Journal of Hydrology: Regional Studies 65 (2026) 103398

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Forecasting change in groundwater storage: A multilayer perceptron model leveraging system dynamics simulation data

Gholamreza Eslamifar [a,*], Ilya Zaslavsky [b], Hamid Balali [a], Vytautas Samalavičius [c], Alexander Fernald [a,d]

[a] New Mexico Water Resources Research Institute, New Mexico State University, Las Cruces, NM 88003, USA
[b] San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093, USA
[c] Faculty of Chemistry and Geosciences of Vilnius University, Vilnius, Lithuania
[d] College of Agricultural, Consumer, and Environmental Sciences, New Mexico State University, Las Cruces, NM 88003, USA

ARTICLE INFO

Keywords:
Groundwater
Modeling
Machine Learning
Artificial Neural Network
System Dynamics
Lower Rio Grande
Water
Agriculture
Water Resources Management
MLP

ABSTRACT

Study region: This study focuses on the Lower Rio Grande (LRG) region of Southern New Mexico, a semi-arid area facing significant water scarcity challenges. Groundwater serves as a critical resource for agriculture, domestic use, and industry in the region, necessitating robust management and predictive tools.

Study focus: The research develops a novel groundwater modeling approach by integrating System Dynamics (SD) and Machine Learning (ML) techniques. A validated SD model of the LRG provided simulation data, which was used to train an ML model employing a feedforward Multilayer Perceptron (MLP) as an artificial neural network (ANN) technique. This approach simplifies groundwater dynamics by concentrating on key variables identified through correlation analysis and simulation results. The model achieved a training RMSLE of 0.031 and R-squared of 0.77, with testing results showing RMSLE of 0.036 and R-squared of 0.71, demonstrating predictive reliability.

New hydrological insights for the region: The proposed model enables accurate forecasting of groundwater storage changes while reducing reliance on extensive data inputs. It pioneers using SD simulation data as a data augmentation method for MLP model training, enhancing predictive capabilities. This integrated methodology supports informed groundwater management and policy-making, offering a transferable framework for other hydrologically similar regions.

1. Introduction

Groundwater is a vital freshwater resource that supports agricultural, domestic, and industrial needs worldwide (Megdal, 2018). Monitoring groundwater levels and changes in storage is essential for understanding past, present, and future conditions, allowing policymakers and water resource managers to design sustainable strategies (Wada et al., 2010). Yet, simulating groundwater storage is challenging due to interactions among climatic, topographic, and hydrogeological factors (Afzaal et al., 2019) and anthropogenic activities such as pumping (Sahoo et al., 2017). Understanding how groundwater responds to both climate variability and human

* Corresponding author.
E-mail addresses: reslami@nmsu.edu (G. Eslamifar), izaslavsky@ucsd.edu (I. Zaslavsky), balali@nmsu.edu (H. Balali), vytautas.samalavicius@chgf.vu.lt (V. Samalavičius), afernald@nmsu.edu (A. Fernald).

https://doi.org/10.1016/j.ejrh.2026.103398
Received 1 October 2025; Received in revised form 25 March 2026; Accepted 26 March 2026
Available online 28 March 2026

influence is crucial, particularly given rising agricultural demands (Wada et al., 2010) and the accelerating effects of climate change on water availability and quality (Gurdak et al., 2007; Green et al., 2011; Taylor et al., 2013). Effective prediction, therefore, requires understanding the dynamic relationships between climate drivers, hydrology, water use, and groundwater responses (Sahoo et al., 2017).

Physically based and numerical models have long been used to simulate groundwater quantity and quality (Tao et al., 2022). However, these models are often limited by their dependence on detailed aquifer data, subsurface geology, and hydrogeological parameters—data that are usually only available at well sites and require expert interpretation (Barnett et al., 2012). Such reliance constrains accuracy and objectivity. To address these limitations, in particular, ML has gained prominence over the past two decades as a powerful tool for groundwater modeling (Tao et al., 2022).

ML methods can overcome data dependency challenges and model nonlinear processes that numerical models cannot easily capture, such as isotope–geochemistry signatures influenced by multiple natural processes (Dupuy et al., 2021; Shiri et al., 2021). Using historical data, approaches such as ANNs (Lallahem et al., 2005), fuzzy logic (Nadiri et al., 2019), tree-based algorithms (Davoudi Moghaddam et al., 2020), genetic programming (Kasiviswanathan et al., 2016), and support vector machines (SVMs) (Zhou et al., 2017) have shown strong potential, in some cases outperforming traditional models. Tao et al. (2022) demonstrated the predictive capabilities of ML methods for groundwater forecasting and documented rapid growth in their use. The dominant methodology involves selecting key variables, running ML algorithms (ANN, SVM, etc.), and comparing outputs against numerical models, which highlights the efficiency of AI-based methods (Tao et al., 2022). Specifically, groundwater ML modeling faces persistent data challenges: incomplete datasets, temporal and spatial variability, and lack of consistent long-term observations (Tao et al., 2022). Sparse data limit model accuracy. To address this, researchers employ synthetic data generation and augmentation, including synthetic time series and probabilistic autoregressive models (Pulla et al., 2024; Taccari et al., 2024). These methods enhance robustness but risk bias if synthetic data fail to capture variability (Nikolenko, 2021; Shorten and Khoshgoftaar, 2019). Overfitting or misrepresentation is possible if artificial data are too simplistic (Reichstein et al., 2019; Gao and Miller, 2019; Zhao et al., 2020).

The complexity of groundwater processes, particularly in agricultural and hydrogeological contexts, makes accurate physical simulations difficult due to extensive calibration requirements (Sahoo et al., 2017). ML efficiently handles nonlinear relationships, enabling improved predictions of groundwater changes even where detailed physical data are lacking (Sahoo et al., 2017).

SD is a simulation framework that represents systems using stocks, flows, feedback loops, and time delays (Forrester, 1970). Importantly, SD models are not restricted by the type of processes represented; they can be applied to physical, biological, or socio-economic processes. Here, our application is closer to a socio-hydrological model, which integrates institutional, economic, and water use processes into the hydrological system. It can generate synthetic datasets by simulating nonlinear relationships and feedback structures based on real-world historical inputs. These datasets can then be used to train ML models, reducing the need for extensive field data while preserving predictive capacity. Combining SD and ML thus offers a hybrid approach that strengthens forecasting and supports decision-making. SD simulations data as synthetic training data for ML models (Rahmandad et al., 2025), leveraging SD principles to integrate hydrological, agricultural, and institutional processes (Sun et al., 2017; Bates et al., 2019) alongside physical model variables. The application of SD in water management has advanced considerably, covering hydrological, agricultural, and socio-economic processes (Guo et al., 2001; Winz et al., 2009; Sun et al., 2017; Balali and Viaggi, 2015; Mashaly and Fernald, 2020; Balali et al., 2026). Studies have integrated SD with other methods to enhance hydrological analyses (Schwaninger, 2006; Lättilä et al., 2010; Baki et al., 2012; Kwakkel and Slinger, 2012; Kwakkel and Timmermans, 2012; Subagadis et al., 2014; Xi and Poh, 2015; Wang et al., 2016; Zomorodian et al., 2017; Zolfagharian et al., 2018).

Traditional physically based models demand large observational datasets and extensive calibration, yet monitoring networks are often sparse (Sun et al., 2017). Pure ML models, while less dependent on physical parameters, require large datasets (Chen et al., 2020; Taccari et al., 2024). Surrogate modeling approximates physical models with ML and includes dimensionality reduction or simplification of the mathematical structure, for instance, by eliminating redundant inputs from a complex model (Asher et al., 2015; Müller et al., 2021; Luo et al., 2023), but still depends on high-fidelity runs and often excludes socio-hydrological feedbacks. By contrast, a hybrid SD-ML approach uses validated SD simulation data as rich training datasets for MLP models, embedding feedback while reducing data demands. In this study, outputs from a validated SD model of the Lower Rio Grande (LRG) were used to train an MLP model. This hybrid model captured complex nonlinear groundwater dynamics with fewer inputs and without rerunning the SD model for each forecast, yielding a more lightweight and transferable forecasting tool for data-scarce contexts. Physically based groundwater models have been widely used to represent flow and storage dynamics; however, their application often requires extensive spatial data, parameter calibration, and long observational records, which are not consistently available in semi-arid agricultural basins. In regions such as the Lower Rio Grande, data limitations and strong human–water feedbacks constrain the practical use of fully distributed numerical models for predictive purposes. As a result, ML approaches have increasingly been adopted as predictive surrogates capable of capturing nonlinear system behavior when data availability is limited.

Several ML modeling approaches have been used previously. Random Forest Regression handles nonlinearities and feature interactions but struggles with extrapolation and can be computationally heavy. Support Vector Regression (SVR) is robust in high-dimensional spaces but scales poorly with large datasets and requires intensive hyperparameter tuning (Rammohan et al., 2024). Physics-Informed Neural Networks (PINNs) incorporate physical laws directly into neural networks, enhancing interpretability (Ehteram and Ghanbari-Adivi, 2023). PINNs may be particularly useful in hybrid SD-ML contexts where some variables are physical, and others are socio-economic, as they can enforce physical consistency while leveraging broader datasets.

This research addresses how comprehensive simulation data can be leveraged to develop an improved, easy-to-use model for predicting changes in groundwater (GW) storage. SD, with its ability to simulate and model complex dynamic systems, offers a more extensive dataset, encompassing a wide range of variables at detailed time intervals. These datasets can be used to train an MLP model, potentially leading to a model that can predict groundwater changes without further reliance on the SD model. This approach not only simplifies future predictions for the case study but also offers a broader application for other regions, adjusting the need for extensive data acquisition and management. Coupled SD-ML modeling could be especially crucial for underdeveloped countries and regions where the data availability and variety are very low. While the underlying SD model is data-intensive and includes a large number of variables, it is used here as an offline data generator based on the structure, dynamics, and relationships it captures. The trained MLP model, in contrast, operates on a reduced set of key inputs, significantly lowering data requirements during application. This distinction enables the model to be reused for the region at any time and also deployed in data-scarce settings without the need to reconstruct the full SD model.

The hypothesis for the research is that we can present a new regional model for predicting groundwater changes, which can streamline future modeling efforts and aid in decision-making with less frequent data updates and reduced resource demands. If it is possible, then it could also be applied to respond proactively to critical challenges such as natural disasters or economic crises by incorporating immediate changes in the SD model.

Despite extensive use of ML and surrogate modeling in groundwater studies, a clear gap remains in leveraging socio-hydrological simulation frameworks as structured data generators for ML. Existing surrogate or meta-modeling approaches typically approximate physically based numerical groundwater models and focus primarily on computational speed-up (e.g., Asher et al., 2015; Müller et al., 2021; Luo et al., 2023), often neglecting institutional, agricultural, and behavioral feedbacks. This study addresses this gap by introducing a hybrid SD-ML framework in which a validated socio-hydrological SD model is used to generate structured synthetic training data for an ANN. The novelty lies not in replacing the SD model, but in transferring its embedded socio-hydrological feedbacks into a reduced-variable, data-driven forecasting tool that can be deployed independently of the full system simulation. This approach enables groundwater storage prediction in data-scarce regions where detailed physical models or SD expertise are not readily available, while retaining key dynamics governing groundwater response. Table 1 summarizes the key differences between common ML-based modeling approaches and the proposed hybrid SD-ML framework, highlighting its contribution to capturing socio-hydrological feedbacks.

There are no specific studies that explicitly utilize SD simulation data as synthetic input for training MLP models. However, the concept of using simulated or synthetic data from mathematical models to train MLP models is gaining increasing attention. For instance, Kratzert et al. (2019) explored the use of synthetic hydrological data generated from hydrological models to train ML models for forecasting and decision-making in hydrological systems. Similarly, Raissi et al. (2019) combined physics-based models to generate synthetic data for training ANNs, demonstrating improvements in model performance. This growing trend highlights the potential of integrating simulation data from established models like SD into ML frameworks for enhanced predictive capabilities. This approach enables the generation of new data from simulations, enriching the training set and improving the MLP model's ability to make predictions in situations where real-world data is scarce. This study primarily contributes a methodological framework by integrating SD and ML for groundwater prediction, while the application to the Lower Rio Grande basin serves to demonstrate its practical relevance and transferability rather than to establish new hydrogeological findings.

2. Study region and context

The agricultural economy and water resources in the LRG of southern New Mexico (Fig. 1) are intertwined. During periods of insufficient surface water from the Rio Grande, the Mesilla Rincon Valley (MRV) increasingly relies on groundwater to sustain its robust agricultural activities. Historically renowned for cultivating green chilies, the MRV has, over the past few decades, emerged as the largest producer of pecans in the United States (NMDA, 2021). In addition to perennial pecans, the region supports a variety of annual row crops, including onions and cotton, alongside forage crops such as alfalfa and winter wheat. The shift toward perennial pecan production has brought economic stability and reduced labor demands. However, this transition has also heightened groundwater dependency to maintain yields and ensure tree survival during surface water shortages (Balali et al., 2025).

The region has a semi-arid climate, with yearly rainfall varying between 203 and 255 mm based on the terrain (Bai et al., 2021). Average temperatures throughout the year range from approximately 16–24 °C (Mokari et al., 2021).

The prolonged drought conditions since 2002, coupled with reduced snowmelt runoff from upstream areas, have significantly depleted surface water resources, increasing reliance on groundwater pumping (Mokari et al., 2021). For example, the Elephant Butte Irrigation District (EBID), which historically allocated around 0.91 m (3 feet) of surface water to farmers, could only provide around

Table 1
Key differences among modeling approaches.

Aspect	Surrogate Modeling	Physics-Informed ML	Hybrid SD–ML (This Study)
Core Idea	Emulates numerical/physical models	Integrates physical laws into ML	Learns system behavior from SD-generated data
Data Source	Physical model simulations	Observations + governing equations	Synthetic data from SD model
Human/System Feedbacks	Not included	Rarely included	Explicitly included (socio-hydrological dynamics)
Model Type	Pure ML (black-box)	ML + physics constraints	SD (process-based) + ML (data-driven)
Data Requirement	High	Moderate–high	Lower (SD-generated data)
Use Case	Speed-up simulations	Physically consistent predictions	Policy-relevant system forecasting
Main Limitation	Ignores human dynamics	Requires known equations	Depends on SD model validity

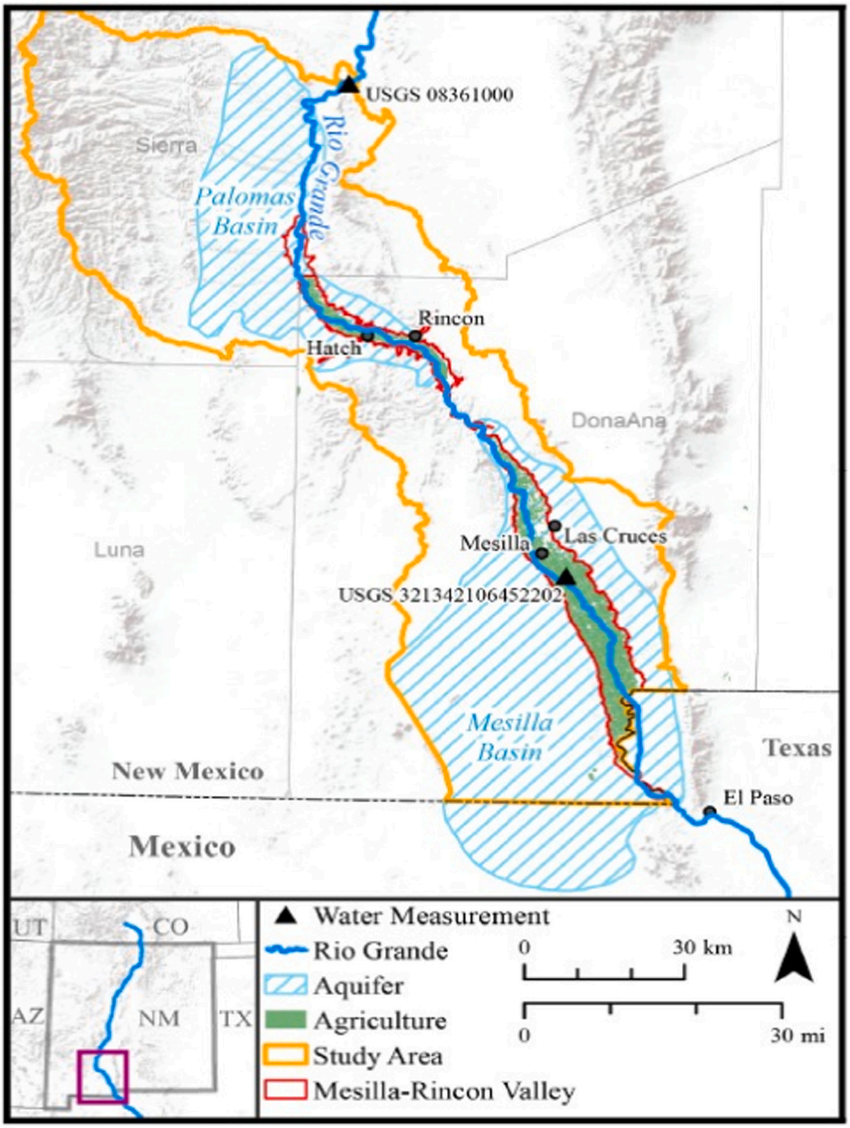

Fig. 1. Lower Rio Grande water planning region, which includes Mesilla Rincon Valley (MRV) in Southern New Mexico (Latitude: 32° 15′ 32″ N (32.270° N), Longitude: 106° 48′ 53″ W (-106.801° W) Elevation: Approximately 3881 feet (1183 m)), reproduced from Eslamifar et al. (2024).

0.1 m (4 in.) in 2021. In the absence of adequate surface water, farmers have turned to the fluvial aquifer beneath the Rio Grande to meet their groundwater needs (Mashaly et al., 2026). This intensified extraction has caused a decline in groundwater levels and has increased river seepage while reducing downstream flow toward Texas. These changes in the water balance led Texas to file a lawsuit against New Mexico, alleging that excessive groundwater pumping has diminished river flow and conveyance to downstream users (Eslamifar et al., 2024).

The SD model used in this research was developed and published before as an output modeling effort in the NM Water Resources Research Institute to study the water scarcity in the LRG region within Dona Ana County, New Mexico. Langarudi et al. (2019) prepared a big and detailed SD model to describe the whole water system. The model includes seven modules, including water, water use, agriculture production, non-agriculture production, population, labor, and wage. The framework allows for the interactive dynamics between groundwater and surface water, capturing the complexities of the socio-hydrological system as shown in Fig. 2. Groundwater and surface water components interact with each other. The SD model consists of a total of 205 variables, including endogenous (stock, flow, auxiliary variables), exogenous inputs, and constant parameters. Endogenous variables represent internal system dynamics, while exogenous variables capture external drivers such as climate and water inflows. Constant parameters, which do not vary over time, were excluded from the MLP training dataset, as they do not contribute to the learning. Thus, only time-varying endogenous and exogenous variables were considered in the feature selection and model training process. There are 97 endogenous variables and 9 exogenous variables (as time-varying data inputs for the SD model). The model was validated and calibrated from the

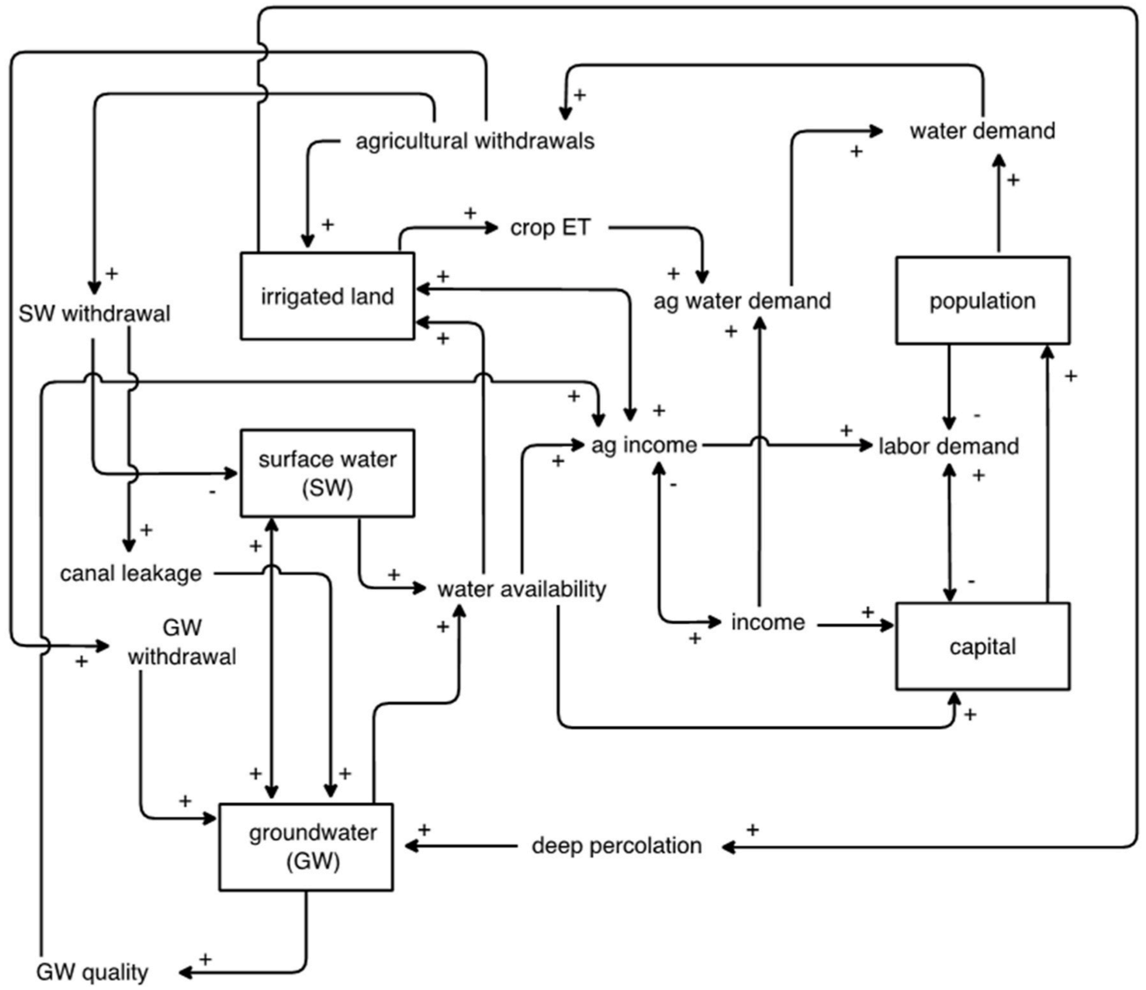

Fig. 2. Major components and flows of the SD model. Reproduced from Bai et al. (2021).

start of 1976 to the end of 2011 under the UKMO (United Kingdom Met Office) climate scenario in terms of different patterns of precipitation and temperature as key exogenous variables for the system. The SD model output and validated results are available in Bai et al. (2021). Model performance was evaluated using Theil's inequality statistics (Bliemel, 1973; Leuthold, 1975), which decompose mean squared error (MSE) into three components: bias (U^M), unequal variance (U^S), and unequal covariance (U^C). A low Theil's coefficient (U) indicates close correspondence between simulated and observed data, while large UC values imply that the model captures the overall mean and trends well. Sterman (1984) recommends this approach for SD models because it indicates both fit quality and potential sources of bias. For most variables, U is below 0.10. Importantly, U^C accounts for the majority of errors in these variables, indicating that the model still captures general trends well. Overall, the validation results suggest that the SD model replicates observed historical patterns across both hydrological and socioeconomic subsystems with acceptable accuracy. Given that the model's primary purpose in this study is to generate synthetic data for MLP model training, its ability to reproduce dynamic patterns with low systematic bias is more critical than a perfect point-to-point fit.

Under the UKMO model assumption applied by Mashaly et al. (2026) for the LRG region in the local model, for the 2017–2050 period, the model indicates wetter conditions relative to historical baselines, with mean annual precipitation reaching 256.5 mm (±68.6 mm), a slight increase compared to the historical average of 254.0 mm. Temperatures during this period rise to 17.0°C (±0.6°C), approximately 0.7°C warmer than historical norms. Notably, surface water inflows surge to 956.5 million cubic meters per year (MCM/yr ±216.8), marking a 14.7% increase over historical levels—the highest inflow among all evaluated models (GFMLP, NCAR) for this timeframe. This suggests heightened water availability but also potential flood risks or storage management challenges.

By the 2051–2099 period, UKMO projects a continued rise in precipitation to 276.9 mm/yr (±61.0 mm), nearly 9% above historical levels and 8% higher than 2017–2050. However, temperatures climb more sharply to 18.7°C (±0.7°C), reflecting 1.7°C of additional warming compared to 2017–2050 and 2.4°C above historical baselines. Despite the wetter conditions, surface inflows decline by 14.9% to 814.4 MCM/yr (±185.1).

3. Method

3.1. SD modeling and simulation

The SD modeling approach utilizes tools capable of incorporating mental models into simulations based on stock-and-flow concepts, which encompass material elements, time lags, and information exchanges (Sterman, 2000). By dynamically replicating the outcomes of evolving systems, SD modeling serves as a decision support tool for testing strategic policies, providing a unique means to evaluate the effects of potential policy interventions and modifications within complex systems (Sterman, 2000). This methodology has a history of application in focusing on complex systems, especially the environment, natural resources, and economic and social systems like water and agriculture (Forrester, 1970, 1971; Meadows et al., 1972; Fiddaman, 1997; Sterman, 2008; Yang et al., 2021; Herrera et al., 2022).

3.2. Process of creating the MLP model with SD simulation results

In this research, MLP was selected for three reasons. First, SD simulations generated large synthetic datasets across thousands of time-steps (1976–2011), enabling deeper networks. Second, groundwater dynamics are nonlinear, shaped by feedbacks between irrigation, recharge, and soil moisture, which MLPs can capture effectively. Third, implementation in TensorFlow/Keras ensures scalability and experimentation with advanced architectures like Long Short-Term Memory (LSTM) networks. To prevent overfitting, dropout regularization was applied, and Mean Squared Logarithmic Error (MSLE) was used as the loss function, well-suited for groundwater datasets spanning multiple magnitudes.

Due to the complexities, including nonlinear behaviors and the large number of input variables needed to define Change in GW storage, ANN (McCulloch and Pitts, 1943; Haykin, 2009) was selected as the supervised ML approach. In this context, the data from selected variables in the SD model presents a regression problem that can be effectively addressed using MLP techniques. This approach allows the model to capture complex relationships within the dataset, facilitating more accurate predictions of groundwater changes. We implemented a feedforward MLP with three hidden layers. To address the research question, we designed a stepwise

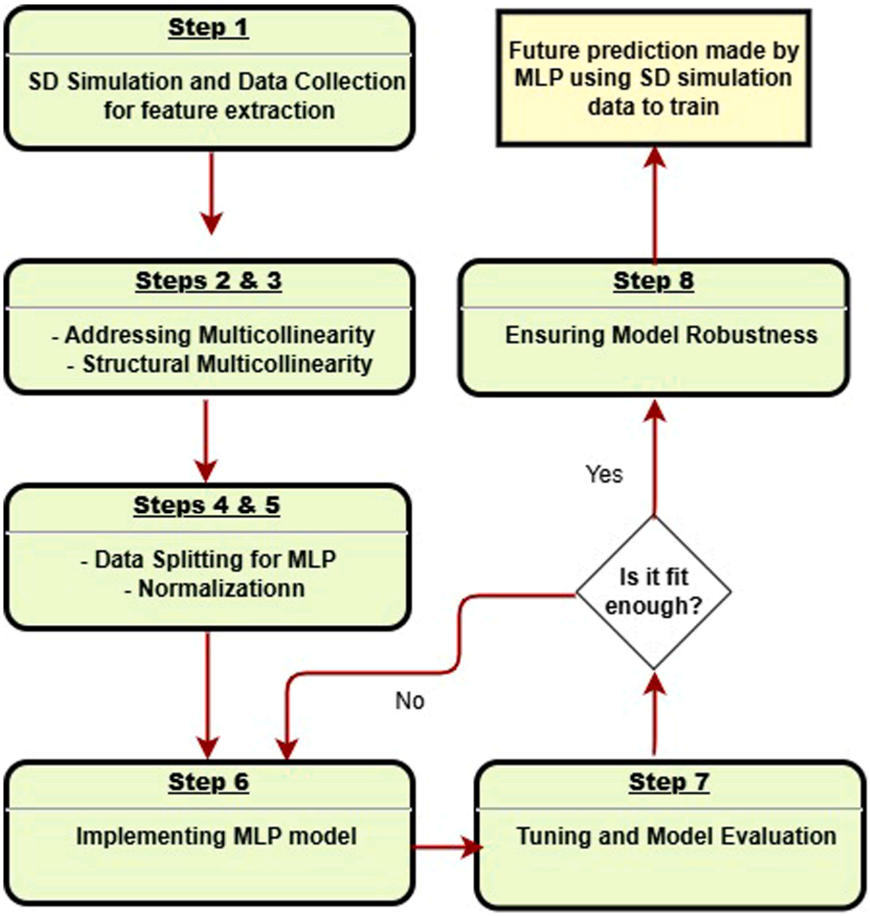

Fig. 3. General overview of workflow to create MLP model with SD data.

process that utilizes SD model data to develop a standalone MLP model for predicting groundwater storage changes. The MLP model inputs are exclusively based on synthetic data generated by the SD model. No independent hydrological or socioeconomic datasets were directly used in the MLP model training. Therefore, the MLP model acts as a surrogate model, approximating the SD model's outputs with fewer features. This is a form of low-fidelity surrogate modeling, where feature elimination is used to reduce dimensionality while retaining predictive performance. The 8-step process, outlined in Fig. 3, includes simulation and data collection for feature extraction, addressing multicollinearity, removing structural multicollinearity, splitting data for training and test, normalizing and training the model, implementing the MLP model, tuning and model evaluation, and ensuring model robustness for reliable groundwater storage prediction as explained in the following steps in detail:

1. SD Simulation and Data Collection for Feature Extraction: We first run the SD model simulation at a precise time step (0.01-year) to ensure that the data points satisfy the requirements for learning of the MLP model. With all variables simulated, we begin Process 1 (Fig. 3), calculating the correlation between the key variable(change in groundwater (GW) storage)and all other variables using historical data. The variable selection process began with an initial pool of 106 variables extracted from the SD model outputs, representing system-wide socio-hydrological dynamics. Pearson's correlation is employed to identify significant correlations ($0.5 < |r| < 0.9$), where high correlations ($|r| > 0.9$) indicate a multicollinearity issue. The p-value for all variables is also computed, with values below 0.05 indicating statistical significance.
2. Addressing Multicollinearity: If the absolute value of Pearson's correlation coefficient between two independent variables exceeds 0.8 or 0.9, multicollinearity is likely. To formally assess this, we calculate the Variance Inflation Factor (VIF) for all variables (with $0.5 < |r| < 0.9$). VIF values greater than 5 signal multicollinearity, and we target those for correction. After calculating R-squared for each correlation, we compute the VIF, marking any values greater than 5. Table 2 summarizes the thresholds for selected variables from steps 1 and 2.
3. Structural Multicollinearity: As feedback loops are central to SD modeling, we ensure that no selected variables have direct structural links with the key variable (change in GW storage). We review each variable's formulation to eliminate structural multicollinearity. For instance, variables like "effect of GW use on agriculture income" and "population growth capacity" require careful examination to avoid direct dependence on groundwater storage.
4. Data Splitting for MLP: The historical data (from SD simulation data) is split into training (1976–2004) and test sets (2005–2011), a standard requirement for MLP models. Given that this is time-series data generated from SD simulations, we employ non-random train-test splits (Medar et al., 2017; Rinderknecht and Klopfenstein, 2021).
5. Normalization: Before training the MLP model, we normalize Change in GW storage data to ensure they are on a similar scale. Z-score normalization (Eq. 1) is applied to the dataset to avoid outputs with larger values dominating the training process (Pedregosa et al., 2011), and the variable called Change in GW storage index.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

To handle errors and reduce the impact of outliers, Mean Squared Logarithmic Error (MSLE) is chosen as the loss function (Eq. 2).

$$\mathrm{MSLE} = \frac{1}{n}\sum_{i=1}^{n}\left(\log(1 + y_i) - \log(1 + \hat{y}_i)\right)^2 \tag{2}$$

6. Implementing MLP model: The model was implemented as an MLP with three hidden layers. Rectified Linear Unit (ReLU) activation functions were used for all hidden layers, followed by a linear output layer. Dropout regularization (rate = 0.2) was applied to mitigate overfitting. Model training employed the Adam optimizer with a learning rate of 0.01. The dataset was divided into training (1976–2004) and testing (2005–2011) periods to evaluate generalization performance. Model evaluation was conducted using an independent testing dataset covering 2005–2011, which was not used during training. The MLP model was designed using TensorFlow (v2.x) and the Keras API, chosen for its flexibility and scalability in developing neural networks. A feedforward MLP was selected for its capability to capture nonlinear relationships in tabular groundwater data while maintaining computational efficiency. The model was constructed using a sequential architecture with three hidden layers consisting of 64, 32, and 16 neurons, respectively, each using the ReLU activation function and initialized with a normal distribution. Dropout layers with a 20% rate were added after the first and second hidden layers to mitigate overfitting. The output layer comprised a single neuron with a linear activation function to predict continuous groundwater change values. The model was compiled using the Adam optimizer with a learning rate of 0.01. Training was conducted over 200 epochs with a batch size of 64.
7. Tuning and Model Evaluation: After tuning the model, we evaluate its performance by calculating RMSLE (Eq. 3). And R-squared (Eq. 4) is used to assess the model's predictive performance during training and testing (Moriasi et al., 2007).

Table 2
Criteria and thresholds to select variables (features).

Criterion	Threshold	Description
Pearson's Correlation Coefficient (r)	$0.5 < \|r\| < 0.9$	Identifies significant correlations
Variance Inflation Factor (VIF)	< 5	Detects and reduces multicollinearity

$$\text{RMSLE} = \sqrt{(\log(y_i + 1) - \log(\hat{y}_i + 1))^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

8- Ensuring Model Robustness: To verify the robustness of the MLP model, we use unseen test data and confirm that the model maintains similar RMSLE values. Also, to assess how uncertainties in the SD model may influence the performance of the trained MLP model, we conduct a sensitivity analysis focusing on the variables selected for MLP model training. A one-at-a-time (OAT) sensitivity analysis was conducted to assess the influence of input variables on predicted groundwater storage changes. Each input variable was independently perturbed by $\pm 5\%$ and $\pm 10\%$ relative to its baseline value while holding other inputs constant. The resulting change in model output was quantified relative to the baseline simulation. In addition, observed groundwater level data from USGS stations that were aggregated in the NM DSWB model (Mashaly et al., 2026) were used to show the SD model calibration, as was done in Bai et al. (2021). The observed data were normalized to create a standardized Change in GW storage index comparable to model outputs. Table 3 summarizes MLP model architecture, training configuration, and hyperparameters.

4. Results and discussion

Based on the process as described in 3.2, we first ran the SD model simulation under the UKMO scenario, using a 0.01-year time step to generate detailed data for MLP. This produced 100 data points per year, resulting in a dataset with 3600 points covering the period from 1976 to 2011. We calculated the correlation between the change in GW storage (cubic meters), as the key variable, and all other variables. After sorting and removing redundancies, 13 variables were identified (Table 4) that exhibited a strong correlation with change in GW storage ($0.5 < |r| < 0.9$). Pearson’s correlation was used to quantify the relationships between variables.

As noted in step 1 of the process for creating an MLP model with SD data (Section 3.2), multicollinearity could introduce issues in regression analysis. This could lead to unreliable estimates of regression coefficients, as one variable may predict the other with high accuracy. Additionally, the p-value for all variables was calculated, with values near zero indicating statistically significant correlations.

Pearson's correlation matrix contains the correlation coefficients between various independent variables related to Change in GW, as shown in Fig. 4.

After calculating the R-squared for each correlation, we computed the VIF, and any values exceeding 5 were flagged as shown in Fig. 5.

Two groups of variables exhibiting multicollinearity can be found in Fig. 5. Within each group, we either needed to select one representative variable or combine the variables to avoid redundancy. In Group A, which included water supply, soil moisture from surface water (SW) irrigation, conveyance recharge, and agricultural SW withdrawal, we selected 'soil moisture from SW irrigation' due to wide data coverage for the region and the greater precision of its measurement methods, compared to, e.g. conveyance recharge, of which only partial/insufficient data was available. In Group B, which comprised river evaporation, SW outflow, river leakage, and surface water, 'river leakage' was chosen based on its physical relevance (direct recharge of groundwater by downward percolation) and broader applicability in other case studies. This refinement reduced the variable set from 13 to 8 variables.

As feedback loops are a central feature of SD modeling, it is essential to ensure that no direct structural multicollinearity exists between the selected variables and the key variable (change in groundwater, GW) in our study. To address this, we reviewed the model

Table 3
MLP model architecture, training configuration, and hyperparameters.

Category	Hyperparameter	Description	Tested Range	Final Value Used
Model Architecture	Input variables	Number of predictor variables selected from the SD model	—	6
	Hidden layers	Number of fully connected hidden layers	2–4	3
	Neurons per layer	Number of neurons in each hidden layer	16–128	64 – 32 – 16
	Activation function	Nonlinear activation	ReLU, tanh	ReLU
	Output activation	Activation for regression output	Linear	Linear
Training Configuration	Optimizer	Optimization algorithm	Adam, RMSprop	Adam
	Learning rate	Step size for gradient descent	0.001–0.01	0.01
	Batch size	Samples per training batch	16–64	32
	Epochs	Maximum number of training iterations	100–500	200
	Loss function	Objective function	MSE, MSLE	MSLE
Regularization	Dropout rate	Fraction of neurons randomly dropped	0.1–0.3	0.2
	Early stopping	Stop training when validation loss stabilizes	Yes / No	Yes
	Early stopping patience	Number of epochs without improvement	10–30	20
Data Handling	Data split	Temporal split for training/testing	—	1976–2004 (train), 2005–2011 (test)
	Data normalization	Feature scaling method	Min–Max, Z-score	Z-score
	Random seed	Reproducibility control	—	Fixed

Table 4
Input variables set for the MLP model and correlation value with the change in GW storage as output.

Variable name	Correlation value (r)
Surface water inflow (sw in)	0.85
Water supply	0.83
Soil moisture (soil moisture from surface water irrigation)	0.82
Conveyance recharge	0.82
Agriculture surface water withdrawal	0.82
Effect of groundwater use on agricultural income	0.72
Infiltration rate (infiltration)	0.67
Population growth capacity (pop growth cap)	0.64
River evaporation	0.64
Surface water outflow	0.64
River leakage	0.64
Surface water	0.64
Irrigation drainage	-0.61

	sw in	water supply	soil moisture from sw irrigation	conveyance recharge	ag sw withdraw	eff of gw use on ag inc	infiltration	pop cap growth	river evaporation	sw outflow	river leakage	surface water	irrigation drainage
sw in	1												
water supply	0.995	1											
soil moisture from sw irrigation	0.892	0.896	1										
conveyance recharge	0.892	0.896	0.873	1									
ag sw withdraw	0.892	0.896	0.742	0.912	1								
eff of gw use on ag inc	0.592	0.570	0.591	0.591	0.591	1							
infiltration	0.781	0.745	0.593	0.593	0.593	0.177	1						
pop cap growth	0.819	0.782	0.755	0.755	0.755	0.104	0.846	1					
river evaporation	0.836	0.821	0.712	0.712	0.712	0.227	0.802	0.896	1				
sw outflow	0.836	0.821	0.712	0.712	0.712	0.211	0.907	0.896	0.855	1			
river leakage	0.836	0.820	0.711	0.711	0.711	0.196	0.908	0.896	0.824	0.866	1		
surface water	0.836	0.821	0.712	0.712	0.712	0.211	0.891	0.896	0.923	0.815	0.942	1	
irrigation drainage	-0.449	-0.397	-0.340	-0.340	-0.340	-0.021	-0.711	-0.547	-0.438	-0.438	-0.439	-0.438	1

Fig. 4. Correlation matrix for selected variables.

formulation of each variable to evaluate the potential influence of the key variable on them.

The two variables, "effect of GW use on agriculture income" and "effect of water availability on agricultural income," were removed to address structural multicollinearity. Even though they passed the statistical test for non-collinearity, they were defined based on GW storage and Change in GW storage and consequently, should be removed.

The final set of six variables as features to train the MLP model was selected as described in Table 5. Precipitation effects are represented indirectly through SD-derived variables such as surface water availability, soil moisture from irrigation, and irrigation demand, and were excluded from the reduced MLP model input set to avoid redundancy.

Following the workflow, data from 1976 to 2004 were used as the training set, while data from 2005 to 2011 were used as the test set. In the next step, given that the selected variables (features for training) had different units and scales, a normalization function was applied to standardize all the data before training. For the test data, an RMSLE of 0.036 and an R-squared of 0.71 confirm the model's reliability without overfitting. After training the MLP model, the Root Mean Square Logarithmic Error (RMSLE) for the training dataset was 0.031, and the R-squared was 0.77. When predicting the test dataset, the RMSLE was 0.036, indicating a good fit with no signs of overfitting. Additionally, the R-square value was 0.71. Model performance during both training and testing periods indicates stable generalization, with comparable R^2 and RMSE values reported in Table 6. The discrepancies between the SD and MLP model outputs (Fig. 6) arise from inherent differences in their methodological frameworks and error propagation mechanisms. Normalization, while essential for MLP training, can propagate systematic errors if the training data's mean and variance do not represent future conditions (e.g., extreme drought events outside the 1976–2004 range). Although the MLP model showed no overfitting (stable RMSLE and R^2

	sw in	water supply	soil moisture from sw irrigation	conveyance recharge	ag sw withdraw	eff of gw use on ag inc	infiltration	pop cap growth	river evaporation	sw outflow	river leakage	surface water	irrigation drainage
sw in	*												
water supply	1	*											
soil moisture from sw irrigation	0	1	*										
conveyance recharge	0	1	1	*									
ag sw withdraw	0	1	1	1	*								
eff of gw use on ag inc	0	0	0	0	0	*							
infiltration	0	0	0	0	0	0	*						
pop cap growth	0	0	0	0	0	0	0	*					
river evaporation	0	0	0	0	0	0	1	1	*				
sw outflow	0	0	0	0	0	0	1	1	1	*			
river leakage	0	0	0	0	0	0	1	1	1	1	*		
surface water	0	0	0	0	0	0	1	1	1	1	1	*	
irrigation drainage	0	0	0	0	0	0	0	0	0	0	0	0	*

Fig. 5. VIF measure matrix to find multicollinearity; 1 means VIF > = 5, and means VIF < 5.

Table 5
Six variables as features to train the MLP model (simulated by SD model to prepare augmented data for the MLP model).

Variable Name	Description	Units
Surface water inflow (sw in)	The volume of water entering the region's surface water systems	Cubic meters per year
Soil moisture (soil moisture from sw irrigation)	Soil moisture from surface water irrigation	Cubic meters per year
Population growth capacity (pop growth cap)	Rate of population increase	Persons per year
River leakage	Water leaking into groundwater from rivers	Cubic meters per year
Infiltration rate (infiltration)	Rate of water infiltration into the ground	Cubic meters per year
Irrigation drainage	Excess water draining from irrigation systems	Cubic meters per year

Table 6
Results of evaluation.

Evaluation Metrics	Description	Final Value
RMSLE	Root Mean Squared Logarithmic Error	0.031 (train), 0.036 (test)
R^2	Coefficient of determination	0.77 (train), 0.71 (test)

across training/test sets), its predictions are contingent on the historical data's representativeness.

Fig. 6 shows the comparison between the MLP model results for the test dataset and the SD simulation for Change in GW storage index. Also, the GW index from observed data, which was extracted from monthly data of NM DSWB (Mashaly et al., 2026) and was used for SD model calibration and validation by Langarudi et al. (2019), compared with both change in GW storage index from SD simulation and prediction of change in GW storage index by MLP. It is important to note that the MLP model is primarily trained on synthetic data generated from the validated SD model, and therefore its performance is first evaluated in terms of its ability to reproduce SD-simulated system behavior. In this sense, the MLP model acts as a surrogate of the SD model. Observational groundwater data are used as a secondary validation step to assess the realism of the coupled SD-ML framework. Given the limited temporal overlap of available observations, this validation is necessarily partial; however, it provides additional confidence that the model captures key groundwater dynamics in the Lower Rio Grande basin. The GW index values are normalized (Z-score normalization as noted in the Method section) to align with the input scales used in the training dataset, ensuring consistent evaluation across model outputs. As the model performance measures (RMSLE and R-squared) didn't change considerably when applied to unseen data (test set) in comparison with training data set, the MLP model is not in an overfit condition. Using historical data to train the MLP model suggests that the model can produce reliable predictions for future groundwater changes.

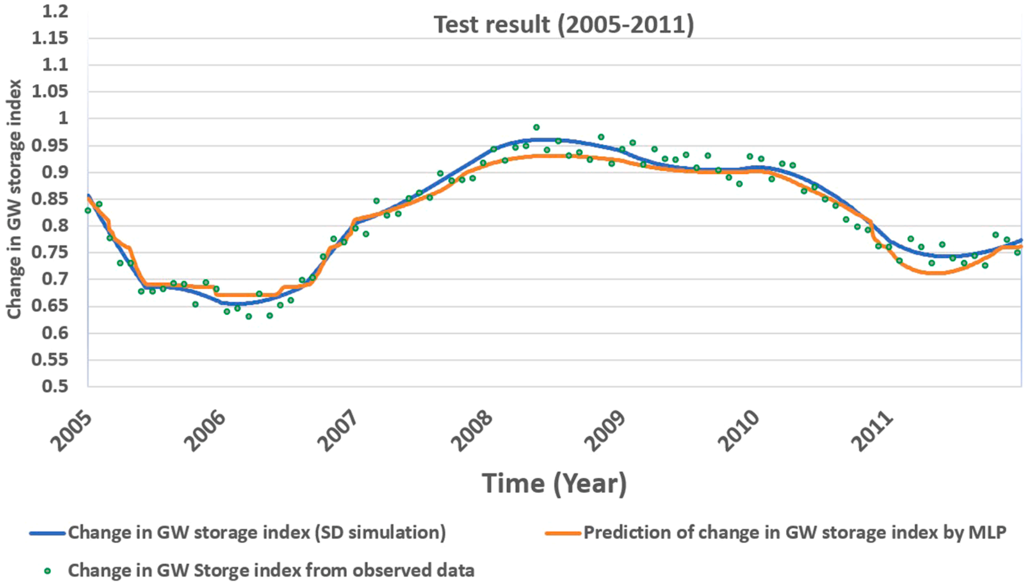

Fig. 6. Out-of-sample comparison between SD-simulated, MLP-predicted GW storage index, and observed data from GW storage index during the testing period (2005–2011). Performance metrics (R^2, RMSE) are shown on Table 6.

This implies that the MLP model can now be used as an independent tool for predicting groundwater changes in various scenarios without relying on the SD model or additional calibration. With only the final six variables, the MLP model can be trained to provide accurate forecasts for future groundwater changes.

By perturbing key 6 variables selected for MLP training within ±%5 and ±%10 (totally 25 runs), we generated a suite of SD simulations and retrained the MLP on each variant. The standard deviation of the MLP model's predictions across these runs was used as a proxy for sensitivity to SD inputs. We found that the MLP model's outputs varied by ±%6 in average around the baseline prediction. This indicates that while the MLP model is generally robust, it inherits and amplifies some of the structural assumptions embedded in the SD simulation. Consequently, careful calibration and validation of the SD model remain essential prior to its use in training data-driven models. This hybrid approach thus benefits from the interpretability of mechanistic models and the predictive strength of data-driven models, but it also requires transparent treatment of error propagation and uncertainty.

Also, sensitivity analysis results indicate that the selected input variables exhibit varying levels of influence on groundwater

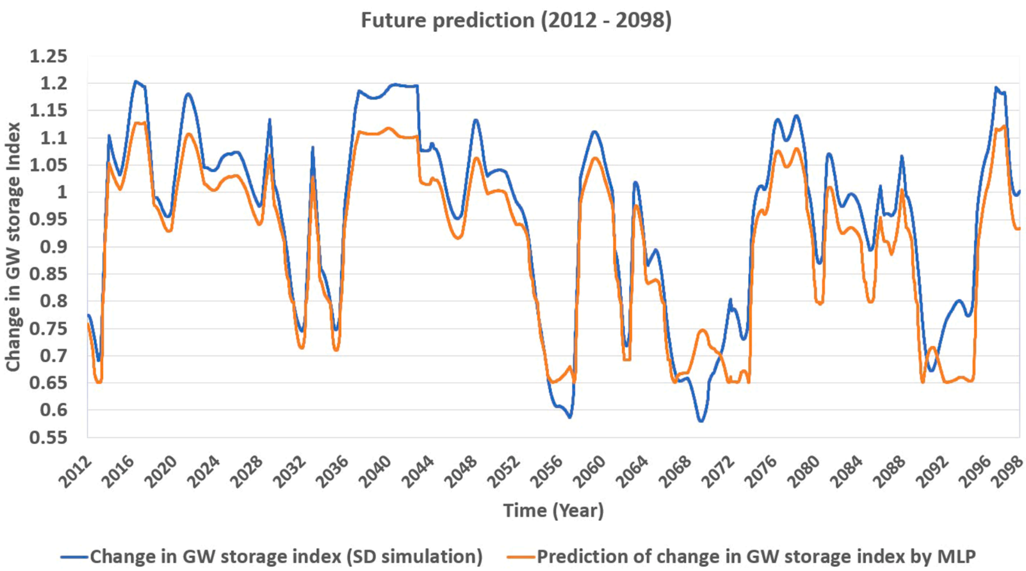

Fig. 7. Future prediction comparison between SD-simulated, MLP-predicted GW storage index, and observed data from GW storage index (2012–2098).

storage. Among them, irrigation demand and surface water inflow show the highest sensitivity, with perturbations leading to the largest changes in predicted groundwater storage. This reflects the dominant role of agricultural water use and surface water availability in the basin. Variables such as soil moisture from irrigation and river leakage demonstrate moderate sensitivity, while other inputs exhibit comparatively smaller effects. Overall, the results highlight the critical importance of conjunctive water use dynamics in shaping groundwater variability in the region.

Thus, we have developed a predictive tool that leverages both SD simulation and MLP. This approach allows us to utilize the MLP model, trained with SD-generated data, for accurate predictions. As the next step, we applied the MLP model to forecast future groundwater changes, using SD simulation data from 2012 onwards, which has not been validated or calibrated. Fig. 7 illustrates the MLP model's forecast for future.

In this process, the majority of the historical SD simulation data was used as training data for the MLP model, while the remaining portion was employed to test the model's performance, which was found to be satisfactory. This demonstrates the MLP model's capability to reliably predict future changes, suggesting that it can be used and replaced for forecasting groundwater dynamics.

The integration of SD modeling with MLP techniques in this research provides a new approach to forecasting groundwater storage changes. Traditional methods for simulating groundwater dynamics, such as physical and numerical models, often struggle with the complexity of groundwater systems, especially in regions with limited data (Tao et al., 2022). The coupled SD-ML approach offers a useful solution, demonstrating several key benefits over conventional modeling methods. It recognizes the limitations of traditional models and explores the potential of AI, with specific attention to ANN and SVM methodologies (Tao et al., 2022).

The study makes several contributions to the existing literature. Firstly, it resulted in a new model a socio-hydrological variable (population growth capacity), addressing a critical gap in traditional groundwater models (Sahoo et al., 2017). The methodology enables the creation of a new regional groundwater model, offering a transferable framework for different cases of study. Selected inputs (six features) are common hydrological and socio-economic parameters that can be reused in on other regions with similar hydrogeological setting. The research capitalizes on the strengths of SD, utilizing its ability to simulate complex dynamic systems, to generate detailed datasets as combination of synthetic data for training MLP algorithms. The performance metrics, with an RMSLE of 0.036 and an R-square of 0.71 for test data set versus RMLSE 0.031 R-square of 0.77 for traning data set, underscore the suitable fit of the model. The new model offers several significant implications.

Reduced Dependence on Physically-Based Groundwater Models: Groundwater systems are inherently complex, partially unobservable, and dynamically evolving, which makes their complete physical characterization impossible (Freeze et al., 1990; Oreskes et al., 1994). Consequently, all groundwater models represent simplified abstractions designed to address specific objectives such as flow, storage, or mass balance estimation. Traditional physically-based models require detailed spatial digitization and extensive parameterization of subsurface properties, leading to substantial data and calibration demands. In contrast, ML approaches can infer system behavior directly from data without explicit representation of physical processes (Chen et al., 2020). This distinction makes ML models particularly valuable in data-limited regions, where constructing and maintaining physically-based models is often impractical.

While the current hybrid SD-ML approach leverages historical dynamics simulated by a SD model to train a MLP framework, future research can benefit from integrating physics-informed machine learning (PIML) methodologies to enhance physical consistency and interpretability. PIML frameworks offer a promising avenue for embedding physical laws like mass balance, Darcy’s law, or governing partial differential equations (PDEs) directly into the learning process of neural networks. Recent studies, such as Cai et al. (2021), (2022), (2024), illustrate how interpretable MLP models can uncover multiscale groundwater drought mechanisms, while Zhan et al. (2024) demonstrate the ability to identify PDEs governing groundwater processes using Lasso-regularized physics-informed regression. Incorporating such approaches into groundwater forecasting models could significantly improve their generalizability to nonstationary or extreme conditions, mitigate overfitting to historical patterns, and increase stakeholder confidence in predictive outcomes.

Improved Predictive Performance under Data Scarcity: Limited availability of high-quality groundwater data remains a major challenge for reliable numerical modeling. While physically-based groundwater models can be formulated within an SD framework using stocks and flows, the SD model applied here explicitly integrates socio-hydrological and institutional feedbacks rather than focusing solely on physical processes. Although the SD model includes more than 205 variables, its structure is designed to capture key feedback mechanisms across hydrological and socio-economic subsystems.

It should be noted that the proposed SD-ML model produces deterministic (point) forecasts without explicitly quantified predictive uncertainty. Therefore, model outputs should be interpreted accordingly. A preliminary ensemble analysis based on multiple MLP training runs indicated limited variability in predictions, suggesting stable model performance; however, this does not replace a formal uncertainty quantification framework. Future work should incorporate probabilistic approaches, such as Monte Carlo dropout or Bayesian neural networks, to provide prediction confidence intervals and better characterize uncertainty under data-scarce conditions.

Reduction of Model Complexity through SD-ML Integration: Coupling SD with MLP approach enables substantial model simplification without sacrificing predictive capability. Traditional groundwater models often require large input sets and intensive calibration efforts. In contrast, the MLP model developed here relies on a reduced subset of influential variables, including surface water inflow, irrigation-induced soil moisture, and river leakage. Groundwater systems exhibit strong nonlinearities and interdependencies, which are effectively captured by neural networks trained on SD-generated synthetic data and validated against observations. While numerical models simulate groundwater behavior based on physical parameters, ANNs learn dynamic input–output relationships (Sahoo et al., 2017). Similar to prior ANN applications in groundwater forecasting (Coppola et al., 2003, 2005), the proposed SD-ML framework achieves high predictive accuracy while generalizing across varying hydrological conditions.

Hydrological Implications for the Lower Rio Grande Basin: The results are consistent with previous studies emphasizing the combined effects of groundwater pumping and climate variability in the Lower Rio Grande (LRG) Basin (Sheng, 2013; Liu and Sheng, 2011). Simulated groundwater storage variability aligns with USGS assessments of surface and shallow groundwater dynamics in the Rio Grande Project area (Driscoll and Sherson, 2016). Although the primary contribution of this study is methodological, the SD-ML framework reveals meaningful hydrological patterns. The modeling results provide region-specific insights into groundwater dynamics in the Mesilla Basin. Consistent with Bai et al. (2021), prolonged drought conditions, reduced upstream surface water deliveries, and expansion of perennial crops have intensified groundwater dependence. Sensitivity analysis indicates that irrigation demand and surface water inflow are the dominant controls on groundwater storage variability, with groundwater declines intensifying during periods of reduced surface water delivery and increased irrigation demand. This finding highlights the critical role of conjunctive water use, where surface water shortages directly translate into increased groundwater withdrawals. Also, these findings highlight the potential effectiveness of demand-side interventions, including improved irrigation efficiency and canal leakage mitigation, in stabilizing groundwater levels during prolonged droughts. The predictive behavior of the SD- ML model reflects key hydrological characteristics of the Lower Rio Grande (LRG) and the Mesilla–Rincon Valley Basin. Groundwater storage changes are strongly influenced by surface water availability from the Rio Grande, irrigation demand, and soil moisture dynamics.

Transferability to Other Groundwater-Stressed Regions: Although this study focuses on the LRG Basin in southern New Mexico, the proposed SD-ML framework is transferable to other regions facing groundwater stress, complex hydrological settings, and limited data availability. The ability to train MLP models using SD-generated simulations reduces dependence on extensive field data and calibration efforts, making the approach particularly relevant for arid and semi-arid regions worldwide where data scarcity constrains effective water management.

Integration of Socio-Hydrological Variables: Unlike conventional groundwater models that emphasize physical hydrology alone, the proposed framework explicitly incorporates socio-hydrological variables such as irrigation practices, soil moisture from irrigation, drainage return flows, and population-driven water demand. Previous studies have highlighted the importance of integrating socio-economic drivers into groundwater modeling (Kumar et al., 2005, 2017). By embedding these factors within the SD structure and transferring their dynamics to the MLP model, the framework provides a more comprehensive representation of groundwater behavior.

Decision Support for Sustainable Groundwater Management: The SD-ML model framework functions as a decision-support tool by providing timely and computationally efficient groundwater forecasts. This aligns with the growing use of AI-based tools for sustainable water management (Daliakopoulos et al., 2005; Krishnan et al., 2022; Farzana et al., 2024). Model outputs can inform policy decisions related to groundwater allocation, conservation strategies, and drought mitigation. Moreover, the framework allows for the exploration of abrupt socio-economic or hydrological changes, supporting proactive responses to emerging water management challenges.

SD Simulation as a Data Augmentation Strategy: A central innovation of this study is the use of SD-generated simulations as a data augmentation technique for MLP model training. Data scarcity is a major limitation in MLP applications for groundwater modeling, often restricting model accuracy and generalizability. While synthetic data generation has been explored using physically-based models (Kratzert et al., 2019; Raissi et al., 2019), this study uniquely leverages SD simulations to generate high-dimensional, internally consistent socio-hydrological data. The dynamic and feedback-rich nature of SD models makes them particularly well-suited for augmenting training datasets and enhancing MLP performance in groundwater forecasting.

5. Conclusion

5.1. Main findings

This study demonstrates that synthetic data generated from a validated socio-hydrological SD model can be effectively used to train an MLP model for predicting groundwater storage changes. The resulting SD-ML model framework captures key nonlinear interactions governing groundwater dynamics in the Lower Rio Grande basin while substantially reducing model complexity and data requirements. This study shows the potential for applying this SD-ML modeling approach to other regions facing similar groundwater challenges. By providing a transferable framework, the model can aid decision-makers in water resource management, especially in regions with limited data availability. Future research could enhance the model's accuracy by integrating real-time data and exploring additional variables to further improve predictive capabilities. The results underline the applicability of combining SD and MLP for groundwater modeling, contributing valuable insights into sustainable water resource management.

5.2. Limitations and future research

Despite its promising results, the study has certain limitations that should be addressed in future research. While the MLP model achieved high predictive accuracy, its reliance on SD-generated data means that any inaccuracies in the SD model could propagate through the MLP model. Ensuring the continued validation and calibration of the SD model, particularly under changing climate conditions, is crucial for maintaining the accuracy of the MLP predictions. Future research could also explore the integration of additional data sources, such as remote sensing or real-time monitoring systems, to further enhance the model's predictive capacity.

A notable limitation of the current MLP model lies in its limited direct evaluation under extreme hydroclimatic conditions, such as prolonged droughts or episodic intense rainfall that may drive substantial groundwater depletion or recharge. While the SD model was previously validated under some extreme conditions, the MLP model's reliance on these simulations may still constrain its responsiveness to novel or more severe future events. Although the training dataset reflects historical variability in key drivers such as climate and irrigation, it may not fully capture the statistical rarity or abrupt nonlinearities associated with future extremes, particularly those that deviate from the observed historical regime. This poses a concern for predictive generalizability, especially in semi-arid regions like the Lower Rio Grande, where increased climate variability is expected under global change. Future work should explore the integration of synthetic stress-testing (e.g., climate projections or hydrological simulations of extreme droughts and recharge events) to assess and strengthen the MLP model's robustness. Combining data-driven models with physically-based components or ensemble scenario approaches may enhance their adaptability to nonstationary and extreme future conditions. The MLP model developed in this study is trained on internally consistent scenarios generated by a socio-hydrological SD framework. As such, ad hoc perturbation of individual climate variables (e.g., precipitation shocks) outside the SD structure may violate coupled hydrologic–institutional feedbacks and lead to unrealistic system states. Consequently, quantitative stress testing under extreme climate conditions is beyond the scope of the present work and is best addressed through future SD-driven climate scenario ensembles that preserve physical and socio-hydrological consistency.

Even though the proposed MLP model demonstrates acceptable performance in reproducing the mean behavior of groundwater storage change, it exhibits limited ability in predicting extreme values, as illustrated in Fig. 7. This limitation is not uncommon in MLP models applied to environmental systems and arises from a combination of data and modeling factors. First, the synthetic dataset generated by the SD model includes relatively few instances of extreme change in GW storage index values, resulting in a skewed distribution that favors moderate changes. This imbalance leads the model to prioritize learning dominant patterns, thereby minimizing loss but underrepresenting rare events. Second, the selected loss function (MSLE) is designed to penalize relative rather than absolute deviations. While this ensures stability across varying magnitudes, it inherently dampens the model's sensitivity to large positive or negative fluctuations. Third, the use of dropout layers for regularization helps prevent overfitting but may also suppress the model's responsiveness to high-variance inputs, further constraining its representation of extremes. To address this, future work may incorporate reweighted loss functions to emphasize rare events, ensemble methods that capture both central trends and outliers, or PINNs that explicitly enforce physical constraints. We have included a dedicated discussion of this limitation in the revised manuscript to provide a more nuanced interpretation of the model's strengths and boundaries.

Although the MLP model demonstrated good predictive performance as measured by RMSLE and R^2, the study does not currently quantify predictive uncertainty through formal statistical frameworks such as prediction intervals or probabilistic inference. This is a limitation, particularly for decision support applications where understanding the confidence bounds around predictions is essential. Uncertainty in MLP model outputs may stem from various sources, including input data variability, biases in the training data derived from the SD model, and the intrinsic stochastic nature of MLP training processes (e.g., weight initialization, optimization paths). While we mitigated some of these uncertainties through robust variable selection and model validation procedures, future research could benefit from incorporating Monte Carlo dropout methods, ensemble predictions, or Bayesian neural networks to produce probabilistic forecasts. These approaches would allow for more transparent risk assessment and greater confidence in groundwater management decisions under uncertainty.

It should be noted that the OAT sensitivity analysis evaluates the effect of individual variables independently and does not account for potential interactions or nonlinear dependencies among inputs. As a result, the combined influence of multiple variables may be underestimated. Future research could apply global sensitivity analysis methods, such as variance-based approaches, to better capture interaction effects and provide a more comprehensive assessment of system behavior.

Also, we will explore the application of post hoc explanation tools such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations), which have been used to interpret feature contributions in black-box models. These tools will allow us to assess whether the influence of predictors (e.g., irrigation, soil moisture, and inflow) on change in GW storage index aligns with known hydrogeological processes.

5.3. Broader impact

The SD-ML framework offers a transferable modeling strategy for data-scarce regions where traditional groundwater models or SD expertise are unavailable. By embedding socio-hydrological feedback into a lightweight forecasting tool, this approach can support adaptive water management, policy analysis, and decision-making across diverse groundwater-dependent systems.

CRediT authorship contribution statement

Alexander Fernald: Writing – review & editing, Visualization, Validation, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Ilya Zaslavsky:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Gholamreza Eslamifar:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Formal analysis, Data curation. **Vytautas Samalavičius:** Writing – review & editing, Visualization, Investigation, Conceptualization. **Hamid Balali:** Writing – original draft, Validation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding for this research was provided by the Foundation for Food & Agriculture Research (FFAR) under Award No. CA21-SS-0000000084. We extend our sincere appreciation to Saeed Tizpaz Niari, Assistant Professor of Computer Science at the University of Illinois Chicago, for his valuable comments and insightful feedback on our paper.

Data availability

Data will be made available on request.

References

Afzaal, H., Farooque, A.A., Abbas, F., Acharya, B., Esau, T., 2019. Groundwater estimation from major physical hydrology components, artificial neural networks, and learning. Water 12 (1), 5.
Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51 (8), 5957–5973.
Bai, Y., Langarudi, S.P., Fernald, A.G., 2021. System dynamics modeling for evaluating regional hydrologic and economic effects of irrigation efficiency policy. Hydrology 8 (2), 61.
Baki, S., Koutiva, I., & Makropoulos, C. (2012). A hybrid artificial intelligence modelling framework for the simulation of the complete, socio-technical, urban water system. In Proceedings of the 6th International Congress on Environmental Modelling and Software, Leipzig, Germany.
Balali, H., Allen, C.R., Ward, F.A., Eslamifar, G., Fernald, A., 2026. Toward convergence in social-ecological systems resilience assessment: A systematic review and conceptual framework. Ecol. Indic. 185, 114792.
Balali, H., Eslamifar, G., Maxwell, C.M., Fernald, A., 2025. Dynamic simulation modeling for resilience assessment of coupled water–agriculture–community systems in a semi-arid region. Sustainability 17 (9), 4006.
Balali, H., Viaggi, D., 2015. Applying a system dynamics approach for modeling groundwater dynamics to depletion under different economic and climate change scenarios. Water 7 (10), 5258–5271.
Barnett, B., Townley, L., Post, V., Evans, R., Hunt, R., Peeters, L., Boronkay, A., 2012. Australian groundwater modelling guidelines. National Water Commission.
Bates, G., Beruvides, M., Fedler, C.B., 2019. System dynamics approach to groundwater storage modeling for basin-scale planning. Water 11 (9), 1907.
Bliemel, F., 1973. Theil's forecast accuracy coefficient: A clarification. J. Mark. Res. 10 (4), 444–446.
Cai, H., Liu, S., Shi, H., Zhou, Z., Jiang, S., Babovic, V., 2022. Toward improved lumped groundwater level predictions at catchment scale: Mutual integration of water balance mechanism and deep learning method. J. Hydrol. 613, 128495.
Cai, H., Shi, H., Liu, S., Babovic, V., 2021. Impacts of regional characteristics on improving the accuracy of groundwater level prediction using machine learning: The case of central eastern continental United States. J. Hydrol. Reg. Stud. 37, 100930.
Cai, H., Shi, H., Zhou, Z., Liu, S., Babovic, V., 2024. Explaining the mechanism of multiscale groundwater drought events: A new perspective from interpretable deep learning model. Water Resour. Res. 60 (7).
Chen, C., He, W., Zhou, H., Xue, Y., Zhu, M., 2020. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, Northwestern China. Sci. Rep. 10 (1), 1–13.
Coppola Jr, E.A., Rana, A.J., Poulton, M.M., Szidarovszky, F., Uhl, V.W., 2005. A neural network model for predicting aquifer water level elevations. Groundwater 43 (2), 231–241.
Coppola Jr., E., Szidarovszky, F., Poulton, M., Charles, E., 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.
Daliakopoulos, I.N., Tsanis, I.K., Koutroulis, A.G., Kourgialas, N.N., Varouchakis, E.A., Karatzas, G.P., Ritsema, C.J., 2005. Groundwater level forecasting using artificial neural networks. J. Hydrol. 309 (1-4), 229–240.
Davoudi Moghaddam, D., Rahmati, O., Haghizadeh, A., Kalantari, Z., 2020. A modeling comparison of groundwater potential mapping in a mountain bedrock aquifer: QUEST, GARP, and RF models. Water 12 (3), 679.
Driscoll, J.M., & Sherson, L.R. (2016). Variability of surface-water quantity and quality and shallow groundwater levels and quality within the Rio Grande Project Area, New Mexico and Texas, 2009–13 (No. 2016-5006). *U.S. Geological Survey*.
Dupuy, M., Garel, E., Huneau, F., Santoni, S., Di Rosa, M., Mattei, A., 2021. Geochemical and isotope characterisation of thermo-mineral springs of Corsica Island: From geological complexity to groundwater singularity. Water 13 (17), 2413.
Ehteram, M., Ghanbari-Adivi, E., 2023. Self-attention (SA) temporal convolutional network (SATCN)-long short-term memory neural network (SATCN-LSTM): An advanced Python code for predicting groundwater level. Environ. Sci. Pollut. Res. 30 (51), 92903–92921.
Eslamifar, G., Balali, H., Fernald, A., 2024. Fallowing Strategy and Its Impact on Surface Water and Groundwater Withdrawal, and Agricultural Economics: A System Dynamics Approach in Southern New Mexico. Water 16 (1), 181.
Farzana, S.Z., Paudyal, D.R., Chadalavada, S., Alam, M.J., 2024. Decision support framework for water quality management in reservoirs integrating artificial intelligence and statistical approaches. Water 16 (20), 2944.
Fiddaman, T.S. (1997). Feedback complexity in integrated climate-economy models (Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA).
Forrester, J.W., 1970. Urban dynamics. Ind. Manag. Rev.
Forrester, J.W., 1971. World dynamics. Wright-Allen Press.
Freeze, R.A., Massmann, J., Smith, L., Sperling, T., James, B., 1990. Hydrogeological decision analysis: A framework. Groundwater 28 (5), 738–766.
Gao, C., Miller, J.A., 2019. Leveraging simulated data for deep learning in medical imaging. IEEE Int. Conf. Big Data.
Green, T.R., Taniguchi, M., Kooi, H., Gurdak, J.J., Allen, D.M., Hiscock, K.M., Aureli, A., 2011. Beneath the surface of global change: Impacts of climate change on groundwater. J. Hydrol. 405 (3–4), 532–560.
Guo, H.C., Liu, L., Huang, G.H., Fuller, G.A., Zou, R., Yin, Y.Y., 2001. A system dynamics approach for regional environmental planning and management: A study for the Lake Erhai Basin. Environ. Manag. 61, 93–111.
Gurdak, J.J., Hanson, R.T., McMahon, P.B., Bruce, B.W., McCray, J.E., Thyne, G.D., Reedy, R.C., 2007. Climate variability controls on unsaturated water and chemical movement, High Plains aquifer, USA. Vadose Zone J. 6 (3), 533–547.
Haykin, S., 2009. Neural networks and learning machines, 3rd ed.). Pearson Education India.
Herrera, H., Schütz, L., Paas, W., Reidsma, P., Kopainsky, B., 2022. Understanding resilience of farming systems: Insights from system dynamics modelling for an arable farming system in the Netherlands. Ecol. Model. 464, 109848.
Kasiviswanathan, K.S., Saravanan, S., Balamurugan, M., Saravanan, K., 2016. Genetic programming-based monthly groundwater level forecast models with uncertainty quantification. Model. Earth Syst. Environ. 2, 1–11.
Kratzert, F., et al., 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol. Earth Syst. Sci. 23 (12), 5089–5110.
Krishnan, S.R., Nallakaruppan, M.K., Chengoden, R., Koppu, S., Iyapparaja, M., Sadhasivam, J., Sethuraman, S., 2022. Smart water resource management using artificial intelligence—a review. Sustainability 14 (20), 13384.

Kumar, R., Singh, R.D., Sharma, K.D., 2005. Water resources of India. Curr. Sci. 89 (5), 794–811.
Kwakkel, J.H., & Slinger, J.H. (2012). A system dynamics model-based exploratory analysis of salt water intrusion in coastal aquifers. In Proceedings of the 30th International Conference of the System Dynamics Society, St. Gallen, Switzerland.
Kwakkel, J.H., & Timmermans, J.S. (2012). Exploratory system dynamics: A directed search for limits to global water use. In Proceedings of the 30th International Conference of the System Dynamics Society, St. Gallen, Switzerland.
Lallahem, S., Mania, J., Hani, A., Najjar, Y., 2005. On the use of neural networks to evaluate groundwater levels in fractured media. J. Hydrol. 307 (1–4), 92–111.
Langarudi, S.P., Maxwell, C.M., Bai, Y., Hanson, A., Fernald, A., 2019. Does socioeconomic feedback matter for water models? Ecol. Econ. 159, 35–45.
Lättilä, L., Hilletofth, P., Lin, B., 2010. Hybrid simulation models—when, why, how? Expert Syst. Appl. 37, 7969–7975.
Leuthold, R.M., 1975. On the use of Theil's inequality coefficients. Am. J. Agric. Econ. 57 (2), 344–346.
Liu, Y., Sheng, Z., 2011. Trend-outflow method for understanding interactions of surface water with groundwater and atmospheric water for eight reaches of the Upper Rio Grande. J. Hydrol. 409 (3–4), 710–723.
Luo, J., Ma, X., Ji, Y., Li, X., Song, Z., Lu, W., 2023. Review of machine learning-based surrogate models of groundwater contaminant modeling. Environ. Res. 238, 117268.
Mashaly, A.F., Fernald, A.G., 2020. Identifying capabilities and potentials of system dynamics in hydrology and water resources as a promising modeling approach for water management. Water 12 (5), 1432.
Mashaly, A.F., Hanson, A.M., Perez, K.A., Fernald, A.G., 2026. Dynamic Water Budget Modeling to Fill Knowledge Gaps and Improve Water Management in Arid and Semi-Arid Regions. Hydrol. Process. 40 (3), e70443.
McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115–133.
Meadows, D.H., Meadows, D.L., Randers, J., Behrens, W.W.I.I.I., 1972. The limits to growth–Club of Rome. Universe Books.
Medar, R., Rajpurohit, V.S., Rashmi, B., 2017. Impact of training and testing data splits on accuracy of time series forecasting in machine learning. Proceedings of the 2017 International Conference on Computing, Communication, Control, and Automation (ICCUBEA). IEEE, pp. 1–6.
Megdal, S.B., 2018. Invisible water: the importance of good groundwater governance and management. NPJ Clean. Water 1 (1), 15.
Mokari, E., Samani, Z., Heerema, R., Ward, F., 2021. Evaluation of long-term climate change impact on the growing season and water use of mature pecan in Lower Rio Grande Valley. Agric. Water Manag. 252, 106893.
Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.
Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Faybishenko, B., Agarwal, D., 2021. Surrogate optimization of deep neural networks for groundwater predictions. J. Glob. Optim. 81, 203–231.
Nadiri, A.A., Naderi, K., Khatibi, R., Gharekhani, M., 2019. Modelling groundwater level variations by learning from multiple models using fuzzy logic. Hydrol. Sci. J. 64 (2), 210–226.
Nikolenko, S.I. (2021). Synthetic data for deep learning. arXiv preprint arXiv:1909.11512.
NMDA. (2021). New Mexico Agricultural Statistics Annual Bulletin. United States Department of Agriculture, National Agricultural Statistics Service, New Mexico Field Office: Las Cruces, NM.
Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263 (5147), 641–646.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, É., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pulla, S.T., Yasarer, H., Yarbrough, L.D., 2024. Synthetic time series data in groundwater analytics: Challenges, insights, and applications. Water 16 (7), 949.
Rahmandad, H., Akhavan, A., Jalali, M.S., 2025. Incorporating Deep Learning Into System Dynamics: Amortized Bayesian Inference for Scalable Likelihood-Free Parameter Estimation. Syst. Dyn. Rev. 41 (1), e1798.
Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.
Rammohan, B., Partheeban, P., Ranganathan, R., Balaraman, S., 2024. Groundwater Quality Prediction and Analysis Using Machine Learning Models and Geospatial Technology. Sustainability 16 (22), 9848.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743), 195–204.
Rinderknecht, M.D., Klopfenstein, Y., 2021. Predicting critical state after COVID-19 diagnosis: Model development using a large US electronic health record dataset. NPJ Digit. Med. 4 (1), 113.
Sahoo, S., Russo, T.A., Elliott, J., Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. Water Resour. Res. 53 (5), 3878–3895.
Schwaninger, M., 2006. System dynamics and the evolution of the systems movement. Syst. Res. Behav. Sci. 23, 583–594.
Sheng, Z., 2013. Impacts of groundwater pumping and climate variability on groundwater availability in the Rio Grande Basin. Ecosphere 4 (1), 1–25.
Shiri, N., Shiri, J., Yaseen, Z.M., Kim, S., Chung, I.M., Nourani, V., Zounemat-Kermani, M., 2021. Development of artificial intelligence models for well groundwater quality simulation: Different modeling scenarios. PLOS ONE 16 (5), e0251510.
Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48.
Sterman, J.D., 1984. Appropriate summary statistics for evaluating the historical fit of system dynamics models. Dynamica 10 (2), 51–66.
Sterman, J.D., 2000. Business dynamics: Systems thinking and modeling for a complex world. McGraw-Hill.
Sterman, J.D., 2008. Risk communication on climate: Mental models and mass balance. Science 322, 532–533.
Subagadis, Y.H., Grundmann, J., Schütze, N., Schmitz, G.H., 2014. An integrated approach to conceptualize hydrological and socio-economic interaction for supporting management decisions of coupled groundwater–agricultural systems. Environ. Earth Sci. 72, 4917–4933.
Sun, Y., Liu, N., Shang, J., Zhang, J., 2017. Sustainable utilization of water resources in China: A system dynamics model. J. Clean. Prod. 142, 613–625.
Taccari, M.L., Wang, H., Goswami, S., De Florio, M., Nuttall, J., Chen, X., Jimack, P.K., 2024. Developing a cost-effective emulator for groundwater flow modeling using deep neural operators. J. Hydrol. 630, 130551.
Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Salim, H., Sungwon, K., Allawi, M.F., 2022. Groundwater level prediction using machine learning models: A comprehensive review. Neurocomputing 489, 271–308.
Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., Treidel, H., 2013. Groundwater and climate change. Nat. Clim. Change 3 (4), 322–329.
Wada, Y., Van Beek, L.P., Van Kempen, C.M., Reckman, J.W., Vasak, S., Bierkens, M.F., 2010. Global depletion of groundwater resources. Geophys. Res. Lett. 37 (20).
Wang, G., Wang, S., Kang, Q., Duan, H., Wang, X.E., 2016. An integrated model for simulating and diagnosing the water quality based on the system dynamics and Bayesian network. Water Sci. Technol. 74, 2639–2655.
Winz, I., Brierley, G., Trowsdale, S., 2009. The use of system dynamics simulation in water resources management. Water Resour. Manag. 23, 1301–1323.
Xi, X., Poh, K.L., 2015. A novel integrated decision support tool for sustainable water resources management in Singapore: Synergies between system dynamics and analytic hierarchy process. Water Resour. Manag. 29, 1329–1350.
Yang, H., Li, X., Ma, L., Li, Z., 2021. Using system dynamics to analyse key factors influencing China's energy-related CO2 emissions and emission reduction scenarios. J. Clean. Prod. 320, 128811.
Zhan, Y., Guo, Z., Yan, B., Chen, K., Chang, Z., Babovic, V., Zheng, C., 2024. Physics-informed identification of PDEs with LASSO regression: Examples of groundwater-related equations. J. Hydrol., 131504
Zhao, Z., Song, Y., & Ermon, S. (2020). Sim-to-real transfer in deep reinforcement learning for robotics: A survey. arXiv preprint arXiv:2006.06858.

Zhou, T., Wang, F., Yang, Z., 2017. Comparative analysis of ANN and SVM models combined with wavelet preprocess for groundwater depth prediction. Water 9 (10), 781.
Zolfagharian, M., Romme, A.G.L., Walrave, B., 2018. Why, when, and how to combine system dynamics with other methods: Towards an evidence-based framework. J. Simul. 12, 98–114.
Zomorodian, M., Lai, S.H., Homayounfar, M., Shaliza, I., Pender, G., 2017. Development and application of coupled system dynamics and game theory: A dynamic water conflict resolution method. PLoS ONE 12, e0188489.