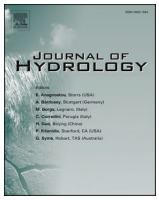
Journal of Hydrology 628 (2024) 130535

Contents lists available at ScienceDirect

Journal of Hydrology

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



Research papers

Reconstructing high-resolution groundwater level data using a hybrid random forest model to quantify distributed groundwater changes in the Indus Basin

Arfan Arshad [a,b], Ali Mirchi [a,*], Javier Vilcaez [c], Muhammad Umar Akbar [a,d], Kaveh Madani [e]

[a] Department of Biosystems and Agricultural Engineering, Oklahoma State University, Stillwater, OK, USA
[b] Department of Irrigation and Drainage, Faculty of Agricultural Engineering and Technology, University of Agriculture Faisalabad, Faisalabad, Pakistan
[c] Boone Pickens School of Geology, Oklahoma State University, Stillwater, OK 74078, USA
[d] Department of Structures and Environmental Engineering, Faculty of Agricultural Engineering and Technology, University of Agriculture Faisalabad, Faisalabad, Pakistan
[e] United Nations University Institute for Water, Environment and Health (UN-INWEH), Hamilton, Ontario, Canada

ARTICLE INFO

Keywords:
Groundwater level
GRACE
machine learning (ML)
Geostatistical method
Gap-filled data
Local covariates
Indus Basin

ABSTRACT

High-resolution, continuous groundwater data is important for place-based adaptive aquifer management. This information is unavailable in many areas due to spatial sparsity of and temporal gaps in groundwater monitoring. This study advances the ability to generate high-resolution (1 km^2), temporally continuous estimates of groundwater level (GWL) changes by incorporating 1 km^2 covariates and existing piezometer observations into predictive modeling. We employed a hybrid machine learning (ML) model, primarily using the geographically weighted random forest (RF_{gw}) model. To assess the performance of the RF_{gw} model, we conducted a comprehensive comparison with the SGS geostatistical method and non-spatial ML models (RF and XGBoost). The framework was implemented across the Indus Basin using biannual (July and Oct) GWL data from piezometers and local covariates from 2003 to 2020. The RF_{gw} model demonstrated superior accuracy in predicting GWLs, improving R^2 by 10 %, 17 %, and 22 % compared to SGS, RF, and XGBoost, respectively. Notably, SGS, RF, and XGBoost substantially underestimated the GWL in deeper wells (7–11 m), whereas RF_{gw} showed a much smaller underestimation (up to ~ 3 m). The 90 % prediction interval revealed that RF_{gw} had less uncertainty (1–3 m) followed by RF (2–5 m), and SGS and XGBoost (up to 8 m) for most testing piezometers. Incorporating high-resolution covariates into RF_{gw} predictive modeling provided reliable estimates of GWL changes for unmonitored sites. Using the reconstructed GWL data, we examined the GWL changes in head (i.e., upstream) and tail (i.e., downstream) farms within canal distributaries, illustrating faster groundwater drawdown in tail farms (e.g., 0.82 m/yr) than head farms (0.02 m/yr in the Hakra canal distributary). Densely populated urban areas (e.g., Lahore, Multan, and Faisalabad) had the highest GWL decline (e.g., up to 0.9 m/yr). The framework can be used in other groundwater-stressed regions to support better aquifer management in the face of limited in-situ observations.

1. Introduction

Groundwater is a critical natural resource that supports various human activities and ecological functions (Aeschbach-Hertig and Gleeson, 2012). Groundwater availability is threatened by unsustainable use and climate change (Cuthbert et al., 2019; Kaur et al., 2021; Ostad-Ali-Askari et al., 2020). Overexploitation of groundwater for irrigation and domestic uses is causing groundwater decline in aquifers worldwide (Asoka et al., 2017; Ostad-Ali-Askari and Shayannejad, 2021; Rodell et al., 2018). In particular, vast areas of the transboundary Indus Basin are facing mounting groundwater stress due to increasing agricultural and urban water demands coupled with shrinkage of available water resources (Smolenaars et al., 2022; Watto and Mugera, 2016; Mehmood et al., 2022). The Indus Basin has witnessed a substantial increase in

* Corresponding author.
E-mail addresses: aarshad@okstate.edu (A. Arshad), amirchi@okstate.edu (A. Mirchi), vilcaez@okstate.edu (J. Vilcaez), umar.akbar@okstate.edu (M. Umar Akbar), keveh.madani@unu.edu (K. Madani).

https://doi.org/10.1016/j.jhydrol.2023.130535

Available online 21 November 2023
0022-1694/Published by Elsevier B.V.

groundwater dependency, leading to a rapid decline in groundwater level (GWL) with an average depletion rate of ~ 0.4 m/yr in the last two decades (Akhtar et al., 2022; MacDonald et al., 2016; Nourani et al., 2023). Effective management of groundwater resources requires understanding of aquifer dynamics and identifying hotspots of groundwater depletion (Du et al., 2020; Ostad-Ali-Askari et al., 2019). Long-term, continuous data about GWL changes facilitates the development of adaptive management plans by illuminating the hydrological behavior of the aquifer under anthropogenic impacts and climate variability (Saito et al., 2021; Stewart, 2015).

Groundwater monitoring initiatives have been undertaken by several water system authorities in the Indus Basin, Pakistan including SMO (Scarp Monitoring Organization), WASA (Water and Sanitation Agency), and DLR (Directorate of Land Reclamation) (Bhatti et al., 2017; Lytton et al., 2021). Previous studies have utilized the piezometer data to evaluate groundwater changes in the Indus Basin (Abbas et al., 2023; MacDonald et al., 2016; Sajjad et al., 2022). Nevertheless, critical knowledge gaps persist because evaluations of groundwater changes have been limited to specific regions due to inadequate and uneven distribution of piezometers. Consequently, assessments are unavailable for many areas facing acute water stress, including urban areas, which often lack extensive groundwater monitoring. Furthermore, defective/discontinuous piezometer monitoring, and the addition of new piezometers at different points in time have resulted in data gaps across both space and time (Basharat et al., 2014; MacAllister et al., 2022). The presence of these data gaps can lead to incomplete and inaccurate assessments of local to regional groundwater resources, which can hamper adaptation decisions. There is an important need to leverage the predictive modeling tools to generate high-resolution, continuous groundwater data to facilitate sustainable water management plans.

Numerous techniques, including physically-based simulations (Deng and Bailey, 2020; Singh, 2014; Talebmorad and Ostad-Ali-Askari, 2022) and data-driven modeling (Sun et al., 2022; Zanotti et al., 2019) have been used to estimate groundwater changes over time and across space. Physically-based hydrological models generally have better reliability and higher accuracy to simulate groundwater dynamics for small-scale applications (Aliyari et al., 2019). However, predicting groundwater changes with high accuracy using physically-based models is challenging in large, complex hydrological systems due to over-simplification of the system and hydrological processes (e.g., aquifer properties and appropriate initial and boundary conditions) and limited availability of input data (Ahmed et al., 2020; Singh, 2014; Zeydalinejad, 2022). Data-driven modeling has emerged as a promising alternative to physically-based models owing to the simplicity in design independent of hydrogeological characteristics of the aquifer and applicability in small to large basins (Wunsch et al., 2018; Yin et al., 2021; Zanotti et al., 2019). Two widely used data-driven approaches to groundwater modeling include: (1) linear methods [e.g., weighted average (Pappas et al., 2014), autoregressive integrated moving average (Oikonomou et al., 2018), and geostatistical methods (Yang and Xing, 2021)]; and (2) non-linear machine learning (ML) methods [e.g. random forest (RF) (Koch et al., 2019; Tang et al., 2019) and artificial neural networks (ANNs) (Collados-Lara et al., 2023; Ostad-Ali-Askari et al., 2017; Zanotti et al., 2019)].

Linear methods are widely used because of their simplified algorithms and fast computational processing (Chung et al., 2019; Sahoo et al., 2017; Valipour et al., 2013). For example, kriging is a powerful linear geostatistical approach for predicting values between data points when certain fundamental conditions of the underlying random function are met (Cui et al., 2016). However, kriging does not offer information on the probability of points exceeding specific data values or thresholds. This leads to a tendency to smoothen the output (Varouchakis et al., 2012), resulting in conditional bias (i.e., overestimation of small values and underestimation of larger ones) in kriged values for un-sampled sites (Karami et al., 2022; Varouchakis et al., 2012). Conditional geostatistical simulation methods such as sequential Gaussian simulation (SGS), a stochastic approach, was developed to overcome this smoothing effect by generating multiple, equally probable stochastic realizations, which effectively preserve local variability in the data (Manchuk and Deutsch, 2012; Mariethoz and Caers, 2014).

Non-linear ML approaches have gained interest in groundwater modeling due to their efficient handling of non-linearity and dynamic behavior of earth system variables (Chen et al., 2020; Sun, 2013; Tao et al., 2022). For example, Chen et al. (2020) simulated groundwater storage changes in the Heihe River Basin using ML algorithms (multilayer perceptron (MLP) and support vector machine (SVM)). They found that ML algorithms performed better than numerical models. Chen et al. (2019) used ANNs to predict long-term groundwater storage variations in the Songhua River Basin by using remote sensing data and in-situ observations. The applications of ANN, RF and XGBoost algorithms for groundwater level prediction have been widely reported in other recent studies (e.g., Rahman et al., 2020; Tao et al., 2022). Among these, RF algorithms have proven to be a robust, efficient regressor that can outperform other ML methods (Zi–chen et al., 2021). RF is a tree-based non-parametric, data-driven model (Breiman, 2001a), which handles multi-collinearity in predicting non-linear functions (Luo et al., 2021). Despite the strengths of RF in predicting groundwater dynamics (Hengl et al., 2018; Koch et al., 2019; Wang et al., 2018), it has some limitations due to its non-spatial, global nature. It ignores the spatial dependencies among the observations which can potentially result in biased predictions (Hengl et al., 2018). One way to overcome this limitation is to use hybrid algorithms that combine classical RF models and soft computing methods (Sahoo et al., 2017), enhancing modeling precision and adaptability of the classic model (Zhang, 2003). Herein, we employed a geographically weighted random forest (RF_{gw}) model, which is a hybrid form of the global RF model (Georganos et al., 2021). The RF_{gw} model combines a spatial weighting matrix with the global RF algorithm using the distance-decay kernel function and a bandwidth parameter to spatially weight the observations (Fotheringham et al., 2003). Thus, the approach allows local calibration of the RF model by fitting sub-models in space, obeying the working principle of geographically weighted regression (GWR) (Fotheringham et al., 2003; Georganos and Kalogirou, 2022).

ML models have been widely used to fill temporal gaps in groundwater observations, and hindcast and forecast groundwater changes (Oikonomou et al., 2018; Pham et al., 2022; Rahman et al., 2020; Zanotti et al., 2019). Previous studies trained ML models to identify temporal patterns concealed in historical groundwater data, and subsequently used those patterns to predict groundwater changes for desired time steps (e.g., Rahman et al., 2020). These models generally performed well when trained with a certain amount of data (Bailly et al., 2022; Taie Semiromi and Koch, 2019) that allowed capturing the complexity of the time series by building sufficient decision trees (Breiman, 2001a; Jha et al., 2018). However, the temporal frequency of piezometer monitoring in the Indus Basin, often biannual (MacAllister et al., 2022), is inadequate for effectively training and constructing precise GWL estimates using temporal predictive modeling. Furthermore, temporal predictive modeling using ML methods often encounters challenges when it comes to providing estimations for locations that lack monitoring data. ML models are not inherently capable of extrapolating beyond the spatial and temporal bounds of their training data (Brenning, 2023; Takoutsing and Heuvelink, 2022). It has been shown that decision-tree based ML methods can be ineffective when applied to regions outside the scope of their training data (Hengl et al., 2018). As a result, previous studies typically focused on only a handful of monitoring wells to predict groundwater changes (Pham et al., 2022; Sun et al., 2022; Wunsch et al., 2018; Yin et al., 2021), limiting the utility of this approach for large-scale applications that involve unmonitored sites.

This paper advances the ability to generate high-resolution (1 km^2), continuous estimates of GWL changes by incorporating 1 km^2 covariates and existing piezometer observations into predictive modeling. To overcome the constraints imposed by limited training data through time,

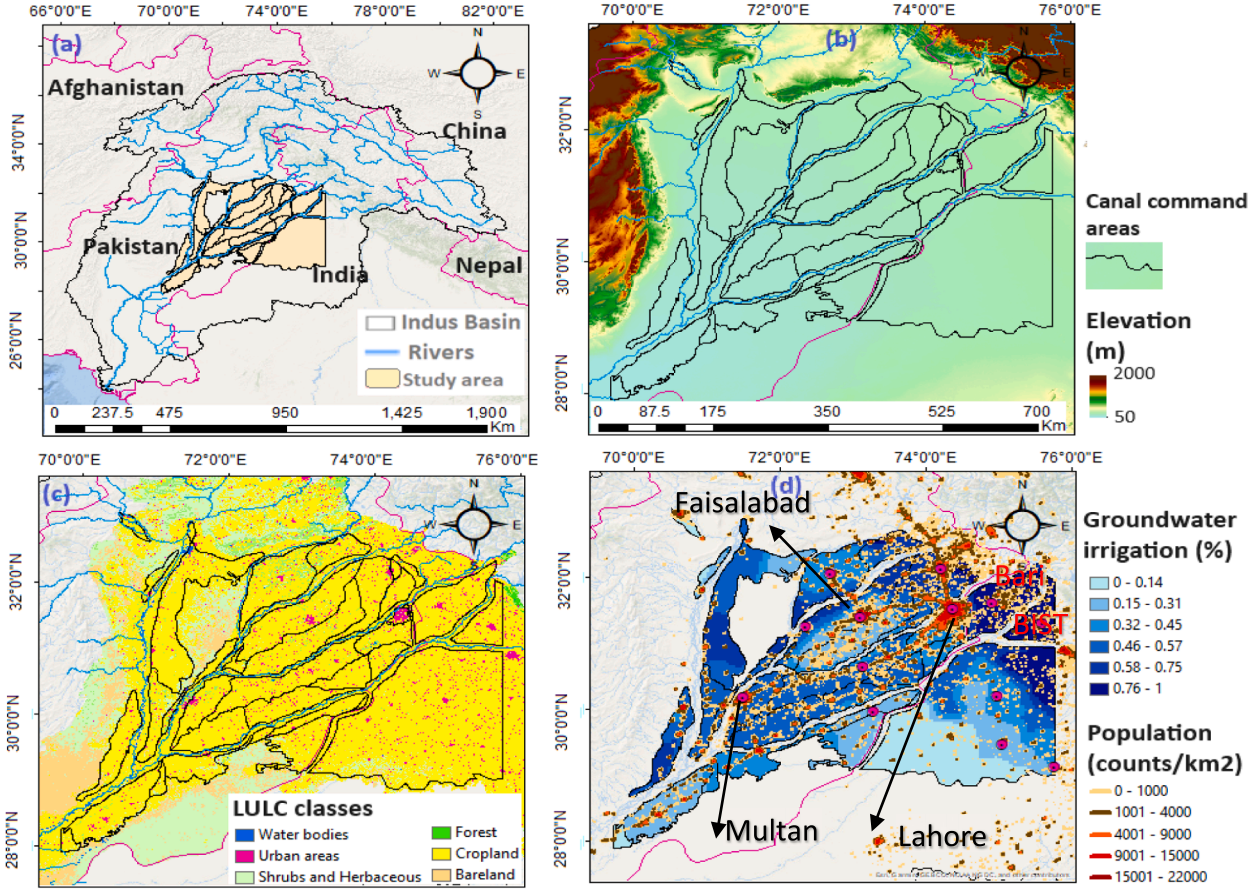


Fig. 1. Map and key attributes of the study area in the Indus Basin: (a) geographical scope, (b) topography, (c) major land use/land cover classes, and (d) groundwater irrigation hotspots and population density.

we introduce a novel framework that employs spatial ML-based predictive modeling iterated over time using a hybrid RF_{gw} model. Instead of relying solely on temporal data and traditional modeling approaches, the framework spatially trains the models at each piezometer based on local covariates and groundwater data from neighboring piezometers. Subsequently, spatial patterns mapped at training piezometers are integrated with gridded (1 km^2) local covariates data to provide high-resolution estimates of GWL changes. Incorporating the high-resolution covariates into predictive modeling facilitates the estimation of GWL changes at unmonitored sites. The primary objectives of this study were threefold: (1) to evaluate the performance of a spatial ML method (RF_{gw}) in comparison to a geostatistical method (SGS) and non-spatial ML models (RF and XGBoost) to predict GWL changes; (2) to reconstruct high-resolution (1 km^2), biannual estimates of GWL changes across the Indus Basin from 2003 to 2020, and (3) to utilize the reconstructed GWL data to assess distributed groundwater changes in unmonitored groundwater stress hotspots. The framework is transferable to other groundwater-stressed data-sparse regions around the world since it can be applied independent of aquifer characteristics.

2. Material and methods

2.1. Study area

The study area occupies ~ 151,970 km^2 in the central part of the Indus Basin, including agro-urban settings (Fig. 1a). Arid and semi-arid continental climatic conditions dominate the region. Annual rainfall is 300–900 mm with monsoon period (Jun-Aug) contributing 70 % of the total rainfall. Five main rivers (Indus, Jhelum, Satlej, Ravi, and Chenab) flow through the study area. The main canals and distributaries receive water from the main rivers and supply water for irrigation in thirty-two canal command areas (Fig. 1a). The elevation within the study area ranges from 2000 m (a.m.s.l.) in the western and northern parts to 50 a.m.s.l. in vast agricultural areas (Fig. 1b). The aquifer is unconfined in nature with a single layer and mainly composed of sand, silt, gravel, and other combinations of rock materials (Umar et al., 2022). The major land use/land cover classes include cropland, urban land, shrubs and forest (Fig. 1c). The region includes pockets of populous metropolitan areas. For example, population density can reach 20,000 capita/km^2 in major cities such as Lahore, Multan, and Faisalabad (Fig. 1d). Groundwater is the main source of water supply for domestic purposes and irrigated agriculture and continuous pumping has caused groundwater to decline by 0.4 m/yr in the last two decades (Arshad et al., 2022; Watto et al., 2021). Groundwater withdrawal for irrigation varies among different canal command areas with higher quantities of pumping reported in the upstream canal command areas (e.g., Bist Doab, Upper Bari) (Fig. 1d).

2.2. Input data

We compiled input data from a variety of sources, including ground-based observations, remote sensing products and previously published data.

Groundwater data from piezometers: As a proxy for GWL, we compiled depth to groundwater data from 3,000 piezometers provided by Punjab Irrigation Department (PID), Pakistan and 969 piezometers by Indian Water Resource Information System (IWRIS) from 2003 to 2020 (Fig. 2a). The data consist of biannual observations (June: pre-monsoon

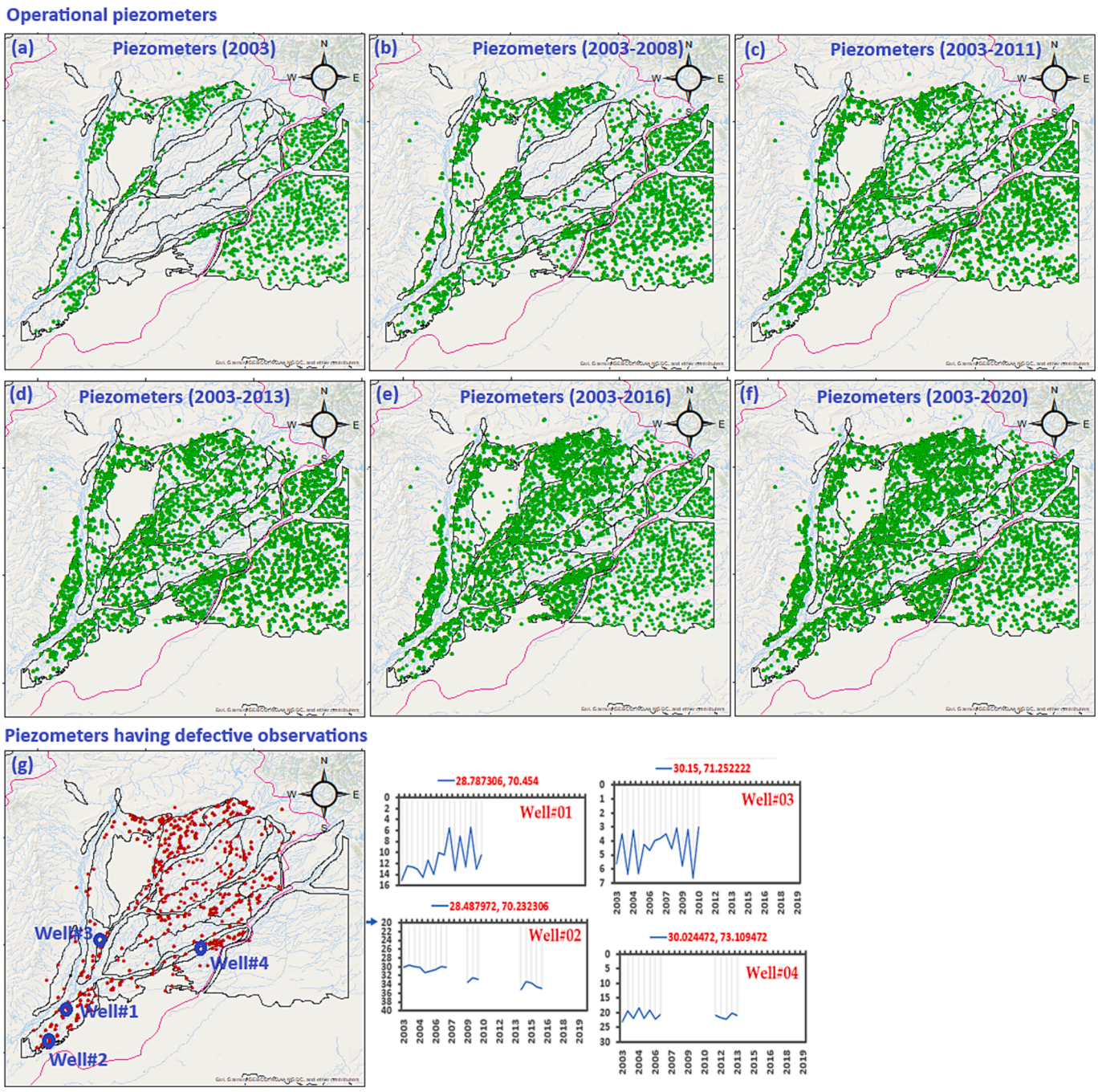


Fig. 2. Spatial distribution of piezometers monitoring groundwater changes: (a-f) operational piezometers, including piezometers that were continuously operating and newly added piezometers between 2003 and 2020, and (g) piezometers with defective observations in different periods between 2003 and 2020 and four examples of discontinuous piezometer records (Well 1 through 4).

and October: post monsoon). A drop in the piezometric water levels represents the change in GWL with reference to the ground surface. We performed a data quality check to identify the total number of wells, wells with complete data, null data, and defective observations (Fig. S1 in Supplementary Material (SM)). In 2003, there were only 753 piezometers with complete data, which gradually increased to 2,820 piezometers by the year 2014 and then decreased to 1,728 in 2020 (Fig. S1 in SM). A larger number of new observations were recorded between 2008 and 2013 at piezometers that were operational, particularly in central Punjab in Pakistan. However, these piezometers lacked complete historical data (Fig. 2b-d). Likewise, continuous biannual GWL data were unavailable for 659 monitoring wells that were defective/discontinued in different periods of time between 2003 and 2020 (Fig. 2g).

Groundwater-related local covariates: We collected monthly (June and October) data for six local covariates (i.e., land surface temperature (LST), normalized difference vegetation index (NDVI), actual evapotranspiration (ET_a), population density (PD), night light (NL), and precipitation (PRECIP)) at high-resolution (1 km^2) in the 2003–2020 period (Fig. S2 in SM). We selected these covariates as indicators of natural or anthropogenic influence on groundwater changes. Socioeconomic development (e.g., increasing population) and vegetation growth are among drivers of increasing water demand, contributing to groundwater depletion (Arshad et al., 2022; Arshad et al., 2019;

Table 1
Summary of variables, their spatiotemporal resolution, and data sources used in this study.

Data Type	Variables	Product	Spatial (temporal) resolution	Sources
Groundwater storage and depletion	GWSA, DEPgw	GRACE JPL-SH (Downscaled)	1 km × 1 km (monthly)	(Arshad et al., 2022) https://doi.org/10.6084/m9.figshare.22301020.v5
Local covariates	Precipitation	TRMM (downscaled)	1 km × 1 km (monthly)	(Noor et al., 2023; Arshad et al., 2023b) https://doi.org/10.6084/m9.figshare.24570397.v4
	ET_a	MODIS	1 km × 1 km (monthly)	https://earlywarning.usgs.gov/fews/product/460
	NDVI	MOD13A3	1 km × 1 km (monthly)	https://lpdaac.usgs.gov/dataset_discovery/modis
	LST	MOD11A2	1 km × 1 km (8 days)	NASA Land Processes Distributed Active Archive Center
	Population density	WorldPop Global High-Resolution Population dataset	1 km × 1 km (yearly)	https://hub.worldpop.org/geodata/listing?id = 64
	NL	DMSP/OLS	0.008 ^{0} × 0.008 ^{0} (yearly)	https://figshare.com/articles/dataset/Harmonization_of_DMSP_and_VIIRS_nighttime_light_data_from_1992-2018_at_the_global_scale/9828827/6
Groundwater	GWL changes reference to ground surface	Piezometers observations	Stations (Biannually: June and October)	PID (Punjab Irrigation Department) and Indian Water Resource Information System (IWRIS)

Kamaran Dastjerdi et al., 2022). While monsoon precipitation helps recharge the aquifer, high evapotranspiration and human water demand in dry periods contribute to falling groundwater (Ahmed and Wiese, 2019; Basharat et al., 2014). More detail about data acquisition, sources and preprocessing of local covariates are provided in TextS1 (See SM) and Table 1.

Gravity Recovery and Climate Experiment (GRACE): High-resolution groundwater storage anomalies (GWSA) and depletion (DEPgw) data at 1 km^2 resolution were acquired from a previous study (Arshad et al., 2022; Arshad et al., 2023a). The data set was produced using data from the Gravity Recovery and Climate Experiment (GRACE) by applying a spatial downscaling approach. The data was validated with in-situ observations (overall R^2 = 81 %) between 2003 and 2020 for the irrigated Indus Basin (see Arshad et al., 2022).

2.3. Description of spatial methods

2.3.1. Sequential geostatistical simulation

The SGS method was used to predict the changes in GWL for unmonitored sites and piezometers that had missing data. SGS works based on the principal assumption of multi-Gaussian random function variable (Delbari et al., 2009). We defined a regularly spaced 1 km^2 grid that covered the study area. The SGS establishes a random path function through all space grid nodes such that each node is visited once in the sequence (Nussbaumer et al., 2018), thus producing multiple possible ensemble realizations of GWL distribution based on the original piezometer data (Manchuk and Deutsch, 2012). Since selecting a fixed number of realizations is debatable, the experiment should be repeated with many iterations until the variance of simulations stabilizes (Metahni et al., 2019). In our case, we selected 250 realizations. Before performing the SGS, we transformed the data into a normal distribution using Gaussian anamorphosis transformation. Thereafter, the transformed data were used to calculate the experimental semi-variogram models with hyperparameters consisting of sill, range and a nugget effect. The simple kriging was performed on normally distributed data to provide an estimate and variance at each location. Furthermore, a conditional simulation step was performed using SGS on a grid size of 1 km^2. Conditional realizations were produced using the same GWL used in the semi-variogram models to ensure that the generated realizations respect the observed values at the sampled locations. The results were back transformed to obtain simulated output in the original units. More detail bout SGS theory and algorithms can be found in Bai and Tahmasebi (2022).

2.3.2. Machine learning methods

We used three ML methods to predict GWLs, namely non-spatial models including extreme gradient boosting (XGBoost), and random forest (RF), and the spatial random forest (geographical weighted random forest (RF_{gw})). These methods are explained as follows:

XGBoost method: We used XGBoost, an ensemble ML approach rooted in the gradient boost tree algorithm. XGBoost builds an ensemble of decision trees, each dependent on the others. XGBoost effectively manages multicollinearity effects and handles complex non-linear relationships within the data. Mathematically, XGBoost prediction can be expressed as:

$$Y_i = Y_i + \sum_{t=1}^{T} h_t(X_i) \quad (1)$$

Y_i is the current prediction of GWL for the i^{th} observation. The predicted GWLs for each observation is updated iteratively through multiple boosting rounds (T), where each round contributes the output of a decision tree (h_t) based on the corresponding feature vector (X_i). The final prediction is the sum of predictions from all boosting rounds. This process allows XGBoost to improve predictions by focusing on the samples that were previously mispredicted.

The hyperparameters of XGBoost regression involve the learning rate, tree depth, the number of boosting rounds, minimum child weight, and gamma, among others. We used an open source "XGBoost" package available in R programming (Chen and Guestrin, 2016).

RF method: The RF machine learning method accounts for the non-linear relationships among variables through independent decision trees (Breiman, 2001b). RF bootstraps data randomly and develops multiple decision trees to make predictions using sub-regression models on each tree. RF helps to avoid overfitting the training dataset and outperforms typical decision tree methods (He et al., 2016). The RF model developed in this study can be expressed as follows:

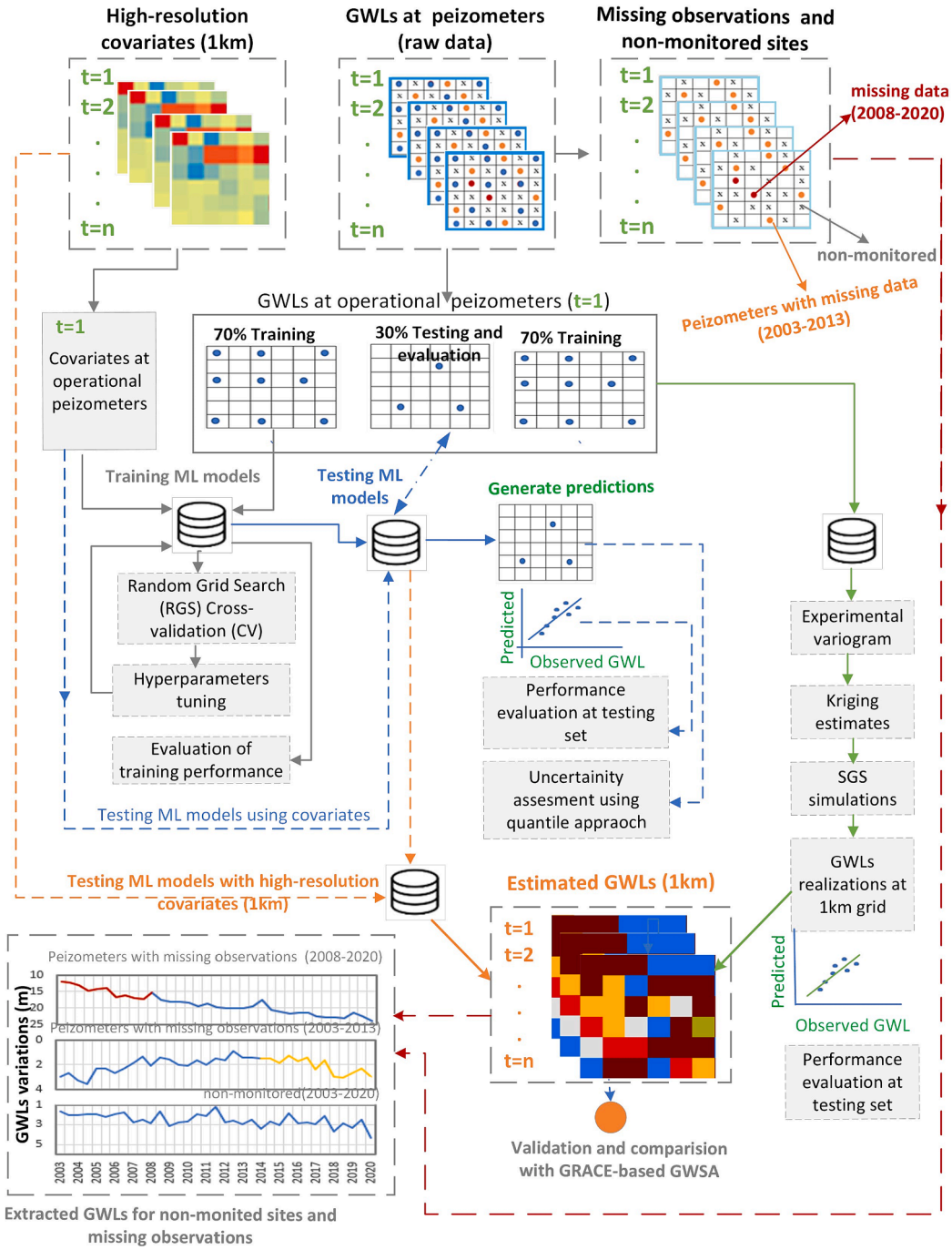


Fig. 3. Schematic diagram to reconstruct high-resolution GWLs, extending estimations to locations with missing observations and unmonitored sites using SGS and ML methods (RF, XGBoost and RF_{gw}). Cells indicate the grid resolution of reconstructed GWLs (i.e., 1 km^2). Models were trained and tested for one time step "t" in each iteration.

$$Y_i = \frac{1}{T}\sum_{t=1}^{T} h_t(X_i) \quad (2)$$

Each decision tree (h_t) in the random forest model provides an individual prediction for each GWL observation based on the corresponding feature vector (X_i). The final prediction is obtained by averaging the predictions from all decision trees (T), resulting in an ensemble prediction that typically offers better generalization and accuracy compared to a single decision tree.

The hyperparameters of a RF regression are mtry "number of variables randomly sampled", ntress "number of trees to grow", and node size" minimum number of observations in a terminal node". The "randomForest" in R is used to perform analysis (Breiman, 2001b).

RF_{gw} method: The XGBoost and RF methods are non-spatial, global models that do not address spatial heterogeneity. The RF_{gw} (or local RF) model utilizes a local version of the RF regression proposed by Georganos et al. (2021). The principle working mechanism of the RF_{gw} model is based on geographically weighted regression (GWR) which uses neighboring data points to weight the observations geographically and fit sub-models at each location. Thus, the RF_{gw} model performs local rather than global estimations. The fitted local sub-models in the RF_{gw} framework help to address the spatial heterogeneity, and outperform non-spatial models, including RF (Fotheringham et al., 2003; Georganos et al., 2021; Georganos and Kalogirou, 2022). The general concept of the

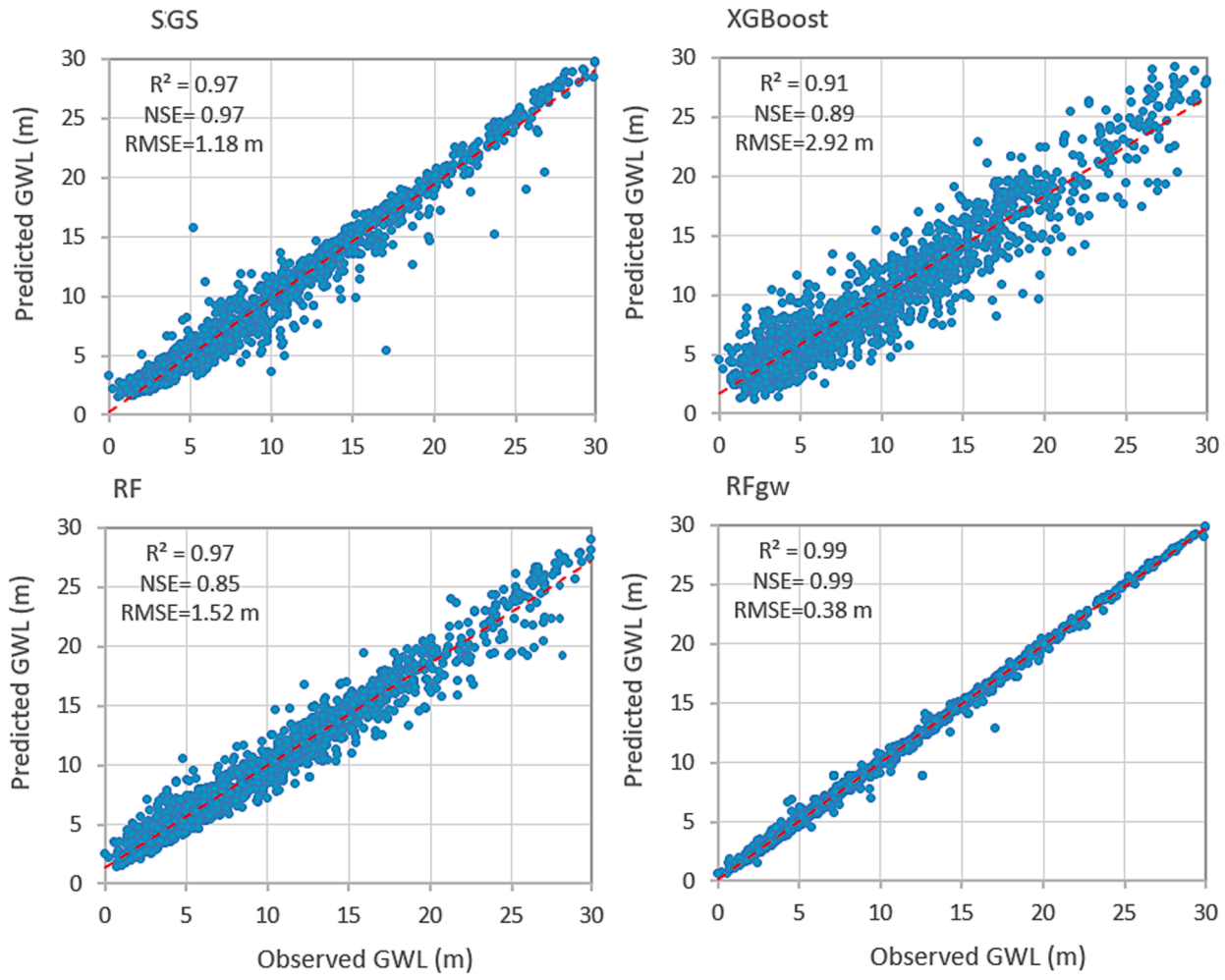


Fig. 4. Comparison of model predicted and observed GWL at training piezometers.

RF_{gw} method can be explained based on simple regression and RF:

$$Y_i = aX_i + e, i = 1 : n \quad (3)$$

Y_i is the GWL for the i^{th} piezometer's observation. aX_i is the global prediction of RF based on X set of covariates (e.g., ET, NDVI, NL, etc.) and e indicates the prediction error. In the RF_{gw} framework, the global RF model is calibrated at each location "i" by considering a local subset of spatial observations within a neighborhood defined by spatial weights. The RF_{gw} method used here can be expressed using the following equation:

$$Y_i = a(u_i, v_i, W)X_i + e, i = 1 : n \quad (4)$$

where, $a(u_i,v_i, W)X_i$ matrix indicated the prediction of RF model calibrated at the i^{th} operational piezometer having geographical coordinates (u_i,v_i) and X set of covariates; u_i, and v_i represent the latitude and longitude, respectively, of the i^{th} observation; and e is the prediction error in the modeled GWL. Spatial weight matrix (SWM) is constructed to spatially weight observations in the RF_{gw} model. The spatial weights (W) for each location are calculated using adaptive Gaussian method as follows:

$$W_{ij} = exp\left[-d_{ij}^2/\theta_{i(k)}^2\right] \quad (5)$$

where, W_{ij} is the spatial weights between location i and j, d_{ij} is the distance between i and j, $\theta_{i(k)}$ is an adaptive bandwidth kernel indicating the k^{th} number of nearest observations near location i.

During the development of a local RF_{gw} model, the number of neighborhood points (or kernel) plays an important role in making reliable predictions. The local RF model uses a certain number of neighborhood observations to train the model. The maximum distance between the kernel and training data point indicates the bandwidth. We used an adaptive kernel to calculate the "k" number of nearest observations. The hyperparameters for RF_{gw} model are ntree,mtry, bandwidth (bw).

2.4. Hyperparameters tuning and optimization

Hyperparameters tuning is performed to select the best combination of hyperparameters. The RGS (Random Grid Search) method was used as tuning strategy for hyperparameters tuning of XGBoost (learning rate, tree depth, the number of boosting rounds, eta and gamma), and RF (ntree, mtry, and nodesize) (Bergstra and Bengio, 2012). We specified a predefined range of values for each hyperparameter, and used the search algorithm to randomly sample combinations of these values (Jiménez et al., 2008). For each combination, the model's performance was evaluated using cross-validation (CV). Thereafter, we identified the hyperparameters values that resulted in the best model performance on unseen data. In the case of the RF_{gw} model, the out-of-bag (OOB) method was applied to attain the optimal value of bandwidth (Georganos and Kalogirou, 2022). We tested OOB accuracy with various bandwidths and selected the optimal bandwidth with the best performance. The optimal bandwidth along with other optimized hyperparameters (mtry, and ntrees) of the RF model were used to train and test the local RF_{gw} model. In the case of SGS, the variogram model's parameters for kriging estimations were selected during the leave-one-out cross-validation process

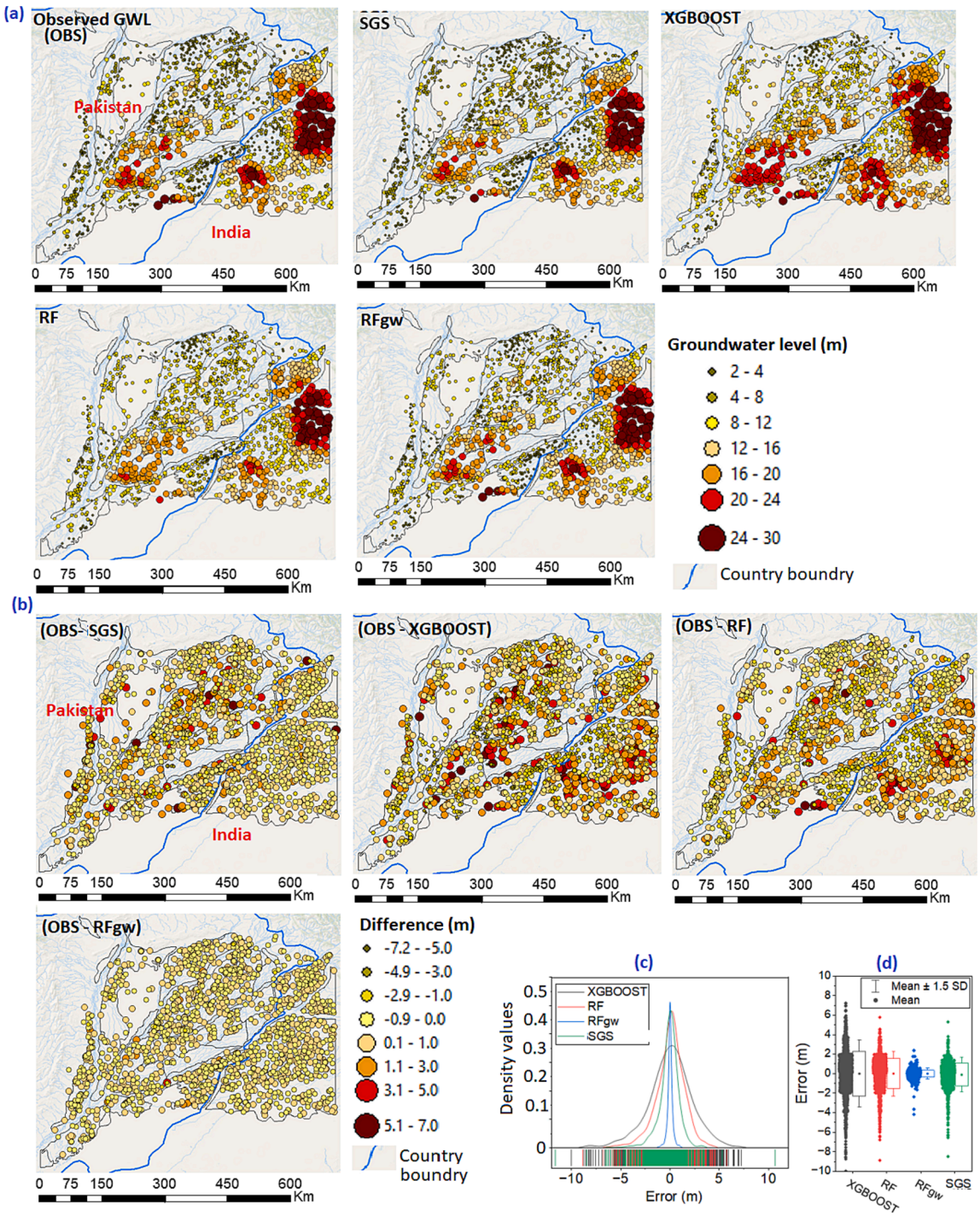


Fig. 5. (a) GWL predictions generated by the four models (SGS, XGBOOST, RF and RF_{gw}) and comparison with observed (OBS) GWL at training piezometers, (b) model prediction error based on the OBS, and (c-d) probability distribution function and error bars indicating the bounds of the models' prediction error.

(Asa et al., 2012). This method involves iteratively leaving out one observation, predicting it based on neighboring data, and calculating prediction errors ('predicted - observed value'). Thus, the leave-one-out cross-validation helps select optimal variogram model parameters (e.g., searching neighborhood and semi-variogram parameters) while ensuring best predictions with minimum RMSE.

2.5. Assumptions and general framework for reconstructing GWL

The proposed framework utilizes two different approaches and four models to reconstruct the high-resolution data, providing distributed (1 km^2) estimates of GWL changes for temporally missing observations at monitored and unmonitored sites. The approaches (and models) are geostatistical (SGS) and ML (XGBoost, RF and RF_{gw}). The specific steps to reconstruct the GWLs for the study area are summarized as follows:

(1) Preparing GWL data and covariates: The biannual (June and Oct) GWL data from piezometers (P; predictand) and covariates were structured in a matrix D_P for a specific month (t) represented as $D_P = \left[\left(Y \times X_i \times P_{J(u \times v)}\right), i \in 1,2,3,..,N; j \in 1,2,3,..,k\right]_t$, where, Y represents predictand (GWL), X_i is covariate *i*, N stands

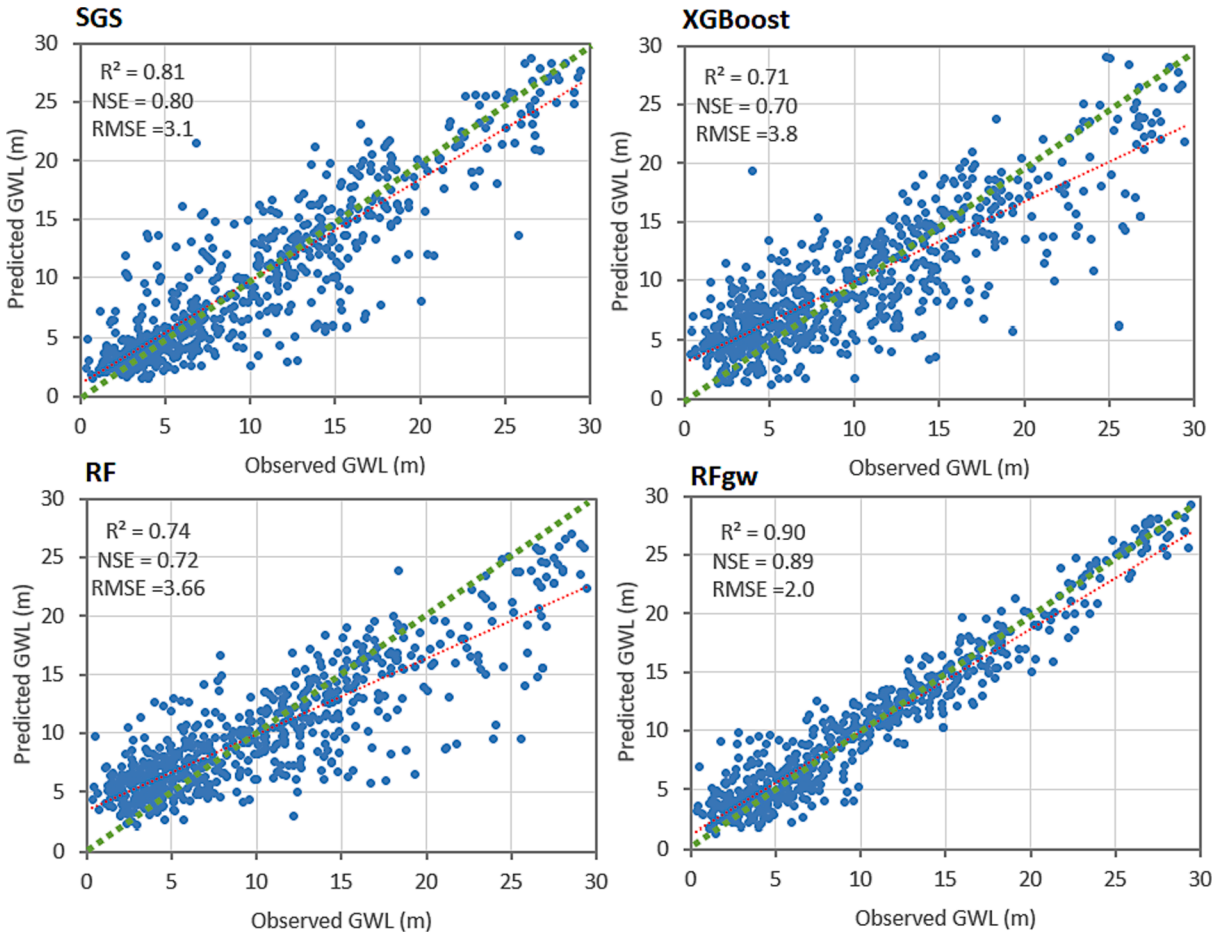


Fig. 6. Comparison of the performance of SGS, XGBoost, RF and RF_{gw} model at testing piezometers. Green dotted line indicates the 1:1 ratio line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for total number of covariates, P_j stands for operational piezometer with spatial coordinate $(u \times v)$ and k stands for total number of operational piezometers for a specific month (t). The data matrix D_P for each specific month (t) has dimensions of "n" rows of operational piezometer locations and "m" columns of total number of variables. This matrix structure was applied for each of the biannual predictions (July and October) from 2003 to 2020 (total 36 predictions).

(2) Preparing training and testing data: We employed stratified sampling which uses the 'createDataPartition' function from the 'caret' package in R to split D_P into training and testing (Hyndman and Athanasopoulos, 2018). We used 70 % of the data for the training set while the remaining 30 % was used for the testing set and validation. Since the number and distribution of operational piezometers vary from year to year, therefore, the stratified sampling was applied to D_P matrix of every year.

(3) Training models at training piezometers: The training set of predictands (GWLs) and local covariates from D_P for a given time step (t: e.g., June 2003) with 1 to k^{th} locations of operational piezometers were used to train the ML models with optimized hyperparameters. In the geostatistical method, an experimental variogram was fitted, and simple kriging was applied to the training set to obtain kriging estimates. These estimates act as the starting point for the SGS method, ensuring accurate GWL predictions at unsampled locations. The training performances of the models were evaluated by comparing the predicted GWLs with observed GWLs at training piezometers.

(4) Testing models at testing piezometers: The trained ML models were tested using only testing covariates from D_P to predict GWLs at testing piezometers. In the geostatistical method, SGS simulations were performed based on kriging estimates to generate multiple equally probable realizations of predicted GWLs at grid size of 1 km^2 covering the entire study area. We validated the predictive performance of the models by comparing their results with observed GWLs at testing piezometers.

(5) Generating GWLs at unmonitored sites and filling missing piezometer data: It was assumed that the spatial model developed with the training data at operational piezometers were applicable at sub-pixels of high-resolution local covariates. So, the 1 km^2 gridded (G) local covariates (X_i) represented as $\left[\left(X_i \times G_{J(u\times v)}\right), \mathrm{i} \in 1,2,3,..,\mathrm{N}; \mathrm{j} \in 1,2,3,..,\mathrm{k}\right]_T$ were imported to test the ML models, generating 1 km^2 GWL data at unmonitored sites and piezometers with missing data.

(6) Repeating Step1-5 for June and October of each year and thus generating 36 predictions per grid cell from 2003 to 2020.

Finally, the 1 km^2 raster output maps generated by the modeling framework were used to: (1) provide high-resolution (information on distributed GWL changes during the 2003–2020 period, (2) provide gap-filled GWL data at desired monitoring wells, and (3) provide GWL estimates for unmonitored sites. The schematic diagram of the modeling steps is presented in Fig. 3.

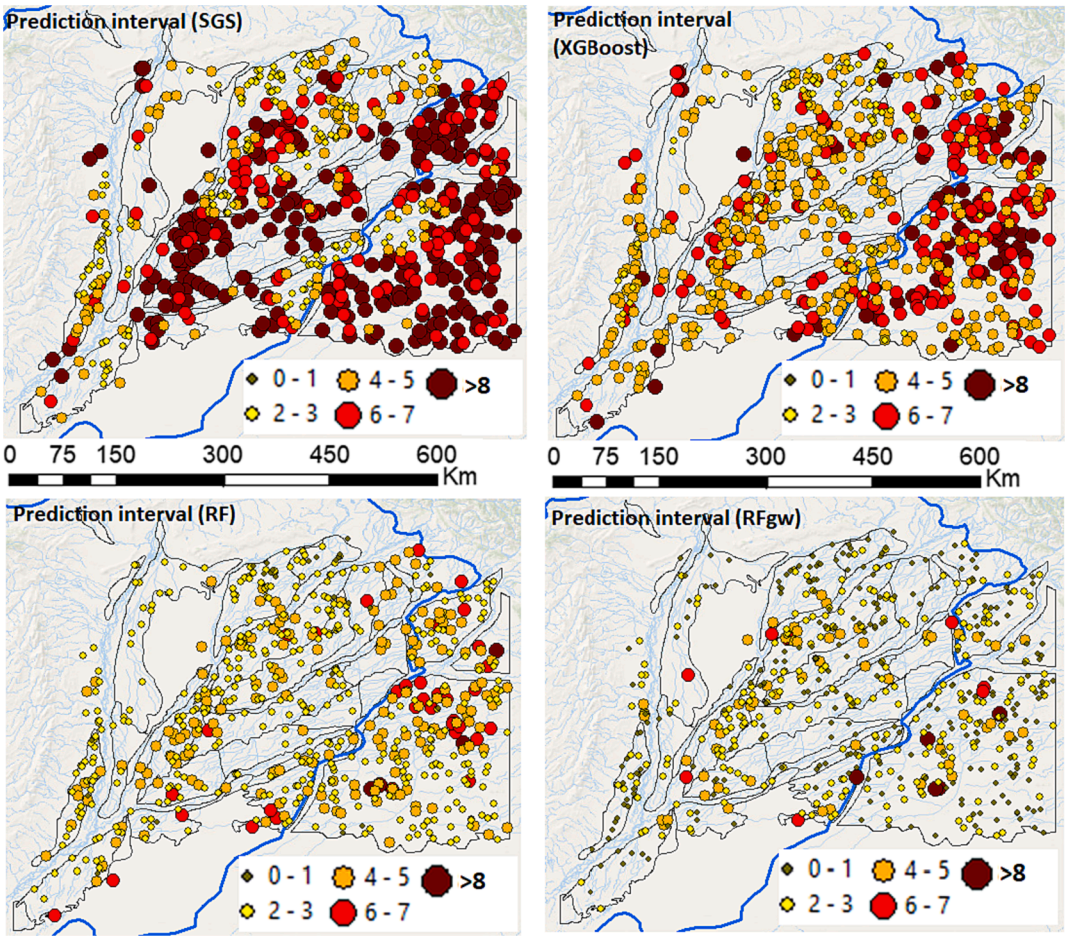


Fig. 7. Width of 90% prediction uncertainty interval for GWLs over the testing piezometers indicated by (a) SGS, (b) XGBoost, (c) RF and (d) RF_{gw}. (Unit: meter).

2.6. Uncertainty quantification

Compounded uncertainty in GWL prediction can result from potential error in monitoring data, covariates from different remote sensing products and various assumptions made in the modeling framework. Therefore, we quantified the total uncertainty of predicted GWL at each piezometer. We trained the ML models using 500 bootstrapping resamples, generating 500 sets of predictions for the testing set. Thereafter, quantile estimation was performed to determine the prediction intervals at 0.05 and 0.95 quantiles (Nourani et al., 2023; Nourani et al., 2022a). The width of prediction intervals was estimated by subtracting the 0.05 quantile prediction from 0.95, which is a common measure of prediction uncertainty (Takoutsing and Heuvelink, 2022; Nourani et al., 2022b). For the SGS method, bootstrapped resampling data were used to generate multiple possible realizations, and the variation among those realizations served as a measure of uncertainty (Delbari et al., 2009).

2.7. Evaluation of prediction skill and model selection

The prediction skill of the four models to reconstruct GWLs for unmonitored sites and piezometers with missing data was assessed by comparing with observed data. We assessed the accuracy in two different ways: (1) how accurately did the model predict GWLs at deep wells based on training data from nearby shallower wells? (2) How was the model's performance affected by increasing the distance between testing piezometers and training piezometers? In other words, how effectively can models make a prediction to fill spatial gaps in GWL data? To do so, we evaluated the prediction accuracy at near (within 8 km), medium distance (8 km to 16 km), and far (beyond 16 km) locations. We quantified predictive accuracy with validation data from piezometers using goodness-of-fit metrics including R^2, MSE (m), RMSE (m), and Nash–Sutcliffe model efficiency coefficient (NSE), (see Table S1 in SM). Furthermore, we compared 1 km^2 estimates of GWL changes with high-resolution GRACE-based GWSA data to validate predictions for unmonitored sites. The model demonstrating the highest accuracy was selected for further applications to study GWL changes in groundwater-stressed areas with limited available data.

3. Results and discussion

3.1. Hyperparameters optimization of ML models

Grid search cross-validation (CV) results indicated that the XGBoost model's prediction error reduced as the boosting iterations "nrounds" increased from 1 to 250. Optimal performance was observed with nrounds (79), eta (0.3) and gamma (0.1) (Fig. S3). Likewise, the RF model's performance improved as the number of ntrees increased up to 250, bringing the MSE down from 3.9 m to 1.6 m. However, additional increases in ntrees had no noticeable influence on the performance (Fig. S4). In many cases, 150–200 ntrees have been reported to render sufficient prediction accuracy (Lopes, 2019). Increasing mtry from 2 to 5 slightly affected the model's prediction error. The optimal values of nodes, ntree, and mtry bootstrapped for the RF model were selected as 16, 250 and 3 (Fig. S4). Optimal values of nodesize, mtry and ntress of the RF model were introduced in the RF_{gw} model. The RF_{gw} model required an appropriate bandwidth to spatially weight the observations. To do so, we evaluated the predictive performance of the RF_{gw} model between 20 and 200 nearest observations using adaptive bandwidth

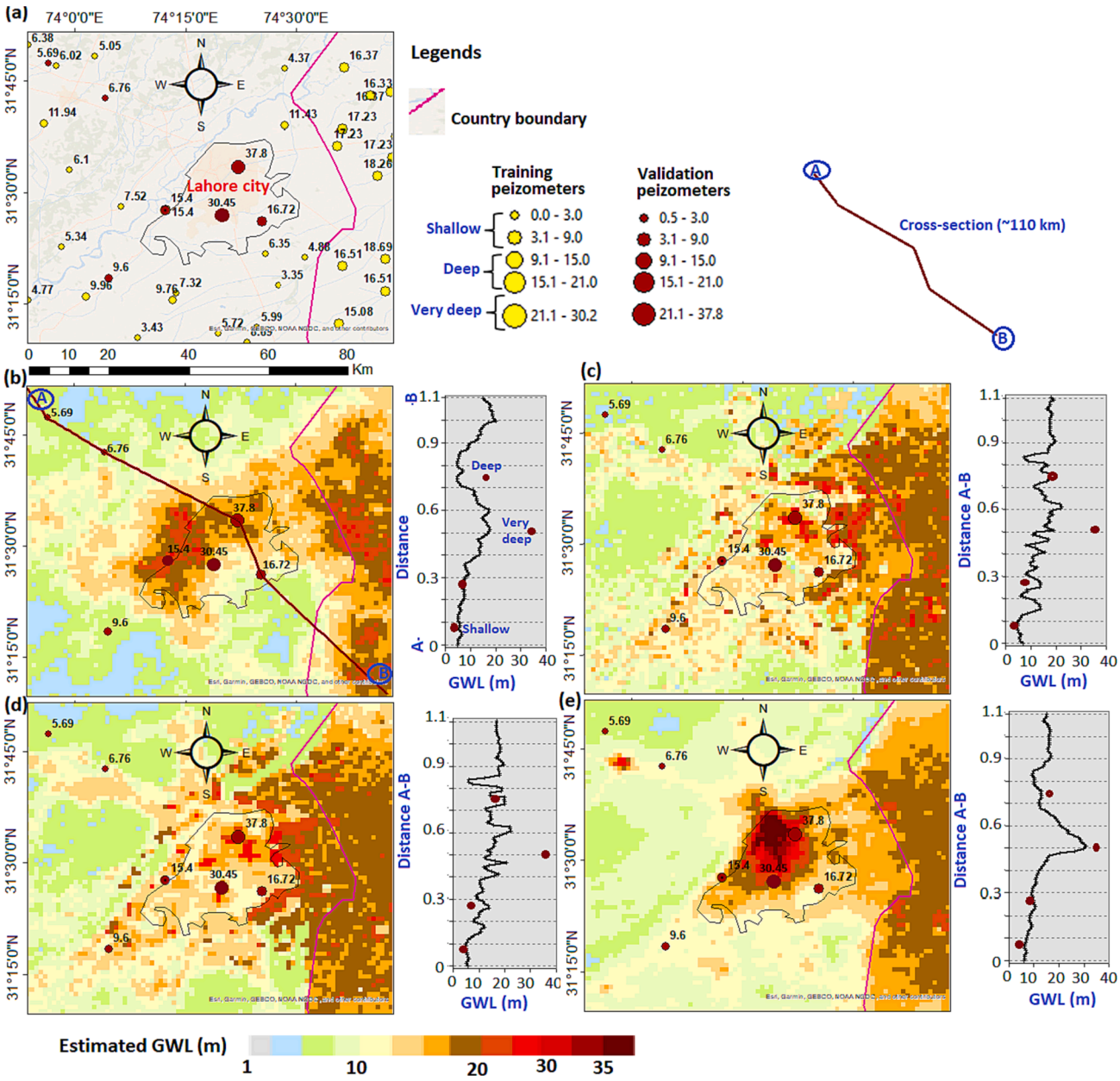


Fig. 8. (a) Spatial distribution of training and validation piezometers in Lahore City, and (b-e) comparative analysis of GWL estimates by the four models—SGS, XGBoost, RF, and RF_{gw} at 1 km^2 resolution. The comparison is based on the spatial patterns of GWLs, and the GWL profiles along the A-B transect.

kernel (Fig. S5). The RF_{gw} model's prediction error reduced from 1.6 to 0.8 m as the bandwidth increased up to 76 nearest observations and then the error started increasing with further increase in the bandwidth as was reported by others (e.g., Fotheringham and Sachdeva, 2022). At the optimal bandwidth of 76 nearest observations, the RF_{gw} model demonstrated superior performance compared to the RF model, achieving a 63 % reduction in the prediction error. Other studies have also reported better performance of RF_{gw} models compared with RF models (Khan et al., 2022).

3.2. *Geographically weighted random forest and comparison with other models*

The predictive accuracy and reliability of each method on the training set is presented in Fig. 4. The RF model demonstrated higher reliability compared to the XGBoost model with an increase of R^2 and NSE from 0.90 to 0.96 and 0.89–0.95, respectively. The RF's superior performance over XGBoost for spatial data was also reported by (Zhang et al., 2023). The SGS geostatistical approach, provided satisfactory results when applied to the training dataset based on R^2 (0.96), NSE (0.97) and RMSE (1.18 m). The comparison between the RF and SGS methods revealed the remarkable superiority of SGS with comparatively smaller RMSE. This outcome underscores the SGS model's ability to harness spatial dependencies, resulting in more accurate and realistic predictions (Zakeri and Mariethoz, 2021). The RF model's non-spatial nature ignores the spatial dependencies among the observations (Heuvelink and Webster, 2022). This can lead to biased predictions, especially when observations have a strong spatial correlation (Takoutsing and Heuvelink, 2022). Further enhancing the RF model's capabilities, we developed the locally and spatially weighted RF_{gw} model. The RF_{gw} model outperformed all the other methods and significantly enhanced the prediction accuracy, resulting in notable improvements in R^2 (0.96 to 0.99) and NSE (0.95 to 0.99), and a substantial reduction in RMSE (1.53 to 0.35 m).

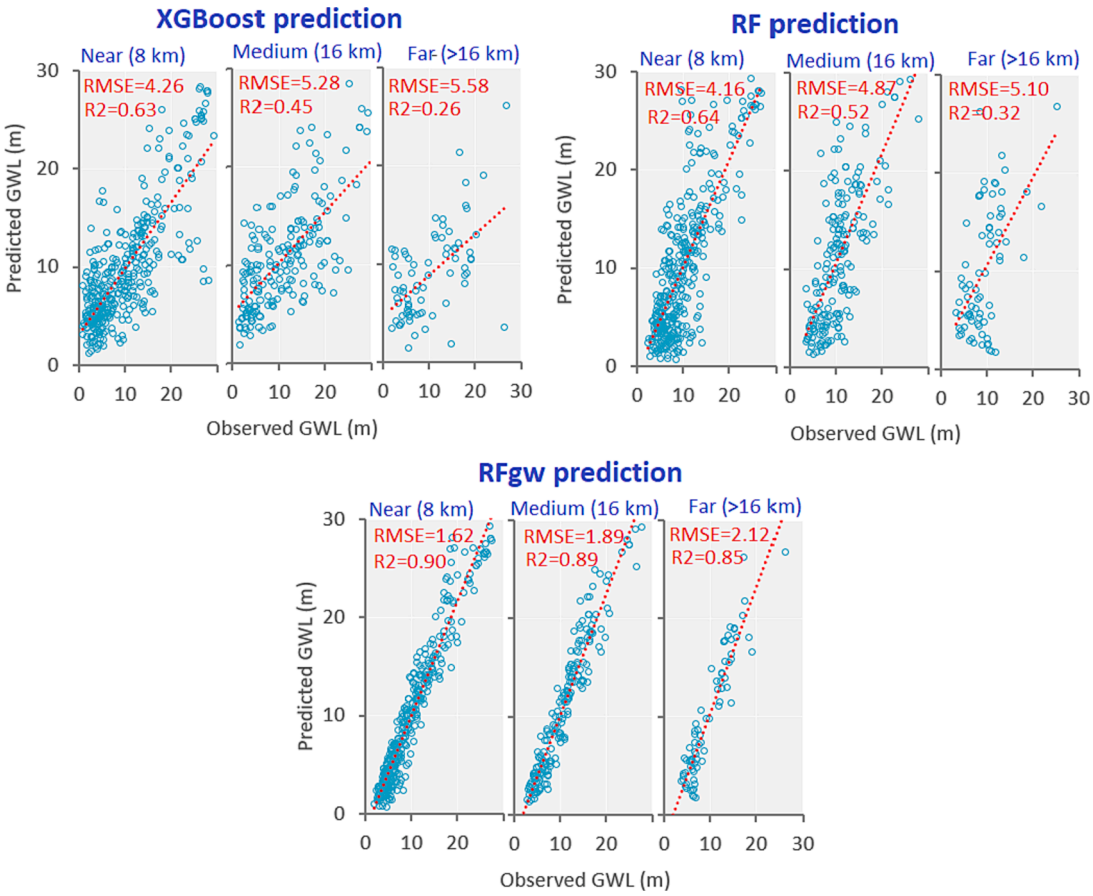


Fig. 9. Evaluation of estimated GWLs based on near (8 km), medium (8–16 km) and far (>16 km) predictions from training piezometers.

Fig. 5 represents the spatial predictions of GWLs and error estimations of SGS, XGBoost, RF and RF_{gw} at training piezometers. The observed values of GWLs were higher in the northern transboundary areas and central Punjab, Pakistan (Fig. 5a), which is characterized by higher water consumption and groundwater storage loss (Arshad et al., 2022; Joshi et al., 2021; Ali et al., 2023). We observed significant difference in the predicted values of GWLs between the XGBoost and RF, particularly across deeper wells (where GWLs > 20 m) in some parts of the study area. The RF_{gw} model effectively captured the spatial pattern and magnitude of the observed GWLs at training piezometers. This is likely because the RF_{gw} model was calibrated locally at each piezometer and accounted for the effects of neighborhood observations to achieve better predictive performance (Georganos et al., 2021; Georganos and Kalogirou, 2022). The error estimations revealed that SGS, RF, and RF approaches tended to overestimate and underestimate groundwater levels (GWLs) across most piezometer observations. Notably, XGBoost and RF demonstrated relatively higher levels of overestimation (up to 7.2 m) and underestimation (down to −7 m) at deeper piezometers on the Indian side (Fig. 5b). The probability density plots and error bars of SGS, XGBoost and RF approaches unveiled the higher spread of predicted values on both sides of zero mean within the range of 7.5 - −8.8 m, indicating relatively lower accuracy. By contrast, the histogram and error bars of the RF_{gw} model displayed a narrower distribution around zero mean, coupled with a higher density value of 0.45, highlighting its better performance in minimizing prediction errors (Fig. 5c&d).

3.3. Statistical evaluation and prediction uncertainty at testing piezometers

The models trained in the previous step were also used to predict the GWLs at testing piezometers to evaluate their prediction skill for unmonitored locations. The predictive performances of SGS, XGBoost, RF and RF_{gw} model at testing piezometers are presented in Fig. 6. Goodness-of-fit metrics confirmed that the SGS approach demonstrated relatively favorable performance compared to RF and XGBoost models based on R^2 (0.81), NSE (0.80), and RMSE (3.1 m). The results of the XGBoost model were the least accurate (R^2 = 0.71, NSE = 0.70, and RMSE = 3.8 m) (Fig. 6). The RF and XGBoost predictions predominantly positioned below the 1:1 ratio line, particularly for deeper wells with GWLs ranging 20–30 m. This indicated underestimations in predicted values for deeper GWLs. In the case of the RF_{gw}, the GWL predictions were less scattered around the 1:1 ratio line, indicating its ability to produce more accurate and consistent estimates of GWLs. Overall, the local RF_{gw} outperformed other approaches with higher values of R^2 (0.90) and NSE (0.89) and lower RMSE (2.0 m) (Fig. 6).

The comparison of prediction uncertainty maps among different modeling approaches revealed distinct variations (Fig. 7). The uncertainty in predictions produced by the SGS method varied across the study area, ranging from 1 m to 8 m. Most testing piezometers indicated higher uncertainty values ranging between 6 and 8 m. The XGBoost model exhibited relatively smaller uncertainty, although they were higher than those of the RF model. This was especially prominent in the case of deeper piezometers located on the Indian side. The RF model's prediction uncertainty depicted a narrower range of values, primarily falling between 2 m and 5 m across most piezometers. Some specific piezometers exhibited relatively higher uncertainty (>7 m). In contrast, the prediction uncertainty map for the RF_{gw} model indicated a reduced range of values compared to the RF model and the other models (Fig. 7). These results underscore that the RF_{gw} method outperforms other ML models and the SGS method in terms of prediction accuracy.

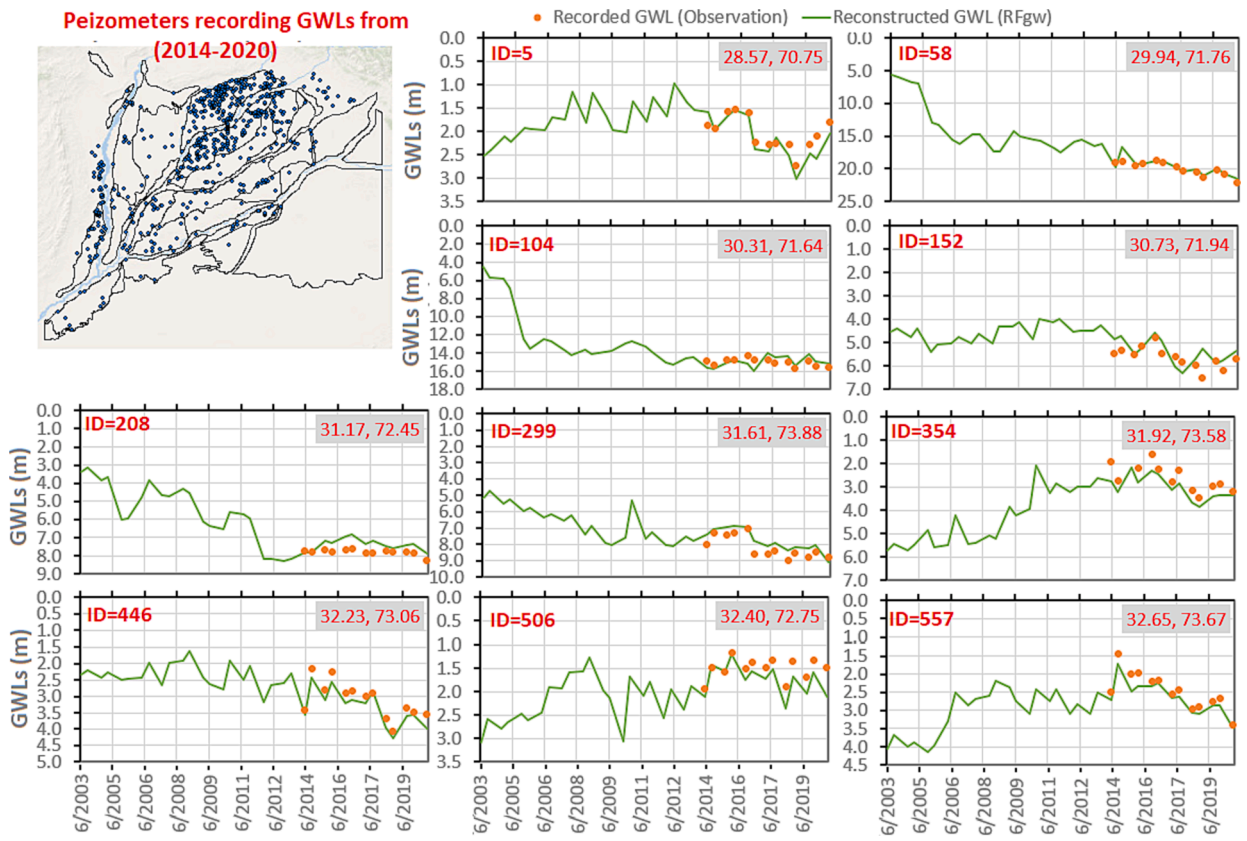


Fig. 10. RF_{gw} based biannually (June and Oct) reconstructed time series of GWL changes at piezometers recording GWLs during the 2014–2020 period.

3.4. Evaluation of high-resolution GWL estimates

The 1 km^2 estimates of GWL were constructed by testing the ML models using high-resolution covariates. In the case of SGS, the estimates of GWL were obtained by conducting simulations on a grid size of 1 km^2 (Fig. S6 in SM). Taking Lahore City as an example, we evaluated the predictive potential of estimated GWLs for deep wells (30–37 m) based on the neighborhood training data encompassing shallower (4–11 m in the west) and deeper wells (GWL ~ 11–19 m in the east) (Fig. 8a). The spatial patterns of estimated GWLs showed good agreement with observational data for shallow to very shallow wells in the suburbs (Fig. 8b-e). However, for deeper wells within the city, SGS, XGBoost, and RF underestimated the GWLs by 7 m, 10 m, and 11 m, respectively, based on the average of three validation piezometers. In contrast, the RF_{gw} model closely estimated GWLs at deeper wells in urban areas with an average deviation of 2 m from the observed values. The RF_{gw} model helped detect hotspots of groundwater drawdown in Lahore City, reflecting higher GWLs in the central parts of the city and shallower GWLs in the suburbs. Analyzing the GWL profile along the A-B transect illustrates the superior performance of the RF_{gw} model in capturing the non-linear aquifer behavior, matching observed levels over all the validation piezometers along the A-B transect (Fig. 8b-e). While the SGS, XGBoost, and RF estimated GWLs with a linear increasing pattern that aligned with the training data, they showed larger underestimations at deeper wells located along the A-B transect. Overall, incorporating anthropogenic factors in the RF_{gw} model improved its accuracy in human-impacted areas, where the SGS model underperformed due to its reliance on neighborhood data alone. Although XGBoost and RF also considered these human factors, their non-spatial approach resulted in less accurate predictions.

We also evaluated the results depending on the distance of tested locations from training piezometers (Fig. 9). Notably, the XGBoost model exhibited comparatively weaker performance, particularly at far distances (>16 km), where R^2 and RMSE were 0.26 and 5.58 m, respectively. The RF model demonstrated improved accuracy, achieving R^2 of 0.32 for far locations and R^2 of 0.64 for those located within 8 km of the training piezometers. Generally, the predictive accuracy of XGBoost and RF decreased when they were tested on far distances from their original training locations. Brenning (2023) also reported that geostatistical (kriging) and ML methods (RF) provided the best estimation near training sites while estimation bias increased with distance from training observations. The RF_{gw} model outperformed both XGBoost and RF, particularly excelling in predicting far GWLs (R^2 = 0.85 and RMSE = 2.12 m (Fig. 9) due to accounting for the geographic autocorrelation of observations in the training set, which helps make reliable prediction on neighborhood locations of testing data (Sekulić et al., 2020). These findings underscore the crucial role of geospatial considerations and model selection in GWL predictions. Based on the superior performance of the RF_{gw} model, we applied this model to estimate GWL at unmonitored locations and at sites whose records had missing data.

3.5. Temporal and map view of reconstructed GWL data from 2003 to 2020

The RF_{gw} model was applied biannually (July and October) to reconstruct high-resolution and continuous GWLs from 2003 to 2020 at unmonitored sites and at piezometers having missing observations. As an example, reconstructed GWLs matched observed GWLs for a series of piezometers recording GWL since 2014. Application of the RF_{gw} model provided GWL data at this site for the 2003–2013 period when GWLs were unavailable (Fig. 10). Testing of the RF_{gw} model at piezometers with available data confirmed reasonable hindcast of GWL estimates. Chen et al. (2019) used a similar assumption while hindcasting groundwater storage changes (1980–2001) before the GRACE period (2002–2019) in the Songhua River Basin using ANNs. To further confirm the accuracy of reconstructed results for non-monitored sites, we also

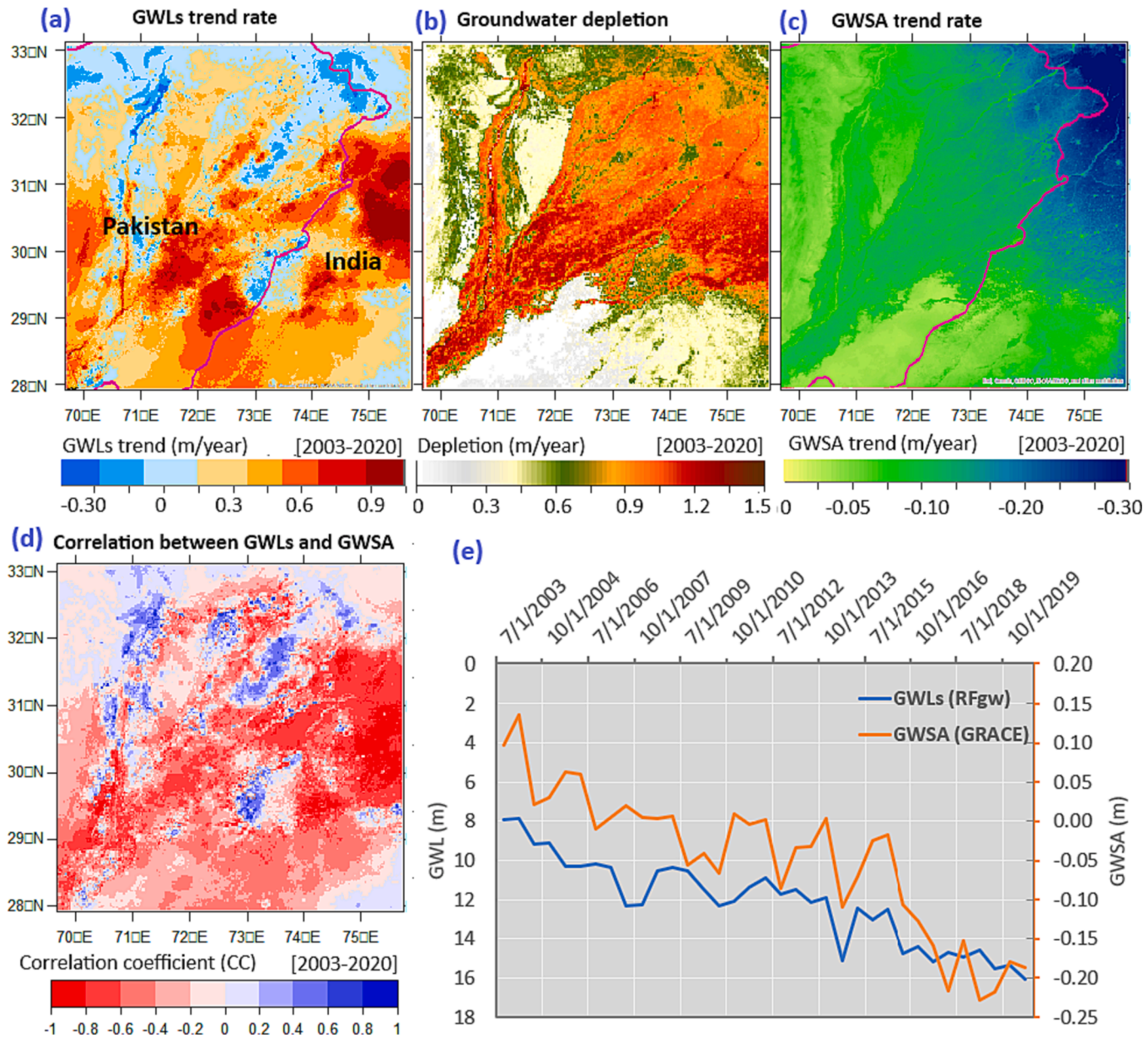


Fig. 11. (a) trend rate of GWL changes, (b-c) average depletion and trend rate of groundwater storage loss based on GRACE GWSA data, (d) grided correlation coefficient, and (e) average temporal changes in GWLs and GWSA from 2003 to 2020 across non-monitored sites.

compared the estimated GWLs with GRACE-based groundwater storage anomalies (GWSA) and depletion changes from 2003 to 2020 (Fig. 11). The GWLs is declining at faster rate by 0.3–0.98 m/year in the eastern side (India) and southern regions of Pakistan (Fig. 11a). The larger decline rates of GWLs were mainly observed in the hotspot regions with larger groundwater depletion (0.75–1.5 m/yr) (Fig. 11b) and groundwater storage loss (GWSA: −0.15 - −0.30 m/year) (Fig. 11 c). The grided correlation coefficient of GWL changes from 2003 to 2020 at most grids showed a negative correlation (-0.60 to −0.96) with GRACE-based GWSA data indicating that groundwater storage loss is significantly associated with the decline rate of GWLs (Fig. 11 d). The increasing rate of groundwater storage loss in the study area has been well documented by previous studies (Ali et al., 2023; Arshad et al., 2022a). Regional average temporal decline in GWLs was consistent with GRACE-based GWSA from 2003 to 2020 across non-monitored sites (Fig. 11e). Thus, the strong correlation between the GRACE-based GWSA and GWL underscores reasonable accuracy of the reconstructed data non-monitored sites.

The reconstructed data showing the GWL changes between 2003 and 2020 for piezometers having defective observations located in different settings (agricultural and *peri*-urban) are provided in the SM (Fig. S7), showing spatiotemporal changes in GWLs (Fig. S8). While GWLs dropped from 2003 to 2020 in many areas, the changes were more pronounced in the south-east (Multan, Khanewal etc.). Also, some hotspots were detected in the upstream urban regions (e.g., Lahore) due to aquifer overdraft for human consumption and irrigation water supply (Aslam et al., 2022). The spatial heterogeneity of GWL variation is also associated with substantial human footprints in the already vulnerable regions as well as arid and semi-arid nature of the study region (Zhu et al., 2021).

3.6. *Application in critical water-stressed and data-sparse regions*

The reconstructed high-resolution GWL data were analyzed to assess

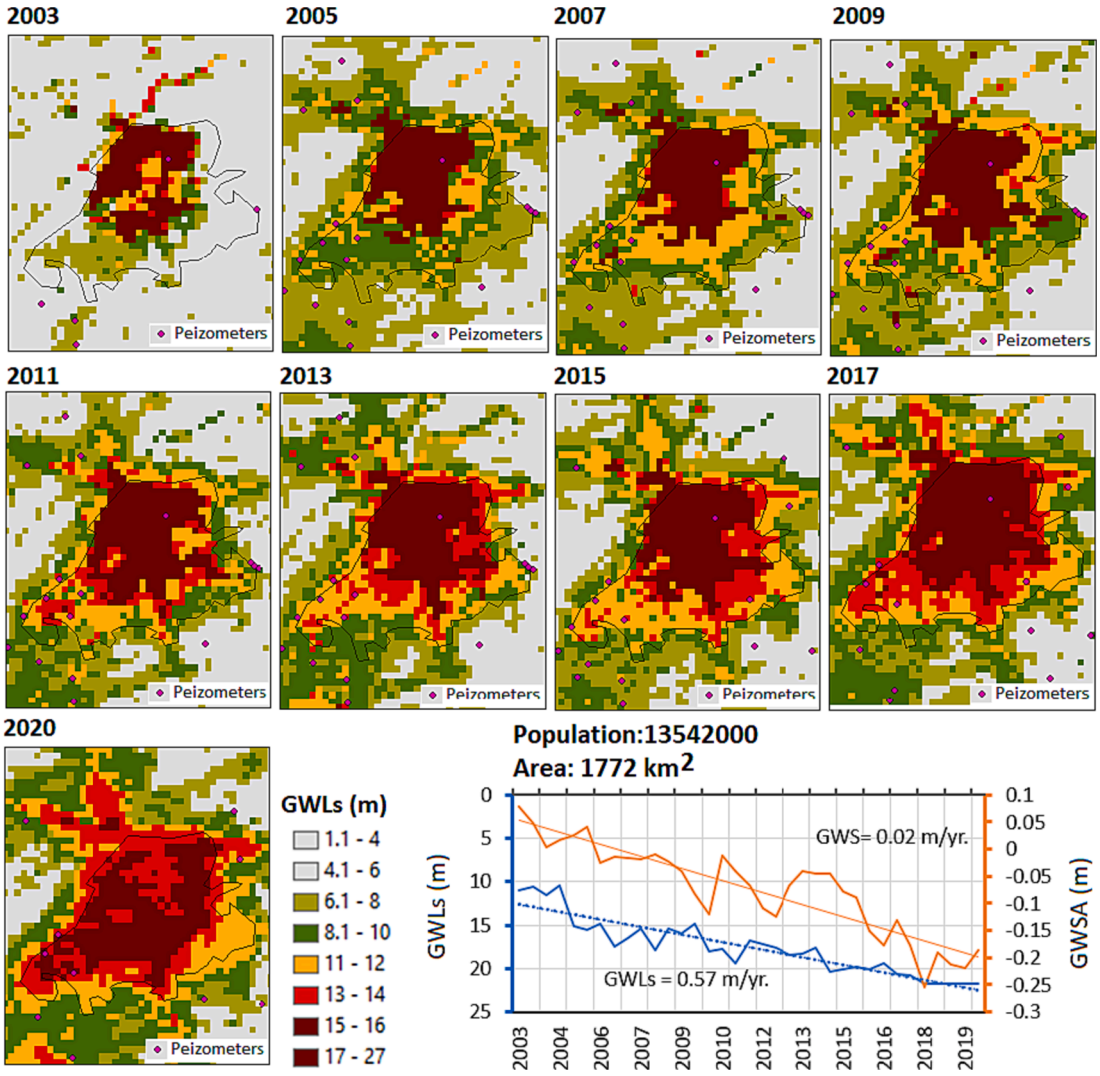


Fig. 12. Spatiotemporal GWL changes in Lahore City and comparison with groundwater storage (GWS) loss indicated by GRACE-based GWSA between 2003 and 2020.

the groundwater changes in regions of higher water stress and limited available data. The selected areas include thirty-two canal command areas, secondary water channels (or distributaries), and densely populated urban areas.

3.6.1. Assessing GWL changes in data-sparse and unmonitored urban regions

The per capita water availability in the region has decreased from 1,700 m^3 (year 1992) to 1,090 m^3 (year 2012) (Kamal et al., 2012). Rapid urbanization driven by population growth and industrialization has put a strain on the available water supplies, and per capita water availability in the basin is projected to decrease to 850 m^3 by the year 2025 (IWMI, 2000). The water consumption in major cities is also projected to increase (e.g., Lahore (from 734 MCM/yr to 1,635 MCM/yr), Faisalabad (from 279 MCM/yr to 623 MCM/yr), and Multan (from 169 MCM/yr to 378 MCM/yr) from 2010 to 2050) (Basharat, 2016). To provide insight about potential vulnerabilities of these major cities, we assessed the GWL changes in Lahore (area: 1,772 km^2, population: 13.5 million), Multan (area: 227 km^2, population: 2.1 million), and Faisalabad (area: 211 km^2, population: 3.6 million) between 2003 and 2020. These urban areas are already hotspots for GWL decline with minimum (and maximum) trend rate of 0.1 m/yr (and 0.9 m/yr) between 2003 and 2020 (Fig. S9). The reconstructed high-resolution data at 1 km^2 illustrates distributed GWL changes in unmonitored sites within and around Lahore City where there were only a few piezometers between 2003 and 2020 (Fig. 12). A decline in groundwater was observed as indicated by gradually increasing depth to groundwater from the ground surface. The pixels with deeper GWLs (e.g., 17–27 m) were mainly detected in the central part of the city in 2003, spreading outward into *peri*-urban regions by the year 2020. The ongoing decline in GWL was consistent with GRACE-based GWS depletion between 2003 and 2020 (Fig. 12) and previous studies (Aslam et al., 2022; Zahran et al., 2023).

3.6.2. Groundwater status in canal command areas

There is a critical need to understand the historical and current groundwater status in the irrigated canal command areas in the Indus Basin to inform sustainable groundwater use by the agriculture sector. The reconstructed biannual GWL data were used to assess groundwater changes in the thirty-two irrigated canal command areas located in the study area. Further, we gained insight into the causes of GWL changes due to shifts in water demand (estimated by actual ET) and supply (estimated by canal water (CW) and precipitation) (Table 2).

Falling GWLs: Fifteen canal command areas showed a statistically significant ($p < 0.05$) increasing rate of GWL decline since 2003 (Table 2 and Fig. S10a). In general, canal command areas irrigated from the eastern rivers (Ravi, and Satluj) had a higher increase in GWL decline

Table 2
GWL changes in thirty-two canal command areas between 2003 and 2020.

Canals	Rivers (Country)	PRECIP (mm/yr)	ET (mm/yr)	CW (mm/yr)	Sen's slope (GWL: m/yr)	Decline status
Muzffgarh Canal	Indus (PK)*	206.00	1239.56	743.00	−0.13	Significant decreasing (P < 0.05)
Upper Sadiqia Canal	Sutlej (PK)	240.00	727.50	865.00	−0.12	
CRBC/Paharpur Canal	Indus (PK)	325.62	992.97	200.00	−0.09	
Rangpur Canal	Chenab (PK)	238.00	1238.44	271.00	−0.08	
Fordwah	Sutlej (PK)	113.51	1283.67	482.00	−0.06	
Haveli Canal	Chenab (PK)	167.34	1295.97	564.00	−0.04	No change (P > 0.05)
Bannu Scarp	Indus (PK)	398.32	984.81	170.00	−0.03	
Thal Canal	Indus (PK)	300.33	829.85	544.00	−0.02	
Upper Bari	Ravi (PK)	658.26	1218.16	500.00	−0.02	
Upper Bahawal & Qaim Canal	Sutlej (PK)	223.00	1294.44	917.00	−0.01	
Gugera	Chenab (PK)	259.54	1182.26	616.00	−0.01	
Upper Chenab Canal	Chenab (PK)	481.56	1219.81	249.00	0.00	
Pakpattan Canal	Sutlej (PK)	250.00	1371.34	607.00	0.00	
Panjnad Canal	Indus (PK)	65.00	1277.85	704.00	0.00	
Marala Ravi Canal	Chenab (PK)	694.44	1195.16	180.00	0.01	
Lower Sadiqia Canal	Sutlej (PK)	240.00	1050.23	865.00	0.02	
Upper Jehlum Canal	Jehlum (PK)	429.00	1212.55	738.00	0.02	
Lower Jehlum Canal	Jehlum (PK)	252.00	1141.48	411.00	0.03	Significant increasing
Raya Branch (BRBD Inetranl)	Chenab (PK)	535.50	1192.77	250.00	0.04	(P < 0.05)
Upper Dipalpur Canal	Ravi (PK)	349.00	1263.89	420.00	0.06	
Bahawal Canal	Sutlej (PK)	150.00	1087.74	763.00	0.08	
Dera Ghazi Khan Canal	Indus (PK)	138.00	1245.41	960.00	0.08	
Jhang (Lower Chenab Canal)	Chenab (PK)	325.76	1121.74	452.00	0.09	
Lower Bari Doab Canal	Ravi (PK)	151.84	1277.50	688.00	0.09	
Ghotki Canal	Indus (PK)	140.00	1192.13	700.00	0.10	
Central Bari Doab Canal	Ravi (PK)	370.58	1106.30	576.00	0.11	
Sidhnai Canal	Ravi (PK)	240.00	1375.63	500.00	0.12	
Abbasia Canal	Chenab (PK)	145.00	980.64	375.00	0.14	
Lower Dipalpur Canal	Ravi (PK)	464.00	1341.84	528.00	0.15	
Mailsi + Lower Pakpattan canals	Ravi (PAK)	183.00	1388.31	472.00	0.17	
Dehli Doab	Sutlej (IND)**	710.00	1275.00	320.00	0.33	
Bist Doab	Sutlej (IND)	1153.00	1350.00	150.00	0.60	

*PK stands for Pakistan.
**IND stands for India.

rate. On the Indian side, the GWL is falling at a statistically significant rate of 0.60 m/yr and 0.33 m/yr in irrigated areas across Bist and Dehli Doab, respectively. In Pakistan, the highest significant increasing rates of GWL decline were observed in the downstream canal command areas such as Mailsi + Lower Pakpattan canals (Multan) followed by Lower Dipalpur, and Abbasia Canal. GWL fell at a rate of 0.17 m/yr in the Mailsi + Lower Pakpattan canals (Multan) between the 2003 and 2020 when the average ET demand (~1,388 mm/yr) was higher than the water supply (precipitation = 183 mm/yr and CW = 432 mm/yr with a difference of 733 mm/yr) (Table 2). As a result of increased irrigation demand and reduced water availability from rain and canal water, groundwater recharge diminished, contributing to the rapid GWL decline in the Mailsi + Lower Pakpattan canals (Fig. S10a). The GWL dropped dramatically by the year 2010, when the irrigated region in Mailsi + Lower Pakpattan canals suffered a reduction in annual precipitation (~-19 % below 2003–2020 average). However, ET remained steady and even increased slightly (~0.5 %) over this period. The rates of GWL decline over the Lower Dipalpur, and Abbasia Canals were recorded as 0.15 m/yr and 0.14 m/yr, respectively (Table 2). The actual ET was higher than the combined water supply from precipitation and canal water by ~ 350 mm/yr (Lower Dipalpur) and ~ 461 mm/yr (Abbasia). Studies reported that the Lower Dipalpur canal always operated in water deficit because deliveries were smaller than entitlements (Siddiqi et al., 2018; Wescoat et al., 2018). In addition, the eastern rivers flow dropped after the Indus Water Treaty (Wescoat et al., 2018) and despite the water transfers from the western rivers through link canals, long stretches of the rivers remain dry most of the year. The lack of surface water has the double devil effect of reducing recharge while also forcing agricultural producers to extract more water from the aquifer to irrigate the crops (Arshad et al., 2022; Hashmi, 2021). In contrast, the rate of GWL decline was smaller (0.03 m/yr) over the Lower Jhelum canal command area (Table 2) with very small fluctuations between 2003 and 2020 (Fig. S10a) as water deliveries were quite close to the entitlements (Wescoat et al., 2018). Managing water demands through canal water supply also assists to reduce the declining rate of groundwater level (Kaur et al., 2023).

Rising GWLs: We identified five canal command areas where GWLs increased between 2003 and 2020 (e.g., Muzffgarh Canal, Upper Sadiqia Canal, CRBC/Paharpur Canal, Rangpur Canal, Fordwah) (Fig. S10b and Table 2). For example, GWLs in Muzaffargarh Canal increased by 0.13 m/yr, although the actual ET exceeded the precipitation and canal water supply by 291 mm/yr. Since this canal command area is located between the junctions of all the rivers near the Panjnad, it is likely that seepage from water channels contributed to increasing groundwater storage. The increasing GWLs in the Upper Sadiqia canal command area may be associated with brackish quality of aquifer as well as precipitation and canal water amount exceeding the actual ET by 377 mm/yr. Our findings agree with other published results about GWL in the study area (Hashmi, 2021; Sanmugan Prathapar et al., 2021).

No change: No significant changes in GWL were observed in twelve canal command areas with p > 0.05 (Table 2). There is a possibility that groundwater consumption in these canal command areas has not exceeded the renewable surface water availability to trigger significant groundwater table drawdown.

3.6.3. *Characterizing GWL changes across secondary channels*

The high-resolution GWL data were further utilized to characterize the groundwater changes in small-scale water distributary channels where irrigation is mainly pumped from the aquifer. To ilustrate this, we provided an example of the Hakra Branch canal (Bahawal) and its seventeen distributaries to characterize GWL changes in head and tail farms for years 2003, 2008, 2013 and 2020 (Fig. 13a). The selection of the Hakra Branch canal as a case study was motivated by its significant role in irrigation management, a history of innovative practices, and its

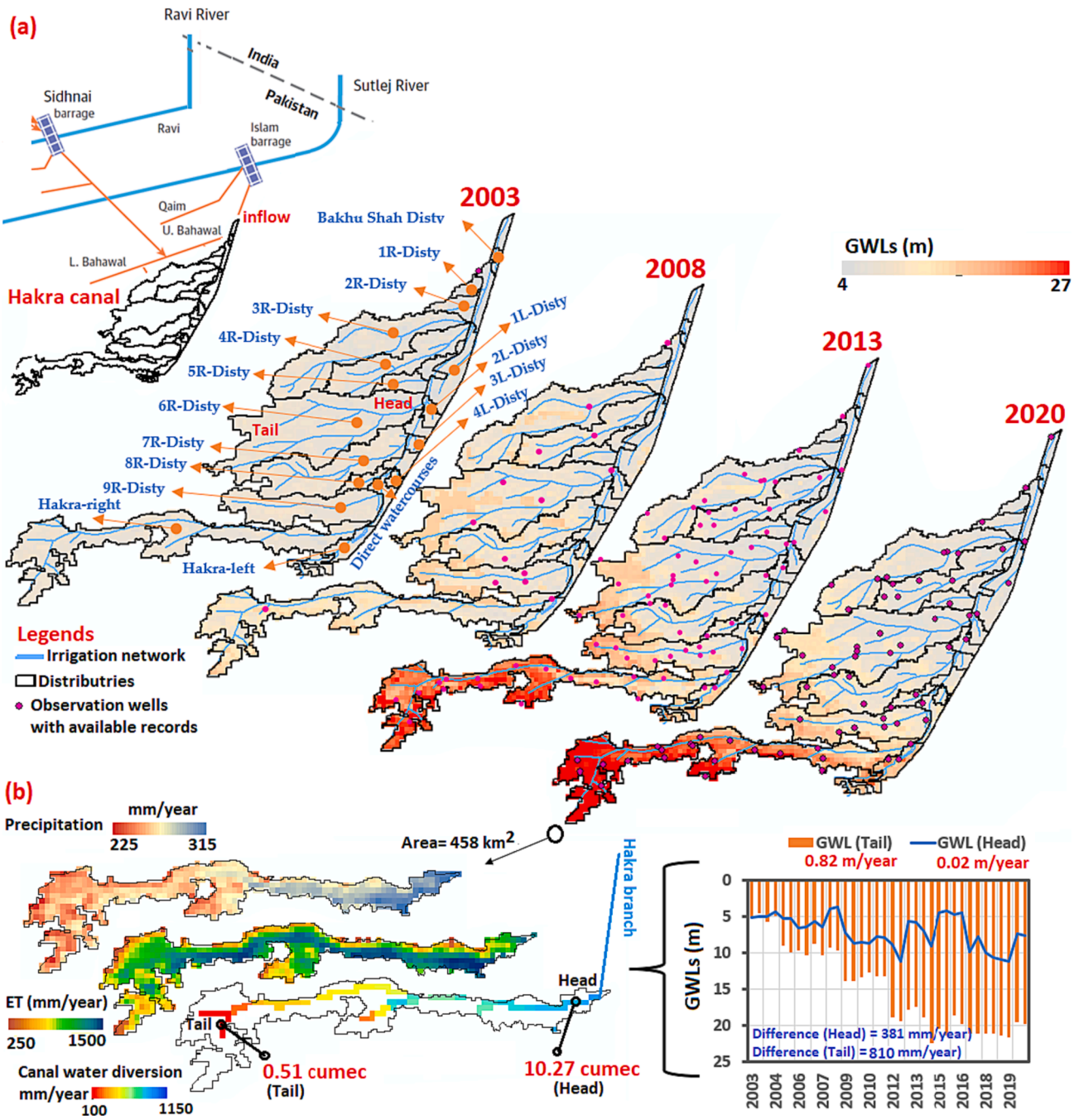


Fig. 13. (a) Spatial-temporal distribution GWL changes in 17 distributires in Hakra Branch canal during the 2003–2020 period, and (b) head–tail (i.e., upsteam-downstream) changes in GWLs relative to water consumption (ET) and supply (precipitation + canal water divrsion) in the Hakra-right distributory. Cumec denotes m^3/sec.

transboundary relevance in the Indo-Pak region. Notably, the region had a sparse groundwater monitoring network, with only 13 monitoring wells in 2003 and 2008, but by 2013, this increased to 70 wells, and in 2020, there were 67 operational monitoring wells. It was observed that during the 2003–2020 period downstream GWLs dropped at a larger rate (0.83 m/yr) in the Hakra-right distributory (Disty) followed by Hakra-left Disty (0.30 m/yr) and 9R-Disty (0.16 m/yr). The GWL decline rate was larger in the tail farms of each distributory as compared to the head farms (Fig. 13a). This is mainly because farmers in the downstream areas of each distributory used ~ 40 % more groundwtaer than the upstream farmers (Awan et al., 2016).

We conducted an evaluation over Hakra-right Disty for a closer examination of the spatial pattern of GWL changes from head to tail farms. This distributary had 458 km^2 of croplands. Surface water diversions were 10.27 m^3/sec (or cumec) in the head farms and 0.51 m^3/sec in the tail farms (Fig. 13b). Based on actual ET, water use was higher in the head farms and moderately lower in the tail farms, which received less precipitation than the head farms (Fig. 13b). On average, GWL dropped at a rate of 0.02 m/yr in the head farms, which received larger canal water diversions. This is because the combined water supply from precipitation (285 mm/yr) and canal water (853 mm/yr) was less than the actual demand (ET: 1519 mm/yr). The deficit (381 mm/yr on average) was compensated by pumping groundwater to meet irrigation demands. The decline rate of GWL in tail farms was much larger (0.82 m/yr) between 2003 and 2020 when the average water deficit was 810 mm/yr (average ET demand = 1150 mm/yr, precipitation = 240 mm/yr, and

CW = 100 mm/yr), increasing the pressure on the aquifer (Fig. 13b). Our results align with those reported by Awan et al. (2016). The reconstructed GWLs at 1 km^2 resolution improves understanding of the spatial–temporal distribution of GWL changes and emphasizes the need of revisitng water allocation in the head and tail (i.e., upsteam and downsteam) farms as highlifgted by a previous study (Qureshi, 2014).

4. Potential limitations and future work

This study's temporal scope is limited to the analysis of historical trends without considering post-2020 changes. This is because the relevant data were only available up to 2020. The temporal resolution of the estimated GWLs is limited to a biannual scale due to the frequency of the available in-situ observational data, potentially missing finer-scale variations. Future research can apply predictive modeling with inputs from CMIP6 climate models to project groundwater levels under climate change. Regional generalization, external factors like land use changes, anthropogenic influences, and the quality of new GWL data at piezometers should be comprehensively investigated in future research for a more holistic view of groundwater changes across the Indus Basin.

5. Conclusions

We presented a comprehensive framework for generating high-resolution (1 km^2), continuous biannual GWLs estimates in data-sparse regions, with a focus on the Indus Basin. Our results from the application of the framework in the Indus Basin demonstrated the superior performance of the RF_{gw} model relative to traditional machine learning (RF and XGBoost) and Geostatistical (SGS) methods. The spatial nature of the RF_{gw} model and leveraging local covariates enhances its predictive accuracy and captures complex heterogeneous patterns of GWL changes, particularly in the human-impacted areas. Furthermore, the framework can be applied to provide high-resolution GWL estimates at unmonitored locations, bridging data gaps and overcoming the scarcity of monitoring wells. Notably, our analysis revealed the RF_{gw} model's ability to accurately predict GWLs even at considerable distances from training piezometers. The temporal analysis revealed a substantial increase in GWL decline in vast areas of the Indus Basin from 2003 to 2020, particularly in regions with higher water consumption. These findings were consistent with GRACE-based estimates of groundwater storage loss. The reconstructed high-resolution estimates of GWL from 2003 to 2020 were utilized in data-scarce regions, including densely populated urban areas such as Lahore, Multan, and Faisalabad. Our results highlighted increasing water consumption trends and their strain on GWL decline. In irrigated regions, fifteen canal command areas displayed an increasing rate of GWL decline associated with reduced surface water supplies (canal and precipitation), increasing water demands and substantial increase in groundwater dependency. High-resolution GWL estimates facilitate better understanding of the spatial–temporal distribution of GWL changes from upstream or head farms to downstream or tail farms within water distributaries. This analysis revealed larger GWL drops in tail farms due to the larger share of groundwater in their total water supply. The new spatiotemporally continuous GWL data provide valuable information to facilitate adaptive water management plans in critically groundwater-stressed regions of the Indus Basin. Our study presents a transferable framework for addressing groundwater monitoring and prediction challenges in data-sparse regions around the world.

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.

Data availability

Data will be made available on request.

Acknowledgement

The authors acknowledge funding from the National Science Foundation (NSF Award 2114701) of the United States. Any opinions, findings, conclusions, or recommendations expressed in this publication are solely those of the authors.

Data Availability:

The datasets generated for this study, including high-resolution groundwater level data for the Indus Basin, are available in the Figshare repository at [https://doi.org/10.6084/m9.figshare.24224674.v1]. These datasets can be accessed for further reference or validation of the research findings.

Arfan Arshad conceived of the ideas, performed the experiments, analyzed the results, and prepared the first draft of the manuscript under the supervision of **Ali Mirchi** and **Javier Vilcaez**. **Muhammad Umar Akbar** assisted with the analysis and data acquisition**. Kaveh Madani** provided oversight, and helped with technical review and editing the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2023.130535.

References

Abbas, M., Arshad, M., Shahid, M.A., 2023. Zoning of groundwater level using innovative trend analysis: Case study at Rechna Doab. Pakistan. Water Resources and Irrigation Management-WRIM 12 (1–3), 64–80. https://doi.org/10.19149/wrim.v12i1-3.3155.

Aeschbach-Hertig, W., Gleeson, T., 2012. Regional strategies for the accelerating global problem of groundwater depletion. Nature Geoscience 5 (12), 853–861. https://doi.org/10.1038/ngeo1617.

Ahmed, W., Rahimoon, Z.A., Oroza, C.A., Sarwar, S., Qureshi, A.L., Framroze Punthakey, J., Arfan, M., 2020. Modelling Groundwater Hydraulics to Design a Groundwater Level Monitoring Network for Sustainable Management of Fresh Groundwater Lens in Lower Indus Basin. Pakistan. Applied Sciences 10 (15), 5200.

Ahmed, M., Wiese, D.N., 2019. Short-term trends in Africa's freshwater resources: Rates and drivers. Science of the Total Environment 695, 133843. https://doi.org/10.1016/j.scitotenv.2019.133843.

Akhtar, F., Nawaz, R.A., Hafeez, M., Awan, U.K., Borgemeister, C., Tischbein, B., 2022. Evaluation of GRACE derived groundwater storage changes in different agro-ecological zones of the Indus Basin. Journal of Hydrology 605, 127369.

Ali, S. et al., 2023. Spatial downscaling of GRACE data based on XGBoost model for improved understanding of hydrological droughts in the Indus Basin Irrigation System (IBIS). Remote Sensing, 15(4): 873. https://doi.org/10.3390/rs15040873.

Aliyari, F., Bailey, R.T., Tasdighi, A., Dozier, A., Arabi, M., Zeiler, K., 2019. Coupled SWAT-MODFLOW model for large-scale mixed agro-urban river basins. Environmental Modelling & Software 115, 200–210.

Arshad, A., Mirchi, A., Samimi, M., Ahmad, B, 2023a. The high-resolution (1km) groundwater storage and depletion maps across Irrigated Indus Basin (IIB) during 2002-2019. figshare. Dataset. https://doi.org/10.6084/m9.figshare.22301020.v5.

Arshad, A,. Zhang, W,. Noor, R,. 2023b. TRMM at 1km-Resolution: High-resolution precipitation data in a data-scarce Indus Basin reconstructed through data-driven spatial downscaling and remote sensing. figshare. Dataset. https://doi.org/10.6084/m9.figshare.24570397.v4.

Arshad, A., Zhang, Z., Zhang, W., Gujree, I., 2019. Long-term perspective changes in crop irrigation requirement caused by climate and agriculture land use changes in Rechna Doab. Pakistan. Water 11 (8), 1567. https://doi.org/10.3390/w11081567.

Arshad, A., Mirchi, A., Samimi, M., Ahmad, B., 2022. Combining downscaled-GRACE data with SWAT to improve the estimation of groundwater storage and depletion variations in the irrigated Indus basin. Science of the Total Environment 156044. https://doi.org/10.1016/j.scitotenv.2022.156044.

Asa, E., Saafi, M., Membah, J., Billa, A., 2012. Comparison of linear and nonlinear kriging methods for characterization and interpolation of soil data. Journal of Computing in Civil Engineering 26 (1), 11–18. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000118.

Aslam, R.A., Shrestha, S., Usman, M.N., Khan, S.N., Ali, S., Sharif, M.S., Sarwar, M.W., Saddique, N., Sarwar, A., Ali, M.U., Arshad, A., 2022. Integrated SWAT-MODFLOW Modeling-Based Groundwater Adaptation Policy Guidelines for Lahore, Pakistan under Projected Climate Change, and Human Development Scenarios. Atmosphere 13 (12), 2001.

Asoka, A., Gleeson, T., Wada, Y., Mishra, V., 2017. Relative contribution of monsoon precipitation and pumping to changes in groundwater storage in India. Nature Geoscience 10 (2). https://doi.org/10.1038/ngeo2869.
Awan, U.K., Anwar, A., Ahmad, W., Hafeez, M., 2016. A methodology to estimate equity of canal water and groundwater use at different spatial and temporal scales: a geo-informatics approach. Environmental Earth Sciences 75 (5), 409. https://doi.org/10.1007/s12665-015-4976-4.
Bai, T., Tahmasebi, P., 2022. Sequential Gaussian simulation for geosystems modeling: a machine learning approach. Geoscience Frontiers 13 (1), 101258. https://doi.org/10.1016/j.gsf.2021.101258.
Bailly, A., Blanc, C., Francis, É., Guillotin, T., Jamal, F., Wakim, B., Roy, P., 2022. Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. Computer Methods and Programs in Biomedicine 213, 106504.
Basharat, M., Umair Ali, S., Azhar, A.H., 2014. Spatial variation in irrigation demand and supply across canal commands in Punjab: a real integrated water resources management challenge. Water Policy 16 (2), 397–421. https://doi.org/10.2166/wp.2013.060.
Basharat, M., 2016. Groundwater Environment in Lahore, Pakistan, Groundwater Environment in Asian Cities. Elsevier, pp. 147-184. https://doi.org/10.1016/B978-0-12-803166-7.00008-8.
Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research 13 (2). https://dl.acm.org/doi/10.5555/2188385.2188395.
Bhatti, M.T., Anwar, A.A., Aslam, M., 2017. Groundwater monitoring and management: Status and options in Pakistan. Computers and Electronics in Agriculture 135, 143–153. https://doi.org/10.1016/j.compag.2016.12.016.
Breiman, L., 2001a. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324.
Breiman, L., 2001b. Random forests. Mach Learn. 2001 (45), 5–32. https://doi.org/10.1023/A:1010933404324.
Brenning, A., 2023. Spatial machine-learning model diagnostics: a model-agnostic distance-based approach. International Journal of Geographical Information Science 37 (3), 584–606. https://doi.org/10.1080/13658816.2022.2131789.
Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.
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. Scientific Reports 10 (1), 1–13. https://doi.org/10.1038/s41598-020-60698-9.
Chen, H., Zhang, W., Nie, N., Guo, Y., 2019. Long-term groundwater storage variations estimated in the Songhua River Basin by using GRACE products, land surface models, and in-situ observations. Science of the Total Environment 649, 372–387. https://doi.org/10.1016/j.scitotenv.2018.08.352.
Chung, S.Y., Venkatramanan, S., Elzain, H.E., Selvam, S., Prasanna, M., 2019. Supplement of missing data in groundwater-level variations of peak type using geostatistical methods. GIS and Geostatistical Techniques for Groundwater Science 33–41. https://doi.org/10.1016/B978-0-12-815413-7.00004-3.
Collados-Lara, A.J., Pulido-Velazquez, D., Ruiz, L.G.B., Pegalajar, M.C., Pardo-Igúzquiza, E., Baena-Ruiz, L., 2023. A parsimonious methodological framework for short-term forecasting of groundwater levels. Science of the Total Environment 881, 163328.
Cui, Y.-Q., Yoneda, M., Shimada, Y., Matsui, Y., 2016. Cost-effective strategy for the investigation and remediation of polluted soil using geostatistics and a genetic algorithm approach. Journal of Environmental Protection 7 (1), 99–115. https://doi.org/10.4236/jep.2016.71010.
Cuthbert, M.O., Gleeson, T., Moosdorf, N., Befus, K.M., Schneider, A., Hartmann, J., Lehner, B., 2019. Global patterns and dynamics of climate–groundwater interactions. Nature Climate Change 9 (2), 137–141.
Delbari, M., Afrasiab, P., Loiskandl, W., 2009. Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. Catena 79 (2), 163–169. https://doi.org/10.1016/j.catena.2009.08.001.
Deng, C., Bailey, R.T., 2020. Assessing causes and identifying solutions for high groundwater levels in a highly managed irrigated region. Agricultural Water Management 240, 106329. https://doi.org/10.1016/j.agwat.2020.106329.
Du, E., Tian, Y., Cai, X., Zheng, Y.i., Li, X., Zheng, C., 2020. Exploring spatial heterogeneity and temporal dynamics of human-hydrological interactions in large river basins with intensive agriculture: A tightly coupled, fully integrated modeling approach. Journal of Hydrology 591, 125313.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons 35 (4). https://doi.org/10.1111/j.1538-4632.2003.tb01114.x.
Fotheringham, A.S., Sachdeva, M., 2022. Modelling spatial processes in quantitative human geography. Annals of GIS 28 (1), 5–14. https://doi.org/10.1080/19475683.2021.1903996.
Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto International 36 (2), 121–136.
Georganos, S., Kalogirou, S., 2022. A forest of forests: a spatially weighted and computationally efficient formulation of geographical random forests. ISPRS International Journal of Geo-Information 11 (9), 471. https://doi.org/10.3390/ijgi11090471.
Hashmi, S.A.P.a.M.A., 2021. Reallocating Canal Water in Punjab Province, Pakistan. 5074031 Asian Development Bank.
He, X., Chaney, N.W., Schleiss, M., Sheffield, J., 2016. Spatial downscaling of precipitation using adaptable random forests. Water Resources Research 52 (10), 8217–8237. https://doi.org/10.1002/2016WR019034.
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518.
Heuvelink, G.B., Webster, R., 2022. Spatial statistics and soil mapping: A blossoming partnership under pressure. Spatial Statistics 50, 100639. https://doi.org/10.1016/j.spasta.2022.100639.
Hyndman, R.J. and Athanasopoulos, G., 2018. Forecasting: principles and practice. OTexts.
IWMI, I., 2000. World water supply and demand: 1995 to 2025. Colombo, Sri Lanka. Colombo: International Water Management Institute.
Jha, D., Ward, L., Paul, A., Liao, W.-K., Choudhary, A., Wolverton, C., Agrawal, A., 2018. Elemnet: Deep learning the chemistry of materials from only elemental composition. Scientific Reports 8 (1). https://doi.org/10.1038/s41598-018-35934-y.
Jiménez, Á.B., Lázaro, J.L., Dorronsoro, J.R., 2008. Finding optimal model parameters by discrete grid search. In: *Innovations in Hybrid Intelligent Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 120–127.
Joshi, S.K., Gupta, S., Sinha, R., Densmore, A.L., Rai, S.P., Shekhar, S., Mason, P.J., van Dijk, W.M., 2021. Strongly heterogeneous patterns of groundwater depletion in northwestern India. Journal of Hydrology 598, 126492. https://doi.org/10.1016/j.jhydrol.2021.126492.
Kamal, S., Amir, P. and Mohtadullah, K., 2012. Development of integrated river basin management for Indus Basin: Challenges and Opportunities. Worldwide Funds for Nature Pakistan (Annual Report), Lahore, Pakistan. Available online at https://www. wwfpak.org/publication/pdf/irbm.
Kamaran Dastjerdi, H., Rajaei, S.A., Mansourian, H., 2022. The effects of urbanization on reduction of groundwater level and changes in vegetation and surface temperature in Iran's desert areas (case study: Yazd Province). International Journal of Environmental Research 16, 1–14.
Karami, S., Jalali, M., Karami, A., Katibeh, H., Fatehi Marj, A., 2022. Evaluating and modeling the groundwater in Hamedan plain aquifer, Iran, using the linear geostatistical estimation, sequential Gaussian simulation, and turning band simulation approaches. Modeling Earth Systems and Environment 1–22. https://link.springer.com/article/10.1007%2Fs40808-021-01295-1.
Kaur, N., Kaur, S., Kaur, P., Aggarwal, R., 2021. Impact of climate change on groundwater levels in Sirhind Canal Tract of Punjab, India. Groundwater for Sustainable Development 15, 100670.
Kaur, N., Kaur, S., Tsolakis, N., Mishra, N., Srai, J.S., 2023. Managing groundwater demand through surface water and reuse strategies in an overexploited aquifer of Indian Punjab. Modeling Earth Systems and Environment 9 (2), 2009–2026.
Khan, S.N., Li, D., Maimaitijiang, M., 2022. A Geographically Weighted Random Forest Approach to Predict Corn Yield in the US Corn Belt. Remote Sensing 14 (12), 2843. https://doi.org/10.3390/rs14122843.
Koch, J., Berger, H., Henriksen, H.J., Sonnenborg, T.O., 2019. Modelling of the shallow water table at high spatial resolution using random forests. Hydrology and Earth System Sciences 23 (11), 4603–4619. https://doi.org/10.5194/hess-23-4603-2019.
Lopes, M.E., 2019. Estimating the algorithmic variance of randomized ensembles via the bootstrap. The Annals of Statistics 47, 1088–1112. https://doi.org/10.1214/18-AOS1707.
Luo, Y., Yan, J., McClure, S., 2021. Distribution of the environmental and socioeconomic risk factors on COVID-19 death rate across continental USA: a spatial nonlinear analysis. Environmental Science and Pollution Research 28, 6587–6599.
Lytton, L., Ali, A., Garthwaite, B., Punthakey, J.F. and Saeed, B., 2021. Groundwater in Pakistan's Indus Basin. http://hdl.handle.net/10986/35065.
MacAllister, D., Krishan, G., Basharat, M., Cuba, D., MacDonald, A., 2022. A century of groundwater accumulation in Pakistan and northwest India. Nature Geoscience 15 (5), 390–396. https://doi.org/10.1038/s41561-022-00926-1.
MacDonald, A., et al., 2016. Groundwater quality and depletion in the Indo-Gangetic Basin mapped from in situ observations. Nature Geoscience 9 (10), 762–766. https://doi.org/10.1038/ngeo2791.
Manchuk, J.G., Deutsch, C.V., 2012. A flexible sequential Gaussian simulation program: USGSIM. Computers & Geosciences 41, 208–216. https://doi.org/10.1016/j.cageo.2011.08.013.
Mariethoz, G., Caers, J., 2014. Multiple-point geostatistics: stochastic modeling with training images. John Wiley & Sons. https://doi.org/10.1002/9781118662953.
Mehmood, K., Tischbein, B., Flörke, M., Usman, M., 2022. Spatiotemporal Analysis of Groundwater Storage Changes, Controlling Factors, and Management Options over the Transboundary Indus Basin. Water 14 (20), 3254.
Metahni, S., Coudert, L., Gloaguen, E., Guemiza, K., Mercier, G., Blais, J.-F., 2019. Comparison of different interpolation methods and sequential Gaussian simulation to estimate volumes of soil contaminated by As, Cr, Cu, PCP and dioxins/furans. Environmental Pollution 252, 409–419.
Noor, R., Arshad, A., Shafeeque, M., Liu, J., Baig, A., Ali, S., Maqsood, A., Pham, Q.B., Dilawar, A., Khan, S.N., Anh, D.T., Elbeltagi, A., 2023. Combining APHRODITE Rain Gauges-Based Precipitation with Downscaled-TRMM Data to Translate High-Resolution Precipitation Estimates in the Indus Basin. Remote Sensing 15 (2), 318.
Nourani, V., Khodkar, K., Gebremichael, M., 2022a. Uncertainty assessment of LSTM based groundwater level predictions. Hydrological Sciences Journal 67 (5), 773–790.
Nourani, V., Khodkar, K., Paknezhad, N.J., Laux, P., 2022b. Deep learning-based uncertainty quantification of groundwater level predictions. Stochastic Environmental Research and Risk Assessment 36 (10), 3081–3107.
Nourani, V., Tapeh, A.H.G., Khodkar, K., Huang, J.J., 2023. Assessing long-term climate change impact on spatiotemporal changes of groundwater level using

autoregressive-based and ensemble machine learning models. Journal of Environmental Management 336, 117653. https://doi.org/10.1016/j.jenvman.2023.117653.
Nussbaumer, R., Mariethoz, G., Gravey, M., Gloaguen, E., Holliger, K., 2018. Accelerating sequential gaussian simulation with a constant path. Computers & Geosciences 112, 121–132. https://doi.org/10.1016/j.cageo.2017.12.006.
Oikonomou, P.D., Alzraiee, A.H., Karavitis, C.A., Waskom, R.M., 2018. A novel framework for filling data gaps in groundwater level observations. Advances in Water Resources 119, 111–124. https://doi.org/10.1016/j.advwatres.2018.06.008.
Ostad-Ali-Askari, K., Shayannejad, M., 2021. Quantity and quality modelling of groundwater to manage water resources in Isfahan-Borkhar Aquifer. Environment, Development and Sustainability 23 (11), 15943–15959. https://doi.org/10.1007/s10668-021-01323-1.
Ostad-Ali-Askari, K., Shayannejad, M., Ghorbanizadeh-Kharazi, H., 2017. Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-rood River, Isfahan. Iran. KSCE Journal of Civil Engineering 21, 134–140. https://doi.org/10.1007/s12205-016-0572-8.
Ostad-Ali-Askari, K., Ghorbanizadeh Kharazi, H., Shayannejad, M., Zareian, M.J., 2019. Effect of management strategies on reducing negative impacts of climate change on water resources of the Isfahan-Borkhar aquifer using MODFLOW. River Research and Applications 35 (6), 611–631. https://doi.org/10.1002/rra.3463.
Ostad-Ali-Askari, K., Ghorbanizadeh Kharazi, H., Shayannejad, M., Zareian, M.J., 2020. Effect of climate change on precipitation patterns in an arid region using GCM models: case study of Isfahan-Borkhar Plain. Natural Hazards Review 21 (2), 04020006. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000367.
Pappas, C., Papalexiou, S.M., Koutsoyiannis, D., 2014. A quick gap filling of missing hydrometeorological data. Journal of Geophysical Research: Atmospheres 119 (15), 9290–9300. https://doi.org/10.1002/2014JD021633.
Pham, Q.B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Islam, A.R.M.T., Talukdar, S., Nguyen, X.C., Ahmed, A.N., Anh, D.T., 2022. Groundwater level prediction using machine learning algorithms in a drought-prone area. Neural Computing and Applications 34 (13), 10751–10773.
Qureshi, A., 2014. Conjunctive water management in the fixed rotational canal system: A case study from Punjab Pakistan. Irrigat Drainage Sys Eng 3 (122), 2. https://doi.org/10.4172/2168-9768.1000122.
Rahman, A.S., Hosono, T., Quilty, J.M., Das, J., Basak, A., 2020. Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms. Advances in Water Resources 141, 103595. https://doi.org/10.1016/j.advwatres.2020.103595.
Rodell, M., Famiglietti, J.S., Wiese, D.N., Reager, J.T., Beaudoing, H.K., Landerer, F.W., Lo, M.-H., 2018. Emerging trends in global freshwater availability. Nature 557 (7707), 651–659.
Sahoo, M., Das, T., Kumari, K., Dhar, A., 2017. Space–time forecasting of groundwater level using a hybrid soft computing model. Hydrological Sciences Journal 62 (4), 561–574. https://doi.org/10.1080/02626667.2016.1252986.
Saito, L., Christian, B., Diffley, J., Richter, H., Rohde, M.M., Morrison, S.A., 2021. Managing groundwater to ensure ecosystem function. Groundwater 59 (3), 322–333.
Sajjad, M.M., Wang, J., Abbas, H., Ullah, I., Khan, R., Ali, F., 2022. Impact of Climate and Land-Use Change on Groundwater Resources, Study of Faisalabad District. Pakistan. Atmosphere 13 (7), 1097.
Sanmugan Prathapar, B.H., Hashmi, M., Arslan, M., 2021. Institutional Transformation of the Punjab Irrigation Department to a Water Resources Department. Punjab Irrigation Department.
Sekulić, A., Kilibarda, M., Heuvelink, G.B., Nikolić, M., Bajat, B., 2020. Random Forest Spatial Interpolation. Remote Sensing 12 (10), 1687. https://doi.org/10.3390/rs12101687.
Siddiqi, A., Wescoat Jr, J.L., Muhammad, A., 2018. Socio-hydrological assessment of water security in canal irrigation systems: a conjoint quantitative analysis of equity and reliability. Water Security 4, 44–55. https://doi.org/10.1016/j.wasec.2018.11.001.
Singh, A., 2014. Groundwater resources management through the applications of simulation modeling: A review. Science of the Total Environment 499, 414–423. https://doi.org/10.1016/j.scitotenv.2014.05.048.
Smolenaars, W.J., Dhaubanjar, S., Jamil, M.K., Lutz, A., Immerzeel, W., Ludwig, F., Biemans, H., 2022. Future upstream water consumption and its impact on downstream water availability in the transboundary Indus Basin. Hydrology and Earth System Sciences 26 (4), 861–883.
Stewart, B., 2015. Measuring what we manage–the importance of hydrological data to water resources management. Proceedings of the International Association of Hydrological Sciences, 366: 80-85. https://doi.org/10.5194/piahs-366-80-2015.
Sun, A.Y., 2013. Predicting groundwater level changes using GRACE data. Water Resources Research 49 (9), 5900–5912. https://doi.org/10.1002/wrcr.20421.
Sun, J., Hu, L., Li, D., Sun, K., Yang, Z., 2022. Data-driven models for accurate groundwater level prediction and their practical significance in groundwater management. Journal of Hydrology 608, 127630. https://doi.org/10.1016/j.jhydrol.2022.127630.
Taie Semiromi, M., Koch, M., 2019. Reconstruction of groundwater levels to impute missing values using singular and multichannel spectrum analysis: application to the Ardabil Plain. Iran. Hydrological Sciences Journal 64 (14), 1711–1726. https://doi.org/10.1080/02626667.2019.1669793.
Takoutsing, B., Heuvelink, G.B., 2022. Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. Geoderma 428, 116192. https://doi.org/10.1016/j.geoderma.2022.116192.
Talebmorad, H., Ostad-Ali-Askari, K., 2022. Hydro geo-sphere integrated hydrologic model in modeling of wide basins. Sustainable Water Resources Management 8 (4), 118. https://doi.org/10.1007/s40899-022-00689-y.
Tang, Y., Zang, C., Wei, Y., Jiang, M., 2019. Data-driven modeling of groundwater level with least-square support vector machine and spatial–temporal analysis. Geotechnical and Geological Engineering 37 (3), 1661–1670. https://link.springer.com/article/10.1007/s10706-018-0713-6.
Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Heddam, S., Kim, S., Sulaiman, S.O., Tan, M.L., Sa'adi, Z., Mehr, A.D., Allawi, M.F., Abba, S.I., Zain, J.M., Falah, M.W., Jamei, M., Bokde, N.D., Bayatvarkeshi, M., Al-Mukhtar, M., Bhagat, S.K., Tiyasha, T., Khedher, K.M., Al-Ansari, N., Shahid, S., Yaseen, Z.M., 2022. Groundwater level prediction using machine learning models: A comprehensive review. Neurocomputing 489, 271–308.
Umar, M., Khan, S.N., Arshad, A., Aslam, R.A., Khan, H.M.S., Rashid, H., Pham, Q.B., Nasir, A., Noor, R., Khedher, K.M., Anh, D.T., 2022. A modified approach to quantify aquifer vulnerability to pollution towards sustainable groundwater management in Irrigated Indus Basin. Environmental Science and Pollution Research 29 (18), 27257–27278.
Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology 476, 433–441. https://doi.org/10.1016/j.jhydrol.2012.11.017.
Varouchakis, E., Hristopulos, D., Karatzas, G., 2012. Improving kriging of groundwater level data using nonlinear normalizing transformations—a field application. Hydrological Sciences Journal 57 (7), 1404–1419. https://doi.org/10.1080/02626667.2012.717174.
Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., Zhang, B.o., 2018. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. Applied Water Science 8 (5). https://doi.org/10.1007/s13201-018-0742-6.
Watto, M.A., Mugera, A.W., 2016. Groundwater depletion in the Indus Plains of Pakistan: imperatives, repercussions and management issues. International Journal of River Basin Management 14 (4), 447–458. https://doi.org/10.1080/15715124.2016.1204154.
Watto, M.A., Mitchell, M., Akhtar, T., 2021. Pakistan's Water Resources: Overview and Challenges. Water Resources of Pakistan: Issues and Impacts 1–12. https://doi.org/10.1007/978-3-030-65679-9_1.
Wescoat Jr, J.L., Siddiqi, A., Muhammad, A., 2018. Socio-hydrology of channel flows in complex river basins: Rivers, canals, and distributaries in Punjab, Pakistan. Water Resources Research 54 (1), 464–479. https://doi.org/10.1002/2017WR021486.
Wunsch, A., Liesch, T., Broda, S., 2018. Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). Journal of Hydrology 567, 743–758. https://doi.org/10.1016/j.jhydrol.2018.01.045.
Yang, R., Xing, B., 2021. A comparison of the performance of different interpolation methods in replicating rainfall magnitudes under different climatic conditions in Chongqing Province (China). Atmosphere 12 (10), 1318. https://doi.org/10.3390/atmos12101318.
Yin, W., Fan, Z., Tangdamrongsub, N., Hu, L., Zhang, M., 2021. Comparison of physical and data-driven models to forecast groundwater level changes with the inclusion of GRACE–A case study over the state of Victoria. Australia. Journal of Hydrology 602, 126735. https://doi.org/10.1016/j.jhydrol.2021.126735.
Zahran, H. et al., 2023. Impact of Urbanization on Groundwater and Surface Temperature Changes: A Case Study of Lahore City. Sustainability, 15(8): 6864. https://doi.org/10.3390/su15086864.
Zakeri, F., Mariethoz, G., 2021. A review of geostatistical simulation models applied to satellite remote sensing: Methods and applications. Remote Sensing of Environment 259, 112381. https://doi.org/10.1016/j.rse.2021.112381.
Zanotti, C., Rotiroti, M., Sterlacchini, S., Cappellini, G., Fumagalli, L., Stefania, G.A., Nannucci, M.S., Leoni, B., Bonomi, T., 2019. Choosing between linear and nonlinear models and avoiding overfitting for short and long term groundwater level forecasting in a linear system. Journal of Hydrology 578, 124015.
Zeydalinejad, N., 2022. Artificial neural networks vis-à-vis MODFLOW in the simulation of groundwater: A review. Modeling Earth Systems and Environment 8 (3), 2911–2932. https://doi.org/10.1007/s40808-022-01365-y.
Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0.
Zhang, W., He, Y., Wang, L., Liu, S., Meng, X., 2023. Landslide Susceptibility mapping using random forest and extreme gradient boosting: A case study of Fengjie. Chongqing. Geological Journal. 58 (5), 2372–2387. https://doi.org/10.1002/gj.4683.
Zhu, Y.u., Liu, S., Yi, Y., Xie, F., Grünwald, R., Miao, W., Wu, K., Qi, M., Gao, Y., Singh, D., 2021. Overview of terrestrial water storage changes over the Indus River Basin based on GRACE/GRACE-FO solutions. Science of the Total Environment 799, 149366.
Zi–chen, G., Tao, W., Shu–lin, L., Wen–ping, K., Xiang, C., Kun, F., Ying, Z., 2021. Comparison of the backpropagation network and the random forest algorithm based on sampling distribution effects consideration for estimating non photosynthetic vegetation cover. International Journal of Applied Earth Observation and Geoinformation 104, 102573.