Estimation of groundwater storage from seismic data using deep learning

Timo Lähivaara[a], Antti Pasanen[b], Leo Kärkkäinen[c,d],
Janne M.J. Huttunen[c], Jan S. Hesthaven[e], Alireza Malehmir[f]

[a]Department of Applied Physics, University of Eastern Finland, Kuopio, Finland
[b]Geological Survey of Finland, Kuopio, Finland
[c]Nokia Bell Labs, Espoo, Finland
[d]Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland
[e]Computational Mathematics and Simulation Science, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[f]Department of Earth Sciences, Uppsala University, Uppsala, Sweden

Abstract

We investigate the feasibility of the use of convolutional neural networks to estimate the amount of groundwater stored in the aquifer and delineate water-table level from active-source seismic data. The seismic data to train and test the neural networks are obtained by solving wave propagation in a coupled poroviscoelastic-elastic media. A discontinuous Galerkin method is used to model wave propagation whereas a deep convolutional neural network is used for the parameter estimation problem. In the numerical experiment, the primary unknowns, the amount of stored groundwater and water-table level, are estimated, while the remaining parameters, assumed to be of less of interest, are successfully marginalized in the convolutional neural networks-based solution.

1 Introduction

Groundwater is the worlds largest readily available freshwater resource [13] and hugely important in both developed and developing countries. A detailed knowledge of the underground water storage (aquifer) properties and subsurface parameters are crucial in aquifer management, e.g. preventing waterlevel drawdown and planning aquifer protection. Traditional approaches for studying aquifers include geophysical surveys, followed by drilling and hydraulic test studies. The methodology presented in this paper can potentially make a significant difference both economically, by reducing the number of boreholes, and for data coverage, by transitioning from point data to continuous data.

Groundwater aquifers are found in porous media such as gravel or sand, or within fractured bedrock. One potential method to characterize and monitor aquifers is to employ seismic data. Seismic signals are generated by vibrators or man-made impacts and propagate through the porous aquifers. Because the seismic wave field interacts with the porous materials, the poroelastic signature of the aquifer can potentially be captured in seismograms. Hence, measured signals can be

used, with efficient numerical tools, to increase the knowledge of the groundwater reservoir state. Porosity, and its subsurface distribution, and the water-table are the factors that define the amount of groundwater stored in an aquifer. Recent progress in computational methods for seismic wave propagation has made it possible, in principle, to attempt the estimation of the key aquifer parameters [6, 23, 24]. It is well-known that porosity not only influences seismic velocities but also reduces the seismic amplitude. Hence, porosity can be estimated from seismic data [20, 28].

Neural networks have been applied to estimate groundwater levels and aquifer parameters. For example, [9], [10], and [33] use inputs such as temperature and well-based water level measurements to build a model for predicting groundwater levels. Furthermore, [4] and [19] are examples of studies where neural networks are used to estimate aquifer transmissivity and storativity from an applicable well-based dataset.

In this work, we consider prediction of the groundwater stored in an aquifer and the water-table level using seismic data. We couple simulations with deep learning for the prediction. More specifically, the presented approach consists of two main components:

I Seismic wave propagation from the source to receivers (i.e. the *forward problem*) in coupled poroviscoelastic-elastic media is simulated using a discontinuous Galerkin (DG) and low-storage Runge-Kutta time stepping methods [16, 26, 12]. The DG method is a well-known high-order accurate numerical technique to numerically solve differential equations and has properties that makes it well-suited for wave simulations (see e.g. [21, 11, 36, 32, 14]). These properties include, for example, the straightforward handling of complex geometries and large discontinuities in the material parameters. In addition, the method has excellent parallelization properties. All of these are essential features for the numerical scheme to be used in complex wave problems.

II The *inverse problem* of estimating the amount of groundwater and water table-level is solved by neural networks techniques. Compared to the conventional inversion techniques, neural networks has an advantage, that after the network has been trained, inferences can be carried out using the network without evaluating the forward model. This can dramatically reduce the computational time. Furthermore, the neural networks provide a straightforward approach to marginalize uninteresting parameters in the inference. In this study, we consider a convolutional neural network (CNN) [30, 5, 29, 7] that employs convolutions instead of matrix multiplications. The convolutional neural networks have shown potential to interpret the data in various estimation problems. Recent studies include electrical impedance tomography [15], aerosol research [17], and ultrasound tomography [27]. In the context of seismic imaging, the deep learning techniques have been studied, for example in [2, 3].

2 Model setup

We consider a 48 m-long model domain $I = [-24, 24]$ m, which consists of one elastic (bedrock) and two poroelastic (air- and water-saturated aquifer) subdomains, see Figure 1. This model is motivated by an on-going seismic experiment carried out in Virttaankangas, Finland [31]. The upper porous layer is air-saturated while the lower layer is water-saturated and the interface is the water-table. We set the free boundary condition on the top surface while other boundaries are modeled as outflow boundaries. Receivers are placed on the ground surface while sources are assumed to be buried at depth of 0.5 m from the ground surface. In addition, 38 receivers are used

to measure the vertical velocity data from 10 seismic source positions. The x-components of the receivers are distributed uniformly over the interval $x \in [-23,\ 23]$ m while x-components of the sources are distributed uniformly over the interval $x \in [-22.378,\ 22.378]$ m.

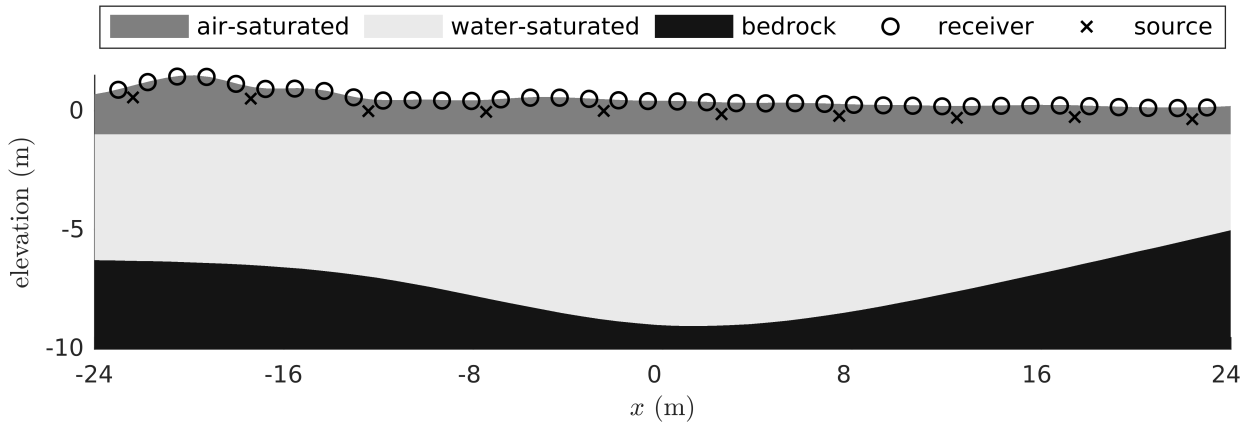

Figure 1: Model used for benchmarking purposes. Circles denote the receivers and the crosses denote the sources.

The seismic sources are modeled as force-type point sources, with the force pointing along the negative vertical-axis. We use the first derivative of a Gaussian pulse as the time-dependent source signal with a frequency $f_0 = 100$ Hz and a time delay $t_0 = 1.2/f_0$. The modeling time was 0.17 s. Note that recorded data are downsampled to a sampling frequency of 1 kHz on each receiver.

As the physical model, we use the coupled poroviscoelastic-elastic model studied in [12]. The aquifer is modeled as a fully saturated porous material based on the Biot theory while the bedrock layer is assumed to have very low porosity and can therefore be modeled as an elastic layer. In the physical model for the poroviscoelastic media, a total of 11 physical parameters must be given. These parameters are: fluid density ρ_{f}, fluid bulk modulus κ_{f}, viscosity η, solid density ρ_{s}, solid bulk modulus κ_{s}, frame bulk modulus κ_{fr}, frame shear modulus μ_{fr}, tortuosity τ, porosity ϕ, permeability k, and quality factor Q_0. In this study, we operate in the Biot's high-frequency regime for which the attenuation is controlled by the quality factor Q_0 [8, 34, 12]. In the elastic layer, we have a total of three unknown physical parameters, namely density ρ_{e}, bulk modulus K_{e}, and shear modulus μ_{e}. For a more detailed discussion of the physical model, we refer to [8] and references therein.

Model parametrization

The basement profile is taken from a ray-tracing based estimate $\hat{b}_y(x)$ [31] to model parametrization and smoothing constraints used for fast model convergence. However, the basement profile is not assumed to be fully accurate. Instead, profiles are perturbed randomly. Basement profiles are sampled as

$$b_y(x) = \hat{b}_y(x) + \delta_M M_c(x) + \delta_H H_{x_H}(x), \quad x \in I, \tag{1}$$

where M_c is a Matern field [35] with $\nu = 3/2$ or, more specifically, a Gaussian process with the covariance function

$$\begin{aligned} C_{\nu=3/2,c}(x,x') &= \mathrm{cov}(M_c(x), M_c(x')) \\ &= \left(1+\sqrt{3}\frac{|x-x'|}{c}\right)\exp\left(-\sqrt{3}\frac{|x-x'|}{c}\right), \quad x,x' \in I, \end{aligned} \tag{2}$$

where c is the correlation length and H is a Heaviside function ($H_{x_H}(x) = 1$ if $x \geq x_H$ and 0 otherwise), used to produce a possible discontinuity in the basement profile.

To generate a sample of a basement profile, we first sample the correlation length c and the standard deviation δ_M as $c \sim \mathcal{U}(3,\ 10)$ and $\delta_M \sim \mathcal{U}(0,1)$, respectively, and then generate a realization of the Matern field corresponding to c. In addition, we sample $\delta_H \sim \mathcal{U}(-2,2)$ and $x_H \sim \mathcal{U}(-24,120)$, i.e. the discontinuity exists with probability $48/144 = 1/3$, and add the discontinuity to the profile. Figure 2 shows examples of the sampled basement profiles. The water-table level W^l is assumed to vary uniformly between -3.7 m and -0.7 m whereas the ground surface remains fixed in each sample.

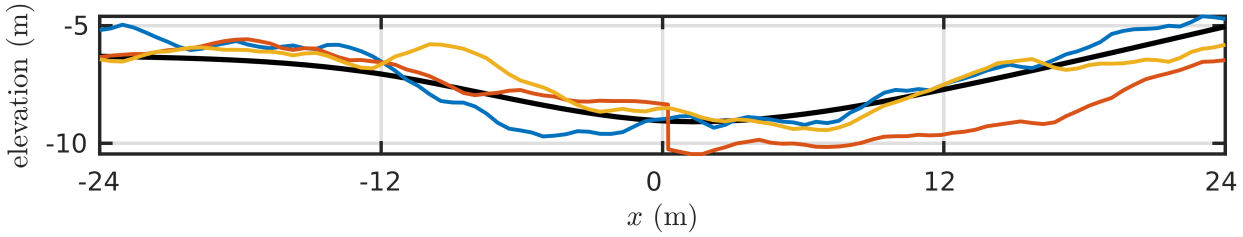

Figure 2: Three realizations and the ray-tracing based value (thick black line) to represent the basement profile.

The fluid parameters for the water-saturated zone are given by: the density $\rho_\mathrm{f} = 1000$ kg/m^3, the fluid bulk modulus $\kappa_\mathrm{f} = 2.1025$ GPa, and the viscosity $\eta = 1.3\text{e-}3$ Pa·s, while in the air-saturated part, we set: $\rho_\mathrm{f} = 1.2$ kg/m^3, $\kappa_\mathrm{f} = 1.3628\text{e}5$ Pa, and $\eta = 1.8\text{e-}5$ Pa·s. All other material parameters of the aquifer are assumed to be unknown. Furthermore, the remaining material parameters are realizations of Gaussian Markov random fields (MRF). The MRFs are generated using an isotropic Ornstein-Uhlenbeck process Π_c with the correlation length c. Each unknown material parameter $\theta = \{\rho_\mathrm{s},\ \kappa_\mathrm{s},\ \kappa_\mathrm{fr},\ \mu_\mathrm{fr},\ \tau,\ \phi,\ k,\ Q_0\}$ is randomized as follows:

$$\theta(x,y) = \theta^* + \delta_\theta \Pi_c(x,y) \tag{3}$$

$$\theta^* \sim \mathcal{U}(0.9\hat{\theta},\ 1.1\hat{\theta}),\ \delta_\theta \sim \mathcal{U}(0,\ 0.1\theta^*),\ c \sim \mathcal{U}(2,\ 20) \tag{4}$$

Mean values of $\hat{\theta}$ are given in Table 1. Furthermore, the unknown parameters are assumed to be uncorrelated.

In the bedrock layer, material fields are assumed to be purely elastic and homogeneous. Most of the energy reflects back from the basement layer and we expect the heterogeneity of the bedrock layer to have a minor effect. Therefore parameters $\theta = \{K_\mathrm{e},\ \mu_\mathrm{e},\ \rho_\mathrm{e}\}$ are assumed to be homogeneous and be sampled as in (3) with $\delta_\theta = 0$. For the mean values we set: $\hat{K}_\mathrm{e} = 66.0$ GPa, $\hat{\mu}_\mathrm{e} = 24.75$ GPa, and $\hat{\rho}_\mathrm{e} = 2750$ kg/m^3.

Table 1: The mean values of the uniform sampling distributions for each unknown physical parameter.

variable name	**symbol**	**value**
solid density	$\hat{\rho}_{\mathrm{s}}$ (kg/m^3)	2400
solid bulk modulus	$\hat{\kappa}_{\mathrm{s}}$ (GPa)	3.0
frame bulk modulus	$\hat{\kappa}_{\mathrm{fr}}$ (GPa)	0.3
frame shear modulus	$\hat{\mu}_{\mathrm{fr}}$ (GPa)	0.2
tortuosity	$\hat{\tau}$	1.8
porosity	$\hat{\phi}$ (%)	30
permeability	$\hat{k}$ (m^2)	5e-8
quality factor	$\hat{Q}_0$	50

Table 2: Calculated wave speeds for each subdomain. c_p^{I} and c_p^{II} denote the fast and slow pressure and c_s the shear wave speed, respectively. Both minimum/maximum values, which are based on sampling the material parameters, are given.

subdomain	c_p^{I} (m/s)	c_p^{II} (m/s)	c_s (m/s)
air-saturated	428/829	205/336	239/514
water-saturated	983/1563	210/470	231/478
bedrock	5434/6598	-	2715/3312

Table 2 lists the calculated wave speeds for each subdomain. The reported wave speed values correspond to the values generated by sampling of the material parameters. For further details of calculating wave speeds, we refer to [12].

The total amount of water stored in the aquifer is one of the primary unknowns in the application and can be computed as

$$V^w = \sum_{\ell=1}^{K^w} \phi_\ell^w A_\ell^w. \tag{5}$$

Here, K^w denotes the number of elements that belong to the water-saturated aquifer and ϕ_ℓ^w and A_ℓ^w are the porosity and area of the ℓth element in the water-saturated aquifer, respectively.

Two randomized porosity fields are shown in Figure 3 (a). Fields are shown for the cases when the amount of stored groundwater is low (left) or high (right). The blue horizontal line shows the water-table level. The visualized porosity fields share the same color scale.

The snapshots of the total solid velocity field are shown in the Figure 3 (b). The source location is shown in the top row. In contrast to the porosity visualization, the color scale is unique for the velocity fields. Snapshots are visualized at the time instant 59 ms. The velocity fields support the previous statement that very low levels of energy penetrate to the elastic subdomain. Furthermore, both of the selected samples contains a sharp discontinuity in the basement profile which can also be seen clearly from the snapshots.

In Figure 3 (c), the corresponding noiseless shot records are shown. As with the snapshots, the source location is given in the top row figures. The modeling results show the direct arrival and a clear reflection from the basement boundary. Water-table appears to produce no clear reflection as expected from the two heterogeneous models. Weak reflections from the boundaries are evident in

the record and the wave field snapshot, particularly for the case with deeper water-table.

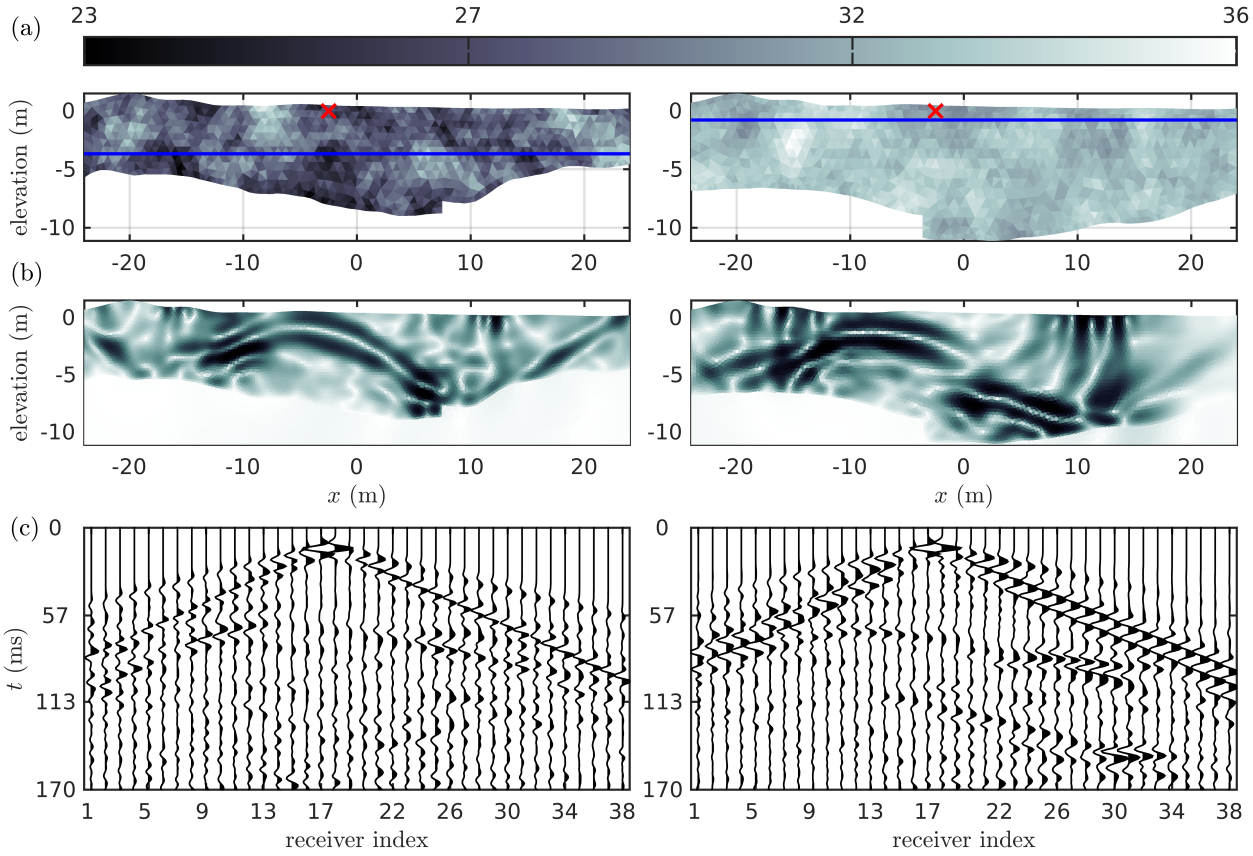

Figure 3: Two example porosity fields (a) and the corresponding recovered vertical solid velocity data (c). Colorbar at top shows the porosity values in percent. Snapshots in the middle (b) show the total solid velocity field for time instant 59 ms. Images correspond to samples where the amount of stored water is low (left) or high (right). The source location (red '×') and water-table level (horizontal blue line) are shown on the top figures.

3 Deep convolutional neural networks

In this work, we apply deep convolutional neural networks to estimate the parameters related to groundwater storage. A neural network $\Theta = \Theta(X; w)$ is a nested composition of functions, from a d-dimensional input space $X \in \mathbb{R}^d$ to $\Theta \in \mathbb{R}^m$, which is the output space of the inferred variable. For example, a neural network, comprising two convolutional layers and two fully connected layers (the architecture used in this work), can be expressed as

$$\Theta(X; w) = \sigma_4(w_4 \cdot \sigma_3(w_3 \cdot \sigma_2(w_2 * \sigma_1(w_1 * X + b_1) + b_2) + b_3) + b_4), \tag{6}$$

where w_1 and w_2 are the parametrized convolution kernels (filter weights) for the convolutional layers, w_3 and w_4 are the weight matrices for the fully connected layers, and b_k are bias terms. The functions σ_k are used to model nonlinearity of the network and in our case are chosen to be the

non-linear Rectified Linear Unit (ReLU), $\sigma = \max(0, X)$. It is to be noted that the convolutional layers also involves pooling layers that reduce the signal dimension.

The input of the network is the recorded vertical solid velocity wave data, expressed as 3D-data $X \in \mathbb{R}^d,\ d = N_t \times N_r \times N_s$, where N_t denotes the number of time steps, N_r the number of receivers, and N_s the number of sources (in our case, the input dimension is $d = 171 \times 38 \times 10$). An example of the velocity data are shown in Figure 4. Note that, since the data are three-dimensional, the convolutions are also three-dimensional. We are interested in the amount of water stored in the aquifer V^w (see Eq. (5)) and the water-table level W^l. We train the networks for the amount of stored water and water-table level separately, meaning that the output dimension is chosen to be $m = 1$ and Θ is either V_ℓ^w or W_ℓ^l. The precise model architecture is shown in Table 3.

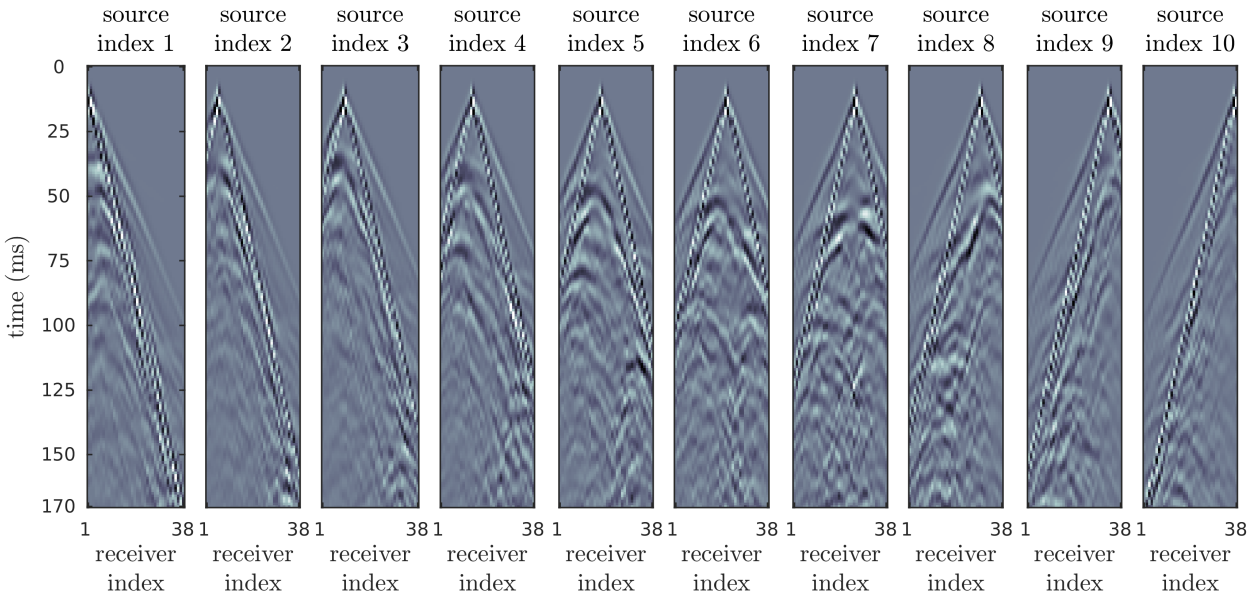

Figure 4: An example 3D-image, expressed as a stack of 2D-images, used for training the deep convolutional neural networks. Note that the deep learning algorithm uses the original pixel values of the image X. In each 2D-image, time is shown on the vertical-axis and the receiver index on the horizontal-axis. Hence, each figure shows data from different sources.

Table 3: The convolutional neural network architecture used in this study.

Layer k	**Type and non-linearity**	**input size**	**output size**
	Input	$171 \times 38 \times 10 \times 1$	$171 \times 38 \times 10 \times 1$
1	Convolution layer ($5 \times 3 \times 3$ filter)	$171 \times 38 \times 10 \times 1$	
	+ ReLU+ Max-pooling ($2 \times 2 \times 2$)		$86 \times 19 \times 5 \times 10$
2	Convolution layer ($5 \times 3 \times 3$ filter)	$86 \times 19 \times 5 \times 10$	
	+ ReLU + Max-pooling ($2 \times 2 \times 2$)		$43 \times 10 \times 3 \times 20$
3	Vectorization	$43 \times 10 \times 3 \times 20$	25800
	Fully connected layer + ReLU	25800	250
4	Fully connected layer	250	1
	Output		1

The neural network is trained using a dataset comprising of physical parameters $\{\Theta_\ell\}$ and corresponding images $\{X_\ell\}$, $\ell = 1, \ldots, N_{nn}$. The generation of such a dataset is described below. The goal is to find weights w and biases b that minimize the discrepancy between the physical parameters $\{\Theta_\ell\}$ and the values predicted by the neural networks $\{\Theta(X_\ell; w, b)\}$. Hence, one is minimizing the quadratic loss function $f(w, b; \{X_\ell\})$ over the simulation dataset

$$f(w, b; \{X_\ell\}) = \frac{1}{N_{nn}} \sum_{\ell=1}^{N_{nn}} (\Theta(X_\ell; w, b) - \Theta_\ell)^2 \tag{7}$$

to obtain the network parameters, weights, and biases of the network. We employ the Adam optimizer [22] for the optimization and the batch size is chosen to be 50 samples. For the network generation, a total of 200 full training cycles in stochastic optimization (epochs) was run. As the computing interface, a Python library TensorFlow [1] was used.

Training and test datasets

For the training of the convolutional neural networks, we generate a dataset comprising 15,000 samples, using computational grids that have ∼3.5 elements per wavelength. The physical parameters and the geometry for each sample are drawn using the framework discussed in Section 2. To further control the numerical accuracy of the forward solver, the order N_ℓ of the polynomial basis functions is allowed to vary in elements ℓ of the computational grid. The order of the basis function in each element is selected from

$$N_\ell = \left\lceil \frac{2\pi \bar{a} h^\ell_{\max}}{\lambda^w_\ell} + \bar{b} \right\rceil, \tag{8}$$

where $\lceil \cdot \rceil$ is the ceiling function, $h^\ell_{\max}$ is the largest distance between two vertices, $\lambda^w_\ell = c^\ell_{\min}/f_0$ is the wavelength, $c^\ell_{\min}$ is the minimum wave speed, and parameters $\bar{a}$ and $\bar{b}$ control the local accuracy on each element. For the generation of training data, we set $(\bar{a},\ \bar{b}) = (1.0294,\ 0.7857)$, see [26, 25] for further details of the nonuniform basis orders.

To train the model to understand the presence of the measurement noise, the samples are corrupted with a simulated Gaussian noise. More precisely, we create five copies of each image in the dataset which are then corrupted as

$$X^{\text{noised}}_\ell = X_\ell + A\alpha\epsilon^A + B|X_\ell|\epsilon^B, \tag{9}$$

where α is the maximum absolute value of the training dataset and $\epsilon^{A/B}$ are independent zero-mean Gaussian random variables. The second term represents additive (stationary) white noise and the last term represents noise relative to the signal amplitude. To include a wide range of noise levels, the coefficients A and B for each sample are randomly chosen such that the standard deviations of the white noise component is between 0.03-5% (varying logarithmically), and the standard deviations of the relative component is between 0-5%. The total number of samples in the training set is $N_{nn} = 5 \times 15000 = 75000$.

Furthermore, an additional test dataset were generated, which comprise 3000 samples. The dataset were generated as the training set, except computational grids were required to have ∼4 elements per wavelength and the non-uniform basis order parameters are $(\bar{a},\ \bar{b}) = (1.2768,\ 1.4384)$ [26] of model (8). The main reason to use different discretization was to avoid an inverse crime [18] related to simulation studies: the use of the same computational model, e.g. same discretizations, to

generate both training and test data could potentially lead to a situation in which severe modelling errors are ignored, yielding unrealistic impressions of the accuracy of the estimates as compared to actual performance with real data. In the test data, noise was added in a more systematic manner to study the performance with different noise levels.

4 Results

We applied the trained network to predict the amount of stored groundwater and water-table level from images of the test data. One must note that the proposed neural networks-based approach enables direct estimates of the amount of stored groundwater in a heterogeneous porous material rather than separately estimating the porosity field and the aquifer geometry. Both the porosity field and the geometry could potentially have many unknowns which increase the overall computational demand.

Figure 5 shows estimates for the test data, contaminated with the white noise component of a moderate level, and Figure 6 shows results for the high noise level. The figures also include relative prediction error histograms.

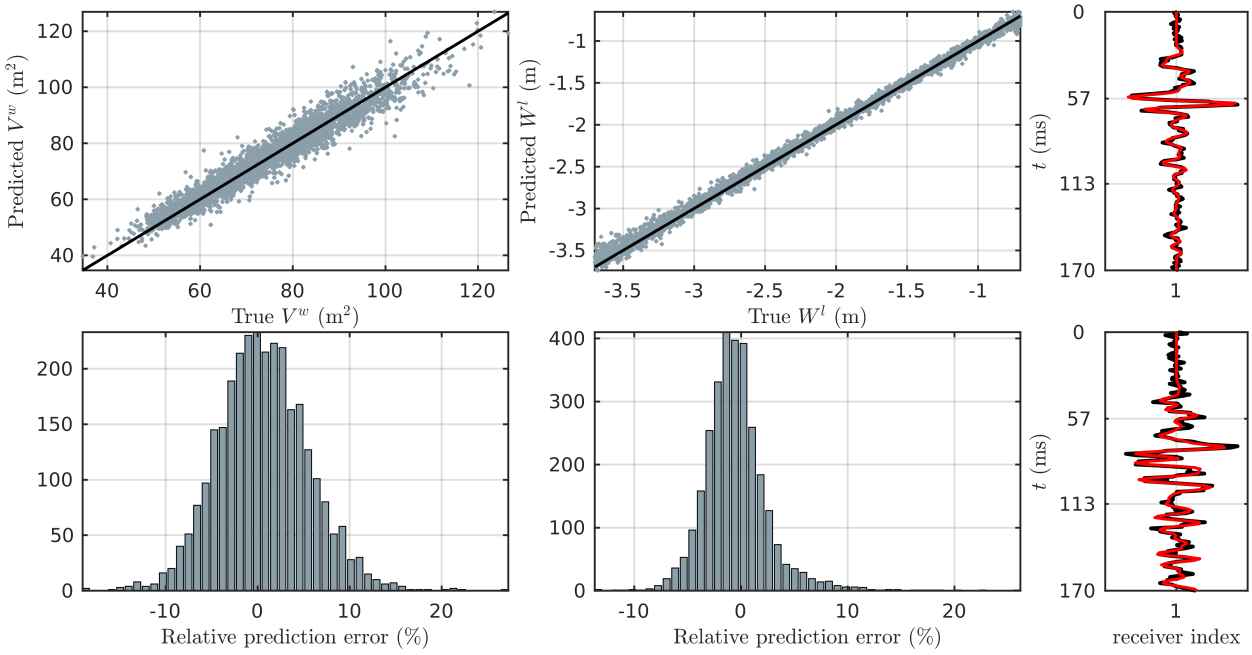

Figure 5: Predicted amounts of stored water (left), water-table levels (middle), and two example signals for noisy (black) and noiseless (red) data for the test data with $A = 0.011$ and $B = 0.248$. Bottom row shows histograms of the relative prediction error (difference between the predicted and true values).

Figure 7 shows a normalized root-mean-squared error (NRMSE) as a function of the parameters A and B for both estimated parameters V^w (top) and W^l (bottom). Results show that the estimation is much more noise sensitive for parameter A than parameter B. In the figures, the selections of A and B, used in Figures 5 and 6 are highlighted with white '$\times$' and '$\circ$' markers, respectively.

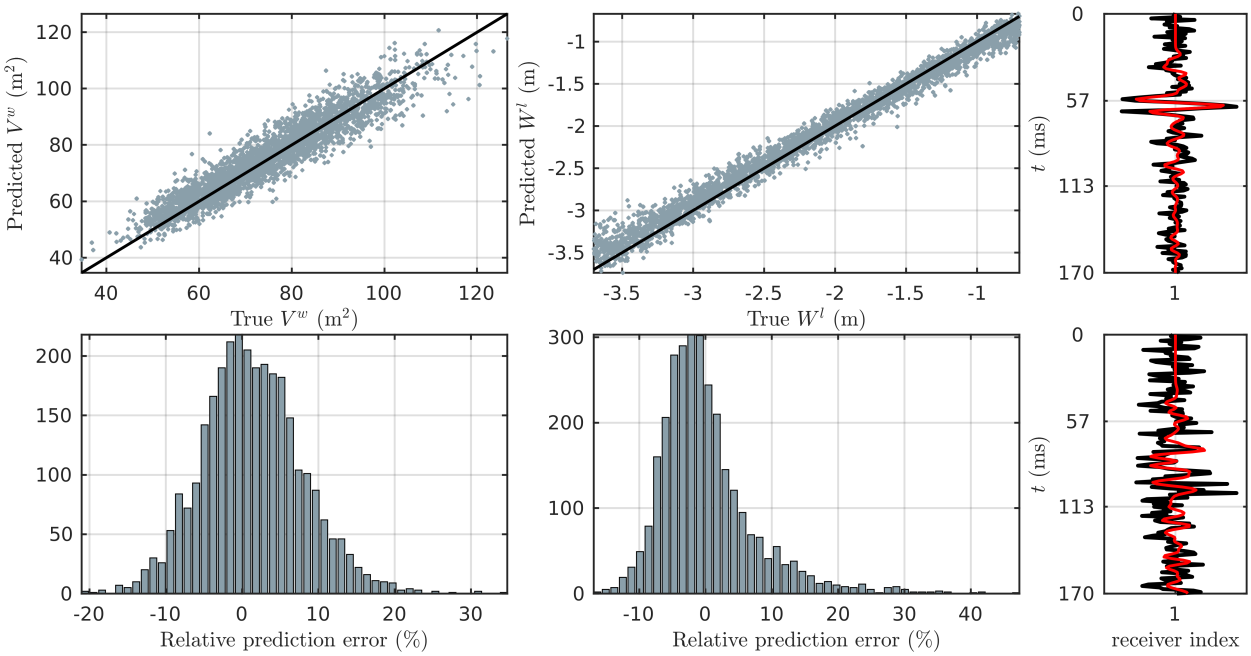

Figure 6: Results with noise parameters $A = 0.045$ and $B = 0.497$. Otherwise same caption as in Figure 5.

Furthermore, the upper bounds for the noise parameters used in the training are highlighted with a dashed white line.

In addition, we have carried out experiments with different network architectures. For instance, increasing the number of layers or neurons did not improve the performance significantly, meanwhile there was some performance decrease with a smaller number of layers or neurons. On the other hand, using 2D-convolutions results into inferior performance compared to the 3D-convolutions.

5 Conclusions

We propose the use of a convolutional neural network (CNN) to estimate the water stored in an aquifer and the water-table level from synthetic seismic data. In the model, seismic data were generated in a heterogeneous three-layered material model. In the material model, two subdomains were modeled as poroviscoelastic while the bedrock layer was modeled as purely elastic. A total of 38 receiving sensors were positioned on the ground surface while a total of 10 separately activated sources were assumed to be located at a depth of 0.5 m from the ground surface. The recorded vertical solid velocity data were represented as 3D-images, which were used as an input data to the CNN.

We estimated the water stored in the aquifer and water-table level while all other material parameters were considered as nuisance (uninteresting) parameters. Based on the results, we argue that these parameters can be estimated with acceptable accuracy with a wide variety of noise levels, while the nuisance parameters are successfully marginalized. The error histograms for both stored water and water-table show promising accuracy in terms of relative prediction error and bias.

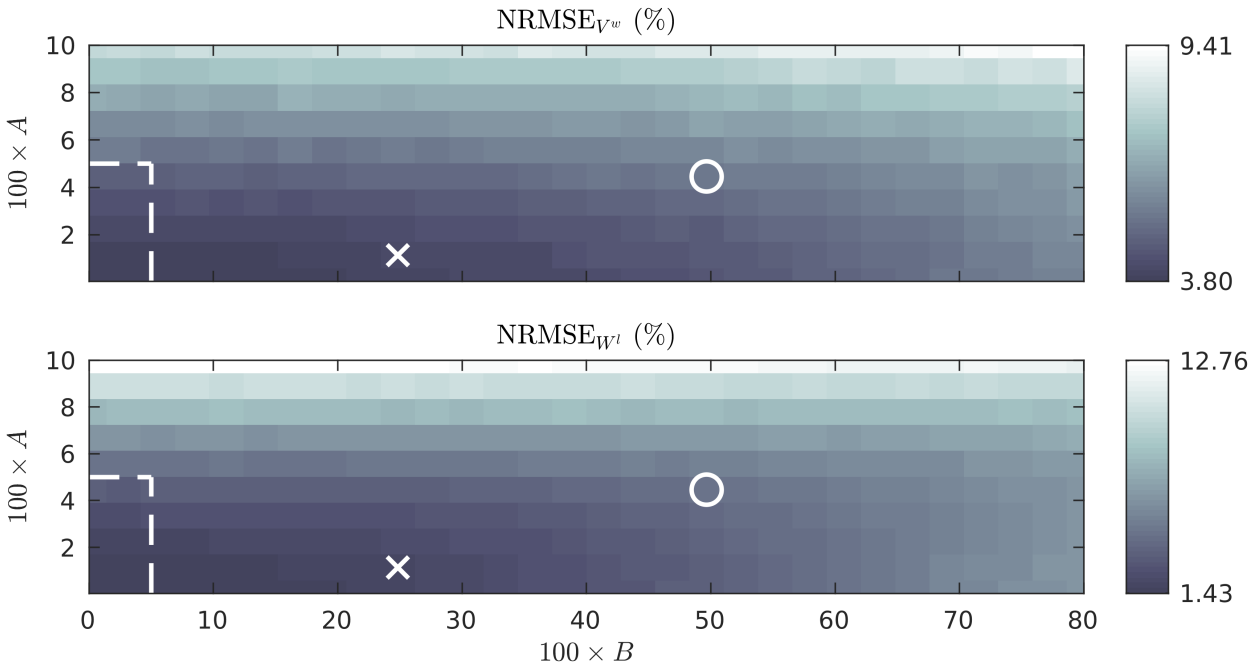

Figure 7: Normalized RMSE as a function of noise parameters A and B for the amount of water stored in the aquifer V^w (top) and the water-table level W^l (bottom). The dashed white box in the bottom-left corner shows the parameter space used in the training data. In addition, the white '$\times$' data are visualized in Figure 5 and the white '$\circ$' in Figure 6.

One benefit of the proposed framework is that it can effectively be used for monitoring aquifer's water storage state without estimating the material fields, which may change as a result of seasonal fluctuations and/or earthquakes, for example.

Future studies should include the extension to field/real measurements. This study is performed with synthetic data, but the future goal is to combine synthetic and real data in such way that the training data can be generated by simulating wave fields while the test data is the actual measurements. These are essential steps to guarantee the effectiveness of the proposed method.

Acknowledgments

This work has been supported by the strategic funding of the University of Eastern Finland and by the Academy of Finland (Finnish Centre of Excellence of Inverse Modelling and Imaging). The authors wish also to acknowledge CSC - IT Center for Science, Finland, for computational resources.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz

Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] M. Araya-Polo, T. Dahlke, C. Frogner, C. Zhang, T. Poggio, and D. Hohl. Automated fault detection without seismic processing. *The Leading Edge*, 36(3):208–214, 2017.

[3] M. Araya-Polo, J. Jennings, A. Adler, and T. Dahlke. Deep-learning tomography. *The Leading Edge*, 37(1):58–66, 2018.

[4] K. S. Balkhair. Aquifer parameters determination for large diameter wells using neural network approach. *Journal of Hydrology*, 265(1):118–128, 2002.

[5] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

[6] M. Bosch, T. Mukerji, and E. F. Gonzalez. Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics*, 75(5):75A165–75A176, 2010.

[7] N. Buduma. *Fundamentals of Deep Learning, Designing Next-Generation Machine Intelligence Algorithms*. O'Reilly Media, 2017.

[8] J. M. Carcione. *Wave Fields in Real Media: Wave propagation in anisotropic, anelastic and porous media*. Elsevier, 1 edition, 2001.

[9] P. Coulibaly, F. Anctil, R. Aravena, and B. Bobée. Artificial neural network modeling of water table depth fluctuations. *Water Resources Research*, 37(4):885–896, 2001.

[10] I. N. Daliakopoulos, P. Coulibaly, and I. K. Tsanis. Groundwater level forecasting using artificial neural networks. *Journal of Hydrology*, 309(1):229 – 240, 2005.

[11] J. de la Puente, M. Dumbser, M. Käser, and H. Igel. Discontinuous Galerkin methods for wave propagation in poroelastic media. *Geophysics*, 73(5):T77–T97, 2008.

[12] N. F. Dudley Ward, T. Lähivaara, and S. Eveson. A discontinuous Galerkin method for poroelastic wave propagation: Two-dimensional case. *Journal of Computational Physics*, 350:690–727, 2017.

[13] C. W. Fetter. *Applied Hydrogeology*. Prentice Hall, 4 edition, 2001.

[14] M. Gaborit, O. Dazel, P. Göransson, and G. Gabard. Coupling of Finite-Element and Plane Waves Discontinuous Galerkin methods for time-harmonic problems. *ArXiv e-prints*, 2018.

[15] S. J. Hamilton and A. Hauptmann. Deep D-bar: Real time Electrical Impedance Tomography Imaging with Deep Neural Networks. *ArXiv e-prints*, 2017.

[16] J. S. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Springer, 2007.

[17] J. Joutsensaari, M. Ozon, T. Nieminen, S. Mikkonen, T. Lähivaara, S. Decesari, M. C. Facchini, A. Laaksonen, and K. E. J. Lehtinen. Identification of new particle formation events with deep learning. *Atmospheric Chemistry and Physics*, 2018. https://doi.org/10.5194/acp-2017-1189, in review.

[18] J. Kaipio and E. Somersalo. Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–504, 2007.

[19] H. Karahan and M. T. Ayvaz. Simultaneous parameter identification of a heterogeneous aquifer system using artificial neural networks. *Hydrogeology Journal*, 16(5):817–827, 2008.

[20] S. Karimpouli, H. Hassani, M. Nabi-Bidhendi, H. Khoshdel, and A. Malehmir. Application of probabilistic facies prediction and estimation of rock physics parameters in a carbonate reservoir from Iran. *Journal of Geophysics and Engineering*, 10(1):015008, 2013.

[21] M. Käser and M. Dumbser. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes - I. The two-dimensional isotropic case with external source terms. *Geophysical Journal International*, 166(23):855–877, 2006.

[22] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *ArXiv e-prints*, 2014.

[23] T. Lähivaara, N. F. Dudley Ward, T. Huttunen, J. Koponen, and J. P. Kaipio. Estimation of aquifer dimensions from passive seismic signals with approximate wave propagation models. *Inverse Problems*, 30(1):015003, 2014.

[24] T. Lähivaara, N. F. Dudley Ward, T. Huttunen, Z. Rawlinson, and J. P. Kaipio. Estimation of aquifer dimensions from passive seismic signals in the presence of material and source uncertainties. *Geophysical Journal International*, 200:1662–1675, 2015.

[25] T. Lähivaara and T. Huttunen. A non-uniform basis order for the discontinuous Galerkin method of the 3D dissipative wave equation with perfectly matched layer. *Journal of Computational Physics*, 229:5144–5160, 2010.

[26] T. Lähivaara and T. Huttunen. A non-uniform basis order for the discontinuous Galerkin method of the acoustic and elastic wave equations. *Applied Numerical Mathematics*, 61:473–486, 2011.

[27] T. Lähivaara, L. Kärkkäinen, J. M.J. Huttunen, and J. S. Hesthaven. Deep convolutional neural networks for estimating porous material parameters with ultrasound tomography. *The Journal of the Acoustical Society of America*, 143(2):1148–1158, 2018.

[28] T. Lähivaara, A. Pasanen, A. Malehmir, and J. P. Kaipio. Full-waveform seismic inversion for estimating aquifer dimensions and hydrologic parameters. In *23rd European Meeting of Environmental and Engineering Geophysics*, 2017. DOI: 10.3997/2214-4609.201702020.

[29] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

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

[31] G. Maries, E. Ahokangas, J. Mäkinen, A. Pasanen, and A. Malehmir. Interlobate esker architecture and related hydrogeological features derived from a combination of high-resolution reflection seismics and refraction tomography, Virttaankangas, southwest Finland. *Hydrogeology Journal*, 25:829–845, 2017.

[32] A. Modave, A. St-Cyr, and T. Warburton. GPU performance analysis of a nodal discontinuous Galerkin method for acoustic and elastic models. *Computers & Geosciences*, 91:64–76, 2016.

[33] S. Mohanty, M. K. Jha, A. Kumar, and K. P. Sudheer. Artificial neural network modeling for groundwater level forecasting in a river island of eastern India. *Water Resources Management*, 24(9):1845–1865, 2010.

[34] C. Morency and J. Tromp. Spectral-element simulations of wave propagation in porous media. *Geophysical Journal International*, 175(1):301–345, 2008.

[35] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[36] L. C. Wilcox, G. Stadler, C. Burstedde, and O. Ghattas. A high-order discontinuous Galerkin method for wave propagation through coupled elastic-acoustic media. *Journal of Computational Physics*, 229:9373–9396, 2010.