Supplementary Material

1 SUPPLEMENTARY DATA

Text S1: Neural Network

A neural network (NN) is a ML algorithm loosely based on the human brain (McCulloch and Pitts, 1943). This algorithm is used for both classification and regression tasks. A simple NN is composed of three parts: an input layer, a hidden layer, and an output layer. The input layer is the input to the model. The hidden layer is composed of neurons (or nodes). Given an input $X = \{x_1, x_2, ..., x_n\}$ of size n, a neuron consists of a weights vector $W = \{w_1, w_2, ..., w_n\}$ of size n and a bias value b. Figure S1 shows an example of a neuron. The neuron uses the input and calculates an output y by the equation:

$$y = f((\sum_{i=1}^{n} w_i x_i) + b) \tag{S1}$$

where f is the activation function. Typically, f is a nonlinear function (i.e sigmoid, rectified linear units) that adds non-linearity to the model. The output of the neuron, y, is then used as part of the input for the next layer, the output layer. The output layer also consists of neurons, but the activation function (or lack there of) is dependent on the model task. For regression models, there is typically no activation function. The output layer can be one value or multiple values depending on the task. NNs are trained by optimizing the weight values over numerous iterations to achieve accurate estimations by minimizing the output of a loss function.

When all the neuron in a layer are connected to all neurons in the adjacent layers, the layer is fully-connected. For one-dimensional neural networks, this is called a dense layer. A DNN is simply a NN with multiple hidden layers of neurons. The number of hidden layers is referred to as the depth of the network. Adding additional hidden layers typically increases model performance by allowing the model to extract higher level abstract features from the input data. With each hidden layer acting as the input to subsequent hidden layer, a hierarchy of abstract features and input data representations can be extracted during the process of DNN model development.

Text S2: ES Analysis

To select the ES, we computed the correlation coefficient and root mean square error between the 60 synthetic true permeability and the corresponding 60 estimations from each of the 537 ensemble posterior, for the different observation errors. Overall, s3 has the highest correlation coefficient and the lowest root mean squared errors for all the errors (Figure S2). It implies that ES has the best estimation on s3, which is consistent with the training results from DNN models. Also, we observe that the correlation and root mean square error generally increases and decreases with observation error, respectively, with the converging or turning point at around 0.05. Therefore, we chose results with relative error of 0.05 for the comparison between DNN models and ES.

Text S3: ATS setup

Critically for simulations at time scales of over a few days, evapotranspiration (ET) is determined in ATS for these problems using a modified Priestly-Taylor approximation based on that of the Precipitation-Runoff Modeling System (PRMS Version 4) (Markstrom et al., 2015). This approach approximates a potential ET based on the Priestley-Taylor equation and a ground flux based on a lagged air temperature. This potential sink is then limited if insufficient water is available within the near surface soil using the approach of the Community Land Model (CLM version 4.5) (Oleson et al., 2013). This limiter is done in two stages: first, a combined factor of water saturation and plant rooting fraction is used to distribute the evapotranspiration flux to the soil, and second, a wilting-point based factor is used to limit fluxes from any given cell based on the local mafic potential.

Prescribed precipitation is partitioned into rain and snow by the air temperature – if the mean daily air temperature is below zero, precipitation is assumed to come as snow. Snow is reserved in a simple bucket model, and snowmelt is reintroduced to the surface water system through a thawing-degree-days approach as is done in PRMS (Markstrom et al., 2015).

To drive the ATS, meteorological data including precipitation as a source of water, air temperature, incoming net radiation, and relative humidity are required by the Priestly-Taylor potential ET, and a subsurface mesh including soil properties are required. ATS then predicts the ET, and solves water conservation equations to determine liquid water pressure in all subsurface and surface grid cells (or equivalently hydrologic head), and the flux of water on both subsurface faces (Darcy flux) and surface faces (overland flow). Boundary conditions are given by no-flux on all subsurface boundaries (except the surface) and a seepage face condition on all surface boundaries that allows discharge out of the domain but allows no water to come onto the domain (as is expected of a headwater system). Initial conditions are given by a spinup process in which the average annual summer precipitation rate was prescribed for 20,000 days across the entirety of the surface, and discharge was monitored until a pseudo-steady-state was reached in which discharge was matched by integrated precipitation. Observations in the code are used to integrate the egress of water from the domain through the outlet of Rock Creek on the surface boundary alone, providing discharge in Rock Creek as a function of time. An artificially high Manning's coefficient of 1.5 is used to simulate surface runoff at reduced computational cost. Sensitivity of discharge to Manning's coefficient was tested by varying Manning's coefficient from 0.5 to 2.5, and the daily averaged discharge was found to be nearly identical. This would likely not be true for hourly or faster precipitation forcing and/or discharge measurements.

2 SUPPLEMENTARY TABLES AND FIGURES

2.1 Tables

Table S1. Hyperparameters varied for each model type. For the "Layer Size(s)" column, the first number is the size of the first hidden layer and the second number is the size of the second hidden layer. Layer sizes with only one number only have one hidden layer.

Learning Rates	Layer Size(s)
0.0001, 0.0002, 0.0003, 0.0004, 0.0005, 0.0006, 0.0007, 0.0008, 0.0009, 0.001	[350, 100], [350, 50], [350, 25], [100, 50], [100, 25], [50, 25], [350], [100], [50], [25]

Table S2. Best parameter combination from table S1 for each DNN type.

Model Type	Predicted Parameter	Layer Size(s)	Learning Rate
DNN A1	s3	[350, 50]	0.0005
DNN A1	s6	[350, 25]	0.0004
DNN A1	g1	[350, 50]	0.0002
DNN A1	g5	[350, 100]	0.0001
DNN A1	g7	[100, 50]	0.0003
DNN A2	all	[350, 100]	0.0001
DNN A3	all	[350, 100]	0.0001

Table S3. Mean squared error of the log10 permeability from the best models for each model type on the training split of ensembles. The ”Mean” column is the mean squared error, the ”STD” column is the standard deviation of the squared error, and the ”R^2” column is the R^2 correlation. Bold numbers in the ”Perm” rows indicated the MSE and R^2 with the best value.

	DNN A1			**DNN A2**			**DNN A3**		
Perm	**MSE**	**STD**	$\boldsymbol{R^2}$	**MSE**	**STD**	$\boldsymbol{R^2}$	**MSE**	**STD**	$\boldsymbol{R^2}$
g1	2.85×10^{-3}	4.94×10^{-3}	0.991	$\mathbf{2.33 \times 10^{-3}}$	4.63×10^{-3}	**0.993**	3.04×10^{-3}	5.01×10^{-3}	0.991
g5	5.19×10^{-4}	7.79×10^{-4}	0.998	1.13×10^{-3}	1.67×10^{-3}	0.997	$\mathbf{4.65 \times 10^{-4}}$	6.72×10^{-4}	**0.999**
g7	$\mathbf{6.88 \times 10^{-5}}$	1.52×10^{-4}	**1.0**	1.90×10^{-4}	3.41×10^{-4}	0.999	1.15×10^{-4}	1.99×10^{-4}	1.0
s3	1.60×10^{-4}	2.53×10^{-4}	1.0	3.35×10^{-4}	5.08×10^{-4}	0.999	$\mathbf{6.43 \times 10^{-5}}$	1.04×10^{-4}	**1.0**
s6	3.71×10^{-2}	6.76×10^{-2}	0.886	$\mathbf{1.20 \times 10^{-2}}$	1.98×10^{-2}	**0.963**	1.43×10^{-2}	2.29×10^{-2}	0.956

Table S4. Mean squared error of the log10 permeability from the best models for each model type on the test split of ensembles (i.e. mean of squared error distribution from Figure 6. The ”Mean” column is the mean squared error, the ”STD” column is the standard deviation of the squared error, and the ”R^2” column is the R^2 correlation. Bold numbers in the ”Perm” rows indicated the MSE and R^2 values for the best performance.

	DNN A1			DNN A2			DNN A3		
Perm	**MSE**	**STD**	R^2	**MSE**	**STD**	R^2	**MSE**	**STD**	R^2
g1	1.98×10^{-2}	4.11×10^{-2}	0.933	$\mathbf{1.85 \times 10^{-2}}$	3.87×10^{-2}	**0.937**	2.02×10^{-2}	3.77×10^{-2}	0.931
g5	$\mathbf{5.94 \times 10^{-3}}$	1.71×10^{-2}	**0.981**	7.21×10^{-3}	1.57×10^{-2}	0.977	6.21×10^{-3}	1.61×10^{-2}	0.981
g7	$\mathbf{1.21 \times 10^{-3}}$	4.74×10^{-3}	**0.996**	1.32×10^{-3}	4.67×10^{-3}	0.996	1.3×10^{-3}	5.59×10^{-3}	0.996
s3	$\mathbf{1.9 \times 10^{-3}}$	3.17×10^{-3}	**0.995**	1.96×10^{-3}	2.96×10^{-3}	0.994	2.38×10^{-3}	3.88×10^{-3}	0.993
s6	1.88×10^{-1}	2.99×10^{-1}	0.498	1.68×10^{-1}	2.73×10^{-1}	0.551	$\mathbf{1.64 \times 10^{-1}}$	2.59×10^{-1}	**0.563**

2.2 Figures

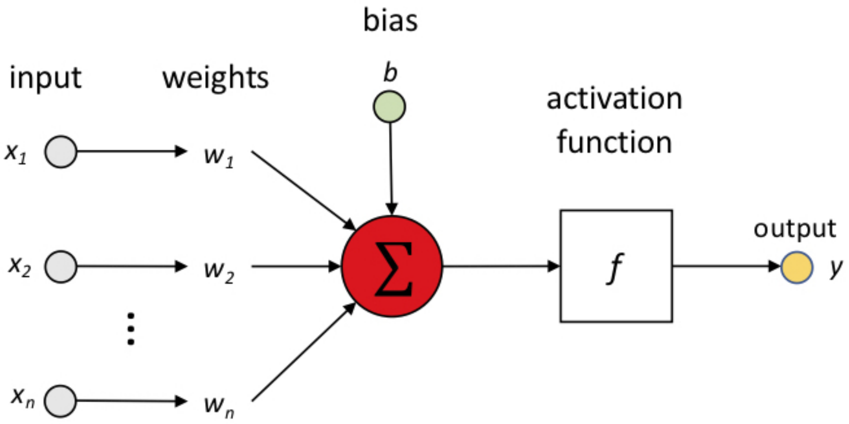

Figure S1. The architecture of a single neuron.

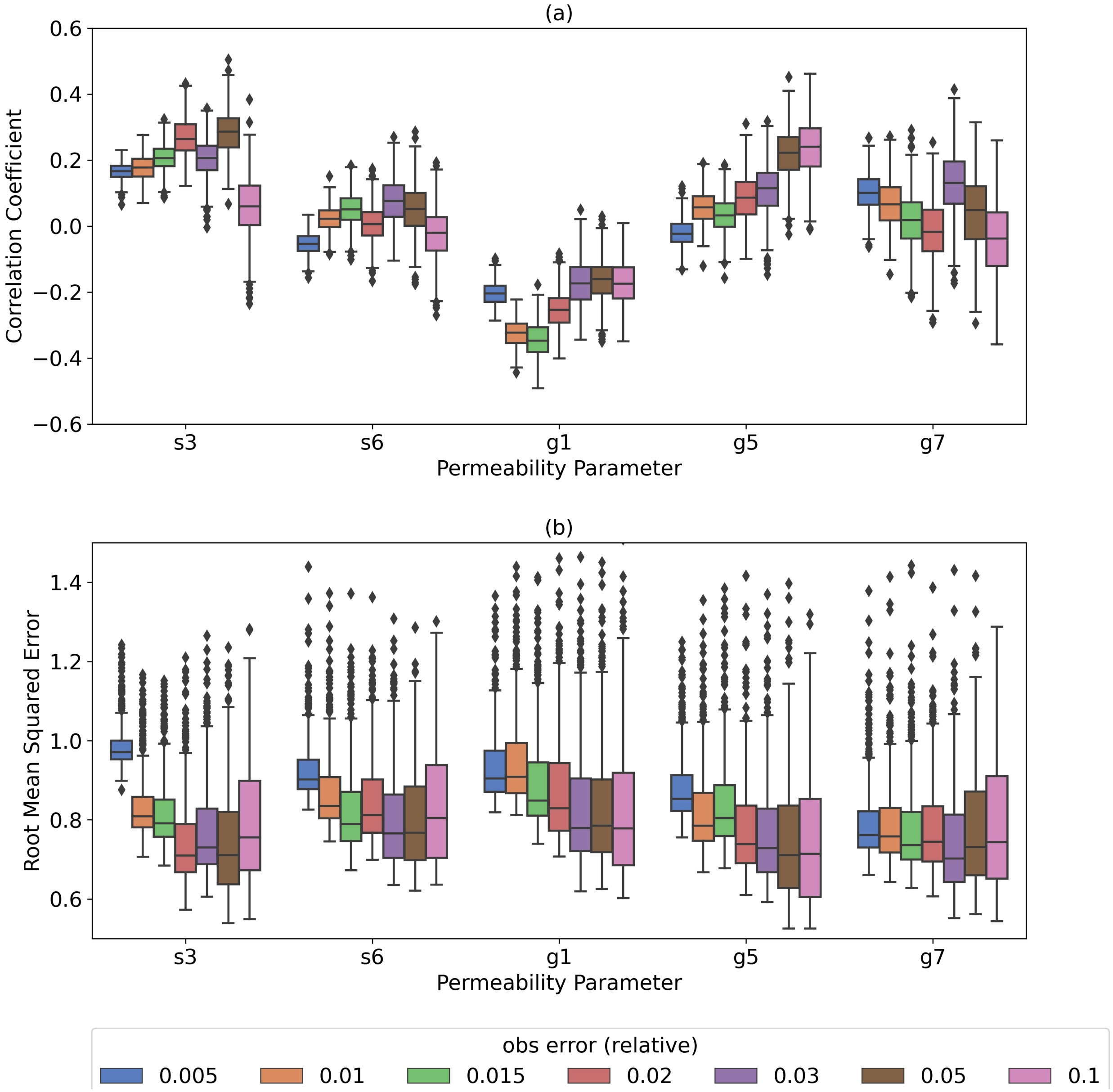

Figure S2. Box plots of correlation coefficient (a) and root mean squared error (b) between the 60 synthetic true permeability and the corresponding estimated permeability from 537 ensemble members using Ensemble Smoother (ES), with different relative observation errors.

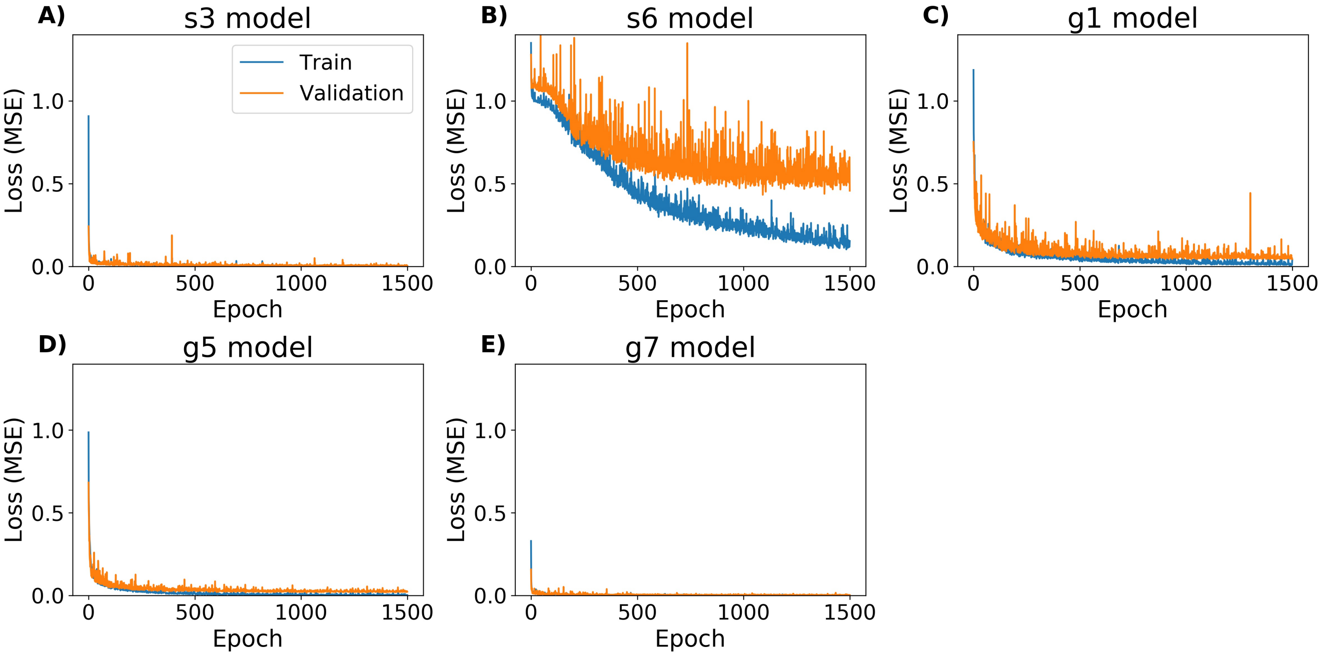

Figure S3. Training loss of the best DNN A1 model per permeability parameter. Each plot is an individual DNN A1 model. The blue line is the loss from the training set and the orange line is the loss from the validations set. The x-axis is the epoch number. The y-axis is the model loss (MSE). A) training loss for the DNN A1 model estimating the s3 permeability parameter; B) training loss for the DNN A1 model estimating the s6 permeability parameter; C) training loss for the DNN A1 model estimating the g1 permeability parameter; D) training loss for the DNN A1 model estimating the g5 permeability parameter; E) training loss for the DNN A1 model estimating the g7 permeability parameter.

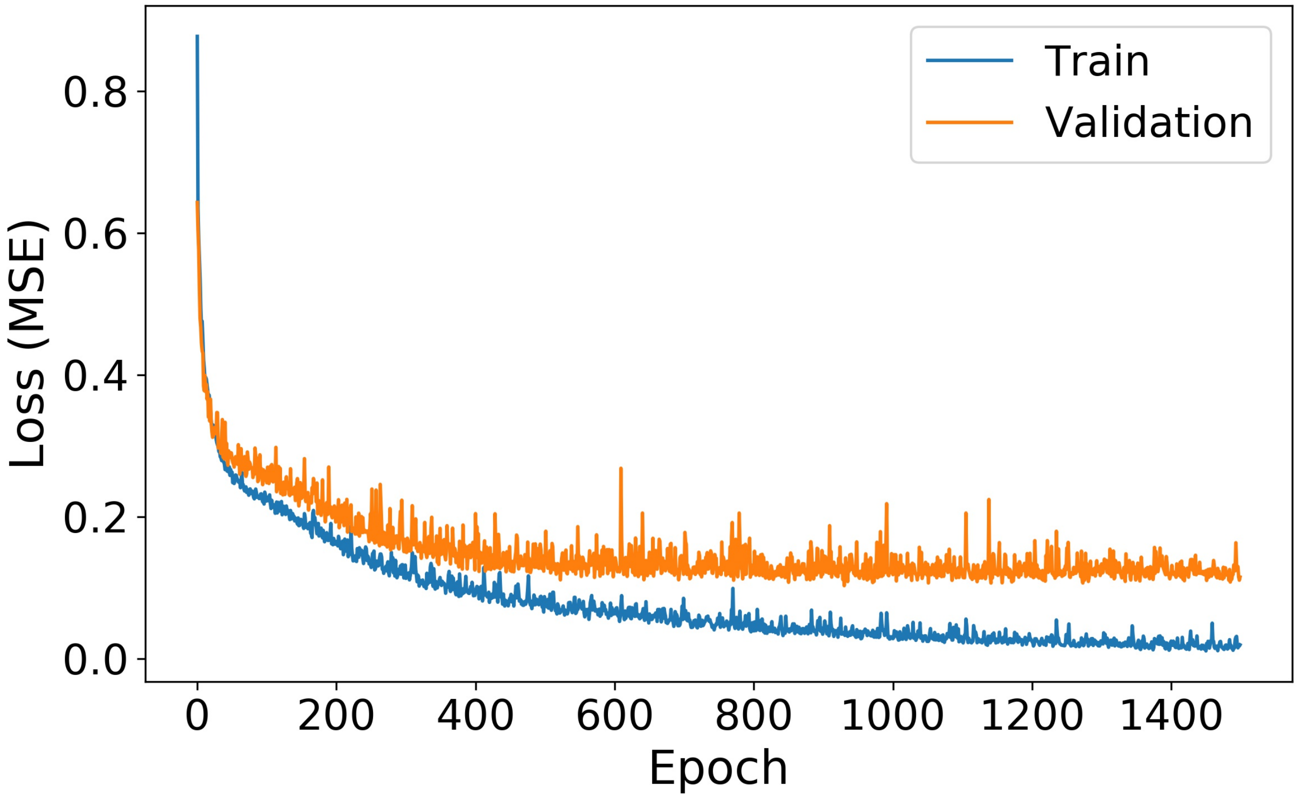

Figure S4. Training loss of best DNN A2 model over the training period. The blue line is the loss from the training set and the orange line is the loss from the validations set. The x-axis is the epoch number. The y-axis is the model loss (MSE). The loss is the average of the loss of the five permeability parameters.

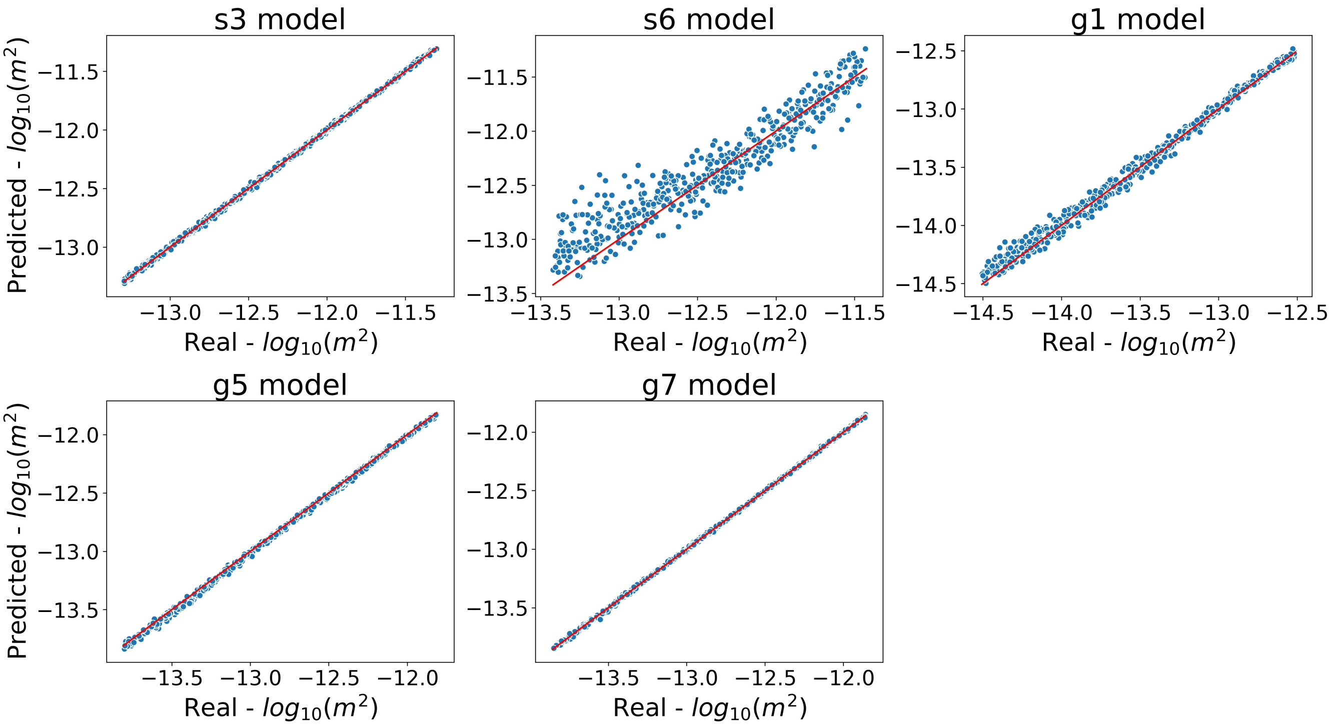

Figure S5. One-to-one plot of the DNN A1 model permeability estimation compare to the real estimation for the train set. Each plot is an individual DNN A1 model estimation for the given permeability parameter. Each dot represents a realization from the test set of ensembles. The x-axis is the log_{10} real permeability value, the y-axis is the estimated permeability value from the model. The red line is the one-to-one line.

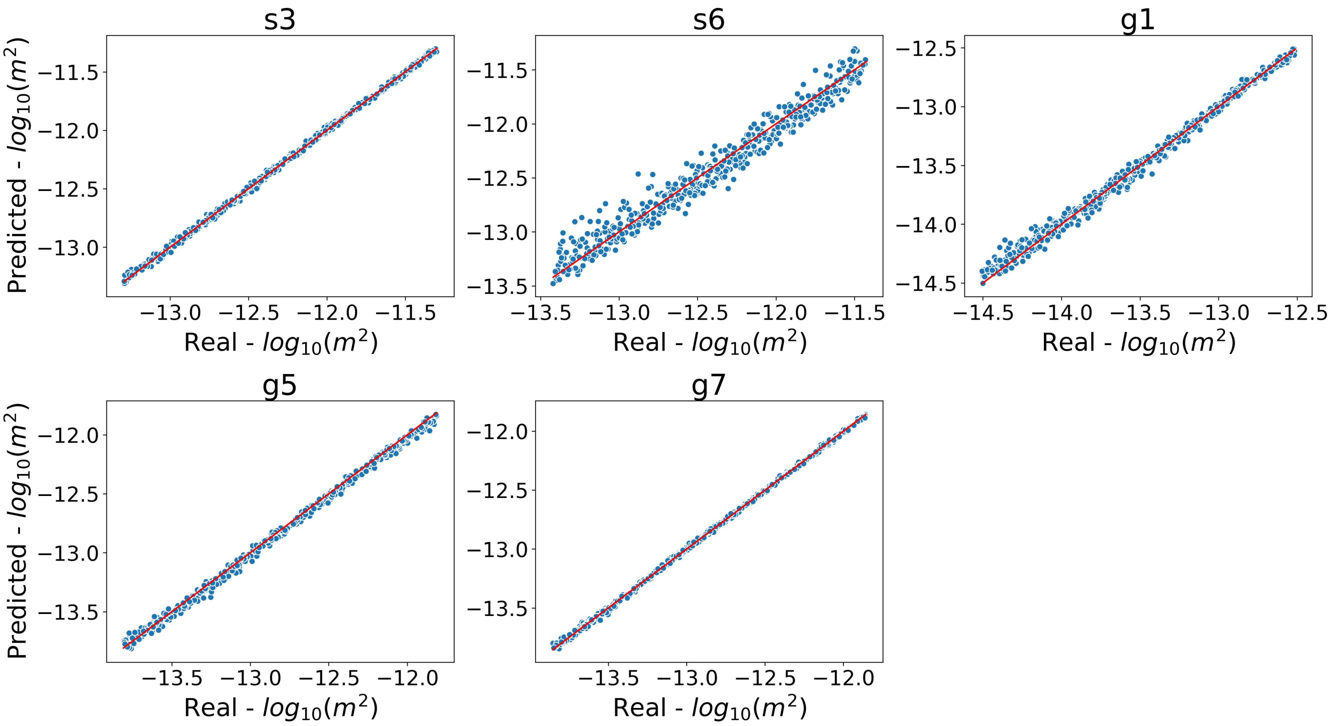

Figure S6. One-to-one plot of the DNN A2 model permeability estimation compare to the real estimation for the train set. Each plot is the DNN A2 model estimation for the given permeability parameter. Each dot represents a realization from the test set of ensembles. The x-axis is the log_{10} real permeability value, the y-axis is the estimated permeability value from the model. The red line is the one-to-one line.

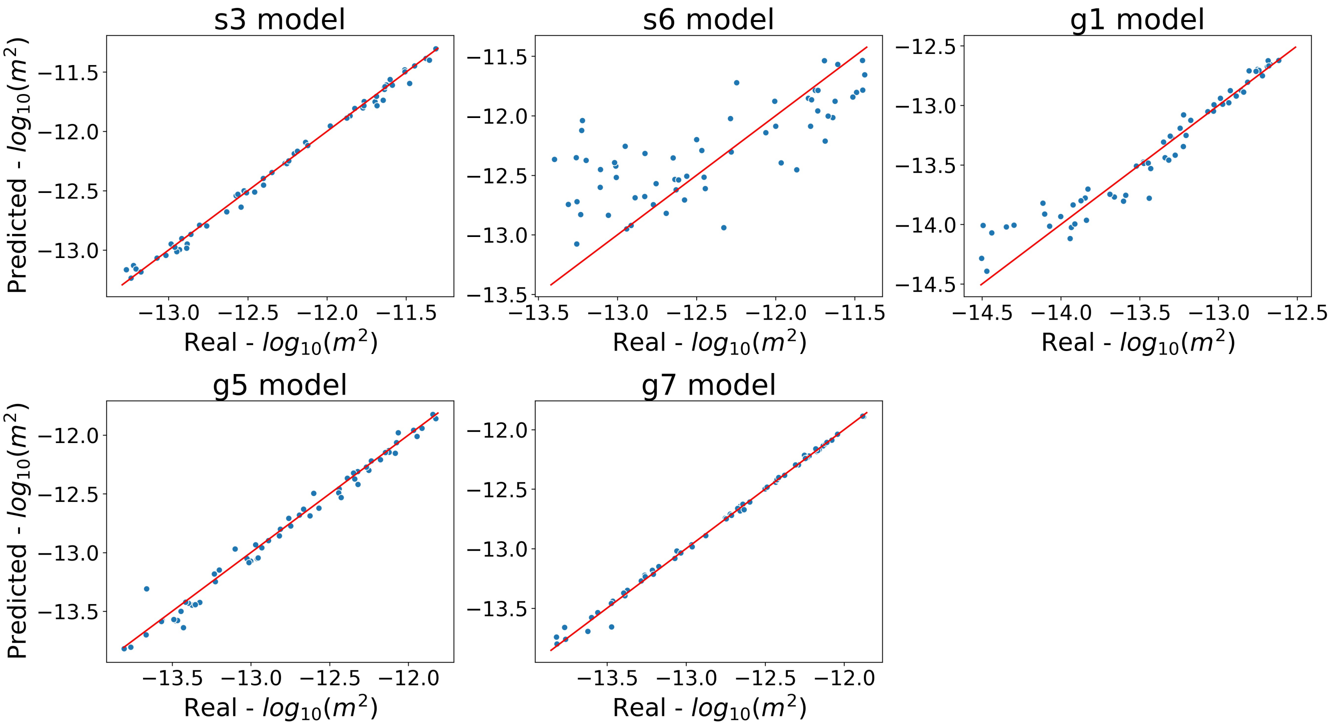

Figure S7. One-to-one plot of the DNN A1 model permeability estimation compare to the real estimation for the test set. Each plot is an individual DNN A1 model estimation for the given permeability parameter. Each dot represents a realization from the test set of ensembles. The x-axis is the log_{10} real permeability value, the y-axis is the estimated permeability value from the model. The red line is the one-to-one line.

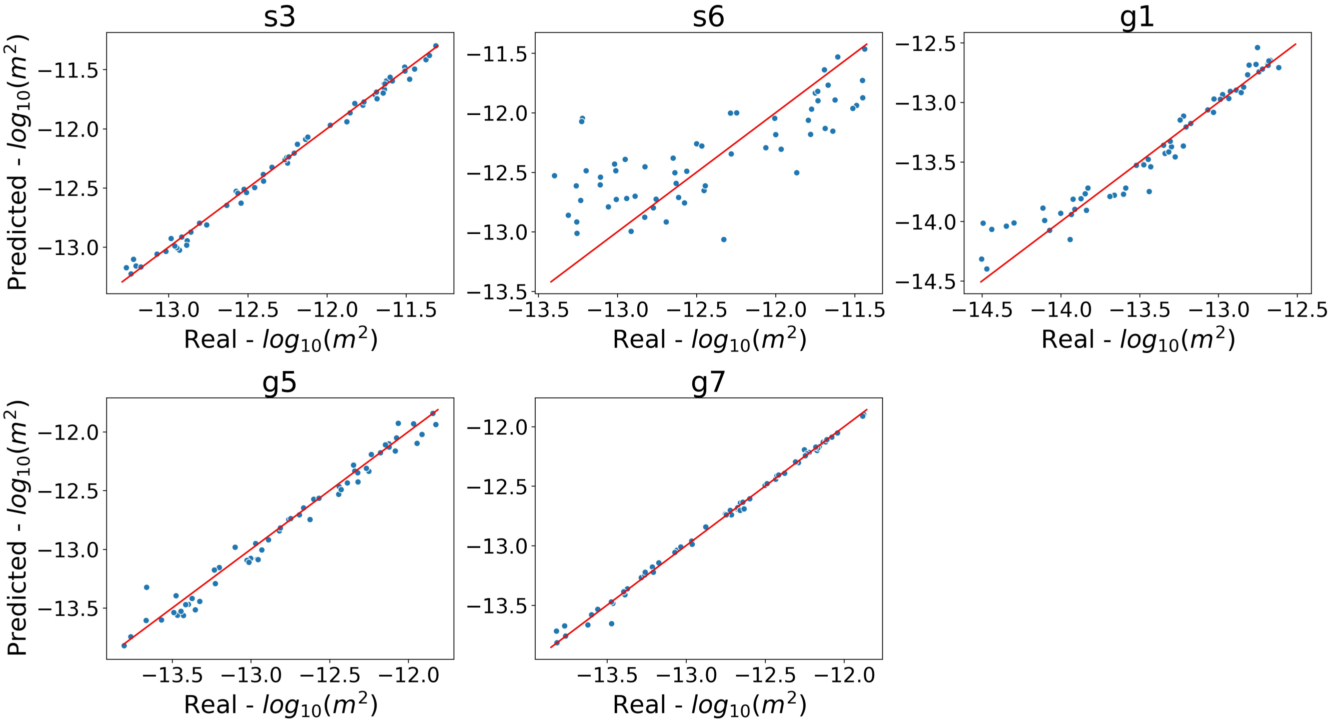

Figure S8. One-to-one plot of the DNN A2 model permeability estimation compare to the real estimation for the test set. Each plot is the DNN A2 model estimation for the given permeability parameter. Each dot represents a realization from the test set of ensembles. The x-axis is the log_{10} real permeability value, the y-axis is the estimated permeability value from the model. The red line is the one-to-one line.

REFERENCES

Markstrom, S., Regan, R., Hay, L., Viger, R., Web, R., Payn, R., et al. (2015). {PRMS-IV}, the {P}recipitation-{R}unoff {M}odeling {S}ystem, version 4. In *U.S. Geological Survey Techniques and Methods*, vol. Book 6. Chap. b7 edn., 158 p.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5, 115–133

Oleson, K., Lawrence, D., Bonan, G., Drewniak, B., Huang, M., Koven, C., et al. (2013). *Technical description of version 4.5 of the Community Land Model (CLM)*. Tech. rep., UCAR/NCAR. doi:10.5065/D6RR1W7M. Artwork Size: 5912 KB Medium: application/pdf