Use of convolutional neural networks with encoder-decoder structure for predicting the inverse operator in hydraulic tomography

Abderrahim Jardani, T. M. Vu, Pierre Fischer

Use of Convolutional Neural Networks with encoder-decoder structure for predicting the inverse operator in hydraulic tomography.

A. Jardani [(1*)], T.M.Vu [(1)] , and P.Fischer [(2)]

(1) Normandy University, UNIROUEN, UNICEAN, UMR CNRS 6143 M2C Morphodynamique Continentale et Côtière, Université de Rouen.
(2) HydroSciences Montpellier, Univ. Montpellier, CNRS, IRD, Montpellier, France

- Corresponding author : abderrahim.jardani@univ-rouen.fr

Abstract: In this manuscript, we discuss the capabilities of a deep learning algorithm implemented with the Conventional Neural Network concept to characterize the hydraulic properties of aquifers. The algorithm called CNN-HT is designed to predict the inverse operator of hydraulic tomography using a synthetic training dataset in which the hydraulic head data associated with pumping tests are linked to hydraulic transmissivity field. This approach relies on an adaptation of the SegNet network that was initially developed to process image segmentation. The SegNet is composed of encoders and decoders networks. In the encoder, sequential operations with multiple filters, as convolution, batch normalization, max-pooling are performed to identify feature maps of the input data. In the decoder, the up-sampling, convolution, batch normalization and regression operations are used to prepare the output by recovering the loss of spatial resolution that occurred in the encoder process. In this adaptation, we used the least-square iterative formulation at the initial iteration with Jacobian matrix to resize the hydraulic head data to match the size of the output (transmissivity field). This protocol was applied to the hydraulic head data computed numerically by solving the groundwater flow equation for a given transmissivity field, generated geostatistically with Gaussian and spherical variograms. A part of this data was used for training the network and the other part to test its performance. The test step confirmed the effectiveness of this tool in reconstructing the main heterogeneities of the hydraulic properties, and its effectiveness is related to the nature and

quantity of the training data. Moreover, the CNN-HT method provided inversion results of the same quality than those obtained with the Gauss-Newton algorithm using the finite difference or adjoint state method in the computation of the Jacobian matrix. However, the computational time is longer in CNN-HT but this time can be less or of the same order as that of Gauss-Newton using finite difference method.

1. Introduction

Since the 1980s, the hydraulic tomography has been adopted in hydrology as an effective technique for mapping the heterogeneity of the hydraulic properties of aquifers (Neuman, 1987; Gottlieb and Dietrich, 1995; Bohling et al., 2002; Yeh and Liu, 2000). Indeed, this approach can effectively provide insights into the spatial variability of the hydraulic transmissivity and storage coefficient of a porous or fractured aquifer through a combined interpretation of hydraulic data obtained from several pumping tests (Bohling et al., 2002; Berg and Illman, 2011; Cardiff et al., 2013; Fischer et al., 2018). The approach consists of using an inversion algorithm, either deterministic or stochastic, to retrieve the best models that could match the observed piezometric data (Fu & Gómez-Hernández, 2009a; Kitanidis, 1997). In general, the inverse solution is not unique for that reason the inverse formulation usually incorporates a regularization term in the optimization of the objective function $\Psi(\boldsymbol{m})$ that is defined as the sum of data misfit and regularization terms, as below (Tarantola and Valette, 1982):

$$\Psi(\boldsymbol{m}) = \underbrace{\left[\mathrm{F}(\boldsymbol{m}) - \mathbf{h}_{\mathrm{obs}}\right]^{\mathrm{T}} \mathbf{C}_{\mathrm{d}}^{-1} \left[\mathrm{F}(\boldsymbol{m}) - \mathbf{h}_{\mathrm{obs}}\right]}_{\text{Data misfit term}} + \underbrace{\left[\boldsymbol{m} - \boldsymbol{m}_{0}\right]^{\mathrm{T}} \mathbf{C}_{\mathrm{m}}^{-1} \left[\boldsymbol{m} - \boldsymbol{m}_{0}\right]}_{\text{constraint term}}, \tag{1}$$

where the data misfit term assesses the pertinence of the model in term of matching the data and is expressed as a sum of square differences between the observed hydraulic data $\mathbf{h}_{\mathrm{obs}}$ and the numerical hydraulic data derived by solving a forward operator $\mathrm{F}(\boldsymbol{m})$. In hydraulic tomography, the forward problem involves a numerical discretization for solving the

groundwater flow equation subject to Darcy's law in transient or steady modes. $\boldsymbol{m} = -\log[T]$ is the spatial distribution of the negative logarithm of the hydraulic transmissivity T, which is regarded as an unknown field and is expressed logarithmically to ensure its positivity during the inverse process. $\mathbf{C}_{\mathrm{d}}$ is a diagonal matrix for incorporating data uncertainty in the optimization. On the other side, the regularization term is expressed with a prior model $\boldsymbol{m}_0$ and its covariance matrix $\mathbf{C}_{\mathrm{m}}$. This term can be obtained from geological, geophysical and tracer investigations to constrain the inversion to offer only plausible solutions (Lochbühler et al., 2013; Zhao et al., 2016; Soueid et al., 2016). In general, the use of geostatistical constraints is still widely applied in hydraulic tomography by using statistical properties such as mean and covariance to obtain an imagery of hydraulic properties with smooth features (Kitanidis, 1997; Yeh and Liu, 2000).

Once the parameters of the objective function are formulated, its optimization is processed iteratively using one of the three categories of optimization algorithms: deterministic (Li et al., 2005), stochastic (Jimenez et al., 2016) or global (Castagna and Bellin, 2009). These algorithms involve a repetitive numerical solving of the forward problem by applying finite elements or finite difference methods to assess the ability of the proposed model to reach the convergence state. Therefore, the computation time is dependent on the numerical tools, the dimensionality of the unknown parameters and the nature of the optimization algorithm employed. In general, deterministic algorithms are the most used in high-dimensional inverse problems due to their advantage of reaching convergence in a short time compared to stochastic or global algorithms by exploiting the geometric properties of the objective function (gradient and Hessian) (Tarantola and Valette, 1982). Nevertheless, their efficiency depends strongly on the choice of the initial model as they only provide local minima. On the other hand, stochastic methods (such as Markov chain Monte Carlo McMC) are based on the concept of sampling where the solution

78 is obtained by exploring the optimization performance of many randomly generated models and
79 selecting the best of them in terms of minimization of the objective function. Most of the time,
80 algorithms belonging to this category are easy to implement and insensitive to the initial model.
81 However, their applicability to inverse problems with a large number of unknown parameters
82 remains limited due to the high repetitive and heavy computation of the forward problem during
83 the sampling process (Oliver et al., 1997; Fu and Gómez-Hernández, 2009b; Wang et al. 2017;
84 Jardani et al., 2012; Elsheikh et al., 2012). Other global optimization algorithms such as particle
85 swarm optimization, genetic, simulated annealing can also be used to identify the best solution
86 by iteratively evaluating the objective function until the state of convergence. However,
87 convergence can be slow, especially in the case of high-dimensional inverse problems (Scales
88 et al., 1990; Fernández-Martínez et al., 2011).

89 In this paper, we test a new generation of deep learning algorithm that can be considered as a
90 global optimizer in which prediction does not depend on an initial model and the computation
91 of the sensitivity matrix as it is the case for gradient-based methods (Zio, 1997). The concept
92 of this approach is different from previous optimization methods as the process focuses on
93 approximating the inverse function by finding a universal relationship linking the input and
94 output data (in the case of hydraulic tomography between hydraulic pressure and hydraulic
95 transmissivity). Indeed, the deep learning is a powerful generalizer of non-linear and complex
96 functions by identifying a set of parameters such as weights and biases allocated to the hidden
97 layer neurons that process the input data and link them to their corresponding output. The first
98 generation of deep learning algorithms uses a set of hidden layers with a large number of
99 neurons fully connected via high-dimensional weights, which complicates computation and
100 requires a lot of time, especially for handling the imaging tasks (Shen, 2018). However, the
101 emergence of new deep learning architectures with Convolutional Neural Networks (CNN)
102 concept that have been proven in high-resolution image processing have opened the way for

broadening their applicability in the hydraulic characterization (LeCun et al., 1998; Sun, 2018; Zhu and Zabaras, 2018). Indeed, the CNN concept has succeeded in reducing drastically the image processing time thanks to the convolution calculation, which allows to retrieve the local features of the image through small filters (Indolia et al., 2018). These filters cover the whole input image but their activation is done zone by zone with local connections between pixels. This local convolution reduces the number of parameters to optimize during the learning process (LeCun et al., 1998). Among the first attempts at adopting the CNN concept in the realm of tomography by inversion in geosciences, we cite: Sun (2018) trained the Generative Adversarial networks for linking the hydraulic head map and the spatial distribution of hydraulic conductivity. Zhu and Zabaras (2018) proposed a deep convolutional encoder-decoder network to reconstruct the image of hydraulic conductivity from hydraulic head map. Laloy et al. (2018) used the Generative Adversarial networks to generate randomly in 2D and 3D binary hydraulic conductivity fields and MCMC to perform the inversion process. In geophysics realm, Wu and Lin et al. (2019) applied CNN with encoder-decoder networks for mapping the subsurface velocity from seismogram data. Puzyrev and Swidinsky. (2019) used the CNN network to determine the vertical subsurface heterogeneity of electrical conductivity by training electromagnetic data.

In this article, we discuss the use of the convolution neural networks formed according to the SegNet architecture to process hydraulic tomography. This technique was initially designed for the semantic segmentation of objects learned on the images (Badrinarayanan, et al. 2017). To illustrate the inversion principle with the CNN-HT structure, we organize this manuscript with the following outline: the first section will be devoted to the generation of data used in training, validation and test, and the second to the introduction of concept of the CNN-SegNet structure. In the application sections, we discuss the relevance of the approach on theoretical cases under different conditions on training data with high and low resolution and contaminated or not by

the noise. We also devote a section to the comparison of the results obtained with CNN-SegNet code with those determined with the Gauss-Newton algorithm.

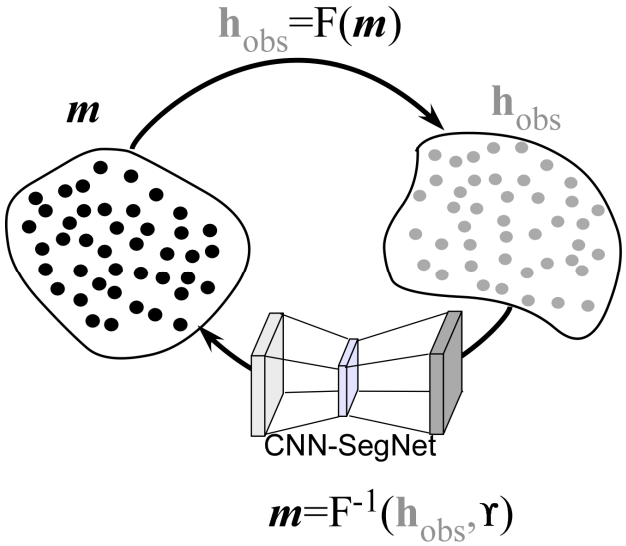
2. Theoretical concept:

2.1.Preparation of Training data:

In deep learning algorithm, the first step concerns the construction of training data that will condition the accuracy of the result obtained from the predicted inversion operator (see figure 1). Given the impossibility of having a real dataset connecting the spatial distribution of hydraulic properties of aquifers and the piezometric responses, we use synthetic models in which the hydraulic parameter fields are generated geostatistically and their corresponding piezometric responses are obtained by solving the numerical groundwater equation in steady state (forward problem).

$$\mathrm{F}(\boldsymbol{m}):\begin{cases}\nabla.(T\nabla h)=q_p\delta\left(\boldsymbol{x}_p-\boldsymbol{x}\right), \\ h=h_0\ , \qquad\qquad \Gamma_{\mathrm{d}}\end{cases}, \tag{2}$$

where $\boldsymbol{m}$ is the negative of logarithm of transmissivity $T=10^{-\boldsymbol{m}}$ randomly generated and distributed on the studied domain. h is the hydraulic head responses due to water extraction represented by a point source term $q_p\delta\left(\boldsymbol{x}_p-\boldsymbol{x}\right)$ defined at pumping well locations $\boldsymbol{x}_p$ (δ is Dirac delta function). The pumping tests will be done sequentially on several wells. The forward problem is solved by finite element technique using the Comsol software with a constant hydraulic level h_0 imposed as Dirichlet boundary at all limit of the domain Γ_{d} that are positioned far from the studied area for reducing their impact.



147

148 **Figure 1**: A sketch describing the concept of hydraulic tomography with the CNN-SegNet. This
149 concept is based on the prediction of inverse operator $\hat{F}^{-1}(\mathbf{h}_{obs},\Upsilon)$ by the CNN-SegNet
150 algorithm using a training dataset in which the hydraulic transmissivity fields are generated
151 geostatistically and the hydraulic head responses due to pumping tests are obtained numerically
152 by solving the groundwater flow equation (Forward problem $F(\boldsymbol{m})$).

153 As the nature of the training models conditions the result of the CNN-HT inversion, we have to
154 use the prior information on the studied field to establish a cluster of models for training and
155 targeting a type of solution with certain characteristics. In this paper, prior information is
156 expressed in statistical terms, with a Gaussian distribution wherein the mean and covariance,
157 are held constants during model generation. Once the parameters of the distribution are defined,
158 we launch the generation of thousands of hydraulic transmissivity fields using the SGeMS code
159 implemented in Matlab (Remy et al., 2009). The generated transmissivity fields are then
160 assigned to a domain of [40 m × 40 m] on which we set up a dense distribution of wells to better
161 cover the heterogeneity of the field. On this configuration, we conduct 5 pumping tests in such
162 a way to hydraulically disturb the whole area. Each pumping test allows us to retrieve 44
163 measurements from the remaining observation wells; in the end, we obtain 220 measurements
164 for each field (see figure 2).

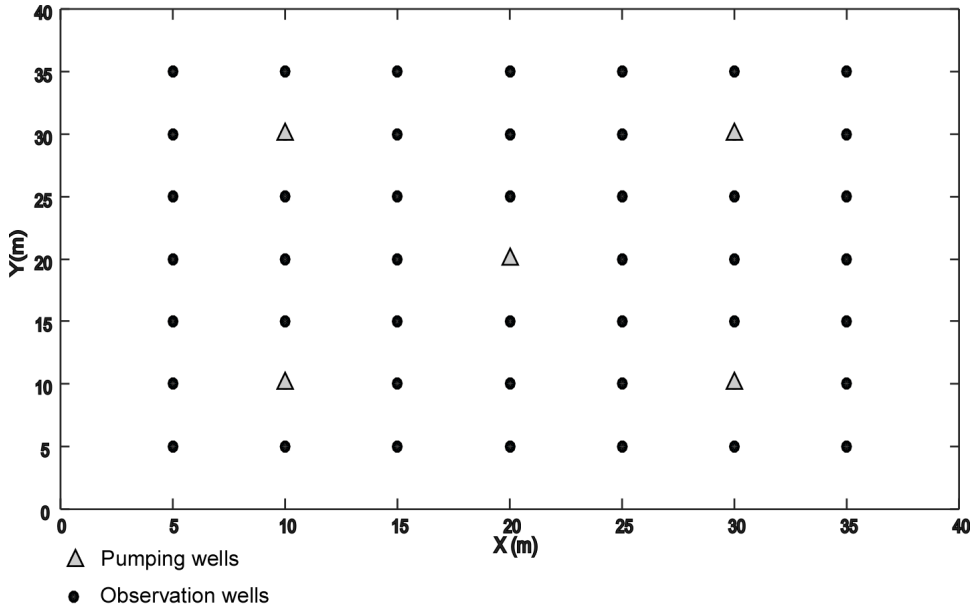


Figure 2: Position of the wells used in the generation of the hydraulic head data. 5 pumping tests were carried out successively in the wells marked with a triangle and the responses of the hydraulic head were collected in 44 wells with the dot symbol. This configuration provides 220 hydraulic data that are then used in the reconstruction of the hydraulic transmissivity.

To predict an inversion operator using convolutional neural networks with the encoder-decoder architecture, we need to reshape the piezometric data into a field with the same dimension than the hydraulic transmissivity field in order to link tow images with the same size. To do this, we can use one of two methods:

(i) The first consists in interpolating the piezometric data recorded on the observation wells for each pumping test in order to form images with the same resolution in the target transmissivity field. In this case, the input data are composed of a number of maps equal to the number of pumping tests. In our example, we have 5 pumping tests and we will then obtain 5 maps that will be treated jointly as 5 channels in convolutional neural networks.

(ii) This second is more compact than the first in which all piezometric data set recorded for different pumping tests are introduced by a single map using a projection operation derived from the Gauss-Newton formulation. This formulation results from the minimization of the

183 objective function presented in the introduction section ($\frac{\partial \Psi(\boldsymbol{m})}{\partial \boldsymbol{m}} = 0$) (Tarantola & Valette,

184 1982):

185 $$\boldsymbol{m}_{i+1} = \boldsymbol{m}_i + \left[\mathbf{J}_i^T \mathbf{C}_d^{-1} \mathbf{J}_i + \mathbf{C}_m^{-1} \right]^{-1} \left[\mathbf{J}_i^T \mathbf{C}_d^{-1} \left(\mathbf{h}_{obs} - F(\boldsymbol{m}_i) \right) + \mathbf{C}_m^{-1} \left(\boldsymbol{m}_i - \boldsymbol{m}_0 \right) \right] , \quad (3)$$

186 where $\mathbf{J}_i$ is Jacobian matrix with a size of $N \times M$, (N and M are number of the data and the

187 unknown parameters respectively). At the first iteration, usually we start with the prior model

188 then:

189 $$\boldsymbol{m}_1 = \boldsymbol{m}_0 + \left[\mathbf{J}_0^T \mathbf{C}_d^{-1} \mathbf{J}_0 + \mathbf{C}_m^{-1} \right]^{-1} \mathbf{J}_0^T \mathbf{C}_d^{-1} \left(\mathbf{h}_{obs} - F(\boldsymbol{m}_0) \right), \quad (4)$$

190 and this equation can be reformulated under this following form:

191 $$\boldsymbol{m}_1 = \boldsymbol{\beta} + \boldsymbol{\Lambda} \mathbf{h}_{obs} , \quad (5)$$

192 where $$\boldsymbol{\beta} = \boldsymbol{m}_0 - \left[\mathbf{J}_0^T \mathbf{C}_d^{-1} \mathbf{J}_0 + \mathbf{C}_m^{-1} \right]^{-1} \mathbf{J}_0^T \mathbf{C}_d^{-1} F(\boldsymbol{m}_0) , \quad (6)$$

193 $$\text{and } \boldsymbol{\Lambda} = \left[\mathbf{J}_0^T \mathbf{C}_d^{-1} \mathbf{J}_0 + \mathbf{C}_m^{-1} \right]^{-1} \mathbf{J}_0^T \mathbf{C}_d^{-1} , \quad (7)$$

194 $\boldsymbol{m}_1$ is the hydraulic transmissivity model resulting from the first iteration; obviously, it does not

195 fit the observed data, but it can be applied as a projection term for all piezometric data in order

196 to convert them into matrices with same size as the transmissivity fields. $\mathbf{J}_0$ is the sensitivity

197 matrix of the prior homogenous model $\boldsymbol{m}_0$ representing the mean of the fields. In this case, the

198 Jacobian can be derived analytically using the Thiem's equation for confined aquifers to express

199 the radial flow field and its adjoint state operator (for details, see Appendix).

To further simplify this projection term so that it is less dependent on prior information, we remove the bias term $\boldsymbol{\beta}$, replace the covariance matrices $\mathbf{C}_m$ and $\mathbf{C}_d$ by simple identity matrices, now the new projection term takes this form:

$$\mathbf{h}_{obs}^{p} = \boldsymbol{\Lambda}\mathbf{h}_{obs} \text{ With } \boldsymbol{\Lambda} = \left[\mathbf{J}_0^T\mathbf{J}_0 + \mathbf{I}_m \right]^{-1} \mathbf{J}_0^T. \quad (8)$$

Thanks to this simple multiplication, the hydraulic data $\mathbf{h}_{obs}$ will be transformed into a vector $\mathbf{h}_{obs}^{p}$ of the same size as the output hydraulic transmissivity field and used in matrix form as the input field in the network. This formulation allows to link linearly $\mathbf{h}_{obs}$ and $\mathbf{h}_{obs}^{p}$ and preserves all the hydraulic information to establish a relationship between $\mathbf{h}_{obs}^{p}$ and transmissivity field.

The comparison of the results of the two methods (linear interpolation and projection) in terms of inverse operator prediction is presented in the last section. However, the first sections all applications are done with the projection method.

2.2. Inversion by Convolutional Neural Networks with encoder-decoder structure

In this article, we will explore for the first time the efficiency of SegNet architecture to process the inversion in hydraulic tomography. The approach uses the Conventional Neural Networks concept built with an encoder-decoder architecture (see figure 3). This approach was originally conceived to process semantic segmentation of images by delineating the shapes of learned objects (Badrinarayanan, et al., 2017).

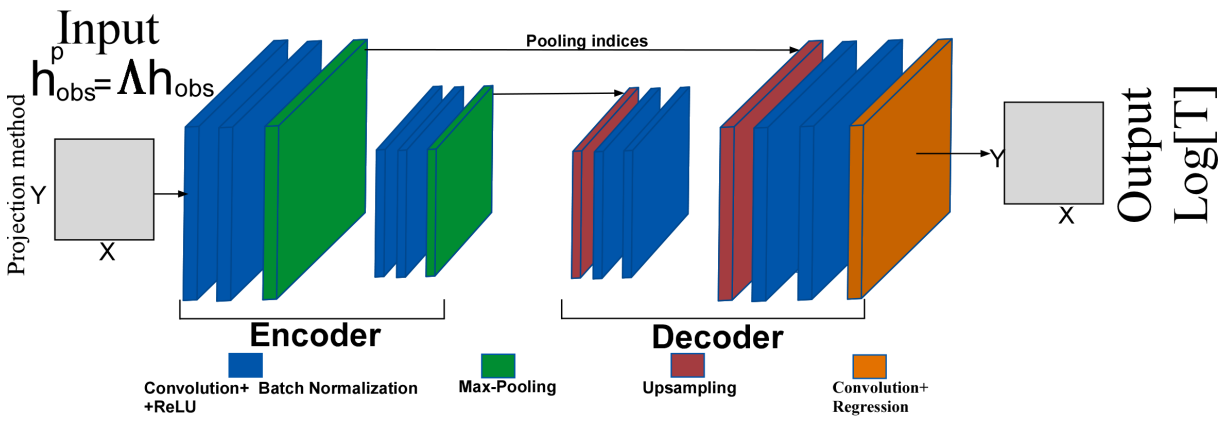


Figure 3: SegNet Convolutional Neural Networks with encoder-decoder structure. *Modified from* (Badrinarayanan, et al., 2017). Here, the piezometric data can be resized to the same dimension of the output field using the projection method

In the encoder network, the input image is processed sequentially by a set of encoders. In our case, we use only 2 encoders chosen after an analysis of the inversion results obtained with networks having 1, 2 and 3 encoders. The network with 2 encoders gave the best result in terms of training and generalization with a low sensitivity to overfitting. Each encoder starts with a convolutional process in which multiple filters with the small size (3×3×64) are convoluted on the input layer zone by zone to identify their features (Badrinarayanan, et al., 2017). This calculus permits to establish a local connectivity between pixels with these shared small filters, which remain simple to predict in the training process. The convolutional operations are followed by Batch normalization which is a linear operation that plays the role of a regularizer in scaling the convolution outputs in order facilitate and accelerate the training (Ioffe and Szegedy, 2015). This step is followed by the application of the ReLU function, which introduces the nonlinearity in the process. Then, the dimensionality of the result of the previous sequence will be considerably reduced in the max-pooling layer with stride of 2 and 2× 2 windows to keep only the main features. Therefore, the construction of feature maps in the encoder network leads to a decrease in the spatial resolution that will be restored in the decoder network. Indeed, the decoder network is configured in a symmetrical form with 2 decoders to

prepare the output by recovering the resolution lost in the decoder with up-sampling operations from the max-pooling indices obtained in the previous max-pooling layers carried out in each encoder (Badrinarayanan, et al., 2017). These results of up-sampling will be convoluted with other filters to build feature maps with more details. Then the Batch normalization and ReLU are applied successively on the convolutional outputs as in the encoder network. At the end of the encoder network, we add a convolution layer having the same size of the output and a regression layer to assess the performance of the training. This last layer replaces the Softmax layer used in the original version of SegNet network for the segmentation task. The set of weights and biases Υ used in the various encoder-decoder operations are chosen in way to match the training data $\left\{\mathbf{m}_{\mathrm{i}}, \mathbf{h}^{\mathrm{p}}_{\mathrm{obs},\mathrm{i}}\right\}$ by minimizing this objective function:

$$\Upsilon = \arg\min\left\{\sum_{\mathrm{i}=1}^{Nt}\left\|\boldsymbol{m}_{\mathrm{i}} - \hat{\mathrm{F}}^{-1}(\mathbf{h}^{\mathrm{p}}_{\mathrm{obs},\mathrm{i}}, \Upsilon)\right\|\right\}, \tag{9}$$

where i is the sample index of the training data set with Nt as size of the training data , $\hat{\mathrm{F}}^{-1}(\mathbf{h}^{\mathrm{p}}_{\mathrm{obs},\mathrm{i}}, \Upsilon)$ denotes the inverse function to identify by using the training data in the prediction of the networks parameters Υ with ADAM optimization algorithm (Kingma and Ba, 2014). This optimizer was run with a batch size of 70 and a learning rate starting at 0.01 and decreasing by 0.1 every 50 epochs. The computation was performed on a workstation with *(Intel(R) Xeon(R) Silver 4110 CPU @ 2.10GHz)* and 128 G of RAM with a single GPU. After training is complete, we use a sample of generated data not used during training phase to check the quality of the inversion results by the CNN-HT code.

3. Applications to synthetic cases

In this section, we apply the inversion concept with CNN encoder-decoder networks on theoretical cases where the log of hydraulic transmissivity fields are generated randomly with

a Gaussian variogram and their hydraulic head responses by the groundwater equation as explained in the data preparation section. This distribution has $\mu = 10^{-5}$ (m^2/s) as mean and an isotropic Gaussian variogram with 0.5 and 10 m as variance and range respectively. With this variogram model, we performed 11000 realizations in which the transmissivity is ranging over $[10^{-8.3}, 10^{-1.6}]$ (m^2/s); however 95% of these models have transmissivities between $10^{-6.4}$ and $10^{-3.5}$ (m^2/s) (see figure 4). Then, these models are used in a forward operator that takes 2s per simulation to get the hydraulic head responses. These realizations aim to evaluate the effectiveness of the CNN-SegNet approach in the reconstruction of hydraulic transmissivity field and to analyze the sensitivity of this reconstruction to the size of the training dataset and the amount and uncertainties of the hydraulic head data collected during the pumping tests.

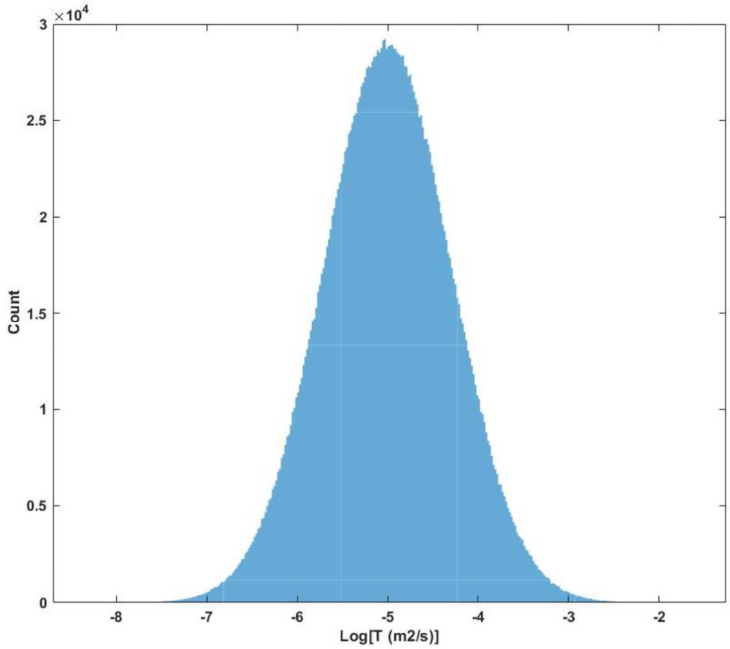
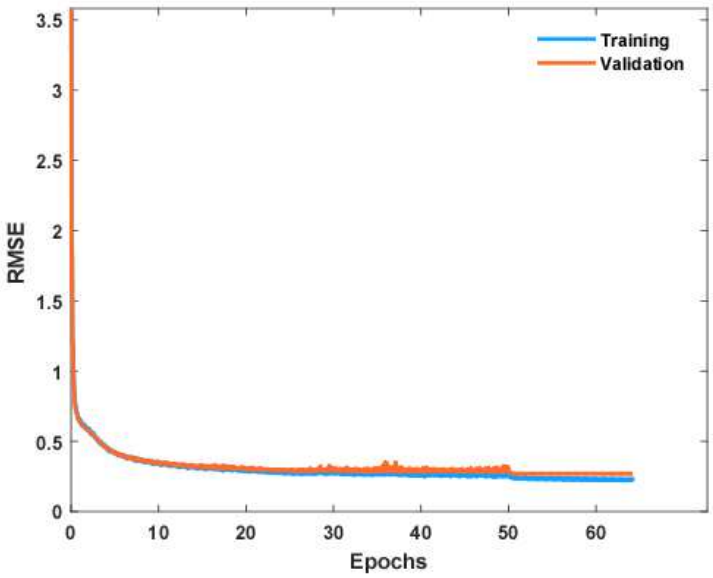


Figure 4: Gaussian distribution of the log of transmissivity fields used in the training of CNN-SegNet network.

3.1. Effect of training data size

In the first analysis, we examine the impact of training data size on the reliability of the inverse operator prediction and determine the amount of data required to obtain a satisfactory prediction. For this purpose, we performed three assessments with different training data sizes: T1=1000, T2=3000, T3=10000, the validation was done with 1000 models and the test with 500 models different from those used in the learning phase (training and validation). During the optimization process, we followed the quality of the minimization of loss functions with the training and validation data until they reached optimal values to stop the process manually or wait until the maximum number of epochs was met (see figure 5). In this way, we avoid the problem of overfitting which appears especially when we tried to build the network with a limited number of training data (T1=1000).

The results obtained with these different sizes of training data are presented in the Table 1a, in which we can easily identify through the analysis of the correlation coefficient between the real and predicted models of test sample that the accuracy of the prediction improved with increasing training data size. Indeed, the reliability of the predictive model derived from a deep learning algorithm is highly dependent on the amount of data used in the training, in which a large amount of data guarantees a good prediction. This implies a long computational time in the generation of training data, in any case the degree of accuracy desired remains linked to the nature of the problem studied.

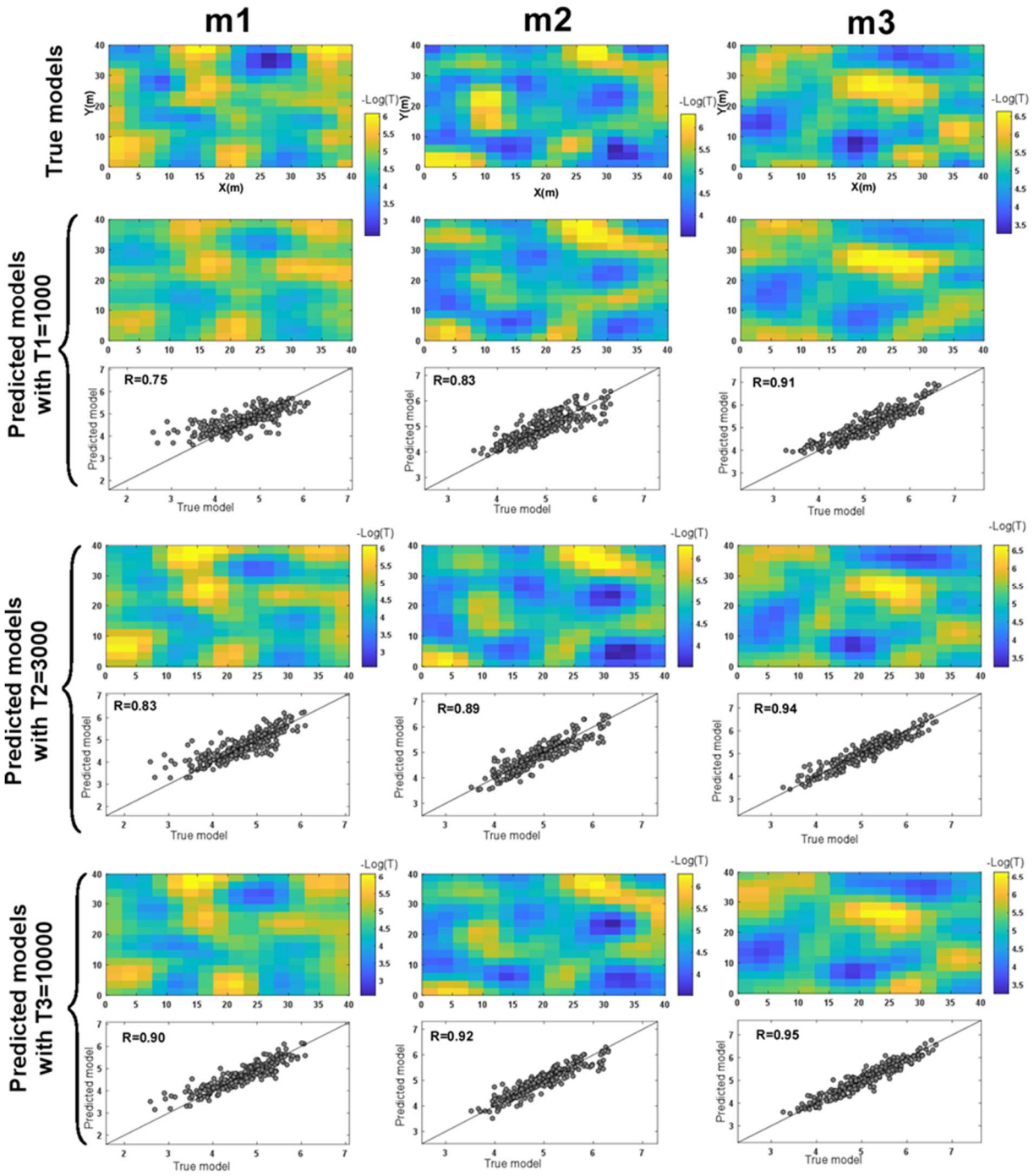


293

294 **Figure 5**: Both curves represent the quality of predictions of the network on validation and
295 training data (10000 models for training, 1000 models for validation) during the optimization
296 process.(RMSE: Root Mean Square Error).

297 In our case, training with T_1=1000 data allowed us to predict 500 test models with a
298 correlation coefficient R between real and predicted models ranging from 0.74 to 0.93 and a
299 mean of 0.86 (see Table 1a). From these models, we choose 3 models: the first one (**m1**) with
300 a relative poor prediction (R=0.75), a second model (**m2**) (R=0.83) close to the mean
301 representing quality of the majority of predicted models, and the last one (**m3**) with a good
302 reconstruction (R=0.91) (see figure 6). These models will be used to analyze the impact of
303 training data size on the quality of their estimations (see figure 6). The first model **m1** is
304 characterized by a high degree of variability with the small heterogeneities at the boundary of
305 the domain and a logarithm of the hydraulic transmissivity that varies between [-6 , -2.5]. Over
306 this range, there are outliers in the Gaussian distribution of training data in particular which are
307 around -2.5. Thus, in this case only a few models have been seen in the training with this feature,
308 which explains the poor reconstruction (see figure 6). The same problem occurs in the
309 prediction of models with low outliers < -6.5. However, the models having variabilities well
310 covered by training models such as **m2** and **m3** with transmissivities values between [- 6.5, -

311 3.5], the reconstructions are more accurate. However, if the number of training models
312 increases, the models with outliers will be more frequent and will be reconsidered in the
313 generalization as shown in the table, where the reconstructions of these models have been well
314 improved.



315

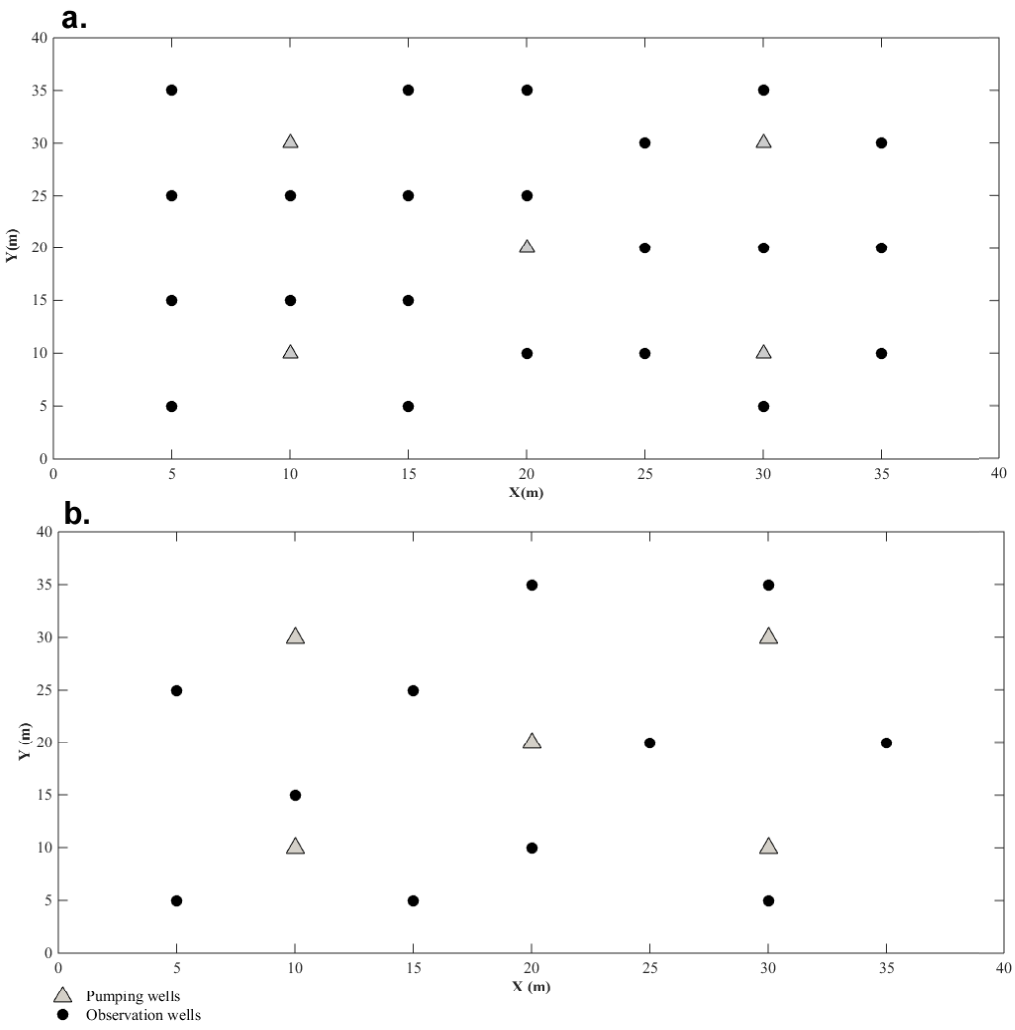
316 **Figure 6**: Three models selected from test models to illustrate the prediction quality of the
317 CNN-HT tool by comparing their true and predicted transmissivities. In this figure, the

318 predictions of these models were made with three different amounts of training data, in which
319 a significant improvement in the reconstructions was obtained when the amount of training data
320 becomes important.

321

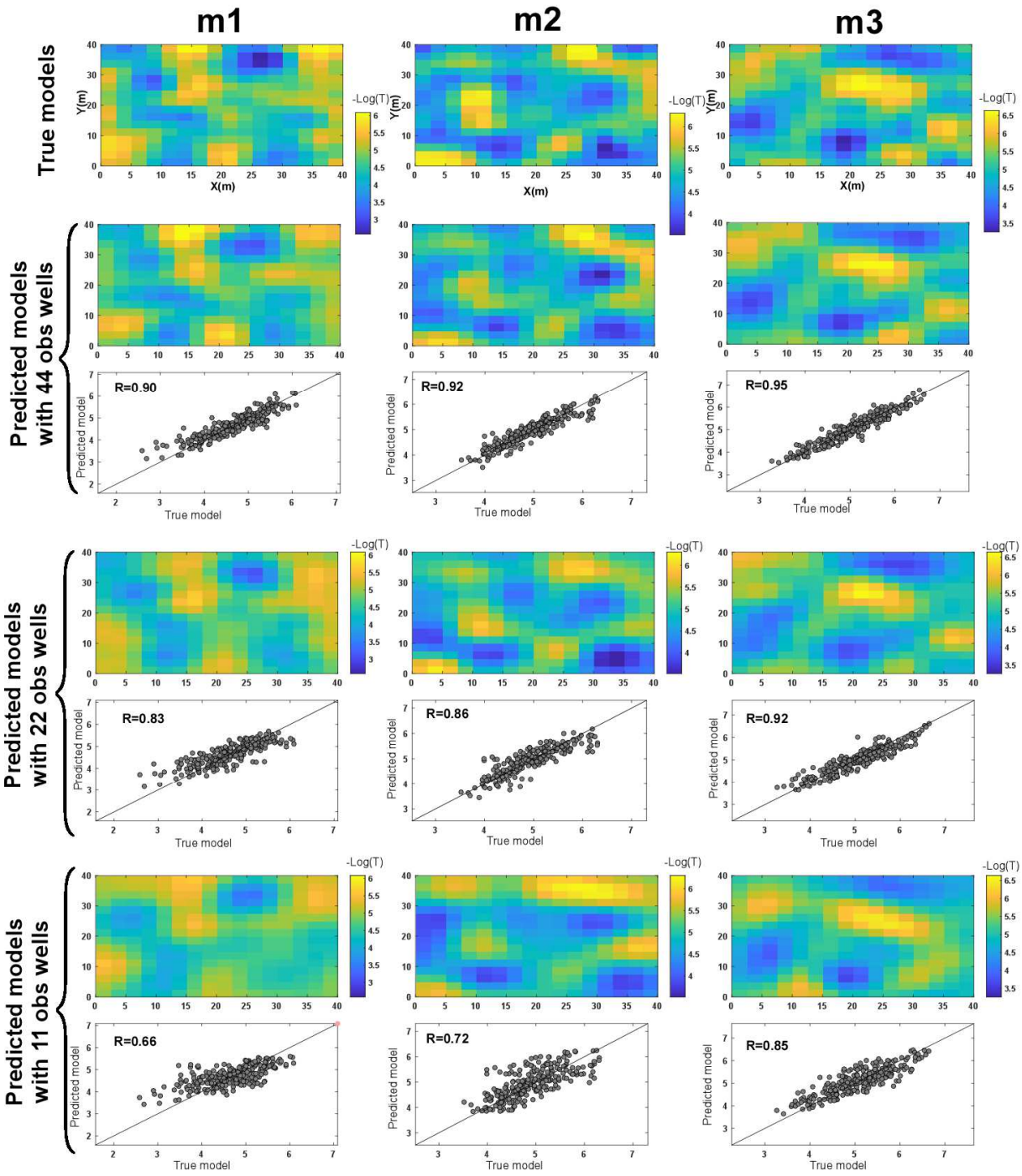
3.2 Effect of amount of observation wells

In the second analysis, we study the influence of the number of hydraulic data collected during the pumping tests on the efficiency of the network to approximate the inverse operator. In this analysis, we first reduce the number of observation wells to half (22 observation wells) and in the second step to one quarter (11 observations wells). However, we maintain the number of pumping tests and the size of the spatial discretization of the transmissivity fields (16 ×16), as used in previous test. These new well configurations are illustrated in the figure 7. In practice, only the hydraulic responses of abandoned wells are removed from previous training data. Thus, these data acquired with the two new observation well configurations are used as training input to determine the corresponding inverse operators, and the results obtained were compared to the previous configuration with 44 wells. The training operations were performed with 10,000 models, and the tests were also done on the previous unseen samples. The reconstruction results of the test models with different series of piezometric data expressed in terms of correlation coefficients reveal a degradation of the accuracy of prediction with the decrease in the number of hydraulic data (see Table 1b). Because the number of piezometers becomes insufficient to cover all the heterogeneities and the complexities of the medium, and this also occurs with classical inversion methods. This is further confirmed by analyzing the spatial aspect of the reconstructions of models (**m1, m2 and m3**), in which we observe a loss of resolution and an increase in smoothness degrees with the reduction of hydraulic data (figure 8). However, the main heterogeneities can still be identified with less data.



342

343 **Figure 7**: Two new observation well configurations were retained to study the effect of the
344 reduction in hydraulic data on the quality of the prediction. In the first configuration, the number
345 of observation wells was reduced by half (22 observation wells) and in the second, by a quarter
346 (11 observation wells) from the initial configuration shown in Figure 2.



347

348 **Figure 8**: Predictions of the three models are provided with a set of training data
349 (T3=10000) in which piezometric responses are collected with 3 different well configurations
350 (44, 22 and 11 wells). These predictions showed a degradation in the reconstructions of
351 hydraulic transmissivity with a decrease in the number of wells used in the survey.

352

353

354

355 **Table 1** Summarizing metric evaluations of the predictions quality (in term of correlation
356 coefficient) of 500 test models with correlation coefficient. **a.** Evaluation of the impact of the
357 size of training data (T) on predictions. **b.** Evaluation of the impact of the number of observation
358 wells on predictions. **c.** Evaluation of the effect of uncertainties in the piezometric data on
359 predictions with incorporation or not of these uncertainties in the learning process.

360

a. Size of training data	**Correlation coefficient R**		
	min	**mean**	**max**
T1=1000	0.74	0.86	0.93
T2=3000	0.77	0.90	0.95
T3=10000	0.84	0.92	0.96
b. Number of observation data	**Correlation coefficient R**		
	min	**mean**	**max**
T3=10000, N° data=220	0.84	0.92	0.96
T3=10000 N° data=110	0.73	0.85	0.93
T3=10000, N° data=55	0.57	0.77	0.91
c. Degree of noise on the data	**Correlation coefficient R**		
	min	**mean**	**max**
T3=10000, N° data=220, **std**=0 free noise	0.84	0.92	0.96
T3=10000, N° data=220, **std**=0.5 on Training and Test data	0.55	0.84	0.93
T3=10000, N° data=220, **std**=0.5 only on Test data	0.49	0.79	0.92

361

362 ### 3.3. Effect of observation uncertainties

363 In this section, we analyze the effect of uncertainties that may be associated with
364 hydraulic pressure measurements collected during pumping tests. This involves contaminating
365 the numerical piezometric data with a random Gaussian noise of 0.5 as the standard deviation,
366 and verifying its impact on the quality of hydraulic field reconstructions of test models.

As a first analysis, we use an inverse operator established on the basis of uncontaminated training data to interpret hydraulic head data from a test set that are contaminated by noise. Here, the training operation did not take into account the noise. The result of this test is presented by the benchmark models in figure 9 and table 1c of correlation coefficients, which reveal a sharp degradation in the quality of the predictions compared to the noise-free test models, with a decrease in the correlation coefficient (from 0.92 to 0.79 mean for all test models). In the second strategy, noise is taken into account in the estimation of the inversion function by adding noise to the hydraulic data used in the learning process. This method has led to a significant improvement in the prediction quality of the test models, as shown by the three models (**m1,m2 & m3**) and the correlation coefficient noted in table 1c with the average R of the set of models increasing from 0.79 to 0.84. However, this prediction is less accurate when compared to the noise-free training and test data, as shown by the R-values. We can, therefore, conclude that uncertainties in the measured data may have an impact on the results of inversion, as is the case with classical optimization techniques, but to reduce their impacts, it is necessary to incorporate these uncertainties in the construction of the inversion model, i.e. in the learning process.

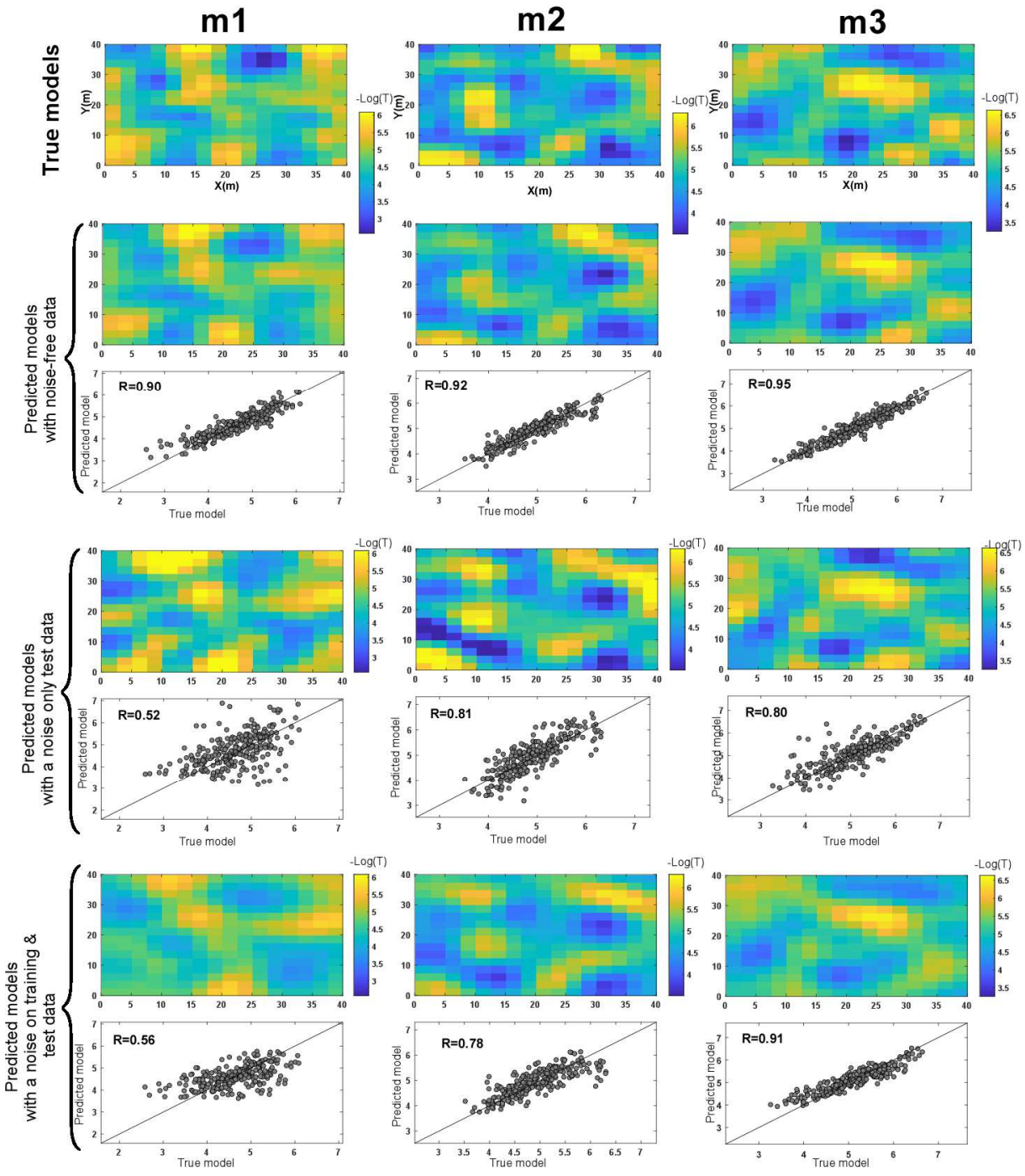


Figure 9: The purpose of these reconstructions is to analyze the effect of noisy data on the efficiency of the inversion with CNN-HT by comparing 3 types of predictions: 1) with the training and test data without noise; 2) with the training data without noise but the test models are noisy; 3) both the training and the test models are noisy. These comparisons show that the uncertainties in the measured data are affecting the quality of the reconstructions; however, their impacts could be attenuated by taking into account this noise in the training process.

4. Comparison of CNN-HT and Gauss-Newton algorithms

In this section, we compare the inversion results obtained with the CNN-HT algorithm with those determined by the traditional Gauss-Newton (GN) method. To this end, we apply both algorithms to the reconstruction of a transmissivity field with a mean of 10^{-5} m^2 s^{-1} and generated with a spherical covariance having 0.5 and 10 m as variance and range respectively. However, this time, the models are highly discretized (32×32) compared to those used in the previous sections. The Gauss-Newton algorithm, as described in the introductory section, is a deterministic approach that relies on the minimization of an objective function (Eq. 1) by iteratively computing the Jacobian matrix until obtaining a local minimum. The calculation of the Jacobian matrix remains the most demanding part of this process, especially when the number of unknowns is important. This calculation is usually done either by the finite difference or the adjoint state technique. The finite difference method is simple to implement, but its computation is time consuming because it involves to solve the forward problem for each unknown parameters (Cardiff & Kitanidis, 2008; Fischer et al., 2017). In our case study, where the model is discretized into 32×32 cells, the construction of the Jacobian will therefore involve solving the forward problem 1024 times. On the other hand, the adjoint state technique requires less computation, by iteratively solving an adjoint operator with the form of the forward problem on the number of observation wells, but its implementation is quite complex (for more details see Sykes et al., 1985; Cardiff & Kitanidis, 2008). In this section, we test the implementation of Gauss-Newton algorithm to invert the synthetic hydraulic data determined with a hydraulic transmissivity field shown in figure 10, (called True model) and using the piezometric coverage shown in figure 2. These data were then contaminated with noise with a variance of $5cm^2$. The inversion of these datasets was constrained, as explained in the introduction, by geostatistical parameters such as covariance and mean. The results of the inversions with the GN algorithm using the finite difference and adjoint state methods are

417 presented in Figure 10 where we can conclude that the GN algorithm succeeded in both cases
418 in identifying the main heterogeneities of the real field with certain smoothness. In the case
419 where the Jacobian is computed with the finite difference method, convergence was achieved
420 after 7 iterations in 4h10min and solving the forward problem 7×(1024+1) =7175 times. In
421 contrast, the adjoint state method took only 1min24s.

422 Regarding the application of the CNN-HT algorithm, we geostatistically generate 11000
423 transmissivity models with the same variogram as the real model used for the comparison. We
424 keep 10000 models for the training, and the rest of the models are equally split for validation
425 and testing. In this case study, we test two different ways of resizing the piezometric data, the
426 first method called projection method uses the Hessian matrix to reshape the input data into the
427 same size as the transmissivity field (output) as described in the section 2. This method requires
428 the calculation of a Jacobian matrix with a homogeneous field for example the mean field value.
429 The second method is based on a linear interpolation of the hydraulic data recorded during each
430 pumping test to form 5 maps of size 32×32 (i.e., one map for each pumping test), so the input
431 data has the form of a matrix of 5 channels (32×32×5).

432 Both approaches were used to build the SegNet networks and then tested on the test
433 sample. The analysis of the predictions obtained on this sample shows that both methods
434 provide almost the same prediction qualities with a small advantage for the projection method
435 according to the correlation coefficients (see table 2). However, when we reduced the training
436 size to T=500 and T=2500 models, the network using the projection method was able to provide
437 accurate generalization while the network with interpolated data failed. We believe this is due
438 to the fact that the interpolation of piezometric data does not preserve all the information
439 masking the degree of variability in hydraulic responses, which makes learning difficult with
440 few training data. In addition, using the interpolation method increases the size of the input

441 data, which requires more memory and makes the training process more cumbersome compared
442 to the projection technique that brings the data into a compact form.

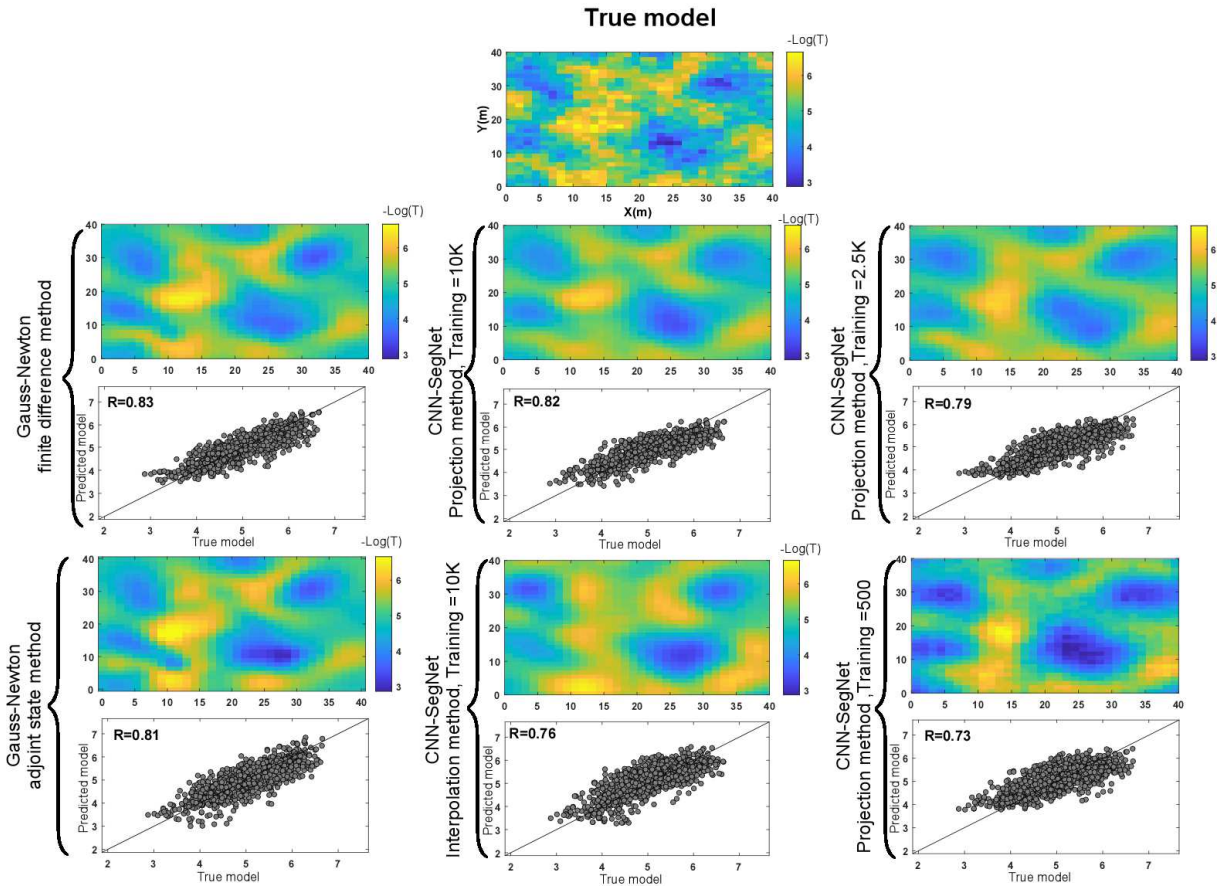
443 Regarding the application of these two networks on the hydraulic data of the real model,
444 we find that the network built with compact data provides a better reconstruction of the spatial
445 variability of transmissivity field with a correlation coefficient R=0.80 (see figure 10).
446 Moreover, this prediction has almost the same accuracy as the result determined by the Gauss-
447 Newton algorithm (R=0.83). We also mention that the time used in the GN with the finite
448 difference method is perhaps comparable to the time spent in the DL method mainly devoted
449 to the construction of the data set. In particular, the DL method becomes faster when we reduce
450 the amount of training data to T=500 and T=2500 (20 min and 1h41 min), because even with
451 these small samples, the transmissivity field can be well reconstructed well in a reasonable
452 amount of time, as demonstrated by the comparison of real and inverted fields (figure10).

453 At the end of this comparison, we conclude that both the CNN-HT and Gauss Newton
454 codes are based on the forward problem for which the computation time depends on the
455 numerical method and the degree of mesh refinement used to solve the groundwater flow
456 equation. The machine learning technique uses this forward problem to generate a training
457 database to build the CNN network. The time spent in this construction can be equivalent or
458 less to the time spent in finding a local minimum with the Gauss Newton using the finite
459 difference method to calculate the Jacobian matrix over a large number of unknown parameters.
460 However, it is difficult to have an idea in advance of the time that will be spent in building this
461 database because the size needed to get accurate results is related to the degree of complexity
462 of the input-output relationship and the type of network. The optimal size as well as the rest of
463 the network parameters (epochs, number of encoders) are only established by a trial-and-error
464 analysis, which can take a considerable amount of time. Therefore, the data size used in this
465 paper cannot be generalized to all cases treated by this network or others.

Both methods also share their dependence on prior information. In the Gauss-Newton algorithm, this prior information is introduced as a constraint to guide the optimization towards realistic models and reduce the uncertainties associated with the non-uniqueness of the solution. On the other hand, the CNN-HT algorithm uses it to generate the training models. The CNN-HT offers the possibility of incorporating multiple prior models into the construction of the training dataset, which is difficult to perform in a deterministic algorithm where the mathematical forms of these constraints must be differentiable. In addition, both methods are sensitive to the uncertainties associated with the measured data. The CNN-HT allows simultaneous learning with data affected by several degrees of uncertainty by adding noise to the training data, then combining this set with the original to double the size of the training data and take into account the impact of noise. Regarding the quantification of the prediction uncertainties, in the Gauss Newton method, we use the a posteriori covariance formulation to get an idea of the uncertainty interval concerning the local minimum (Tarantola and Valette, 1982). However, in the CNN-HT algorithm it is difficult to construct an uncertainty map of the predicted model. We can only build an uncertainty map on the model used in the test.

Table 2: We report the metric evaluations (correlation coefficient R) and computational time of inversions obtained with the CNN-HT and Gauss-Newton algorithms. The CNN-HT network was trained with hydraulic data introduced either by the linear interpolation method or by the projection method. We tested 3 different sizes of training data (500, 2500, 10000), but the network with the interpolation method failed in the training process when the data size became small (500 and 2500). We also present the evaluation of the Gauss-Newton metric using the finite difference method and adjoint state on a test model shown in figure 10.

		Computation Time	**R True Model (figure 10)**	**R Test Sample**		
				min	**mean**	**max**
CNN-HT	T1=500 Projection method	20min	0.73	0.54	0.69	0.79
	T2=2500 Projection method	1h41	0.79	0.65	0.77	0.86
	T3=10000 Projection method	6h	0.82	0.69	0.80	0.88
	T3=10000 Interpolation method	6h38min	0.76	0.68	0.80	0.89
GN	Finite Difference	4h 10 min	0.83			
	Adjoint state method	1 min 24s	0.81			



489

490 **Figure 10**: The results obtained when applying CNN-HT and Gauss-Newton algorithm
491 in the reconstruction of the "True model". The results obtained with GN using the finite
492 difference and adjoint state methods recovered the main heterogeneities of the field. The same
493 reconstruction quality was obtained with CNN-HT, especially using the projection method for
494 the preparation of the input data. Even with a small amount of data (T=500 and T=25000), this
495 CNN-HT network provided satisfactory reconstructions.

496

497 ## 5. Conclusion and Summary

498 In this work, we explore the relevance of a deep learning tool called the SegNet network
499 in the mapping of spatial variability of hydraulic parameters by interpreting the piezometric
500 data recorded during pumping tests. This approach relies on the approximation of the non-linear
501 inverse operator connecting the hydraulic head data to the hydraulic transmissivity field. This
502 approach is the outcome of an adaptation of the SegNet network that has been successfully
503 applied in the realm of image segmentation by identifying the objects studied on the pixels of
504 images. The SegNet is designed to establish a relationship between two images sharing the

same size and certain similarities, which is not the case for hydraulic head data, which are less numerous than the hydraulic transmissivity values to be estimated. To meet this point, we test the projection and interpolation methods. Projection method is based on the classical iterative formulation of a least-square method at the initial iteration to build a projection operator with the Hessian and Jacobian matrices derived from the mean value of transmissivity fields used in the training data to resize the hydraulic data into the hydraulic transmissivity field size. Interpolation method involves reshaping the data size by interpolating the hydraulic head data to form a map with the same size as the output (transmissivity field) for each pumping test. As a result, the interpolation significantly increases the size of the input data, which can slow down the numerical calculation without improving the quality of the prediction compared to the projection method, which better preserves the hydraulic information in compact form.

In the projection method, the operator is applied to the totality of the hydraulic pressure data used in training which are determined numerically as responses to pumping tests carried out on geostatistically generated hydraulic transmissivity fields with certain statistical properties. Once the training data has been gathered, the inverse operator is established by identifying the different filters used in the SegNet, which is composed of encoders and decoders networks. The encoder network allows, through multiple application of filters and sequential calculations of convolution, Batch normalization, ReLU and max-pooling operations to extract the main features of the input data. This encoder process induces a drop in spatial resolution, which will be recovered in the decoder network, which is designed to prepare the output with the applications of up-sampling, convolution, Batch normalization and ReLU operations. This network ends with a regression layer to evaluate the performance of the inverse operator approximation.

The quality of this approximation is highly dependent on the amount of data used in the training with an accurate approximation when the dataset is large. This makes the assembly of

the training data the most expensive step in time, requiring several hours compared to the learning process that in this case does not exceed 30 minutes. In practice, it is preferable to launch the learning with few data and check whether or not the desired accuracy is achieved before generating large amounts of data unnecessarily. In terms of comparison, the time spent in forming a training database can be equivalent to the time spent to reach a local minimum with Gauss Newton using the finite difference method to compute the Jacobian matrix

In addition to the amount of training data, the nature of this data also controls the prediction results of the network. Indeed, the trained network is only intended to handle certain types of models with similar characteristics to the models used in training. Thus, it is by no means a universal inversion operator, which makes the choice of the nature of the training models a crucial step that must be based on realistic prior information. In real application case, it should also be verified that the transmissivity fields used in the training allow to generate hydraulic data with the magnitude of the piezometric data observed in the field. It is also necessary to ensure that the degree of heterogeneity of the generated models can be easily mapped to the piezometric data during training process. Because if the piezometric coverage does not allow to cover transmissivity models with high heterogeneity, the learning process will fail to establish this relation. The result of the inversion with CNN-HT depends on the prior information and this particularity is also found with classical inversion techniques such as Gauss-Newton.

Noise can alter the information carried by the piezometric data and is a frequent problem in inversion algorithms. The deep learning algorithms are no exception, but its impact can be reduced by integrating it into the learning step by contaminating the input of training data. The number of wells and their spatial configurations used in the study also had a determining effect on the effectiveness of the CNN-SegNet inversion tool, with more reliable reconstructions when

554 these piezometers are well distributed over the study area, which was also observed in the use
555 of conventional deterministic and stochastic methods.

556 At the end of this conclusion, we enumerate and discuss the main advantages and
557 limitations of the proposed method. Regarding the advantages, we point out that:

558 ✓ The CNN-HT method uses the forward problem as a black box to generate the training
559 data with user-selected transmissivity fields, thus avoiding numerical instability and outliers in
560 hydraulic head simulations related to its solving. These numerical problems can be encountered
561 when using conventional inversion methods (stochastic and deterministic) if they are ill-
562 constrained.

563 ✓ Once the inverse operator is trained, the network permits to interpret the piezometric
564 data in a few seconds without the user having to add new parameters.

565 ✓ The CNN-HT code provides predictions that require any initial hydraulic transmissivity
566 models or computation of sensitivity matrix, as is the case with deterministic methods. This
567 advantage makes CNN-HT an effective inversion tool for dealing with highly nonlinear inverse
568 problems with large-scale parameterization that are difficult to handle with deterministic codes,
569 especially when the computation of sensitivity is complex.

570 ✓ The CNN-HT code performs reconstructions of the transmissivity field that depend on
571 the a priori model used in the generation of transmissivity field for the training data. This
572 dependence to the a priori model is a property shared with other inversion methods. However,
573 in this CNN-HT tool, the mathematical form of this a priori model is not subject to any
574 condition, which is not the case with deterministic methods where the model must have a
575 quadratic and differentiable form.

576

✓ The CNN-HT algorithm also provides great flexibility for the user to incorporate multiple prior models, especially when in doubt about which a priori information to use.

✓ The proposed method is based on a simple network that can be generalized and applied to handle hydraulic tomography in 3D or in time domain in order to identify the storativity coefficient as well, by only modifying the training dataset. It is also easily adaptable to deal with hydraulic tomographies with complex heterogeneities having for example multiple discontinuous hydrofacies with contrasting transmissivities, or parameterized in discrete fracture network mode. These types of problems can be treated as a classification task by replacing the regression layer of the network with the Softmax layer. In this way, the network can identify fracture geometries or hydrofacies shapes.

✓ The code allows the incorporation of uncertainties on the boundary conditions imposed in the numerical solving of the forward problem by randomly choosing their values within a confidence interval when building the training data set.

✓ This code, like other deep learning algorithms that use the convolutional neural network technique are able to better reduce the effects of noise in the data. This is thanks to the convolution concept that processes the input data with some local connectivity and not in a pointwise manner as in other conventional inversion methods (Vu and Jardani, 2021). The impact of noise could be minimized and accounted for in the inversion results by contaminating the hydraulic data used in the training data with a noise signal.

Regarding the disadvantages of this method, we mention that

✓ It is based on a repetitive and time-consuming numerical resolution of the forward problem, which is involved in the construction of the training database. The size of this database conditions the quality of the predictions, and its optimal size can only be obtained by a long trial-and-error procedure.

✓ The method requires important computing resources to carry out the learning.

602 ✓ The code also uses a large number of parameters that intervene in the construction of
603 the networks (such as: size of the decoder-encoder, number of filters and their size), and in the
604 optimization algorithm used in the training phase (number of epochs, regularization parameters,
605 initialization of filters, Batchsize, Dropout). The determination of all these parameters implies
606 a tedious sensitivity analysis with several training operations to choose the best parameters.
607 ✓ The method does not provide an uncertainty assessment of the predicted transmissivity
608 fields. This is a major weakness of this method compared to classical approaches. There are
609 some attempts in the literature to estimate the uncertainty of the prediction determined with DL
610 algorithms by performing the learning in a probabilistic way (Kendall et al., 2016; Gal and
611 Ghahramani, 2016). However, this type of approach requires a lot of computational time and it
612 does not take into account all the uncertainties that can come from all the parameters used in
613 the learning.
614 ✓ The method is highly dependent on prior information. This dependence can be an
615 obstacle to its application, especially when the network fails in the learning phase to relate the
616 transmissivity fields to their piezometric data. This can occur when the transmissivity fields
617 have complex heterogeneities that are not captured by low piezometric resolution. This
618 dependence can also lead to a misinterpretation, even with a well-trained inversion operator,
619 when the magnitudes of the piezometric data used in the training are so different from those we
620 desire to predict. To avoid both of these problems, it is necessary to ensure that generated
621 transmissivity fields have a degree of heterogeneity that can be mapped with the piezometer
622 coverage and that they produce hydraulic head data that include the magnitude of the hydraulic
623 observations to be interpreted.
624 In conventional inversion methods, the predicted transmissivity distribution is the result of a
625 compromise between the a priori model and the piezometric data. However, here, the result is

626 strongly linked to the a priori model and the piezometric data to be interpreted come only at the
627 end of the learning process.

628 .

629

630 **Acknowledgements :** We thank Normandy region for its financial support to our consortium
631 on the hydraulic characterization of aquifers.

632 Appendix

633 Here we present the mathematical formulation to analytically calculate the Jacobian matrix of
634 a homogeneous transmissivity field. This matrix is used in the projection method to transform
635 the hydraulic data to the same size as the transmissivity field to form the Segnet network.

636 We recall that in the joint state method, the sensitivity matrix can be expressed as an integral
637 (Sykes et al.,1985):

638 $$\frac{\partial h_i}{\partial m_j} = -\log 10 \iint_{\Omega_j} \left[T \nabla h \nabla \phi \right] \mathrm{d}\Omega \quad ; \qquad \text{A1}$$

639 Where h is the hydraulic head measured at location $\mathrm{O}(x_{\mathrm{obs}}, y_{\mathrm{obs}})$ due to a pumping source

640 placed at $\mathrm{P}(x_{\mathrm{p}}, y_{\mathrm{p}})$ with a rate q_p . ϕ is the adjoint operator with a source term $q_{adj} = -1$

641 placed at the observation point $\mathrm{O}(x_{\mathrm{obs}}, y_{\mathrm{obs}})$. T is transmissivity hydraulic with m=-log(T).

642 When the transmissivity is constant, h and ϕ can be determined analytically with Thiem's
643 equation for a confined aquifer

644 $$h = h_0 + \left(\frac{q_p}{2\pi T} \right) . \ln \left(\frac{R}{r_p} \right) \text{ and } \phi = \phi_0 + \left(\frac{q_{adj}}{2\pi T} \right) . \ln \left(\frac{R}{r_{obs}} \right) ; \qquad \text{A2}$$

645 with $r_{\text{obs}} = \sqrt{(x_{\text{obs}} - x)^2 + (y_{\text{obs}} - y)^2}$ and $r_{\text{p}} = \sqrt{(x_{\text{p}} - x)^2 + (y_{\text{p}} - y)^2}$

646 h0 and ϕ_0 are the values imposed as the boundary of the domain located at a distance R, away
647 from the study area. *r* is radial distance. Then the sensitivity of hydraulic head at point O to a
648 variation of the transmissivity at cell Ω_j can be approximated with this integral:

649 $$\frac{\partial h_i}{\partial m_j} = -\log(10)\frac{q_p\, q_{adj}}{(4\pi^2 T)}\iint_{\Omega_j}\frac{(x_{obs}-x)(x_p-x)+(y_{obs}-y)(y_p-y)}{r_{obs}^2 .\, r_p^2}\mathrm{d}\Omega \quad \text{A3.}$$

650

651 **References**:

652 Alcolea A., Castro E., Barbieri M., Carrera J., Bea S., 2007. Inverse modeling of coastal
653 aquifers using tidal response and hydraulic tests, *Groundwater*, 45(6), 711-722.

654 Badrinarayanan V., Kendall A., Cipolla R., 2017. Segnet: A deep convolutional encoder-
655 decoder architecture for image segmentation, *IEEE transactions on pattern analysis and*
656 *machine intelligence*, 39(12), 2481-2495.

657 Berg S.J., Illman W.A., 2011. Three-dimensional transient hydraulic tomography in a highly
658 heterogeneous glaciofluvial aquifer-aquitard system. *Water Resources Research*, 47, W10507.

659 Bohling G.C., Zhan X., Butler J.J.J., Zheng L., 2002. Steady shape analysis of tomographic
660 pumping tests for characterization of aquifer heterogeneities, *Water Resources Research*,
661 38(12), 1-15.

662 Cardiff M., Kitanidis P., 2008. Efficient solution of nonlinear, underdetermined inverse
663 problems with a generalized PDE model. *Computers & Geosciences*, 34(11), 1480-1491.

664 Cardiff M., Barrash W., Kitanidis P., 2013. Hydraulic conductivity imaging from 3-D transient
665 hydraulic tomography at several pumping/observation densities, *Water Resources Research*,
666 49(11), 7311-7326.

667 Castagna M., Bellin A., 2009. A Bayesian approach for inversion of hydraulic tomographic
668 data, *Water Resources Research*, 45(4), W04410.

669 Elsheikh A.H., Jackson M.D., Laforce T.C., 2012. Bayesian Reservoir History Matching
670 Considering Model and Parameter Uncertainties, *Mathematical Geosciences*, 44, 515-543.

671 Fernández-Martínez J.L., Mukerji T., García-Gonzalo E., Fernández-Muñiz Z., 2011.
672 Uncertainty assessment for inverse problems in high dimensional spaces using particle swarm
673 optimization and model reduction techniques. *Mathematical and Computer Modelling*, 54(11-
674 12), 2889-2899.

675 Fischer P., Jardani A., Soueid Ahmed A., Abbas M., Wang X., Jourde H., Lecoq N., 2017.
676 Application of large-scale inversion algorithms to hydraulic tomography in an alluvial aquifer.
677 *Groundwater*, 55(2), 208-218.

678 Fischer P., Jardani A., Lecoq N., 2018. Hydraulic tomography of discrete networks of conduits
679 and fractures in a karstic aquifer by using a deterministic inversion algorithm, *Advances in*
680 *Water Resources*, 112, 83-94.

681 Fu J., Gómez-Hernández J., 2009a. A blocking Markov Chain Monte Carlo method for inverse
682 stochastic hydrogeological modeling. *Mathematical Geosciences*, 41, 105–128.

683 Fu J., Gómez-Hernández J., 2009b. Uncertainty assessment and data worth in groundwater flow
684 and mass transport modeling using a blocking Markov chain Monte Carlo method. *Journal of*
685 *Hydrology*, 364(3-4), 328-341.

686 Gal Y., Ghahramani Z., 2016. Dropout as a Bayesian approximation: Representing model
687 uncertainty in deep learning. Proceedings of the 33rd International Conference on Machine
688 Learning.

689 Gernez S., Bouchedda A., Gloaguen E., Paradis D., 2019. Comparison between hydraulic
690 conductivity anisotropy and electrical resistivity anisotropy from tomography inverse
691 modeling, *Frontiers in Environmental Science*, 7, 67.

692 Gottlieb J., Dietrich P., 1995. Identification of the permeability distribution in soil by hydraulic
693 tomography, *Inverse Problems*, 11(2), 353–360.

694 Indolia S., Goswami A.-K., Mishra S.P., Asopa P., 2018. Conceptual understanding of
695 convolutional neural network - a deep learning approach, *Procedia computer science*, 132, 679-
696 688.

Ioffe S., Szegedy C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift, Proceedings of the 32nd International Conference on Machine Learning, 37, 448-456.

Jimenez S., Mariethoz G., Brauchler R., Bayer P., 2016. Smart pilot points using reversible-jump Markov-chain Monte Carlo, *Water Resources Research*, 52, 3966-3983.

Jardani A., Dupont J.-P., Revil A., Massei N., Fournier M., Laignel B., 2012. Geostatistical inverse modeling of the transmissivity field of a heterogeneous alluvial aquifer under tidal influence, *Journal of Hydrology*, 472, 287-300.

Kendall A., Cipolla R., 2016. Modelling uncertainty in deep learning for camera relocalization, 2016 IEEE international conference on Robotics and Automation (ICRA), 4762-4769.

Kingma D.P., Ba J., 2014. Adam: A method for stochastic optimization, International Conference on Learning Representations (ICLR).

Kitanidis P.K., 1995. Quasilinear geostatistical theory for inversing, *Water Resources Research*, 31(10), 2411-2419.

Kitanidis P.K, 1997. *Introduction to geostatistics: applications in hydrogeology*, Cambridge University Press.

Laloy E., Hérault R., Jacques D., Linde N., 2018. Training-image based geostatistical inversion using a spatial generative adversarial neural network, *Water Resources Research*, 54(1), 381-406.

LeCun Y., Bottou L., Bengio Y., Haffner P., 1998. Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86(11), 2278-2324.

Li W., Nowak W., Cirpka O., 2005. A Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown, *Water Resources Research*, 41(8).

Lochbühler T., Doetsch J., Brauchler R., Linde N., 2013. Structure-coupled joint inversion of geophysical and hydrological data, *Geophysics*, 78(3).

Neuman S.P., 1987. Stochastic continuum representation of fractured rock permeability as an alternative to the R.E.V. and fracture network concepts, *Rock Mechanics: Proceedings of the 28th*, 533-561.

Oliver D.S., Cunha L.B., Reynolds, A.C., 1997. Markov chain Monte Carlo methods for conditioning a permeability field to pressure data, *Mathematical Geology*, 29, 61-91.

Puzyrev V., Swidinsky A., 2021. Inversion of 1D frequency-and time-domain electromagnetic data with convolutional neural networks, *Computers & Geosciences*, 149, 104681.

Remy N., Boucher A., Wu J., 2009. *Applied Geostatistics with SGeMS: A User's Guide*, Cambridge University Press.

Rizzo D.M., Dougherty D.E., 1994, Characterization of aquifer properties using artificial neural networks: Neural kriging, *Water Resources Research*, 30(2), 483-497.

Scales J.A., Smith M.L., Fischer T.L., 1992. Global optimization methods for multimodal inverse problems, *Journal of Computational Physics*, 103(2), 258-268.

Shen C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists, *Water Resources Research*, 54(11), 8558-8593.

Shigidi A., Garcia L.A., 2003. Parameter estimation in groundwater hydrology using artificial neural networks, *Journal of Computing in Civil Engineering*, 17(4), 281-289.

Soueid A., Jardani A., Revil A., Dupont J.-P., 2016. Specific storage and hydraulic conductivity tomography through the joint inversion of hydraulic heads and self-potential data, *Advances in Water Resources*, 89, 80-90.

Sun A.Y., 2018. Discovering State-Parameter Mappings in Subsurface Models Using Generative Adversarial Networks, *Geophysical Research Letters*, 45(20), 11137-11146.

Sykes J.F., Wilson J.L., Andrews R.W., 1985. Sensitivity analysis for steady state groundwater flow using adjoint operators, *Water Resources Research*, 21(3), 359-371.

Tarantola A., Valette B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Reviews of Geophysics*, 20(2), 219-232.

Tso C.H.M., Za Y., Yeh T.C.J., Wen J.C., 2016. The relative importance of head, flux, and prior information in hydraulic tomography analysis, *Water Resources Research*, 52(1), 3-20.

Vu M.T, Jardani A, 2021. Convolutional neural networks with SegNet architecture applied to three-dimensional tomography of subsurface electrical resistivity: CNN-3D-ERT, *Geophysical Journal International*, 225 (2), 1319–1331, https://doi.org/10.1093/gji/ggab024

753 Wang X., Jardani A., Jourde H., 2017. A hybrid inverse method for hydraulic tomography in
754 fractured and karstic media, *Journal of Hydrology*, 551, 29-46.

755 Wu Y., Lin Y., 2018. InversionNet: A Real-Time and Accurate Full Waveform Inversion with
756 CNNs and continuous CRFs, arXiv preprint arXiv:1811.07875.

757 Yeh T.C.J., Liu S., 2000. Hydraulic tomography: Development of a new aquifer test
758 method. *Water Resources Research*, *36*(8), 2095-2105.

759 Zhao Z., Illman W.A., Berg S.J., 2016. On the importance of geological data for hydraulic
760 tomography analysis: laboratory sandbox study, *Journal of Hydrology*, 542, 156-171.

761 Zhu Y., Zabaras N., 2018, Bayesian deep convolutional encoder–decoder networks for
762 surrogate modeling and uncertainty quantification, *Journal of Computational Physics*, 366,
763 415-447.

764 Zio E., 1997, Approaching the inverse problem of parameter estimation in groundwater models
765 by means of artificial neural networks, *Progress in Nuclear Energy*, 31(3), 303-315.

766