AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2016WR019933

# Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S.

S. Sahoo[1], T. A. Russo[1], J. Elliott[2,3], and I. Foster[2,3,4]

[1]Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania, USA, [2]Computation Institute, University of Chicago, Chicago, Illinois, USA, [3]Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, Illinois, USA, [4]Department of Computer Science, University of Chicago, Chicago, Illinois, USA

**Key Points:**
- Groundwater level change can be modeled with high accuracy using machine learning methods
- Model framework does not require subsurface parameters and simulates comparable groundwater levels to numerical models of physical flow
- Seasonal irrigation demand has the highest relevance to groundwater level change compared to climate and streamflow inputs for most wells

**Supporting Information:**
- Supporting Information S1

**Correspondence to:**
T. A. Russo,
russo@psu.edu

**Citation:**
Sahoo, S., T. A. Russo, J. Elliott, and I. Foster (2017), Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S., *Water Resour. Res.*, *53*, 3878–3895, doi:10.1002/2016WR019933.

Received 12 OCT 2016
Accepted 15 APR 2017
Accepted article online 19 APR 2017
Published online 13 MAY 2017



**Abstract** Climate, groundwater extraction, and surface water flows have complex nonlinear relationships with groundwater level in agricultural regions. To better understand the relative importance of each driver and predict groundwater level change, we develop a new ensemble modeling framework based on spectral analysis, machine learning, and uncertainty analysis, as an alternative to complex and computationally expensive physical models. We apply and evaluate this new approach in the context of two aquifer systems supporting agricultural production in the United States: the High Plains aquifer (HPA) and the Mississippi River Valley alluvial aquifer (MRVA). We select input data sets by using a combination of mutual information, genetic algorithms, and lag analysis, and then use the selected data sets in a Multilayer Perceptron network architecture to simulate seasonal groundwater level change. As expected, model results suggest that irrigation demand has the highest influence on groundwater level change for a majority of the wells. The subset of groundwater observations not used in model training or cross-validation correlates strongly ($R > 0.8$) with model results for 88 and 83% of the wells in the HPA and MRVA, respectively. In both aquifer systems, the error in the modeled cumulative groundwater level change during testing (2003–2012) was less than 2 m over a majority of the area. We conclude that our modeling framework can serve as an alternative approach to simulating groundwater level change and water availability, especially in regions where subsurface properties are unknown.

## 1. Introduction

Groundwater dynamics are determined by physical hydrogeological properties and system boundary conditions, including climate variability and pumping. However, nonlinear interactions, spatial heterogeneity, and temporal lags between these processes can be challenging to characterize in real-world systems. Improved understanding of groundwater level response to climate patterns and pumping is essential for sustainable planning and management, particularly in the context of increasing groundwater demands in agriculture [*Wada et al.*, 2010] and climate change impacts on groundwater quantity and quality [*Gurdak et al.*, 2007; *Green et al.*, 2011; *Taylor et al.*, 2012]. Indeed, quantifying the availability of groundwater and long-term impacts of climate variability on groundwater is more complex than for surface water [*Alley et al.*, 2002; *Holman*, 2006]. A key challenge is to predict how changes in one system component will impact others. Such predictions require that we quantify relationships among climate drivers, surface hydrology, agricultural water use, and groundwater responses.

The complexity of this system of systems makes accurate physical process simulations challenging due to the large agricultural and hydrogeological data requirements for model development and calibration. Thus, it is appealing to consider data-driven and machine learning methods based on nonlinear interdependencies that may be able to predict groundwater level change without deep knowledge of the underlying physical parameters. Machine learning methods recognize patterns hidden in historical data and then apply those patterns to predict future scenarios. Researchers have applied a variety of machine learning models in groundwater modeling, including artificial neural network (ANN) [*Coulibaly et al.*, 2001; *Adamowski and Chan*, 2011; *Sahoo and Jha*, 2013; *Nourani et al.*, 2015], fuzzy theory [*Kurtulus and Razack*, 2010; *Güler et al.*, 2012], genetic programming [*Shiri and Kisi*, 2011; *Kasiviswanathan et al.*, 2016], autoregressive models [*Knotters and Bierkens*, 2001; *Bidwell*, 2005; *Chang et al.*, 2016], and support vector machine [*Behzad et al.*, 2010; *Yoon et al.*, 2011].

Assessments of machine learning applications in hydrology suggest that such methods can achieve performance comparable to, or even more accurate than, that of numerical models of physical transport [*Coppola et al*., 2003; *Parkin et al*., 2007; *Nikolos et al*., 2008; *Chu and Chang*, 2009]. Overall, these studies have demonstrated the ability of statistical and machine learning methods for capturing the nonlinear relation between climate variables and groundwater, particularly in areas in which physically based models would be challenging to implement. However, most of these studies were applied at the regional scale and considered only a handful of wells for prediction, thereby potentially restricting their large-scale applicability. Furthermore, as they did not consider groundwater extraction data as an input, they were not able to demonstrate strong model performance in agricultural regions where groundwater pumping is a significant control.

Input variable selection is crucial to the development of data-driven models and is particularly relevant in water resources modeling [*Maier and Dandy*, 2000; *Galelli et al*., 2014; *Quilty et al*., 2016]. The nonlinearity and complexity associated with hydrological systems make it ineffective to apply widely preferred simple and linear correlation analysis to identify suitable predictors [*May et al*., 2011]. Moreover, preprocessing of input time series is an effective way to deal with high seasonal variability and noise in time series, which also helps improve model precision [*Hanson et al*., 2004; *Wu et al*., 2009; *Wang et al*., 2014].

In the work reported here, we combine both data preprocessing and an improved input variable selection method with a nonlinear regression model to facilitate improved groundwater level prediction. Our method employs concepts from mutual information theory to capture nonlinear dependencies between explanatory variables by using their joint and marginal probability instead of a linear correlation. Mutual information has been successfully applied to find nonlinear relations among variables by many researchers in water resources applications [e.g., *Maier et al*., 2006; *Shrestha et al*., 2009; *Gong et al*., 2013; *Mishra et al*., 2013]. However, it has the disadvantage that even if a predictor has a strong connection with the model output, this information might be redundant if the same information is already provided by another predictor [*Sharma*, 2000; *Fernando et al*., 2009; *Quilty et al*., 2016]. To overcome this problem, we combined mutual information theory and a genetic algorithm to minimize entropy. A variety of genetic algorithm approaches have been employed in the past to address hydrologic problems, such as groundwater remediation [*Kalwij and Peralta*, 2006; *Yan and Minsker*, 2006; *Chang et al*., 2007], reservoir operation [*Kerachian and Karamouz*, 2006], groundwater level prediction [*Jha and Sahoo*, 2015], and water quality [*Yeh et al*., 2006; *Jalalkamali*, 2015]. Our proposed method uses genetic algorithms to identify predictors with less redundancy and more relevance to modeling the desired output parameter.

This study had two main objectives: (1) to quantify the relative influence of climate variability, crop irrigation demand, and streamflow on groundwater level change and (2) to develop an empirical (data-driven) model of the hydrologic system. We used simulated crop irrigation demand as a model input in lieu of unavailable groundwater pumping data. We developed and tested our model in the context of two major agriculture-dominated aquifers, for which we used the model to predict groundwater level change (seasonal, 1980–2012) using high-resolution and noisy input variables. While we tested and applied these methods on well-studied aquifers, we intend that the method be applied in regions lacking high-resolution subsurface information (e.g., hydraulic conductivity and storativity).

Our model differs from previous models in two key aspects: (1) we use a novel input variable selection technique based on singular spectrum analysis (SSA), mutual information, and genetic algorithms to extract the best set of predictors based on their highest significant nonlinear dependence with the predictand and (2) we propose and evaluate an automated hybrid artificial neural network (HANN) model, in an ensemble modeling framework to simulate seasonal groundwater level change. In the following, we present the model structure, the uncertainty in model predictions, and the sensitivity of model predictions to input parameters. We also discuss potential applications arising from the model predictions.

## 2. Materials and Methods

### 2.1. Data Description

To assess performance of the proposed ensemble method, we used the HANN model to predict seasonal groundwater level change in the High Plains aquifer system (HPA) and the Mississippi River Valley alluvial aquifer (MRVA). The HPA is one of the primary agricultural regions in the USA and underlies approximately 450,000 $km^2$ spanning eight states: Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota,

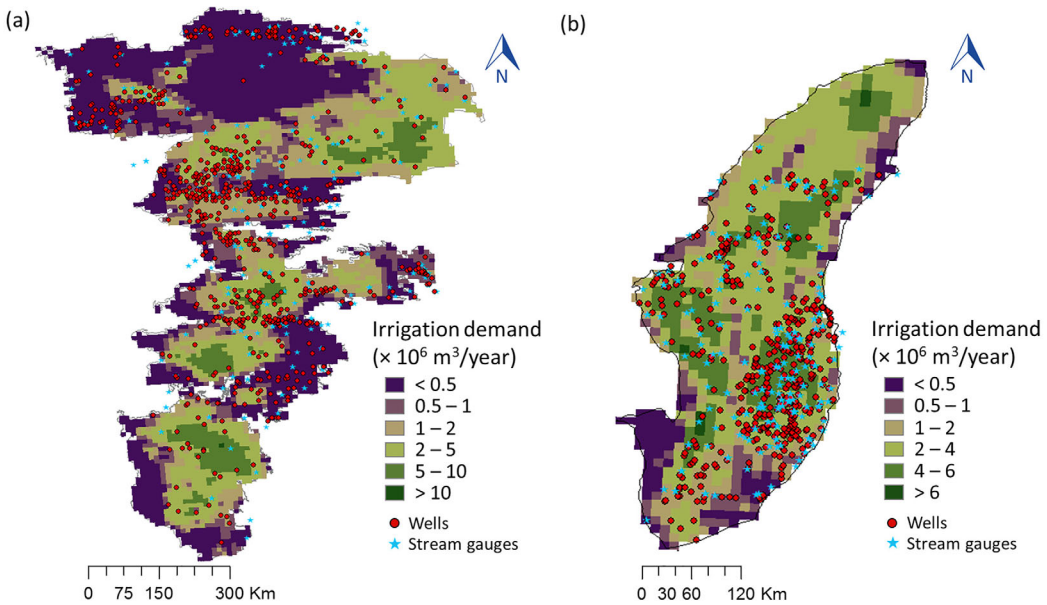


**Figure 1.** (a) HPA and (b) MRVA showing observation wells (red circles), stream gauges (blue stars), and DSSAT-simulated annual average irrigation demand ($m^3$/yr) at 5 arc min resolution. The irrigation demand in the HPA is shown for seven crops: corn, sorghum, soybean, cotton, barley, spring wheat, and winter wheat, and accounts for 83% of the total irrigated cropland. The irrigation demand in the MRVA is shown for five crops: cotton, sorghum, soybean, corn, and winter wheat, and accounts for 93% of the total irrigated cropland.

Texas, and Wyoming. Within the HPA, groundwater is generally under unconfined conditions, with saturated thickness ranging from less than 15 to 366 m [*McGuire*, 2013]. The MRVA underlies 82,880 $km^2$ of Missouri, Kentucky, Tennessee, Mississippi, Louisiana, and Arkansas and supplies increasingly large volumes of water for agriculture [*Clark et al*., 2011]. The MRVA is the uppermost shallow aquifer of the Mississippi Embayment aquifer system and is the primary source of groundwater for irrigation in the agricultural region [*Clark et al*., 2011]. In the HPA and MRVA, groundwater comprises 89 and 94%, respectively, of irrigation water [*Maupin et al*., 2014].

We acquired groundwater level data from 1980 to 2012 (33 years) for 687 wells within the HPA (Figure 1a) and 437 wells within the MRVA (Figure 1b) from the USGS National Water Information System [*USGS*, 2015]. These wells were selected from the complete USGS record based on minimum observation requirements. All wells included in this study had 33 years of monthly groundwater records not exceeding a set threshold of missing values (25%) in the time series. In both the HPA and MRVA, few wells have continuous monthly measurements from 1980 to 2012. Eighteen wells in the HPA and eight wells in the MRVA with records $\geq$25 years ($\geq$300 months) were extracted and interpolated using a spline method, thereby preserving the temporal patterns. The remaining 669 wells in HPA and 429 wells in MRVA with larger and discontinuous gaps during the study period were interpolated at a monthly scale from nearby wells using the Inverse Distance Weighting method (IDW) method. Gridded precipitation and average temperature time series from 1980 to 2012 were acquired from Daymet [*Thornton et al*., 2014]. Monthly stream discharge records from 1980 to 2012 were obtained for 153 stations overlying the HPA and 132 stations overlying the MRVA from the USGS National Water Information System [*USGS*, 2015] (Figure 1). Input data sets were selected based on their presumed influence on groundwater storage, and their availability in the HPA, MRVA, and numerous major agricultural aquifers around the world.

We selected the Multivariate ENSO index (El Niño Southern Oscillation, 2–6 year cycle), Standardized PDO index (Pacific Decadal Oscillation, 10–25 year cycle), and NAO index (North Atlantic Oscillation, 3–6 year cycle) as potential predictor variables. These climate indices, representing coupled ocean-atmospheric phenomena over the Pacific and Atlantic Oceans, were obtained from the National Oceanic and Atmospheric Administration (NOAA) [*NOAA*, 2016a, 2016b, 2016c].

We used the Decision Support System for Agrotechnology Transfer (DSSAT) model [*Jones et al*., 2003] (Figure 1) to simulate irrigation demand. DSSAT simulates the full farm system at high spatial resolution using

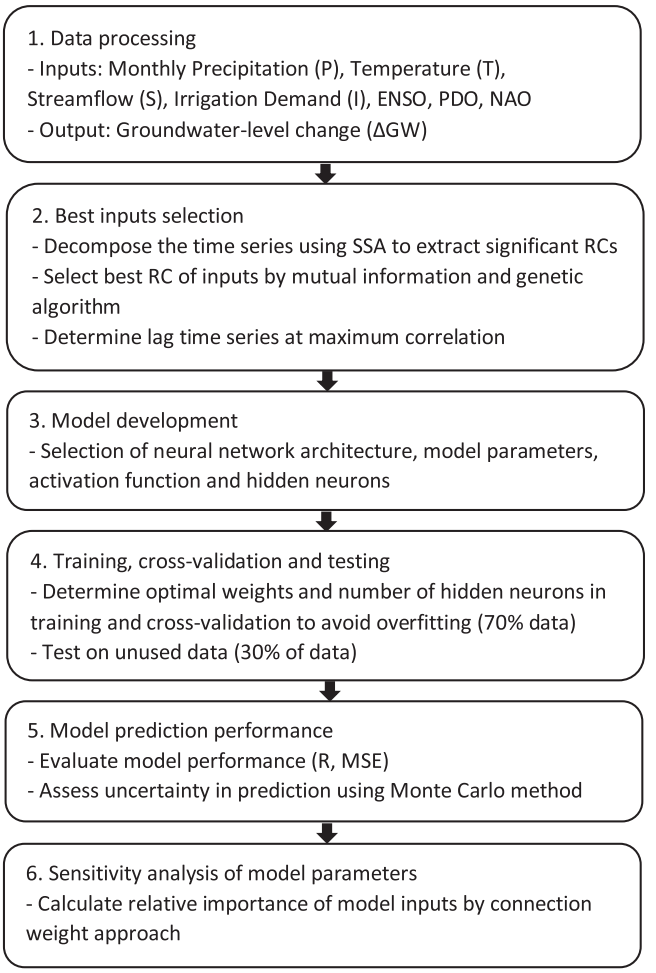


**Figure 2.** Methodological framework of the proposed groundwater model.

the parallel System for Integrating Impacts Models and Sectors (pSIMS) [*Elliott et al*., 2014]. We simulated irrigation demand at 5 arc min resolution (0.08333° or ~10 km grids) from 1980 to 2012 for the HPA using seven major crops (corn, sorghum, soybean, cotton, barley, spring wheat, and winter wheat) and for the MRVA using five crops (cotton, sorghum, soybean, corn, and winter wheat). pSIMS-DSSAT takes historic crop areas, precipitation, temperature, solar radiation, wind speeds, fertilizer application levels, sowing dates and densities, and cultivar characteristics as inputs and produces annual irrigation demands as output. To produce the seasonal demands required by the hydrologic model, we then partitioned the annual total irrigation output from DSSAT by expected demand in winter (DJF), spring (MAM), summer (JJA), and fall (SON) seasons, where expected demand is proportional to the fraction of growing period days in each season and the developmental stages of the crops based on FAO crop water requirements [*Allen et al*., 1998].

### 2.2. Modeling Framework

The HANN model is a machine learning based groundwater model using environmental and human drivers (Figure 2). The HANN model combines a hybrid data preprocessing method which integrates SSA, mutual information, genetic algorithm, and lag analysis with ANN. We present here the model's input variable selection and processing, the model structure, and a method for evaluating the relative importance of the input parameters.

#### 2.2.1. Input Variable Preprocessing

Input variables were selected based on their physical significance as well as on the basis of their statistical relevance to the response variable. Gridded and point input variables were extracted to point-specific data using the Haversine equation corresponding to the groundwater data locations [*Sinnott*, 1984; *Morrison and Munster*, 2015]. The input data were prepared using a three-step preprocessing scheme: singular spectrum analysis (SSA), mutual information, and a genetic algorithm. SSA is a time series frequency analysis technique defined by *Vautard et al*. [1992] and applied by *Dettinger et al*. [1995]. It is a form of principal component analysis in lag-time domain that is used to detect periodic signals in noisy time series data. SSA decomposes time series data into separable and nominally interpretable reconstructed components (RCs) [e.g., *Hanson et al*., 2006; *Kuss and Gurdak*, 2014]. *Hanson et al*. [2004] suggested that the variability in most hydrologic time series can be adequately described in terms of the first 10 RCs; for example, seasonal or diurnal fluctuations may be represented by individual RCs. In this study, the first 10 RCs (e.g., Figure 3) were extracted for each of the time series anomalies (departure from the mean).

Next, we passed these RCs through two steps combining mutual information and a genetic algorithm to extract the best RC. The mutual information method attempts to find the nonlinear dependence between the inputs (RCs) and the output (groundwater level). In this study, mutual information $I(X;Y)$ between each

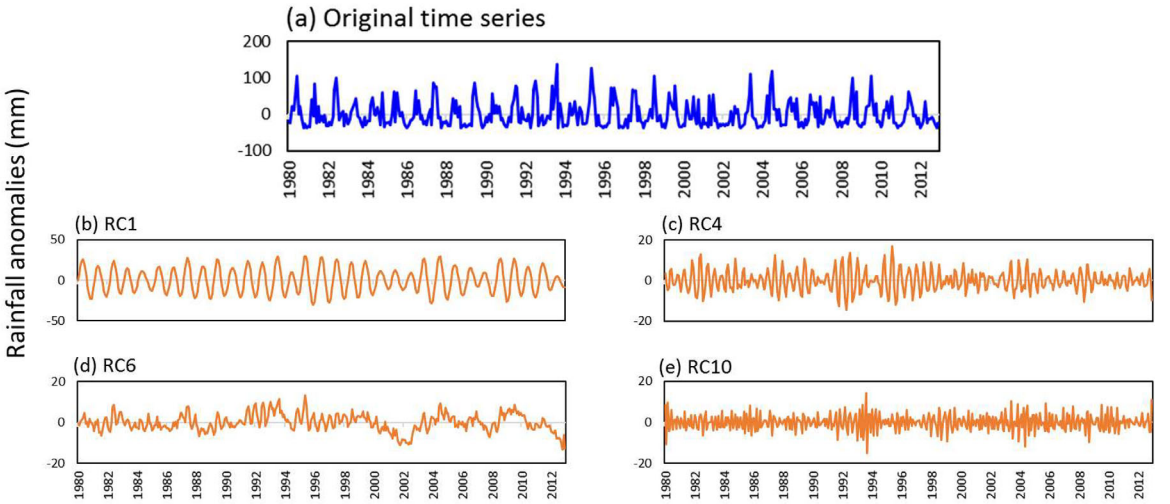


**Figure 3.** (a) Original time series of the rainfall anomalies. (b-e) Decomposed time series (4 RCs shown here out of 10 RCs) using SSA.

input $[X]$ and output $[Y]$ was estimated using Shannon's information theory [*Shannon and Weaver*, 1949; *Cover and Thomas*, 2005], which is expressed as follows:

$$I(X;Y)=H(X)+H(Y)-H(X,Y)=H(X,Y)-H(X|Y)-H(Y|X), \quad (1)$$

where $H(X)$ and $H(Y)$ are the marginal entropies, $H(X|Y)$ and $H(Y|X)$ are the conditional entropies, and $H(X,Y)$ is the joint entropy of $X$ and $Y$. Marginal entropy is the average information provided by observing a variable ($X$ or $Y$), conditional entropy is the average additional information provided by observing $X$, given we have already observed $Y$, and joint entropy is the average total information provided by observing two variables ($X$ and $Y$). More details on the information theory can be found in *MacKay* [2008].

For each input parameter, the most relevant RC was identified with these mutual information values and a genetic algorithm using the principles of maximum relevance ($V$; equation (2)) and minimum redundancy ($P$; equation (3)),

$$V=\frac{1}{n}\sum_{i=1}^{n} I[X_i;Y], \quad (2)$$

$$P=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I[X_i;X_j], \quad (3)$$

$$Z=V-P, \quad (4)$$

where $V$ is the mean of the mutual information values, $I[X_i;Y]$, $X_i$ is the individual input $i$, representing a subset of RCs, and $Y$ is the output, representing groundwater level change anomaly. $P$ is the mean of all the mutual information values, $I[X_i;X_j]$, where $X_i$ and $X_j$ are individual inputs. We solved this optimization problem by using a genetic algorithm to maximize the fitness function, $Z$ (equation (4)), thus maximizing the relevance and minimizing the redundancy of the inputs.

The genetic algorithm was initiated with 50 randomly generated chromosomes with 10 indices each, where each index represents a randomly selected RC. The chromosomes were generated from a uniform distribution of RCs. The fitness of each chromosome is calculated (equation (4)) and the best chromosomes (called parents) are selected based on the ranking of their fitness values. Following ranking of chromosomes by fitness, crossover was performed by selecting pairs of parents to create a new generation. The new chromosomes repeat the procedure of fitness evaluation, ranking, and crossover until the change in fitness between generations reaches a specified closure criterion, or the maximum number of generations (in our case, 100). We also applied the elitist selection method, which replicates the best chromosome in the new generation. The algorithm outputs the indices of the RCs that maximize the fitness function. We evaluated the two best RCs for each input parameter and ultimately selected the single best to use as input in the

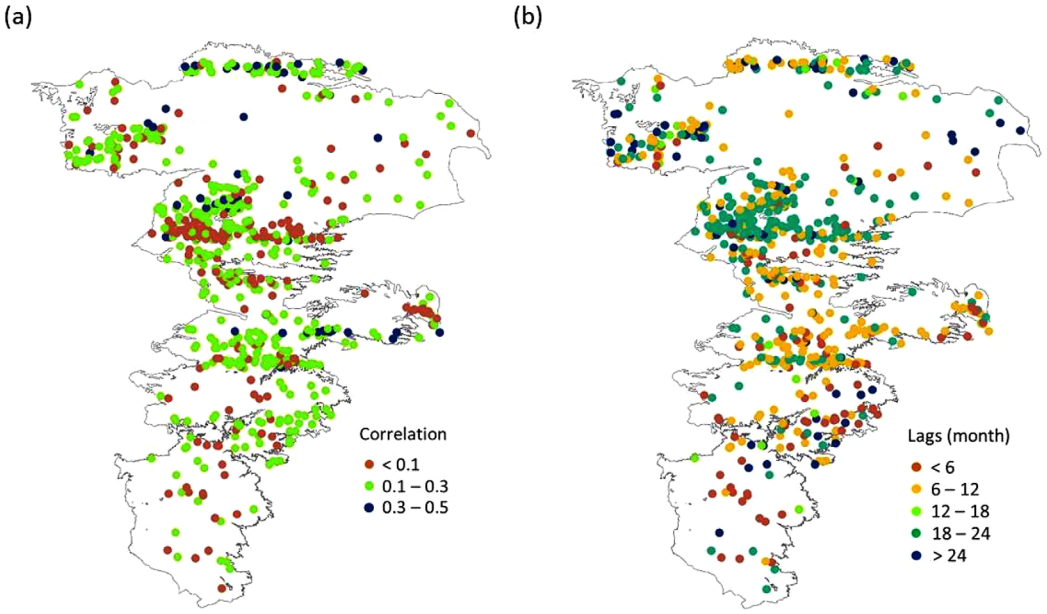


**Figure 4.** (a) Correlations and (b) respective time lags between monthly groundwater level and precipitation over the HPA for 687 well locations.

ANN model. More detailed information on the feature selection by genetic algorithms can be found in *Ludwig and Nunes* [2010].

After selection of the best RC, the time lag components were determined using cross-correlation analysis. We analyzed lag correlations between the best RC of groundwater and the best RC of precipitation, temperature, streamflow, and climate indices, respectively. The resulting time lags for precipitation varied from 6 to 24 months over the HPA, with 6–12 month lags being predominant in the southern and central regions of the aquifer (Figure 4). Similarly, we determined time series lags for temperature, streamflow, ENSO, PDO, and NAO for both the aquifers (Supporting Information Figures S2 and S3). No lag components of irrigation demand were calculated because we assumed that groundwater extraction would have an immediate local impact on groundwater level. For each input, seasonal values were then calculated by averaging monthly values over 3 month periods (December–February, March–May, June–August, and September–November) to get corresponding seasonal time series. For output, seasonal groundwater-level changes were estimated as the difference between first month and last month groundwater levels of each season, winter (DJF), spring (MAM), summer (JJA), and fall (SON). Since a monthly time series was not obtained for the irrigation demand, the model uses seasonal irrigation demand as an input.

### 2.2.2. Model Structure and Parameterization

We developed nonlinear regression models for prediction of site-specific groundwater level change using a Multilayer Perceptron (MLP) network architecture. MLP is considered a universal approximator [*Hornik et al*., 1989] and has been successfully used in hydrologic modeling [e.g., *ASCE*, 2000; *Maier and Dandy*, 2000]. An MLP network is a type of feedforward ANN where information is fed forward from one layer to the next. The model consists of an input layer, one or more hidden layers, and an output layer, with each layer consisting of one or more neurons (Figure 5). Every neuron in each layer has connections to every neuron in the adjacent layers, with an individual weight assigned to each interlayer link. The input signals propagate through the network in a forward direction, layer by layer. Each neuron input value is multiplied by the connection weight and then the sum of the products is passed through a nonlinear transfer function to produce a result [*Haykin*, 2008].

The ANNs developed in this study were single hidden layer MLPs (Figure 5). The input layer consisted of six neurons for HPA wells and seven neurons for MRVA wells; the MRVA models had the same inputs as the HPA models, plus an input representing the NAO index. The output was one neuron that represented seasonal groundwater level change. We constructed the MLP network by connecting layers through a series of two transformation types. First, connections between the input variables and hidden layer ($w_{ji}$) were

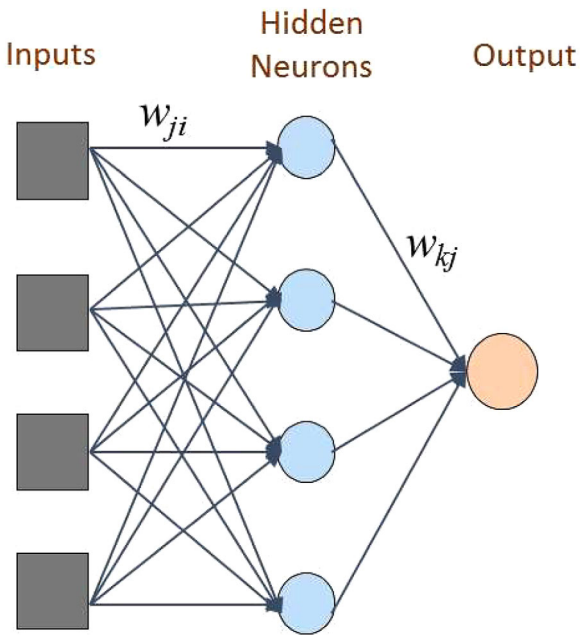


**Figure 5.** Structure of a feedforward ANN model used in this study showing inputs, one layer of hidden neurons, and the output.

established. A logistic sigmoid transfer (activation) function was used for the hidden layer, which ranged between [0, 1] and is the most commonly used activation function [*Haykin*, 2008]. Second, the connections from the hidden layer to the output layer ($w_{kj}$) were established using a linear transfer function, as suggested by *Maier and Dandy* [2000]. The activation function acts as a compressing function and determines the response of a neuron to the total input signal it receives. The sigmoid function is a bounded, monotonic, nondecreasing function that provides a graded, nonlinear response [*Shamseldin*, 1997], whereas a linear transfer function calculates a neuron's output by simply returning the value passed to it.

The prediction from the ANN model ($\hat{y}_k$) can be given as [*Haykin*, 2008],

$$\hat{y}_k = f\left(\sum_{j=1}^{m} w_{kj} h_j + b_k\right), \tag{5}$$

where $w_{kj}$ is the connection weight between output neuron $k$ and hidden neuron $j$, $h_j$ is the output from neuron $j$ in the hidden layer, $m$ is the number of neurons in the hidden layer, $f$ is the activation function of the output layer, and $b_k$ is the bias of the output neuron $k$. "Bias" in an ANN model is equivalent to the intercept in a regression model and is generally assigned a value of 1 [*Haykin*, 2008]. The output from each neuron in the hidden layer is expressed as,

$$h_j = \varphi\left(\sum_{i=1}^{n} w_{ji} x_i + b_j\right), \tag{6}$$

where $n$ is the number of input neurons, $\varphi$ is the activation function of the hidden layer, $w_{ji}$ is the connection between input $i$ and hidden neuron $j$, $x$ is the input layer neuron, and $b_j$ is the bias of the hidden neuron $j$.

The number of hidden neurons is critical to ANN performance. If too many hidden neurons are used, the network may overfit; too few hidden neurons may result in underfitting. Overfitting occurs when the neural network does not learn the underlying statistical properties of the data, but rather memorizes the patterns and any noise they may contain. This results in neural networks which perform well in training but poorly in out-of-sample data. A rule of thumb suggests that the number of hidden neurons should be between half and twice the number of predictors [*Minns and Hall*, 1996; *Berry and Linoff*, 2004], or in this case between 3 and 12 (14) neurons for the HPA (MRVA). In this study, the optimal number of neurons in the hidden layer was determined by testing all reasonable possible values for each individual well until the best match was achieved based on the correlation coefficient (R) and mean squared error (MSE) between observed and predicted groundwater level changes. The maximum number of neurons in the hidden layer was set at 15, although no further performance improvement was gained beyond 12 neurons. The Levenberg-Marquardt algorithm [*Levenberg*, 1944; *Marquardt*, 1963], considered to be one of the most efficient learning algorithms [*Maier and Dandy*, 2000], was selected for backpropagation and updating the connection weights.

This method was repeated to develop site-specific ANN models for each well in the HPA and MRVA. The neural network model for each well underwent training, cross-validation, and testing to predict seasonal groundwater level change between 1980 and 2012. The input and response variables were divided into three temporal subsets using contiguous blocks of the original data set: (1) training from 1980 to 1996, (2) cross-validation from 1997 to 2002, and (3) testing from 2003 to 2012. The data were split into the training and cross-validation periods according to the early-stopping approach [*Prechelt*, 2012] where the cross-

validation step was introduced to validate that the network was generalizing and to stop training before overfitting. The intent of cross-validation is to choose the condition in which the error of the validation data sets reaches the minimum. The testing period was used to assess the network's predictive ability on data not used for either training or cross-validation. The ANN model was retrained up to 500 times using the same inputs and output and was stopped once it attained $R > 0.8$, or reached 500 iterations. The final model parameters and model prediction performance (R, MSE) were recorded once the termination criteria were achieved.

The propagation of uncertainty from inputs to model prediction was estimated in a Monte Carlo framework. In this method, 1000 random samples were generated from the observed data. The simulation was repeated 1000 times to obtain a distribution of model output values, and the mean and standard error were calculated. Uncertainty bounds of model groundwater level change simulations were determined as standard errors of the distribution of predictions at 95% confidence intervals.

We compare results from the ANN and HANN models to multivariate linear regression (MLR) and nonlinear regression (MNLR) models, to evaluate the predictive ability of our new models relative to alternative empirical models. The MLR model is used to simulate the relationship between selected predictors (original inputs, same as ANN) and groundwater level change by fitting a linear regression equation to the observations [*Pedhazur*, 1997; *Sahoo and Jha*, 2015], whereas the MNLR models fit a nonlinear regression to the observed data [*Seber and Wild*, 2003]. The MLR (equation (7)) and MNLR (equation (8)) equations are given as,

$$Y_{MLR} = c + \beta_1 X_1 + \cdots + \beta_n X_n, \tag{7}$$

$$Y_{MNLR} = c + \beta_1 X_1^{\beta_2} + \beta_3 X_2^{\beta_4} + \cdots + \beta_m X_n^{\beta_{m+1}}, \tag{8}$$

where $Y$ is the response (dependent) variable, $c$ is the intercept, $\beta$ is the slope or coefficient, $X$ is the predictor (independent) variable, $m$ is the number of coefficients in the nonlinear models (Supporting Information Figure S1), and $n$ is the number of input variables (six for HPA models, seven for MRVA models). For comparison with our HANN model, we also define HMLR and HMNLR models which use transformed inputs ($X$) and thus combine the data preprocessing, variable selection, and lag analysis mechanisms described above with the MLR and MNLR regression models, respectively.

#### 2.2.3. Relative Importance of Model Inputs

We employed the connection weight approach [*Olden et al.*, 2004] to quantify the importance of the various input variables for model prediction. *Olden et al.* [2004] compared nine different methods to determine the relative contribution of inputs and found that this approach was the most successful of the nine methods studied for identifying the importance of all the variables in the neural network. The connection weight approach calculates the product of the hidden-input ($w_{ji}$) and hidden-output ($w_{kj}$) connection weights between each input neuron and output neuron and sums the products across all hidden neurons. The relative importance ($RI$) of the $i$th input variable is defined as follows:

$$RI_i = \frac{|z_i|}{\sum_i |z_i|}, \quad \text{where } z_i = \sum_{j=1}^{n} w_{ji} * w_{kj} \tag{9}$$

Based on the relative importance value, the input variables were ranked. The input variable with the highest absolute relative importance value has the most influence on the output variable, while the lowest value corresponds to the input variable with the least influence. Accordingly, the most to least influential parameters were identified for prediction of groundwater level change at each site.

## 3. Results and Discussion

### 3.1. Implementation and Evaluation of Models

#### 3.1.1. Distribution of Hidden Neurons and Model Performance

The models each have one hidden layer, but the number of hidden neurons varies across each study area. The optimal number of hidden neurons, the associated weights, and model performance indicators were all derived for each well individually. Though ANNs are "black box" models, some useful information about the model parameters and physical system can be interpreted from the connection weights and the number of hidden neurons. Hidden neurons may be considered "learned feature detectors" [*Touretzky and Pomerleau*,

1989], with the number of hidden neurons increasing with the number of input parameters and training samples [*Maier and Dandy*, 2000]. According to *Lipták* [2005], more complex model processes require a greater number of neurons in the hidden layer. The optimal number of hidden neurons ranges from 3 to 11 in the HPA, and from 3 to 12 in the MRVA (Figures 6a and 6c). In the HPA, 60% of the wells have three to five hidden neurons and 20% have five to seven hidden neurons (Figure 6a). In the MRVA, 38% of the wells have three to five hidden neurons, and 28% have seven to nine hidden neurons (Figure 6c). The greater percentage of wells in the MRVA with more hidden neurons may be due the additional input parameter, or a more complex underlying statistical relationship between the inputs and groundwater level change.

We evaluate the performance of the HANN model using two statistical indices: correlation coefficient (R) and mean squared error (MSE). We see good agreement between HANN model output values and observations in most wells during the study period. R values decrease by approximately 0.1 between the training and testing periods; all correlation coefficients reported herein use only the testing data set (2003–2012) (Figures 6b and 6d). R values for the HPA vary from 0.65 to 0.87 for the testing period (Figure 6b): 49% of wells have $R > 0.85$, 39% have R between 0.8 and 0.85, and 8% have R between 0.75 and 0.8 (Figure 6b). The remaining 4% of wells have R values between 0.65 and 0.75. In the MRVA, R values for the testing period vary from 0.58 to 0.89; 48% of wells have $R > 0.85$, 35% have R between 0.8 and 0.85, 14% have R between 0.7 and 0.8, and the remaining 3% have R values between 0.58 and 0.7 (Figure 6d). In both aquifer systems, nearly half of the well sites have $R > 0.85$, demonstrating strong predictive ability. There does not appear to be any spatial coherence or clustering for either the number of hidden neurons or R value (Figure 6).

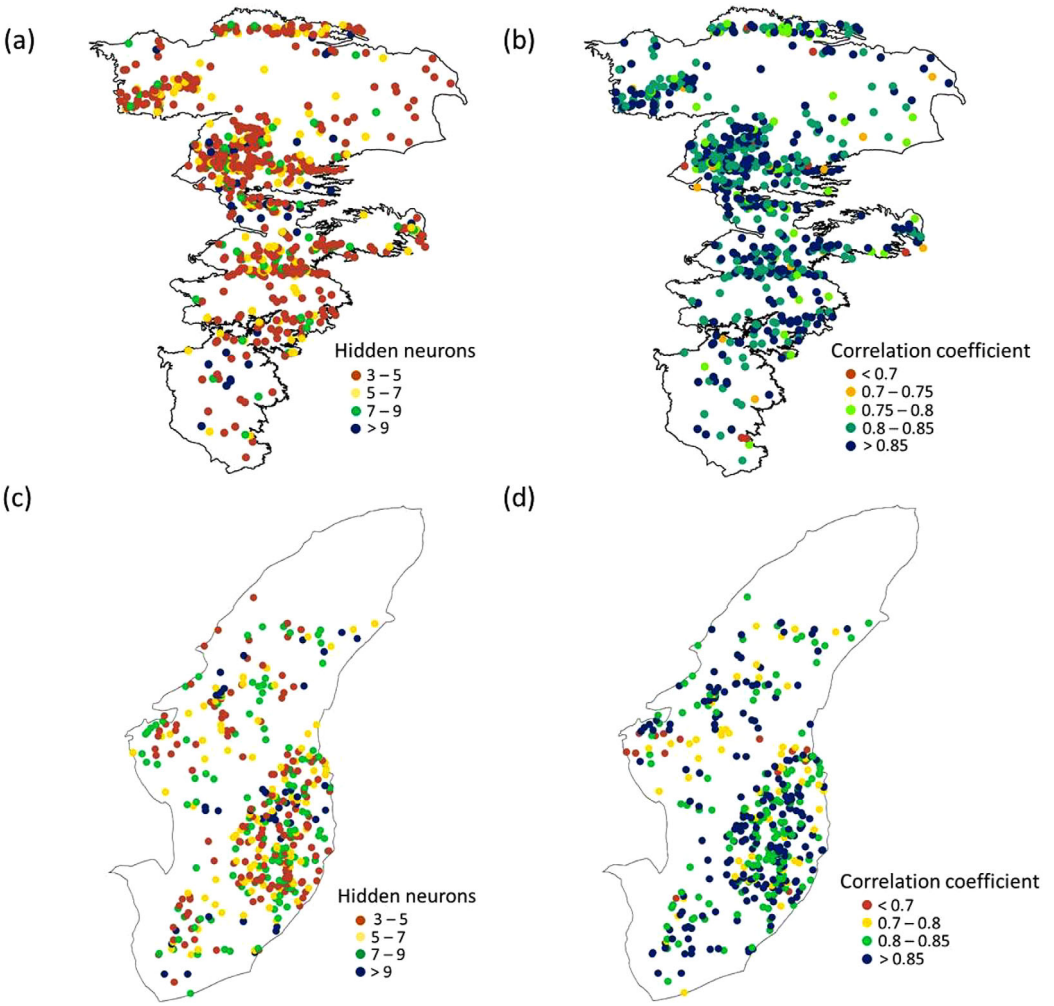


**Figure 6.** (a, c) The number of hidden neurons and (b, d) the correlation coefficient between observed and predicted seasonal groundwater level change using the HANN model for 687 wells in the HPA and 437 wells in the MRVA, respectively.

### 3.1.2. Model Residuals and Uncertainty

Figure 7 shows the distribution of the groundwater level change residuals obtained from the HANN model for all wells over the HPA and MRVA for each model period: training, cross-validation, and testing. Residuals are calculated for each model time step (DJF, MAM, JJA, and SON) by subtracting the observed groundwater level change from the simulated groundwater level change. A positive change in groundwater level equals a rise in groundwater level; therefore, a negative residual represents an underestimate of groundwater rise or overestimate of decline, conversely a positive residual represents an overestimate of groundwater rise or an underestimate of decline. The median residual value is nearly zero (unbiased) for all cases (Figure 7). The distributions of residuals are roughly normal, except in the case of the testing period for the MRVA which has mostly negative residuals and a positive skew, indicating model overestimation of groundwater decline (or underestimation of groundwater rise) at most wells. This difference may be due to errors in the irrigation area or crop area data, or complexities that the model does not account for, such as an increase in the amount of surface water capture as groundwater development increased rapidly in the MRVA during the study period [*Maupin et al.*, 2014]. The distribution of residuals for both aquifers is largest for the cross-validation period, because this period is used to test the predictive models developed in training to determine the optimal number of hidden neurons and a stopping point for the MLP method. As such, the overall errors may be higher during cross-validation because it is evaluating which model and final settings provide the best fit, without overfitting. The errors on the testing set are smaller because they benefit from the final model determined based on the training and the cross-validation sets.

The HANN model uncertainty was quantified by calculating the standard error of the 1000 Monte Carlo predictions. A majority of the sites have an average seasonal groundwater level uncertainty between 0 and 0.4 m in both aquifers (Figure 8). A small subset of the wells, 23 in HPA (Figure 8a) and 22 in MRVA (Figure 8b), have higher uncertainty in prediction (0.6–1 m), which may be due to uncertainty in input data, interpolation methods, unaccounted for external factors, or data deficiency. The uncertainty varies across both aquifers, with no obvious spatial pattern clustering regions of high or low uncertainty. If our model were missing one or more significant regional controls, we would expect to see the uncertainty increase in those regions. Because the spatial distribution of uncertainty values is relatively uniform, we conclude that we are capturing the key input parameters. However, deficiencies in the model may still be incrementally improved by adding input parameters for all sites.

### 3.1.3. Model Comparison

We compared results from the ANN and HANN models to results of the multivariate linear regression (MLR) and multivariate nonlinear regression (MNLR) models using both original and transformed (SSA with mutual information, genetic algorithm, and lags) input variables (Table 1). The original model inputs consist of untransformed time series of precipitation, temperature, climate indices, irrigation, and streamflow data.

We find that the hybrid models trained or calibrated with transformed input data have higher prediction accuracy than the models with the original inputs (Table 1). Moreover, the overall prediction accuracy of the

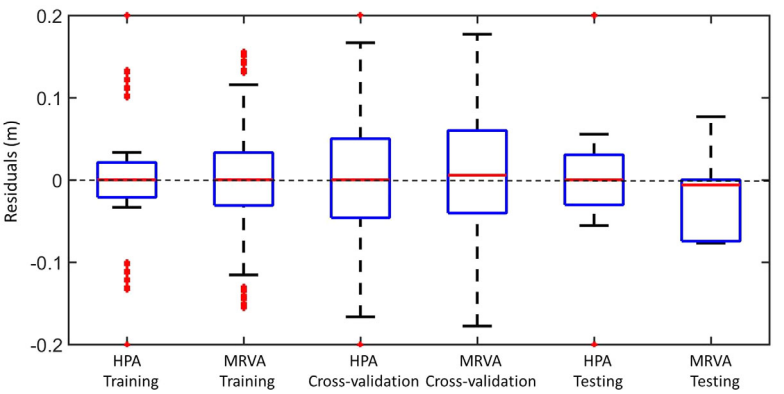


**Figure 7.** Box plots of residuals (simulated minus observed) for the 687 HPA wells and 437 MRVA wells during the training (1980–1996), cross-validation (1997–2002), and testing periods (2003–2012). The boxes show the median and the 25th and 75th percentiles. The whiskers in black correspond to approximately $\pm 2.7\sigma$, and the outliers are plotted individually (red dots). The minimum and maximum outlier residuals span −9 to 8, though a truncated plot is shown to illustrate the detail of a majority of the data.

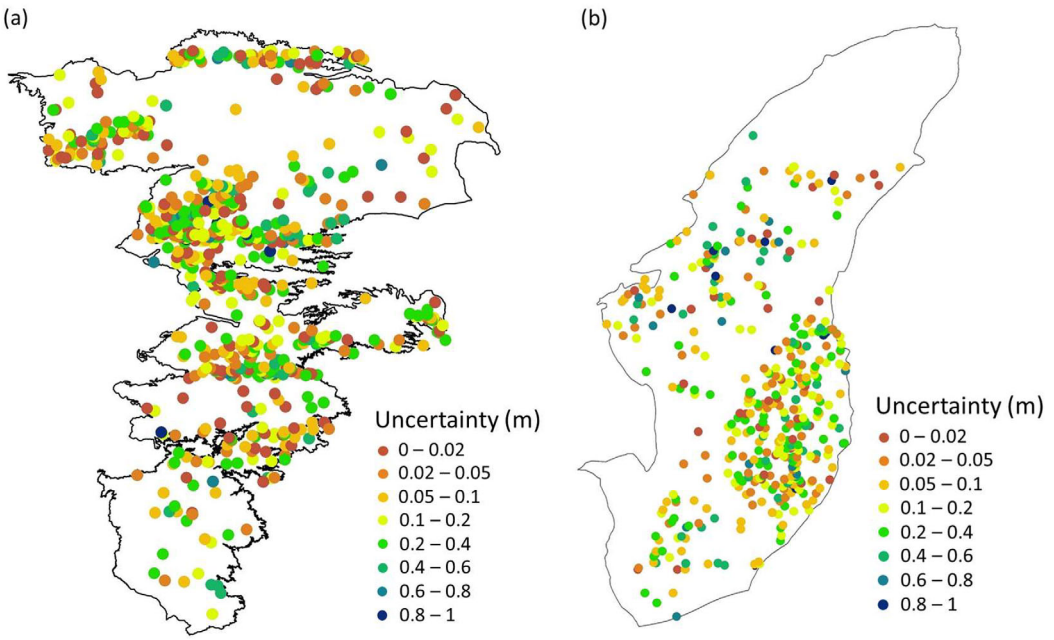


**Figure 8.** Average seasonal groundwater level uncertainty in model prediction for (a) 687 wells in the HPA and (b) 437 wells in MRVA calculated using HANN model and Monte Carlo method.

HANN models is superior to those of the linear (HMLR) and nonlinear (HMNLR) regression models. For example, the HMNLR and HANN models of the HPA, the R value increased from 0.67 to 0.81, and the MSE decreased from 1.81 to 0.28 $m^2$, respectively, suggesting an overall better fit to observations and lower prediction error. The performance statistics for the three model types using either original or transformed input data for the MRVA also highlight the higher prediction accuracy for the HANN models. Although ANN models are type of nonlinear regression models [*Maier and Dandy*, 2000], they have the added leverage of hidden neurons and connection weights to improve model performance, whereas the linear models have only one set of regression coefficients to predict the groundwater level change. In addition, the retraining approach improved model accuracy significantly when compared to a single neural network approach. It is not surprising that the ANN model outperforms the regression models, but it is an important comparison nonetheless; if the regression models were sufficient there would be no justification for the ANN model.

### 3.2. Relative Contribution of Model Inputs

We calculated the relative importance of each input parameter from the connection weights of the trained ANN models (equation (3)). We show in Figures 9a and 9b the relative importance of each input parameter for two individual wells, one in the HPA and one in the MRVA. The most important input at the HPA

**Table 1.** Comparison of Performance Statistics for Three Model Types Using Either Original or Transformed (Prefixed With an H for "Hybrid") Input Data for the HPA and the MRVA[a]

| | | Test Results | | | |
|---|---|---|---|---|---|
| | | HPA | | MRVA | |
| Model | Model Inputs | R | MSE [$m^2$] | R | MSE [$m^2$] |
| 1. Multivariate linear regression model (MLR) | a. Original (MLR) | 0.41 | 5.8 | 0.47 | 4.71 |
| | b. Transformed (HMLR) | 0.60 | 2.64 | 0.57 | 3.12 |
| 2. Multivariate nonlinear regression model (MNLR) | c. Original (MNLR) | 0.62 | 2.20 | 0.60 | 2.37 |
| | d. Transformed (HMNLR) | 0.67 | 1.81 | 0.64 | 2.01 |
| 3. Artificial neural network model (ANN) | e. Original (ANN) | 0.74 | 1.14 | 0.69 | 1.62 |
| | f. Transformed (HANN) | 0.81 | 0.28 | 0.76 | 0.43 |

[a]Statistics represent mean of 687 wells and 437 wells in the HPA and the MRVA, respectively.

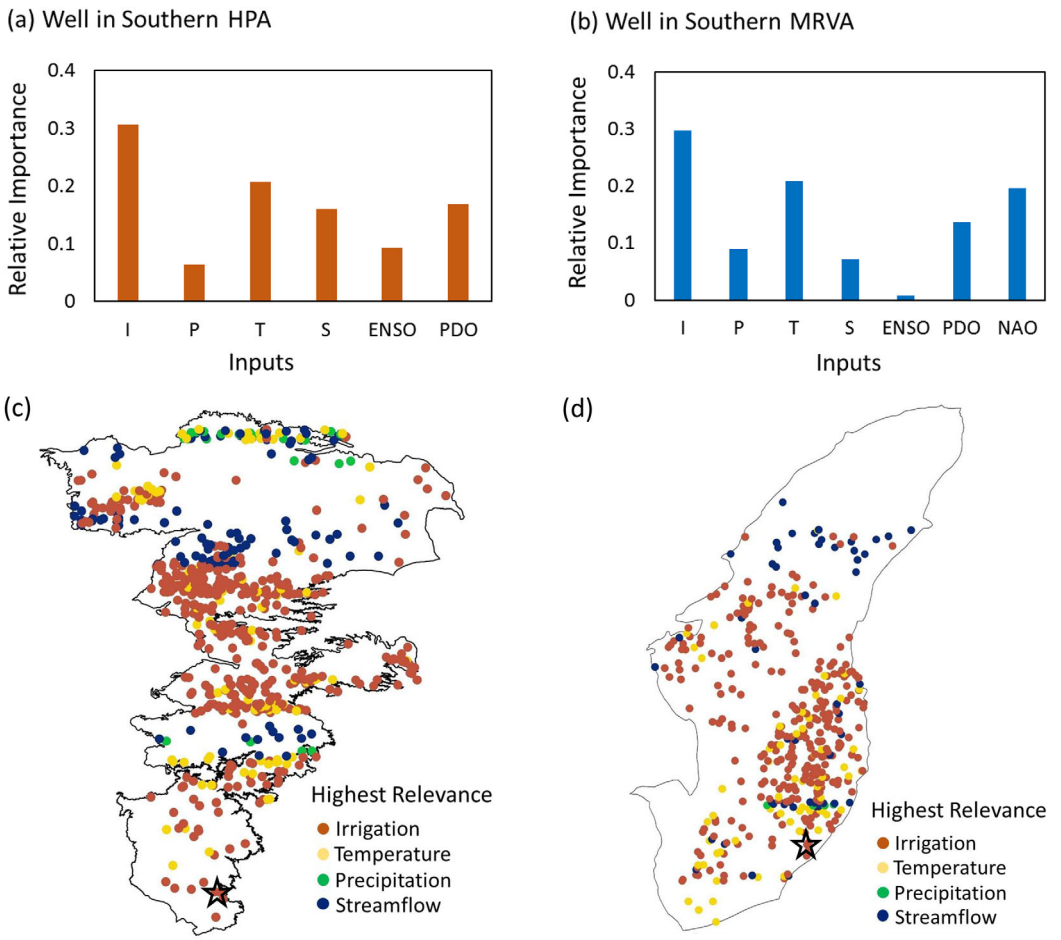


**Figure 9.** Relative importance of model inputs to groundwater level change prediction by connection weight approach for the example wells in the (a) HPA and (b) MRVA (I: irrigation demand, P: precipitation, T: temperature, S: stream discharge), and the input parameter with the highest relevance at each well location using the HANN model for (c) 687 wells in the HPA and (d) 437 wells in the MRVA. The black stars in the HPA and MRVA indicate the locations of two example sites. The ranking for all of the input parameters and percentage of wells that fall under each category is included in Table 2.

example site is irrigation demand, followed by temperature, PDO, streamflow, ENSO, and lastly precipitation (Figure 9a). At the MRVA example site, irrigation demand is also the most important input, followed by temperature and two of the climate variability indices, NAO, and PDO (Figure 9b).

**Table 2.** Ranking of Input Parameters As Obtained From the Connection Weight Approach for (a) HPA and MRVA

| | | Percentage of Wells Within Each Rank for Six (or Seven) Inputs | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Rank | Irrigation | Temperature | Precipitation | Streamflow | ENSO | PDO | NAO |
| HPA | 1 | 66 | 21 | 3 | 10 | 0 | 0 | N/A |
| | 2 | 27 | 39 | 8 | 12 | 1 | 13 | N/A |
| | 3 | 3 | 32 | 19 | 17 | 11 | 18 | N/A |
| | 4 | 2 | 6 | 20 | 18 | 19 | 35 | N/A |
| | 5 | 1 | 1 | 16 | 20 | 38 | 24 | N/A |
| | 6 | 1 | 1 | 34 | 23 | 31 | 10 | N/A |
| MRVA | 1 | 72 | 15 | 1 | 12 | 0 | 0 | 0 |
| | 2 | 22 | 41 | 7 | 8 | 2 | 8 | 12 |
| | 3 | 2 | 31 | 6 | 9 | 7 | 21 | 24 |
| | 4 | 1 | 9 | 19 | 16 | 13 | 25 | 17 |
| | 5 | 1 | 2 | 22 | 21 | 15 | 20 | 19 |
| | 6 | 1 | 1 | 18 | 20 | 28 | 19 | 13 |
| | 7 | 1 | 1 | 27 | 14 | 35 | 7 | 15 |

Irrigation demand has the highest relative importance at the two example wells (Figures 9a and 9b), as well as at a majority of the wells in each aquifer (Figures 9c and 9d and Table 2). These results demonstrate the connections between climate, crop water demand, and groundwater level change previously shown at the local scale by *Gurdak et al.* [2007] and suggested for the continental U.S. by *Russo and Lall* [2017]. While the connection between pumping and groundwater level change is rooted in simple physics, it was not initially obvious whether the climate parameters (e.g., precipitation and temperature) would have a stronger influence on seasonal groundwater level than irrigation demand, given that irrigation demand was calculated based on those same climate inputs. The dependency of simulated irrigation demand on temperature and precipitation introduces redundancy in the HANN input parameters; however, additional constraints within the DSSAT model result in irrigation demand estimates with enough independent information to be ranked more important than either precipitation or temperature at a majority of the wells.

In the HPA, irrigation demand has the greatest influence on groundwater level in 66% of the wells, followed by temperature in 21%, streamflow in 10%, and precipitation in 3% (Figure 9c and Table 2); the wells for which streamflow has the highest impact are located near major rivers including the North and South Platte River in the northern HPA and the Canadian River in the central HPA. For several regions in the HPA, temperature or precipitation is the most important input corresponded to areas with no or low rates of irrigation. In these cases, temperature and precipitation are likely controlling groundwater level by impacting recharge, rather than pumping. For the MRVA, irrigation demand has the greatest influence at 72% of the wells, followed by temperature at 15%, streamflow at 12%, and precipitation at 1% (Figure 9d and Table 2); the wells for which streamflow has the highest impact are linearly clustered in the north and southeast regions of the aquifer. The sites where temperature has the highest importance were relatively well distributed across the aquifer. Though a majority of irrigation in the two regions comes from groundwater, there are small areas where surface water is relied upon. The use of surface water, when available, for irrigation may influence groundwater by either reducing extraction rates or increasing recharge via return flow. Overall, in these predominantly groundwater-supported agricultural regions we find that irrigation demand has the greatest influence on predicting groundwater level as compared to temperature, precipitation, and streamflow.

### 3.3. Modeled Groundwater Level Change

We summed simulations from the HANN model during the model testing period (2003–2012) to estimate the cumulative groundwater level change. We used Barnes analysis [*Barnes*, 1964] to interpolate the modeled cumulative groundwater level change across 687 wells in the HPA and 437 wells in the MRVA (Figures 10a and 10c). The residual was calculated as the modeled minus the observed groundwater level change and was likewise interpolated and shown for each aquifer (Figures 10b and 10d).

Groundwater level change model results indicate declines in the HPA across much of the aquifer except in the upper portion of Nebraska and South Dakota. This result is in agreement with the observed groundwater change calculated from several studies [*Scanlon et al.*, 2012; *McGuire*, 2014; *Sahoo et al.*, 2016]. For the MRVA, we see groundwater level declines across much of the aquifer except for small regions along the edges of the MRVA (Figure 10c).

We show in Figures 10b and 10d the cumulative residual (modeled minus observed) groundwater level change error over the 10 year testing period for the HPA and MRVA. Regions with positive residual values indicate model underestimation of groundwater decline or overestimation of rise, and conversely regions with negative residual values indicate model overestimation of groundwater decline or underestimation of rise. In the HPA, the model performs reasonably well with a majority of residuals over 10 years of testing period fall within $\pm 2$ m. Isolated green patches (positive residuals) in the central and northern part of HPA indicate underestimation of declines or overestimation of rise during testing period. In the MRVA, a majority of the residuals are near zero or negative (–4 to $+2$ m), which indicates overestimation of declines or underestimation of rise (Figures 7 and 10d). The negative residuals (–4 to $-2$ m) in the northernmost part of MRVA are partly due to insufficient data in the area (Figure 1). Groundwater use is rapidly increasing in the MRVA, and it is possible that our input data do not reflect actual irrigated area, or changes in surface or subsurface water inputs to the system. The MRVA model may be refined by adding additional recent years and extending the training period.

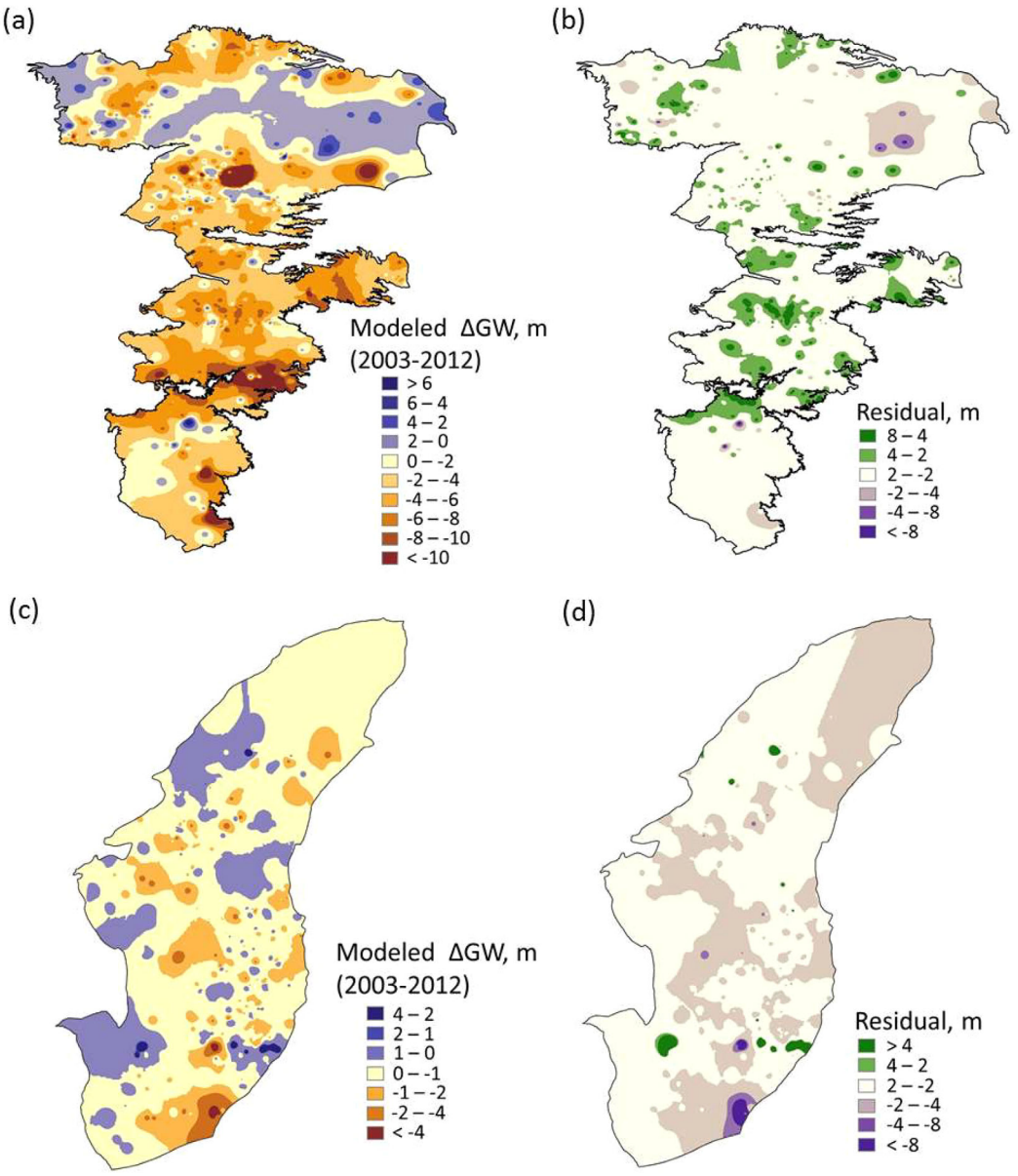


**Figure 10.** Modeled cumulative groundwater level change (2003–2012) for the (a) HPA and (c) MRVA, and residual (modeled minus observed) groundwater level change for (b) HPA and (d) MRVA.

Numerical models simulate the system water balance using physical properties, whereas ANN models model the system dynamics based on the relationships between inputs and observed outputs mapped in the training process. The HANN model results can be compared to regional groundwater level simulations using the numerical model MODFLOW [*Harbaugh*, 2005]. Results from a groundwater flow model of the central HPA developed by *Luckey and Becker* [1999] showed a MSE of 29.8 $m^2$ for 162 observation wells within Oklahoma for the 1946–1998 simulation period. *Peterson et al*. [2008] simulated groundwater levels for the Elkhorn and Loup River Basins in Nebraska with a MSE of 1.96 $m^2$ for 584 wells from 1995 to 2005, a MSE of 3.79 $m^2$ for 42 wells from 1945 to 2005, and a MSE of 1.7 $m^2$ for 2033 wells from 1940 to 2005. In contrast, our model achieves an average MSE of 0.25 and 0.21 $m^2$ using only testing period results for the same regions of Nebraska and Oklahoma, respectively. *Clark et al*. [2011] developed a numerical model for the MRVA as a part of Mississippi Embayment Regional Aquifer Study. The MSE for the simulation of groundwater level in the alluvial aquifer was 26.6 $m^2$. Using the HANN model, the MSE for the testing period in MRVA varies from 0.04 to 1.21 $m^2$. Groundwater storage changes from previous studies are compared to

**Table 3.** Comparison of Groundwater Storage Changes With Past Studies for Both HPA and MRVA

| Authors | Timeframe | Changes in Storage ($km^3$) | Method | This Study |
|---|---|---|---|---|
| *Konikow* [2013] | 1900–2000 | −259.1 | Groundwater level change and storativity | N/A |
| | 1900–2008 | −340.9 | | N/A |
| | 2001–2008 | −81.8 | | −93.7 |
| *McGuire* [2013] | 1950–2011 | −303.4 | Groundwater level change and storativity | N/A |
| *McGuire* [2004] | 1980–2002 | −82.4 | | −105.3 |
| *McGuire* [2014] | 2003–2013 | −93.6 | | −127.5 (2003–2012) |
| *Sahoo et al.* [2016] | 2003–2013 | −142 | Groundwater level change and storativity | −127.5 (2003–2012) |
| *Breña-Naranjo et al.* [2014] | 2003–2013 | −125 | GRACE-derived groundwater storage with a correction for irrigation | −127.5 (2003–2012) |
| *Steward and Allen* [2015] | 1980–1999 | −135 | Logistic regression from Hubbert's curve | −95.1 |
| *Haacker et al.* [2015] | 1950–2011 | −385 | Saturated thickness and specific yield | N/A |
| *Clark et al.* [2011] | 1870–2007 | −183.8 | Numerical modeling, MERAS model, | N/A |
| | 1980–2007 | −160.4 | USGS MODFLOW-2005 | −144.2 |

this study using the same study time periods (Table 3). We show similar estimates of storage changes, given expected variability due to differences in interpolation methods and uncertainty arising from the input parameters and number of wells used for each study. The relatively strong performance of the HANN model supports previous direct comparisons of empirical and numerical models [e.g., *Coppola et al.*, 2003].

We attribute some of the HANN model's strength to its structure. Because the model is developed and optimized for each individual well, we benefit from training values at each well. It is therefore not surprising that the MSE is lower in the HANN model compared to the physically based models, which were calibrated using a subset of wells, and were constrained by the ability to represent heterogeneity in the system. On the other hand, while the HANN model has a lower groundwater level prediction error than the physical models, it does not offer many of the benefits of a physical model, such as flux estimates and residence time calculations. Additionally, accounting for changes to the hydrologic system not included in our input parameters, for example imported water or managed aquifer recharge, would require a new empirical model, while the physical model can be amended to include these additional sources and processes. Therefore, there are environments and applications for which each model type excels. Future improvements to the HANN model may include using spatial covariance statistics or multivariate geostatistical accounting of model parameters. These additions would improve prediction of groundwater levels by accounting for uncertainties arising from the spatial heterogeneity and spatial coherence of controlling factors.

## 4. Conclusions

We propose an automated hybrid artificial neural network (HANN) model with a new input data processing method for simulating groundwater level change. We apply, evaluate, and discuss the model capability for predicting seasonal groundwater level change at 1124 wells across two major agricultural regions of the US: the HPA and MRVA. In both regions, groundwater pumping has produced groundwater level declines across large areas and is a growing concern for water resource managers. We elected to test the HANN model on these two well-studied aquifers to allow for model output comparison. Ultimately, the HANN model may be most useful in regions lacking known hydrogeologic subsurface parameters.

The HANN model uses precipitation, temperature, streamflow and climate indices as input parameters. In addition, irrigation demand simulated with the DSSAT model is used in lieu of unavailable groundwater pumping records. We evaluated the performance of both our artificial neural network (ANN) and conventional regression models in two configurations, one using raw input data and the other using input data transformed by SSA, mutual information, genetic algorithm, and maximum-correlated lag. The models trained with transformed input data, distinguished in their names by the prefix "Hybrid," produce higher prediction accuracy than the models with original (untransformed) inputs. The overall performance of the HANN model is superior to that of hybrid linear regression (HMLR) and hybrid nonlinear (HMNLR) regression models. The nonlinear and complex relationships between climatic variables, irrigation demand, streamflow, and groundwater levels justify the use of ANNs.

We find irrigation demand to be the most important parameter when modeling groundwater levels in HPA and MRVA, followed by temperature, rainfall, and streamflow. This result indicates that when projecting groundwater availability, it will be important to quantify and account for agricultural intensification and climate changes that increase irrigation demands. Among the climate indices, PDO plays an important role in long term fluctuations of groundwater level in the HPA as had been shown previously [*Gurdak et al*., 2007] while NAO has higher influence on groundwater fluctuations in the MRVA.

We attribute the strong performance of the HANN method to its integration of robust data processing and input variable selection techniques with the use of ANN models for capturing the impacts of the potential predictor variables on groundwater level change at multiple-well locations. Monte Carlo uncertainty analysis suggests that the prediction is within acceptable range (up to 0.4 m error over 33 years at most sites), providing confidence in model prediction results. Applications of the HPA and MRVA models may include prediction of future groundwater levels based on projections of climate, streamflow, and irrigation demand. These results would be useful for identifying locations that are vulnerable to groundwater depletion under the influences of a changing climate and increased agricultural water demand. Relative to physical modeling methods, we expect the HANN method to be particularly useful in agricultural regions where groundwater level measurements are abundant but subsurface parameters are broadly unknown, such as Punjab, India [*Russo et al*., 2015]. Overall, the HANN model is a valuable tool for predicting groundwater level changes, especially for regions where it is difficult to develop a physical hydrogeologic model.

**Acknowledgments**
This work was supported by the University of Chicago 1896 Pilot Project, the Center for Robust Decision Making on Climate and Energy Policy (RDCEP), NSF award 0951576, and DOE contract DE-AC02-06CH11357. We would like to thank two anonymous reviewers for comments that greatly improved the manuscript. We also thank Debashish Sahu for his helpful suggestions on the model development. The groundwater level, climate, and streamflow data used in this study are publically available from the sources listed in the article. All data sets are available from the corresponding author on request.

## References

Adamowski, J., and H. F. Chan (2011), A wavelet neural network conjunction model for groundwater level forecasting, *J. Hydrol.*, *407*(1–4), 28–40, doi:10.1016/j.jhydrol.2011.06.013.

Allen, R. G., L. S. Pereira, D. Raes, M. Smith, and W. Ab (1998), Crop evapotranspiration—Guidelines for computing crop water requirements, *Irrig. Drain.*, *300*(56), 1–15, doi:10.1016/j.eja.2010.12.001.

Alley, W. M., R. W. Healy, J. W. LaBaugh, and T. E. Reilly (2002), Flow and storage in groundwater systems, *Science*, *296*(5575), 1985–1990, doi:10.1126/science.1067123.

ASCE (2000), Artificial neural networks in hydrology. I: Preliminary concepts, *J. Hydrol. Eng.*, *5*(2), 115–123, doi:10.1061/(ASCE)1084-0699(2000)5:2(115).

Barnes, S. L. (1964), A technique for maximizing details in numerical weather map analysis, *J. Appl. Meteorol.*, *3*(4), 396–409, doi:10.1175/1520-0450(1964)003 < 0396:ATFMDI>2.0.CO;2.

Behzad, M., K. Asghari, and E. A. Coppola (2010), Comparative study of SVMs and ANNs in aquifer water level prediction, *J. Comput. Civ. Eng.*, *24*(5), 408–413, doi:10.1061/(ASCE)CP.1943-5487.0000043.

Berry, M. J., and G. S. Linoff (2004), Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, John Wiley, Indianapolis, Indiana.

Bidwell, V. J. (2005), Realistic forecasting of groundwater level, based on the eigenstructure of aquifer dynamics, *Math. Comput. Simul.*, *69*(1–2), 12–20, doi:10.1016/j.matcom.2005.02.023.

Breña-Naranjo, J. A., A. D. Kendall, and D. W. Hyndman (2014), Improved methods for satellite-based groundwater storage estimates: A decade of monitoring the high plains aquifer from space and ground observations, *Geophys. Res. Lett.*, *41*, 6167–6173, doi:10.1002/2014GL061213.

Chang, F. J., L. C. Chang, C. W. Huang, and I. F. Kao (2016), Prediction of monthly regional groundwater levels through hybrid soft-computing techniques, *J. Hydrol.*, *541*, 965–976, doi:10.1016/j.jhydrol.2016.08.006.

Chang, L.-C., H.-J. Chu, and C.-T. Hsiao (2007), Optimal planning of a dynamic pump-treat-inject groundwater remediation system, *J. Hydrol.*, *342*(3–4), 295–304, doi:10.1016/j.jhydrol.2007.05.030.

Chu, H. J., and L. C. Chang (2009), Application of optimal control and fuzzy theory for dynamic groundwater remediation design, *Water Resour. Manage.*, *23*(4), 647–660, doi:10.1007/s11269-008-9293-1.

Clark, B. R., R. M. Hart, and J. J. Gurdak (2011), Groundwater availability of the Mississippi Embayment, *U.S. Geol. Surv. Prof. Pap.*, *1785*, 62 p.

Coppola, E., F. Szidarovszky, M. Poulton, and E. Charles (2003), Artificial neural network approach for predicting transient water levels in a multilayered groundwater system under variable state, pumping, and climate conditions, *J. Hydrol. Eng.*, *8*(6), 348–360, doi:10.1061/(ASCE)1084-0699(2003)8:6(348).

Coulibaly, P., F. Anctil, R. Aravena, and B. Bobee (2001), Artificial neural network modeling of water table depth fluctuations, *Water Resour. Res.*, *37*(4), 885–896, doi:10.1029/2000WR900368.

Cover, T. M., and J. A. Thomas (2005), *Elements of Information Theory*, John Wiley, Hoboken, N. J.

Dettinger, M. D., M. Ghil, C. M. Strong, W. Weibel, and P. Yiou (1995), Software expedites singular-spectrum analysis of noisy time series, *Eos Trans. AGU*, *76*(2), 12–21, doi:10.1029/EO076i002p00012.

Elliott, J., D. Kelly, J. Chryssanthacopoulos, M. Glotter, K. Jhunjhnuwala, N. Best, M. Wilde, and I. Foster (2014), The parallel system for integrating impact models and sectors (pSIMS), *Environ. Modell. Software*, *62*, 509–516, doi:10.1016/j.envsoft.2014.04.008.

Fernando, T. M. K. G., H. R. Maier, and G. C. Dandy (2009), Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach, *J. Hydrol.*, *367*(3–4), 165–176, doi:10.1016/j.jhydrol.2008.10.019.

Galelli, S., G. B. Humphrey, H. R. Maier, A. Castelletti, G. C. Dandy, and M. S. Gibbs (2014), An evaluation framework for input variable selection algorithms for environmental data-driven models, *Environ. Modell. Software*, *62*, 33–51, doi:10.1016/j.envsoft.2014.08.015.

Gong, W., H. V. Gupta, D. Yang, K. Sricharan, and A. O. Hero (2013), Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach, *Water Resour. Res.*, *49*, 2253–2273, doi:10.1002/wrcr.20161.

Green, T. R., M. Taniguchi, H. Kooi, J. J. Gurdak, D. M. Allen, K. M. Hiscock, H. Treidel, and A. Aureli (2011), Beneath the surface of global change: Impacts of climate change on groundwater, *J. Hydrol.*, *405*(3–4), 532–560, doi:10.1016/j.jhydrol.2011.05.002.

Güler, C., M. A. Kurt, M. Alpaslan, and C. Akbulut (2012), Assessment of the impact of anthropogenic activities on the groundwater hydrology and chemistry in Tarsus coastal plain (Mersin, SE Turkey) using fuzzy clustering, multivariate statistics and GIS techniques, *J. Hydrol.*, *414–415*, 435–451, doi:10.1016/j.jhydrol.2011.11.021.

Gurdak, J. J., R. T. Hanson, P. B. McMahon, B. W. Bruce, J. E. McCray, G. D. Thyne, and R. C. Reedy (2007), Climate variability controls on unsaturated water and chemical movement, High Plains Aquifer, USA, *Vadose Zone J.*, *6*(3), 533–547, doi:10.2136/vzj2006.0087.

Haacker, E. M. K., A. D. Kendall, and D. W. Hyndman (2015), Water level declines in the High Plains Aquifer: Predevelopment to resource senescence, *Groundwater*, *54*(2), 231–242, doi:10.1111/gwat.12350.

Hanson, R. T., M. W. Newhouse, and M. D. Dettinger (2004), A methodology to ASESS relations between climatic variability and variations in hydrologic time series in the southwestern United States, *J. Hydrol.*, *287*(1–4), 252–269, doi:10.1016/j.jhydrol.2003.10.006.

Hanson, R. T., M. D. Dettinger, and M. W. Newhouse (2006), Relations between climatic variability and hydrologic time series from four alluvial basins across the southwestern United States, *Hydrogeol. J.*, *14*(7), 1122–1146, doi:10.1007/s10040-006-0067-7.

Harbaugh, A. W. (2005), MODFLOW-2005, The U.S. Geological Survey modular ground-water model—The ground-water flow process, *U.S. Geol. Surv. Tech. Methods*, *6-A16*, 252 pp.

Haykin, S. (2008), *Neural Networks and Learning Machines*, Prentice Hall, Upper Saddle River, N. J.

Holman, I. P. (2006), Climate change impacts on groundwater recharge-uncertainty, shortcomings, and the way forward?, *Hydrogeol. J.*, *14*(5), 637–647, doi:10.1007/s10040-005-0467-0.

Hornik, K., M. Stinchcombe, and H. White (1989), Multilayer feedforward networks are universal approximators, *Neural Networks*, *2*(5), 359–366, doi:10.1016/0893-6080(89)90020-8.

Jalalkamali, A. (2015), Using of hybrid fuzzy models to predict spatiotemporal groundwater quality parameters, *Earth Sci. Inf.*, *8*, 885–894, doi:10.1007/s12145-015-0222-6.

Jha, M. K., and S. Sahoo (2015), Efficacy of neural network and genetic algorithm techniques in simulating spatio-temporal fluctuations of groundwater, *Hydrol. Processes*, *29*(5), 671–691, doi:10.1002/hyp.10166.

Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie (2003), The DSSAT cropping system model, *Eur. J. Agron.*, *18*(3–4), 235–265, doi:10.1016/S1161-0301(02)00107-7.

Kalwij, I. M., and R. C. Peralta (2006), Simulation/optimization modeling for robust pumping strategy design, *Ground Water*, *44*(4), 574–582, doi:10.1111/j.1745-6584.2006.00176.x.

Kasiviswanathan, K. S., J. He, K. P. Sudheer, and J.-H. Tay (2016), Potential application of wavelet neural network ensemble to forecast streamflow for flood management, *J. Hydrol.*, *536*, 161–173, doi:10.1016/j.jhydrol.2016.02.044.

Kerachian, R., and M. Karamouz (2006), Optimal reservoir operation considering the water quality issues: A stochastic conflict resolution approach, *Water Resour. Res.*, *42*, W12401, doi:10.1029/2005WR004575.

Knotters, M., and M. F. P. Bierkens (2001), Predicting water table depths in space and time using a regionalised time series model, *Geoderma*, *103*(1–2), 51–77, doi:10.1016/S0016-7061(01)00069-6.

Konikow, L. F. (2013), Groundwater depletion in the United States (1900–2008), U.S. Geol. Surv. Sci. Invest. Rep., 2013-5079, 63 p., doi: 10.1111/gwat.12306.

Kurtulus, B., and M. Razack (2010), Modeling daily discharge responses of a large karstic aquifer using soft computing methods: Artificial neural network and neuro-fuzzy, *J. Hydrol.*, *381*(1–2), 101–111, doi:10.1016/j.jhydrol.2009.11.029.

Kuss, A. J. M., and J. J. Gurdak (2014), Groundwater level response in U.S. principal aquifers to ENSO, NAO, PDO, and AMO, *J. Hydrol.*, *519*, 1939–1952, doi:10.1016/j.jhydrol.2014.09.069.

Levenberg, K. (1944), A method for the solution of certain problems in least squares, *Q. Appl. Math.*, *2*, 164–168.

Lipták, B. (2005), *Instrument Engineers' Handbook: Process Control and Optimization*, vol. 2, CRC Press, Boca Raton, Fla.

Luckey, R. R., and M. F. Becker (1999), Hydrogeology, water use, and simulation of flow in the High Plains aquifer in northwestern Oklahoma, southeastern Colorado, southwestern Kansas, northeastern New Mexico, and northwestern Texas, U.S. Geol. Surv. Water Resour. Invest. Rep., 99-4104, 73 p.

Ludwig, O., and U. Nunes (2010), Novel maximum-margin training algorithms for supervised neural networks, *IEEE Trans. Neural Networks*, *21*(6), 972–984, doi:10.1109/TNN.2010.2046423.

MacKay, D. J. (2008), *Information Theory, Inference, and Learning Algorithms*, Cambridge Univ. Press, Cambridge, U. K.

Maier, H. R., and G. C. Dandy (2000), Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications, *Environ. Modell. Software*, *15*(1), 101–124, InTech, Rijeka, Croatia-European Union, doi:10.1016/S1364-8152(99)00007-9.

Maier, H. R., A. C. Zecchin, L. Radbone, and P. Goonan (2006), Optimising the mutual information of ecological data clusters using evolutionary algorithms, *Math. Comput. Modell.*, *44*(5–6), 439–450, doi:10.1016/j.mcm.2006.01.004.

Marquardt, D. W. (1963), An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, *11*(2), 431–441, doi: 10.1137/0111030.

Maupin, M., J. Kenny, S. Hutson, J. Lovelace, N. Barber, and K. Linsey (2014), Estimated use of water in the United States in 2010, *U.S. Geol. Surv. Circ.*, *1405*, 56 p.

May, R., G. Dandy, and H. Maier (2011), Review of input variable selection methods for artificial neural networks, in *Artificial Neural Networks—Methodological Advances Biomedical Applications*, vol. 362, InTech, Rijeka, Croatia-European Union, doi:10.5772/644.

McGuire, V. (2013), Water-level and storage changes in the High Plains aquifer, predevelopment to 2011 and 2009–11, *U.S. Geol. Surv. Sci. Invest. Rep.*, *2012-5291*, 15 p.

McGuire, V. L. (2004), Water-level changes in the High Plains aquifer, predevelopment to 2002, 1980 to 2002, and 2001 to 2002, *U.S. Geol. Surv. Fact Sheet*, *2004-3026*, 6 p.

McGuire, V. L. (2014), Water-level changes and change in water in storage in the High Plains aquifer, predevelopment to 2013 and 2011–13, *U.S. Geol. Surv. Sci. Invest. Rep.*, *5218*, 14 p., doi:10.3133/sir20145218.

Minns, A., and M. Hall (1996), Artificial neural networks as rainfall-runoff models, *Hydrol. Sci. J.*, *41*(3), 399–417, doi:10.1080/02626669609491511.

Mishra, A. K., A. V. M. Ines, V. P. Singh, and J. W. Hansen (2013), Extraction of information content from stochastic disaggregation and bias corrected downscaled precipitation variables for crop simulation, *Stochastic Environ. Res. Risk Assess.*, *27*(2), 449–457, doi:10.1007/s00477-012-0667-9.

Morrison, D., and C. Munster (2015), Groundwater app to determine flow direction and gradient, *Groundwater*, *53*(2), 342–346, doi:10.1111/gwat.12211.

Nikolos, I. K., M. Stergiadi, M. P. Papadopoulou, and G. P. Karatzas (2008), Artificial neural networks as an alternative approach to groundwater numerical modelling and environmental design, *Hydrol. Processes*, *22*(17), 3337–3348, doi:10.1002/hyp.6916.

NOAA (2016a), *El Niño Southern Oscillation (ENSO) Data*, Earth Syst. Res. Lab., Phys. Sci. Div., Boulder, Colo. [Available at http://www.esrl.noaa.gov/psd/enso/mei/#data.]
NOAA (2016b), *Pacific Decadal Oscillation (PDO) Data*, Natl. Cent. for Environ. Inf., Asheville, N. C. [Available at https://www.ncdc.noaa.gov/teleconnections/pdo/.]
NOAA (2016c), *North Atlantic Oscillation (NAO) Data*, Natl. Weather Serv., Clim. Predict. Cent., College Park, Md. [Available at http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml.]
Nourani, V., M. T. Alami, and F. D. Vousoughi (2015), Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling, *J. Hydrol.*, *524*, 255–269, doi:10.1016/j.jhydrol.2015.02.048.
Olden, J. D., M. K. Joy, and R. G. Death (2004), An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data, *Ecol. Modell.*, *178*(3–4), 389–397, doi:10.1016/j.ecolmodel.2004.03.013.
Parkin, G., S. J. Birkinshaw, P. L. Younger, Z. Rao, and S. Kirk (2007), A numerical modelling and neural network approach to estimate the impact of groundwater abstractions on river flows, *J. Hydrol.*, *339*(1–2), 15–28, doi:10.1016/j.jhydrol.2007.01.041.
Pedhazur, E. J. (1997), Multiple Regression in Behavioral Research, New York Harcourt Brace Coll., New York.
Peterson, S. M., J. S. Stanton, A. T. Saunders, and J. R. Bradley (2008), Simulation of ground-water flow and effects of ground-water irrigation on base flow in the Elkhorn and Loup River Basins, Nebraska, *U.S. Geol. Surv. Sci. Invest. Rep., 2008-5143*, 65 p.
Prechelt, L. (2012), Early stopping—But when?, in *Neural Networks: Tricks of the Trade*, edited by G. Montavon and K.-R. Müller, pp. 53–67, Springer, Berlin.
Quilty, J., J. Adamowski, B. Khalil, and M. Rathinasamy (2016), Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling, *Water Resour. Res.*, *52*, 2299–2326, doi:10.1002/2015WR016959.
Russo, T. A., and U. Lall (2017), Depletion and response of deep groundwater to climate-induced pumping variability, *Nat. Geosci.*, *10*, 105–108, doi:10.1038/ngeo2883.
Russo, T. A., N. Devineni, and U. Lall (2015), Assessment of agricultural water management in Punjab, India using Bayesian methods, in *Sustainability of Integrated Water Resources Management: Water Governance, Climate and Ecohydrology*, edited by S. Setegn and M. Donoso, pp. 147–162, Springer, Switzerland.
Sahoo, S., and M. K. Jha (2013), Groundwater-level prediction using multiple linear regression and artificial neural network techniques: A comparative assessment, *Hydrogeol. J.*, *21*(8), 1865–1887, doi:10.1007/s10040-013-1029-5.
Sahoo, S., and M. K. Jha (2015), On the statistical forecasting of groundwater levels in unconfined aquifer systems, *Environ. Earth Sci.*, *73*(7), 3119–3136, doi:10.1007/s12665-014-3608-8.
Sahoo, S., T. Russo, and U. Lall (2016), Comment on "Quantifying renewable groundwater stress with GRACE" by Alexandra S. Richey et al., *Water Resour. Res.*, *52*, 4184–4187, doi:10.1002/2015WR018085.
Scanlon, B. R., C. C. Faunt, L. Longuevergne, R. C. Reedy, W. M. Alley, V. L. Mcguire, and P. B. Mcmahon (2012), Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, *Proc. Natl. Acad. Sci. U. S. A.*, *109*(24), 9320–9325, doi:10.1073/pnas.1200311109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1200311109.
Seber, G. A. F., and C. J. Wild (2003), *Nonlinear Regression*, John Wiley, New York.
Shamseldin, A. Y. (1997), Application of a neural network technique to rainfall-runoff modelling, *J. Hydrol.*, *199*(3–4), 272–294, doi:10.1016/S0022-1694(96)03330-6.
Shannon, C. E., and W. Weaver (1949), *The Mathematical Theory of Communication*, Univ. of Ill. Press, Urbana.
Sharma, A. (2000), Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1—A strategy for system predictor identification, *J. Hydrol.*, *239*(1–4), 232–239, doi:10.1016/S0022-1694(00)00346-2.
Shiri, J., and O. Kisi (2011), Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations, *Comput. Geosci.*, *37*(10), 1692–1701, doi:10.1016/j.cageo.2010.11.010.
Shrestha, D. L., N. Kayastha, and D. P. Solomatine (2009), A novel approach to parameter uncertainty analysis of hydrological models using neural networks, *Hydrol. Earth Syst. Sci.*, *13*(7), 1235–1248, doi:10.5194/hess-13-1235-2009.
Sinnott, R. W. (1984), Virtues of the Haversine, *Sky Telescope*, *68*(2), 159.
Steward, D. R., and A. J. Allen (2015), Peak groundwater depletion in the High Plains Aquifer, projections from 1930 to 2110, *Agric. Water Manage.*, *170*, 36–48, doi:10.1016/j.agwat.2015.10.003.
Taylor, R. G., et al. (2012), Ground water and climate change, *Nat. Clim. Change*, *3*(4), 322–329, doi:10.1038/nclimate1744.
Thornton, P. E., M. M. Thornton, B. W. Mayer, N. Wilhelmi, Y. Wei, R. Devarakonda, and R. B. Cook (2014), Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2, Oak Ridge Natl. Lab. Distrib. Act. Arch. Cent., Oak Ridge, Tenn., doi:10.3334/ORNLDAAC/1219.
Touretzky, D., and D. Pomerleau (1989), What's hidden in the hidden layers, *Byte*, *14*(8), 227–233.
USGS (2015), National Water Information System Data: USGS Water Data for the Nation, U.S. Geol. Surv., doi:10.5066/F7P55KJN. [Available at https://waterdata.usgs.gov/nwis/gw.]
Vautard, R., P. Yiou, and M. Ghil (1992), Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Physica D*, *58*(1–4), 95–126, doi:10.1016/0167-2789(92)90103-T.
Wada, Y., L. P. H. van Beek, C. M. van Kempen, J. W. T. M. Reckman, S. Vasak, and M. F. P. Bierkens (2010), Global depletion of groundwater resources, *Geophys. Res. Lett.*, *37*, L20402, doi:10.1029/2010GL044571.
Wang, Y., S. Guo, H. Chen, and Y. Zhou (2014), Comparative study of monthly inflow prediction methods for the Three Gorges Reservoir, *Stochastic Environ. Res. Risk Assess.*, *28*(3), 555–570, doi:10.1007/s00477-013-0772-4.
Wu, C. L., K. W. Chau, and Y. S. Li (2009), Methods to improve neural network performance in daily flows prediction, *J. Hydrol.*, *372*(1–4), 80–93, doi:10.1016/j.jhydrol.2009.03.038.
Yan, S., and B. Minsker (2006), Optimal groundwater remediation design using an adaptive neural network genetic algorithm, *Water Resour. Res.*, *42*, W05407, doi:10.1029/2005WR004303.
Yeh, M.-S., Y.-P. Lin, and L.-C. Chang (2006), Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms, *Environ. Geol.*, *50*(1), 101–121, doi:10.1007/s00254-006-0190-8.
Yoon, H., S.-C. Jun, Y. Hyun, G.-O. Bae, and K.-K. Lee (2011), A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, *J. Hydrol.*, *396*(1–2), 128–138, doi:10.1016/j.jhydrol.2010.11.002.