**Predicting Future Well Performance for Environmental Remediation Design using Deep Learning**

**Authors:** Xuehang Song[a,*], Huiying Ren[a], Zhangshuan Hou[a], Xinming Lin[a], Marinko Karanovic[b], Matt Tonkin[b], Vicky L. Freedman[a,c], Inci Demirkanli [a,*], and Rob Mackley[a]

**Affiliations:**

[a] Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99354

[b] S.S. Papadopulos and Associates Inc., 505 N. Pine St., Williamsfield, IL, 61489

[c] Sealaska, 1200 6th Ave, Suite 800, Seattle, WA 98101

*Corresponding Authors: xuehang.song@pnnl.gov and inci.demirkanli@pnnl.gov

**Abstract**

In this study, we developed a deep learning (DL) framework with a multi-channel three-dimensional convolutional neural network (MC3D-CNN) to predict well performance and thereby assist future environmental remediation design. Such prediction of extraction well performance at designated locations is critical for configuring pump-and-treat (P&T) well network design and operation, setting reasonable target closure dates for overall remedying, and estimating remedy costs. The framework is developed with operational and monitoring data routinely collected during P&T remedy operations, including well extraction and injection rates as well as in situ contaminant concentrations. Traditionally, the collected data were rarely used for purposes other than assessing **past well performance** and the accuracy of the conceptual site model. However, recent advances in data-driven computational approaches enable better use of the large datasets to inform **future well performance**, enhance site characterization, and improve remediation planning. In this study, we established a DL framework to integrate transient three-dimensional contaminant plumes and multiple aquifer properties (e.g., hydraulic conductivity and hydrostratigraphic maps) to identify characteristic patterns controlling and representing extraction well mass recovery, aiming at providing future mass recovery estimates for existing wells and candidate wells at any proposed locations. We evaluated our framework by using a realistic synthetic dataset generated from a well-calibrated flow and transport model used in the 200 West Area of the U.S. Department of Energy's Hanford Site in southeastern Washington state. The multi-channel feature in our framework allows integration of various types and temporal densities of training datasets for DL model development. Overall, we found that the trained DL model achieved an accuracy of over 90% in ranking extraction well performance in validation datasets, and over 80% in predicting high-performance-ranking well locations. This data-informed approach provides a flexible tool to support adaptive site management, streamline decision-making, and potentially reduce remediation time and costs. Our DL framework can be used as a filtering tool to improve the current P&T network optimization design by reducing the number of candidate well locations.

**Key words**

multi-channel three-dimensional convolutional neural network (MC3D-CNN), Deep Learning, groundwater remediation, multi-class classification, pump-and-treat, remediation performance assessment, well performance ranking,

## 1. Introduction

In this study, we detail a new framework for improving prediction of future well performance by using a deep learning (DL) model to discover the relationship between historical pump-and-treat (P&T) records and related contamination distribution, hydrostratigraphic units, and hydraulic conductivity maps. P&T is one common approach used to hydraulically contain and remediate groundwater contaminant plumes at many waste sites (McKinney and Lin, 1996; National Research, 2013; Truex et al., 2017). Statistics show that P&T is employed in approximately 40% of contaminated groundwater sites (EPA, 2002). In a typical P&T system, the contaminated groundwater is extracted to the ground surface by pumping and then treated using a filtering or stripping system. The treated water is then injected into the aquifer for groundwater recharge and/or to hydraulically contain the remaining contaminant plumes (EPA, 2005). The efficiency of P&T systems heavily depends on the performance of extraction wells, i.e., how much contaminant mass can be pumped from the aquifer within a reasonable time frame. Knowing extraction well performance at designated locations is critical for planning and modifying P&T systems, setting reasonable target closure dates for remedy operation, and estimating remedy costs (Zheng and Wang, 2002). While large amounts of historical well performance data are often routinely recorded under regulatory requirements, such information is mainly used for monitoring purposes only.

Over the past four decades, a variety of methods have been developed to simulate/predict the performance of extraction wells for remedy design, including the classical "batch flushing" analytical equation (Haley et al., 1989; National Research, 1994), analytic element method (Gaur et al., 2011; Majumder and Eldho, 2016; Matott et al., 2006), boundary element method (Kontos and Katsifarakis, 2017), semi-analytical solutions (Ameli and Craig, 2018; Cardiff et al., 2010), and more complex numerical fate and transport (F&T) models (such as MODFLOW, eSTOMP, PFLOTRAN, ITOUGH2, and many other simulators) (Finsterle, 2006; Finsterle and Zhang, 2011; Hammond and Lichtner, 2010; Neville and Tonkin, 2004; White and Oostrom, 2003). Among these approaches, the numerical F&T models usually provide more comprehensive representations of complex site features, such as three-dimensional nonuniform distribution of the contaminant plume, heterogeneous aquifer hydrogeological properties, dynamic groundwater gradient, and geochemical and/or biogeochemical reactions in the aquifer (Finsterle and Zhang, 2011; Huang and Mayer, 1997; Minsker et al., 2004). However, calibrating a numerical model is not trivial and can be time-consuming and site-specific. In many cases, site measurements are inadequate to fully constrain model parameters, which results in non-unique solutions and high predictive uncertainties of numerical models (Carrera, 1993; Singh and Minsker, 2008; Wu and Zeng, 2013). Reactive transport modeling is especially hard because of its large problem dimension (e.g., number of reactants and possible pathways) and high computational costs (Mayer et al., 2001; Steefel et al., 2015; Tsang et al., 2015). Because of these difficulties, simplified well models are still widely used for remedy planning (EPA, 2005). In addition, one commonly used phased strategy in remedy design is to move from simple calculations to analytical models and finally to a more detailed P&T system design (McMahon et al., 2001). In spite of the great value and wide application of simplified well models in remedy design, it is widely acknowledged that the existing simplified models lack representation of some important site features, especially the nonuniform distribution of contaminant plumes and heterogeneous aquifer hydrogeological properties, which can lead to overestimation of extraction well performance and then underestimation of cleanup time and cost (Brusseau, 1996; Hadley and

Newell, 2012; National Research, 1994). These over-simplifications can be alleviated by adjusting empirical parameters and adding additional components to the analytical solutions to mimic more complex site behaviors (Sváb et al., 2008).

Recently, machine learning (ML) techniques have drawn increasing attention in water and geoscience research (Shen, 2018; Tahmasebi et al., 2020). One popular application of ML methods in geoscience is to construct surrogate models to substitute for computationally expensive F&T models (Razavi et al., 2012; Yadav et al., 2018). In such cases, hundreds or a few thousand realizations of physical F&T were created by perturbing model parameters, well configurations, and/or initial conditions, and then the simulated modeling results were used as training datasets to feed into regression ML models. The trained ML models have several advantages, including high accuracy in reproducing the F&T model training dataset, low computational requirements relative to F&T models, and the capability to generate new predictions almost instantaneously. Mainly due to their high efficiency, the ML-based surrogate models became popular in computationally demanding F&T optimization applications. A variety of ML methods, such as artificial neural networks (Gaur et al., 2013; Rogers and Dowla, 1994; Yan and Minsker, 2006), extreme learning machine (Majumder and Lu, 2021), and deep neural network models (Chen et al., 2022; Yu et al., 2020), have been used to train surrogate models and coupled with global optimization models (e.g., evolution algorithms) for P&T network design. To alleviate the potential bias of surrogate models, it was also proposed to combine multiple types of ML models and form more reliable ensemble surrogates (Yin and Tsai, 2020; Zounemat-Kermani et al., 2021). One drawback of these surrogating approaches is that they lack physical representation and may produce physically unrealistic results (e.g., violating mass conservation law), and there is significant interest in enforcing physical laws to these "black box" models through physics-informed ML methods, such as employing metamodels (Soriano et al., 2021) or minimizing the residual of physical equations (Tartakovsky et al., 2020; Wang et al., 2021), among others. In addition to being employed as emulators for F&T simulations, ML methods are also used as novel inverse models in subsurface science by creating bidirectional mappings between physical parameters (e.g., hydraulic conductivity and facies structures) and model state variables (e.g., hydraulic head) (Mo et al., 2019; Sun, 2018; Wang et al., 2021). ML models have also been used to improve monitoring and site characterization of active and closed P&T sites by optimizing monitoring network design (Kontos et al., 2022; Meray et al., 2022), improving plume source identification (Kontos et al., 2022), and filling data gaps (Ren et al., 2022).

One recent intriguing topic of ML application in groundwater remediation is the use of pure data-informed approaches. Instead of explicitly implementing physical constraints in ML models or learning from physical F&T simulation results, these data-informed approaches seek to directly link the model outcome of interest and its controlling factors under ML frameworks for better groundwater contaminant estimation and remediation design. The selection of the controlling factors is not arbitrary, but rather is based on the understanding of causal relations of the aquifer system and physical laws. Thus, the ML algorithms are performed as a way of data mining to extract and formulate hidden relations and patterns between the variables of concern and their controlling factors. A predictive model of groundwater nitrate pollution was built using random forest (R.F.) regression by examining nitrate concentrations with 24 related site parameters, such as intrinsic hydrogeologic properties, driving forces, remotely sensed variables, and physical-chemical variables (Rodriguez-Galiano et al. (2014)). McConnell and others trained regression-based ML models (Prophet model and damped Holt's exponential smoothing model) to predict future carbon tetrachloride ($CCl_4$) plumes using historical $CCl_4$ concentration samples (McConnell et al., 2022). Their models can achieve satisfactory prediction of site closure time without solving governing transport equations with sufficient spatial and temporal density of data. Wu and others applied R.F. methods for classifying high health risk areas using various groundwater chemistry measurements, and demonstrated that an R.F. model with four types of the most important chemistry measurements can achieve a classification accuracy of 88.21% for groundwater quality (Wu et al., 2020). The aforementioned applications focused on estimating the extent of a current contaminant plume and/or predicting future plume migration; no studies has been reported that predict the performance of the P&T system using data driven-approaches.

With the rapid advance in ML applications in contaminant migration modeling and remediation design, one important dataset, historical P&T records, has been under-utilized. Under regulatory requirements, large amounts of extraction and injection records, such as flow rate and contaminant concentration, are routinely collected as a standard practice in most P&T remediation sites. These data are mainly used to monitor contaminant rate/mass removal while they directly reflect the response of the aquifer, and potentially can be mined to support better decision-making for future remedies (Brusseau, 2013; Truex et al., 2017). On the contrary, the historical well production data have already been proven to be very useful in the gas and oil industry to predict future oil production rates and guide future well drilling; e.g., (Hirschmiller et al., 2019); (Li et al., 2019).

Our DL approach allows integration of multi-type multi-resolution P&T monitoring and site characterization data, addresses the limitations of the analytical and semi-analytical solutions in representing heterogenous field characteristics, and avoids solving expensive governing transport equations. An image-based DL model, convolutional neural networks (CNNs), was developed and integrated to extract the hidden spatiotemporal correlations between physical control factors and historical records of P&T well performance. The model is evaluated using a realistic synthetic dataset generated from a calibrated F&T model for a site with historical contamination. The effects of training dataset type and temporal data density are also interrogated to understand potential improvements in model performance. Given the large amount of P&T records generated at many waste sites for monitoring purposes, and the increasing automatic collection and digitalization of these records, data-informed approaches such as our multi-channel CNN create opportunities to improve our understanding of contaminant transport and site management, streamline decision-making, and potentially reduce future remediation costs.

## 2. Methods

We developed a multi-channel 3D-CNN (MC3D-CNN) DL architecture to extract important features from transient plume distributions and aquifer hydrogeological properties that control extraction well performance. The trained DL model can be used as a prediction model that provides favorable locations for future wells to maximize contaminant mass recovery, shorten the operational time of the P&T system, and reduce total remediation cost. Section 2.1 introduces the DL background and key configurations of CNN, and Section 2.2 describes the architecture of the MC3D-CNN model (Figure 1) and its major hyperparameters. Implementation of the CNN classification model for predicting well performance with physical model simulations and physical properties is illustrated in Figure 2 and discussed in Section 2.3.

## 2.1. Deep learning and CNN configurations

DL is a sub-field of ML, which has been designed to reveal the hidden controlling mechanism in high-dimensional and nonlinear complex systems. Typical DL approaches, such as fully connected neural networks (FC-NNs), CNN, and long short-term memory, can automatically find the most salient features to be learned. These approaches have demonstrated tremendous success in a variety of applications, such as speech recognition, computer vision, and natural language processing CNN has outperformed other DL methods in predictive capability in many image-related applications, including medical imaging, material structure, object recognition, and others (Rao and Liu, 2020).

CNN was first developed for visual imagery analysis and feature extraction (LeCun et al., 2015). The Visual Geometry Group (VGG) block-wise model architecture is adopted to push the model depth toward high accuracy (Simonyan and Zisserman, 2015). The VGG model architecture includes a series of convolutional blocks containing multiple convolutional layers followed by batch normalization, pooling, and dropout layers within each block, and then connected to flatten and dense layers. VGG model architecture needs to be adjusted according to the characteristics of the datasets. The number of convolutional filters is incremented layer by layer to make sure that the increasingly richer features are properly extracted. Such layered organizations can learn hierarchical representations. The neurons of adjacent layers are connected by assigning weights and biases $\{W_i, b_i\}_{i=1}^{m}$, where $m$ is the number of layers in neural network $NN_m$. The initial layer is the input layer constructed in image sets and the last layer is the output defined as the classification labels. The predicted outcome is compared with the label and a measure is calculated representing the performance of the CNN. The categorical cross-entropy class is chosen for the multi-label classification problems. It computes the cross-entropy loss between the labels and model predictions, and the calculation of the loss function requires that the last dense layer be configured with the total number of classes; this enables softmax activation to predict the probability for each class. In between are hidden layers transforming the feature space of the input such that it matches the output. Max pooling is performed to reduce the data size using spatial down-sampling, while preserving discriminant information. The normalization of output from previous layers allows the neural network to learn the pattern more independently. Dropout as a common regularization technique is also used to introduce stochasticity to make model performance more robust and prevent overfitting.

### 2.2 Multi-channel 3D-CNN architecture

Three-dimensional (3D)-CNN is needed to take labeled 3D images for feature extraction, but it is computationally and memory exhausting because of its much larger number of trainable parameters compared to the regular two-dimensional (2D)-CNN variant. Recent advances in computational hardware, especially general-purpose graphics processing units, have made 3D-CNN computationally affordable (Zhao et al., 2019). In a typical 3D-CNN design, a 3D image passes through a series of blocks of convolutional layers to extract feature maps, as shown by the architecture illustrated in Figure 1. Multiple 3D image datasets, including different data types and their temporal and spatial components, are fed into our multi-channel 3D-CNN. Under such architecture, the ensemble of sub-CNNs per channel are trained simultaneously to learn the spatiotemporal features ingested from various sources (e.g., plume distribution, hydrostratigraphic unit map) and match the features to the predefined labels (e.g., well performance ranking index).

Hyperparameter searching is needed to optimize the MC3D-CNN model configuration. A series of configuration parameters were explored, including batch size, kernel size, number of layers, number of neurons in each convolutional block, and dropout rate. The optimal configuration was then chosen by comparing the performance metrics of various hyperparameter combinations on training and validation datasets. The final MC3D-CNN model was then evaluated with an independent testing dataset.

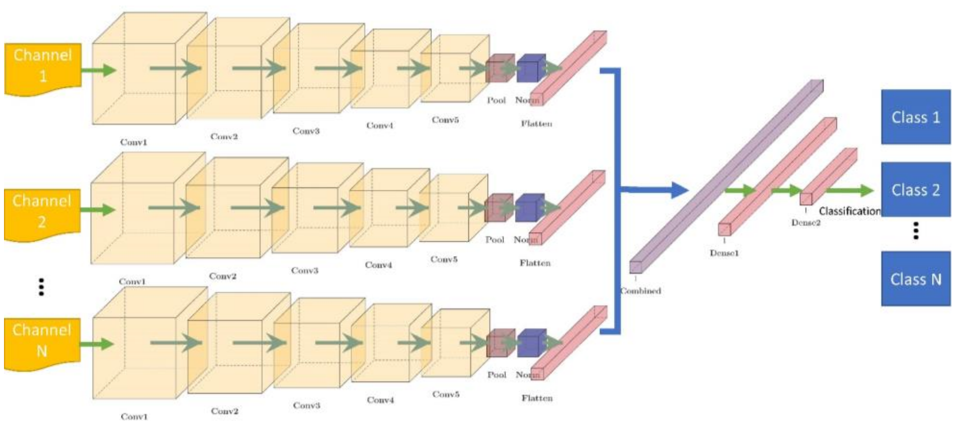


Figure 1. Architecture of the MC3D-CNN classification model.

**2.3 DL framework for well performance classification**

Figure 2 illustrates the overall workflow for training MC3D-CNN and predicting future performance ranking. Three physical attributes: plume concentration, hydrostratigraphic unit, and hydraulic conductivity, were used to compose the 3D training images. In this study, a synthetic modeling dataset was used to generate the 3D training images for each extraction well, as detailed in section 3.2.2. The accumulative mass recovery was categorized into three levels of performance, which serve as the image labels for CNN supervised learning. The raw pixel data from the 3D training images were fed into the 3D-CNN model, which can integrate various sources of data representing different aspects of system behaviors and dynamics related to well performance~~.~~.

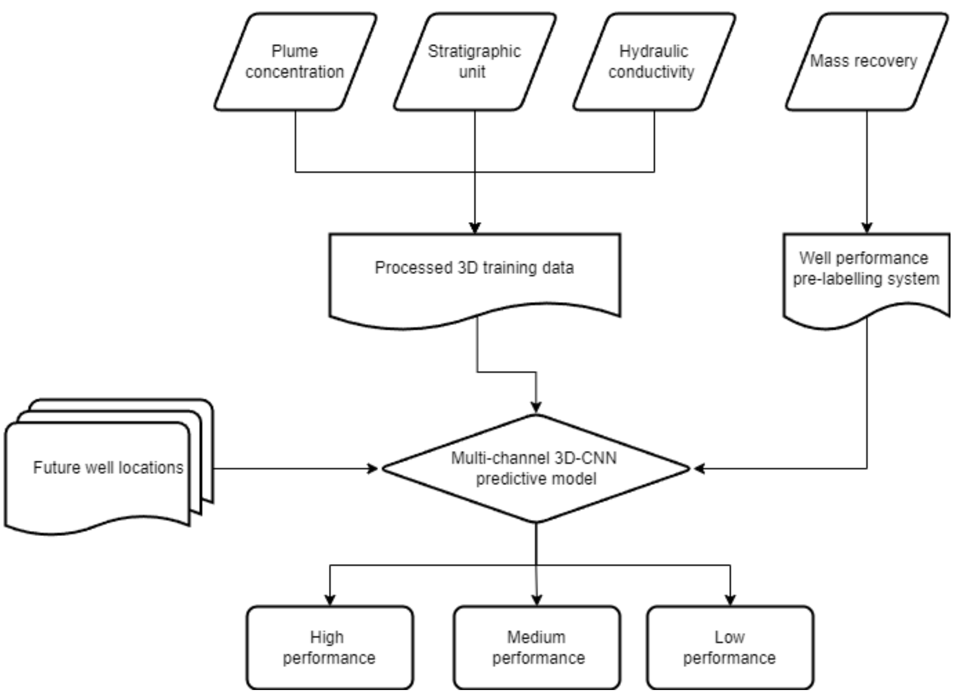

Figure 2. DL-based workflow for ranking well performance.

## 3. Case Design and Model Evaluation Criteria

A suitable example for demonstrating the MC3D-CNN model is the remediation of the Central Plateau at the U.S. Department of Energy (DOE) Hanford Site (Section 3.1). Here, we extracted the simulated $CCl_4$ plumes from an F&T model covering the Central Plateau (Section 3.2) to compose a synthetic dataset for model testing. The advantage of using synthetic data from the F&T model is that it can provide a known reference for evaluating the accuracy of the DL model. Section 3.3 presents the details of the DL model setup and model parameters for the synthetic case. The criteria used to measure the performance of the DL model are described in Section 3.4. Section 3.5 illustrates how to use the trained DL model to predict well performance ranking.

### 3.1. Site description

The DOE Hanford Site, located in southeastern Washington State, holds radioactive waste from the disposal of nuclear fuel fabrication wastes from 1943 to 1975 (Figure 3). The Central Plateau is an informal geographic designation given to the broad central portion of the Hanford Site that encompasses the 200 West and 200 East Areas. The Central Platea area is one of the most complex environmental remediation sites in the world, with shallow sources (e.g., waste tanks), persisting and recalcitrant deep vadose zone residual sources, large-scale groundwater plumes [e.g., carbon tetrachloride ($CCl_4$), technetium-99 (Tc-99), iodine-129 (I-129) and nitrate ($NO_3$)], and subsurface heterogeneity (Demirkanli and Freedman, 2021). Groundwater with several contaminants of concern (COCs) has been treated by the 200 West P&T Facility in the Central

Plateau since 2012. The 200 West P&T Facility is designed to capture and treat contaminated groundwater to reduce the mass of selected COCs, such as $CCl_4$, $NO_3$, Tc-99, and I-129, by at least 95% within 25 years from the startup (Demirkanli et al., 2018). The locations of extraction wells are shown as green diamonds in Figure 3b.

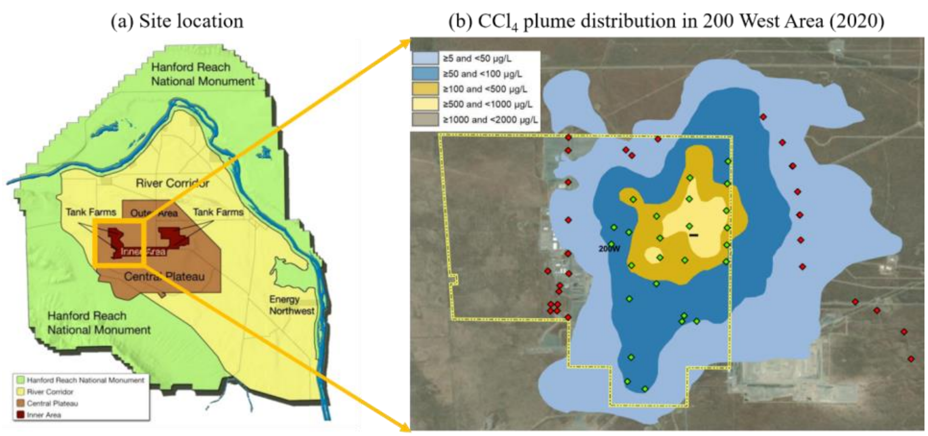


Figure 3. (a) Site location (source: https://www.hanford.gov/files.cfm/Attachment_5_Approach_CP_Cleanup_handout.pdf ); (b) $CCl_4$ plume distribution in 200 West Area (downloaded from https://www.hanford.gov/page.cfm/PHOENIX). The yellow polygon shows the boundary of the 200 West Area. The green diamonds are existing extraction wells and the red diamonds are existing injection wells.

### 3.2. Synthetic dataset generation

The Plateau-to-River groundwater model (P2R model) (Budge and Nichols, 2020) has been developed for the Central Plateau and extends eastward to the Columbia River (Figure S1). The

P2R model was calibrated using hundreds of monitoring wells. The model primarily provides the computational basis for simulating the F&T of contaminants in groundwater within the near- and far-field portion of the affected aquifer in the Central Plateau, and is currently used to support ongoing remedial activities on the Central Plateau. More details of the P2R model can be found elsewhere (Budge and Nichols, 2020).

#### 3.2.1. Well performance ranking

While the P2R model simulated groundwater F&T of multiple contaminants in the Central Plateau, this study only focused on the $CCl_4$ removal data in the 200 West Area for demonstration purposes. The performance of each well is ranked as high, medium, or low according to the well's $CCl_4$ recovery. These well rankings were used as labels in the DL classification model. It is noticed that the ranking of each well varies over time, where a typical well might be ranked as a high-performance well in its early years of operation, then its performance will decrease over the years with the removal of surrounding $CCl_4$. Therefore, performance of each well is distinguished and labeled at multiple time segments. Figure 4 shows the simulated annual $CCl_4$ mass recovery of 28 existing extraction wells from the P2R model.

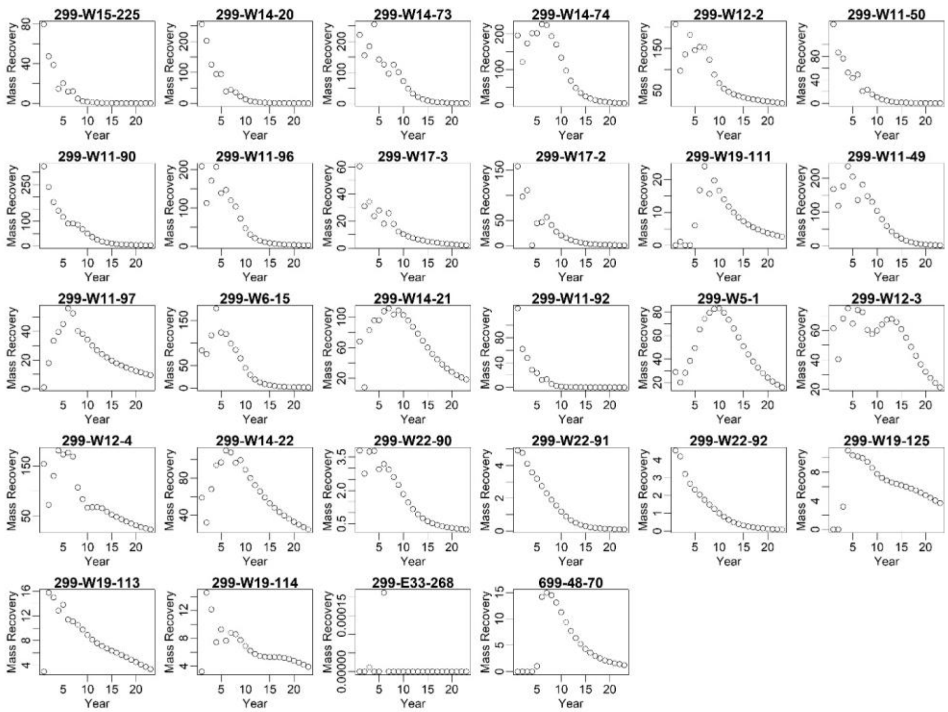


Figure 4. Yearly mass recovery of 28 wells extracted from P2R model simulation.

Each of the 28 extraction wells was then labeled as high, medium, or low performance according to its 5-year cumulative mass recovery (Eq 1), as:

$$
\begin{gathered}
R_{Low}: M_{i,j} < C_1 \\
R_{Medium}: C_1 \leq M_{i,j} \leq C_2 \\
R_{High}: C_2 \leq M_{i,j}
\end{gathered}
\tag{1}
$$

where $C_1$[Kg] and $C_2$[Kg] are predefined threshold values. $R_{Low}$ , $R_{Medium}$, and $R_{High}$ denote the performance indicator values corresponding to low, medium, and high performance, respectively. $M_{i,j}$ is a moving sum calculated from the annual $CCl_4$ mass recovery:

$$M_{i,j} = \sum_{k=j+1}^{j+n} m_{i,k} \text{ , for } j = N - n \tag{2}$$

where $m_{i,k}$[Kg] is the $CCl_4$ mass recovery of well $i$ in year $k$ ; $M_{i,j}$ is the $CCl_4$ mass recovery of well $i$ for the following $n$ year starting in year $j$; $n$[year] is the time window for moving sum, which is 5 years in this study; and $N$[year] is total number of years. For example, the "future" performance rank of a well in 2012 is evaluated using its total mass recovery between 2012 and 2017. We chose a 5-year moving time window because more than half of the wells reached their peak performance around year 5 (Figure 4). The values of C1 and C2 are 30[Kg] and 200[Kg], respectively, where the selections are based on the quantile and distribution of the cumulative mass recovery and the pre-designed equal distributed number of members in each class (Figure S2). The purpose is to make a balanced classification system to improve the predictive performance and avoid model bias. Figure 5 shows the calculated ranking for each well in different years. It is not surprising that most of the 28 wells move from medium/high performance to low/medium performance rankings over decades of remediation. However, there are still a few wells that move from low/medium to medium/high rankings (e.g., 299-W11-49 and 200-W11-97). This is because the plume spreading increases the $CCl_4$ concentration in those wells and thus improves their $CCl_4$ extraction rates.

**Formatted:** Subscript

**Formatted:** Subscript

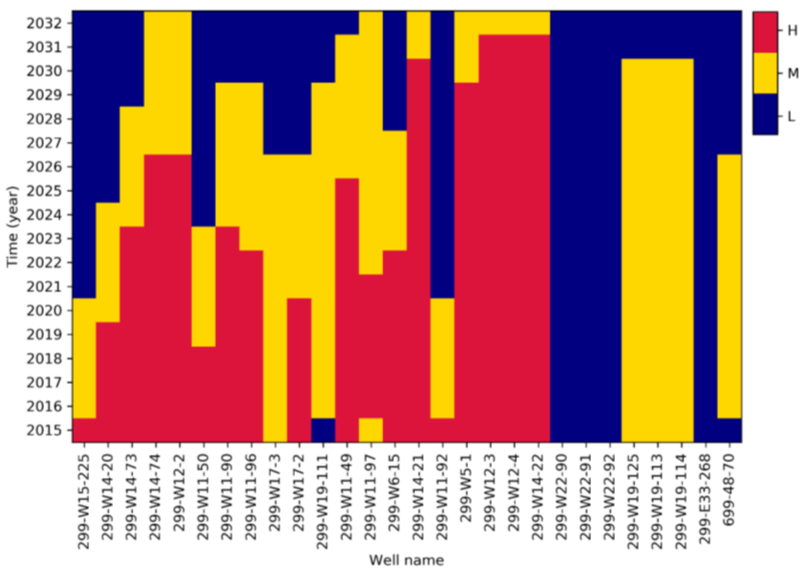


Figure 5. Well performance rankings for DL model training (year 2015~2028) and testing (year 2029 ~2032).

#### 3.2.2. 3D image sets under different scenarios

Around each well, three types of model parameters and variables – plume concentration, hydraulic conductivity, and hydrostratigraphic unit – were extracted from the P2R model configuration and output files. The horizontal and vertical lengths of the 3D images are 1000 m and 80 m, respectively, as selected by model sensitivity test. Three scenarios have been designed to test different combinations of these datasets on DL model performance:

- Scenario 1 (S1) uses the current snapshot of plume concentration only.
- Scenario 2 (S2) adds two earlier time steps to S1 as the model inputs. Such a setup considers the temporal impact for model prediction.

- Scenario 3 (S3) adds hydraulic conductivity and hydrostratigraphic units to S2. Both hydraulic conductivity and hydrostratigraphic units are static over time.

Examples of different data sources are illustrated in Figure 6 at three wells representing different performance indicators. In the year 2021, the performance wells 299-W14-20, 299-W14-73, and 299-W-15-225 are ranked as medium, high, and low, respectively. In the year 2026, the performance rankings of the three wells change to low, medium, and low, respectively. In the year 2031, all three wells are ranked as low performance.

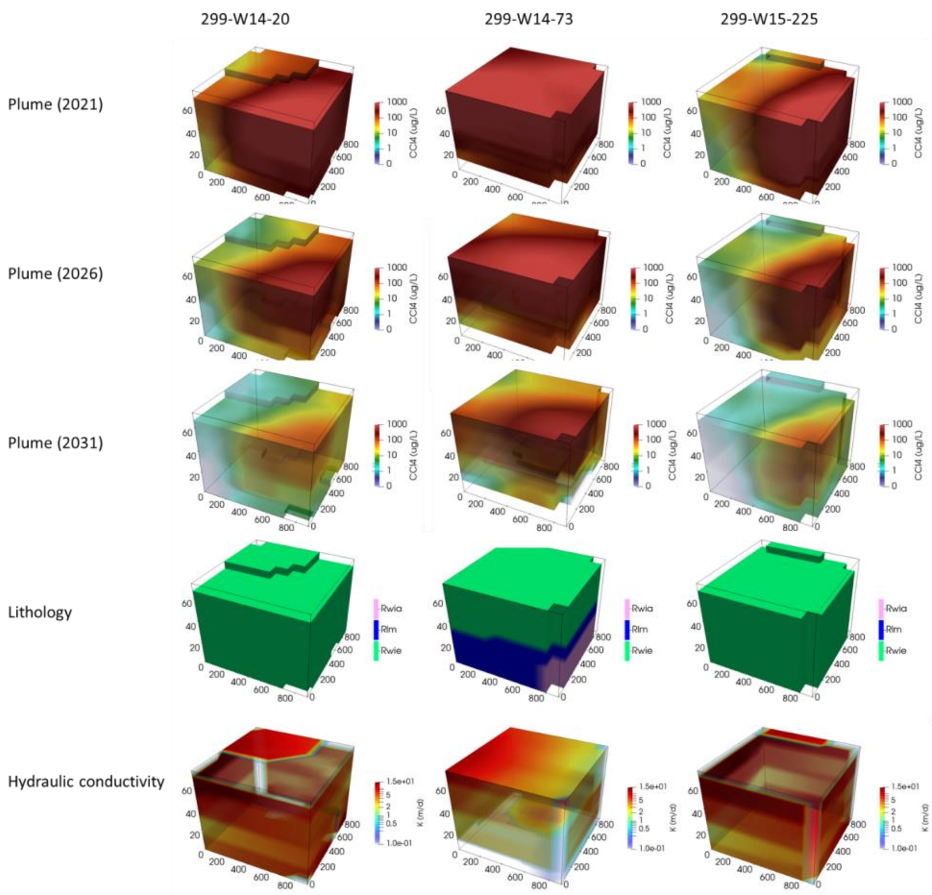


Figure 6~~-~~. Examples of 3D training image sets from multiple data sources at three different wells. Each well is located at the center of the training images.

### 3.3. DL model setup

Based on the performance ranking rule defined in Section 3.1, a total of 486 pairs of label (performance ranking) and training image datasets were created for the 28 wells at 18 timesteps. Note that not all wells were operated throughout the entire simulation time frame. We first used

80% of the dataset for model training and validation (year 2015~2029) and the remaining 20% of the dataset for model testing (year 2029~2032). Each scenario described in Section 3.2 shares the same 3D-CNN model architecture and grid searching in the hyperparameter tuning process to make a fair comparison across different scenarios.

### 3.4. Model accuracy measures

The accuracy of the trained ML model was evaluated in the testing datasets, which are the last 20% of the synthetic data (year 2029~2032). The direct model predictions were plotted in corresponding years to examine misclassified testing points spatiotemporally. The commonly used confusion matrix was applied to quantify the MC3D-CNN model performance using the testing set. The confusion matrix visualizes the accuracy of a classifier by comparing the actual and predicted classes (e.g., well performance rankings in this study). For a binary classification problem, the four types of events in its confusion matrix and their meanings are as follows:

(1) True Positive (T.P.), correctly predicted true positives.
(2) False Negative (F.N.), true positives predicted as negative values.
(3) False Positive (F.P.), true negatives values predicted as positives
(4) True Negatives (T.N.), correctly predicted true negatives

For multi-class classification, T.P., F.N., F.P., and T.N. need to be determined for each class separately by lumping all other classes into one class. Then, the confusion matrix was used to assess the accuracy, precision, sensitivity, and specificity of each class to diagnose the model performance. Accuracy represents the percentage of correctly labeled events within the whole pool of events. Precision uses the portion between T.P. and (T.P.+F.P.) to assess how good the model is at assigning positive events to the positive classes. Sensitivity is the fraction between T.P. and (T.P.+F.N.), which measures how proper the model is for detecting events in the positive class. Specificity is the ratio of T.N. and (T.N.+F.P.), which evaluates how exact the assignment to the negative class is.

### 3.5. Future well performance prediction

The trained DL model was used to predict the future well performance ranking in year 2022 for the entire model domain to demonstrate the usage of this model. The unconfined aquifer in the 200 West Area was first discretized to a structured grid with a spatial resolution of $100 \times 100 \times 5\ m$. Imaginary wells were placed in each grid node with a predefined screen length of 48 m, which is the medium value of the existing wells. Same as the training dataset, a series of $10 \times 10 \times 13$ image sets were extracted from $CCl_4$ plume, hydraulic conductivity, and hydrostratigraphic unit datasets as inputs for model prediction. The future well performance ranking was then generated using the trained MC3D-CNN model for each of the gridded locations, which were then used to create a map of well performance ranking.

## 4. Results

We trained DL models for the three designed model configurations, and in Section 4.1 we evaluate their accuracy. The trained models achieved over 90% accuracy on the training and validation datasets, and provided satisfactory results on the testing set. The trained DL models were then used to provide a site-wide ranking map to illustrate the usage of this method (Section 4.2).

### 4.1. Model evaluation under different scenarios

The CNN model was trained with various configurations for well performance evaluation and the optimal model configuration was applied to the third independent testing dataset. The model training history and class statistics were calculated and are illustrated in Figure 7. Figure 7 (a-c) represents the influence of the optimized model configurations (e.g., the model accuracy vs. epochs curves) with three settings(scenarios) of predictors: single plume (S1), multi-step plume (S2), and multi-step plume with field properties (S3). Both the training and validation accuracy increased with epochs and converged at or above 90% for all three scenarios. With more data channels added to the training pool, the model accuracy increased. Although the S3 model has the highest training and validation accuracy, the overall averaged accuracy for all three models was satisfactory without noticeable over-fitting. It is not surprising that S2 and S3 yielded similar learning results because (1) plume distribution is the most important control factor on extraction well performance and (2) the impacts of hydraulic conductivity and hydrostratigraphic units have been implicitly represented in the multi-year plume variations.

Figure 7 (d-f) is a multi-class confusion matrix for the testing set; specifically, the diagonal line stands for the matched cases between predictions and targets for each class, the upper corners are the overestimated cases, and the lower corners are the underestimated cases. In general, the results show that the testing accuracy across the three scenarios is lower than training and validation accuracy. The reason is that the model was trained using 12 years of historical data to predict 4 years of future behaviors. This indicates that the aquifer conditions were nonstationary due to the continuous decrease of $CCl_4$ inventory. For the single-plume scenario (S1), the overall model accuracy for the testing dataset is 67%. The model has about 23% underestimation on well performance, where 7 high-performance-ranking cases and 14 medium-performance-ranking cases were classified as low-performance-ranking. The model has good performance in controlling the overestimation, with only 8 low-performance-ranking cases classified as medium. After adding multi-time channel data (S2 and S3), the averaged model accuracy increases to 73% and the improvement is observed for both medium and high classes. The underestimated rate was reduced to 14%, which is 8.6% lower than S1. The overestimated rate was 2% higher compared to the single-plume scenario. The DL models trained by S2 and S3 have comparable performance.

Figure 7 (g-h) shows class statistics using four metrics calculated from the confusion matrix. For the single-plume scenario (S1), the sensitivity for the high and medium classes is low because the number of T.P. for both classes is small, which means the model needs to be improved to better detect high- and medium-performance cases. The specificity of the low class is 46%, which means the model cannot assign low class exactly; in this case, medium- and high-performance-ranking cases are likely to be predicted as low performance. All four performance statistics improve after adding multi-step temporal information (S2 model). In the high-performance-ranking class of the S2 model, accuracy and precision reached 82% and 80%, respectively. The high precision in predicting the high-performance-ranking class means that the model is good at predicting high-performance-ranking cases. The S3 model demonstrates slightly better sensitivity for the low-performance-ranking class, and its overall statistical performance is similar to that of the S2 model. Based on the comparisons of the above statistical performance metrics, S3 was selected and applied to the testing dataset and the field prediction described in the following sections.

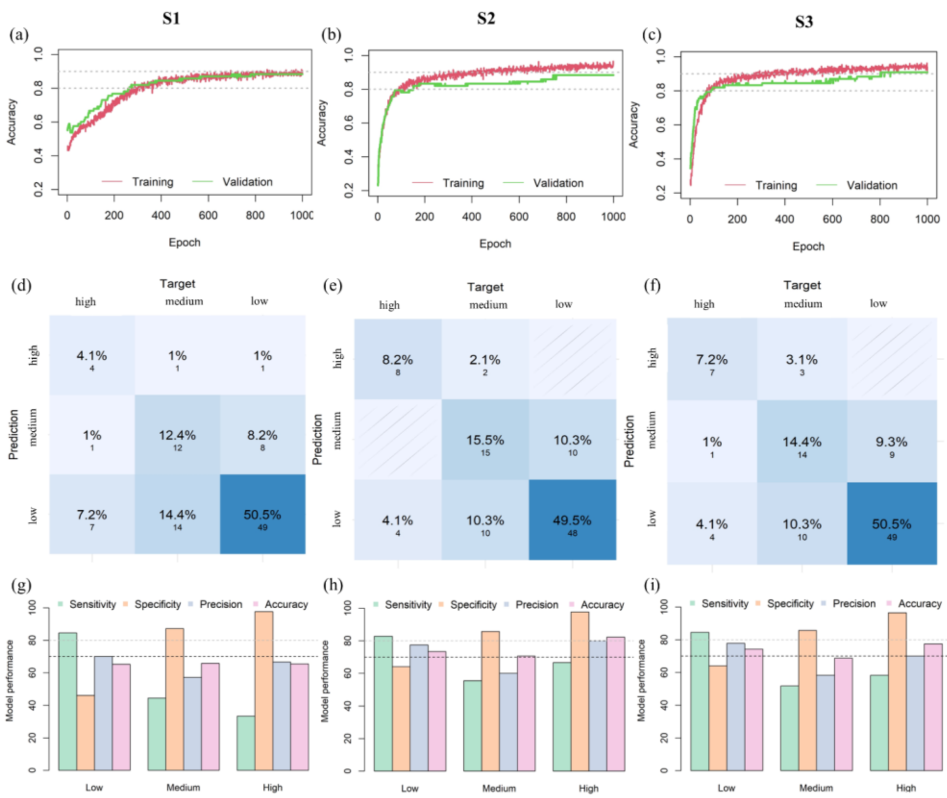


Figure 7. Model optimization and performance evaluation for different scenarios; each column represents a pre-designed scenario: (a-c) model training history under the most suitable configuration; (d-f) multi-class confusion matrix; (g-i) model performance metrics on the testing set.

Figure 8 shows the spatial distributions of correct prediction, underestimation, and overestimation cases by the S3 model. Each dot under the subplots represents a testing well location for a particular year. In general, model predictions match the label references for most wells. However, since the mass recovery data extracted from the P2R model tends to decrease over time and the model predicts future well performance, the S3 model tends to underestimate for later years. These results suggest that the deep learning model could be further improved by including more representative spatiotemporal samples into the training set.

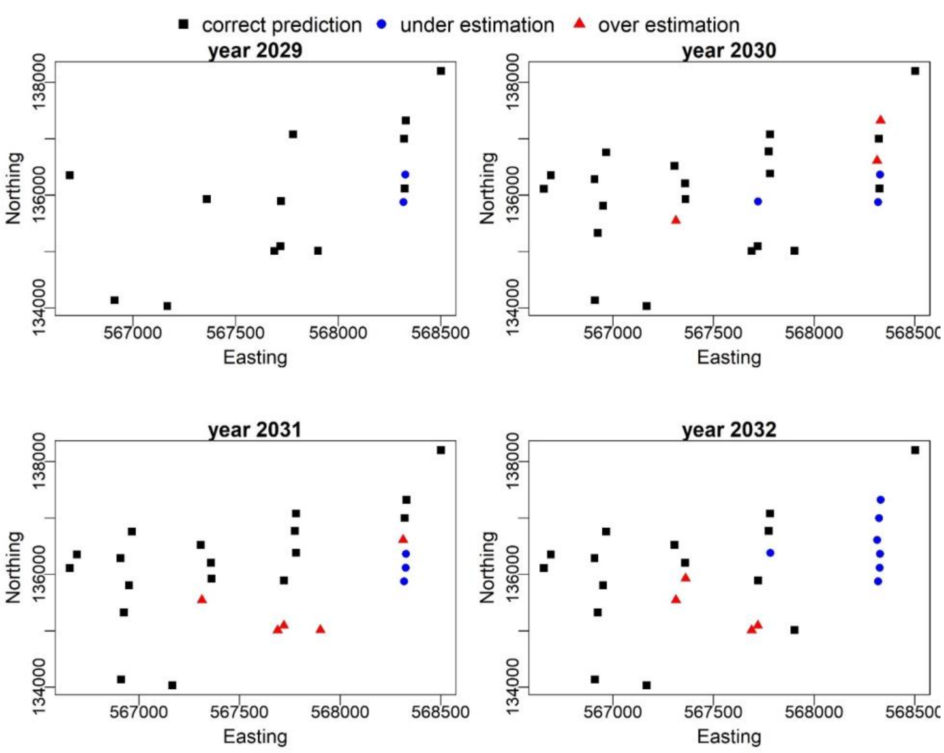

Figure 8. Spatial distribution of predictive well performance ranking obtained from the S3 model at the testing time period (year 2029~2032).

### 4.2. Well performance ranking prediction

The predicted well performance ranking map covering the entire field is obtained using the well-trained model. Based on the input dataset requirements of our model, the entire field is sliced into 10x10x13 image sets within the 200 West Area. Each image set is assigned a performance ranking for the next 5 years, predicted by the MC3D-CNN model as a potential well location, with a given plume concentration image at any time step, hydrostratigraphic unit, and hydraulic conductivity. Figure 9 presents the resulting performance ranking prediction, which can be used to assist in future well planning.

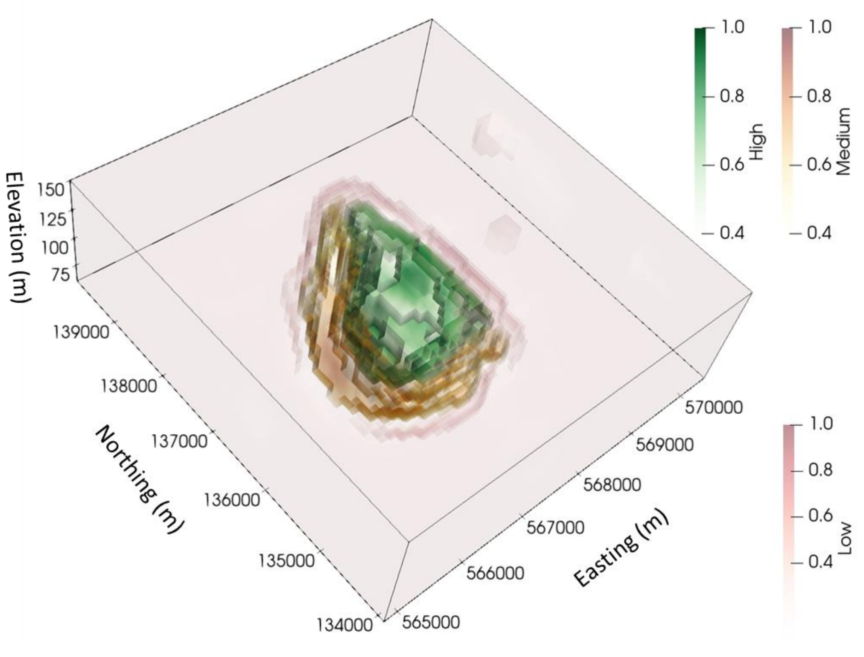

Figure 9. 3D performance ranking map. Green, orange, and red indicate the locations of high, medium, and low well performance rankings, respectively. For each category, the level of transparency indicates the confidence of the CNN model prediction. The more transparency, the less confidence there is in CNN model prediction. The transparency of the low-ranking areas (red) is further scaled down to 10% to show the internal high- and medium-ranking areas in the domain.

The centroid plume concentration of each grid in Figure 9 was clustered according to its performance ranking. The results are shown in Figure 10 using a violin plot, which is a hybrid of a boxplot and kernel density plot that can visualize the distribution of a numeric variable in different groups. It is not surprising that the low-ranking grids have much lower concentration compared to the medium- and high-ranking grids. The concentration distributions of the medium and high classes are similar. It is interesting to see that many grids with high concentration were ranked as medium while many other grids with lower concentration were ranked as high. This highlights the value of using image-based classification instead of point measurement for decision-making.

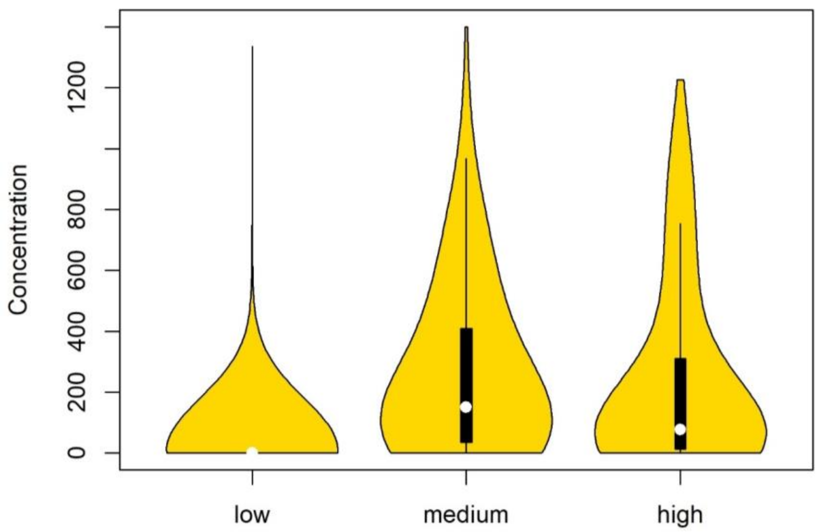


Figure 10. Violin plot of the cell grid concentration vs. predictive performance ranking (white dot is the median of each group, the black box is the boxplot with whisker range, the yellow violin shows the probability density).

## 5. Conclusion and Discussion

We propose a DL framework that can predict the future performance of extraction wells by mining the hidden relationship between historical P&T records and contaminant plume distribution, and aquifer hydrogeological properties. Inspired by the success of DL applications in computer image recognition applications, we formulated the well performance prediction as a classification problem similar to image recognition. The resulting DL model can identify key patterns of subsurface properties that control well extraction by learning from pre-labeled historical records and paired subsurface property images. The trained DL model can capture such key patterns around new wells and classify them as high-, medium-, and low-performance-ranking locations. The advantage of this DL method compared to analytical well models is that it can look deep into the heterogeneous nature of aquifer conditions. Although rigorous physical constraints were not explicitly imposed in the DL framework, they were implicitly included by selecting the key physical factors as the training image for model inputs. The deep learning framework is also adaptive to recruit new data whenever they become available to retrain and improve the deep learning model with small computational cost. Compared to numerical F&T models, this approach is much more portable and makes full use of historical P&T records, and thus is suitable for data-driven decision-making and adaptive site management (ASM) to reduce remediation time and cost, as discussed in Section 5.1. The future development of this method is also discussed in Section 5.1, and the limitations are discussed in Section 5.2.

### 5.1. Application scenario of the well performance ranking tool

The data-driven and portable features of the DL approach made it a useful tool for ASM. Here, we refer to ASM as a systematic and iterative management strategy that routinely re-evaluates and prioritizes site remedial actions and characterization activities to expedite the remediation of large and/or complex sites (Demirkanli and Freedman, 2021). Prediction of extraction well performance at specific locations is critical for ASM for evaluating and revising P&T well network design and operation, setting reasonable remedy targets, and estimating remedy costs. Although this method is data-driven, it doesn't incur additional costs for collecting new data because the well mass recovery records are often routinely recorded during the P&T operation, subject to the regulatory requirement. The DL model can also be easily re-trained with better accuracy for continuous planning whenever new data are available. During the routine evaluation of remediation progress, the updated well performance ranking map can assist in planning new well locations, rebalancing the pumping rates for existing wells, or even turning off some low-performance-ranking wells to better use their treating volume for high-performance wells.

Another application scenario for this performance prediction model is to integrate it with optimization workflows for P&T well network design. There is a long history of developing and applying integrated simulation-based optimization approach for P&T system design (Khan et al., 2004; Maskey et al., 2002; Mayer et al., 2002; McKinney and Lin, 1996; Wagner and Gorelick, 1987; Zheng and Wang, 2002). In a typical simulation-based optimization application, optimization search algorithms [e.g., differential evolution (Bayer et al.), generic evolution (Park, 2016), particle swarm (Mategaonkar and Eldho, 2014), firefly (Kazemzadeh-Parsi et al., 2015), and others] are used to drive a groundwater simulator iteratively to check whether the environmental and/or hydraulic constraints were met with certain P&T configuration parameters (e.g., new well location, pumping rates, and pumping duration), and then adjust these configuration parameters accordingly. The computational cost of the optimization problem grows exponentially with the number of parameters and makes formal optimization nearly impossible for complex waste sites with a large number of wells. The data-driven MD3D-CNN model is much more portable comparing to expensive numerical F&T models, and it is also more adaptive to recruit new data whenever they become available. These make it ideal as~~The MD3D-CNN well performance prediction model can be used as~~ a filtering tool to reduce the number of candidate well locations for the optimization search algorithms so that limited computational resources can be concentrated on more promising well installation plans. The performance prediction model can also be integrated into groundwater simulators (e.g., the Hanford Site's P2R model) and provide direct on-the-fly optimization. In such cases, the model can be used as a wrapper of the groundwater simulator that pauses the simulator periodically and then rebalances the extraction rates among wells based on their performance ranking to achieve better mass recovery.

### 5.2. Limitations and future development

We demonstrated this performance prediction model using the model-simulated $CCl_4$ of a real complex remediation project located on the Hanford Site. The model simulation results provided a known answer so that accuracy and mismatch of the DL models could be precisely measured and traced. Although the scope of this study is limited to developing and demonstrating the new method, one question to answer is how to apply this method in field as there are no known "subsurface images" such as the exact plume distribution provided by the calibrated P2R model. Our next step will address this question with the following three approaches:

- ~~Geospatial~~ Geostatistical simulation and ensemble prediction. Due to the limited sampling and monitoring data in any real remediation site, it is clearly impossible to have exact plume distribution for the performance prediction model as model input. However, a remediation project always has some type of estimation of the plume distribution, which is essential for decision-making. Such a plume distribution estimation can be used as a training image for the performance ranking tool. An even better method would be to train multiple DL models using the geostatistical realization of the plume distribution (Murray and Bott, 2008) to augment the training pool to correct model bias. The site uncertainty can be reduced by combining all the DL model results to provide a more representative or accurate ensemble prediction.

- Incorporating multiple data inputs. Although the demonstration case only used three key aquifer properties as inputs, the MC3D-CNN architecture is flexible and can be readily extended to include other variables. One important potential training image dataset is geophysical investigations, which can provide high spatial-temporal resolution snapshots of subsurface measurements.
- Training the DL model with the aid of numerical model results. In sites with very limited subsurface measurement, numerical simulation results can be added as a supplementary dataset for model training.

This study provides a DL framework to make better use of P&T records for future remediation design. P&T remedy monitoring and operational data, along with site investigation information, applied to data-informed approaches such as the one tested in this study, can create opportunities to improve our understanding of contaminant transport, provide flexible tools for site management, streamline decision-making, and potentially reduce remediation costs.

## Acknowledgments

Funding for this work was provided by the U.S. Department of Energy Richland Operations Office under the Deep Vadose Zone – Applied Field Research Initiative. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the Department of Energy under Contract DE-AC05-76RL01830.

**Supporting information**

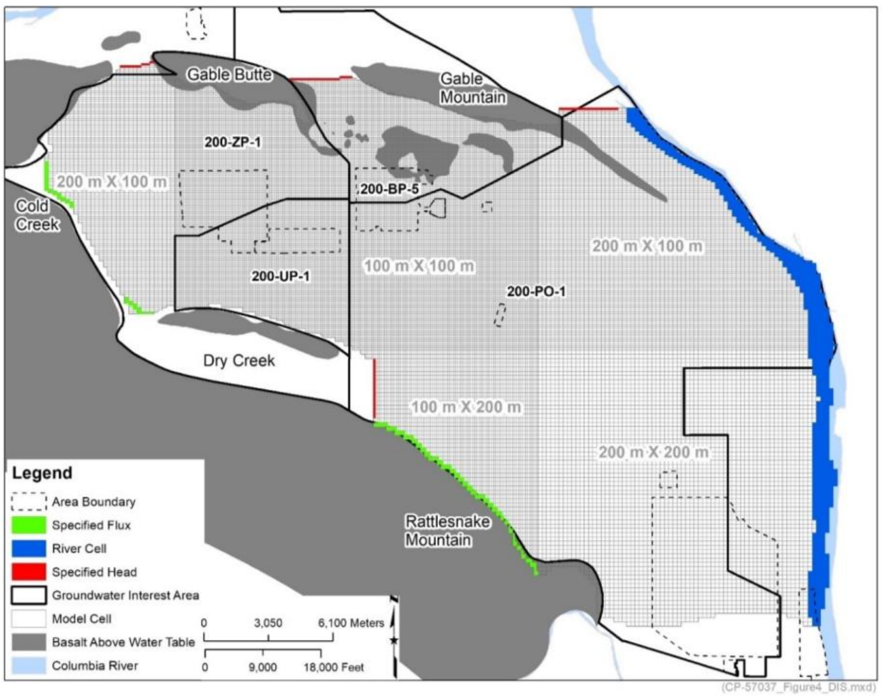


Figure SI. P2R version 8.3 model extent and boundary conditions. (Source: CP-57037, Model Package Report: Plateau to River, Version 8.3, Rev. 2.)

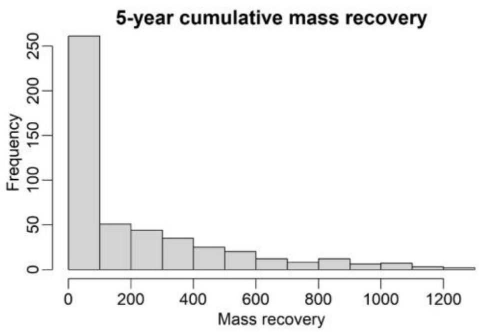


Figure S2. Histogram of 5-year cumulative mass recovery.

**References**

Ameli, A.A., Craig, J.R., 2018. Semi-analytical 3D solution for assessing radial collector well pumping impacts on groundwater–surface water interaction. Hydrology Research, 49(1): 17-26. DOI:10.2166/nh.2017.201

Bayer, P., Paly, d.M., Bürger, C.M., 2010. Optimization of high-reliability-based hydrological design problems by robust automatic sampling of critical model realizations. Water Resour. Res., 46(5). DOI:10/d8sv7r

Brusseau, M.L., 1996. Evaluation of Simple Methods for Estimating Contaminant Removal by Flushing. Groundwater, 34(1): 19-22. DOI:10.1111/j.1745-6584.1996.tb01860.x

Brusseau, M.L., 2013. Use of Historical Pump-and-Treat Data to Enhance Site Characterization and Remediation Performance Assessment. Water Air Soil Pollut, 224(10): 1741. DOI:10/gjhfq9

Budge, T., Nichols, W., 2020. Model Package Report: Plateau to River Groundwater Model Version 8.3. CP-57037-Rev.2.

Cardiff, M., Liu, X., Kitanidis, P.K., Parker, J., Kim, U., 2010. Cost optimization of DNAPL source and plume remediation under uncertainty using a semi-analytic model. Journal of Contaminant Hydrology, 113(1): 25-43. DOI:10/c75h97

Carrera, J., 1993. An overview of uncertainties in modelling groundwater solute transport. Journal of Contaminant Hydrology, 13(1): 23-48. DOI:10.1016/0169-7722(93)90049-X

Chen, Y., Liu, G., Huang, X., Meng, Y., 2022. Groundwater Remediation Design Underpinned By Coupling Evolution Algorithm With Deep Belief Network Surrogate. Water Resour Manage, 36(7): 2223-2239. DOI:10.1007/s11269-022-03137-w

Demirkanli, D.I., Freedman, V.L., 2021. Adaptive Site Management Strategies for the Hanford Central Plateau Groundwater. PNNL-32055.

Demirkanli, D.I. et al., 2018. Assessment of Pump-and-Treat System Impacts on 200 West Aquifer Conditions. PNNL--28063, 1490801.

EPA, 2002. Groundwater Remedies Selected at Superfund Sites. EPA-542-R-01-022, Washington, D.C.

EPA, 2005. Cost-effective design of pump and treat systems. EPA 542-R-05-008, Washington, D.C.

Finsterle, S., 2006. Demonstration of optimization techniques for groundwater plume remediation using iTOUGH2. Environmental Modelling & Software, 21(5): 665-680. DOI:10.1016/j.envsoft.2004.11.012

Finsterle, S., Zhang, Y., 2011. Solving iTOUGH2 simulation and optimization problems using the PEST protocol. Environmental Modelling & Software, 26(7): 959-968. DOI:10.1016/j.envsoft.2011.02.008

Gaur, S., Ch, S., Graillot, D., Chahar, B.R., Kumar, D.N., 2013. Application of Artificial Neural Networks and Particle Swarm Optimization for the Management of Groundwater Resources. Water Resour Manage, 27(3): 927-941. DOI:10/f4mwmv

Gaur, S., Chahar, B.R., Graillot, D., 2011. Analytic elements method and particle swarm optimization based simulation–optimization model for groundwater management. Journal of Hydrology, 402(3): 217-227. DOI:10/cz25k2

Hadley, P.W., Newell, C.J., 2012. Groundwater Remediation: The Next 30 Years. Groundwater, 50(5): 669-678. DOI:10.1111/j.1745-6584.2012.00942.x

Haley, J.L., Lang, D.J., Herrinton, L., 1989. EPA's approach to evaluating and cleaning up ground water contamination at Superfund sites. Ground Water Monitoring Review; (USA), 9:4. DOI:10.1111/j.1745-6592.1989.tb01027.x

Hammond, G.E., Lichtner, P.C., 2010. Field-scale model for the natural attenuation of uranium at the Hanford 300 Area using high-performance computing: MODEL FOR NATURAL ATTENUATION OF URANIUM. Water Resour. Res., 46(9). DOI:10.1029/2009WR008819

Hirschmiller, J., Biryukov, A., Groulx, B., Emmerson, B., Quinell, S., 2019. The Importance of Integrating Subsurface Disciplines with Machine Learning when Predicting and Optimizing Well Performance – Case Study from the Spirit River Formation, Day 2 Tue, October 01, 2019, Calgary, Alberta, Canada, pp. D021S025R004. DOI:10/gkq3mb

Huang, C., Mayer, A.S., 1997. Pump-and-treat optimization using well locations and pumping rates as decision variables. Water Resour. Res., 33(5): 1001-1012. DOI:10/bjzdwq

Kazemzadeh-Parsi, M.J., Daneshmand, F., Ahmadfard, M.A., Adamowski, J., 2015. Optimal Remediation Design of Unconfined Contaminated Aquifers Based on the Finite Element Method and a Modified Firefly Algorithm. Water Resour Manage, 29(8): 2895-2912. DOI:10/f7ccpn

Khan, F.I., Husain, T., Hejazi, R., 2004. An overview and analysis of site remediation technologies. Journal of Environmental Management, 71(2): 95-122. DOI:10.1016/j.jenvman.2004.02.003

Kontos, Y.N. et al., 2022. Machine learning for groundwater pollution source identification and monitoring network optimization. Neural Comput & Applic. DOI:10.1007/s00521-022-07507-8

Kontos, Y.N., Katsifarakis, K.L., 2017. Optimal management of a theoretical coastal aquifer with combined pollution and salinization problems, using genetic algorithms. Energy, 136: 32-44. DOI:10/gb4sv6

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature, 521(7553): 436-444.

Li, Y., Sun, R., Horne, R., 2019. Deep Learning for Well Data History Analysis. DOI:10.2118/196011-MS

Majumder, P., Eldho, T.I., 2016. A New Groundwater Management Model by Coupling Analytic Element Method and Reverse Particle Tracking with Cat Swarm Optimization. Water Resour Manage, 30(6): 1953-1972. DOI:10/f8hgcp

Majumder, P., Lu, C., 2021. A novel two-step approach for optimal groundwater remediation by coupling extreme learning machine with evolutionary hunting strategy based metaheuristics. Journal of Contaminant Hydrology, 243: 103864. DOI:10.1016/j.jconhyd.2021.103864

Maskey, S., Jonoski, A., Solomatine, D.P., 2002. Groundwater Remediation Strategy Using Global Optimization Algorithms. Journal of Water Resources Planning and Management, 128(6): 431-440. DOI:10.1061/(ASCE)0733-9496(2002)128:6(431)

Mategaonkar, M., Eldho, T.I., 2014. Multiobjective Groundwater Remediation Design Using a Coupled MFree Point Collocation Method and Particle Swarm Optimization. Journal of Hydrologic Engineering, 19(6): 1259-1263. DOI:10/f54mkm

Matott, L.S., Rabideau, A.J., Craig, J.R., 2006. Pump-and-treat optimization using analytic element method flow models. Advances in Water Resources, 29(5): 760-775. DOI:10/fq9dkh

Mayer, A.S., Kelley, C.T., Miller, C.T., 2002. Optimal design for problems involving flow and transport phenomena in saturated subsurface systems. Advances in Water Resources, 25(8): 1233-1256. DOI:10.1016/S0309-1708(02)00054-4

Mayer, K.U., Blowes, D.W., Frind, E.O., 2001. Reactive transport modeling of an in situ reactive barrier for the treatment of hexavalent chromium and trichloroethylene in groundwater. Water Resour. Res., 37(12): 3091-3103. DOI:10.1029/2001WR000234

McConnell, L. et al., 2022. Forecasting Groundwater Contaminant Plume Development Using Statistical and Machine Learning Methods. Groundwater Monit R. DOI:10.1111/gwmr.12523

McKinney, D.C., Lin, M.-D., 1996. Pump-and-Treat Ground-Water Remediation System Optimization. Journal of Water Resources Planning and Management, 122(2): 128-136. DOI:10.1061/(ASCE)0733-9496(1996)122:2(128)

McMahon, A., Heathcote, J., Carey, M., Erskine, A., 2001. Guide to good practice for the development of conceptual models and the selection and application of mathematical models of contaminant transport processes in the subsurface. National Groundwater & Contaminated Land Centre. Environment Agency. UK. Report NC/99/38, 2.

Meray, A.O. et al., 2022. PyLEnM: A Machine Learning Framework for Long-Term Groundwater Contamination Monitoring Strategies. Environ. Sci. Technol., 56(9): 5973-5983. DOI:10.1021/acs.est.1c07440

Minsker, B., Zhang, Y., Greenwald, R., Peralta, R., Zheng, C., 2004. Application of Flow and Transport Optimization Codes to Groundwater Pump and Treat Systems- Volume III, Fort Belvoir, VA.

Mo, S., Zabaras, N., Shi, X., Wu, J., 2019. Deep Autoregressive Neural Networks for High-Dimensional Inverse Problems in Groundwater Contaminant Source Identification. Water Resour. Res., 55(5): 3856-3881. DOI:10.1029/2018WR024638

Murray, C., Bott, Y.-J., 2008. Revised Geostatistical Analysis of the Inventory of Carbon Tetrachloride in the Unconfined Aquifer in the 200 West Area of the Hanford Site. DOI:10.2172/945229

National Research, C., 1994. Alternatives for Ground Water Cleanup.

National Research, C., 2013. Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites.

Neville, C., Tonkin, M., 2004. Modeling multiaquifer wells with MODFLOW. Ground water, 42: 910-9. DOI:10.1111/j.1745-6584.2004.t01-9-.x

Park, Y.-C., 2016. Cost-effective optimal design of a pump-and-treat system for remediating groundwater contaminant at an industrial complex. Geosci J, 20(6): 891-901. DOI:10/gjhfmk

Rao, C., Liu, Y., 2020. Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization. Computational Materials Science, 184: 109850. DOI:10.1016/j.commatsci.2020.109850

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res., 48(7). DOI:10.1029/2011WR011527

Ren, H., Cromwell, E., Kravitz, B., Chen, X., 2022. Technical note: Using long short-term memory models to fill data gaps in hydrological monitoring networks. Hydrology and Earth System Sciences, 26(7): 1727-1743. DOI:10.5194/hess-26-1727-2022

Rodriguez-Galiano, V., Mendes, M.P., Garcia-Soldado, M.J., Chica-Olmo, M., Ribeiro, L., 2014. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). Science of The Total Environment, 476-477: 189-206. DOI:10.1016/j.scitotenv.2014.01.001

Rogers, L.L., Dowla, F.U., 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. Water Resour. Res., 30(2): 457-481. DOI:10.1029/93WR01494

Shen, C., 2018. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resour. Res., 54(11): 8558-8593. DOI:10.1029/2018WR022643

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs].

Singh, A., Minsker, B.S., 2008. Uncertainty-based multiobjective optimization of groundwater remediation design. Water Resour. Res., 44(2). DOI:10.1029/2005WR004436

Soriano, M.A. et al., 2021. Assessment of groundwater well vulnerability to contamination through physics-informed machine learning. Environ. Res. Lett., 16(8): 084013. DOI:10.1088/1748-9326/ac10e0

Steefel, C.I. et al., 2015. Reactive transport codes for subsurface environmental simulation. Comput Geosci, 19(3): 445-478. DOI:10.1007/s10596-014-9443-x

Sun, A.Y., 2018. Discovering State-Parameter Mappings in Subsurface Models Using Generative Adversarial Networks. Geophysical Research Letters, 45(20): 11,137-11,146. DOI:10.1029/2018GL080404

Sváb, M., Zilka, M., Müllerová, M., Kocí, V., Müller, V., 2008. Semi-empirical approach to modeling of soil flushing: model development, application to soil polluted by zinc and copper. Sci Total Environ, 392(2-3): 187-197. DOI:10.1016/j.scitotenv.2007.12.001

Tahmasebi, P., Kamrava, S., Bai, T., Sahimi, M., 2020. Machine learning in geo- and environmental sciences: From small to large scale. Advances in Water Resources, 142: 103619. DOI:10.1016/j.advwatres.2020.103619

Tartakovsky, A.M., Marrero, C.O., Perdikaris, P., Tartakovsky, G.D., Barajas-Solano, D., 2020. Physics-Informed Deep Neural Networks for Learning Parameters and Constitutive Relationships in Subsurface Flow Problems. Water Resour. Res., 56(5): e2019WR026731. DOI:10.1029/2019WR026731

Truex, M. et al., 2017. Performance Assessment of Pump-and-Treat Systems. Groundwater Monit R, 37(3): 28-44. DOI:10/gc2hr8

Tsang, C.-F., Neretnieks, I., Tsang, Y., 2015. Hydrologic issues associated with nuclear waste repositories. Water Resour. Res., 51(9): 6923-6972. DOI:10.1002/2015WR017641

Wagner, B.J., Gorelick, S.M., 1987. Optimal groundwater quality management under parameter uncertainty. Water Resour. Res., 23(7): 1162-1174. DOI:10.1029/WR023i007p01162

Wang, N., Chang, H., Zhang, D., 2021. Deep-Learning-Based Inverse Modeling Approaches: A Subsurface Flow Example. Journal of Geophysical Research: Solid Earth, 126(2): e2020JB020549. DOI:10.1029/2020JB020549

White, M.D., Oostrom, M., 2003. STOMP subsurface transport over multiple phases version 3.0 User's guide.

Wu, C., Fang, C., Wu, X., Zhu, G., 2020. Health-Risk Assessment of Arsenic and Groundwater Quality Classification Using Random Forest in the Yanchi Region of Northwest China. Expo Health, 12(4): 761-774. DOI:10.1007/s12403-019-00335-7

Wu, J., Zeng, X., 2013. Review of the uncertainty analysis of groundwater numerical simulation. Chin. Sci. Bull., 58(25): 3044-3052. DOI:10.1007/s11434-013-5950-8

Yadav, B., Mathur, S., Ch, S., Yadav, B.K., 2018. Data-based modelling approach for variable density flow and solute transport simulation in a coastal aquifer. Hydrological Sciences Journal, 63(2): 210-226. DOI:10.1080/02626667.2017.1413491

Yan, S., Minsker, B., 2006. Optimal groundwater remediation design using an Adaptive Neural Network Genetic Algorithm. Water Resour. Res., 42(5). DOI:10.1029/2005WR004303

Yin, J., Tsai, F.T.C., 2020. Bayesian set pair analysis and machine learning based ensemble surrogates for optimal multi-aquifer system remediation design. Journal of Hydrology, 580: 124280. DOI:10/gk7s8f

Yu, X. et al., 2020. Deep learning emulators for groundwater contaminant transport modelling. Journal of Hydrology, 590: 125351. DOI:10.1016/j.jhydrol.2020.125351

Zhao, X. et al., 2019. A Multi-Branch 3D Convolutional Neural Network for EEG-Based Motor Imagery Classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(10): 2164-2177. DOI:10.1109/TNSRE.2019.2938295

Zheng, C., Wang, P.P., 2002. A Field Demonstration of the Simulation Optimization Approach for Remediation System Design. Groundwater, 40(3): 258-266. DOI:10.1111/j.1745-6584.2002.tb02653.x

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: A review. Journal of Hydrology, 598: 126266. DOI:10.1016/j.jhydrol.2021.126266