing resources overcomes the marginal loss of information. For example, for the regional steady-state hydraulic heads simulation, the average gain in computing time between a full 3D and a 2D model is about 98% (51 s per model compared to 1 s per model), for a difference in response of a few centimeters (see Figure 6 and discussion in Section 3.3).

### 2.3. Data Assimilation With ESMDA

ArchPy allows us to rapidly generate a large number of plausible models. This ensemble represents our prior knowledge; it includes the regional geological concept expressed through the stratigraphic pile and all the borehole data. However, the compatibility of each model with indirect field data (geophysical or hydrogeological) is not guaranteed. To adjust the models based on all available data, we implemented the ESMDA algorithm introduced by Emerick and Reynolds (2013). In this section only a brief introduction to ESMDA is provided; readers are directed to the reference paper (Emerick & Reynolds, 2013) for more details on the method.

This stochastic data assimilation algorithm performs iterative corrections of a finite ensemble of $N$ models $\{m_i^{pr}, \ldots, m_N^{pr}\}$. At iteration 0, the models are conditioned solely by the boreholes. In the next iterations, the members of the ensemble are the corrected models of the previous iteration. At iteration $k$, the correction requires computing the Kalman gain matrix $K$ from the experimental covariance matrix between the data and the model parameters $C_{MD}^K$ combined with the data auto-covariance matrix $C_{DD}^k$ and the expected uncertainty of the data $C_{err}$. $C_{err}$ can be estimated using multiple stacking of the data in the field.

$$K = \left(C_{MD}^k \left(C_{DD}^k + \alpha_{k+1} C_{err}\right)^{-1}\right) \tag{2}$$

The parameter $\alpha$ is the noise inflation ratio. This parameter is added to address one common issue with Ensemble Smoothers and Ensemble Methods in general: inbreeding (Houtekamer & Mitchell, 1998). This term is used to describe the underestimation of uncertainty due to the fact that the ensemble used for the calculation of the Kalman Gain $K$ is the same as the one used to estimate the error. The individual models within the set are becoming more and more coupled after each iteration. Emerick and Reynolds (2013) showed that this inbreeding tends to be reduced with an increased number of members, and proposed the use of an inflation factor $\alpha$. This factor must respect the following relationship:

$$\sum_{k=1}^{N_{iter}} \frac{1}{\alpha_k} = 1 \tag{3}$$

with $N_{iter}$ being the number of iterations. Changing $\alpha$ through iterations does not significantly increase the quality of assimilation (Emerick, 2016). Consequently, we kept it unchanged at $\alpha = N_{iter}$ for all iterations. The correction for each member at iteration $k$ is determined as follows:

$$m_i^{k+1} = m_i^k + K \odot LM \cdot \left(d_{obs,i}^k - g\left(m_i^k\right)\right) \tag{4}$$

$$d_{obs,i}^k = d_{obs} + \sqrt{\alpha_{k+1}} C_{err}^{1/2} z_{d,i} \text{with } z_{d,i} \sim N(0,1) \tag{5}$$

where $d_{obs,i}^k - g\left(m_i^k\right)$ is the Euclidean distance between the observed measurements and the predictions computed by the forward operator $g$ using the current model parameters. To account for the imprecision in the measurement of the observed data, $d_{obs}$ are perturbed with random Gaussian noise drawn from $N\left(0, \sqrt{\alpha_{k+1}} C_{err}\right)$. The noise realizations are resampled for each iteration and ensemble member. As shown in Evensen (2018), each iteration of ES-MDA implicitly minimizes a different cost function. It combines on the one hand the difference between the parameters before and after the given iteration weighted by the inverse of the prior parameters covariance; and on the other hand, the misfit between the observations and the predictions. To reduce the risk of spurious correlations, a localization matrix LM is introduced. The localization matrix multiplies the Kalman gain elementwise. Spurious correlations occur mainly because

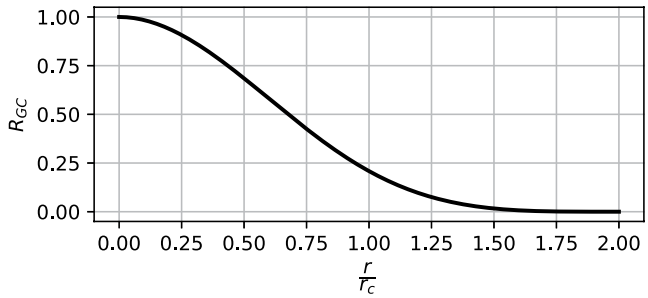


**Figure 3.** Gaspari Cohn correlation function (Gaspari & Cohn, 1999) value versus the ratio between the distance $r$ and the critical distance $r_c$.

ESMDA approximates the covariance matrix based on a finite number of models (Anderson, 2001; Evensen, 2009; Wen & Chen, 2005). Therefore, zero entries are difficult to reproduce in the covariance matrix, and spurious correlation can occur. One way to address this is either to increase indefinitely the number of members or to infuse prior knowledge in order to exclude correlations that are physically impossible. For example, we can expect no correlation between an EM sounding (data) and the resistivity value several kilometers away (parameter). The localization matrix varies between 0 and 1, and is inversely proportional to the spatial distance between a parameter and a data point. The Kalman Gain $K$ will become negligible for a distant set of parameter-data. Several functions exist to scale the localization matrix with distance. In this study, we implemented the widely used correlation function introduced by Gaspari and Cohn (1999). They propose the following piecewise rational covariance function $R_{GC}$ for data assimilation:

$$R_{GC}(r) = \begin{cases} 1 - \frac{5}{3}\left(\frac{r}{r_c}\right)^2 + \frac{5}{8}\left(\frac{r}{r_c}\right)^3 + \frac{1}{2}\left(\frac{r}{r_c}\right)^4 - \frac{1}{4}\left(\frac{r}{r_c}\right)^5 & 0 \le \frac{r}{r_c} < 1 \\ 4 - \frac{2}{3}\left(\frac{r}{r_c}\right)^{-1} - 5\left(\frac{r}{r_c}\right) - \frac{5}{3}\left(\frac{r}{r_c}\right)^2 + \frac{5}{8}\left(\frac{r}{r_c}\right)^3 + \frac{1}{2}\left(\frac{r}{r_c}\right)^4 + \frac{1}{12}\left(\frac{r}{r_c}\right)^5 & 1 \le \frac{r}{r_c} < 2 \\ 0 & \frac{r}{r_c} \ge 2 \end{cases} \quad (6)$$

where $r$ is the distance matrix of the Euclidean distance between the data observation points and the parameters, and $r_c$ is the critical distance beyond which the effect of a parameter on an observation is assumed to be negligible. Figure 3 represents $R_{GC}$ in 1D as a function of the ratio between $r$ and $r_c$. The critical distance of the localization matrix should be chosen according to the expected correlation length. An analysis of the evolution of the correlation with distance for both observation types is available in Supporting Information S1 (Figure S1). In our case, we set the critical distance of geophysical data to 150 m, since a stabilization of correlation is visible around this range. No significant evolution in correlation with the parameters was observed for the hydraulic head measurements as a function of the distance. Therefore, the LM was not implemented for these observations and is set to 1 for all hydraulic heads measurements. The summary of the LM parameters is available in Supporting Information S1 (Table S2).

## 3. Application to the Aare Valley

To illustrate the applicability of the proposed methodology on a real site, we used a test area in central Switzerland. The upper Aare valley is filled with Quaternary deposits, with a thickness ranging from a few meters on the sides of the valley to a few hundred in the center (Bandou et al., 2022). The investigation zone is located north of the upper Aare valley, just upstream of the city of Bern. It is a 12 km$^2$ area where a large amount of drinkable water is exploited. Several glacial cycles occurred, causing a complex interwinding of glacial and interglacial deposits (Bandou et al., 2022; Graf & Burkhalter, 2016; Schorpp et al., 2022). At least two different levels of aquifer have been identified in interglacial fluvial deposits (Schorpp et al., 2022). A shallow aquifer at the surface and a deeper one that has not been extensively studied. Despite this lack of knowledge, the upper Aare valley is widely used. In 60 km$^2$, we denote 4 quarries and more than 6,000 wells (shallow geothermic or water production). Despite its importance, the exact dynamics of the aquifer and its extent on the regional scale are not well known. Due to its importance, this site has been extensively studied with different data types (see Figure 5). It illustrates well how the abundance of data does not necessarily lead to an improved underground model due to the lack of an easy and applicable integration algorithm in hydrogeology. In the area, we denote the presence of 130 lithologically described boreholes, 35 hydraulic head measurements, and about 6,500 tTEM soundings. While the hydraulic heads measurements are more or less homogeneously distributed in the domain, there is a clear spatial division for the other two data sets. The boreholes are drilled most of the time within the cities or around existing buildings (within a few hundreds of meters) in areas where the electromagnetic coupling prohibits the use of inductive methods such as EM. The two data sets then cover different areas and are, by nature, complementary.

**Table 1**
*Geostatistical Parameters Used by ArchPy to Generate the Geological Models*

| | | | | |
|---|---|---|---|---|
| Stratigraphic surface model | | | | |
| Unit | 2 | 3 | 4 | |
| Variogram type | Linear | Linear | Linear | |
| Range $x$ and $y$ [m] | 1,000 | 1,000 | 1,000 | |
| Sill [$m^2$] | 70 | 150 | 150 | |
| Litho-facies model | | | | |
| Unit | 1 | 2 | 3 | 4 |
| Type | Gravel-sand | Clay | Gravel-sand | Clay |
| Uniform facies | | | | |
| Resistivity model | | | | |
| Facies | Mean ($\log_{10}$[Ωm]) | Range $x$ (m) | Range $z$ (m) | Sill ($\log_{10}$[$\Omega^2m^2$]) |
| Gravel-sand | 2–2.7 | 200 | 20 | 0.05 |
| Clay | 1.3–1.8 | 20 | 4 | 0.01 |
| Hydraulic conductivity model | | | | |
| Facies | Mean ($\log_{10}$[m/s]) | Range $x$ (m) | Range $z$ (m) | Sill ($\log_{10}$[$m^2/s^2$]) |
| Gravel-sand | −5 to −1.5 | 200 | 20 | 0.01 |
| Clay | −9 to −6 | 200 | 20 | 0.05 |

*Note.* The variogram type used for all the litho-facies, resistivity, and hydraulic conductivity models is exponential.

### 3.1. ArchPy Model

To define the conceptual model of the area, we rely on expert knowledge, existing boreholes, and previous publications (Graf & Burkhalter, 2016; Schorpp et al., 2022; Volken et al., 2016). Our zone of interest is the first 100 m of the valley filling. The conceptual model of the area includes a very shallow aquifer followed by impermeable clay. Below the clay, a second aquifer can be present. Finally, a deeper clay layer is present everywhere. When the second aquifer is absent, the two clay layers become one, even if they are still considered two distinct units. The top surface of the first layer is the topography. Therefore, there are only three surfaces delimiting the four units. For each surface, a variogram model was derived from the analysis of the borehole data and expert knowledge. All details of the parameters are shown in Table 1. The surfaces are simulated as 2D Gaussian Random Fields (GRF) with inequalities using the Sequential Gaussian Simulations (SGS) method (Chilès & Delfiner, 2009; Deutsch & Journel, 1992; Freulon & de Fouquet, 1993). In each layer, we also define the variogram for the simulations of petrophysical parameters. Once the units are generated, the lithologies are assumed to be uniform within each unit: the aquifer units are assumed to be made of gravel to sandy-gravel, and the other units are assumed to be made of clay. We expect spatial variations in terms of petrophysical properties within each lithology as a result of sedimentological changes, while we do not expect major spatial variations of lithologies within each unit. The two petrophysical parameters of interest are electrical resistivity ($\rho$) and hydraulic conductivity ($K$). They are modeled within the units using 3D GRF (e.g., Chilès & Delfiner, 2009) simulated with the circulant embedding spectral method (Lang & Potthoff, 2011). This method is faster for large 3D models compared to SGS when inequalities are not needed. In addition, as mentioned before, pilot points are implemented to represent a possible non-stationarity within the domain. They are used to parameterize the local spatial mean for the 3D GRF simulation during the data assimilation. Figure 4 shows a cross-section through one realization of an ArchPy petrophysical model. The three simulated surfaces and the topographic surface define the four units. Two are considered to be mainly clay and 2 are considered aquifer. Thanks to the ArchPy hierarchical approach, we can simulate different parametric fields for each of the units, according to the prior range of values.

The borehole logs were digitized by the Swiss Geological Survey during the GeoQuat project (Volken et al., 2016). As discussed above, we distinguish only aquifer formations from aquicludes. Note that most of the boreholes were drilled for the exploitation of the shallow aquifer. Therefore, most of them stop before or when they reach the bottom of the upper aquifer. The vertical depths are then usually between 3 and 12 m. Only 3 wells reach the lower aquifer

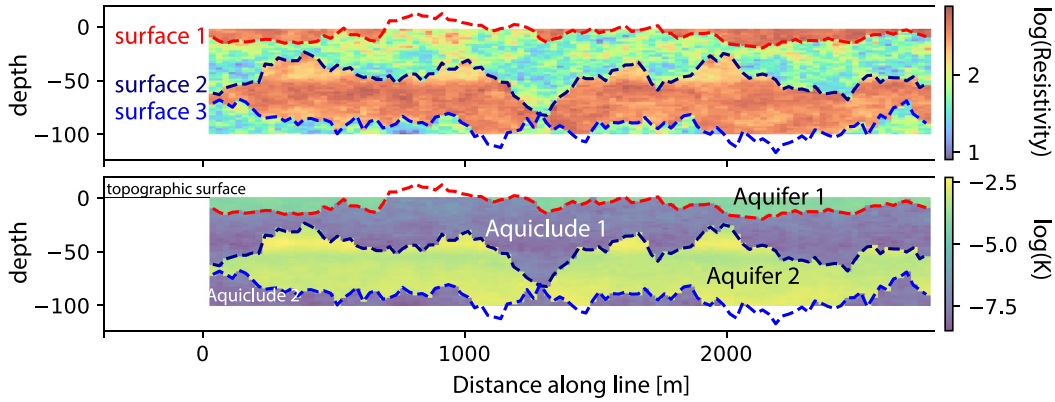


**Figure 4.** Cross section through one realization of a borehole-only ArchPy petrophysical model of electrical resistivity and hydraulic conductivity. In addition, the surfaces bounding the units are outlined. Note that there is a vertical exaggeration.

at more than 40 m depth. All of these wells were incorporated into ArchPy and used to constrain the geostatistical simulations. An important feature of ArchPy is the integration of inequalities (Schorpp et al., 2022) during geostatistical simulations. A well that does not reach the top of a certain unit may still have an effect on the local uncertainty since it indicates that this surface must be absent at that location, or deeper than the bottom of the well. The complete ArchPy model used for the prior generation can be found in the Zenodo repository, as a Python object (Neven & Renard, 2023). A summary of the parameters subject to assimilation and their prior distribution is available in Supporting Information S1 (Table S1).

### 3.2. Geophysical Data

The tTEM data come from a large-scale acquisition carried out in 2020. Data are publicly available. All characteristics of the system, data processing, cleaning, and validation, were documented in Neven et al. (2021). Data are used in raw electromagnetic decay sounding form ($dB/dt$), and therefore are not affected by any inversion choice. The forward code used to simulate the models is a 1D Time-Domain EM implemented in Fortran in *AarhusInv* (Auken et al., 2015). The use of a 1D forward simulation code is common when dealing with

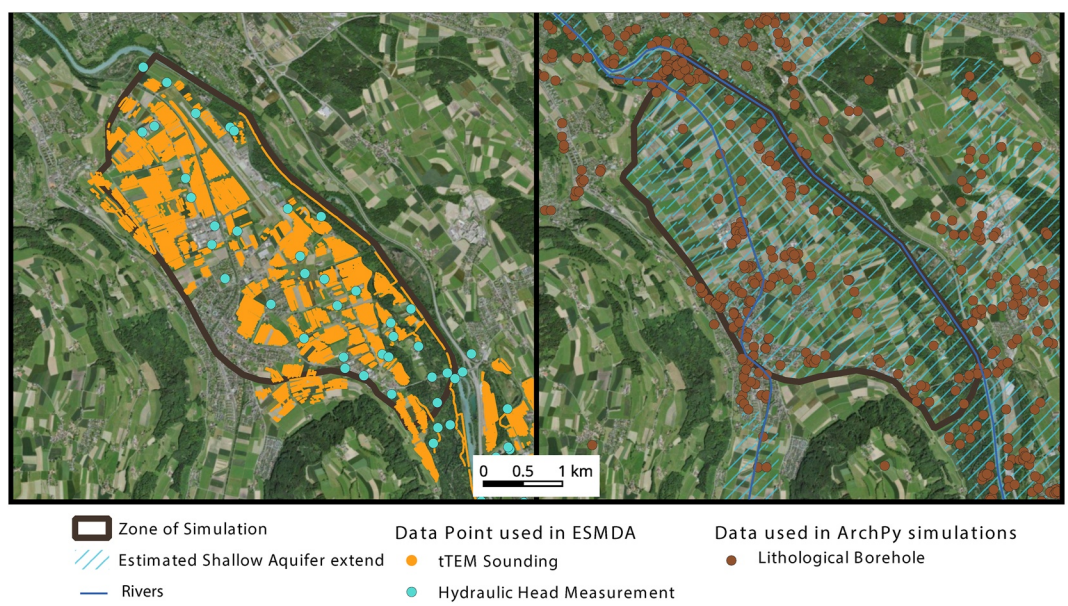


**Figure 5.** Extent of the simulation with the available data illustrated besides. The base map is published open access by the Swiss Topographic Office.

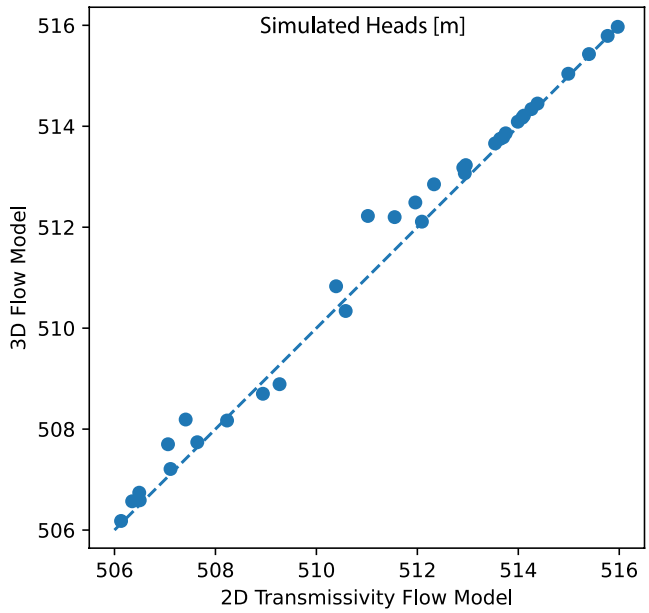


**Figure 6.** Comparison of the hydraulic heads in the 35 measuring stations computed using a complete 3D flow model and the 2D equivalent transmissivity model.

EM technology (e.g., Auken et al., 2015; Farquharson & Oldenburg, 1993; Scholl et al., 2009). The relying assumption is that locally the subsurface can be considered in his vertical dimension only. The full system geometry is used for the calculation of the synthetic responses. These responses can then be compared with the raw electromagnetic decay measured on the field. As mentioned in the data descriptor, the uncertainty associated with the observations in the data is derived from the multiple stacking of soundings during the acquisition, directly in the instrument. As mentioned in Auken et al. (2019), the tTEM instrument calculate the uncertainty by stacking few hundreds measurement. In the area of Belp, 6,400 soundings of 29 gates each are integrated (see Figure 5).

### 3.3. Hydrogeological Data

Hydraulic head measurements are from a large-scale piezometric campaign (Kellerhals et al., 1981). At the same time, the heads were measured at hundreds of points in the valley, 35 of which are in our area of interest. The levels of the two rivers were also measured. Recent but less dense measurements acquired in 2020–2022 showed that those previous measurements are still representative of current aquifer conditions. Due to the lack of sufficient time series to compare with, we decided to use only a steady-state model in this paper. The forward groundwater model is two-dimensional and is based on ModFlow6 (Hughes et al., 2017). Figure 6 shows the comparison between the heads simulated using a 2D transmissivity model and a complete 3D model for an ArchPy realization. We denote only a marginal loss of information with an average difference in heads of less than 12 cm. We can denote that the heads further away from the boundary conditions, which also present an intermediate head value on Figure 6, are also those that present more variations between the 3D and the equivalent transmissivity simulation.

Recharge was estimated using precipitation measured at the site and was set spatially uniform to 652 mm of rain per year. River boundary conditions were imposed along the two stream paths (see Figure 5) using the altitude profile of the streams and the hydraulic heads measured during a piezometric campaign in the river (Kellerhals et al., 1981). A head is imposed with a stream conductance damping (Hughes et al., 2017). In addition, two constant head boundary conditions were assigned: one along the southeast boundary of our model and one along the north boundary. Their value was set to, respectively, 506 and 516 m, following the iso-head of multiple measurements. All known water pumping wells present in the area were added as flow boundary conditions, and their estimated flow is provided in Kellerhals et al. (1981). A conceptual representation of the groundwater flow model is available as (Figure S2 in Supporting Information S1). All Modflow 6 files are also available in the Zenodo repository Neven and Renard (2023). During data assimilation, only the hydraulic conductivity field and the geometry of the aquifers, and consequently the resulting transmissivity, are corrected. The other parameters are kept constant. They could be included in the data assimilation procedure in the future to improve the performance of the model.

### 3.4. Validation and Indicators

In applying data assimilation methodology to real data, unlike a synthetic test, the true geology, parameter fields, and uncertainty are unknown. To quantitatively measure the validity of our results, we implemented serveral indicators.

The first indicator used to benchmark the quality of the models is the data misfit. It benchmarks the agreement between the predicted data and the observations. This indicator is not used during the inversion since ES-MDA does not need an objective function. But, as mentioned before, ES-MDA will try to minimize a combination of terms that includes the misfit. It is used here to assess the posterior compatibility of the data. The residual at a given iteration is:

$$DataMisfit = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{(d_{obs,i} - d_{frw,i})^2}{\sigma_{d,i}^2}} \quad (7)$$

where $d_obs$ is the observed data, $d_frw$ is the simulated data, $\sigma_d$ is the uncertainty of the observed data and $N$ is the total number of data point. The indicator is dimensionless and allows us to analyze different types of data on a comparable basis.

To have a better representation of the performance of the parameters prediction, the next indicators rely on a cross-validation algorithm based on the borehole data. This involves selectively withholding a portion of the conditioning data set from the simulation to create a basis for comparison and evaluate the prediction quality against these intentionally unaccounted-for boreholes. However, the spatial correlation inherent in the data means that a random selection of boreholes could lead to an overestimation of algorithm performance if the omitted boreholes are close to each other from a different fold. To counteract this, we employ the K-Means clustering algorithm (Hartigan & Wong, 1979) to the spatial coordinates of the boreholes, segregating them into 20 groups. This ensures boreholes in close proximity fall into the same group, providing a more accurate representation of the algorithm's performance.

From these 20 groups, we randomly assemble five similar-sized folds. For each fold, we generate a unique ArchPy prior ensemble, each time excluding the data of one fold, accounting for approximately 20% of the total data. Following this, we apply the data aggregation algorithm independently to all five sets, allowing for a comprehensive evaluation of its performance. The results of each set are then compared against their corresponding excluded boreholes to assess the prediction quality. These models are created only for the purpose of quality assessment and function independently from the complete models, which incorporate all available borehole data. This separation guarantees an unbiased evaluation of performance while ensuring that the final models yield the most accurate representation possible, given the available data.

Since our main objective is to illustrate and test the application of multiple data types assimilation to hydrogeology, we will also benchmark the performance of our models to outline the boundary of the aquifer. Therefore, we define three quality indicators for each borehole: the mean error, the mean normalized error, the Continuous Rank Probability Score (CRPS). The three indicators are applied on the prediction of the thickness of the upper aquifer, which corresponds to the altitude of the surface that defines the limit between the Unit 1 and the Unit 2. This limit is also the one the most frequently defined in the boreholes. The mean error allows controlling that there is no constant positive or negative bias in the prediction and that the correct value is predicted on average. The mean error is defined as:

$$\epsilon_{x,y}^{k} = \frac{1}{N}\sum_{i=1}^{n}\left(z_{i,x,y}^{k} - z_{x,y}^{borehole}\right) \quad (8)$$

where $\epsilon_{x,y}^{k}$ is the mean error in iteration $k$ at position $x$ and $y$. $N$ is the number of members, $z_{i,x,y}^{k}$ is the prediction of the thickness at the position $x$, $y$ in the model $i$ and $z_{x,y}^{borehole}$ is the true thickness of the aquifer from the description of the well. Therefore, the error is defined at each borehole position and should converge to zero on average. The second indicator is the normalized error. It is defined as the mean error divided point-to-point by the experimental variance of the simulation results. It shows if the predicted magnitude of the uncertainty is comparable with the actual ensemble errors, assuming a Gaussian distribution of errors. Normalized error is then defined as:

$$E_{x,y}^{k} = \frac{\left(\epsilon_{x,y}^{k}\right)^2}{\left(\sigma_{x,y}^{k}\right)^2} \quad (9)$$

where $E_{x,y}^{k}$ is the normalized error in iteration $k$ at position $x$ and $y$, and $\sigma_{x,y}^{k}$ is the experimental standard deviation calculated over the simulation results of the $n$ members at position $x$ and $y$. The normalized error should converge on average to 1, meaning that the uncertainty is in the same order of magnitude as the ensemble error. Finally, we used the Continuous Rank Probability Score (CRPS) (Gneiting et al., 2007). The CRPS is sensitive to both bias and uncertainty. It is defined as being the integral between the Cumulative Distribution Function (CDF) of the ensemble of the predictions at a given point and the step-function CDF of the discrete true value at the same point. It is expressed as:

$$CRPS_{x,y}^{k}\left(F, z_{x,y}^{borehole}\right) = \int_{-\infty}^{+\infty}\left(F(y) - \mathbb{1}\left(y > z_{x,y}^{borehole}\right)\right)^2 dy \quad (10)$$

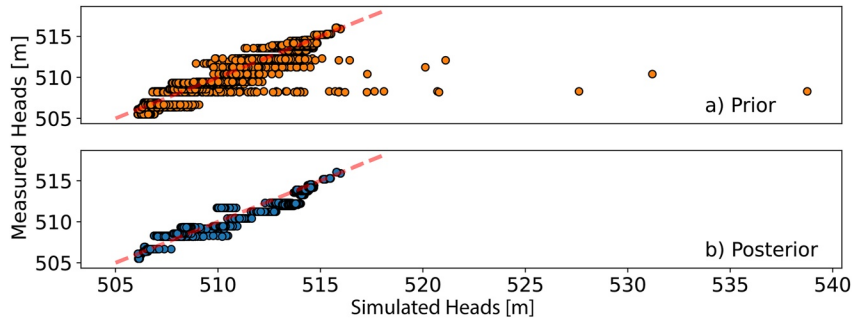


**Figure 7.** Simulated Heads vs. Measured Heads in the 35 groundwater monitoring stations for the (a) 500 Prior models and (b) 500 posterior models. The dashed line is the perfect prediction.

where $\mathbb{1}$ is the Heaviside function, so that $\mathbb{1}\left(y > z_{x,y}^{borehole}\right) = 1$ if $\left(y > z_{x,y}^{borehole}\right)$ and 0 otherwise, and $F$ is the Cumulative Density function of the ensemble of predictions at a given point. The minimum CRPS score is zero, and is reached if the prediction is correct with no uncertainty. If the prediction has a small uncertainty and is not biased, the score will be relatively small. On the contrary, as soon as the prediction is biased or the uncertainty large, the CRPS increase. Unlike normalized error, CRPS has the advantage of not assuming a normal distribution of errors. All these indicators are point indicators, and will be calculated on all excluded boreholes positions for all five folds.

## 4. Results

Five hundred initial members were generated using ArchPy. This set reflects the prior knowledge and will be referred to later as the prior. Figure 8 shows the evolution of the observation misfit thought the iterations. Ten ESMDA iterations were carried out. Iteration 0 corresponds to the prior observations. We denote a global decrease of the residual, with a mean total joint residual of the data (geophysical and hydrogeological) decreasing from 15.1 to 4.1. Of course, since the number of geophysical data points is much higher than the number of hydrological measurements, the total misfit closely follows the geophysical misfit. However, the hydrological misfit also shows a decrease in terms of variability and average value. We can see that the residual tends to greatly reduced in the first iterations, meaning that good results could probably be achieved in less than 10 iterations.

The computing time per iteration is about 4h20, for about 3 million active cells. On this total time, about 75% is dedicated to the forward calculation and the rest to the update of the model. The ensemble of models after data assimilation will be referred to as posterior. Figure 7 compares the heads measured in the valley with the heads predicted from the 500 models at the same position, both prior and posterior. A decrease in terms of variability of the heads can be denoted, and the final ensemble of prediction is closer to the perfect prediction line. It suggests that the posterior ensemble is capable of predicting heads in the domain fairly well.

Figure 9 shows a cross-section through the model with all the boreholes taken into account. Seventy-five members drawn out of the 500 of the prior and the posterior are compared in terms of the depth of the upper aquifer. The reduced uncertainty at the 4,600 m position is due to the presence of a conditioning borehole and is therefore consistently present in both prior and posterior sets. The locations of the geophysical EM data are outlined in yellow in the posterior cross-section. Only the points that are collocated are shown, but the effect of neighboring points can occur. We denote a clear reduction in uncertainty in the posterior compared to the prior, where geophysical data have been acquired. The confidence interval is significantly reduced. Where no geophysical data is present, the uncertainty logically tends to be equivalent to the one in the prior. When applying ESMDA, one common risk is the collapse of the posterior, where all members converge to a similar model, potentially leading to an underestimation of uncertainty. In the present case, despite 11 iterations, each model in the posterior ensemble exhibits notable differences, suggesting that the uncertainty estimation is reasonably accurate. However, it is important to acknowledge that validating the correctness of uncertainty estimates can be challenging.

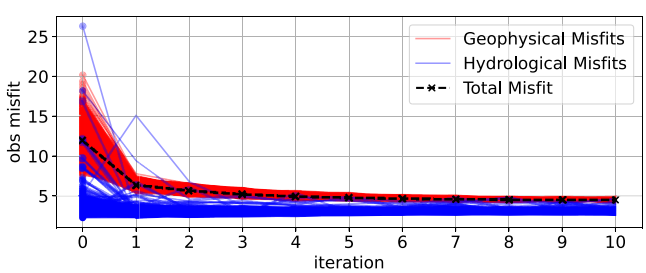


**Figure 8.** Evolution of misfit between the simulated observations and the measured observations. Iteration 0 corresponds to the prior models simulation.

One clear advantage of the presented methodology is that the final models are hierarchical. For example, for most applications in hydrology, the thickness

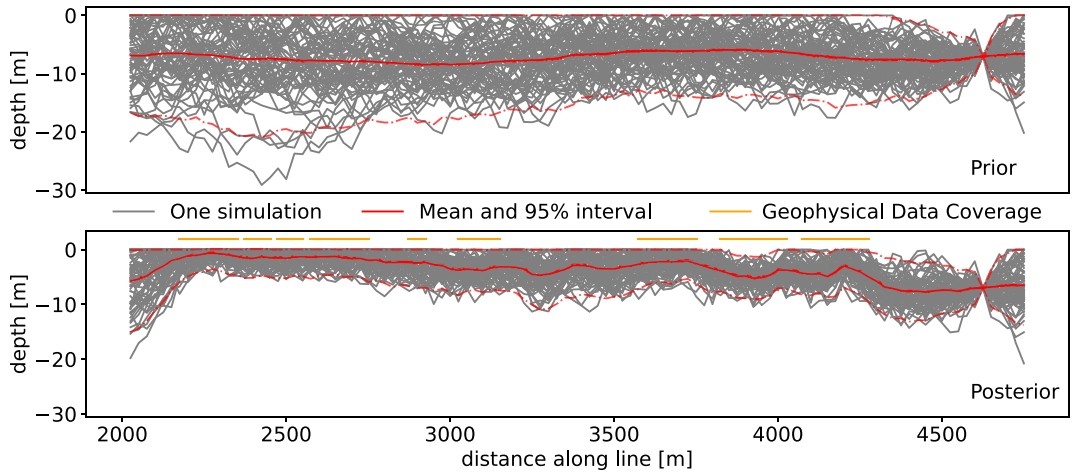


**Figure 9.** Example of a cross-section through the prior and posterior models showing the depth of the aquifer along a 2.5 km line. The ensemble of simulations is shown in gray, and the arithmetic mean is shown in red with the 95% confidence interval envelope (2.5% and 97.5% quantile of the ensemble). The decreasing uncertainty at 4,600 m is due to a conditioning borehole.

and extent of the aquifer are of key interest. Unlike the 3D electrical resistivity field, in this workflow, the thickness of the aquifer is part of the model and can be easily extracted. Figure 10 shows the mean depth of transition between the shallow aquifer and the aquiclude over the 500 members. The estimated uncertainty is also outlined. We can denote the zero uncertainty points in both the posterior and the prior, which correspond to the borehole positions where the depth is considered certain. The results appear to be spatially consistent. First, the variations in terms of aquifer depth show continuity over the range of a data point update in the Localization Matrix, and therefore beyond the effect of a geophysical point. The critical distance was set to 150 m, which means that above twice this distance, for example, a geophysical measurement cannot have an effect during the update. Large areas with lower or higher depth are due to information that is actually brought by the data. It suggests that distant geophysical data are consistent and are successfully incorporated during the ESMDA inversion. Second, the continuous lower or higher depths are elongated in a SE-NW orientation, which is consistent with the deposition phases estimated at this site, despite the fact that the variograms used for the generation of the prior are isotropic. Third, we note that the uncertainty is reduced where and only where geophysical data are present and kept unchanged far from them. Finally, the small artifacts that we can see mainly in the Prior are due to neighboring wells that carry a significant difference in terms of aquifer depth, probably unrealistic due to their close position from each other, and cause an abrupt change.

Cross-validation was performed using 5 distinct folds, 500 members, and 10 iterations each. The variable cross-validated here is the thickness of the first aquifer. The aggregated results are shown in Figure 11. Each fold is amputated of 20% of the borehole data set. The purpose is not only to benchmark the results of the data integration algorithm, but also to compare them with the borehole-only approach. For this reason, cross-validation was also performed using only prior knowledge before data integration. Several results can be highlighted. First, compared to the prior models, the error range is smaller and is centered on zero (see Figure 11a). This unbiased error is a critical point and suggests that, unlike the prior, we tend to equally over- and underestimate the thickness of the aquifer. This proves that this methodology can perform well even if the prior is uncertain or not totally correct. Figure 11b shows that a global reduction in uncertainty is clearly visible almost everywhere in the domain, especially in places that are surrounded or covered by tTEM points. No negative change in uncertainty was denoted between the prior and the posterior. Despite this global decrease in uncertainty, the normalized error is improved (Figure 11c). The combination of these two indicators shows that the uncertainty in the posterior is smaller and is better scaled to the ensemble error. Furthermore, some remaining outliers points are in fact far from any measurements and consequently will only be slightly changed between prior and posterior. The distribution is asymmetric; with most of the values being lower than one; it suggests that in some locations the data assimilation algorithm still tends to slightly overestimate the uncertainty.

Unlike the normalized error, the CRPS shown in Figure 11d has a target value that is also the arithmetic minimum: 0. CRPS is used to compare a probabilistic prediction with a deterministic truth by integrating their cumulative

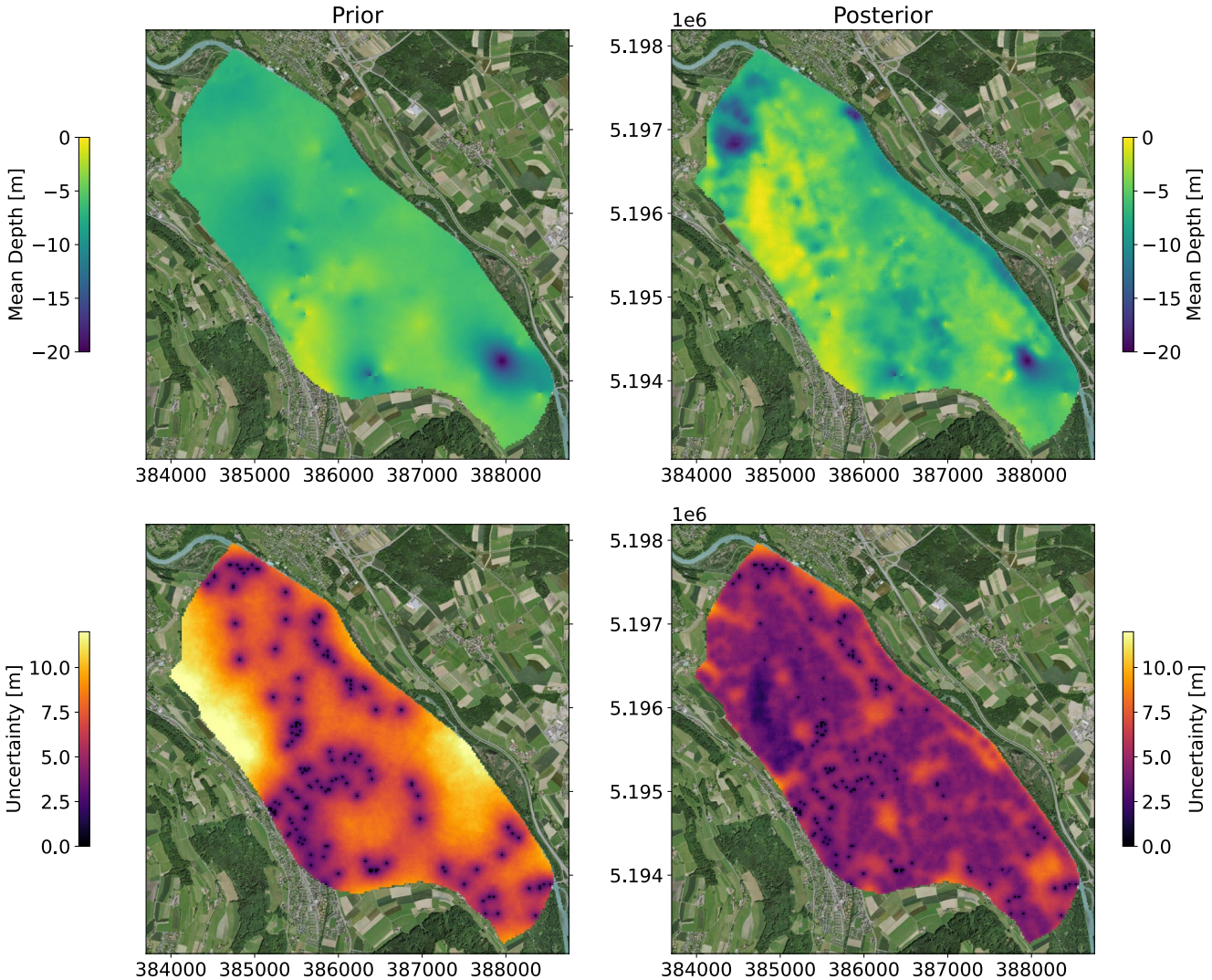


**Figure 10.** Mean depth of transition between the shallow aquifer and the aquiclude over the 500 posterior simulations (after data integration) and prior simulations (boreholes only), and associated uncertainty. The uncertainty at the location of the boreholes. The coordinates are in UTM32 (epsg: 32632). Basemap freely obtained from the Swiss Federal Office of Topography.

density function. The CRPS has the advantage of not assuming a Gaussian distribution of the error and considers the probabilistic forecast as a whole. In our case, we denote that the CRPS is distinctly better in the posterior than in the prior, both in terms of global distribution or in terms of arithmetic mean. We can then assume that the posterior models give a better forecast in terms of aquifer thickness than a borehole-only-based model, without any manual integration of geophysical data.

In the above results, the thickness of the upper aquifer was used since it is the information that is the most present in the boreholes and can therefore be used for cross-validation and comparison. However, our data integration in the discrete geological model approach can also provide various other information. First, the distribution of the petrophysical parameters can be retrieved. Figure 13 shows an E-W cross section through one posterior realization (Figures 13a and 13c) and the mean petrophysical fields (Figures 13b and 13d). Since the petrophysical models are generated from the discrete unit model, where each cell is assigned a specific layer, we denote sharp transitions between the different facies. On the other hand, the mean realization is much smoother, since it is averaged over the members. For the resistivity mean cross section (Figure 13d), we can see that averaged smooth petrophysical properties are closer to what can be achieved with classic deterministic inversion, compared to a single model (Figure 13c). Figure 13e compares the smooth inversion of the same tTEM data published by Neven

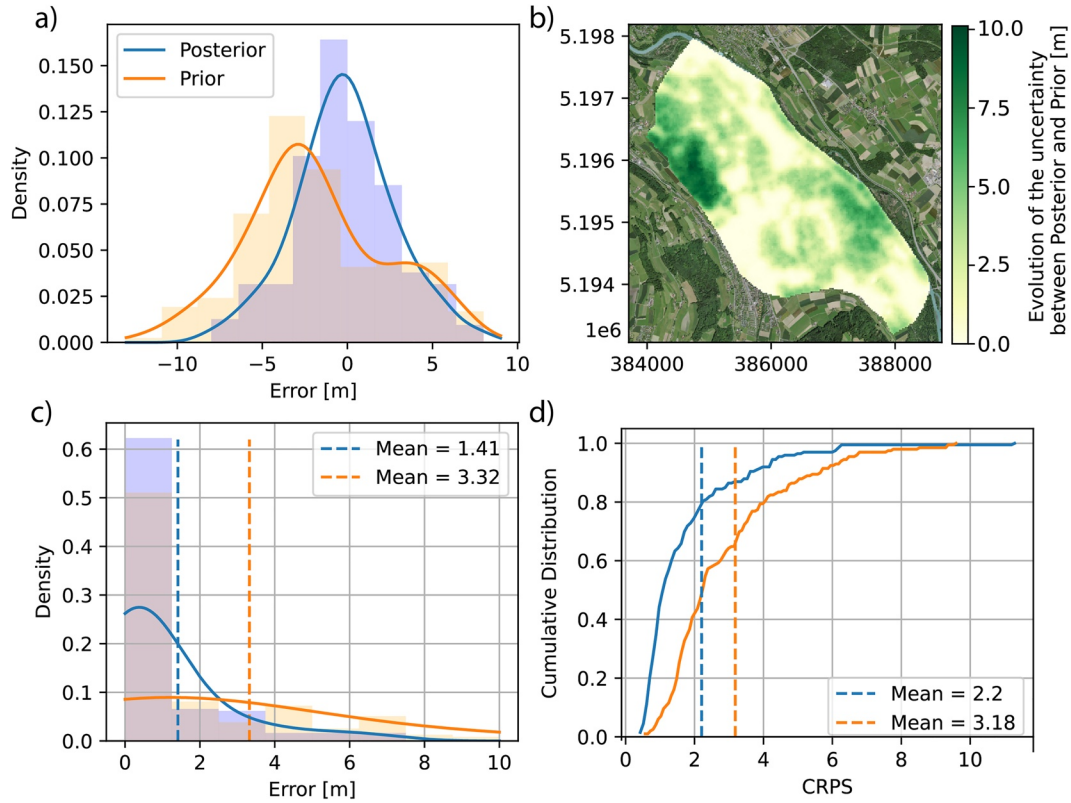


**Figure 11.** Results of the Cross-Validation: (a) Error distribution for prior and posterior models (b) Evolution of the uncertainty ($2\sigma$) between the posterior and the prior. A Positive value means a reduced uncertainty in the posterior, where a null value means that so significant changes has occurred. (c) Distribution of the Normalized error (d) Cumulative distribution of the Continuous Ranked Probability Score (CRPS).

et al. (2021). To facilitate the comparison of the structures, the mean resistivity field over the realizations is added in grayscale. The agreement between the mean resistivity field and the 1D resistivity models is satisfactory.

Figure 12a shows the joint distribution of resistivity versus hydraulic conductivity in the upper aquifer after inversion. As a reminder, in the prior both proprieties were assumed to be normally distributed around uniformly distributed random means drawn at the pilot points position. Recall that the uniform range for the pilot points was for resistivity and hydraulic conductivity, respectively, $100 - 500$ $(10^2 - 10^{2.7})$ $\Omega \cdot$ m and $10^{-5} - 10^{-1.5}$ m·s$^{-1}$. The resulting prior petrophysical distributions for the upper aquifer are shown in orange in Figure 12a. Logically, the maximum of the prior distribution of parameters corresponds roughly to the average of the initial pilot point values. The predicted value range in the posterior is smaller compared to the large prior range. Interestingly, the results of the data integration show two distinct populations not present in the prior (see Figure 12a). We can denote that the two populations are clear in terms of hydraulic conductivity but overlap in terms of resistivity. The values for both resistivity and hydraulic conductivity are within the expected range in the area. Finally, Figure 12b shows the probability of encountering the second aquifer above a certain depth. Since only 4 boreholes reach this second aquifer, the probability in the prior is almost uniform at 50%. The data aggregation algorithm has provided a significant amount of information, reducing the probability in some areas of the valley and predicting with high confidence the presence of the aquifer in other places.

Another interesting result is the possibility to assess the vulnerability of the second aquifer. Shallow aquifers in quaternary settings are generally more affected by pesticides, fertilizers, or permanent contaminants (PFAS e.g.,) than the deeper ones. In Switzerland, these problems lead some water authorities to start targeting deeper aquifers to produce drinking water. Quantifying the potential connection between the shallow and deeper aquifers is therefore an important practical issue. The modeling framework proposed in this paper allows estimating this thickness in every simulation and post-processing them to obtain some statistical estimates. Figure 14, summarizes such results. The thinner the aquiclude, the higher the risk of communication between the upper and lower aquifer, and

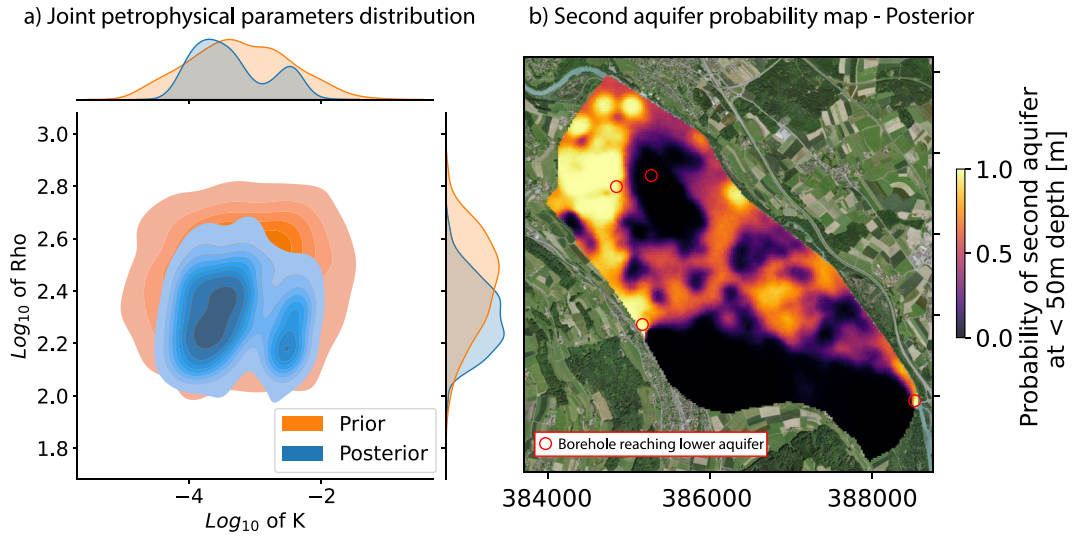


**Figure 12.** (a) Joint Posterior Distribution of the two petrophysical parameters of interest in the aquifer: Electrical Resistivity (Rho) and Hydraulic Conductivity (K). (b) Probability of the presence of the second aquifer at a depth less than 50 m with the position of the boreholes reaching the second aquifer.

consequently the vulnerability of the second aquifer to surface pollutants. Figure 14 shows that the uncertainty is relatively high in the set of prior simulations. It is low only around the seven borehole locations. Note that among these boreholes only three are touching the deeper aquifer and are used as conditioning points, while the others have not reached the second aquifer and are used as inequalities in the conditioning of the geostatistical simulations. The use of the EM data allows to constrain better the thickness map and reduce significantly the uncertainty (Posterior results in Figure 14). But we observe that the posterior uncertainty still remains much higher than the uncertainty of the thickness of the shallow aquifer. The uncertainty reduction is lower, this is certainly due to the higher depth and the presence of a thick clay layer because the EM method is only poorly sensitive to structures at a large depth.

Finally, we show the effect of the data assimilation on different particle tracking scenarios to illustrate the impact of the method on possible solute transport predictions. The prior and posterior models are compared. Using the simulated heads of the groundwater model, the advective path is calculated for each of the members. We used the same hydrogeological forward model as the one used for inversion. As shown earlier (Figure 7), the hydraulic heads are much less dispersed in the posterior simulations. Figure 15 shows four situations with a different initial location for the contamination source and shows their respective prior and posterior advective paths. In all the cases, the uncertainty has been reduced by integrating the geophysical and groundwater data. We observe that all the posterior paths are contained within the prior. This suggests that a model based only on boreholes could be used to estimate the travel path uncertainty, but the uncertainty would likely be much too broad. For example, in the case of scenarios A and B (Figures 15a and 15b), the river in which the pollutant may appear is not identified with confidence in the prior since it could reach either the Aare river on the east or the Grube in the middle of the domain when the posterior distribution has clearly identified the most probable case. Figures 15c and 15d illustrates how the thicker and more permeable aquifer areas identified through the assimilation of the data affect the path. The thicker aquifer northwest of the area (see Figure 10) drains most of the particles along an impermeable boundary of the valley.

## 5. Discussion

The methodology and results presented in this study have proven the feasibility of integrating geological, geophysical, and hydrogeological field data to obtain consistent models at the scale of a part of a regional aquifer. The use of two different data types through joint inversion has proven to be a valuable approach. The two distinct hydraulic conductivity populations (Figure 12a) probably correspond to two distinct sub-aquifer regional granulometry changes, which is typical of highly heterogeneous Quaternary environments. The two populations cannot be distinguished from resistivity only, since their respective ranges overlap, despite an order-of-magnitude

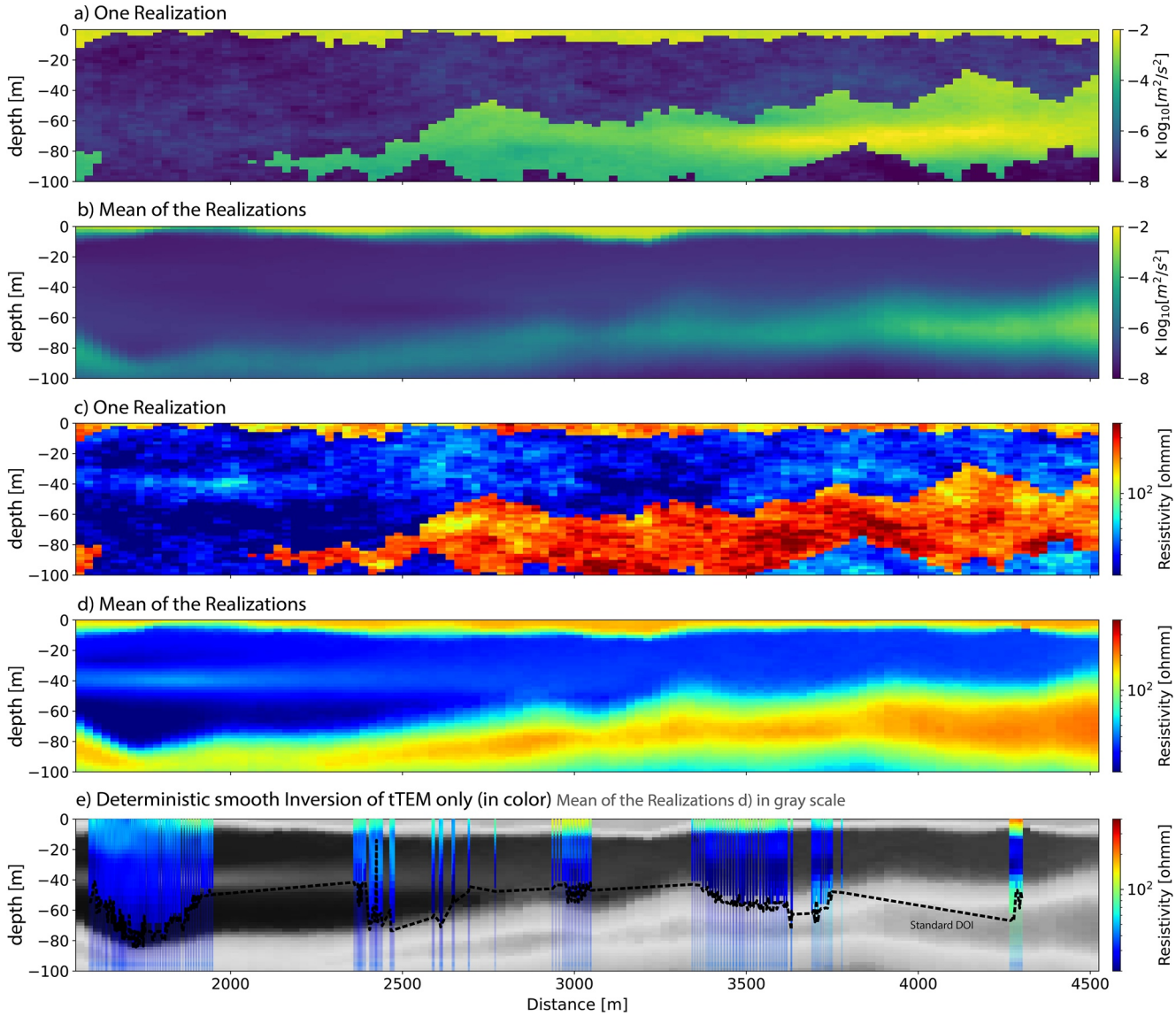


**Figure 13.** E-W Cross-Section ($y$ = 5,194,568 m UTM32N) comparing the hydraulic conductivity field of one posterior model (a), the mean hydraulic conductivity over the posterior members (b), the resistivity field of one posterior model (c), the mean resistivity over the posterior members (d), and the 1D resistivity models from a deterministic smooth inversion of the tTEM data (e) from Neven et al. (2021). The dashed black line is the standard depth of investigation, retrieved from the inversion. The gray scale base map of (e) is the mean resistivity over the posterior members (d).

difference in hydraulic conductivity. However, geophysical measurements constrain the thickness of the aquifer. Consequently, the space of uncertainty of the transmissivity is reduced, and the space of uncertainty for the hydraulic conductivity is limited. Without joint inversion, it is probable that these areas would have been considered to be a shallower aquifer section, instead of a less permeable one.

To accelerate the forward computations, we used a simplified 2D groundwater model. This is not a requirement for the proposed methodology, but it constitutes a reasonable simplification in the aquifer that we studied and with the available integrated head data. In a thicker aquifer with precise head measurements collected at different depths, this 2D approach would not be adequate to represent the system since the vertical flow components across the geological formations may have an important impact. In these situations, a complete 3D groundwater model would be required. Similarly, the use of 1D EM forward for inversion is widespread but significant differences can occur between 2D or 3D modeling in the case of complex geology such as seawater intrusion for example, (S. Kang et al., 2015). However, our method lies in between since even if a 1D forward code is used, the underlying model is still 3D and has some constrains.

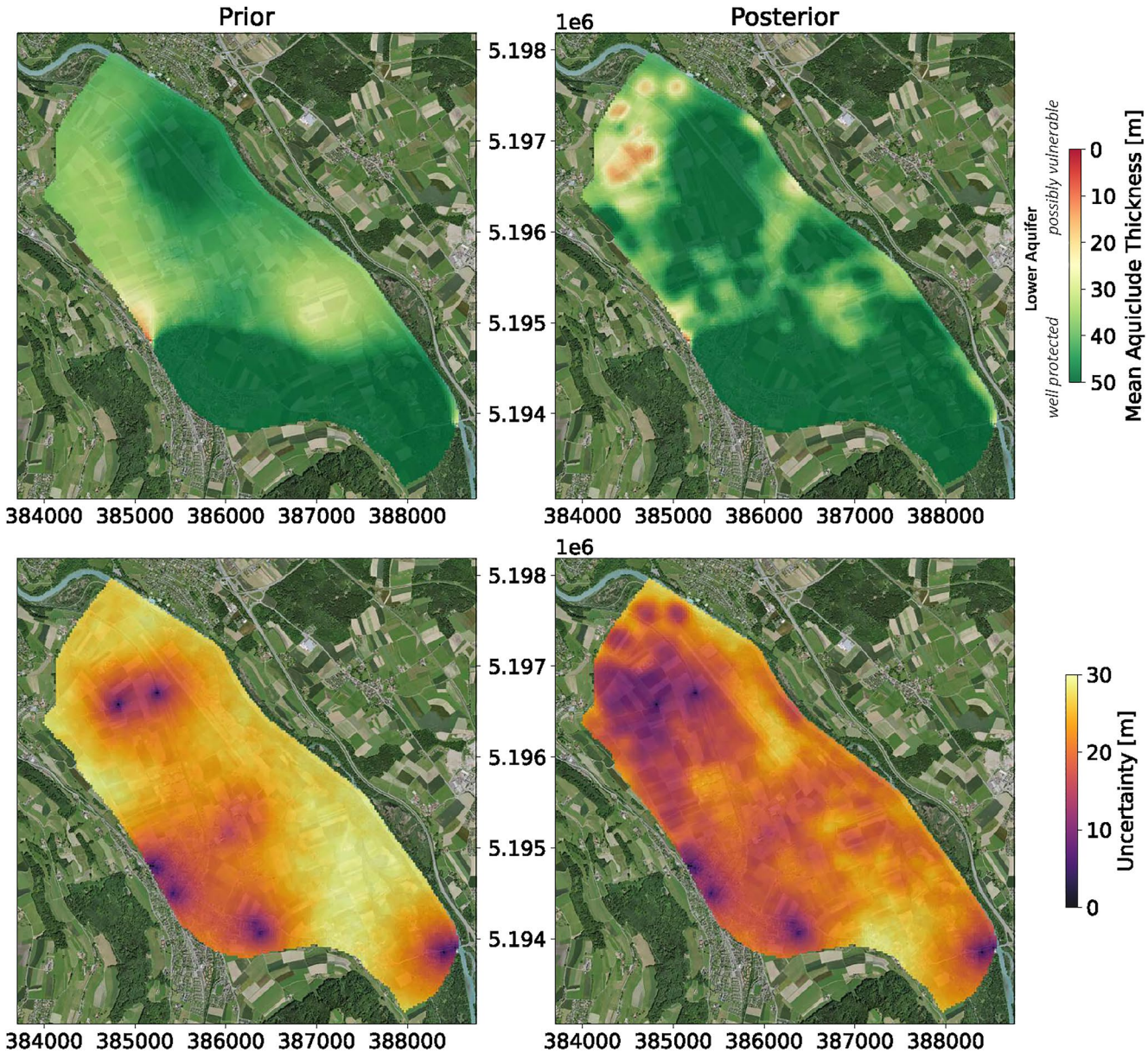


**Figure 14.** Thickness of the aquiclude, acting as the separation of the surface and shallow aquifers. Low thicknesses are correlated with a higher risk of pollutant contamination from the shallow aquifer to the lower protected aquifer. Zero-incertitude points are boreholes. The coordinates are in UTM32 (epsg: 32632). Basemap freely obtained from the Swiss Federal Office of Topography.

The proposed workflow has the advantage of being flexible and could easily be extended to account for more physical processes and parameters. The data assimilation method could include other types of geophysical or groundwater measurements and allow identifying additional model parameters such as storativity, porosity, or river conductance, as well as boundary conditions or source terms. Combining more information would require developing appropriate prior distributions for all these additional unknowns but could help to better quantify the overall uncertainty in groundwater models. The downside could be that adding more parameters, especially global ones, could lead to more spurious correlations in the ESMDA procedure and could also increase significantly the computing time required for the assimilation algorithm.

Compared to a classical geophysical inversion, the data residual of the proposed approach is much higher (4.1 vs. less than 1 in the previous inversion by Neven et al. (2021)). Indeed, the classical deterministic inversion is only constrained by the regularization applied to the inversion algorithm. These fairly loose constraints allow the

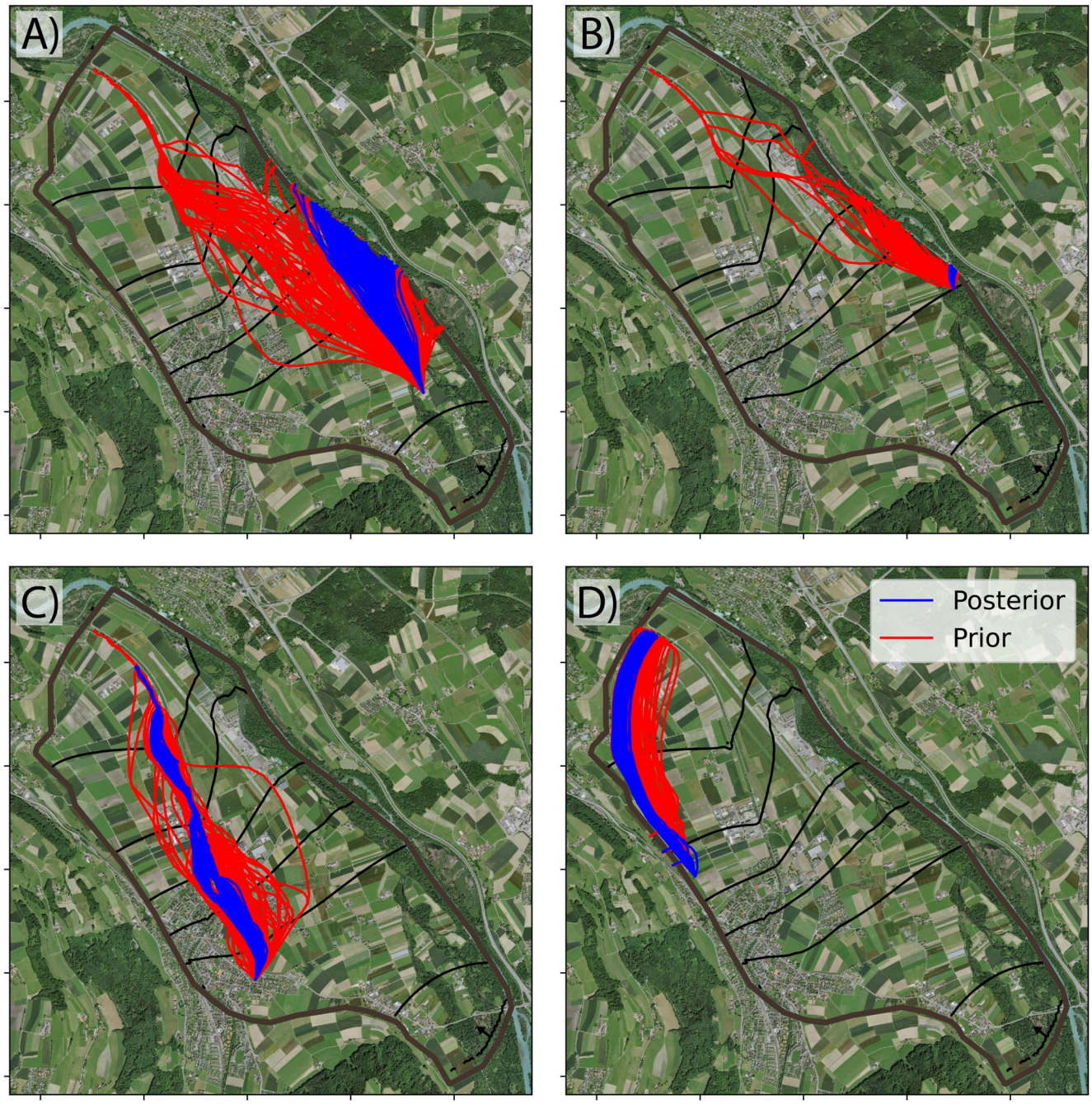


**Figure 15.** Advective particle path computed using the Prior and Posterior models for 4 different starting sites. For readability, only 100 members are represented.

minimization scheme to fit the data almost perfectly. In contrast, our inversion scheme involves much more prior knowledge, such as the number of layers, the distribution of the parameters, and their spatial continuity. The final models are therefore the best models that fit both all the data and the conceptual knowledge of the area. For example, some local areas may show a thin additional layer of clay within the upper aquifer; this is not considered by the conceptual model and will result in a higher residual close to the point. Additionally, as shown in Figure 13, the mean of the realizations is very similar to the deterministic inversion. It is a good indicator that the space of uncertainty around this mean optimum is correctly represented by the ensemble of posterior models. It is also possible that the variogram models chosen for the parameter fields simulation are not suited, or that the number of pilot point should be increased. The balance between the representation of the detailed complexity and the

large-scale structures always depends on the purpose of the final model, and one should be aware that the prior model will need to reflect the level of expected details. For hydrological observations, we denote only a marginal decrease of the residual. The fact that only a steady state model was used limits the amount of information which can be gathered from the hydrological data. The use of dynamic data such as tracer test or piezometric time series could probably increase the contribution of the hydraulic data.

We do not assume a specific petrophysical relationship a priori. The geostatistical simulations for each petrophysical parameter are independent. The only prior knowledge that links resistivity and hydraulic conductivity comes from the prior distribution of the mean values for each layer (and rock types) drawn at the pilot point locations. This means that a permeable layer will always be associated with high resistivities; however, extremely high values within the high aquifer resistivity range will not necessarily be associated with high permeability within the aquifer values range.

The cross-validation performed on the prior models has revealed a bias of about −4 m on average. Even if this error might seem negligible, for an aquifer that has a mean thickness of about 5–6 m, it represents an important bias in terms of available resources. We interpret this result as a classical example of sampling bias and should be carefully considered when making geological models. The prior mean depth used for the geostatistical simulation is calculated by averaging the available thickness data from the boreholes. However, when the aquifer is deep, most drilled wells for geothermal heat pumps or irrigation do not reach the bottom of the aquifer. Therefore, the exact thickness of the aquifer is unknown at these locations. This leads to a sampling bias toward smaller thicknesses and consequently to a bias in the prior model. We note that the ESMDA approach successfully identified the bias and recenter the distribution. Choosing a correct prior when bias exists in the available data is a general issue when dealing with Bayesian approaches, and was previously noted in the review by Linde et al. (2015). Properly capturing the heterogeneity and statistical parameters from the wells is straightforward with variogram-based methods and the simplicity of the approach makes the transcription of prior conceptual knowledge into statistics easier than with other more advanced object-based or multiple-point statistics modeling methods. Combining variogram-based geostatistical methods with inversion offers a good balance between simplicity and realism. If the prior is not totally statistically correct (such as the bias in the mean altitude of a geological interface in this study), data assimilation may correct it. The variogram of the posterior distribution of the base surface of the upper aquifer was compared to the original variogram model from the prior simulation in Figure 16. It reveals that data aggregation has significantly changed the variogram, its sill has been reduced by a factor of almost two. This posterior variogram could help to scale accordingly similar areas that are not covered by geophysical measurement by assuming that the statistics are similar. Getting reliable and data-based variogram models in geology is always difficult, and applying this methodology to other environments could create potential databases of analogs.

As mentioned above, the total computing time was about 4h20 per iteration. The complete workflow was parallelized on 30 CPUs. Despite the reduced-dimensionality strategy, the forward simulations still represent 75% of the computing time. To accelerate even further the algorithm, one possible future research direction could be to employ non-physical surrogate modeling, such as, for example, machine learning (Bording et al., 2021; Juda & Renard, 2021). Using a less computationally expensive approach for the first iterations, or the complete inversion, could significantly speed up the computation.

## 6. Conclusion

This paper has shown the feasibility of integrating geological observations, an explicit geological concept, hydrogeological, and geophysical data on a shallow quaternary aquifer. Using the ESMDA algorithm, the methodology is able to effectively integrate all these data and provide robust uncertainty estimation.

The approach consists in generating and updating a consistent ensemble of 3D geological models that are obtained using the ArchPy hierarchical approach. To accelerate the ESMDA computation, we propose to run the forward physical simulations in a reduced dimension that is different for the geophysical and hydrogeological problems. In addition, the nature of the ESMDA algorithm allows for benefiting heavily from parallel computing. The application of the methodology to the upper Aare Valley in Switzerland, with actual field data, demonstrates the applicability of the proposed workflow in real situations. The stochastic nature of the method and the existence of the underlying complete 3D geological model allows for estimating uncertainty, with the advantage of being

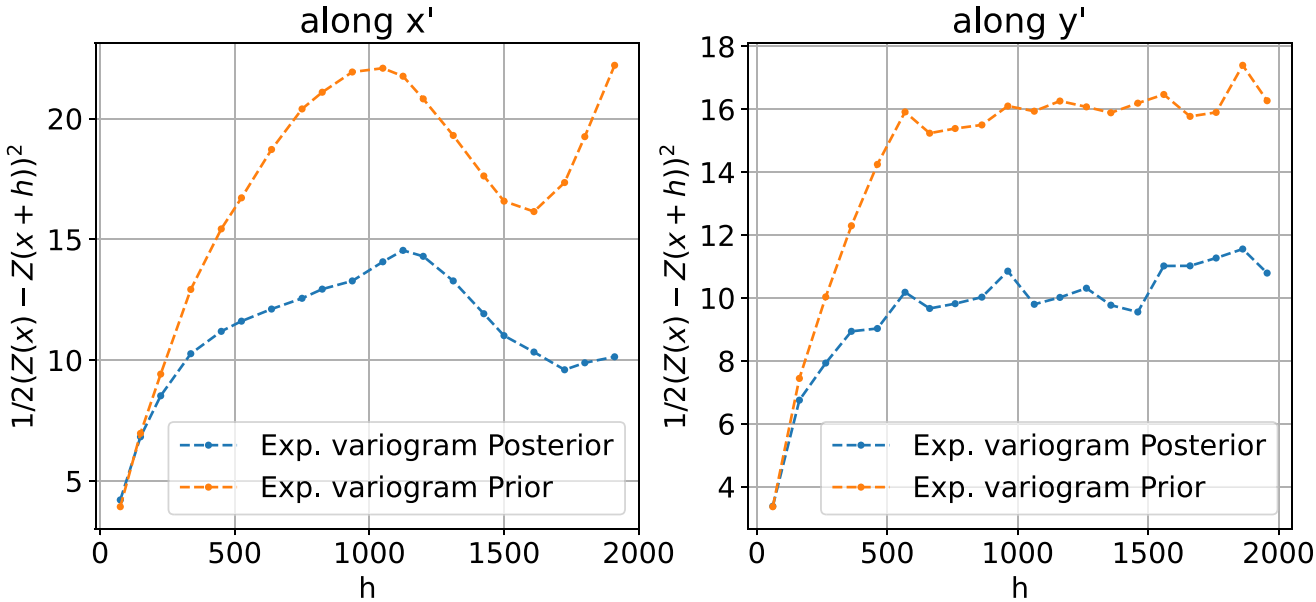


**Figure 16.** Experimental variogram on the bottom of the aquifer surface (surface 1) along the *X* and *Y* axis for the posterior and the prior.

able to retrieve not only petrophysical fields but also some underlying quantities derived from the model, such as the thickness of the aquifer, or the probability of occurrence of a protection layer above an aquifer. This feature can be used to deliver specific critical information for decision-makers which would not be simple to derive at the aquifer scale from geophysical data only.

To conclude, the results presented in this paper demonstrate that coupling the ArchPy methodology with ESMDA is a promising approach that could be applied to a wide variety of problems. All the components of this methodology are available as open-source Python-based software that can easily be modified. For example, any other forward simulator could be coupled with the existing code to extend it to a wide variety of situations to constrain the geological models with other types of geophysical observations or other hydrological processes such as heat or solute transport.

## Data Availability Statement

All the data and codes used in this study are previously published and can be accessed. The tTEM and the heads data used for inversion in the study are available respectively at Neven et al. (2021) and Kellerhals et al. (1981). The heads data, gathered from Kellerhals et al. (1981), are also available in Supporting Information S1 (Table S3). The ArchPy modeling software is available at Schorpp et al. (2022). Finally, the two forward simulations codes can be obtained, respectively, in Auken et al. (2009) and Hughes et al. (2017). To facilitate reproducibility, the ArchPy models (prior and posterior) and the flow model used for the flow simulation can be found in the zenodo repository (https://zenodo.org/record/8047723, Neven and Renard (2023)).

**Acknowledgments**
The authors warmly thank Ludovic Schorpp for his help in coupling ArchPy with ESMDA, and Julien Straubaar for developing some geostatistical simulation tools used within this publication. In addition, the authors want to thank Anders Vest Christiansen for his help related to the geophysical forwards, and the Swiss Geological Survey (Swisstopo) for their help related to data standardization. This research was supported by the Swiss National Science Foundation through the project Phenix (Grant 182600). Finally, we would like to express our sincere gratitude to the three anonymous reviewers for their feedback and constructive comments, which greatly enhanced the quality of this paper. Open access funding provided by Universite de Neuchatel.

## References

Adams, K. H., Reager, J. T., Rosen, P., Wiese, D. N., Farr, T. G., Rao, S., et al. (2022). Remote sensing of groundwater: Current capabilities and future directions. *Water Resources Research*, *58*(10), e2022WR032219. https://doi.org/10.1029/2022wr032219

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:aeakff>2.0.co;2

Asher, M. J., Croke, B. F., Jakeman, A. J., & Peeters, L. J. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, *51*(8), 5957–5973. https://doi.org/10.1002/2015wr016967

Auken, E., Christiansen, A. V., Kirkegaard, C., Fiandaca, G., Schamper, C., Behroozmand, A. A., et al. (2015). An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data. *Exploration Geophysics*, *46*(3), 223–235. https://doi.org/10.1071/eg13097

Auken, E., Christiansen, A. V., Westergaard, J. H., Kirkegaard, C., Foged, N., & Viezzoli, A. (2009). An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system. *Exploration Geophysics*, *40*(2), 184–192. https://doi.org/10.1071/eg08128

Auken, E., Foged, N., Larsen, J. J., Lassen, K. V. T., Maurya, P. K., Dath, S. M., & Eiskjær, T. T. (2019). tTEM—A towed transient electromagnetic system for detailed 3D imaging of the top 70 m of the subsurface. *Geophysics*, *84*(1), E13–E22. https://doi.org/10.1190/geo2018-0355.1

Bandou, D., Schlunegger, F., Kissling, E., Marti, U., Schwenk, M., Schläfli, P., et al. (2022). Three-dimensional gravity modelling of a quaternary overdeepening fill in the Bern area of Switzerland discloses two stages of glacial carving. *Scientific Reports*, *12*(1), 1441. https://doi.org/10.1038/s41598-022-04830-x

Barfod, A. A. S., Møller, I., Christiansen, A. V., Høyer, A.-S., Hoffimann, J., Straubhaar, J., & Caers, J. (2018). Hydrostratigraphic modeling using multiple-point statistics and airborne transient electromagnetic methods. *Hydrology and Earth System Sciences*, *22*(6), 3351–3373. https://doi.org/10.5194/hess-22-3351-2018

Bording, T. S., Asif, M. R., Barfod, A. S., Larsen, J. J., Zhang, B., Grombacher, D. J., et al. (2021). Machine learning based fast forward modelling of ground-based time-domain electromagnetic data. *Journal of Applied Geophysics*, *187*, 104290. https://doi.org/10.1016/j.jappgeo.2021.104290

Bouzaglou, V., Crestani, E., Salandin, P., Gloaguen, E., & Camporese, M. (2018). Ensemble Kalman filter assimilation of ERT data for numerical modeling of seawater intrusion in a laboratory experiment. *Water*, *10*(4), 397. https://doi.org/10.3390/w10040397

Chen, Y., & Oliver, D. S. (2013). Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences*, *17*(4), 689–703. https://doi.org/10.1007/s10596-013-9351-5

Chilès, J.-P., & Delfiner, P. (2009). *Geostatistics: Modeling spatial uncertainty*. John Wiley & Sons.

Christensen, N. K., Minsley, B. J., & Christensen, S. (2017). Generation of 3-D hydrostratigraphic zones from dense airborne electromagnetic data to assess groundwater model prediction error. *Water Resources Research*, *53*(2), 1019–1038. https://doi.org/10.1002/2016wr019141

Christiansen, A. V., Auken, E., & Sørensen, K. (2009). The transient electromagnetic method. In R. Kirsch (Ed.), *Groundwater geophysics: A tool for hydrogeology* (pp. 179–226). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88405-7_6

Deutsch, C. V., & Journel, A. G. (1992). *GSLIB. Geostatistical software library and user's guide*. Oxford University Press.

Dumont, M., Reninger, P., Pryet, A., Martelet, G., Aunay, B., & Join, J. (2018). Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies. *Journal of Applied Geophysics*, *157*, 1–9. https://doi.org/10.1016/j.jappgeo.2018.06.020

Emerick, A. A. (2016). Analysis of the performance of ensemble-based assimilation of production and seismic data. *Journal of Petroleum Science and Engineering*, *139*, 219–239. https://doi.org/10.1016/j.petrol.2016.01.029

Emerick, A. A., & Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, *55*, 3–15. https://doi.org/10.1016/j.cageo.2012.03.011

Engebretsen, K. W., Zhang, B., Fiandaca, G., Madsen, L. M., Auken, E., & Christiansen, A. V. (2022). Accelerated 2.5-D inversion of airborne transient electromagnetic data using reduced 3-D meshing. *Geophysical Journal International*, *230*(1), 643–653. https://doi.org/10.1093/gji/ggac077

Evensen, G. (2009). *Data assimilation*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03711-5

Evensen, G. (2018). Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, *22*(3), 885–908. https://doi.org/10.1007/s10596-018-9731-y

Farquharson, C. G., & Oldenburg, D. W. (1993). Inversion of time-domain electromagnetic data for a horizontally layered Earth. *Geophysical Journal International*, *114*(3), 433–442. https://doi.org/10.1111/j.1365-246x.1993.tb06977.x

Fitterman, D. V. (1987). Examples of transient sounding for ground-water exploration in sedimentary aquifers. *Ground Water*, *25*(6), 685–692. https://doi.org/10.1111/j.1745-6584.1987.tb02209.x

Freulon, X., & de Fouquet, C. (1993). Conditioning a Gaussian model with inequalities. *Geostatistics Tróia'92*, *1*, 201–212.

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B*, *69*(2), 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Graf, H. R., & Burkhalter, R. (2016). Quaternary deposits: Concept for a stratigraphic classification and nomenclature—An example from northern Switzerland. *Swiss Journal of Geosciences*, *109*(2), 137–147. https://doi.org/10.1007/s00015-016-0222-7

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, *28*(1), 100. https://doi.org/10.2307/2346830

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., & Madsen, B. (2003). Methodology for construction, calibration and validation of a national hydrological model for Denmark. *Journal of Hydrology*, *280*(1–4), 52–71. https://doi.org/10.1016/S0022-1694(03)00186-0

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, *126*(3), 796–811. https://doi.org/10.1175/1520-0493(1998)126⟨0796:dauaek⟩2.0.co;2

Høyer, A.-S., Jørgensen, F., Sandersen, P., Viezzoli, A., & Møller, I. (2015). 3D geological modelling of a complex buried-valley network delineated from borehole and AEM data. *Journal of Applied Geophysics*, *122*, 94–102. https://doi.org/10.1016/j.jappgeo.2015.09.004

Hughes, J. D., Langevin, C. D., & Banta, E. R. (2017). *Documentation for the MODFLOW 6 framework*. US Geological Survey. https://doi.org/10.3133/tm6a57

Irving, J., & Singha, K. (2010). Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. *Water Resources Research*, *46*(11), W11514. https://doi.org/10.1029/2009wr008340

Jardani, A., Revil, A., & Dupont, J. (2013). Stochastic joint inversion of hydrogeophysical data for salt tracer test monitoring and hydraulic conductivity imaging. *Advances in Water Resources*, *52*, 62–77. https://doi.org/10.1016/j.advwatres.2012.08.005

Jørgensen, F., Høyer, A.-S., Sandersen, P. B., He, X., & Foged, N. (2015). Combining 3D geological modelling techniques to address variations in geology, data type and density—An example from Southern Denmark. *Computers & Geosciences*, *81*, 53–63. https://doi.org/10.1016/j.cageo.2015.04.010

Jørgensen, F., Møller, R. R., Nebel, L., Jensen, N.-P., Christiansen, A. V., & Sandersen, P. B. E. (2013). A method for cognitive 3D geological voxel modelling of AEM data. *Bulletin of Engineering Geology and the Environment*, *72*(3–4), 421–432. https://doi.org/10.1007/s10064-013-0487-2

Juda, P., & Renard, P. (2021). An attempt to boost posterior population expansion using fast machine learning algorithms. *Frontiers in Artificial Intelligence*, *4*, 624629. https://doi.org/10.3389/frai.2021.624629

Juda, P., Straubhaar, J., & Renard, P. (2022). Comparison of three recent discrete stochastic inversion methods and influence of the prior choice. *Comptes Rendus Geoscience*, *355*(S1), 1–26. https://doi.org/10.5802/crgeos.160

Kang, S., Cockett, R., Heagy, L. J., & Oldenburg, D. W. (2015). Moving between dimensions in electromagnetic inversions. In *SEG technical program expanded abstracts 2015*. Society of Exploration Geophysicists. https://doi.org/10.1190/segam2015-5930379.1

Kang, S., Knight, R., Greene, T., Buck, C., & Fogg, G. (2021). Exploring the model space of airborne electromagnetic data to delineate large-scale structure and heterogeneity within an aquifer system. *Water Resources Research*, *57*(10), e2021WR029699. https://doi.org/10.1029/2021wr029699

Kang, X., Shi, X., Revil, A., Cao, Z., Li, L., Lan, T., & Wu, J. (2019). Coupled hydrogeophysical inversion to identify non-Gaussian hydraulic conductivity field by jointly assimilating geochemical and time-lapse geophysical data. *Journal of Hydrology*, *578*, 124092. https://doi.org/10.1016/j.jhydrol.2019.124092

Kellerhals, P., Haefeli, C., & Tröhler, B. (1981). Grundlagen für Schutz und Bewirtschaftung der Grundwasser des Kantons Bern Hydrogeologie Aaretal, zwischen Thun und Bern. (Technical Report). *Bern: Wasser- u. Energiewirtschaftsamt des Kantons Bern*.

Knight, R., Smith, R., Asch, T., Abraham, J., Cannia, J., Viezzoli, A., & Fogg, G. (2018). Mapping aquifer systems with airborne electromagnetics in the Central Valley of California. *Groundwater*, *56*(6), 893–908. https://doi.org/10.1111/gwat.12656

Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, *29*(7), 945–958. https://doi.org/10.1016/j.advwatres.2005.08.006

Lam, D.-T., Renard, P., Straubhaar, J., & Kerrou, J. (2020). Multiresolution approach to condition categorical multiple-point realizations to dynamic data with iterative ensemble smoothing. *Water Resources Research*, *56*(2), e2019WR025875. https://doi.org/10.1029/2019wr025875

Lang, A., & Potthoff, J. (2011). Fast simulation of Gaussian random fields. *Monte Carlo Methods and Applications*, *17*(3). https://doi.org/10.1515/mcma.2011.009

Lemieux, J.-M., Sudicky, E. A., Peltier, W. R., & Tarasov, L. (2008). Dynamics of groundwater recharge and seepage over the Canadian landscape during the Wisconsinian glaciation. *Journal of Geophysical Research*, *113*(F1), F01011. https://doi.org/10.1029/2007JF000838

Li, N., McLaughlin, D., Kinzelbach, W., Li, W., & Dong, X. (2015). Using an ensemble smoother to evaluate parameter uncertainty of an integrated hydrological model of Yanqi basin. *Journal of Hydrology*, *529*, 146–158. https://doi.org/10.1016/j.jhydrol.2015.07.024

Linde, N., & Doetsch, J. (2016). Joint inversion in hydrogeophysics and near-surface geophysics. In *Integrated imaging of the earth* (pp. 117–135). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118929063.ch7

Linde, N., Renard, P., Mukerji, T., & Caers, J. (2015). Geological realism in hydrogeological and geophysical inverse modeling: A review. *Advances in Water Resources*, *86*, 86–101. https://doi.org/10.1016/j.advwatres.2015.09.019

Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, *100*(B7), 12431–12447. https://doi.org/10.1029/94jb03097

Neven, A., Christiansen, A. V., & Renard, P. (2022). Automatic stochastic 3D clay fraction model from tTEM survey and borehole data. *Scientific Reports*, *12*(1), 17112. https://doi.org/10.1038/s41598-022-21555-z

Neven, A., Maurya, P. K., Christiansen, A. V., & Renard, P. (2021). tTEM20aar: A benchmark geophysical data set for unconsolidated fluvioglacial sediments. *Earth System Science Data*, *13*(6), 2743–2752. https://doi.org/10.5194/essd-13-2743-2021

Neven, A., & Renard, P. (2023). *Archpy/ES-MDA inverse results*. Zenodo. https://doi.org/10.5281/ZENODO.8047723

Neven, A., Schorpp, L., & Renard, P. (2022). Stochastic multi-fidelity joint hydrogeophysical inversion of consistent geological models. *Frontiers in Water*, *4*. https://doi.org/10.3389/frwa.2022.989440

Paine, J. G. (2003). Determining salinization extent, identifying salinity sources, and estimating chloride mass using surface, borehole, and airborne electromagnetic induction methods. *Water Resources Research*, *39*(3), 1059. https://doi.org/10.1029/2001wr000710

Reuschen, S., Jobst, F., & Nowak, W. (2021). Efficient discretization-independent Bayesian inversion of high-dimensional multi-Gaussian priors using a hybrid MCMC. *Water Resources Research*, *57*(8), e2021WR030051. https://doi.org/10.1029/2021wr030051

Ringrose, P., & Bentley, M. (2015). *Reservoir model design*. Springer Netherlands. https://doi.org/10.1007/978-94-007-5497-3

Scholl, C., Helwig, S. L., Tezkan, B., Goldman, M., & Kafri, U. (2009). 1-D multimodel joint inversion of TEM-data over multidimensional structures. *Geophysical Journal International*, *176*(1), 81–94. https://doi.org/10.1111/j.1365-246x.2008.03973.x

Schorpp, L., Straubhaar, J., & Renard, P. (2022). Automated hierarchical 3D modeling of quaternary aquifers: The ArchPy approach. *Frontiers in Earth Science*, *10*. https://doi.org/10.3389/feart.2022.884075

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717921

United Nations Educational Scientific and Cultural Organization. (2022). *The united nations world water development report 2022 the united nations world water development report 2022*. United Nations Educational Scientific and Cultural.

van Leeuwen, P. J., & Evensen, G. (1996). Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, *124*(12), 2898–2913. https://doi.org/10.1175/1520-0493(1996)124⟨2898:daaimi⟩2.0.co;2

Volken, S., Preisig, G., & Gaehwiler, M. (2016). GeoQuat: Developing a system for the sustainable management, 3D modelling and application of quaternary deposit data. *Swiss Bulletin for Applied Geology*, *21*, 3–16. https://doi.org/10.5169/seals-658182

Wen, X.-H., & Chen, W. (2005). Real-time reservoir model updating using ensemble Kalman filter. In *SPE reservoir simulation symposium* (p. SPE92991). SPE. https://doi.org/10.2118/92991-ms

Xu, T., Gómez-Hernández, J. J., Chen, Z., & Lu, C. (2021). A comparison between ES-MDA and restart EnKF for the purpose of the simultaneous identification of a contaminant source and hydraulic conductivity. *Journal of Hydrology*, *595*, 125681. https://doi.org/10.1016/j.jhydrol.2020.125681