

# Bayesian Evidential Learning: an alternative to hydrogeophysical coupled inversion

Thomas HERMANS<sup>1\*</sup>, Nicolas COMPAIRE<sup>1,2,3</sup>, Robin THIBAUT<sup>1</sup>, Nolwenn LESPARRE<sup>3</sup>

<sup>1</sup>Ghent University, Department of Geology <sup>2</sup>Institut Supérieur de l'Aéronautique et de l'Espace (ISAE)

<sup>3</sup>Université de Strasbourg, CNRS, EOST, ENGEES, Institut Terre et Environnement de Strasbourg

## Summary

Deterministic geophysical inversion suffers from a lack of realism because of the regularization, while stochastic inversion allowing for uncertainty quantification is computationally expensive. In this contribution, we propose to use Bayesian Evidential Learning as an alternative to hydrogeophysical coupled inversion. We demonstrate the ability of the approach to successfully predict a hydrogeological target from time-lapse ERT data in the context of a heat injection and storage experiment.

## Introduction

The current paradigm in geophysical imaging relies on inverse modelling yielding the “best” solution reproducing the data within its estimated noise level. Since geophysical inverse problems are ill-posed, and solutions are non-unique, the most commonly used approach is to solve a regularized version of the problem, where the data misfit is minimized under some model constraints (Tikhonov and Arsenin, 1977). The image produced is generally the result of a non-linear inversion or process and represents only one possibility amongst an infinite number of solutions (Linde et al., 2015). Recently, the nature and extent of the non-unicity of the solution, related to the intrinsic uncertainty of the subsurface and the noisy nature of the data, has been acknowledged. The need to account for uncertainty at each step of the imaging process has been recognized (Linde et

al., 2015). Therefore, the development of stochastic methodologies for the inversion of geophysical data has become mandatory (e.g., Wathelet, 2008; Trainor-Guitton and Hoversten, 2011; Lochbühler et al., 2014). The currently most widespread approach to stochastically solving the geophysical inverse problem is to use Markov chain Monte Carlo (MCMC) methods (Sambridge and Mosegaard, 2002; Vrugt et al., 2013). MCMC methods are based on algorithms sampling the posterior distribution of the model parameter space. However, these methods are very computationally demanding, because they typically require hundreds of thousands of forward model runs based on a chain of models dependent on each other. Each step in the process requires to solve the forward problem to estimate the likelihood of the current model (i.e., its data misfit), and hundreds of thousands of iterations might be needed to converge, especially for complex subsurface models (e.g., Irving and Singha, 2010).

However, obtaining a single image is often not the end-goal of a geophysical study. Instead, a geologist, hydrogeologist or reservoir-engineer often wants to derive some interesting (hydro)geological features from it. Post-processing techniques are then used to extract features from this image. Interpretation often relies on uncertain empirical petrophysical relationship (Rubin and Hubbard, 2005; Paasche et al. 2006) and suffers from the spatially-dependent resolution of the image (Day-Lewis et al., 2005). Taking this into consideration when interpreting a geophysical image

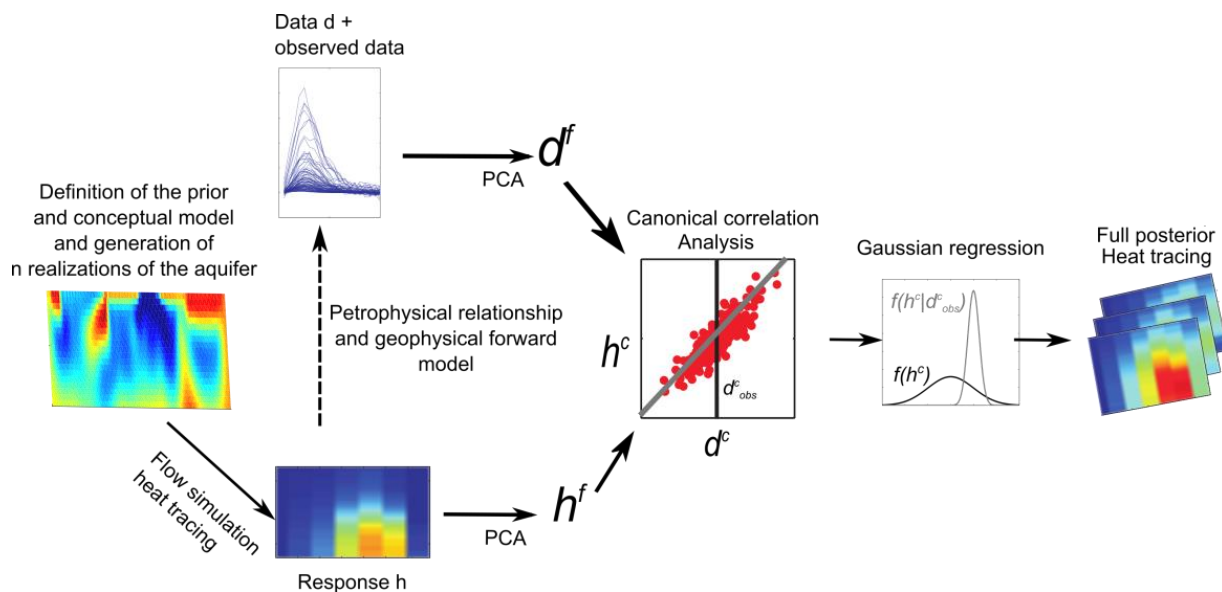


Figure 1: Illustration of the different steps of the BEL framework

## BEL: an alternative to coupled inversion

generally requires the re-introduction of the concept of uncertainty or stochasticity (e.g., Mooney et al., 2005; Singha and Mooney, 2006; Hermans and Irving, 2017)

Uncertain parameter	Range
Log K mean	U[-1, -4], K in m/s
$\sigma_{\log K}$	U[0.05, 2], K in m/s
Effective porosity	U[0.05, 0.3]
Variogram main range	U[1, 10] m
Anisotropy ratio	U[0, 0.5]
Orientation	U[0, $\pi$ ]
Natural gradient	U [0.0005, 0.00167]

Table 1. Parameters of the prior model

The spatially-dependent resolution of geophysical images and the difficulty to extract quantitative features out of it gave birth to coupled inversion approaches, where geophysical data are integrated to hydrogeological (or other dynamical) inverse problems through a petrophysical relationship (Irving and Singha, 2010; Christensen et al., 2016). Such an approach allows to produce subsurface models consistent with both geophysical and hydrogeological data and to integrate prior geological information. However, parameterizing and calibrating such a model remains very complex and coupled hydrogeophysical inversion often require simplifying assumptions (Christensen et al., 2016) or require computationally expensive stochastic approaches (Irving and Singha, 2010; Hermans et al., 2015).

In this contribution, we propose to circumvent coupled hydrogeophysical inversion through Bayesian Evidential Learning (BEL; Hermans et al., 2016, 2018). The rationale behind this approach is that model inversion is rarely the final objective of modelling but rather a mean to generate

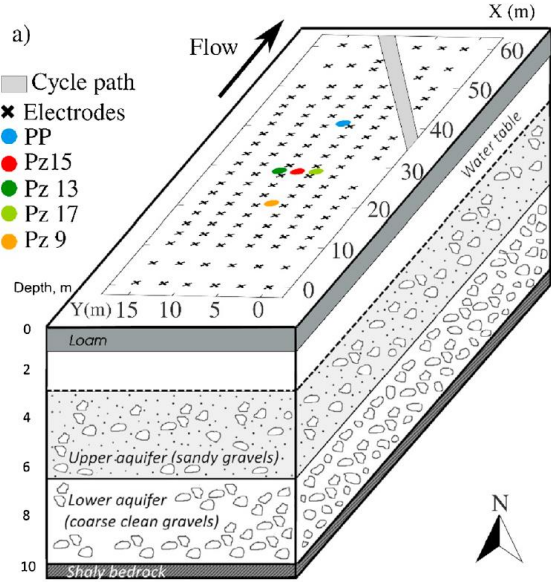


Figure 2: Experimental set-up

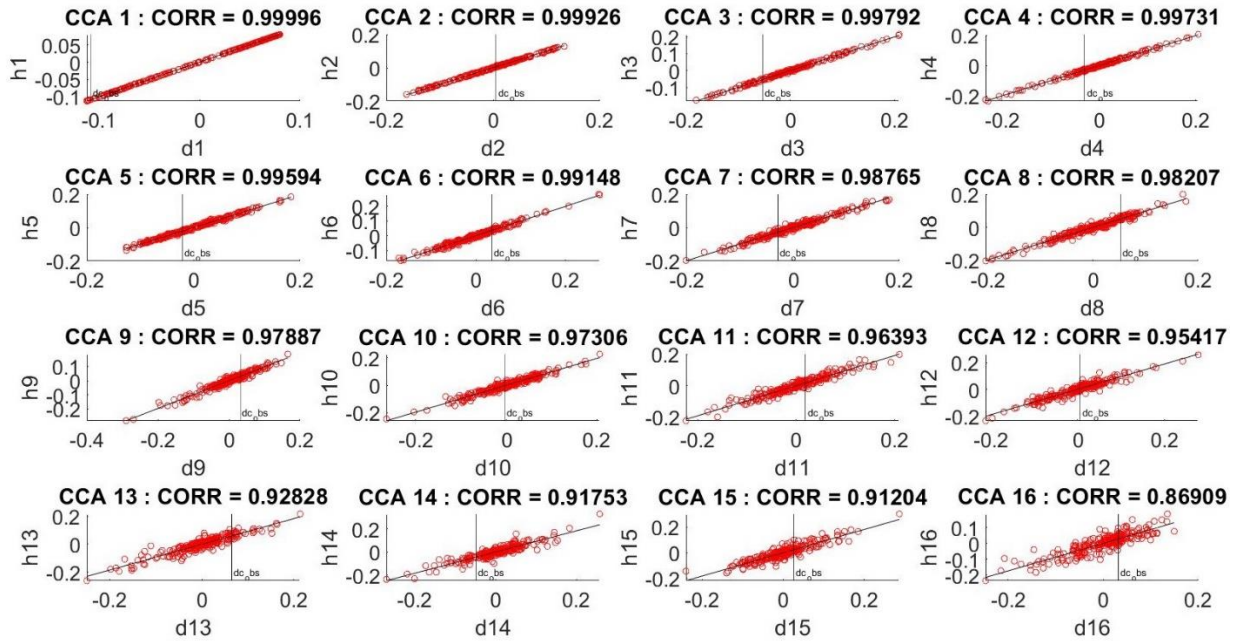


Figure 3: Results of canonical correlation analysis (16 first components)

## BEL: an alternative to coupled inversion

some predictions related to the study site. In BEL, a statistical relationship is sought between data and prediction through physically-based forward modelling, using a limited number (typically a few hundreds) of realizations sampled from the prior model space. This allows to maintain the geological realism of the model and prediction, while replacing the complex inversion in the model space by a simple regression in the joint data-prediction space. This allows to estimate the posterior distribution of the prediction at limited costs. In this contribution, we illustrate the method to predict the thermal affected zone during a heat injection and storage experiment carried out in an alluvial aquifer (Lesparre et al., 2019).

### Method

We refer to Hermans et al. (2016, 2018) for a detailed description of BEL applied to geophysical data. Here, we only shortly describe the different steps (Figure 1):

- 1) The prior model is defined, i.e., a prior uncertainty range is given for all model parameters. In our cases, the hydraulic conductivity field is simulated by sequential Gaussian simulation, with uncertain mean, standard deviation and variogram range, anisotropy and orientation. In addition, the amplitude of the natural gradient and the effective porosity are also uncertain (Table 1). We sample 250 realizations from the prior.
- 2) The experiment is simulated for each sample. It consists in the injection of heated water ( $\Delta T = 28.6 K$ ) at a rate of  $3 \text{ m}^3/\text{h}$  during 6 hours at 5 m depth in a 10 m thick aquifer (Figure 2), followed by a storage period (91 h), a pumping period (15.5 h at  $3 \text{ m}^3/\text{h}$ ) and a final resting period (Lesparre et al., 2019). Note that if data and prediction correspond to different experiments, both should be simulated.
- 3) The temperature distribution is extracted from the heat transfer simulation at every time step to define the prediction (thermal affected zone) and the data (time-lapse ERT) collected using 6 parallel profiles of 21 electrodes spaced by 2 m (Figure 2). We use a linear relationship for the dependence of the electrical conductivity to the temperature (Hermans et al., 2014).
- 4) The original data set contains 1948 quadrupoles, and the prediction is the temperature in a volume composed of 3808 cells, both for 106 time-steps. We first reduce the dimensions of both variables using principal component analysis (PCA). 99.9% of the data and 91.3 % of the prediction are explained by 30 dimensions.
- 5) We apply canonical correlation analysis (CCA) (Figure 3). This process yields a set of independent

- 6) We sample the joint data-prediction distribution in the reduced dimension space by applying the same dimension reduction to the observed data. Noise is propagated using a Monte Carlo approach to estimate the covariance matrix in the reduced space (Hermans et al., 2016). In contrast to Hermans et al. (2016, 2018), we do not use a linear regression because the distributions are not Gaussian. Instead, we use Kernel Density Estimation (Hermans et al., 2019, Michel et al., 2020). This provides the posterior distribution of the prediction in the reduced space.
- 7) The posterior distribution is back-transformed in the original physical space.

### Results

The validity of the approach is demonstrated using a synthetic case. One of the realizations of the prior is taken as the true model and the corresponding data set as the observed field data set.

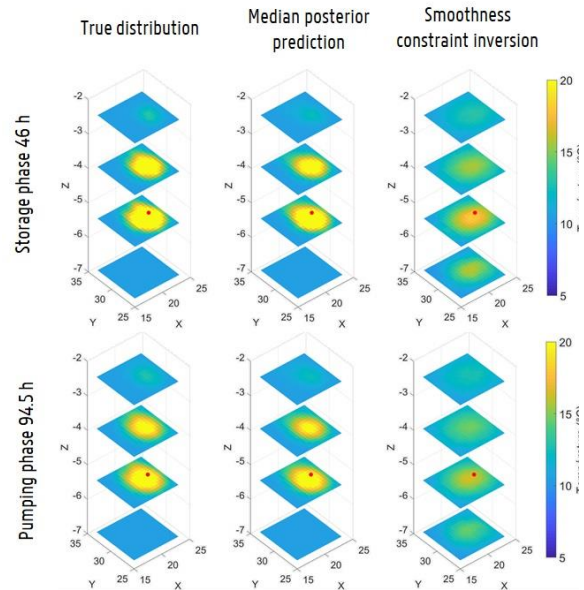


Figure 4: This caption is placed inside the frame

Figure 4 shows a comparison between the true prediction, the median posterior prediction and a deterministic inversion using the standard smoothness constraint. We can clearly observe that samples from the posterior distribution are

## BEL: an alternative to coupled inversion

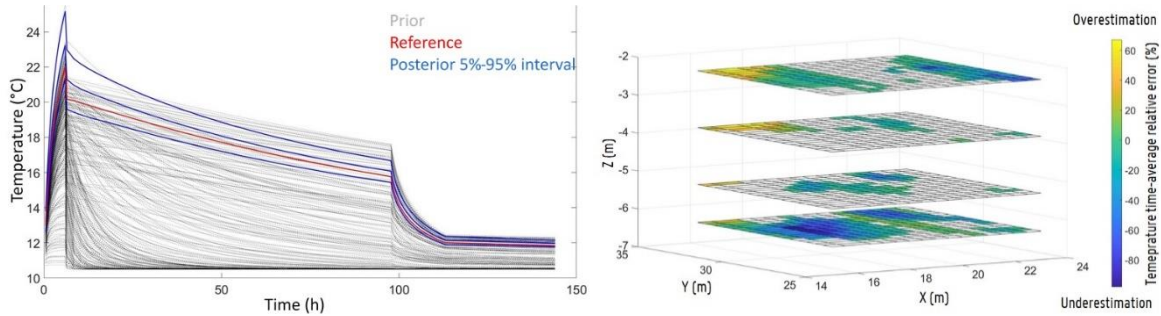


Figure 5: Estimation of the average temperature around the injection well throughout the experiment (left) and distribution of the average error of estimation (right).

hydrogeologically realistic. They correctly identify the center of the hot plume, its spatial extent and the amplitude of the temperature variation at any time of the experiment. In particular, the hot plume is vertically limited to the thickness of the screen interval. In contrast, the smoothness constraint inversion underestimates the amplitude of the maximum temperature and spreads out the thermal affected zone over a much larger volume, illustrating the typical drawback of deterministic inversions.

BEL also allows to estimate the uncertainty range, as it is easy to sample any number of samples from the posterior distribution. Figure 5 shows the temporal evolution of the temperature in the vicinity of the injection borehole for the prior realizations, the true prediction and the 5%-95% confidence interval deduced from BEL posterior distribution. We see that the median posterior sample is close to the true model and that the uncertainty range is maximum at the end of the injection phase and in the storage phase, with a value about 1°C. Note that 1°C correspond to a change of electrical resistivity of about 2%. The uncertainty range is much narrower during the pumping phase and for the final resting period.

We estimated the average estimation error of BEL through time (Figure 5, right). The percentage on the figure indicates the number of time steps for which the true prediction is not in the 5-95% interval. Blank cells are cells that are always within this range. We see that deviations occur mostly for grid cells located outside of the center of the plume (Figure 3), i.e., for grid cells that are not experiencing large changes of temperature. In other words, those cells correspond to zones with a very small variation amplitude, which is beyond the resolution of ERT ( $< 0.5\text{-}1\text{ }^{\circ}\text{C}$ ). The few cells located in the middle of the model close to the injection borehole are due to the slight overestimation tendency after the pumping phase (Figure 5, left). In comparison, the smoothness constraint inversion lies outside of the confidence interval for every time step.

## Conclusion

In this contribution, we propose to use Bayesian Evidential Learning as an alternative to hydrogeophysical inversion. Instead of inverting the geophysical data set to get the posterior distribution of the model parameters, which would be extremely difficult and computationally expensive, we circumvent the inversion by learning a direct relationship between the geophysical data (time-lapse ERT) and the hydrogeological prediction (temperature distribution). It is then straightforward to generate the posterior distribution of the prediction for the observed data set.

This approach is particularly efficient. We only need 250 samples out of the prior distribution to learn the statistical relationship. Therefore, the posterior distribution can be calculated by running only 250 times the forward heat transport model. This is several orders of magnitude less than what would be necessary for an MCMC approach. This is possible because the prediction is much less complex than the model parameter distribution. The latter is a heterogeneous hydraulic conductivity distribution with a complex spatial correlation (see Table 1), while the temperature distribution is relatively smooth and of limited volume because of the conduction processes taking place in the subsurface (heat exchange between the solid matrix and the fluid).

The framework is easily adaptable, so not only can it be applied to hydrogeological data (Hermans et al., 2019), but also could be extended to combine geophysical and hydrogeological data. It can also be applied to prediction corresponding to another experiment or even another location in the aquifer (Hermans et al., 2018). Therefore, BEL is a suitable framework to replace hydrogeophysical inversion when the distribution of the model parameters is not an objective of the study.

**BEL: an alternative to coupled inversion**