JOURNAL OF
EMERGING INVESTIGATORS

Article

Groundwater prediction using artificial intelligence: Case study for Texas aquifers

Vedant Sharma[1], Shilpi Sharma[2]
[1] Awty International School, Houston, Texas
[2] London Stock Exchange Group (LSEG), Houston, Texas

SUMMARY

Groundwater resources are highly dependent upon recharge from precipitation; hence, variability in precipitation patterns is important for sustainable groundwater management. Climate change is not only increasing the frequency and intensity of extreme hydro-meteorological events like hurricanes and cyclones, but is also having an even more complex impact on other variables such as precipitation, which can be measured by the Palmer Modified Drought Index (PMDI). Therefore, it is imperative to understand how climate change impacts groundwater management. This research has created a preliminary model to predict the future groundwater levels based on the critical causality and regression analysis using artificial intelligence. In this study, we hypothesized that tree-based automated artificial intelligence models, would perform best in predicting future groundwater levels. Unlike other machine learning models, tree-based models can accommodate correlated datasets like climate variables and changing demographics. We acquired water aquifer data from the United States Geological Survey (USGS) for Texas aquifers, climate data from the National Oceanic and Atmospheric Administration (NOAA), and population data from the US census to train artificial intelligence models using MATLAB. This research project identified key trends that groundwater levels were down 50% on a normalized basis for the selected dataset and established correlation of groundwater changes, weather related changes (temperature changes, precipitation changes and drought indices) and demographic changes. Understanding these key trends allowed us to select the right predictors for the AI model. Tree-based AI models predicted the future groundwater levels with the most accuracy and least root mean square error (RMSE) as compared to other AI models like linear regression, neural network, support vector machines.

INTRODUCTION

Scarcity of freshwater is an increasingly critical public health problem in many parts of the world. Inadequate access to safe freshwater contributes to waterborne disease, malnutrition, poverty, economic and political instability, and conflict between countries or groups within countries. Approximately 97.5% of all water is either salt water or polluted water unsafe for human consumption. Of the remaining 2.5%, nearly 70% is frozen in glaciers and polar ice caps. Less than 0.01% of all water worldwide is directly available for human consumption and resides in lakes, rivers, and reservoirs (1).

Surface water is above the ground and participates in water cycle movement to and from the Earth's surface. Water that seeps into the ground is called groundwater. Groundwater and surface water are constantly interacting with each other as the weather changes over the year from summer to winter. Groundwater resources provide a buffer against depletion of freshwater resources due to climate variability. Groundwater is essential to the survival of 49% of the global population, who solely depend on it to meet their basic daily needs (2). Groundwater is also a critical resource to produce food and accounts for around 25% of all water withdrawn for irrigation, serving 38% of the world's irrigated land (3). The rate of global aggregated groundwater storage depletion is considerable and more prominent in areas with explosive population and economic growth like China and India, increasing from 158km3/year in 1950 to 959 km^3/year in 2017 (4).

Access to water and sanitation is a fundamental human right recognized by United Nations. However, according to the World Bank, water scarcity affects 40% of the global population (5). According to the research done by the World Bank, because of world population growth, there will be a 40% shortfall between demand and available supply of water by 2030, which will require a 15% increase in water withdrawals from our scarce ground water resources (6).

A geological study showed that the Houston area has experienced one of the fastest rates of subsidence, a gradual sinking of the land. The study showed that subsidence was linked to excessive groundwater usage (7). Similar research in China links groundwater depletion to land subsidence (8).

Aquifers are a body of permeable rocks which can store and transmit groundwater. Groundwater is a term which is used to describe the water which has infiltrated the soil and collected in the rocks below the surface. Groundwater additions and subtractions are best explained by The University of New Hampshire's (UNH) Water Balance Model, which simulates the global water cycle and includes the water extraction for human consumption (9). The model shows the major fluxes and storage by overlaying the water demands (subtractions) attributed to irrigation, domestic, industrial and livestock needs, with the water inflows from reservoirs via glaciers (additions). It also factors recharging the groundwater aquifers, while understanding the impact of precipitation (addition), evapotranspiration (subtraction) on the overall groundwater balance and flow. The UNH's water balance model can be used to create a specific map of an aquifer by interlinking the flows attributed to surface runoff and base-flow to fluid river water and by factoring in man-made dams and reservoirs. Based on the historical

https://doi.org/10.59720/23-200

research conducted, groundwater recharge (additions) and discharge (subtractions) are impacted by environmental (e.g., precipitation, ambient temperature), geographical (e.g., density of streams, sub surface geology, depth to water path (e.g., surface infiltration, surface runoff, spillover to ocean), and human variables (e.g., population, intensity of land use) (10).

As climate change drives increasing variability in precipitation patterns, it is imperative to understand how groundwater resources get impacted due to delay in the recharge from precipitation and extended draught conditions (11). The depth of water aquifers indicates the level at which groundwater is available from the surface. As our consumption of groundwater continues to increase with population, we have depleted our water resources and need to dig deeper to find scarce water resources (12). Impacts of changing climate on variables like precipitation, Palmer Hydrological Drought Index (PHDI), and temperature of groundwater is even more complex and difficult to establish in a static model. Climate change and groundwater research has been conducted to identify the impacts of climate change on specific aquifers, but the use of artificial intelligence (AI) to create prediction models is still in its infancy (13). In many AI models, researchers studied water resources management using neural networks, but in some cases machine learning models were used to create water variables. The initial models' accuracy was then improved by applying complex optimization algorithms. The reasons may be that sufficient historical data are required to analyze the characteristics of climate change. Moreover, sometimes the most important parameters like PHDI are not readily available in a standardized format, across geographies and long timespan, to be consumed in the data intensive AI models. To date, the driving forces that cause extreme hydro-meteorological events and gradual changes due to demographics have not been combined in one model (14).

It is necessary to assess the impact of future climate changes coupled with different groundwater abstraction resulting from human activities and water diversion scenarios (15) (16). Key limitations of existing research are that it is in silos (of geology or weather-related or demographics) and current prediction models do not factor all variables in an interdependent causation; additionally, the models are not scalable across areas of study. These models required complex numerical modeling by understanding the impact of geology on porous media and is heavily dependent on the unique understanding of the data related to local topography. However, AI algorithms have advanced dramatically, allowing us to create models linking complex interdependent variables by processing enormous non-structured and multidimensional data and applying optimization techniques to create a new understanding of the complex interlinkages (17). We need to research the water levels across the aquifers to best understand the critical parameters other than precipitation which impact groundwater levels; identifying key parameters will allow us to develop a model to predict the future groundwater levels.

The first step of our study was to identify and select key parameters that impact groundwater level. We collected, cleansed and normalized data collected from United States Geological Survey (USGS), National Oceanic Atmospheric Administration (NOAA) and US Census to select the right predictors for groundwater depth. After the right parameters were selected, the AI model based on the selected dataset helped us establish the linkage between demographics and climate variables to understand the long-term impact on groundwater. This study allows us to predict the impact of future climate changes on groundwater resources and make our groundwater resources more resilient by identifying the right water infrastructure.

In this study, we hypothesized that a tree-based automated AI model would outperform all other machine learning models we tested and reveal the impact of climate and demographic variables to groundwater depth. This was because the climate variables are interdependent on each other, and demographics are impacting the climate directly. We also conducted detailed literature review of the comparative study of artificial intelligence models and statistical methods for groundwater level prediction (18). Our research revealed that the relevant AI models for groundwater prediction can be categorized into neural networks, regression, scaler vector machines, and gaussian process regression. We observed the key parameters and process for defining the model comparison parameters, by inputting the timeseries dataset and using statistical methods to identify input and output parameters. We were also able to understand the best practices in terms of defining training and testing data and compare various models like multiple linear regression, neural networks and extreme learning machines, by understanding the key statistical indices like root mean square error (RMSE).

Our next step was to analyze data to select prediction model parameters which were fed in a regression learner application of MATLAB to create initial models. We trained different model types in MATLAB with a subset of data and were able to confirm that fine tree model has the least RMSE of all the selected models. Our findings also correctly established the direct correlation with population, temperature, and inverse correction with temperature and Palmer hydrological drought index (PHDI).

RESULTS

To test our hypothesis, we collected Texas state data from aquifer database from USGS, climate database from NOAA and population data from US Census, normalized it for ingestion in MATLAB to create a model and then tested the model in regression learner to test our hypothesis **(Figure 1)**. We were able to normalize and harmonize the data and assign the appropriate climate and demographic data to the respective aquifer. Our analysis allowed us to identify key trends and correlations amongst the predictors and groundwater depth. We then imported the harmonized data into MATLAB in a matrix. Our detailed analysis in MATLAB visualization using graphs and causation analysis allowed us to finalize the predictors which had a high degree of causation. We were then able to import this matrix into Regression Learner module of MATLAB to test our hypothesis by testing a subset of the dataset with the trained model. We were able to reduce the error in predictions by refining different models by iterating on the predictors and adjusted the parameters.

Based on the USGS dataset that was selected for Texas aquifers, groundwater-depth continues to increase, highlighting the depleting groundwater resources. Normalized groundwater-depth from has increased from 104 feet in 1940 to 181 feet in 2022, on a normalized basis **(Figure 2A)**. The raw data plotted in MATLAB which shows the USGS data

JOURNAL OF
EMERGING INVESTIGATORS

https://doi.org/10.59720/23-200

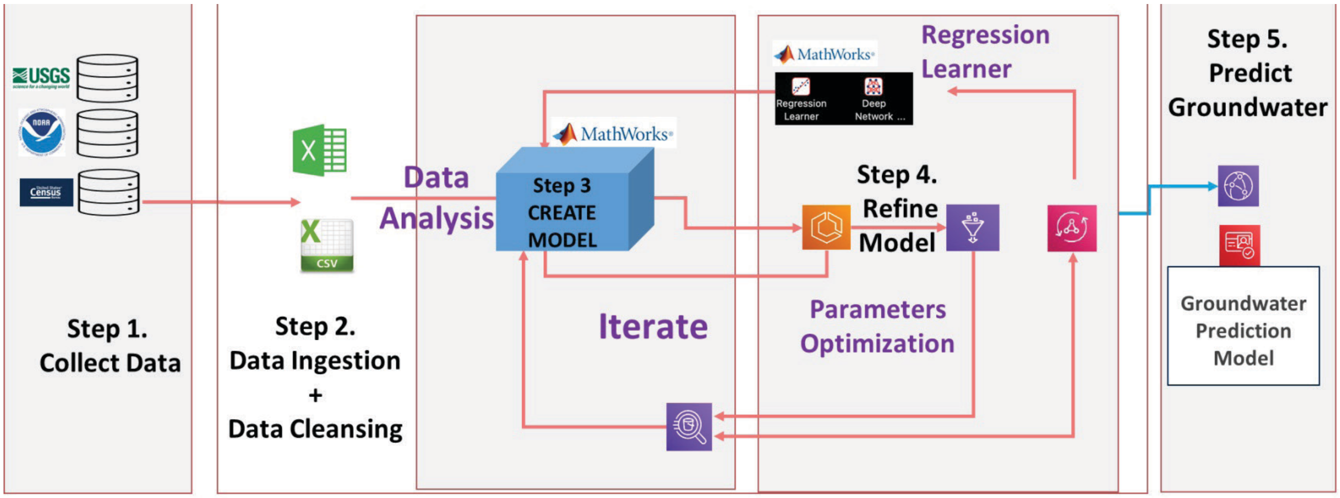

Figure 1: Research methodology. Step by step guide to creating groundwater prediction model. 1. collect data; 2.data cleansing; 3.create model 4.refine model 5. predict groundwater values.

of the 171 Aquifers and how raw groundwater depth has also continued to increase by 74% over the last 80 years **(Figure 2B)**. The trends by month and year of groundwater-depth also show similar trend whereby the groundwater depth is continuously increasing across the months over the last 80 years **(Figure 2C)**. Average groundwater-depth (actual reading) is a close approximation of the trend related to the average normalized groundwater-depth, based on the analysis done in excel.

We used linear-regression model to determine correlations between different aspects of demographic data and climate data, including temperature, precipitation, and Palmer Hydrological Drought Index (PHDI) **(Figure 3)**. Temperature for the State of Texas did not show significant variation by month **(Figure 3A)**. Based on the precipitation data, the trend indicated that the precipitation average had increased by 0.42 inches from 2000 onwards, as compared to the same timeframe prior, with unusual spikes driven by destructive hurricanes like Harvey **(Figure 3B)**. The PHDI and PMDI had statistically significant material differences in the last 30 years as the drought like conditions have increased in the recent years **(Figure 3C and Figure 3G)**. Temperature anomaly showed an increasing trend in the last 20 years **(Figure 3D)**. Precipitation anomaly increased over the observation period

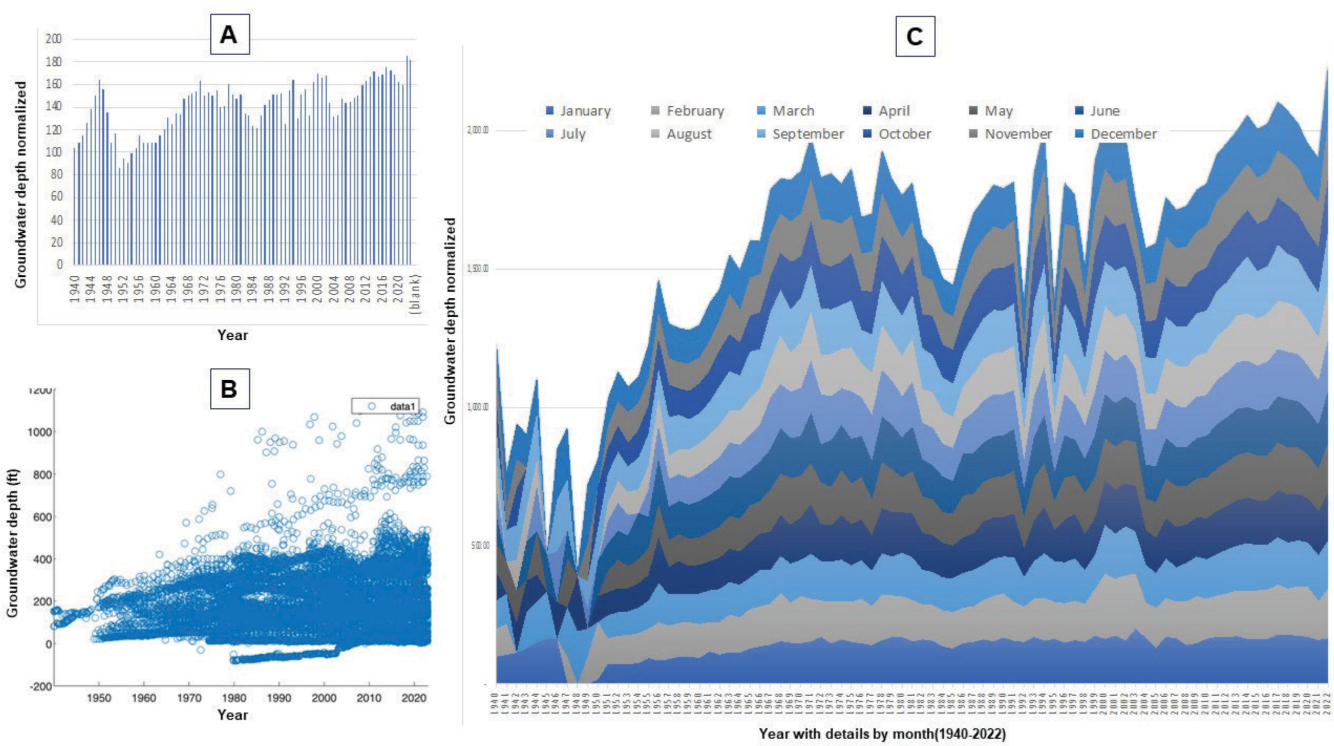

Figure 2: Ground water depth from 1940 to 2022. A) normalized in Excel; B) raw data in MATLAB; C) normalized by month and year in excel.

https://doi.org/10.59720/23-200

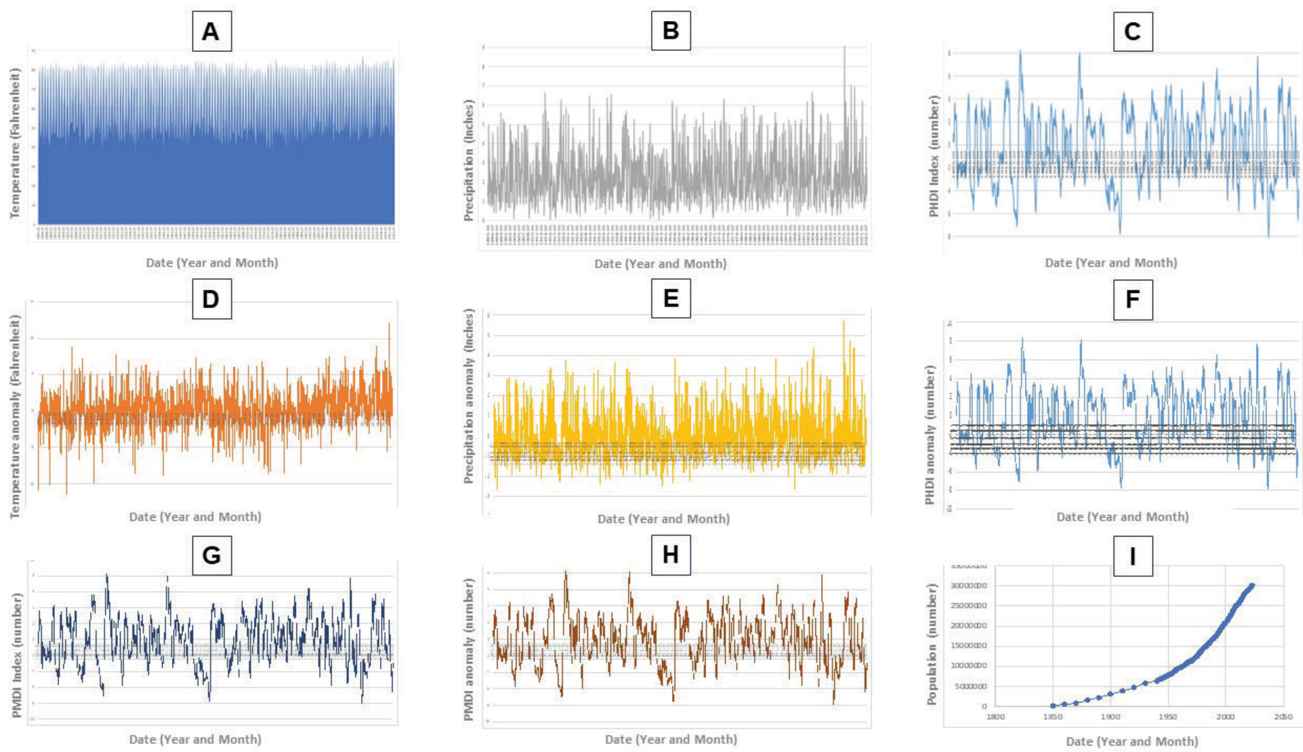

Figure 3: Climate parameters (Temperature, Precipitation, PHDI) and climate anomalies and population by year and month (1850-2022). A) Temperature, B) Precipitation, C) PHDI, D) Temperature anomaly, E) Precipitation anomaly, F) PMDI anomaly, G) PMDI, 3H PMDI anomaly, I) Population.

(Figure 3E). PHDI anomaly and PMDI anomaly show increasingly positive values with more positive values then negative values in recent years **(Figure 3F and Figure 3H)**. Population data population of Texas shows an exponential increase from approx. 5 million in 1930s to approximately 30 million in the last census **(Figure 3I)**.

We harmonized the data and selected the key variables to create an array in MATLAB for the Texas aquifers to predict the "groundwater depth" in the future. Thirty percent of the data was selected as the training data for the models and then the actual results were plotted against the models. Over 26 different model types were trained in MATLAB across linear regression, regression trees, support vector machines (SVM), gaussian process regression, kernel approximation, ensembles of trees and neural network methods. After multiple iterations, the fine tree regression model allowed us to predict the groundwater data with the least Root Mean Square Error (RMSE). The fine-tree model also has the best value of R-squared, which denotes how well the data fit the selected model.

After the fine tree mode was selected with the right parametrization selection, research results validated the dependencies of the demographic (population) and climate change (ambient temperature, precipitation, PHDI, PMDI, temperature anomaly, precipitation anomaly) data and established the correlation by analyzing partial dependence plots with respect to the predicted variable of "groundwater-depth", modeled as "DepthToWaterBelowLandSurfaceInFT" in MATLAB. Any increase in the groundwater depth below land surface denotes that the groundwater resources had depleted. It was also important to understand the linkage of anomaly variables as they predict the correlation with extreme hydrological events and allow us to do various scenario modeling in the future.

Partial dependence plots (PDPs) in MATLAB establish causation and visualize the marginal effect of each predictor on the predicted response of a trained regression model. We trained the fine tree model and realized that population, temperature, and precipitation anomaly have a direct correlation with increasing groundwater depth, while PHDI index have inverse correlations with groundwater depth **(Figure 4)**.

There is a direct correlation of population with the groundwater depth. According to the PDP the impact of the population is the most significant, accounting for most of the groundwater depth increase **(Figure 4A)**. This is consistent with other studies which correlate an exponential increase in population in China and India with the most depletion of groundwater resources (19) (20). Population is also a surrogate indirectly correlated to the amount of vegetation and deforestation which leads to less area available to absorb the precipitation and therefore increasing the groundwater depth.

Higher ambient temperature is directly correlated to increasing groundwater depth **(Figure 4B)**. Temperatures above 50 degrees Fahrenheit (1st Stair) and 63 degrees Fahrenheit (2nd Stair) average cause a steep jump and then it increases gradually after 70-degrees Fahrenheit. As the temperature goes even higher in the future because of climate change (not shown in the figure), based on PDP predictions, the impact of hypothetical 3rd (80-90 degrees Fahrenheit) and 4th step (100-110 degrees Fahrenheit) jumps will lead to

https://doi.org/10.59720/23-200

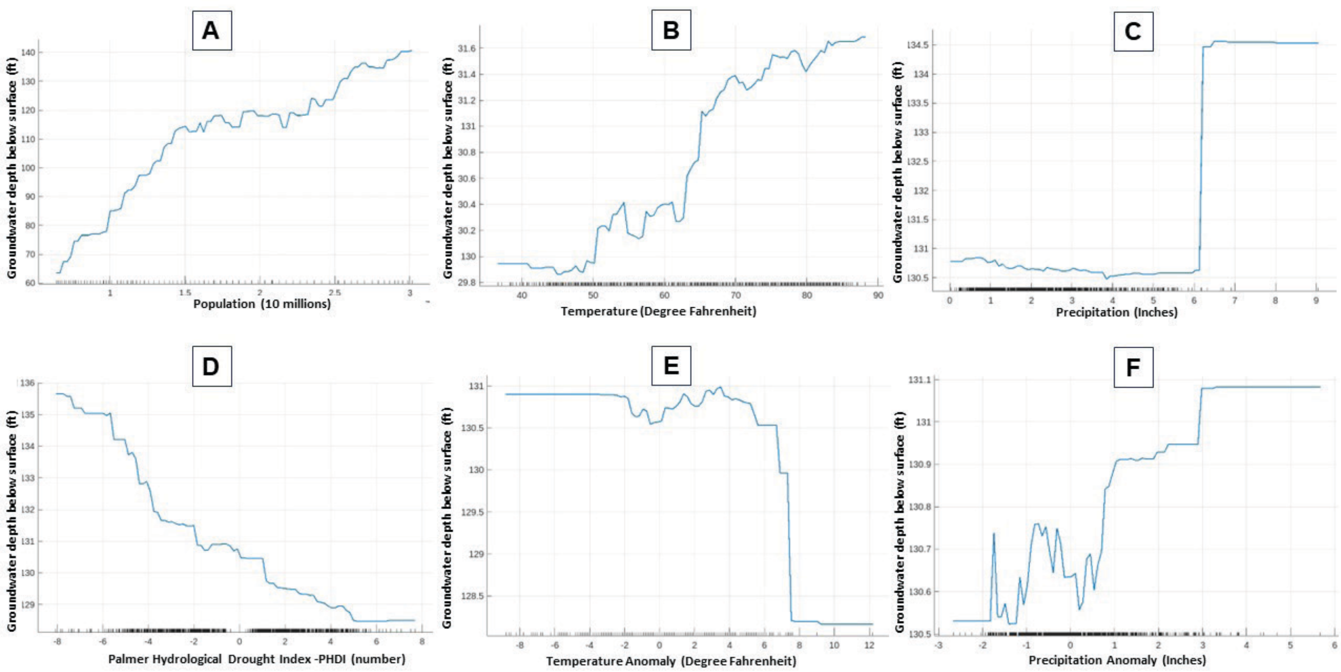

Figure 4: Partial Dependent plots (PDP) correlation to the groundwater depth. A) Population; B) Temperature; C) Precipitation; D) PHDI; E) Temperature Anomaly; F) Precipitation Anomaly.

significantly more impact to the groundwater depth.

Based on past studies, precipitation is assumed to be inversely correlated to groundwater depth, but according to the PDPs in our regression model, precipitation does not have an impact at all **(Figure 4C)**. However, the precipitation value after 6 inches indicates a direct correlation with groundwater-depth. It can only be explained by a combination of other factors like run-offs and the increase in population and loss of farmland. More datapoints need to be generated across the US to confirm this finding.

Palmer Hydrological Drought Index (PHDI) allows us to calculate when a drought will end based on precipitation needed by using a ratio of moisture received to moisture required to end a drought. This is one of the most common parameters which is used to assess the impact on groundwater resources on longer timescales. PHDI values less than -2 indicate a drought. PHDI below -2 (conditions of draught) has a direct correlation with groundwater depth (i.e., as the drought conditions worsen and go below -2, the groundwater depth goes down) at a consistent pace with a flattening (reduced rate) after PHDI reaches -6 **(Figure 4D)**.

Temporal climate variability (denoted by the variables temperature anomaly and precipitation anomaly), especially variability in precipitation, can have substantial effects on recharge and groundwater levels. The increased variability in precipitation and temperature that is predicted under many climate-change scenarios will likely have variable effects on different aquifers and different locations within an aquifer depending on spatial variability in hydraulic properties and distance from the recharge areas. According to the National Oceanic and Atmospheric Administration's (NOAA) national weather service, temperature or precipitation anomaly means a departure from a reference value or long-term average, often over the thirty-year period for that region. A positive anomaly indicates that the observed temperature was warmer than the reference value, while a negative anomaly indicates that the observed temperature was cooler than the reference value. According to the regression model, the groundwater variation was not directly correlated to temperature anomalies in the trained regression model **(Figure 4E)**. The depth ranged from 128 feet to 131 feet across the wide range of temperature. This indicates that wider swings in temperature do not materially change the ground water depth. This finding requires more research and a wider data set to validate the results. Precipitation Anomaly between -1 and +2 have the most impact on the model leading to a steep jump **(Figure 4F)**. More data needs to be included to understand if anomalies increasing significantly will generate new steps which will deplete the groundwater resources further.

In MATLAB the models are evaluated based on the RMSE (Root Mean Square Error), R-Squared (Coefficient of determination), MSE (Mean Square Error) and MAE (Mean Absolute Error) in addition to the elegance parameters like prediction speed, training time and model size. In addition, you can also evaluate models using residuals plots which show the comparison of the residuals based on the validation data set after training the model and are ideally symmetrically distributed around zero. All the models selected for comparison used cross-validation schema with 5 cross-validation folds, which splits the data into 5 distinct sets. The validation schema trains a model using the training-fold observations, which are not in the validation fold and then assesses the model performance using validation-fold data for training. 10% of the data was set aside to test the created model. The

https://doi.org/10.59720/23-200

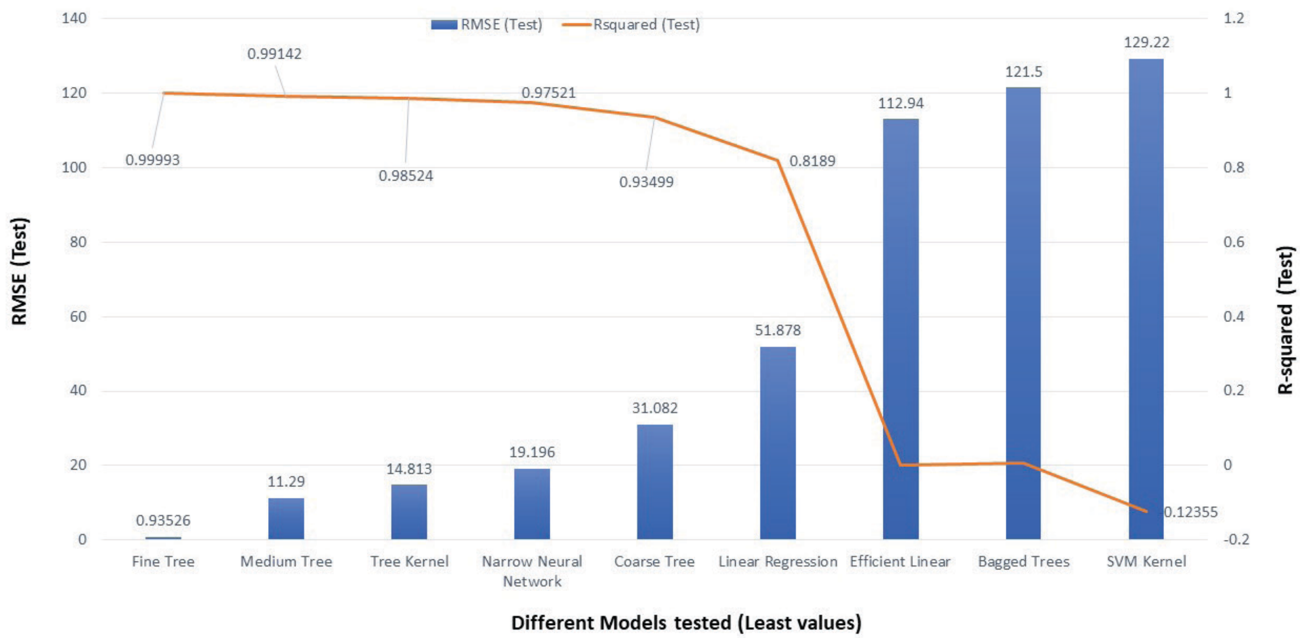

Figure 5: Best Root Mean Square Error (RMSE) (less is better) and R-square values (closer to 1 is better) of all the different models trained. Fine Tree model has the best fit with the least RMSE and best R-square values.

key features which were found to be relevant in the data set as per the MRMR algorithm in decreasing order of importance were precipitation, population, Palmer Hydrological Draught Index (PHDI), PHDI anomaly, Palmer Modified Draught Index (PMDI) anomaly, temperature, PMDI, and temperature anomaly from the selected parameters in the model. The model with the least RMSE and most R-square (0.99) was the fine tree model **(Figure 5)**. The neural network model was the closest, but the ensemble and kernel models were found to be ten times worse (based on RMSE) compared to the fine tree model. The fine tree predictor model correctly established the direct correlation with population, temperature, and inverse correction with temperature and PHDI anomaly.

DISCUSSION

A key aspect of this research was to develop an approach that uses the latest advancements in artificial intelligence whereby data creates the model versus creating a model based on the equations. Due to advancements in machine learning, we can leverage existing data sets and support adaptation by adding new variables and parameters and create a modular approach to simulate different scenarios. Here, we hypothesized that tree-based automated artificial intelligence models, would perform best in predicting future groundwater levels.

Our prediction model used large amounts of data to predict groundwater levels based on actual measurements at aquifers by correlating different parameters. The large dataset allowed us to create a groundwater model which can be constantly improved as we gather more data, and we can increase the confidence of our prediction by expanding the geographical footprint beyond Texas. Typically, microstudies are conducted when severe drought like conditions persist for a longer time to predict the groundwater levels. However, microstudies use many resources and take years to complete after considering geological factors. The circular nature of geological parameters which impact the groundwater levels makes the studies obsolete very quickly. Also, different climate variables that should be embedded in the model are difficult to consider. The complexity of different predictors and inability to homogenize the data is another problem due to lack of standardization across the globe. There is a tremendous value in identifying unique parameters which are consistently identified as predictors in the numerous case studies done across the world to predict groundwater levels especially in countries which are grappling with water scarcity like China, India, and Iran (21).

One key limitation of the model developed from our study is that population of Texas was used as a surrogate for the Texas aquifers. Texas population concentration in metropolitan areas like Houston and Dallas will lead to a variation in population by county, which is currently not factored in the model. We can improve the model accuracy further by using the county population in which aquifer is located. For some years, population data was not available. So, the population was projected between the two census years when it was not available. For example, the population for years 1952 to 1959 was projected based on the starting population in 1950 and the ending population in 1960. This smoothing using linear regression could have had an under or overestimation impact on the model depending on the variable rate of increase. In the current model, yearly population was assumed for all the months. Monthly increases and decreases of population will also have an impact on the model accuracy.

Another limitation of the prediction model was that different aquifers were monitored from groundwater-depth perspective on a different time horizons. For example, the number of readings increased from 15 in 1940 to 9803 in 2021. So, it was imperative for us to normalize the water depth to understand the trends and averages across the different aquifers, due to lack of complete data availability across the aquifers for all years. The normalized data allowed us to visualize the overall trend across the years and select the predictor variables. It is important to note that unique weather across all the Aquifers would have an impact on the localized water aquifers, which was not factored into the overall model. The key challenge with expanding this research is to factor in the parameters which impact the groundwater value other than the core attributes like hydraulic conductivity, basin type, infiltration rate, basin characterization model, and runoff estimation. Further investigation and more data are necessary to continue to refine the model.

In the future, we can use this model to develop a conceptual predictive model for groundwater aquifers across the world by adding more data and more variables which complete the model (like catchment, impact of floods, discharge data) and predict subsidence related to groundwater reduction or predict key metrics related to food sustenance which are dependent on aquifers for irrigation. There is also a need to expand the dataset and remove any biases or hallucinations in the dataset. This model can be expanded by including the following data sources in the future: International Water Management Institute (IWMI), IHE Delft Institute of Water Education, World Water Assessment Program (WWAP), and World Water Development Report II (WWDRII) database.

Our fine-tree AI model, with the least RMSE, identified the correlation of the variables and allows us to predict the change in groundwater resources in the future. We can use this model as a tool to preserve our groundwater resources, by identifying vulnerabilities and conducting risk-assessment ahead of time. We can also conduct scenario modeling based on different weather and climate predictions, including the impact of future drought scenarios.

MATERIALS AND METHODS

Data Collection and Cleansing

The first step in model creation was to collect different datasets and harmonize them into one matrix across the selected parameters **(Figure 1)**. We retrieved **groundwater dataset** from USGS, which collects dataset monthly for all the US Aquifers including groundwater depth, aquifer type and geography details. The focus of our study was the state of Texas to understand the patterns. Finally, after cleansing and randomizing the aquifers across the state, 171 Texas Aquifers were included in the study leading to a 142773 X 16 matrix of monthly data points from 1940 to 2022. **Climate dataset** was collected from NOAA, which included temperature, precipitation, PMDI, and PHDI data from 1895 to 2022 for Texas. This resulted in a matrix of 1536 X 13. **Population dataset** from the US census was used as a surrogate to determine the human consumption correlation. The US Census data for the state of Texas from 1850 to 2022 were gathered. This ultimately resulted in a 32 X 256 matrix after smoothing over the population changes over the census period when the data was not available. All this data resulted in a 142773 X 25 normalized matrix which had the data of the aquifers enriched by the climate parameters and population data for the year.

Analyze Data to Select Prediction Model Parameters

Excel and MATLAB were used to conduct detailed data analysis to understand the key parameters which should be incorporated as part of the model. Historical data analysis was used followed by bias correction and removal of outliers and statistical analysis of data and identification of patterns to form analysis. Preliminary analysis was done in Excel. After normalizing the raw data, the matrix was imported into MATLAB. The three matrices were merged into MATLAB and the matrix that was used to train the model was 142733 X 25. Since aquifer data had different start dates for measurement, the depth was normalized by assuming the start of the reading as 100. The key values which were observed as having a trend were then analyzed in MATLAB to understand the detailed statistical parameters and identify the key trends.

MATLAB Regression Learner Creates a Model

We used regression learner to import data and then trained the dataset to predict future parameters of the dataset based on “predictors". For training, linear regression, regression trees, support vector machines, gaussian process regression models, kernel approximation regression models, ensembles of trees and neural networks were used. Regression learner was used to train different models and then we tested a portion of the data (30%), which is set aside at the beginning of the test. After selecting the model, we adjusted the features (predictors) iteratively to achieve a better result.

Iterate across Different Prediction Models (Neural Networks, Trees, Regression, Support Vector Machines) and Select the one with the Least Root Mean Square Error (RMSE)

We then trained different models and based on the least RMSE we selected the model which works best. Since the variables that were selected to model the groundwater prediction are correlated to each other, tree regression models were the most optimal as it allows us to predict continuous valued outputs instead of discrete outputs. Medium trees, narrow neural networks, boosted trees, bagged trees and fine-tree models were iteratively tested after optimizing the parameters and selecting the right number of predictors **(Figure 6)**. Predictions using fine-tree AI models achieved the most symmetry against the predicted response **(Figure 6E)**. Details of the model 7 which had the least RMSE of all the models were analyzed to understand the predictor’s accuracy (e.g. temperature, population) based on the selected fine-tree models with the test data set aside at the beginning of training the model **(Figure 7)**. Also, the residuals plot of the model 7 against various parameters which are symmetrical around 0, which is a sign of a good, trained model.

Export the Model as a Function in MATLAB

After training the model, we exported the model as a function (GW_Vedant), which has all the coding related to the training inbuilt in the model (Figure 8). The details of training data used, and the exact model details are also exported in the function. After exporting the model to the workspace from Regression Learner, or run the code generated from the app, we were able to create a structure that can be used to make predictions using new data. The structure contains a model object and a function for prediction. This exported model allows us to make predictions by using new data as an input to predict the future groundwater depth.

https://doi.org/10.59720/23-200

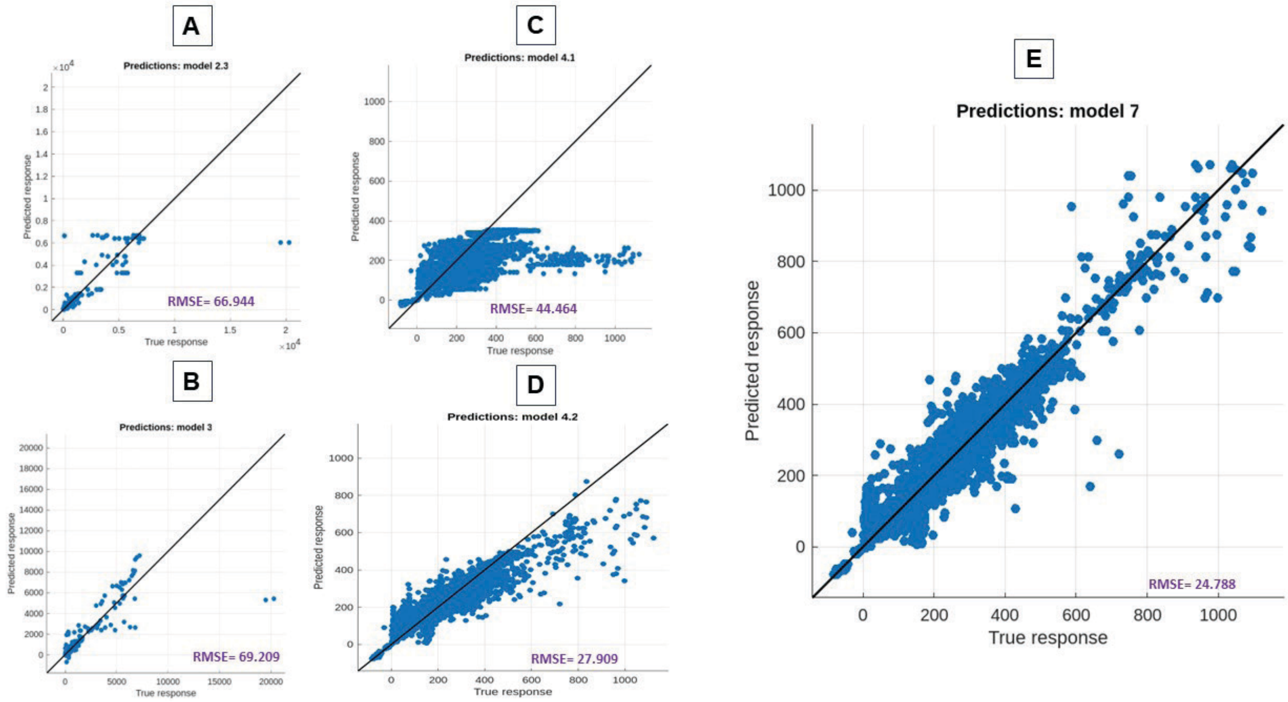

Figure 6: Prediction versus Actual Plots for different models (with RMSE): Perfect prediction is on the line and closest predictions are closest to the line. A) Medium Trees, B) Narrow Neural Network, C) Boosted Trees D) Bagged Trees, E) Fine Tree Model

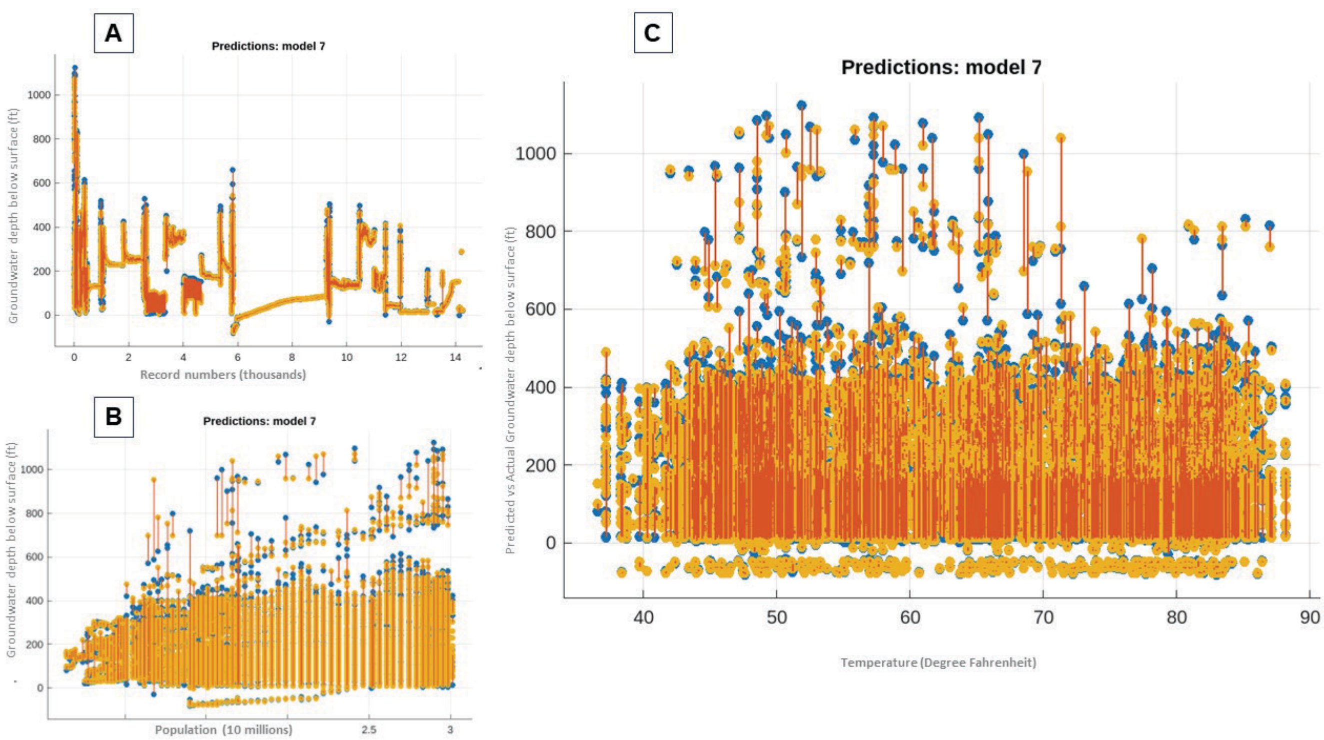

Figure 7: Predictions mapped against the trained model. A) Record numbers against groundwater depth, B) Population prediction vs actual values against groundwater depth, C) Temperature prediction vs actuals against groundwater depth. Predicted values are shown in orange while Actual values are shown in blue dots. Errors are shown in red.

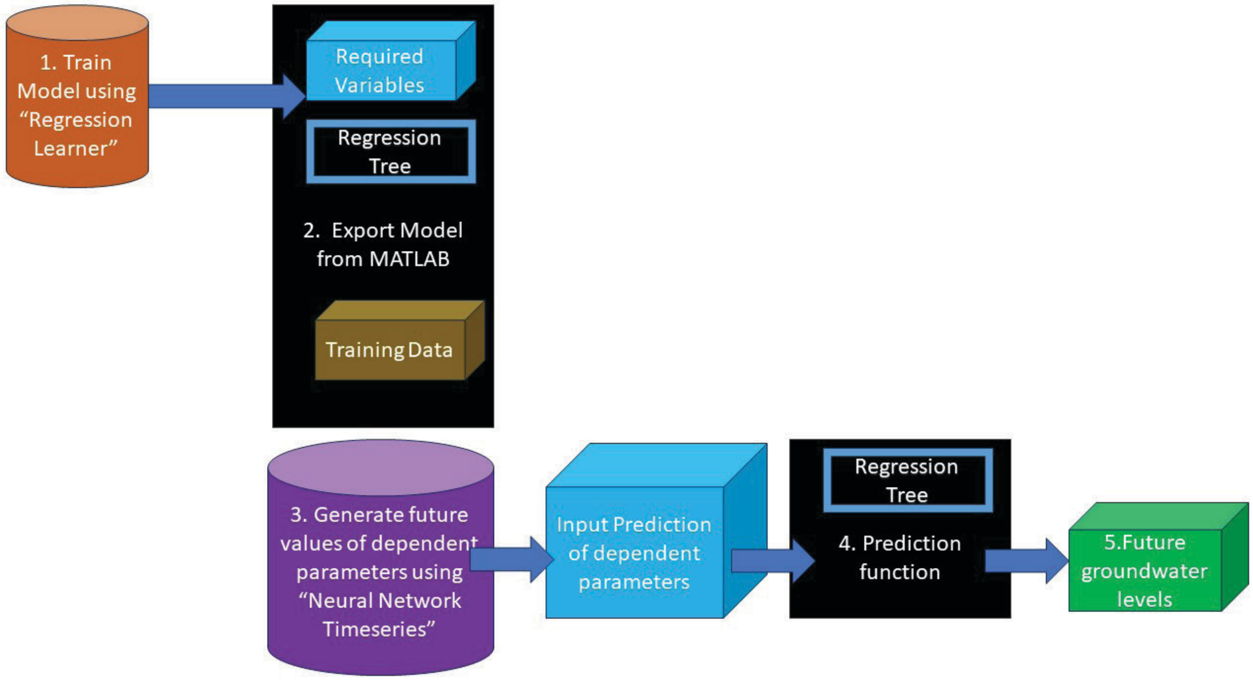

Figure 8: Exporting model with details of the training data and predicting with new parameters.

Use Prediction Function to Predict the Groundwater Depth

We were able to test the AI model by predicting the aquifer's depth for "City of Douglasville" by inputting the mandatory predictor parameters as input variables **(Figure 8)**. "Neural Network" Timeseries in MATLAB was also used to predict future values based on past values. In addition, we were also able to create "What If" scenarios and get different groundwater depths depending on the simulations of predictor parameters like extreme heat and cold.

Received: July 9, 2023
Accepted: December 19, 2023
Published: April 19, 2024

REFERENCES

1. Levy, Barry S, and Victor W Sidel. "Water rights and water fights: preventing and resolving conflicts before they boil over." *American Journal of Public Health* vol. 101,5 (2011): 778-80. https://doi.org/10.2105/AJPH.2010.194670.
2. UNESCO World Water Assessment Programme. "The United Nations World Water Developmental Report 2022: groundwater: making the invisible visible" *unesdoc.* unesco.org/ark:/48223/pf0000380721
3. Margat, Jean & van der Gun, Jac. (2013). *Groundwater around the World: A Geographic Synopsis*. https://doi.org/10.1201/b13977.
4. Margat, Jean & van der Gun, Jac. (2013, fig. 5.4, p. 128). *Groundwater around the World: A Geographic Synopsis*. https://doi.org/10.1201/b13977.
5. World Bank. *Water resources management: context and challenges.* worldbank.org/en/topic/waterresourcesmanagement
6. Rodella, Aude-Sophie; Zaveri, Esha Dilip; Bertone, Francois. *The Hidden Wealth of Nations: The Economics of Groundwater in Times of Climate Change - Executive Summary (English).* Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/099145503202323072/P178601171e7ffac1ea0714b5e187c0122449517b07d
7. Khan, Shuhab D., *et al.* "Surface Deformation Analysis of the Houston Area Using Time Series Interferometry and Emerging Hot Spot Analysis." *Remote Sensing,* vol. 14, no. 15, Aug. 2022, p. 3831. https://doi.org/10.3390/rs14153831.
8. Chai, J. C., *et al.* "Land Subsidence Due to Groundwater Drawdown in Shanghai." *Géotechnique,* vol. 54, no. 2, Mar. 2004, pp. 143–47, https://doi.org/10.1680/geot.2004.54.2.143.
9. Grogan, D. S., *et al.* "Water Balance Model (WBM) v.1.0.0: A Scalable Gridded Global Hydrologic Model with Water-Tracking Functionality." *Geoscientific Model Development,* vol. 15, no. 19, 2022, pp. 7287–323, https://doi.org/10.5194/gmd-15-7287-2022.
10. Hartmann, Andreas. "The Hydrology of Groundwater Systems – From Recharge to Discharge." *Encyclopedia of Inland Waters,* Elsevier, 2022, pp. 324–30. https://doi.org/10.1016/B978-0-12-819166-8.00097-9.
11. Lorenzo-Lacruz, Jorge, *et al.* "Groundwater Level Responses to Precipitation Variability in Mediterranean Insular Aquifers." *Journal of Hydrology,* vol. 552, Sept. 2017, pp. 516–31. https://doi.org/10.1016/j.jhydrol.2017.07.011.
12. Ebi, Kristie L., *et al.* "Extreme Weather and Climate Change: Population Health and Health System Implications." *Annual Review of Public Health,* vol. 42, no. 1, Apr. 2021, pp. 293–315. https://doi.org/10.1146/annurev-publhealth-012420-105026.

https://doi.org/10.59720/23-200

13. Allen, D. M., *et al.* "Groundwater and Climate Change: A Sensitivity Analysis for the Grand Forks Aquifer, Southern British Columbia, Canada." *Hydrogeology Journal,* vol. 12, no. 3, June 2004. https://doi.org/10.1007/s10040-003-0261-9.
14. Pittock, Jamie, *et al.* "Groundwater Management Under Global Change: Sustaining Biodiversity, Energy and Food Supplies." *Integrated Groundwater Management,* edited by Anthony J. Jakeman *et al.,* Springer International Publishing, 2016, pp. 75–96. https://doi.org/10.1007/978-3-319-23576-9_4.
15. Veldkamp, T. I. E., *et al.* "Water Scarcity Hotspots Travel Downstream Due to Human Interventions in the 20th and 21st Century." *Nature Communications,* vol. 8, no. 1, June 2017, p. 15697. https://doi.org/10.1038/ncomms15697.
16. Green, Timothy R., *et al.* "Beneath the Surface of Global Change: Impacts of Climate Change on Groundwater." *Journal of Hydrology,* vol. 405, no. 3–4, Aug. 2011, pp. 532–60. https://doi.org/10.1016/j.jhydrol.2011.05.002.
17. Huntingford C, Jeffers ES, Bonsall MB, Christensen HM, Lees T, Yang H. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environ Res Lett* 14(12):124007, 2019. https://doi.org/10.1088/1748-9326/ab4e55.
18. Poursaeid, M., Poursaeid, A.H. & Shabanlou, S. A Comparative Study of Artificial Intelligence Models and A Statistical Method for Groundwater Level Prediction. *Water Resource Management* 36, 2022, 499–1519. https://doi.org/10.1007/s11269-022-03070-y.
19. Xiyue Jia, Deyi Hou, Liuwei Wang, David O'Connor, Jian Luo, The development of groundwater research in the past 40 years: A burgeoning trend in groundwater depletion and sustainable management, *Journal of Hydrology,* Volume 587, 2020, 125006, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2020.125006.
20. Swain, S., Taloor, A.K., Dhal, L. et al. Impact of climate change on groundwater hydrology: a comprehensive review and current status of the Indian hydrogeology. *Appl Water Sci* 12, 120, 2022. https://doi.org/10.1007/s13201-022-01652-0.
21. Hai Tao, Mohammed Majeed Hameed, *et al.* Groundwater level prediction using machine learning models: A comprehensive review, *Neurocomputing,* Volume 489, 2022, pp 271-308, https://doi.org/10.1016/j.neucom.2022.03.014.