

Water Table Depth Estimates over the Contiguous United States Using a Random Forest Model

by Yueling Ma^{1,2,3} , Elena Leonarduzzi^{1,2}, Amy Defnet^{1,2,4}, Peter Melchior^{5,6}, Laura E. Condon⁷ , and Reed M. Maxwell^{1,2,4} 

Abstract

Water table depth (WTD) has a substantial impact on the connection between groundwater dynamics and land surface processes. Due to the scarcity of WTD observations, physically-based groundwater models are growing in their ability to map WTD at large scales; however, they are still challenged to represent simulated WTD compared to well observations. In this study, we develop a purely data-driven approach to estimating WTD at continental scale. We apply a random forest (RF) model to estimate WTD over most of the contiguous United States (CONUS) based on available WTD observations. The estimated WTD are in good agreement with well observations, with a Pearson correlation coefficient (r) of 0.96 (0.81 during testing), a Nash-Sutcliffe efficiency (NSE) of 0.93 (0.65 during testing), and a root mean square error (RMSE) of 6.87 m (15.31 m during testing). The location of each grid cell is rated as the most important feature in estimating WTD over most of the CONUS, which might be a surrogate for spatial information. In addition, the uncertainty of the RF model is quantified using quantile regression forests. High uncertainties are generally associated with locations having a shallow WTD. Our study demonstrates that the RF model can produce reasonable WTD estimates over most of the CONUS, providing an alternative to physics-based modeling for modeling large-scale freshwater resources. Since the CONUS covers many different hydrologic regimes, the RF model trained for the CONUS may be transferrable to other regions with a similar hydrologic regime and limited observations.

¹High Meadows Environmental Institute, Princeton University, Princeton, NJ, USA

²Integrated GroundWater Modeling Center, Princeton University, Princeton, NJ, USA

³Corresponding author: High Meadows Environmental Institute, Princeton University, Princeton, NJ 08544; ym5379@princeton.edu

⁴Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

⁵Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

⁶Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA

⁷Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

Article impact statement: This paper provides reliable water table depth estimates over the contiguous United States using a random forest model and observations.

Received January 2023, accepted September 2023.

© 2023 The Authors. *Groundwater* published by Wiley Periodicals LLC on behalf of National Ground Water Association.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

doi: 10.1111/gwat.13362

Introduction

Groundwater is an important terrestrial compartment of the hydrologic cycle with broad implications for human health, food security, terrestrial ecosystems, and energy production (van der Gun 2020). Water table depth (WTD) refers to the depth of the upper surface of the water saturated aquifers to the land surface and plays a critical role in the linkage between groundwater dynamics and land surface processes (Kollet and Maxwell 2008). During times of drought, shallow groundwater provides baseflows to surface water bodies such as rivers and lakes and maintains water in the soil and vegetations, which later acts as a source term of evapotranspiration (ET) to the atmosphere (York et al. 2002; Maxwell et al. 2007, 2011; Fan et al. 2013; Maxwell and Condon 2016; Condon and Maxwell 2019; Furusho-Percot et al. 2019; Hartick et al. 2022; Ryken et al. 2022). Moreover, due to unsustainable anthropogenic activities, groundwater is being depleted extensively in many parts of the world, which is reflected in the depletion of WTD (Koch et al. 2019). Therefore, a high-resolution WTD map helps better characterize freshwater availability and aid in our

understanding of the connection between groundwater dynamics and land surface processes.

Lack of WTD observations is a global challenge (Fan et al. 2013; Rust et al. 2018; Liesch and Wunsch 2019; Ma et al. 2022). In the United States, we are fortunate to have about 1 million WTD monitoring wells from the U.S. Geological Survey (USGS) and local environmental agencies as well as remotely sensed data from the Gravity Recovery and Climate Experiment (GRACE) satellite mission launched in 2002. However, the WTD monitoring wells are sparse at the scale of individual watersheds or farms (where local decision making happens). GRACE terrestrial water storage anomalies are of good spatial coverage, but their spatial resolution is coarse, around 300 km, and they account for variations in both near-surface water and groundwater storage (Chen et al. 2016).

Traditionally, physically-based groundwater models have been used to estimate WTD at large scales, such as Fan et al. (2013) and de Graaf et al. (2015) for the globe, and Maxwell et al. (2015) for most of the contiguous United States (CONUS). Owing to the computational expense of running these simulations, large-scale groundwater models are typically not calibrated (Condon et al. 2021; Gleeson et al. 2021). Reinecke et al. (2020) compared simulated WTD from several groundwater models with well observations in Canterbury, New Zealand, including those of Fan et al. (2013) and de Graaf et al. (2015). They showed that all simulated WTD had large discrepancies when compared to observations.

As an alternative to physically-based modeling, machine learning (ML) methods are able to learn complex nonlinear relationships between groundwater dynamics and atmospheric and land surface processes from historical data (Ma et al. 2021a, 2021b, 2022). Their low background knowledge requirement, fast simulation time, and acceptable accuracy have led to their increasing use in groundwater resource modeling, as documented in the review articles, Rajaei et al. (2019) and Osman et al. (2022). Among the available studies, most of them focus on the use of ML techniques to predict the temporal variations in the groundwater system at aquifer and watershed scales, for example, Sun (2013); Gholami et al. (2015); Zhang et al. (2018); Vu et al. (2021); and Wunsch et al. (2022). However, few studies (Bechtold et al. 2014; Koch et al. 2019, 2021; Gonzalez and Arsanjani 2021; Lendzioch et al. 2021; Schneider et al. 2022) have assessed the ability of data-driven modeling to estimate the spatial changes in WTD. These studies focus on small areas with limited hydrogeological settings.

The successful application of advanced ML methods such as convolutional neural networks and Long Short-Term Memory networks relies on large amounts of spatially or temporally continuous data, but they do not apply to the estimation of the spatial changes in WTD that has sparse point observations. Random forest (RF) is a tree-based ML method that has a relatively simple architecture (Osman et al. 2022). As such, RF

is useful in the case of limited training data. Koch et al. (2019) have successfully employed a RF model to predict a WTD map for a wintertime minimum event in Denmark ($\sim 15,000 \text{ km}^2$), demonstrating the promising model performance in reproducing the spatial details of WTD.

In this study, we apply a RF model to map WTD for the same study domain as Maxwell et al. (2015), that is, most of the CONUS ($\sim 6.3 \text{ million km}^2$). The study domain incorporates a wide range of geologic, climatic, and topographic conditions (Gleeson et al. 2011), leading to many different hydrogeological settings and sophisticated groundwater systems, which poses challenges for such large-scale reconstruction. The WTD data estimated by the RF model are compared to WTD observations. An additional feature of RF is the ability to estimate parameter sensitivity and uncertainty. As such, we conduct a sensitivity analysis to input variables, and evaluate the uncertainty of the RF model using quantile regression forests. The comprehensive analysis of the performance of the RF model over most of the CONUS provides new insights for groundwater hydrology.

Methods

Study Area and Data

This study is performed in the same area as Maxwell et al. (2015), that is, most of the CONUS. The study domain (refer to Figure 1 for the domain extent) spans approximately 6.3 million km^2 ($3342 \text{ km} \times 1888 \text{ km}$) at 1-km lateral grid spacing, encompassing the majority of eight major river basins in the United States at high resolution (Maxwell and Condon 2016; O'Neill et al. 2021).

Figure 1 displays long-term mean WTD observations over the study domain. In general, WTD observations are sparse at the watershed scale, and are unevenly distributed within the study domain. The observed WTD mean data were derived from historical WTD measurements over 1914 to 2023 at 263,417 USGS WTD monitoring wells and long-term mean WTD measurements over 1927 to 2009 at 262,724 wells from Fan et al. (2013), with a minor overlap ($\sim 10\%$) between the wells in the two data sets. Approximately 61% of the USGS WTD monitoring wells (159,887 wells) only have a single observation. For wells with one observation, we use its observation as the long-term mean value for that well, which may introduce additional uncertainty. All WTD data have been validated using at least two of the following four criteria. First, the attribute “reliability_cd” in the USGS metadata is “C,” indicating that the data have been checked by the reporting agency. Second, the data are limited in range from 0 to 300 m to exclude deep confined aquifers that have little connections to the physics of the water system. Third, the z score for the difference between the data at the studied time step and the data at the last step is ≤ 3 , so that 99.7% of the data at a well are included. Fourth, the data at a well are considered to have a strong

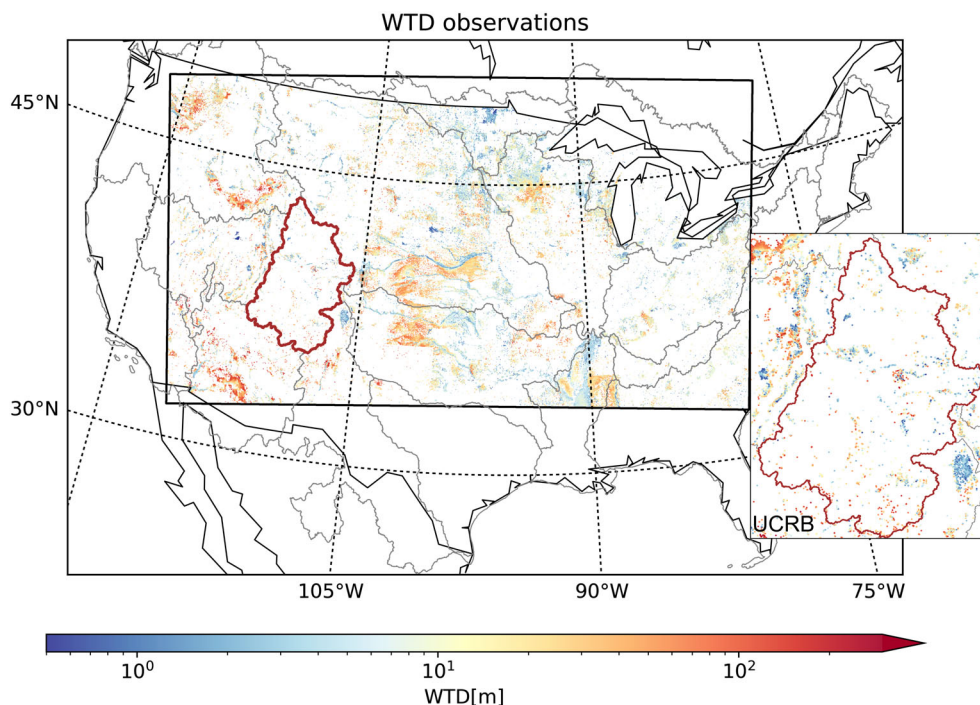


Figure 1. Map of long-term mean water table depth (WTD) observations over most of the CONUS with an inset zooming into the Upper Colorado River Basin (UCRB), one of the principal headwater basins in the United States. There are no WTD observations in the blank areas.

similarity with its neighbors within a spatial distance of 0.02° (~ 2 km), according to the results of the Local Outlier Factor algorithm (Breunig et al. 2000). The spatial distance of 0.02° is selected to account for the 1-km grid cell where the well is located and its adjacent 1-km grid cells.

The inset of Figure 1 zooms into the Upper Colorado River Basin (UCRB), one of the principal headwater basins in the United States. The elevations in the UCRB change substantially from peaks higher than 3300 m at the Rocky Mountains to around 900 m at Lee's Ferry in Arizona. The big differences in elevations lead to spatial variations in the climate, from alpine conditions in the east to semiarid in the west. The Colorado River provides water for major cities such as Los Angeles, Phoenix, and Las Vegas and more than 40 million people (Tran et al. 2020).

Here, the RF model is constructed using 1-km gridded mean WTD observations, which are estimated by averaging the long-term mean WTD observations from all the wells located at the same 1-km grid cell. WTD observations are available for a total of 208,665 grid cells ($\sim 3\%$ of all the grid cells). A map showing the number of wells per 1-km grid cell over most of the CONUS is provided in Figure S1. The number of wells at a grid cell with WTD observations varies from 1 to 594, with a median of 2 and a standard deviation of 3.38. For grid cells with more than one WTD monitoring wells, the standard deviation of WTD at a grid cell varies from 0 to 134.75 m, with a median of 0.16 m. In the UCRB, only 2246 grid cells have WTD observations, about 0.77% of all the grid

cells in the basin. The lack of WTD observations impedes the understanding of its hydrologic system.

The input variables of the RF model are annual mean precipitation, annual mean temperature, precipitation minus evapotranspiration (PME), elevation, topographic slope, natural log of hydraulic conductivity ($\ln K$), x , and y locations. The x and y locations represent the locations of a grid cell in the study domain on the x and y axes, respectively. Anthropogenic impacts are not considered in the inputs, due to the lack of data. Figure S2 illustrates the Spearman's rank correlation coefficients between the input variables. The strongest positive and negative correlations are found between annual mean precipitation and PME (0.89) and between annual mean temperature and y location (-0.83), respectively. All the input variables are available at a spatial resolution of 1 km over most of the CONUS. The input data are the same as those used by Maxwell et al. (2015), except for annual mean precipitation, annual mean temperature, and $\ln K$. The annual mean precipitation and temperature data were computed from the precipitation and temperature data interpolated from the North American Land Data Assimilation System Phase 2 (NLDAS-2; Xia et al. 2015; Cosgrove et al. 2003). The $\ln K$ data were derived from a continental-scale subsurface K dataset developed by Tijerina-Kreuzer et al. (2023). Table 1 presents an overview of the input variables used in the RF model. The input variables are further divided into three groups based on physical relationships, namely climatological-related variables, geology-related variables, and location, for which information is also provided in Table 1.

Table 1
Overview of the Input Variables used in the Random Forest Model.

Input Variable	Data Source	Group
Annual mean precipitation	North American Land Data Assimilation System Phase 2 (NLDAS-2; Xia et al. 2015; Cosgrove et al. 2003)	Climatology-related
Annual mean temperature		
PME	Maxwell et al. (2015)	Geology-related
Elevation		
Topographic slope	Tijerina-Kreuzer et al. (2023)	Location
lnK		
x	Not applicable	Location
y		

RF Model

We build a RF model (Figure 2) for the entire study domain (i.e., most of the CONUS). RF is a ML method proposed by Breiman (2001), consisting of a collection of decision trees. In this study, a number of decision trees are constructed to learn the linkage between the input variables (i.e., climatological-related variables, geology-related variables, and location) and the output variable (i.e., long-term mean WTD) during training, and then the median prediction of these trees is calculated as the estimated WTD, similar to other approaches for soil moisture (Abbaszadeh et al. 2019). To increase the diversity within the ensemble of decision trees, we produce a unique bootstrap sample of the original training set for each decision tree. Based on sampling with replacement, each bootstrap sample contains a part of the original samples but remains the original sample size. This technique is known as “bagging.” In addition, a randomly selected subgroup of input variables is used to train each decision tree. In the training process, each decision tree recursively divides a random subset of the training data obtained from the aforementioned processes into more homogenous groups by decision rules at nodes (circles in Figure 2; Koch et al. 2019). The ensemble of the slightly different decision trees in the RF model results in a robust prediction for WTD estimates.

As aforementioned, compared with other ML techniques such as artificial neural networks, RF has a less complicated structure and fewer hyperparameters (i.e., adjustable parameters that control model behavior and complexity), and thus it requires less data for training. Furthermore, RF often works well without heavy hyperparameter tuning and data normalization, which simplifies the training process. On the other hand, RF faces a trade-off between model performance and computational cost. A more accurate prediction necessitates more decision trees, leading to slower computation of results (Müller and Guido 2017).

Here, we randomly split the data at 208,665 grid cells with WTD observations into a training set (80% of the data, 166,932 grid cells) and a test set (20% of the data, 41,733 grid cells) for the training and testing processes, respectively. The random splitting of

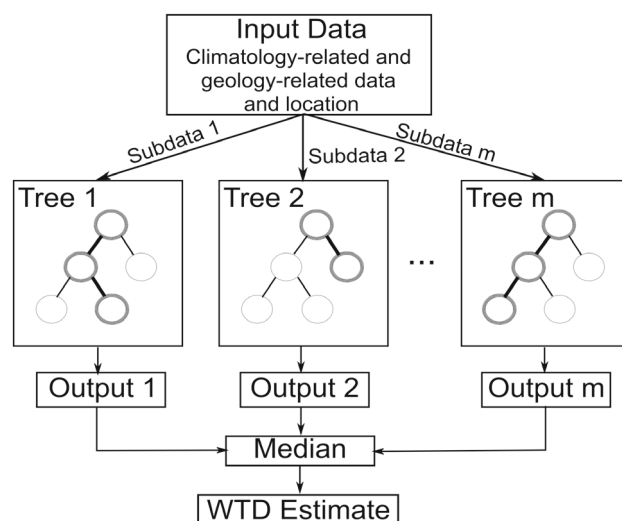


Figure 2. Schematic diagram showing how the random forest (RF) model works in this study. The RF model is an ensemble of decision trees. Each tree is slightly different from each other, resulting in various outputs. The median of the outputs from all decision trees in the RF model is computed as estimated water table depth (WTD).

the data guarantees the generalizability of the RF model in estimating WTD (i.e., the ability of the trained RF model to handle previously unobserved data). To search for the best hyperparameter configuration of the RF model, we perform the five-fold cross-validation during training. In the five-fold cross-validation, the original training set is equally partitioned into five small sets (named folds, 20% of the original training set). Of the five folds, a single fold is preserved as the validation data for evaluating the performance of the trained RF model, and the remaining four folds (80% of the original training set) are used as training data. The cross-validation process is repeated five times, with each fold used once as the validation data. The optimal values for the hyperparameters listed in Table 2 are selected based on average validation accuracies. The cross-validation approach utilizes all the training data in hyperparameter tuning, which is useful in this study where the grid cells used for training are sparse over the study domain. Finally, the entire training set is used to train

Table 2
Hyperparameter Setting of the Random Forest Model.

Hyperparameter	Description	Range, Optimal Value
max_samples	Percentage of samples used to train each tree	0.1–1.0, 0.8
max_features	Number of features used to train each tree	1–8, 5
n_estimators	Total tree number	100–1000, 300
max_depth	Maximum depth of a tree	100–1000, 900
min_samples_leaf	Minimum required number of observations to split a node	1–5, 1

the RF model with the optimal hyperparameter setting, and the resulting model is tested on the test set for final evaluation. A WTD map over most of CONUS is also generated by the trained model. In this study, the RF model is developed using scikit-learn, an open-source ML python library (Pedregosa et al. 2011), with slight modifications to obtain the median of tree outputs. Using two AMD EPYC 7402 24-core processors, it takes about 4 min to train the optimal RF model and another 4 min to simulate the CONUS WTD.

Permutation Importance of Input Variables

The calculation of permutation importance is common practice to analyze the sensitivity of the RF model to input variables, for example, Schneider et al. (2022) and Koch et al. (2019). The permutation importance is defined as the decrease in model accuracy when randomly shuffling the values of a single input variable (Breiman 2001), reflecting the contribution of the input variable on model performance. Here, the model accuracy is assessed by the root mean square error (RMSE). To eliminate bias in the outcome caused by a single permutation, we perform 30 different permutations of an input variable and compute the relative mean decrease in model accuracy (i.e., the mean increase in RMSE divided by the original RMSE) as the permutation importance of the input variable.

Some of our input variables are closely linked (illustrated in Figure S2), which may lead to an underestimation of their importance. To overcome this weakness, we also study the permutation importance of three groups of the input variables that are physically related (i.e., climatological-related variables, geology-related variables, and location), as suggested by Koch et al. (2019). The permutation importance of a group of input variables is calculated as the relative increase in RMSE when collectively permuting the input variables in the group.

Quantile Regression Forests

The RF model provides information about the full distribution of the estimated WTD generated by the involved decision trees, not only about the median. This is the basis of implementing quantile regression forests introduced by Meinshausen (2006). The main idea of the approach is to address the uncertainty associated with the RF model based on the quantiles of the distribution of the tree outputs at each grid cell (Meinshausen 2006; Koch

et al. 2019). Here, we utilize the coefficient of variation of the outputs from the 300 decision trees in the RF model to express the uncertainty, which is calculated by the standard deviation of the 300 tree outputs divided by their mean.

Results and Discussion

Estimated WTD Map over the Study Domain

Figure 3 displays long-term mean WTD estimates generated by the RF model for the study domain. WTD varies in space, from 0 to 299.59 m. The WTD mean and median are 18.29 m and 11.43 m, respectively. In general, deeper WTD is in the more arid western regions, while shallower WTD is in the more humid eastern regions. In addition, shallow WTD exists along river channels. These findings are consistent with Maxwell et al. (2015). The inset of Figure 3 presents the estimated WTD in the UCRB, varying from 0 m in the east to 299.31 m in the west, which is reasonable. As aforementioned, because of the spatial heterogeneity of elevations, there are distinct climates in the basin. The larger precipitation (1000 mm/year) in the eastern regions due to the alpine climate results in more recharge to the local groundwater systems and thus shallower WTD. Moreover, it is worth mentioning that the high-resolution UCRB WTD map is produced using the RF model based on the sparse WTD observations shown in the inset of Figure 1 (only about 0.77% of the grid cells with WTD observations). Yet, there also appear to be WTD artifacts in locations such as the Ohio River Basin, which require further investigation.

The scatter plot in Figure 4A and 4B demonstrates good agreement between the estimated and observed WTD for all 208,665 grid cells with WTD observations and 41,733 grid cells used for testing, where most data points are concentrated along the 1:1 line. As shown in Figure 4C, the estimated and observed WTD have similar distributions, with the peaking at 2 to 5 m. Overall, the RF model achieves a Pearson correlation coefficient (r) of 0.96, a Nash–Sutcliffe efficiency (NSE) of 0.93, and a RMSE of 6.87 m for all grid cells with WTD observations, and a r of 0.81, a NSE of 0.65, and a RMSE of 15.31 m in the testing process. Maxwell et al. (2015) used a physically-based groundwater model to simulate WTD in the same study domain, gaining a r of 0.25 and a RMSE of 30.03 m at all grid cells with WTD

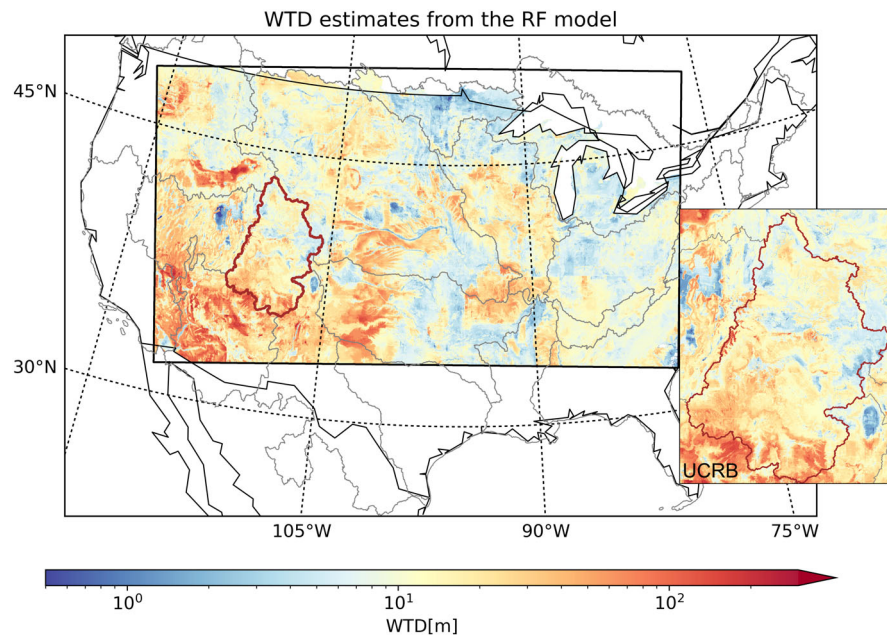


Figure 3. Map of long-term mean water table depth (WTD) estimates from the RF model for most of the CONUS with an inset zooming into the Upper Colorado River Basin (UCRB).

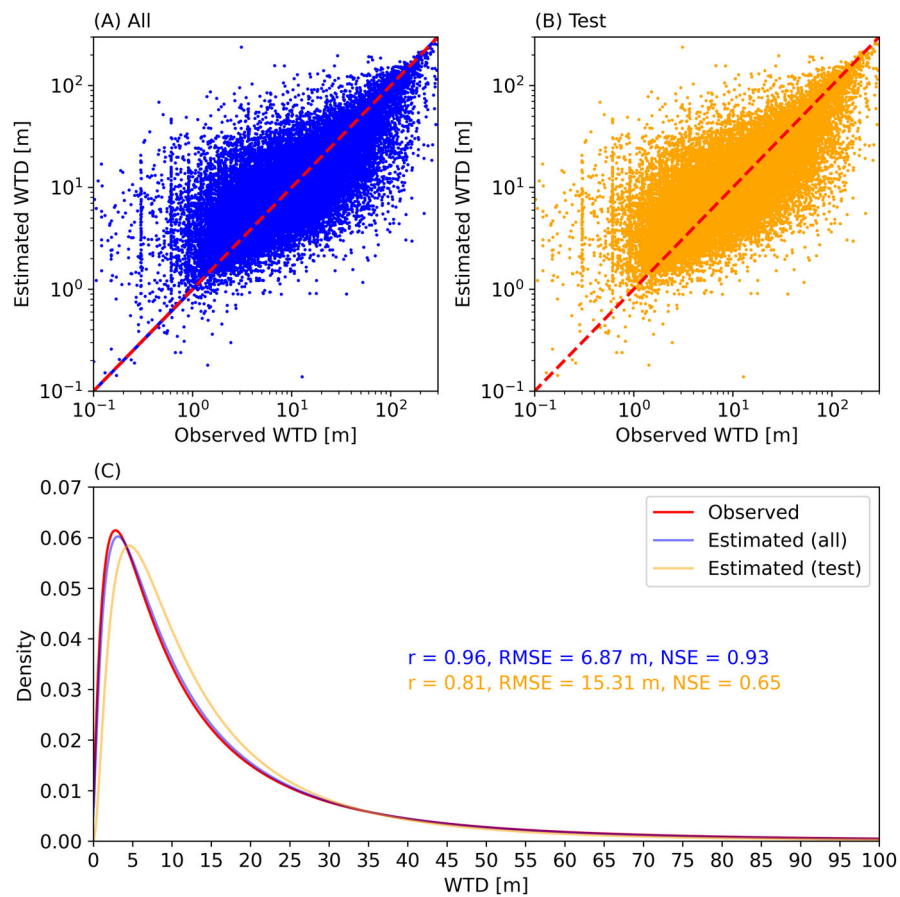


Figure 4. (A) Scatter plot of all estimated versus observed water table depth (WTD), (B) scatter plot of test estimated versus observed WTD, and (C) probability density function plot of estimated and observed WTD for most of the CONUS, where the estimated WTD are produced by the RF model at all grid cells with WTD observations (in total 208,665 grid cells) and the grid cells used for testing (in total 41,733 grid cells). The Pearson correlation coefficient (r), Nash–Sutcliffe efficiency (NSE), and root mean square error (RMSE) between estimated and observed WTD are presented here. Note the log scale used for the axes in (A) and (B).

observations. There is no obvious relationship between the density of WTD observations in a water basin and the RF model performance (Figure S3). Although we do not explicitly include anthropogenic impacts (e.g., groundwater pumping) in the RF model, the model preserves its performance at wells with documented pumping activities (not shown here). Thus, we suspect that the RF model might be capable of learning or inferring pumping from WTD observations; however, that is not explored in any detail in this study. Moreover, we compare the estimated and observed WTD in the Ohio River Basin where WTD artifacts are observed in Figure 3 (Figure S4). In the Ohio River Basin, the RF model performs relatively poorly during testing with a r of 0.58. The distribution of the estimated WTD also shows significant differences from the observed WTD at the grid cells used for testing (Figure S4C).

Sensitivity Analysis of the RF Model to Input Variables

The sensitivity of the RF model to input variables is assessed based on permutation importance, which is expressed by the relative increase in RMSE. Figure 5 summarizes the permutation importance of the eight input variables and the three groups of input variables computed on the test set, revealing the contribution of input variables or groups of input variables to the generalizability of the trained RF model. The input variables in order of decreasing importance are x location (65.47%), y location (60.02%), elevation (43.10%), annual mean temperature (21.91%), annual mean precipitation (17.37%), PME (8.55%), lnK (8.46%), and topographic slope (1.21%). Due to the importance of x and y locations in the input variables, their group location (79.98%) also plays the most critical role in estimating WTD over the study domain. The RF model used in this study are not able

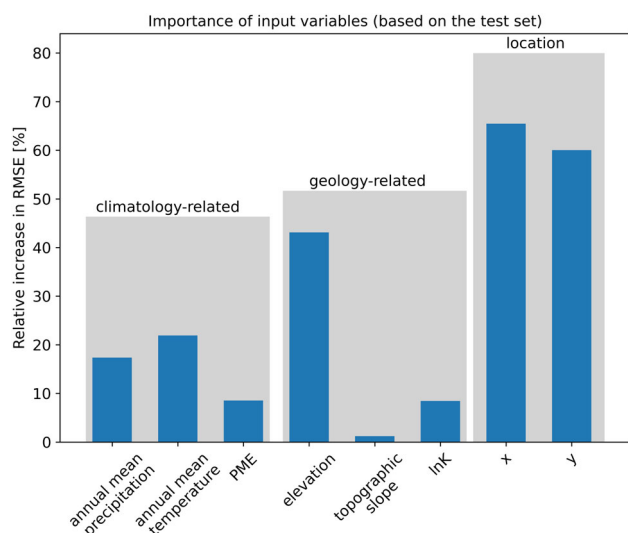


Figure 5. Input variable importance of the trained random forest model. The permutation importance is calculated based on the relative increase in the test root mean square error (RMSE). Permutation is applied to both single input variables (blue) and groups of input variables (gray).

to exploit the spatial dependencies in the data, and the location of each grid cell might be a surrogate for spatial information in the model, which is important in modeling groundwater flow. The WTD is calculated from the difference between local elevation and the upper surface of the water saturated aquifers (i.e., water table). As such, it is reasonable that elevation is the third important input variable. When x and y locations are removed from the input, the RF model still achieves good performance in estimating WTD over most of the CONUS, with a test NSE of 0.57, and elevation plays the most important role. Condon et al. (2015) also demonstrated the importance of elevation on WTD in eight major river basins over CONUS using a modified k-regression algorithm. The geology-related variables (51.66%) to which elevation belongs are the second important group of input variables. In the same group, there are also lnK and topographic slope. Topographic slope is found to have the least contribution to the estimation of WTD over CONUS in this study, contrary to the findings of Condon et al. (2015) and Condon and Maxwell (2015). This discrepancy might be explained by the latter two studies focusing on linear correlations between variables. Finally, the climatology-related variables (46.34%) are considered as less important, consisting of annual mean precipitation, annual mean temperature, and PME. As aforementioned, PME is precipitation minus ET, which is somewhat redundant with annual mean precipitation and temperature. Hence, PME shows less importance here. The permutation importance calculated on the training set (Figure S5) provide similar results.

Uncertainty Analysis

Figure 6 displays the resulting uncertainty based on quantile regression forests, which is represented by the coefficient of variation of the outputs of all 300 decision trees in the RF model at each grid cell. Similar to the estimated WTD presented in Figure 3, the uncertainty changes spatially, from 0.02 to 17.29, with a mean of 0.98 and a median of 0.96. Uncertainties are higher in most western regions including the Great Basin, UCRB, and Pacific Northwest Basin, and lower in some eastern regions such as the Lower Mississippi River Basin and Great Lakes Basin. High uncertainties are also observed along river channels, though not all river channels as we see changes in the uncertainty as we move downstream from, for example, the Platte River to the Mississippi. Figure S6 shows a close connection between the uncertainty and the estimated WTD. In general, high uncertainties are linked to the locations with a shallow WTD and vice versa. It is important to note that the uncertainty in WTD is quite large (over 10) in some regions. To reduce the uncertainty, we may improve the performance of individual decision trees in the RF model by, for example, increasing input variables and training data. These uncertainties are an important outcome of the statistical approach used here and provide more transparency in the model estimates and indication of where additional characterization data or

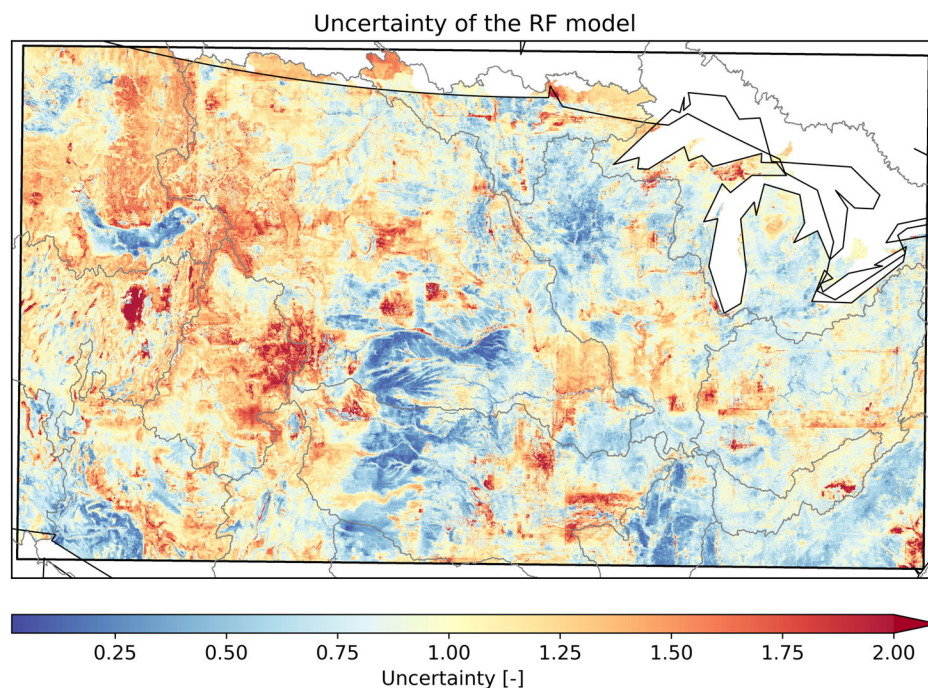


Figure 6. Uncertainty map showing the coefficient of variation of the outputs of all 300 decision trees in the random forest (RF) model over most of the CONUS.

observations are needed. Moreover, we observe more evident horizontal and vertical artifacts in Figure 6 compared to Figure 3. The uncertainty reflects the distribution of tree outputs in the RF model and is therefore more sensitive to outliers.

Conclusions

In this study, we develop a RF model to study the spatial variations in WTD over most of the CONUS based on available WTD observations and climatological and geological datasets. Using the RF model, we generate a map of long-term mean WTD estimates over most of the CONUS with a spatial resolution of 1 km. We compare the estimated and observed WTD in terms of data distribution, correlation, and evaluation metrics r , NSE, and RMSE. The RF model performs very well for all 208,665 grid cells with WTD observations, gaining a r of 0.96, a NSE of 0.93, and a RMSE of 6.87 m. For 41,733 grid cells used for testing, the RF model achieves a r of 0.81, a NSE of 0.65, and a RMSE of 15.31 m. The good performance can be attributed to the fact that the RF model is fully data-driven and provides a more direct approach to interpolate WTD values over the study domain. In addition, the location of each grid cell contributes most to the WTD estimation of the RF model over the study domain, which might be a substitute for spatial information. This is an important aspect as the RF model is not mass conservative yet learns the changes in WTD due to lateral flow likely using the locations of grid cells as a surrogate for flow. Moreover, there appears to be a strong linkage between uncertainty (represented by the correlation of variation of 300 tree outputs in the RF model) and estimated WTD,

with high uncertainties typically occurring in locations with a shallow WTD.

There is still much room of improvement in the developed RF model. The estimated WTD map (Figures 3 and 6) appears to be artifacts in some areas such as the Ohio River Basin, which require further investigation. In addition, large uncertainties (over 10) exist in the WTD estimates from the RF model. Nevertheless, our study shows the ability of RF in estimating spatial variations in WTD in a larger study domain with more complex hydrologic systems, that is, most of the CONUS. Therefore, this study can be considered as an extension of existing work, such as Koch et al. (2019), for a more challenging region. The study demonstrates that the developed RF model can produce good WTD estimates over the study domain, thereby providing alternative estimates of large-scale freshwater resources. In future, we plan to extend the study to the entire CONUS. RF does not account for the spatial resolution of input data. As such, without additional training, the trained RF model may generate a map with any spatial resolution by adjusting the spatial resolution of its input data. The resulting WTD map can be applied as the initial condition of a physically-based groundwater model to improve model performance. Particularly, for a region that has a similar hydrologic regime with a region in the study domain, the trained RF can be potentially transferred to the region, even if observational data is limited.

Acknowledgments

This research has been supported by the U.S. National Science Foundation Convergence Accelerator Program

(grant no. CA-2040542). Data products will be made available via the HydroGEN project (<https://hydrogen.princeton.edu/>) upon final publication. We thank the Editor in Chief (L. Konikow), Guest Editor (M. Tonkin), and three anonymous reviewers for their constructive comments, which have added to the quality and clarity of this work.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article. Supporting Information is generally *not* peer reviewed.

Figure S1. Map of the number of water table depth monitoring wells per 1-km grid cell over most of the contiguous United States.

Figure S2. Correlation heatmap showing Spearman's rank correlation coefficients between input variables.

Figure S3. Pearson correlation coefficients (r) for all grid cells (blue) and grid cell used for testing (yellow) in water basins with various densities of water table depth (WTD) observations.

Figure S4. (A) Scatter plot of all estimated versus observed water table depth (WTD), (B) scatter plot of test estimated versus observed WTD, and (C) probability density function plot of estimated and observed WTD for the Ohio River Basin, where the estimated WTD are produced by the RF model at all grid cells with WTD observations and the grid cells used for testing.

Figure S5. Input variable importance of the trained random forest model.

Figure S6. Relationship between estimated water table depth (WTD) and the uncertainty of the random forest (RF) model.

References

- Abbaszadeh, P., H. Moradkhani, and X. Zhan. 2019. Downscaling SMAP radiometer soil moisture over the CONUS using an ensemble learning method. *Water Resources Research* 55, no. 1: 324–344. <https://doi.org/10.1029/2018WR023354>
- Bechtold, M., B. Tiemeyer, A. Laggner, T. Leppelt, E. Frahm, and S. Belting. 2014. Large-scale regionalization of water table depth in peatlands optimized for greenhouse gas emission upscaling. *Hydrology and Earth System Sciences* 18, no. 9: 3319–3339. <https://doi.org/10.5194/hess-18-3319-2014>
- Breiman, L. 2001. Random forests. *Machine Learning* 45, no. 1: 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breunig, M.M., H.-P. Kriegel, R.T. Ng, and J. Sander. 2000. LOF: Identifying density-based local outliers. *ACM SIGMOD Record* 29, no. 2: 93–104. <https://doi.org/10.1145/335191.335388>
- Chen, J., J.S. Famiglietti, B.R. Scanlon, and M. Rodell. 2016. Groundwater storage changes: Present status from GRACE observations. *Surveys in Geophysics* 37, no. 2: 397–417. <https://doi.org/10.1007/s10712-015-9332-4>
- Condon, L.E., A.S. Hering, and R.M. Maxwell. 2015. Quantitative assessment of groundwater controls across major US river basins using a multi-model regression algorithm. *Advances in Water Resources* 82: 106–123. <https://doi.org/10.1016/j.advwatres.2015.04.008>
- Condon, L.E., S. Kollet, M.F.P. Bierkens, G.E. Fogg, R.M. Maxwell, M.C. Hill, H.H. Fransen, et al. 2021. Global groundwater modeling and monitoring: Opportunities and challenges. *Water Resources Research* 57, no. 12: e2020WR029500. <https://doi.org/10.1029/2020WR029500>
- Condon, L.E., and R.M. Maxwell. 2015. Evaluating the relationship between topography and groundwater using outputs from a continental-scale integrated hydrology model. *Water Resources Research* 51, no. 8: 6602–6621. <https://doi.org/10.1002/2014WR016774>
- Condon, L.E., and R.M. Maxwell. 2019. Simulating the sensitivity of evapotranspiration and streamflow to large-scale groundwater depletion. *Science Advances* 5, eaav4574. <https://doi.org/10.1126/sciadv.aav4574>
- Cosgrove, B.A., D. Lohmann, K.E. Mitchell, P.R. Houser, E.F. Wood, J.C. Schaake, A. Robock, C. Marshall, J. Sheffield, Q. Duan, L. Luo, R.W. Higgins, R.T. Pinker, J.D. Tarpley, and J. Meng. 2003. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres* 108, no. D22: 2002JD003118. <https://doi.org/10.1029/2002JD003118>
- de Graaf, I.E.M., E.H. Sutanudjaja, L.P.H. van Beek, and M.F.P. Bierkens. 2015. A high-resolution global-scale groundwater model. *Hydrology and Earth System Sciences* 19, no. 2: 823–837. <https://doi.org/10.5194/hess-19-823-2015>
- Fan, Y., H. Li, and G. Miguez-Macho. 2013. Global patterns of groundwater table depth. *Science* 339, no. 6122: 940–943. <https://doi.org/10.1126/science.1229881>
- Furusho-Percot, C., K. Goergen, C. Hartick, K. Kulkarni, J. Keune, and S. Kollet. 2019. Pan-European groundwater to atmosphere terrestrial systems climatology from a physically consistent simulation. *Scientific Data* 6, no. 1: 320. <https://doi.org/10.1038/s41597-019-0328-7>
- Gholami, V., K.W. Chau, F. Fadaee, J. Torkaman, and A. Ghafari. 2015. Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers. *Journal of Hydrology* 529: 1060–1069. <https://doi.org/10.1016/j.jhydrol.2015.09.028>
- Gleeson, T., L. Marklund, L. Smith, and A.H. Manning. 2011. Classifying the water table at regional to continental scales. *Geophysical Research Letters* 38, L05401. <https://doi.org/10.1029/2010GL046427>
- Gleeson, T., T. Wagener, P. Döll, S.C. Zipper, C. West, Y. Wada, R. Taylor, B. Scanlon, R. Rosolem, S. Rahman, N. Oshinlaja, R. Maxwell, M.H. Lo, H. Kim, M. Hill, A. Hartmann, G. Fogg, J.S. Famiglietti, A. Ducharme, I. de Graaf, M. Cuthbert, L. Condon, E. Bresciani, and M.F.P. Bierkens. 2021. GMD perspective: The quest to improve the evaluation of groundwater representation in continental- to global-scale models. *Geoscientific Model Development* 14, no. 12: 7545–7571. <https://doi.org/10.5194/gmd-14-7545-2021>
- Gonzalez, R.Q., and J.J. Arsanjani. 2021. Prediction of groundwater level variations in a changing climate: A Danish case study. *ISPRS International Journal of Geo-Information* 10, no. 11: 792. <https://doi.org/10.3390/ijgi10110792>
- Hartick, C., C. Furusho-Percot, M.P. Clark, and S. Kollet. 2022. An interannual drought feedback loop affects the surface energy balance and cloud properties. *Geophysical Research Letters* 49, no. 22: e2022GL100924. <https://doi.org/10.1029/2022GL100924>
- Koch, J., H. Berger, H.J. Henriksen, and T.O. Sonnenborg. 2019. Modelling of the shallow water table at high spatial resolution using random forests. *Hydrology and Earth System Sciences* 23, no. 11: 4603–4619. <https://doi.org/10.5194/hess-23-4603-2019>
- Koch, J., J. Gotfredsen, R. Schneider, L. Trolldborg, S. Stisen, and H.J. Henriksen. 2021. High resolution water table

- modeling of the shallow groundwater using a knowledge-guided gradient boosting decision tree model. *Frontiers in Water* 3: 701726. <https://doi.org/10.3389/frwa.2021.701726>
- Kollet, S.J., and R.M. Maxwell. 2008. Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research* 44, W02402. <https://doi.org/10.1029/2007WR006004>
- Lendzioch, T., J. Langhammer, L. Vlček, and R. Minařík. 2021. Mapping the groundwater level and soil moisture of a montane peat bog using UAV monitoring and machine learning. *Remote Sensing* 13, no. 5: 907. <https://doi.org/10.3390/rs13050907>
- Liesch, T., and A. Wunsch. 2019. Aquifer responses to long-term climatic periodicities. *Journal of Hydrology* 572: 226–242. <https://doi.org/10.1016/j.jhydrol.2019.02.060>
- Ma, Y., C. Montzka, B. Bayat, and S. Kollet. 2021a. Using long short-term memory networks to connect water table depth anomalies to precipitation anomalies over Europe. *Hydrology and Earth System Sciences* 25, no. 6: 3555–3575. <https://doi.org/10.5194/hess-25-3555-2021>
- Ma, Y., C. Montzka, B. Bayat, and S. Kollet. 2021b. An indirect approach based on long short-term memory networks to estimate groundwater table depth anomalies across Europe with an application for drought analysis. *Frontiers in Water* 3, 723548. <https://doi.org/10.3389/frwa.2021.723548>
- Ma, Y., C. Montzka, B.S. Naz, and S. Kollet. 2022. Advancing AI-based pan-European groundwater monitoring. *Environmental Research Letters* 17, no. 11: 114037. <https://doi.org/10.1088/1748-9326/ac9c1e>
- Maxwell, R.M., F.K. Chow, and S.J. Kollet. 2007. The groundwater–land–surface–atmosphere connection: Soil moisture effects on the atmospheric boundary layer in fully-coupled simulations. *Advances in Water Resources* 30, no. 12: 2447–2466. <https://doi.org/10.1016/j.advwatres.2007.05.018>
- Maxwell, R.M., and L.E. Condon. 2016. Connections between groundwater flow and transpiration partitioning. *Science* 353, no. 6297: 377–380. <https://doi.org/10.1126/science.aaf7891>
- Maxwell, R.M., L.E. Condon, and S.J. Kollet. 2015. A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geoscientific Model Development* 8, no. 3: 923–937. <https://doi.org/10.5194/gmd-8-923-2015>
- Maxwell, R.M., J.K. Lundquist, J.D. Mirocha, S.G. Smith, C.S. Woodward, and A.F.B. Tompson. 2011. Development of a coupled groundwater–atmosphere model. *Monthly Weather Review* 139, no. 1: 96–116. <https://doi.org/10.1175/2010MWR3392.1>
- Meinshausen, N. 2006. Quantile regression forests. *Journal of Machine Learning Research* 7: 983–999.
- Müller, A.C., and S. Guido. 2017. Ensembles of decision trees. In *Introduction to Machine Learning with Python: A guide for data scientists*, 1st ed., 85–94. Sebastopol: O'Reilly Media, Inc.
- O'Neill, M.M.F., D.T. Tijerina, L.E. Condon, and R.M. Maxwell. 2021. Assessment of the ParFlow–CLM CONUS 1.0 integrated hydrologic model: Evaluation of hyper-resolution water balance components across the contiguous United States. *Geoscientific Model Development* 14, no. 12: 7223–7254. <https://doi.org/10.5194/gmd-14-7223-2021>
- Osman, A.I.A., A.N. Ahmed, Y.F. Huang, P. Kumar, A.H. Birima, M. Sherif, A. Sefelnasr, A.A. Ebraheem, and A. El-Shafie. 2022. Past, present and perspective methodology for groundwater modeling-based machine learning approaches. *Archives of Computational Methods in Engineering* 29: 3843–3859. <https://doi.org/10.1007/s11831-022-09715-w>
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12: 2825–2830. <https://doi.org/10.48550/arXiv.1201.0490>
- Rajaei, T., H. Ebrahimi, and V. Nourani. 2019. A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology* 572: 336–351. <https://doi.org/10.1016/j.jhydrol.2018.12.037>
- Reinecke, R., A. Wachholz, S. Mehl, L. Foglia, C. Niemann, and P. Döll. 2020. Importance of spatial resolution in global groundwater modeling. *Groundwater* 58, no. 3: 363–376. <https://doi.org/10.1111/gwat.12996>
- Rust, W., I. Holman, R. Corstanje, J. Bloomfield, and M. Cuthbert. 2018. A conceptual model for climatic teleconnection signal control on groundwater variability in Europe. *Earth-Science Reviews* 177: 164–174. <https://doi.org/10.1016/j.earscirev.2017.09.017>
- Ryken, A.C., D. Gochis, and R.M. Maxwell. 2022. Unraveling groundwater contributions to evapotranspiration and constraining water fluxes in a high-elevation catchment. *Hydrological Processes* 36, no. 1: e14449. <https://doi.org/10.1002/hyp.14449>
- Schneider, R., J. Koch, L. Trolborg, H.J. Henriksen, and S. Stisen. 2022. Machine-learning-based downscaling of modelled climate change impacts on groundwater table depth. *Hydrology and Earth System Sciences* 26, no. 22: 5859–5877. <https://doi.org/10.5194/hess-26-5859-2022>
- Sun, A.Y. 2013. Predicting groundwater level changes using GRACE data. *Water Resources Research* 49, no. 9: 5900–5912. <https://doi.org/10.1002/wrcr.20421>
- Tijerina-Kreuzer, D., J.S. Swilley, H.V. Tran, J. Zhang, B. West, C. Yang, L.E. Condon, & R.M. Maxwell (2023). Continental scale hydrostratigraphy: basin-scale testing of alternative data-driven approaches. *Groundwater*. Portico. <https://doi.org/10.1111/gwat.13357>
- Tran, H., J. Zhang, J. Cohard, L.E. Condon, and R.M. Maxwell. 2020. Simulating groundwater–streamflow connections in the Upper Colorado River Basin. *Groundwater* 58, no. 3: 392–405. <https://doi.org/10.1111/gwat.13000>
- van der Gun, J. 2020. Groundwater resources sustainability. In *Global Groundwater*, 1st ed., ed. A. Mukherjee, B. Scanlon, A. Aureli, S. Langan, H. Guo, and A. McKenzie, 331–345. Amsterdam, Netherlands: Elsevier. <https://doi.org/10.1016/b978-0-12-818172-0.00024-4>
- Vu, M.T., A. Jardani, N. Massei, and M. Fournier. 2021. Reconstruction of missing groundwater level data by using long short-term memory (LSTM) deep neural network. *Journal of Hydrology* 597: 125776. <https://doi.org/10.1016/j.jhydrol.2020.125776>
- Wunsch, A., T. Liesch, and S. Broda. 2022. Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. *Nature Communications* 13, no. 1: 1221. <https://doi.org/10.1038/s41467-022-28770-2>
- Xia, Y., M.T. Hobbins, Q. Mu, and M.B. Ek. 2015. Evaluation of NLDAS-2 evapotranspiration against tower flux site observations. *Hydrological Processes* 29, no. 7: 1757–1771. <https://doi.org/10.1002/hyp.10299>
- York, J.P., M. Person, W.J. Gutowski, and T.C. Winter. 2002. Putting aquifers into atmospheric simulation models: An example from the Mill Creek watershed, northeastern Kansas. *Advances in Water Resources* 25, no. 2: 221–238. [https://doi.org/10.1016/S0309-1708\(01\)00021-5](https://doi.org/10.1016/S0309-1708(01)00021-5)
- Zhang, J., Y. Zhu, X. Zhang, M. Ye, and J. Yang. 2018. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology* 561: 918–929. <https://doi.org/10.1016/j.jhydrol.2018.04.065>