Advances in Water Resources 138 (2020) 103545

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Seeing macro-dispersivity from hydraulic conductivity field with convolutional neural network

Zhengkun Zhou, Liangsheng Shi*, Yuanyuan Zha

State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, Hubei 430072, China

ARTICLE INFO

Keywords:
Macrodispersivity
Convolutional neural work
Hydraulic conductivity field
Contaminant transport
Heterogeneity
Groundwater

ABSTRACT

Building a quantitative relation between the spatial heterogeneity of the hydraulic conductivity fields and the macroscale behavior of solute transport is fundamental for groundwater environment problem. In this work, the deep learning technique is explored to build the functional mapping between the hydraulic conductivity field and the longitudinal macro-dispersivity. We examine the capability of the deep neural network in estimating macro-dispersivities of conductivity fields with different variances. The universality of the trained deep neural network is investigated. Comparisons of the neural network results and the reference values (macro-dispersivities from transport simulation) suggest the promising potential of deep learning technique in porous media with moderate heterogeneity. For a given size of training datasets, the deep neural network produces better macro-dispersivity estimation for the conductivity field with smaller variance. The trained neural network by conductivity fields with larger variance has stronger universality for macro-dispersivity estimation. This study demonstrates that deep neural network can be an effective alternative for estimating macroscale behavior of solute transport by directly interpreting hydraulic conductivity fields.

1. Introduction

Describing solute transport in porous media is fundamental for many groundwater contamination problems (de Barros et al., 2016; Han et al., 2016; Guo et al., 2019). The dispersive transport models are the widely used models at various scales. Therefore, estimating the dispersivity at corresponding scale is essential for any efforts in predicting the spreading of a contaminant plume. At field-scale, the structure of natural porous media usually exhibits spatial variability, which has a profound impact on solute transport (Fiori et al., 2016; Cvetkovic et al., 2016). For applying the advection-dispersion models under field conditions, hydro-geologists have proven that the magnitude of field-scale dispersivity (macro-dispersivity) can be several orders of magnitude higher than lab-scale value for the same material (Fiori et al., 2017). This increase mainly attributes to the spatial variability of aquifer structure which can be generally described by the spatial distribution of the hydraulic conductivity.

Considering the heterogeneous distribution of hydraulic conductivity as a random field and relating flow and transport to its statistical moments has been one of the primary goals of the field of stochastic modelling (e.g. Dagan, 1989; Gelhar, 1993; Rubin, 2003). A fundamental issue addressed by these works is how macro-dispersivity can be related to the statistical properties of the hydraulic conductivity field (e.g. Rehfeldt et al., 1992; Fiori et al., 2015a; Zech et al., 2015). However, the general applicability of the stochastic approach is sometimes questionable due to several foundational assumptions. And the first and second-order spatial statistics cannot provide sufficient information on estimating of macro-dispersivity (e.g. Zinn and Harvey, 2003; Zheng and Gorelick, 2003; Zheng et al., 2011; Molinari et al., 2015; Sanchez-Vila and Fernàndez-Garcia, 2016). The conductivity fields with the same first two moments may produce very different solute spreading because of the spatial patterns that are not characterized by these statistics (e.g. Zinn and Harvey, 2003; Bianchi and Pedretti, 2018). The concepts of connectivity and geological entropy then emerge as other attempts to characterize the transport behavior from the heterogeneous conductivity fields (e.g. Renard and Allard, 2013; Rizzo and de Barros, 2017; Fiori, 2014; Fiori and Jankovic, 2012; Freixas et al., 2017; Fernàndez-Garcia et al., 2010; Bianchi and Pedretti, 2017; 2018). In short, researchers have made great efforts to predict solute transport behavior only from a characteristic description of the conductivity field. Despite the helpfulness of these works in understanding the correlation between the heterogeneity of conductivity field and the transport behavior, a direct and efficient functional mapping between the conductivity field and the transport behavior for predictive purposes remains to be solved.

Deep learning methods are representation-learning methods which can be fed with raw data and automatically discover the representations needed for classification. The convolutional neural network (CNN), which is a particular network type of deep learning, is designed to

* Corresponding author.
E-mail address: liangshs@whu.edu.cn (L. Shi).

https://doi.org/10.1016/j.advwatres.2020.103545
Received 18 September 2019; Received in revised form 14 January 2020; Accepted 23 February 2020
Available online 24 February 2020

process data with multiple arrays and has dominated the field of image classification and achieved many practical success (LeCun et al., 2015). Researchers have adopted CNN to solve various problems in water science (Shen, 2018), e.g., extracting flow features from flow fields (Ströfer et al., 2019), serving as a surrogate model for uncertainty quantification of partial differential equations (Zhu and Zabaras, 2018; Mo et al., 2018), reconstructing porous media (Mosser et al., 2017; Cang et al., 2017; Wang et al., 2018) and reducing dimensionality of complex geological models (Laloy et al., 2017; 2018; Chan and Elsheikh, 2017). Particularly, CNN has also been used to predict the effective properties of heterogeneous materials directly (Srisutthiyakorn, 2016; Cang et al., 2018; Wu et al., 2018; Vasilyeva and Tyrylgin, 2018). Srisutthiyakorn (2016) demonstrated the possibility of predicting permeability from rock images by CNN. Wu et al. (2018) improved the performance of CNN in predicting permeability by involving physical information. Vasilyeva and Tyrylgin (2018) constructed a CNN to accelerate the calculation of elasticity tensors of random poroelastic media. Since the macro-dispersivity is supposed to be formation-specific (Zech et al., 2015), it should be possible to develop a surrogate model that directly map a conductivity field to corresponding macro-dispersivity by CNN. Regarding conductivity fields as abstract images of aquifers and resembling the task of image classification (LeCun et al., 1995; Krizhevsky et al., 2012), CNN takes a conductivity field as input and gives the classification label (macro-dispersivity) as output by recognizing the features in the field.

In this work, a CNN of eight layers is set up to learn a map between stochastic conductivity field and longitudinal macro-dispersivity based on synthetic two-dimensional conductivity fields. A direct and efficient functional map between macro-dispersivity and conductivity field is built. Such surrogate model has the potential to enable fast estimation of macro-dispersivities of aquifers without performing direct simulations of physical models which are usually time-consuming and need many input parameters. To our knowledge, this is the first study to interpret macro-dispersivity from conductivity field by deep neural network directly. Two-dimensional isotropic multi-Gaussian fields are generated with a sequential Gaussian simulation (Bellin and Rubin, 1996). Reference macro-dispersivities are calculated by conducting numerical trace tests. The performance of the proposed neural network architecture is validated through various training and test datasets. For a specific size of training datasets, the estimating capability of the neural network declines with increasing variances (namely increasing heterogeneity of conductivity fields). Notably, the universality of the deep neural network in estimating macro-dispersivities of conductivity fields with different degree of heterogeneity is also explored. Assessments of the proposed neural network architecture show that directly estimating macro-dispersivity from distribution of hydraulic conductivity is promising.

The rest of the paper is organized as follows: Section 2 recapitulates the entire computation framework, the CNN architecture and generation of training dataset. Results of numerical tests and discussions are presented in Section 3. Finally, Section 4 concludes the paper.

2. Methodology

Since the conductivity fields can be regarded as abstract images of aquifers, and the macro-dispersivity, which is an effective coefficient to characterize field-scale dispersion process in dispersive subsurface mass-transport models, can be considered as an inherent property of an aquifer. Using the deep learning methods to estimate macro-dispersivity from conductivity field is worth a try. In this section, a framework to train a convolutional neural network is set up, and the detailed steps are presented.

2.1. Sketch of training framework

The objective of the training framework presented here is to train a deep-learning model for fast estimation of macro-dispersivity from the spatial distribution of the hydraulic conductivity. This training framework consists of the following steps, as illustrated in Fig. 1:

(1) Generating training datasets. Two-dimensional random fields of the hydraulic conductivity are generated. Direct simulations with the random walk particle tracking method (Salamon et al., 2006b) are then used to compute the macro-dispersivities of the generated conductivity fields. The field-dispersivity pairs consist of the training datasets for the deep neural network model. The details of generating the training datasets are presented in Section 2.2.
(2) Training the CNN. The training datasets from the previous step are then used to train our CNN that takes a heterogeneous conductivity field as input and gives macro-dispersivity as output. Section 2.3 specifies the CNN architecture and training process.
(3) Estimating macro-dispersivities. The trained CNN is then used to estimate macro-dispersivities of new conductivity fields that are not in the training datasets.

It should be noted that this work is established on two-dimensional synthetic conductivity fields and the macro-dispersivity here is referred to as longitudinal macro-dispersivity. Three-dimensional synthetic or real-world conductivity field will be the goal of future works.

2.2. Building of training datasets

Simulations of uniform flow trace tests are conducted on the conductivity fields to estimate the corresponding macro-dispersivities. Square bi-dimensional confined aquifers with the uniform mean flow in the x-direction is considered, and the flow is driven by a mean hydraulic gradient equal to J=0.0176. The computational domain used in the study is illustrated in Fig. 2. The domain is discretized into square cells each of dimension $0.25I_{lnK}$. There are 140 grid cells in each coordinate direction corresponding to an overall grid dimension of $35I_{lnK}$. For boundaries parallel to the mean flow, boundaries conditions are no flux. The rest boundaries are the constant head. We have to emphasize that the grid block number per correlation scale is fixed to 4 in this work. Only training sets of different variances are considered. The effects of training sets of different correlation lengths are not investigated here. And there may be a distribution shift that links to the I_{lnK} change (Liu and Ziebart, 2014). This issue will be investigated in the future study.

2.2.1. Generating hydraulic conductivity fields

The synthetic aquifers are characterized by spatially varying hydraulic conductivity $K(\mathbf{x})$, and the $\ln K$ fields follow a Gaussian random function. The isotropic exponential covariance model is considered:

$$C_{lnK}(r) = \sigma^2_{lnK} \exp\left(-\frac{r}{I_{lnK}}\right) \tag{1}$$

where r is the separation vector between two points of the aquifers, $\ln K$ denotes the log-hydraulic conductivity, σ^2_{lnK} is the variance of $\ln K$ and I_{lnK} is the correlation length. The log-conductivity fields are generated through HYDRO-GEN (Bellin and Rubin, 1996). The geometric mean of the $\ln K$ fields is set equal to 116.7 m/day (K_g = 116.7m/day), which is a reasonable value for a sandy aquifer.

2.2.2. Generating velocity fields

Before the transport simulations, velocity fields must be developed at first. Groundwater flow is described by the flow mass conservation equation and Darcy's law. Under steady-state flow conditions, the flow equation can be formally expressed as

$$\nabla \cdot (K(\mathbf{x})\nabla h(\mathbf{x})) = 0 \tag{2}$$

where $h(\mathbf{x})$ is the hydraulic head, and $K(\mathbf{x})$ is the spatially variable hydraulic conductivity. A finite-difference groundwater flow model, MODFLOW (Harbaugh et al., 2000) is used to solve the flow problem and

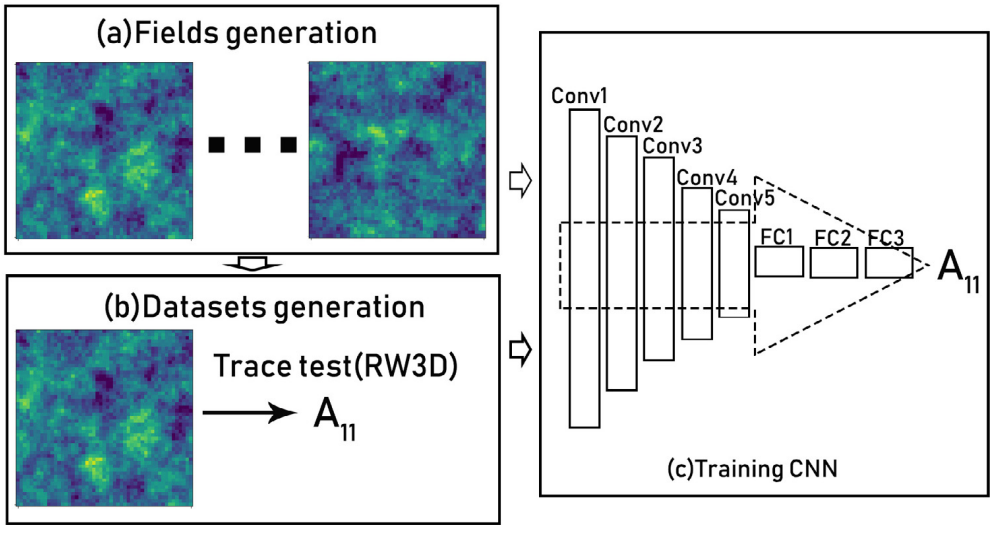

Fig. 1. Sketch of the framework, including (a) generating heterogeneous conductivity fields, (b) building datasets by performing numerical trace test with particle tracking method, (c) using the datasets to train the CNN.

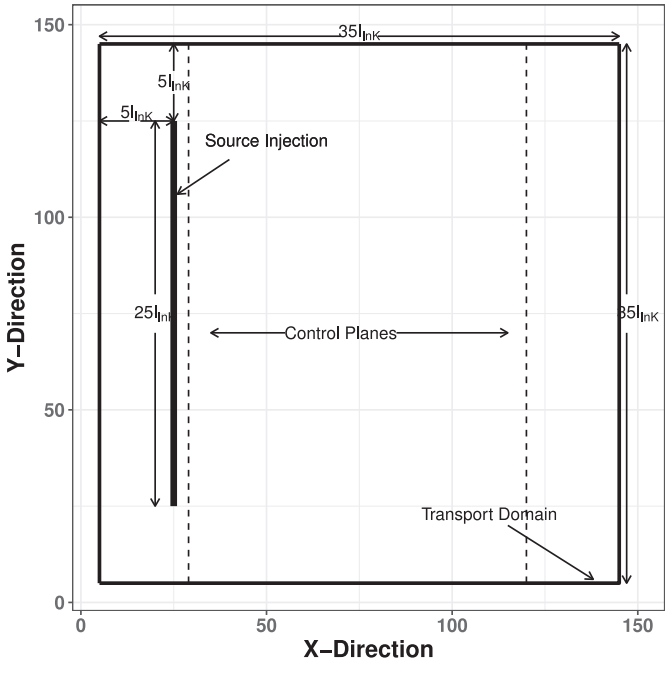

Fig. 2. Sketch of domain of transport simulations showing the initial location of trace source, the control planes.

the velocity field corresponding to each conductivity field is then determined with the mean hydraulic gradient. The Darcy's velocity distribution is computed by

$$\mathbf{q}(\mathbf{x}) = -K(\mathbf{x})\nabla h(\mathbf{x}) \tag{3}$$

2.2.3. Transport simulation with particle tracking

Using the solute mass conservation equation and Fick's law at the local scale, solute transport through porous media is locally governed by the advection-dispersion equation and can be written as

$$n\frac{\partial C(\mathbf{x},t)}{\partial t} = -\nabla\cdot(\mathbf{q}(\mathbf{x})C(\mathbf{x},t)) + \nabla\cdot(n\mathbf{D}\nabla C(\mathbf{x},t)) \tag{4}$$

where $\mathbf{q}$ is the Darcy velocity, n is the porosity, C is the solute concentration, and $\mathbf{D}$ is the local hydro-dynamic dispersion tensor. Solute transport simulations are conducted based on velocity fields from Eq. (2), through a random walk particle tracking transport code (RW3D) (Fernàndez-Garcia et al., 2005; Salamon et al., 2006b; 2006a; Henri and Fernàndez-Garcia, 2014; 2015). The porosity is assumed to be spatially homogeneous with a value of 0.35. The tracer plume migration is simulated by partitioning the tracer mass into a large number of representative mass particles. Advection is simulated by moving particles within the velocity field, whereas dispersion is emulated by Brownian motion. This Lagrangian method is free of numerical dispersion. More details about the numerical algorithm can be found in the work of Fernàndez-Garcia et al. (2005) and Salamon et al. (2006b). At the beginning of simulations, a number of particles (1000) are uniformly distributed in a vertical line perpendicular to the mean flow direction. This line is located $5I_{lnK}$ away from the up-gradient boundary to relieve boundary effects (Rubin and Dagan, 1988; 1989). The source size is $25I_{lnK}$ in the transverse direction to the mean flow and is centered with the transverse direction. For estimating longitudinal macro-dispersivities, a control plane transverse to the mean flow direction is placed $3I_{lnK}$ away from the down-gradient boundary. It should be noted that 1000 particles are enough to offer an accurate estimate of macro-dispersivity in this work. The validation of the chosen particle number is displayed in Appendix A.

2.2.4. Estimating of macro-dispersivities from temporal moments

The first arrival time passing through the control plane is tracked until particles exited the lower constant head boundary. Then, the estimation of associated temporal moments of the breakthrough curve (BTC) can be conducted without having to evaluate the actual shape of the BTC. The nth-absolute temporal moment $M_n(x)$ is expressed as the expected value of the arrival time of particles at the control plane to the nth power (Fernàndez-Garcia et al., 2005; Fadili et al., 1999; Shapiro and Cvetkovic, 1988)

$$M_n(x) = \frac{1}{m_{\text{total}}}\int_0^\infty t^n Q C_{\text{f}}(x,t)\mathrm{d}t \approx \frac{1}{\sum_{k=1}^{NP} m_{\text{p}^{(k)}}}\sum_{k=1}^{NP} m_{\text{p}^{(k)}}\left(t_{\text{p}}^{(k)}(x)\right)^n \tag{5}$$

where m_{total} is the total injected tracer mass, Q is the total water flux passing through the control plane, x is the mean flow direction coordinate, C_{f} is the flux concentration of tracer passing through the control plane, $m_{\text{p}^{(k)}}$ is the mass of a particle, $t_{\text{p}^{(k)}}$ is the first arrival passage time of the kth particle and NP is the total number of particles arrival at the x-control plane. The nth central temporal moment $M_{T,n}(x)$ is then calculated by the relationship between central and absolute temporal moments (Fernàndez-Garcia et al., 2005; Kendall and Stuart, 1977)

$$\begin{aligned} M_{T,n}(x) &= \frac{1}{m_{\text{total}}}\int_0^\infty (t - M_1(x))^n Q C_{\text{f}}(x,t)\mathrm{d}t \\ &= \sum_{r=0}^{n}\binom{n}{r} M_{n-r}(x)\left(-M_1(x)\right)^r \end{aligned} \tag{6}$$

Macro-dispersivities from temporal moments are calculated as (Fernàndez-Garcia and Gómez-Hernández, 2007; Fernàndez-Garcia et al., 2005; Aris, 1958; Goltz and Roberts, 1987)

$$A_{11}(x_p) = \frac{x_p}{2}\frac{M_{T,2}(x_p)}{(M_1(x_p))^2} \tag{7}$$

where x_p is distance between the tracer source and the control plane, A_{11} is the longitudinal macro-dispersivity. Macro-dispersivity from Eq. (7) is the apparent longitudinal macro-dispersivity that corresponds to a specific hydraulic conductivity field. The apparent longitudinal macro-dispersivity is viewed as equivalent value in homogeneous porous media. Applying it to the classical advection-dispersion equation will lead to the same first two temporal moments of the BTC as observed in the simulations for heterogeneous porous media (Fernàndez-Garcia et al., 2005). Finally, the estimated macro-dispersivities and the corresponding hydraulic conductivity fields constitute the training and validation datasets. Other details on the evaluation of macro-dispersivities from temporal moments can be found in the work of Fernàndez-Garcia et al. (2005) and Fernàndez-Garcia and Gómez-Hernández (2007). Furthermore, it should be noted that the calculated macro-dispersivities did not reach its asymptotic value due to the limitation of the computation domain.

2.2.5. Validation of physical model simulation

Fiori et al. (2015b) have shown that for purely advective transport the mean advective velocity $\bar{U}$ calculated from the results of the transport simulation is equal to the mean value of the component of the Eulerian velocity field parallel to the mean flow direction (U_{mean}). The same as the work of Bianchi and Pedretti (2018) and Jankovic et al. (2017), the accuracy of the numerical simulation is evaluated by checking the equality between the two mean velocities. For each simulation, the mean advective velocity $\bar{U}$ is calculated by

$$\bar{U} = L/\bar{t} \tag{8}$$

where L is the distance between the injection source and the control plane, $\bar{t}$ is the average arrival time of the particles. The components of the Eulerian velocity field at the center of each grid block are obtained by averaging the specific discharge across the numerical grid block interfaces under homogeneous effective porosity. The fluxes are derived from the cell-by-cell balance of the groundwater flow model (MODFLOW). For this work, 50 simulations with different conductivity fields are conducted to assess the equality of $\bar{U}$ and U_{mean}. The rations between $\bar{U}$ and U_{mean} are plotted in Fig. 3 and are all close to one.

Besides, the effective macro-dispersivities from numerical simulations are compared with analytical models of Dagan (1984) and Gelhar and Axness (1983). Dagan (1984) and Gelhar and Axness (1983) derived analytical models of ensemble macrodispersivity for moderate heterogeneity ($\sigma^2_{lnK} < 1$). The same setup of simulation as the work of Fernàndez-Garcia and Gómez-Hernández (2007) is used at first. The effective macro-dispersivities of 50 realizations are compared with Dagan (1984) analytical model. The result is displayed in Fig. 4. Then another 50 simulations with the setup of Section 2.2 are conducted. The effective and apparent macro-dispersivities of 50 realizations are compared with Dagan (1984) and Gelhar and Axness (1983) analytical models in Fig. 5. The effective macro-dispersivity is the averaged of apparent macro-dispersivities from 50 realizations. From both figures, it can be seen that the effective macro-dispersivity generally agrees well with the model of Dagan but has not reached the asymptotic limit of Gelhar. Overall, these results show that the physical model simulations are accurate.

2.3. Architecture of the CNN

Neural networks are a set of algorithms that are composed of some highly interconnected processing elements (neurons) working through a hierarchy of layers (Schmidhuber, 2015). A regular CNN consists of

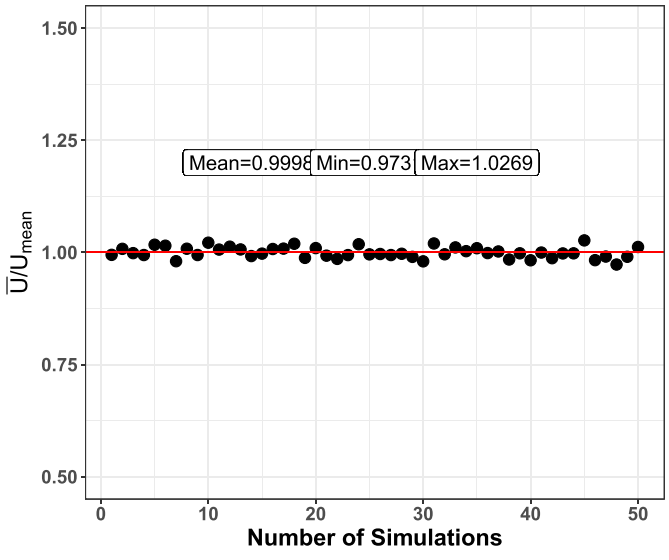

Fig. 3. Comparison of $\bar{U}$ and U_{mean} for 50 simulations.

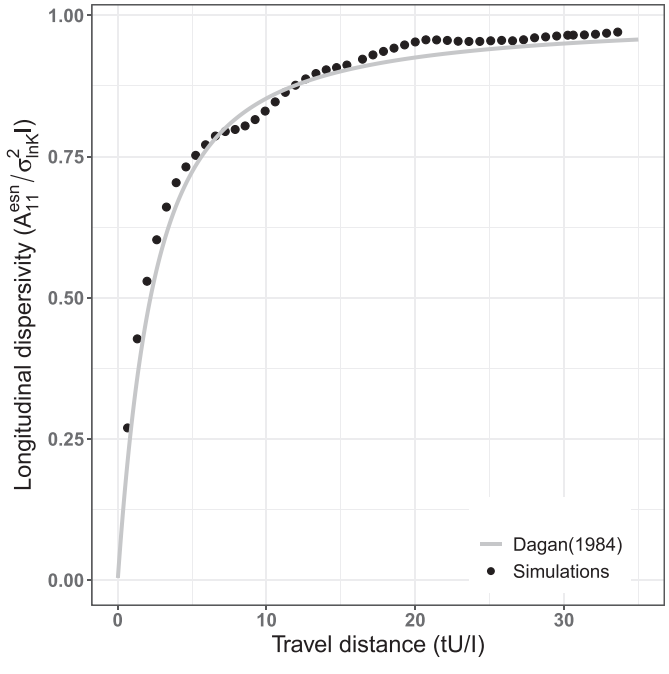

Fig. 4. Comparison of effective macrodispersivity with the Dagan (1984) model ($\sigma^2_{lnK} = 0.1$).

some convolutional layers, pooling layers and fully connected layers (Fig. 1). A convolutional layer is composed of some convolution kernels which are used to compute the feature maps. When a 2D image is inputted, a convolutional layer G is obtained by employing some filters $\tau^q \in \mathrm{R}^{k_i,k_j}$, where q is the number of filters and k is kernel size, to evolve an input pixel or point value $h_{m,n}$ to get the feature value $G^q_{m,n}(h_{m,n})$ at location (m, n) as

$$G^q_{m,n}(h_{m,n}) = f\left(\sum_{i=1}^{k_i}\sum_{j=1}^{k_j}\tau^q_{i,j}h_{m+i,n+j}\right) \tag{9}$$

This results in a convolutional lay G consisting of q feature maps. Two other important parameters for the convolutional layer are the stride and padding. Stride determines the distance between two successive moves of the filter and padding specifies the padding of the borders

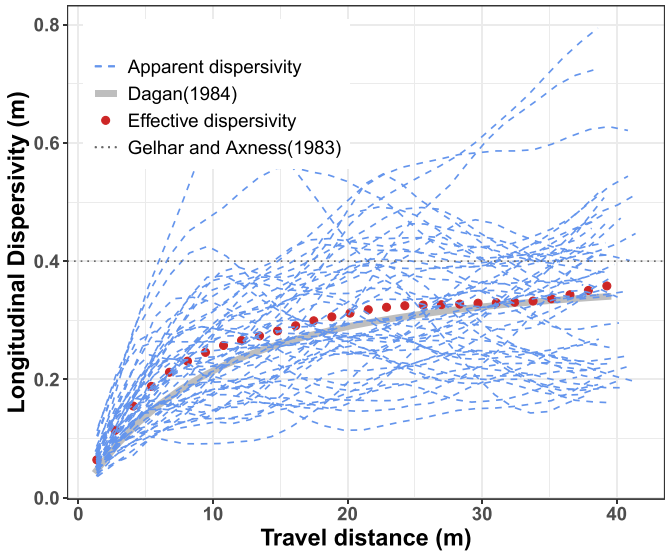

Fig. 5. Comparison of effective and apparent macrodispersivities with the Dagan (1984) and Gelhar and Axness (1983) analytical models ($\sigma^2_{lnK} = 0.1$).

of the input image with zeros for size preservation. Pooling layers combine the outputs of neuron clusters from the convolutional layer into a single neuron, and so the learnable parameters are reduced. In the problem of estimating macro-dispersivity from the conductivity field as concerned in this study, the convolutional layers extract features from an input conductivity field by using small squares of the input field. The feature maps of the conductivity field encompass the correlation structure of conductivity field as well as the higher-order moments. The pooling layers reduce the dimensionality of each feature map but retain the vital information for estimating the macro-dispersivity. The fully connected layers flatten the feature map matrixes and combine these features to create a model with macro-dispersivity as the final output. The CNN architecture proposed here is implemented in the machine learning framework PyTorch (Paszke et al., 2019) and is based on the AlexNet (Krizhevsky et al., 2012). The CNN here includes five convolutional layers, each followed by a batch normalization operation (Krizhevsky et al., 2012), a ReLU activation function and a pooling layer, and three consecutive fully-connected layers (Fig. 1). The numbers of layers, channels and neurons are empirically selected to ensure enough complexity of the neural network and minimize the mean squared error of the estimated macro-dispersivities for the training datasets. The first convolutional layer has 16 feature maps. The first four convolutional layers all have a convolutional kernel of size 7×7 to extract different features from the corresponding input. The last convolutional layer has a convolutional kernel of size 5×5. In the pooling layers, the max-pooling function and a kernel size of 2×2 with strides 2×2 are adopted. The three fully-connected layers all have 256 neurons. The details of the deep neural network architecture are displayed in Table 1. The other details of network architecture and hyperparameter selection of the particular problem considered are presented in Appendix B. The computational cost is of the order of seconds for estimating the macro-dispersivities of 1000 conductivity fields with 140×140 data points by using the trained CNN and is three orders of magnitude lower than the simulations of above-mentioned physical models.

2.4. Performance criteria

The root square error (*RMSE*) and the coefficient of determination (R^2), are used to evaluate the performance of our deep neural network. The results of simulations with random walk particle tracking method are taken as reference values. For *RMSE*, a value of 0 means a perfect fit between the reference values and the estimated values. The smaller *RMSE* value stands for better performance. The definition of *RMSE* is shown below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(A_i^{\text{sim}} - A_i^{\text{CNN}})^2}{n}} \tag{10}$$

where A_i^{sim} is the ith simulated macro-dispersivity in validation/test datasets, A_i^{CNN} is the ith estimated macro-dispersivity by the CNN.

The definition of the coefficient of determination is presented as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(A_i^{\text{sim}} - A_{\text{CNN}}^{\text{mean}})^2}{\sum_{i=1}^{n}(A_i^{\text{sim}} - A_{\text{sim}}^{\text{mean}})^2} \tag{11}$$

where $A_{\text{CNN}}^{\text{mean}}$ represents the average of the estimating macro-dispersivities by the CNN, $A_{\text{sim}}^{\text{mean}}$ is the average of the reference macro-dispersivities in the test dataset. The R^2 values range between 0 and 1, and the value of 1 indicates a perfect correlation.

3. Synthetic experiments setting

Synthetic experiments are conducted to demonstrate the capability of the neural network to estimate the macro-dispersivity by considering different variances of conductivity fields.

3.1. Datasets

We evaluated the CNN using datasets of conductivity fields produced with increasing variances of 0.1, 0.2, 0.3 and 0.5 (called Var0.1, Var0.2, Var0.3, and Var0.5, respectively). The discretized field realizations vary highly with increasing variance, and the intrinsic dimensionality of the random field is 19600 for 140×140 grids which are the cases presented here. These field characteristics create a significant challenge for data-driven models to capture. Our current data includes three sets: the training, validation and test sets. Each training set contains 4000 training data and the validation and test set each consist of 1000 input conductivity fields. All datasets are organized in the form of images. The training set is limited to 4000 samples because each simulation run of the random walk has high-computational cost. A sample costs about 10 min (CPU: Intel Xeon E5-2630 v3 2.4GHz 16 cores RAM: 96GB DDR4 2133MHz ECC). A training set of 4000 samples costs almost a month. Also, the data on field-scale dispersion in real aquifers is even more limited (Zech et al., 2015; Gelhar et al., 1992). Generating more input realizations of hydraulic conductivity using only the training dataset is another critical problem (Shen, 2018). The interest of this work is to validate the potential possibility of building the field-to-dispersivity mapping with the limited training dataset.

3.2. Training

The CNN is trained with Adam (Kingma and Ba, 2014), a variant of stochastic gradient descent, with the loss function being L_2 regularized mean square error(MSE) which is implemented as weight decay in PyTorch. The initial learning rate is 0.001; the weight decay is 0.0005; the batch size is 64. A learning rate scheduler which drops 10 times on plateau of the rooted MSE is also used. The model is trained 200 epochs. The estimating performance is evaluated with validation and test fields that are not included in the training datasets. The macro-dispersivities estimated from the simulations of the random walk particle tracking method are taken as reference values.

3.3. Experiments

At first, CNN is trained with the four datasets of different variances, respectively. The estimation performance of these trained neural networks is tested with the corresponding test set. The estimating capability of the proposed deep neural network is demonstrated, and the influence of variance of conductivity field on estimation accuracy of macro-dispersivity is revealed.

Table 1
8 layers neural network architecture.

Layer	Type	Maps and Neurons	Kernel Stride Padding
0	Input	1 field of 140 × 140 data points	–
1	Convolutional[a]	16 feature maps	7 2 1
2	Convolutional[a]	32 feature maps	7 1 2
3	Convolutional[a]	64 feature maps	7 1 2
4	Convolutional[a]	128 feature maps	7 1 2
5	Convolutional[a]	256 feature maps	5 1 2
6	Fully connected[b]	256 neurons	–
7	Fully connected[b]	256 neurons	–
8	Fully connected	256 neurons	–

[a] The convolutional layers are followed by a bacth normalization, a ReLU activation and a max pooling.
[b] The fully connected layers are followed by a ReLU activation.

Table 2
Setup of experiments investigated in this work.

Experiment	Training set	Test set[a]
1	Var0.1	Var0.1
	Var0.2	Var0.2
	Var0.3	Var0.3
	Var0.5	Var0.5
2	Var0.2	Var0.1
	Var0.3	Var0.1 Var0.2
	Var0.5	Var0.1 Var0.2 Var0.3
3	Var0.1	Var0.2 Var0.3 Var0.5
	Var0.2	Var0.3 Var0.5
	Var0.3	Var0.5

[a] Each test set has 1000 samples which are not in the training set.

Next, the performance of the neural networks trained with conductivity fields of relatively large variance (training sets of Var0.5, Var0.3 and Var0.2) are tested with the test sets of conductivity fields of relatively small variance (test sets of Var0.1, Var0.2 and Var0.3). It is found that the neural networks trained by large variances can estimate conductivity fields with small variances.

Finally, the estimating capability and universality of the trained neural networks (training set of Var0.1, Var0.2 and Var0.3) are verified by using test datasets with large variances (test sets of Var0.2, Var0.3 and Var0.5). The universality of the neural network to estimate conductivity fields with relatively large variances is validated. Detailed setup about the three experiments is presented in Table 2.

4. Results and discussion

4.1. Experiment 1: Training CNN with datasets of different variances

Fig. 6 displays the training process with epochs for the training sets of different variances. It is seen that the validation *RMSE* increases with increasing variance and the validation R^2 decreases with increasing variance. However, the CNN trained by Var0.5 still has a decent R^2 value 0.73 on the validation set. These trained neural networks are then tested with the test set of the same variance as the training set. All test sets are consists of 1000 samples of conductivity fields. The test results are illustrated in Fig. 7. It should be noted that the values of *RMSE* and R^2 are calculated from the test set of 1000 samples, but only 50 samples of conductivity field which are randomly sampled from test set are displayed. From Fig. 7, it is found that the neural networks trained with different variances all have achieved accurate estimation of macro-dispersivity in terms of *RMSE* and R^2. For all the four cases, most macro-dispersivities estimated from the neural networks fall in the error range of ± 20% (marked as baby blue shaded regions in figures). The CNN trained by the dataset Var0.1 has smallest *RMSE*=0.05, and largest $R^2 = 0.869$ and the CNN trained by dataset Var0.5 has largest *RMSE* = 0.375 and smallest $R^2 = 0.72$. From the two criteria, it can be seen that the estimation performance generally gets worse with the increasing variance of dataset. However, for Var0.1 and Var0.2, the decreasing of estimation performance is not obvious. On the whole, all the four trained CNNs have a satisfactory estimation of macro-dispersivity with the limited size of the training set.

From this experiment, it is found that CNN can achieve acceptable estimations of macro-dispersivities for conductivity fields with weak heterogeneity ($\sigma^2_{lnK} \leq 0.5$) and increasing variances of conductivity fields would make a functional mapping difficult for the CNN with the limited size of training datasets. Larger training datasets may introduce better estimation performance.

4.2. Experiment 2: Test trained CNNs with test sets of small variances

In experiment 2, we will use the trained CNNs in experiment 1 to estimate macro-dispersivities of conductivity fields with smaller variances than the training set. For the neural network trained by conductivity fields of $\sigma^2_{lnK} = 0.2$, an only test set of Var0.1 are used (Fig. 8). In Fig. 8, nearly all estimated macro-dispersivities are within ± 20% error range of the reference values. The large value of R^2 (0.831) confirms the results and the *RMSE* is close to the results of Fig. 7(a1). The CNN trained by Var0.2 can give accurate estimates of the test set of Var0.1 as same as the test set of Var0.2 [Fig. 7(b1)].

Fig. 9 shows the results of using CNN trained by Var0.3 to estimate macro-dispersivities for test sets of Var0.1 and Var0.2. For the test set of Var0.1 [Fig. 9(a1),(b1)], underestimation of macro-dispersivities can be observed for most data points. Nevertheless, there are still at least half of the data points are within the error range. The $R^2 = 0.496$ validates the estimation performance. For the test set of Var0.2 [Fig. 9(a2),(b2)], the CNN has achieved quite good estimation performance. Most estimated macro-dispersivities are within the error range of the reference values, and the R^2 is 0.829. The *RMSE* (0.114) is almost equal to the result (0.110) from the neural network trained by Var0.2. These results show that the CNN trained by Var0.3 can estimate test set Var0.2 and Var0.1. However, the estimation performance drops with the increasing disparity of variances. The results of using CNN trained by Var0.5 to estimate macro-dispersivities for the test sets of Var0.1, Var0.2 and Var0.3 are displayed in Fig. 10. For the test fields of Var0.1 [Fig. 10(a1),(b1)], more than 50% of estimations fall within the ± 20% error range. While for the test set fields of Var0.2 and Var0.3, most estimations fall within the ± 20% error range. The R^2 values for the test sets generally increase with increasing variances of conductivity fields. Specifically, for the test set of Var0.3 (R^2=0.74), the CNN trained by Var0.5 can provide almost the same estimation performance as the test set of Var0.5 (R^2=0.72). Similar to the results of Fig. 9, the estimation performance drops with the increasing difference of variances.

In summary, experiment 2 firstly demonstrates that the CNN trained by fields with a specific variance has some ability to estimate conduc-

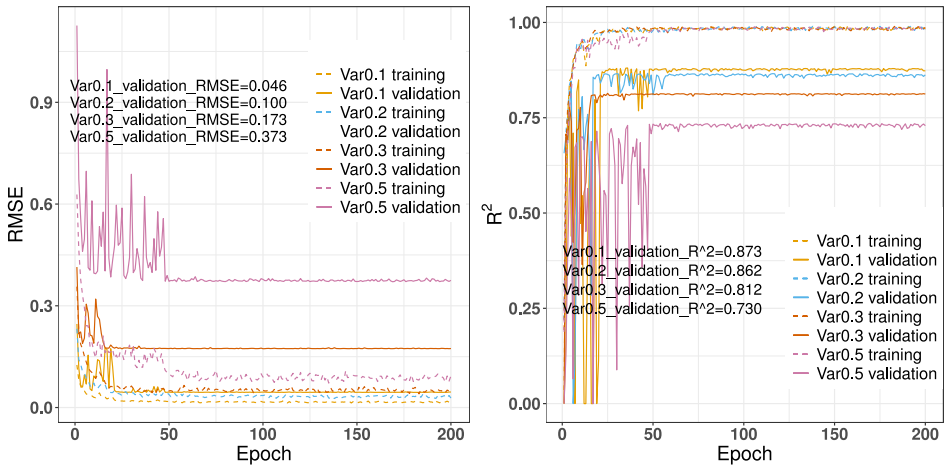

Fig. 6. Training process of the neural network with the training sets of different variances.

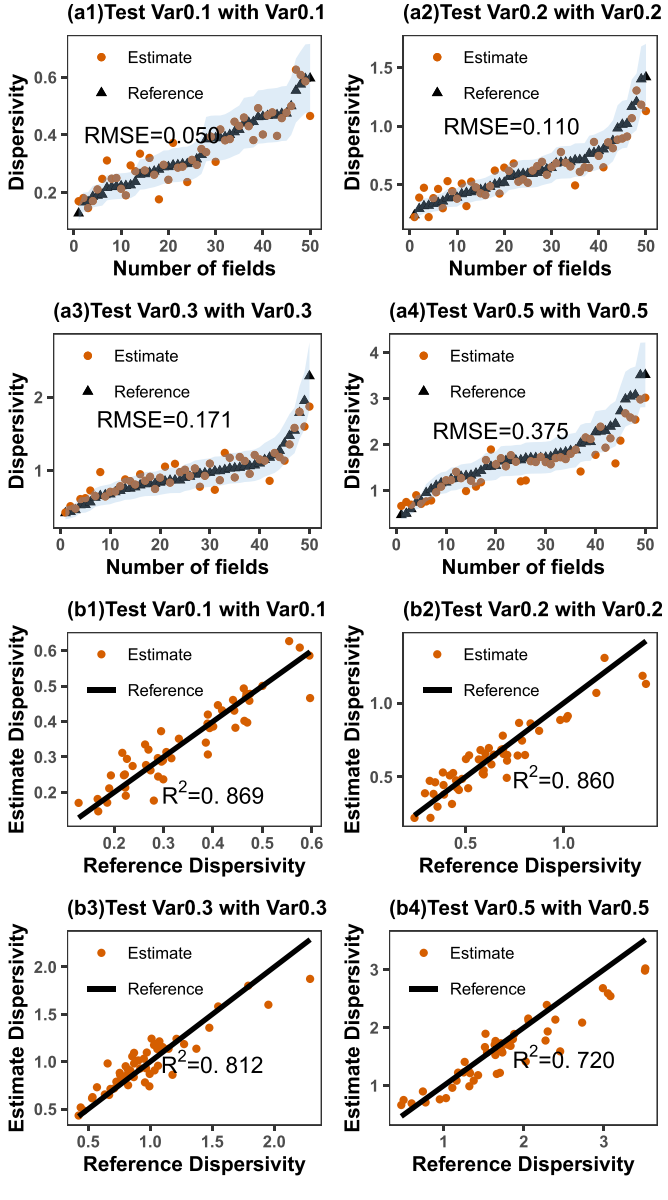

Fig. 7. The comparison of the estimate and the reference macro-dispersivities for training datasets of different variances.

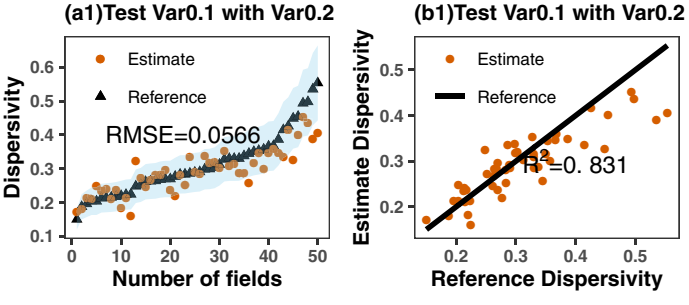

Fig. 8. Estimating macro-dispersivities of fields with $\sigma^2_{lnK} = 0.1$ using neural network model trained by fields with $\sigma^2_{lnK} = 0.2$.

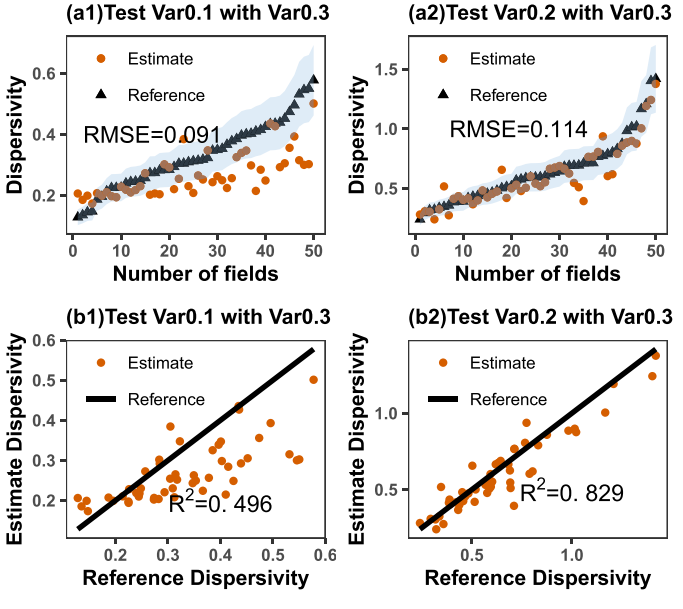

Fig. 9. Estimating macro-dispersivities of fields with σ^2_{lnK}=0.1 and 0.2 using neural network model trained by fields with $\sigma^2_{lnK} = 0.3$.

tivity fields with the other smaller variances and the estimating performance decreases with increasing differences of variances. Secondly, the neural network trained by large variance has a stronger capacity in estimating macro-dispersivities of conductivity fields with small variances. More specifically, it seems that the CNN trained by Var0.5 has better universality than the CNN trained by Var0.3. Because, the CNN trained

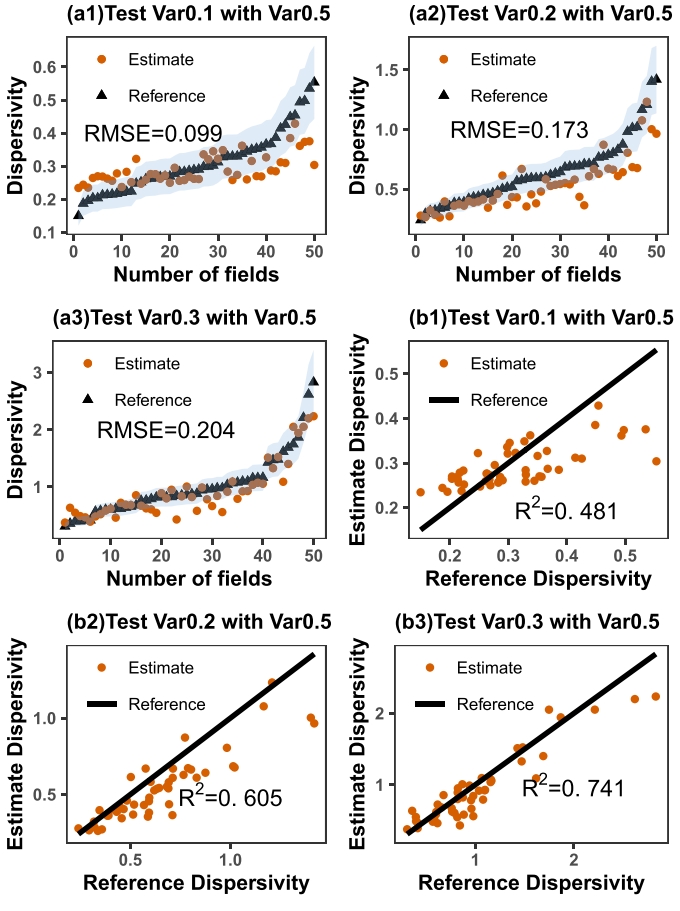

Fig. 10. Estimating macro-dispersivities of fields with σ^2_{lnK} =0.1, 0.2 and 0.3 using neural network model trained by fields with σ^2_{lnK} = 0.5.

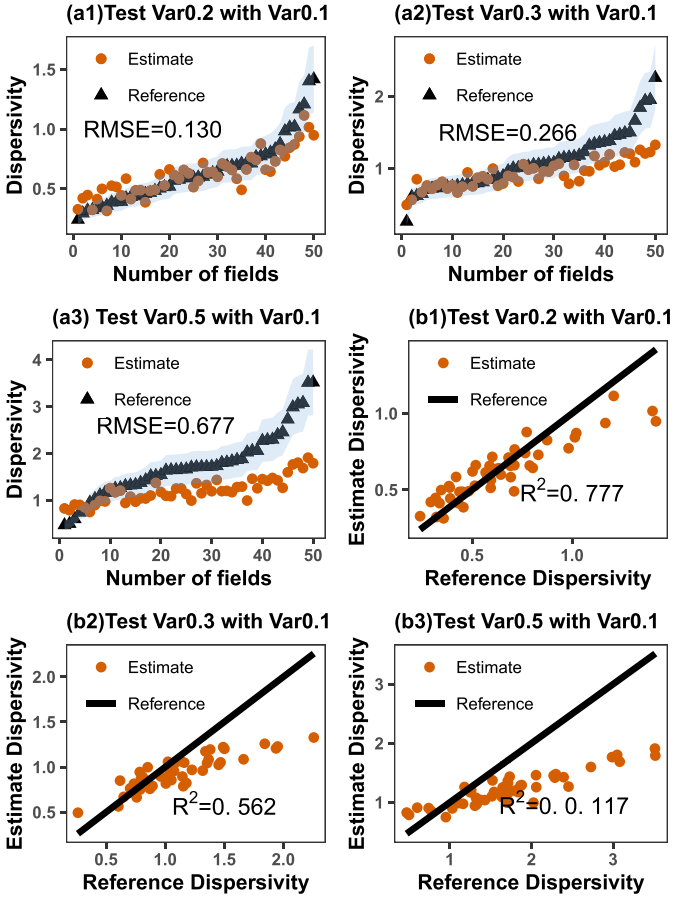

Fig. 11. Estimating macro-dispersivities of fields with σ^2_{lnK} =0.2, 0.3 and 0.5 using neural network model trained by fields with σ^2_{lnK} = 0.1.

by Var0.5 works much better in estimating test set of Var0.3 than using the CNN trained by Var0.3 to estimate test set of Var0.1, although the disparity between variances of train set and test set are 0.2 for the two cases. Furthermore, the CNN trained by Var0.5 still has better performance in estimating test set of Var0.2 than estimating test set of Var0.1 with the CNN trained by Var0.3.

4.3. Experiment 3: Test trained CNNs with test sets of large variances

In experiment 3, we will use the trained CNNs in experiment 1 to estimate macro-dispersivities of conductivity fields with larger variances than the training set. Fig. 11 illustrates the results of estimating macro-dispersivities of conductivity fields with σ^2_f =0.2, 0.3 and 0.5 by using the CNN trained by train set of σ^2_f=0.1. For the test set of Var0.2 [Fig. 11(a1),(b1)], almost all data points fall within the error range. The neural network has achieved fairly good estimations (R^2 = 0.777). With increasing variance of test set [Fig. 11(a2),(b2)], underestimations of macro-dispersivities happen for large reference values. For the test set of Var0.5 [Fig. 11(a3),(b3)], the underestimations of macro-dispersivities for large reference values become prominent. Correspondingly, the values of R^2 decrease with increasing variances.

In Fig. 12, the neural network trained by Var0.2 is used to estimate the test set of Var0.3 and Var0.5. Similar to Fig. 11, the estimation performance of the neural network decreases with increasing variance. Comparing Fig. 12 with Fig. 11(a2,a3,b2,b3), it is found that, the neural network trained by Var0.2 has stronger estimating ability than the neural network trained by Var0.1. For example, the neural network trained by Var0.1 generally underestimates the macro-dispersivities for the test set of Var0.5 (R^2 = 0.117), while the neural network trained by

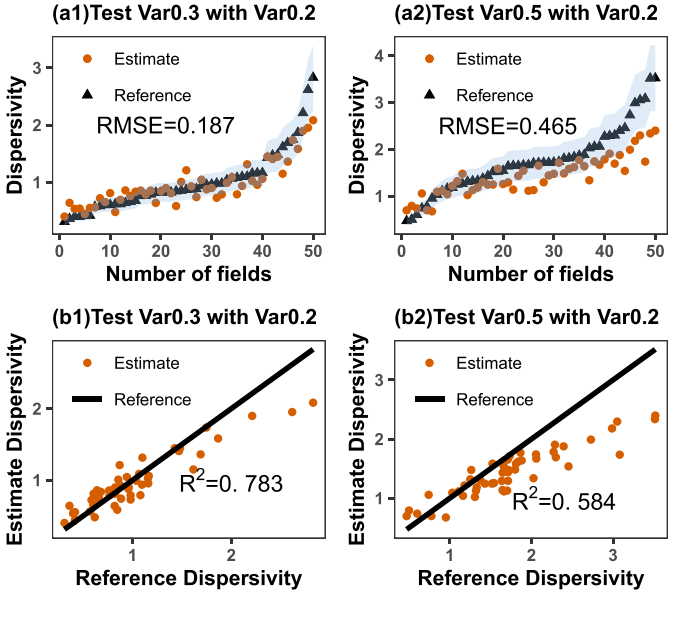

Fig. 12. Estimating macro-dispersivities of fields with σ^2_{lnK} =0.3 and 0.5 using neural network model trained by fields with σ^2_{lnK} = 0.2.

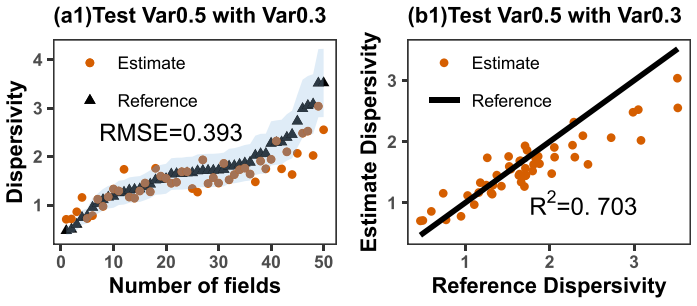

Fig. 13. Estimating macro-dispersivities of fields with σ^2_{lnK} =0.5 using neural network model trained by fields with $\sigma^2_{lnK} = 0.3$.

Var0.2 can get more satisfactory estimations for the test set of Var0.5 ($R^2 = 0.584$). In Fig. 13, the neural network trained by Var0.3 is used to estimate macro-dispersivities for the test set of Var0.5. The *RMSE* and R^2 are almost consistent with the results of Fig. 7(a4,b4) in which the macro-dispersivities for the test set of Var0.5 are estimated by the neural network trained with the train set of Var0.5 ($RMSE = 0.375, R^2 = 0.72$). The estimating performance of the neural network trained by Var0.3 is better than the neural networks trained by Var0.2 and Var0.1. Comparing Fig. 13 with Fig. 11(a3,b3) and Fig. 12(a2,b2), the R^2 of macro-dispersivities for test set of Var0.5 from the neural network trained by Var0.3 is the largest ($R^2 = 0.703$).

Generally speaking, experiment 3 illustrates that the neural network trained by conductivity fields with a specific variance can be used to estimate conductivity fields with the larger variances and the estimating performance decreases with increasing variances of conductivity fields. The neural network which trained by large variance may have a large capacity in estimating macro-dispersivity.

4.4. Features from the CNN

In this section, results from the above three experiments are summarized, and some discussion about the features extracted by the CNN are presented.

Fig. 14 displays the comparison of R^2 from the above three experiments. Generally, the CNN trained by conductivity fields with relatively large variances can achieve better performance on estimations of macro-dispersivities for conductivity fields with relatively small variances. The neural network trained by highly heterogeneous fields seems to have a high ability to extract features of heterogeneity. In Fig. 14(a4), the neural network trained by Var0.5 has much better performance than the neural network trained by Var0.1, although the neural network trained by Var0.5 has worse performance in its own test set. However, there seems to be an exception. From Fig. 14(a3), the neural network trained by Var0.1 has better performance than the neural network trained by Var0.3. We think part of the reason may be that the neural network of Var0.1 has been better trained than the neural network of Var0.3. Besides, the difference between conductivity fields of Var0.1 and Var0.3 is relatively small, which makes the two trained neural networks have similar abilities to extract features of heterogeneity.

Overall, CNN can provide estimations of macro-dispersivities with the limited size of training data. It is mainly because the macro-dispersivity is indeed a macroscopic description of the impact of the conductivity field (aquifer structure) on solute transport and a functional mapping from the conductivity field to macro-dispersivity is expected to exist. Some insights on how the CNN estimates macro-dispersivity form conductivity field and why the CNN has universality in estimating conductivity fields with different degree of heterogeneity are discussed below. In the CNN, feature maps are obtained by doing convolution on the input conductivity field with filters (convolutional weights). These filtered results are usually visualized to illustrate what CNN learns. Sixteen typical filtered results from the first convolutional layer are presented in Fig. 15(b) for estimating macro-dispersivity of the conductivity field shown in Fig. 15(a). It should be noted that pixels of these features have been mapped to the range (0,1) with Sigmoid function and are multiplied by the mean value of the ln$K$ field. The feature maps show how the trained CNN analyzes a conductivity field to make the estimation. It can be seen that in Fig. 15(b) the CNN attempts to identify different features of the heterogeneous conductivity field. Apparently, we can see that, feature 1, 3, 10, 13, 16 mainly capture the distributions of relatively high ln$K$ values of the original ln$K$ field. While feature 2, 4, 7, 11, 12, 14 mainly capture the distributions of relatively low ln$K$ values of the original ln$K$ field. Other features also reflect some other heterogeneous patterns that are not prominent. By capturing the multiple hierarchic features of the heterogeneity in the conductivity field, CNN is able to build a functional mapping between macro-dispersivities and conductivity fields. The ability to extract these features may be the reason why a CNN trained by a specific variance of conductivity fields can be used to estimate fields of the other variances because CNN has already learned how to extract different heterogeneous patterns. It should be noted that the features presented in Fig. 15 are only for a qualitative visualization of how the neural network learns the heterogeneity from the conductivity fields. In practice (mathematically definition of the learning process), the parameters (e.g. convolutional weights, biases) of the CNN are learned by minimizing the estimation error of the training datasets with back-propagation and preventing overfitting by imposing sparsity.

5. Conclusion

Building a relationship between macro-dispersivity and conductivity field is of significant practical importance to characterize the transport behavior. In this work, a direct and efficient map between macro-dispersivity and conductivity field is built by the convolutional neural network based on synthetic field datasets. The effectiveness of CNN is validated through various training and test datasets. The universality of CNN is also investigated, that is to say, whether or not the CNN trained by a specific group of conductivity fields can be used to estimate the other types of conductivity fields. The following conclusions can be drawn from these experiments:

(1) The estimating performance of CNN generally drops with increasing variances of conductivity fields (increasing heterogeneity) for the given size of training datasets (4000 fields) and data points (140×140). There are many factors that may influence the performance of CNN-based model, such as CNN architecture, input data points and dataset size. The effects of these factors need further research.
(2) The CNN trained by conductivity fields with a specific variance has universality in estimating macro-dispersivity to a certain extent because a well trained CNN will have the capacity to extract different patterns of heterogeneity. Consequently, the trained CNN can extract some standard heterogeneous features of conductivity fields for estimating macro-dispersivities.
(3) Furthermore, the universality of the trained CNN decreases with the increasing disparity between variances of conductivity fields in training set and test set. And the CNN trained by conductivity fields with relatively large variances can have stronger universality of estimating. The reason may be that the neural networks trained by conductivity fields with large variances (high heterogeneity) can extract more heterogeneous patterns than the neural network trained by conductivity fields with small variances (low heterogeneity) since conductivity fields with large variances have more complicated spatial heterogeneity than conductivity fields with small variances. This guess also needs to be further verified.
(4) In general, the deep neural network is a very promising approach in building direct mapping between complicated subsurface structure and solute transport behavior.

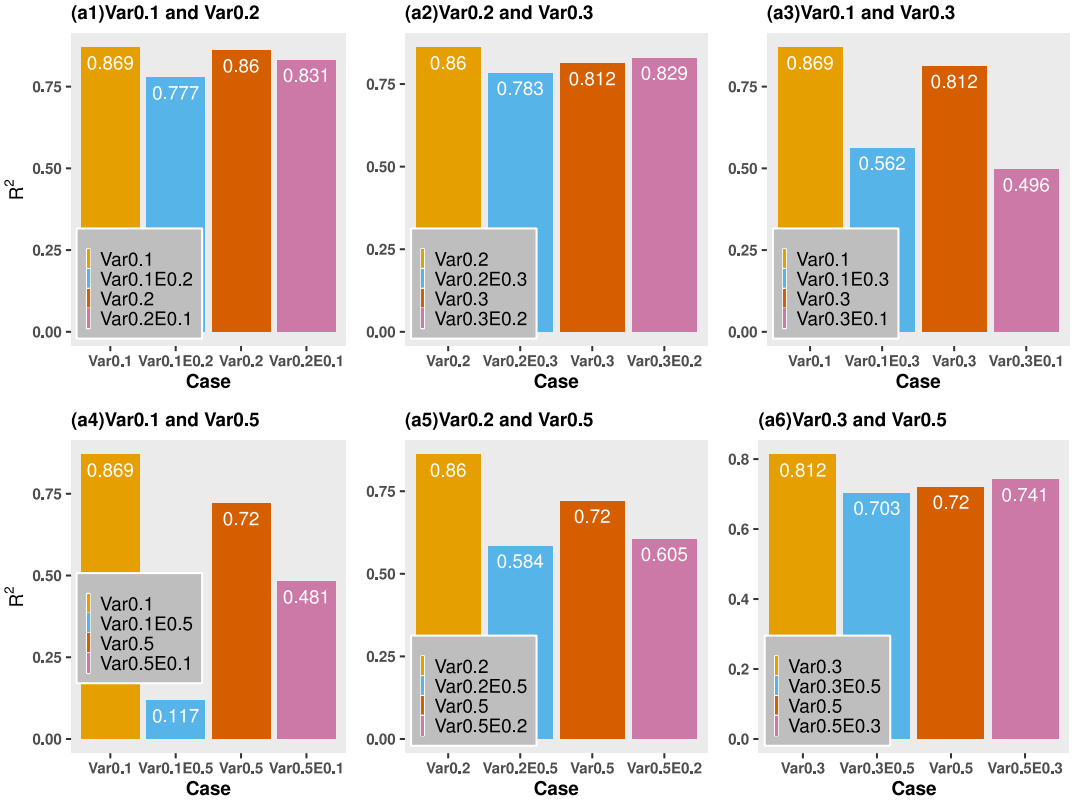

Fig. 14. Comparison of R^2 for different cases; Var0.1 represents the results of using the CNN trained by the training set of Var0.1 to estimate conductivity fields in the test set of Var0.1; Var0.1E0.2 represents the results of using the CNN trained by the training set of Var0.1 to estimate conductivity fields in the test set of Var0.2; The other labels have similar meanings.

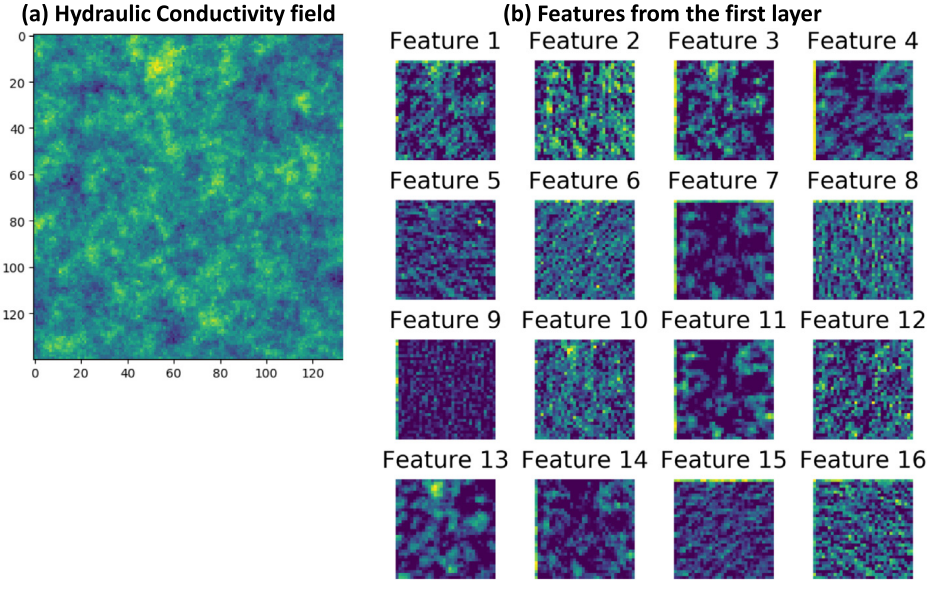

Fig. 15. Illustration of features learned by the CNN, including (a) an input lnK field and (b) corresponding features obtained from the first convolutional layer.

As the first attempt at interpreting macro-dispersivity from hydraulic conductivity field with deep neural network, there still exist some shortcomings or questions that need to be solved in following work. The conductivity fields used in this work are synthetic and do not correspond to any real-world data. Hence, at this time, we cannot project the performance of CNN with real conductivity fields. This work is limited to weak heterogeneity, and the macro-dispersivity studied in this work varied by only one order of magnitude, because of the synthetic nature of the fields. Real aquifers have macro-dispersivities covering several orders of magnitudes. Moreover, the requisite size of the training datasets is reasonably large. Getting such a large size of the dataset in real-world problems is very difficult. As a data-driven model, the estimation of CNN is limited to the information provided by the training dataset.

These deficiencies may be improved by adding physical information into the neural network (Raissi et al., 2017; 2019) or using more complicated neural networks. The main goal of this work is on the

proof-of-concept for adopting image recognition and specifically CNN in estimating macro-dispersivity of heterogeneous formation. The deep neural network should have a high potential in fast estimating macro-dispersivity from conductivity field for engineering applications. However, we must note that there have also been some advancements in the computational performance of physically-driven models. For example, Rizzo et al. (2019) have developed the parallel random walk particle tracking code *PAR*^{2}, an efficient GPU parallelized particle tracking code. The GPU-accelerated solute transport simulator can drastically reduce the total simulation time of the random walk particle tracking method.

Over all, the deep learning method provides another way to interpret macro-dispersivity directly from the information of subsurface structure and has the potential of leading to high-impact developments.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Editors and reviewers are greatly acknowledged for providing extremely helpful comments. This study was supported by the National Natural Science Foundation of China Grant 51861125202 and 51629901. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University. The code and data used in this work will be made available at https://github.com/chinazzk/CNN-for-Macrodispersivity upon publication of this manuscript.

Appendix A. Choosing the number of particles

In order to choose an appropriate number of particles, several RW3D simulations are carried out. The setup of transport simulation is the same as Section 2.2. By changing the number of particles for trace source, six trace tests are conducted. The calculated apparent dispersivities from control planes are displayed in Fig. A.16. It can be seen that 1000 particles are enough to give accurate estimations of apparent macro-dispersivities.

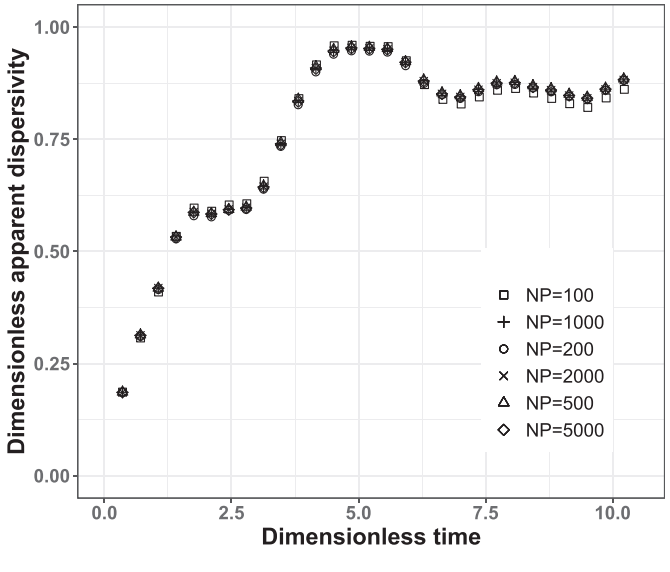

Fig. A.16. The dispersivities from trace tests with different particle numbers.

Appendix B. Optimization of the network hyperparameters

B1. Architecture of CNN

The basic structure of CNN is from a Tutorial of Pytorch in GitHub (https://github.com/yunjey/pytorch-tutorial).

The original basic CNN, which has convolutional kernel 5×5, is designed for MNIST dataset and does not work for the presented problem. The dataset Var0.1 is used here to evaluate different neural network hyperparameters. The training set has 4000 samples, and the validation set has 1000 samples. A lot of training experiments are conducted to modify the architecture of CNN for applying to the problem considered. The network hyperparameters are then empirically optimized by following the general architecture of AlexNet. Firstly, we experiment with different sizes of the convolutional kernel (3×3, 5×5 and 7×7) in the basic CNN. The convolutional kernel 7×7 has the best performance. And then three architectures of CNN, which have a different number of convolutional layers, are designed for the problem (Fig. B.17). The kernel size of last convolutional layer in CNN-5Lays is 5×5 due to the small size of the feature map from the fourth convolutional layer. The three fully connected layers all have the same number of neurons. The fully connected layers in CNN-3Lays has 3136 neurons. The fully connected layers in CNN-4Lays and CNN-5Lays has 512 and 256 neurons, respectively. The number of neurons in fully connected layer is just the number of neurons in the last convolutional layer for each CNN. The loss function is taken as the regularized MSE and the validation metric as the R^2-score in Eq. (11). Adam optimizer is used for training 200 epochs, with learning rate 0.001 and a plateau scheduler on the test *RMSE* in Eq. (10). The batch size is 64. The weight decay is set at 0.0005. The three networks displayed in Fig. B.17 are validated on the validation set. The variation of *RMSE* and R^2-scores with epoch are displayed

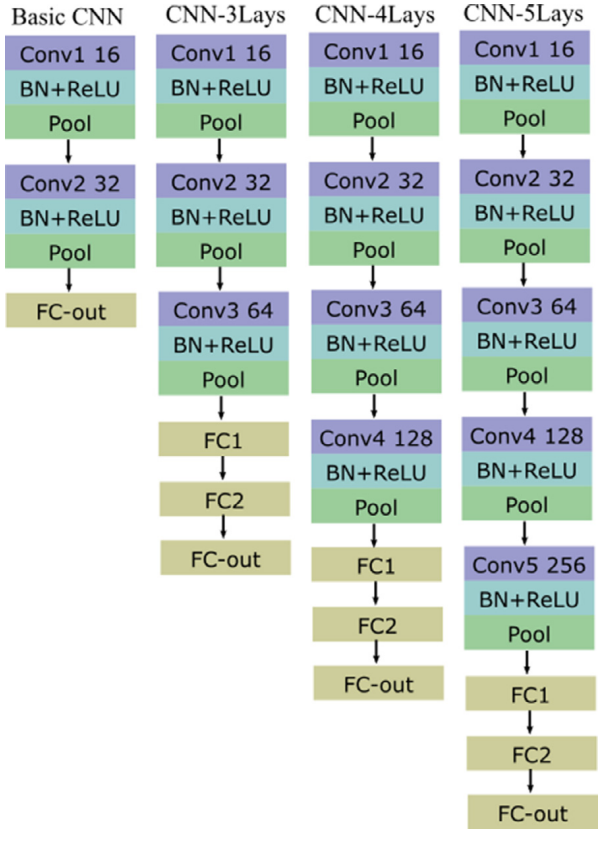

Fig. B.17. Four architectures of CNN; 3Lays, 4Lays and 5Lays indicate the number of convolutional layers in the CNN; Except for the output layer, the fully connected layers are all followed by a ReLU activation function.

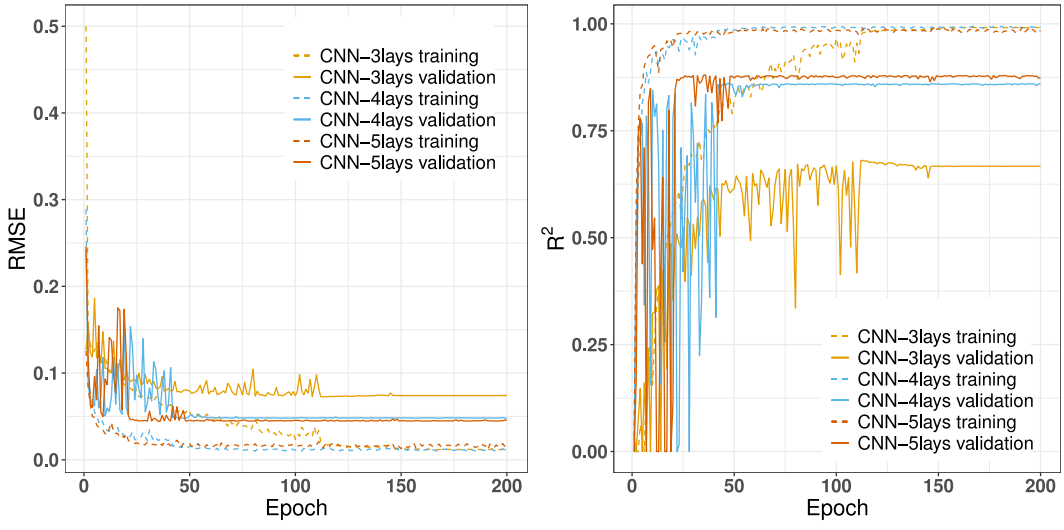

Fig. B.18. Training process for different neural networks.

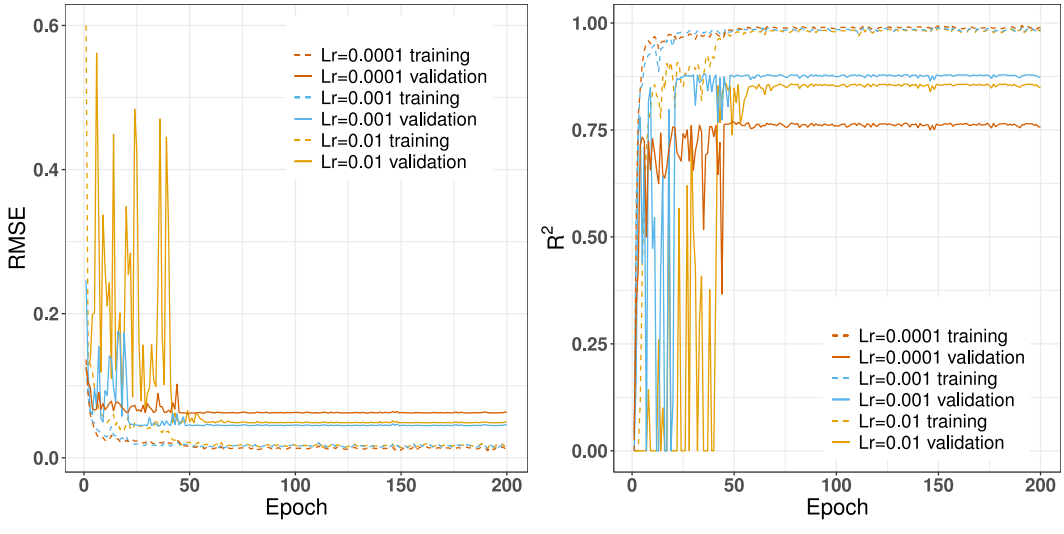

Fig. B.19. Training process for different learning rates.

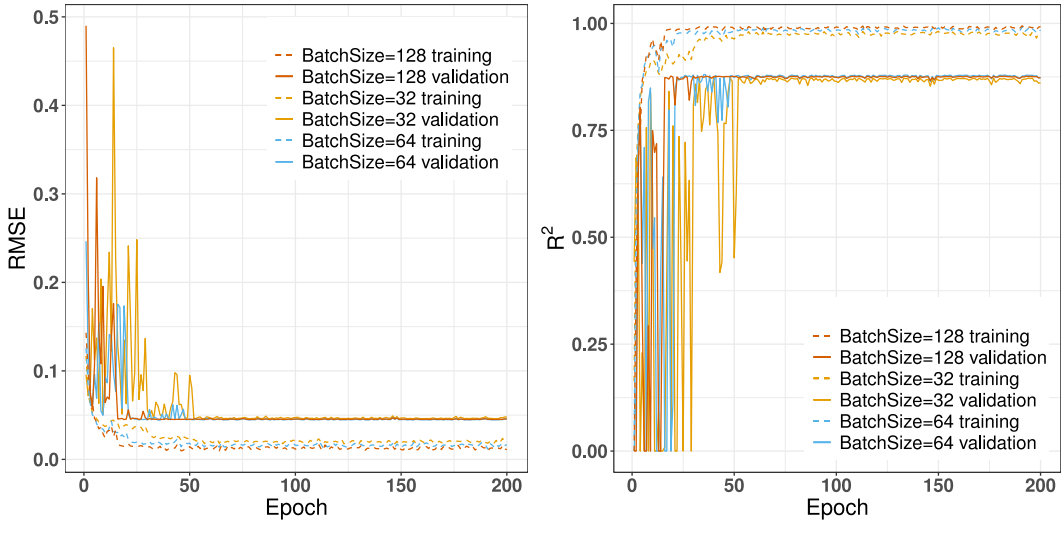

Fig. B.20. Training process for different batch sizes.

in Fig. B.18. In Fig. B.18, CNN-5Lays has the best performance on the validation set. Besides, we have also compared the CNN-5Lays with an 18-layer ResNet (He et al., 2016). The CNN-5Lays has also illustrated a better performance on the validation set. Estimation and learning curve for the selected network CNN-5Lays are presented in Section 4.1.

B2. Hyperparameters for training process

After the selection of CNN architecture, we have tried many configurations of initial learning rate, weight decay and batch size. At last, the learning rate is set as 0.001, weight decay is set as 0.0005 and batch size is set as 64. By trial and error, this is the best configuration that we can find. In Fig. B.19, the different performances of three learning rates are displayed. It can be seen that the learning rate 0.001 has the best estimation performance on the validation dataset. In Fig. B.20, the different performances of three batch sizes are displayed. The impact of batch size on estimation performance is minimal.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2020.103545.

References

Aris, R., 1958. On the dispersion of linear kinematic waves. Proceed. R. Soc. Lond. Series A Math. Phys. Sci. 245 (1241), 268–277.

de Barros, F., Bellin, A., Cvetkovic, V., Dagan, G., Fiori, A., 2016. Aquifer heterogeneity controls on adverse human health effects and the concept of the hazard attenuation factor. Water Resour. Res. 52 (8), 5911–5922. https://doi.org/10.1002/2016WR018933.

Bellin, A., Rubin, Y., 1996. Hydro_gen: a spatially distributed random field generator for correlated properties. Stochastic Hydrol. Hydraul. 10 (4), 253–278. https://doi.org/10.1007/BF01581869.

Bianchi, M., Pedretti, D., 2017. Geological entropy and solute transport in heterogeneous porous media. Water Resour. Res. 53 (6), 4691–4708. https://doi.org/10.1002/2016WR020195.

Bianchi, M., Pedretti, D., 2018. An entrogram-based approach to describe spatial heterogeneity with applications to solute transport in porous media. Water Resour. Res. 54 (7), 4432–4448. https://doi.org/10.1029/2018WR022827.

Cang, R., Li, H., Yao, H., Jiao, Y., Ren, Y., 2018. Improving direct physical properties prediction of heterogeneous materials from imaging data via convolutional neural network and a morphology-aware generative model. Comput. Mater. Sci 150, 212–221. https://doi.org/10.1016/j.commatsci.2018.03.074.

Cang, R., Xu, Y., Chen, S., Liu, Y., Jiao, Y., Ren, M.Y., 2017. Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design. J. Mech. Des. 139 (7), 71404.

Chan, S., Elsheikh, A.H., 2017. Parametrization and generation of geological models with generative adversarial networks. arXiv: 1708.01810.

Cvetkovic, V., Fiori, A., Dagan, G., 2016. Tracer travel and residence time distributions in highly heterogeneous aquifers: coupled effect of flow variability and mass transfer. J. Hydrol. (Amst.) 543, 101–108. https://doi.org/10.1016/j.jhydrol.2016.04.072.

Dagan, G., 1984. Solute transport in heterogeneous porous formations. J. Fluid Mech. 145, 151–177. https://doi.org/10.1017/S0022112084002858.

Dagan, G., 1989. Flow and transport in porous formations. Springer Berlin.

Fadili, A., Ababou, R., Lenormand, R., 1999. Dispersive particle transport: identification of macroscale behavior in heterogeneous stratified subsurface flows. Math. Geol. 31 (7), 793–840. https://doi.org/10.1023/A:1007572700358.

Fernàndez-Garcia, D., Gómez-Hernández, J., 2007. Impact of upscaling on solute transport: traveltimes, scale dependence of dispersivity, and propagation of uncertainty. Water Resour. Res. 43 (2). https://doi.org/10.1029/2005WR004727.

Fernàndez-Garcia, D., Illangasekare, T.H., Rajaram, H., 2005. Differences in the scale-dependence of dispersivity estimated from temporal and spatial moments in chemically and physically heterogeneous porous media. Adv. Water Resour. 28 (7), 745–759. https://doi.org/10.1016/j.advwatres.2004.12.011.

Fernàndez-Garcia, D., Trinchero, P., Sanchez-Vila, X., 2010. Conditional stochastic mapping of transport connectivity. Water Resour. Res. 46 (10). https://doi.org/10.1029/2009WR008533.

Fiori, A., 2014. Channeling, channel density and mass recovery in aquifer transport, with application to the made experiment. Water Resour. Res. 50 (12), 9148–9161. https://doi.org/10.1002/2014WR015950.

Fiori, A., Bellin, A., Cvetkovic, V., de Barros, F., Dagan, G., 2015. Stochastic modeling of solute transport in aquifers: from heterogeneity characterization to risk analysis. Water Resour. Res. 51 (8), 6622–6648. https://doi.org/10.1002/2015WR017388.

Fiori, A., Cvetkovic, V., Dagan, G., Attinger, S., Bellin, A., Dietrich, P., Zech, A., Teutsch, G., 2016. Debatesstochastic subsurface hydrology from theory to practice: the relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. what is characterization and stochastic theory good for? Water Resour Res 52 (12), 9228–9234. https://doi.org/10.1002/2015WR017525.

Fiori, A., Jankovic, I., 2012. On preferential flow, channeling and connectivity in heterogeneous porous formations. Math. Geosci. 44 (2), 133–145. https://doi.org/10.1007/s11004-011-9365-2.

Fiori, A., Zarlenga, A., Gotovac, H., Jankovic, I., Volpi, E., Cvetkovic, V., Dagan, G., 2015. Advective transport in heterogeneous aquifers: are proxy models predictive? Water Resour. Res. 51 (12), 9577–9594. https://doi.org/10.1002/2015WR017118.

Fiori, A., Zarlenga, A., Jankovic, I., Dagan, G., 2017. Solute transport in aquifers: the comeback of the advection dispersion equation and the first order approximation. Adv. Water Resour. 110, 349–359. https://doi.org/10.1016/j.advwatres.2017.10.025.

Freixas, G., Fernàndez-Garcia, D., Sanchez-Vila, X., 2017. Stochastic estimation of hydraulic transmissivity fields using flow connectivity indicator data. Water Resour. Res. 53 (1), 602–618. https://doi.org/10.1002/2015WR018507.

Gelhar, L.W., 1993. Stochastic subsurface hydrology. Prentice-Hall.

Gelhar, L.W., Axness, C.L., 1983. Three-dimensional stochastic analysis of macrodispersion in aquifers. Water Resour. Res. 19 (1), 161–180. https://doi.org/10.1029/WR019i001p00161.

Gelhar, L.W., Welty, C., Rehfeldt, K.R., 1992. A critical review of data on field-scale dispersion in aquifers. Water Resour. Res. 28 (7), 1955–1974. https://doi.org/10.1029/92WR00607.

Goltz, M.N., Roberts, P.V., 1987. Using the method of moments to analyze three-dimensional diffusion-limited solute transport from temporal and spatial perspectives. Water Resour. Res. 23 (8), 1575–1585. https://doi.org/10.1029/WR023i008p01575.

Guo, J., Lu, W., Yang, Q., Miao, T., 2019. The application of 0–1 mixed integer nonlinear programming optimization model based on a surrogate model to identify the groundwater pollution source. J. Contam. Hydrol. 220, 18–25. https://doi.org/10.1016/j.jconhyd.2018.11.005.

Han, Z., Ma, H., Shi, G., He, L., Wei, L., Shi, Q., 2016. A review of groundwater contamination near municipal solid waste landfill sites in china. Sci. Total Environ. 569, 1255–1264. https://doi.org/10.1016/j.scitotenv.2016.06.201.

Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. Modflow-2000, the u. s. geological survey modular ground-water model-user guide to modularization concepts and the ground-water flow process. Open-file Report. U. S. Geological Survey (92) 134.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Henri, C.V., Fernàndez-Garcia, D., 2014. Toward efficiency in heterogeneous multispecies reactive transport modeling: a particle-tracking solution for first-order network reactions. Water Resour. Res. 50 (9), 7206–7230. https://doi.org/10.1002/2013WR014956.

Henri, C.V., Fernàndez-Garcia, D., 2015. A random walk solution for modeling solute transport with network reactions and multi-rate mass transfer in heterogeneous systems: impact of biofilms. Adv. Water Resour. 86, 119–132. https://doi.org/10.1016/j.advwatres.2015.09.028.

Jankovic, I., Maghrebi, M., Fiori, A., Dagan, G., 2017. When good statistical models of aquifer heterogeneity go right: the impact of aquifer permeability structures on 3d flow and transport. Adv. Water Resour. 100, 199–211. https://doi.org/10.1016/j.advwatres.2016.10.024.

Kendall, M., Stuart, A., 1977. The advanced theory of statistics. 3d ed. vol. 1.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv: 1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105.

Laloy, E., Hérault, R., Jacques, D., Linde, N., 2018. Training-image based geostatistical inversion using a spatial generative adversarial neural network. Water Resour. Res. 54 (1), 381–406. https://doi.org/10.1002/2017WR022148.

Laloy, E., Hérault, R., Lee, J., Jacques, D., Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. Adv. Water Resour. 110, 387–405. https://doi.org/10.1016/j.advwatres.2017.09.029.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436. https://doi.org/10.1038/nature14539.

LeCun, Y., Bengio, Y., et al., 1995. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks 3361 (10), 1995.

Liu, A., Ziebart, B., 2014. Robust classification under sample selection bias. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Advances in neural information processing systems. Curran Associates, Inc., pp. 37–45.

Mo, S., Zhu, Y., Zabaras, N., Shi, X., Wu, J., 2018. Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media https://doi.org/10.1029/2018WR023528. arXiv: 1807.00882.

Molinari, A., Pedretti, D., Fallico, C., 2015. Analysis of convergent flow tracer tests in a heterogeneous sandy box with connected gravel channels. Water Resour. Res. 51 (7), 5640–5657. https://doi.org/10.1002/2014WR016216.

Mosser, L., Dubrule, O., Blunt, M.J., 2017. Reconstruction of three-dimensional porous media using generative adversarial neural networks. Phys. Rev. E 96 (4), 43309. https://doi.org/10.1103/PhysRevE.96.043309.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, pp. 8024–8035.

Raissi, M., Perdikaris, P., Karniadakis, G., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv: 1711.10561.

Rehfeldt, K.R., Boggs, J.M., Gelhar, L.W., 1992. Field study of dispersion in a heterogeneous aquifer: 3. geostatistical analysis of hydraulic conductivity. Water Resour. Res. 28 (12), 3309–3324. https://doi.org/10.1029/92WR01758.

Renard, P., Allard, D., 2013. Connectivity metrics for subsurface flow and transport. Adv. Water Resour. 51, 168–196. https://doi.org/10.1016/j.advwatres.2011.12.001.

Rizzo, C.B., de Barros, F.P., 2017. Minimum hydraulic resistance and least resistance path in heterogeneous porous media. Water Resour. Res. 53 (10), 8596–8613. https://doi.org/10.1002/2017WR020418.

Rizzo, C.B., Nakano, A., de Barros, F.P., 2019. Par2: parallel random walk particle tracking method for solute transport in porous media. Comput. Phys. Commun. 239, 265–271. https://doi.org/10.1016/j.cpc.2019.01.013.

Rubin, Y., 2003. Applied stochastic hydrogeology. Oxford University Press.

Rubin, Y., Dagan, G., 1988. Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers: 1. constant head boundary. Water Resour. Res. 24 (10), 1689–1697. https://doi.org/10.1029/WR024i010p01689.

Rubin, Y., Dagan, G., 1989. Stochastic analysis of boundaries effects on head spatial variability in heterogeneous aquifers: 2. impervious boundary. Water Resour. Res. 25 (4), 707–712. https://doi.org/10.1029/WR025i004p00707.

Salamon, P., Fernàndez-Garcia, D., Gómez-Hernández, J., 2006. Modeling mass transfer processes using random walk particle tracking. Water Resour. Res. 42 (11). https://doi.org/10.1029/2006WR004927.

Salamon, P., Fernàndez-Garcia, D., Gómez-Hernández, J.J., 2006. A review and numerical assessment of the random walk particle tracking method. J. Contam. Hydrol. 87 (3–4), 277–305. https://doi.org/10.1016/j.jconhyd.2006.05.005.

Sanchez-Vila, X., Fernàndez-Garcia, D., 2016. Debates stochastic subsurface hydrology from theory to practice: why stochastic modeling has not yet permeated into practitioners? Water Resour. Res. 52 (12), 9246–9258. https://doi.org/10.1002/2016WR019302.

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Shapiro, A.M., Cvetkovic, V.D., 1988. Stochastic analysis of solute arrival time in heterogeneous porous media. Water Resour. Res. 24 (10), 1711–1718. https://doi.org/10.1029/WR024i010p01711.

Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resour. Res. 54 (11), 8558–8593. https://doi.org/10.1029/2018WR022643.

Srisutthiyakorn, N., 2016. Deep-learning Methods for Predicting Permeability from 2D/3D Binary-segmented Images. In: SEG technical program expanded abstracts 2016. Society of Exploration Geophysicists, pp. 3042–3046. https://doi.org/10.1190/segam2016-13972613.1.

Ströfer, C.M., Wu, J., Xiao, H., Paterson, E., 2019. Data-driven, physics-based feature extraction from fluid flow fields using convolutional neural networks. Commun. Comput. Phys. 25, 625–650. https://doi.org/10.4208/cicp.OA-2018-0035.

Vasilyeva, M., Tyrylgin, A., 2018. Machine learning for accelerating effective property prediction for poroelasticity problem in stochastic media. arXiv: 1810.01586.

Wang, Y., Arns, C.H., Rahman, S.S., Arns, J.-Y., 2018. Porous structure reconstruction using convolutional neural networks. Math. Geosci. 50 (7), 781–799. https://doi.org/10.1007/s11004-018-9743-0.

Wu, J., Yin, X., Xiao, H., 2018. Seeing permeability from images: fast prediction with convolutional neural networks. Sci. Bull. 63 (18), 1215–1222. https://doi.org/10.1016/j.scib.2018.08.006.

Zech, A., Attinger, S., Cvetkovic, V., Dagan, G., Dietrich, P., Fiori, A., Rubin, Y., Teutsch, G., 2015. Is unique scaling of aquifer macrodispersivity supported by field data? Water Resour. Res. 51 (9), 7662–7679. https://doi.org/10.1002/2015WR017220.

Zheng, C., Bianchi, M., Gorelick, S.M., 2011. Lessons learned from 25 years of research at the made site. Groundwater 49 (5), 649–662. https://doi.org/10.1111/j.1745-6584.2010.00753.x.

Zheng, C., Gorelick, S.M., 2003. Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale. Groundwater 41 (2), 142–155. https://doi.org/10.1111/j.1745-6584.2003.tb02578.x.

Zhu, Y., Zabaras, N., 2018. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. J. Comput. Phys. 366, 415–447. https://doi.org/10.1016/j.jcp.2018.04.018.

Zinn, B., Harvey, C.F., 2003. When good statistical models of aquifer heterogeneity go bad: a comparison of flow, dispersion, and mass transfer in connected and multivariate gaussian hydraulic conductivity fields. Water Resour. Res. 39 (3). https://doi.org/10.1029/2001WR001146.