

## Multi-model ensemble machine learning-based downscaling and projection of GRACE data reveals groundwater decline in Saudi Arabia throughout the 21st century



Arfan Arshad <sup>a,e</sup>, Muhammad Shafeeqe <sup>b,c</sup>, Thanh Nhan Duc Tran <sup>d</sup>, Ali Mirchi <sup>a,\*</sup>, Zaichen Xiang <sup>a</sup>, Cenlin He <sup>e</sup>, Amir AghaKouchak <sup>f,g,h</sup>, Jessica Besnier <sup>d</sup>, Md Masudur Rahman <sup>i</sup>

<sup>a</sup> Department of Biosystems and Agricultural Engineering, Oklahoma State University, Stillwater, OK, USA

<sup>b</sup> Alfred Wegener Institute for Polar and Marine Research, Bremerhaven 27570, Germany

<sup>c</sup> Climate Lab, Institute of Geography, University of Bremen, Bremen 28359, Germany

<sup>d</sup> Department of Civil and Environmental Engineering, University of Virginia, Charlottesville, VA 22904, USA

<sup>e</sup> Research Applications Laboratory, NSF National Center for Atmospheric Research, Boulder, CO, USA

<sup>f</sup> Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

<sup>g</sup> Department of Earth System Science, University of California, Irvine, USA

<sup>h</sup> United Nations University Institute for Water, Environment and Health (UNU-INWEH), Richmond Hill, Ontario, Canada

<sup>i</sup> Interdisciplinary Research Center for Aviation and Space Exploration (IRC-ASE), King Fahd University of Petroleum and Minerals, Dhahran 31261, Kingdom of Saudi Arabia

### ARTICLE INFO

#### Keywords:

Groundwater depletion  
GRACE  
Multi-model ensemble downscaling  
Machine learning  
Google Earth Engine  
Climate projections

### ABSTRACT

#### Study region:

Saudi Arabia.  
**Study focus:** The major goal of this study is to downscale GRACE (Gravity Recovery and Climate Experiment) groundwater storage (GWS) anomalies to assess the local-scale vulnerabilities of groundwater changes across western regions of Saudi Arabia (Al Jumum, Makkah, Jeddah, and Bahrah). This was accomplished by using multi-model ensemble machine learning (ML) approach leveraging Random Forest, CART, and Gradient Tree Boosting algorithms within Google Earth Engine (GEE). Additionally, we used the downscaled GWS and CMIP6 climate data with the Generalized Additive Model (GAM) to project the future GWS changes under climate change.

**New hydrological insights for the region:** The ensemble results demonstrated robust performance ( $R^2 = 0.92$  and RMSE = 20 mm) compared to the individual model ( $R^2 = 0.84$ –0.88 and RMSE = 25–28 mm). The areas of higher groundwater depletion were predominantly observed in Jeddah and Makkah, with average annual rates of – 165 mm/year and – 150 mm/year, respectively, from 2002 to 2023. The total volumetric losses range from 11.38 km<sup>3</sup> to 15.31 km<sup>3</sup> across different sub-regions. Seasonally, the peak GWS drop (-90 to – 125 mm) was detected during the summer months (April–July), aligning with periods of maximum water demand. Several key drivers that control the GWS changes were also identified, including anthropogenic effects, local climate anomalies, and large-scale climate oscillations. Projections for GWS reveal an irreversible decline throughout the 21st Century with potential reductions surpassing – 216 mm/year in

\* Corresponding author.

E-mail addresses: [aarshad@okstate.edu](mailto:aarshad@okstate.edu) (A. Arshad), [muhammad.shafeeqe@awi.de](mailto:muhammad.shafeeqe@awi.de) (M. Shafeeqe), [syu3cs@virginia.edu](mailto:syu3cs@virginia.edu) (T.N.D. Tran), [amirchi@okstate.edu](mailto:amirchi@okstate.edu) (A. Mirchi), [zaichen.xiang@okstate.edu](mailto:zaichen.xiang@okstate.edu) (Z. Xiang), [cenlinhe@ucar.edu](mailto:cenlinhe@ucar.edu) (C. He), [amir.a@uci.edu](mailto:amir.a@uci.edu) (A. AghaKouchak), [jb8qv@virginia.edu](mailto:jb8qv@virginia.edu) (J. Besnier), [mdmasudur.rahaman@kfupm.edu.sa](mailto:mdmasudur.rahaman@kfupm.edu.sa) (M.M. Rahman).

high-emission scenarios (SSP5-8.5). The developed approach is transferable to other regions for quantitative assessment of local groundwater changes.

## 1. Introduction

Groundwater is an essential resource that plays an important role in maintaining ecological development and ensuring food and water security for billions of people around the world (Gleeson et al., 2019; Bhattacharai et al., 2023). The reliance on groundwater has exacerbated groundwater overdrafts due to substantial population growth and increased water consumption (Jasechko et al., 2024). Groundwater in major cities is increasingly pumped through private wells for public and industrial water demands (Jasechko et al., 2024). The projected rise in environmental hazards and global changes, including elevated temperatures and heightened flash droughts, is expected to intensify groundwater depletion by 2100, potentially having profound effects on human health and the environment (Taylor et al., 2013; Mishra et al., 2024; Cotterman et al., 2018; Jódar et al., 2024; Tran et al., 2023).

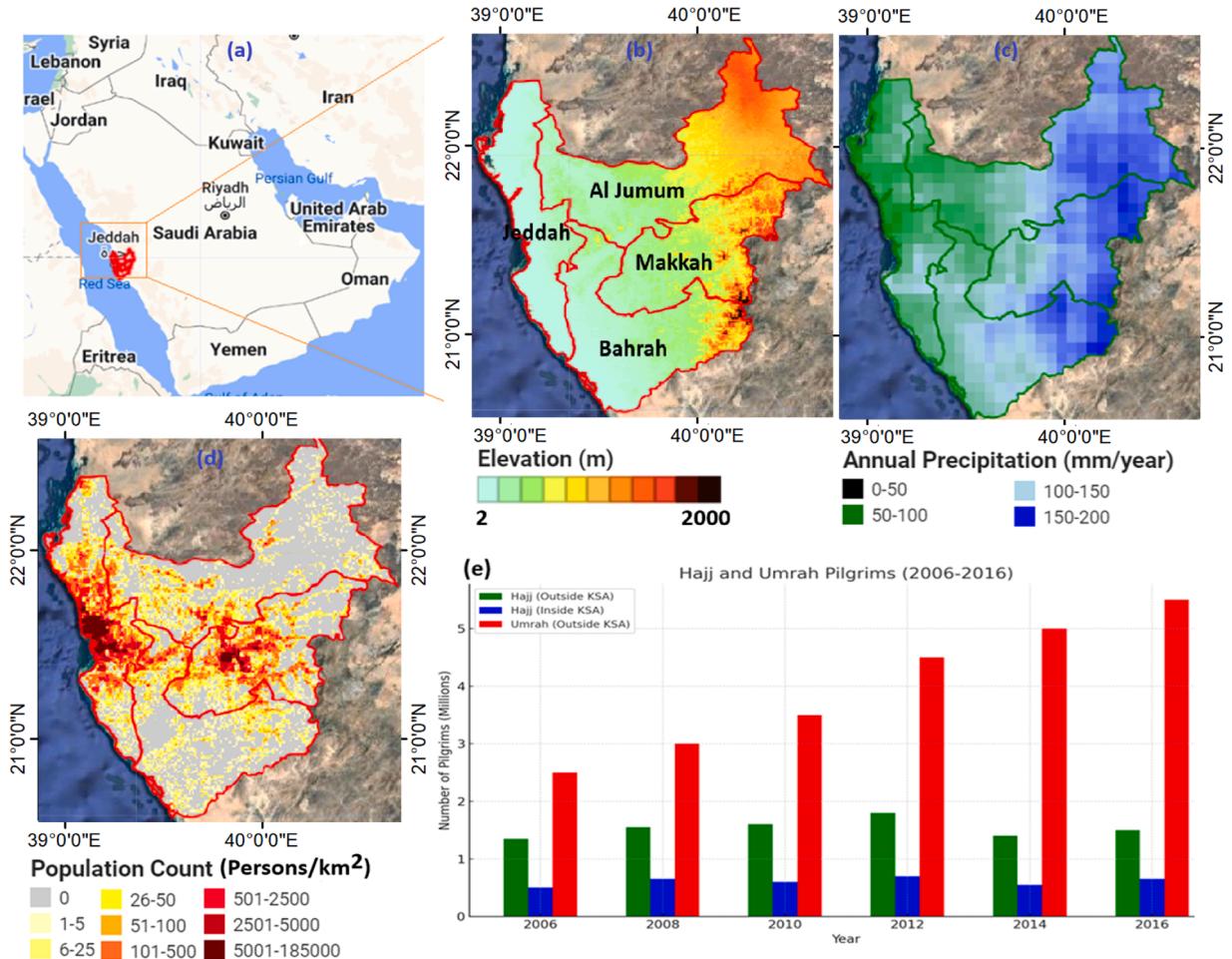
Groundwater sustainability challenges are most noticeable in the western coastal areas of Saudi Arabia (e.g., Makkah and Jeddah), facing heightened water scarcity driven by their growing populations and development initiatives under Saudi Vision 2030 (Altayyar et al., 2024; Abdella et al., 2024). Further dry climate conditions in the coastal areas exacerbate groundwater depletion when there is minimum opportunity for annual recharge. Every year, millions of people come to Makkah to perform religious activities (Hajj and Umrah), which puts more stress on the city's water supplies. Cities regularly need water, and religious tourists sometimes come in large numbers, which makes groundwater loss even worse (Ascoura et al., 2013; Abonomi et al., 2022). However, groundwater use in major urbanized regions is poorly monitored, leaving government authorities unaware of the issues related to the sustainability of groundwater-dependent rural and urban communities and the long-term availability of this resource. Lack of monitoring data also hinders the knowledge of how climate variability and environmental changes may impact groundwater availability. To address these concerns, high-resolution groundwater data is required to improve our understanding of localized groundwater changes and to identify areas that urgently require sustainable water management practices.

Several approaches have been employed to assess and monitor groundwater dynamics, including ground-based monitoring networks (MacAllister et al., 2022) and computational modeling techniques (Rajaeian et al., 2024). Ground-based monitoring data generally provides accurate estimations, while limited spatial coverage of in-situ observations compromises the large-scale applicability of decision-support water management models. In recent decades, advancements in remote sensing and geospatial approaches have proven reliable in unlocking new insights into groundwater dynamics (Liu et al., 2022; Ibrahim et al., 2024). The GRACE (Gravity Recovery and Climate Experiment) and its follow-on missions (GRACE-FO), launched by NASA in April 2002, provide valuable information on the monthly global terrestrial water cycle across time and space with global coverage using twin satellites (Tapley et al., 2004; Do et al., 2024a). GRACE/GRACE-FO raw data has been further processed by different research centers (e.g., JPL: Jet Propulsion Laboratory, CSR: Center of Space Research) to provide the equivalent water height or total water storage (TWS) data which includes groundwater, soil moisture, canopy water, snow/ice, lake and river water storage. The integration of GRACE/GRACE-FO datasets with land surface and hydrological modeling tools has effectively facilitated the study of groundwater storage (GWS) dynamics at both basin and watershed scales (Chen et al., 2019; Ahmed, 2020; Ahmed et al., 2021; Kim et al., 2024a, 2024b; Jawadi et al., 2024). Nevertheless, the current GRACE products exhibit insufficient spatial detail (e.g., ~ 55 km JPL Mascon) limits their direct application in local-scale water management decisions (Arshad et al., 2024).

Spatial downscaling can improve the resolution of remote sensing products to more effectively meet community information needs by leveraging remote sensing capabilities at decision-relevant scales. Previous studies have employed downscaling approaches to improve the existing information from remote sensing products to investigate vulnerabilities due to various environmental changes (Liu et al., 2020; Chen et al., 2020; Luo et al., 2023). Spatial downscaling relies on the assumption that predictive relationships can be established between coarse-resolution remote sensing data (e.g., groundwater) and other accessible high-resolution remote sensing variables (e.g., precipitation, soil moisture), which are subsequently utilized to refine the coarse-resolution datasets to a finer scale (Q. Chen et al., 2020; C. Chen et al., 2020). Implementing spatial downscaling with ML and artificial intelligence (AI) techniques facilitates enhancing the capabilities of existing remote sensing datasets and improves assessment of groundwater vulnerability at scales that are relevant for local decision-making (Chen et al., 2020; Agarwal et al., 2023; Wang et al., 2024; Uz et al., 2024; Marshall et al., 2025). Various ML/AI approaches have been employed to develop predictive relationships for downscaling. For example, Q. Chen et al. (2020), C. Chen et al. (2020) downscaled Soil Moisture Active Passive (SMAP) satellite from 36 km to 1 km resolution using Random Forest (RF) algorithm to study localized soil moisture changes. Ali et al. (2022) downscaled GRACE data to explore the local scale vulnerabilities of groundwater droughts in the irrigated Indus Basin, reporting that Extreme Gradient Boosting (XGBoost) performed slightly better than Artificial Neural Network (ANN). Gou and Soja (2024) proposed a globally generalizable deep learning-based assimilation model, which significantly improved GRACE-derived total water storage anomalies (TWSAs) at 0.5° resolution. Yin et al. (2022) used multivariable linear regression (LR) and random forest (RF) to enhance GRACE-based water storage estimates from ~ 110 km<sup>2</sup> to ~ 27 km<sup>2</sup>, and their results suggested superiority of RF over LR. The majority of published studies mainly focus on single model applications in downscaling GRACE data. However, each of these ML/AI approaches have distinct strengths and inherent limitations based on various computing algorithms and the complexity of data (Senanayake et al., 2024; Akbar et al., 2025). Different training algorithms can result in distinct spatial patterns in downscaled outputs, which exhibit deviation from original data (Li et al., 2019a, 2019b; Arshad et al., 2024; Khosravi et al., 2025). Combining the strength of individual ML models offers an opportunity to generate robust downscaled outputs.

The advancements in Multi-Model Ensemble (MME) approaches have gained interest in climate and hydrological research. MME can reduce biases inherent in individual models by combining outputs from different models and yield improved estimates (Wang et al., 2023; Akbar et al., 2025). Studies documented that combining outputs of ML algorithms in MME approach significantly improve the accuracy of regional hydro-climate projections (Bilbao-Barrenetxea et al., 2024; Khosravi et al., 2025). Three ensemble methods including Bagging, Boosting, and Stacking have been widely used in MME frameworks (Dou et al., 2020). Among these, stacking algorithms via a meta-learner have received much attention due to their simplicity and capability to integrate predictions from multiple ML-base models (Rice and Emanuel, 2017; Zhai and Chen, 2018; Sun and Trevor, 2018; Khosravi et al., 2025). Despite the growing application of MME in hydro-climate research, the use of stacking MME approach in downscaling GRACE data from multiple ML models remains unexplored, particularly in the context of Saudi Arabia. The integration of MME in GRACE downscaling framework is particularly important for improving the reliability of downscaled-GWS estimates from GRACE data to support more effective water management strategies in data-scarce, vulnerable regions. Recognizing the need for a more robust downscaling predictive framework, a major goal of this research is to apply the Super Learner stacking ensemble approach and reduce model-specific biases in downscaled GWS estimates.

Implementing spatial downscaling with AI/ML is inherently a computationally demanding task. Researchers have utilized several open-source tools (ArcGIS) (Yang et al., 2021) and programming languages (e.g., RStudio) (Hossein Abadi et al., 2023; Roy et al., 2024). Nonetheless, acquiring data from various sources and preparing it for input into ML/AI models introduces challenges for researchers attempting to implement downscaling over big regions. Advancements in cloud computing platforms such as Google Earth Engine (GEE) facilitate accessing different remotely sensed datasets at various spatial and temporal scales and make them ready for various applications (Amani et al., 2020). The integration of ML/AL tools with spatial downscaling in GEE can diminish the expenses associated with data downloading and preprocessing (such as resampling and reprojecting) while efficiently expediting the



**Fig. 1.** (a) Geographic location of the study area in Saudi Arabia, (b) elevation distribution (compared to mean sea level; m.s.l.), (c) annual precipitation (mm/year), (d) population counts (estimate: persons/km<sup>2</sup>), and (e) number of people visiting the study area for religious activities. Precipitation estimates are based on ERA5-land products. Hajj and Umrah Pilgrims data are taken from published literature, and the graph is reproduced (Yezli et al., 2017). Population estimates are based on LandScan data.

downscaling process. In this study, we take an opportunity to implement the ML/AI-based spatial downscaling in the GEE. To our knowledge, there is no prior study that has downscaled GRACE data to 1 km<sup>2</sup> for advancing groundwater risk assessment in Saudi Arabia, especially using ensemble ML methods.

This paper advances the ability to provide high-resolution (1 km<sup>2</sup>) information on groundwater storage (GWS) changes between 2002 and 2023 by using multi-model ensemble machine learning-based downscaling and GRACE data in the GEE. Further, we incorporate the CMIP6 (Climate Model Intercomparison Project Phase 6) climate data between 2024 and 2099 to project GWS changes. To overcome the limitations and uncertainties of individual ML regressors, we introduce a novel framework that employs multi-model ensemble downscaling by combining the outputs of various ML regressions such as Random Forest, Gradient Tree Boost, and CART. The overall goal is to explore the local scale vulnerabilities of groundwater risk associated with climate change and anthropogenic stressors, facilitating local decision-making for effective water resource management. The primary contributions of the paper are three-fold: (1) developing a multiple-model ensemble downscaling framework combining RF, CART, and GBDT algorithms to produce robust downscaling results, (2) utilizing downscaled data to assess spatial-temporal changes in GWS, and (3) projecting future GWS changes through 2099 under different Greenhouse Gas (GHG) emission scenarios represented by Shared Socioeconomic Pathways (SSPs 1–2.6, 2–4.5, 3–7.0, and 5–8.5). Our downscaling approach has the advantage of not requiring complex parameters, making them easily adaptable to other global regions.

## 2. Material and methods

### 2.1. Study area

The research areas are the Saudi Arabia, focusing on four interconnected urban regions that include Jeddah, Makkah, Al Jumum, and Bahrah (Fig. 1a). Jeddah is the largest port city located along the Red Sea and serves as the main entry point for pilgrims traveling to the holy cities (Almazroui et al., 2013). Makkah is located 65 km away from Jeddah and as the holiest city in Islam, it hosts millions of pilgrims from around the world each year for the Hajj and Umrah (Gazzawe and Albahar, 2024). Elevations range from 2 m in Jeddah to approximately 2000 m in the mountains east of Makkah (Fig. 1b). The study area experiences a hot desert climate with mean annual temperatures surpassing 30°C in coastal areas and relatively lower in high-altitude regions (Mir and Ashraf, 2023). The study area receives annual precipitation up to ~ 200 mm (Fig. 1c). Makkah and Jeddah cities host millions of residents with population counts exceeding 5000 people/km<sup>2</sup> (Fig. 1d). Additionally, millions of international pilgrims visit Makkah to perform religious activities (Hajj and Umrah) (Fig. 1e).

### 2.2. GRACE/GRACE-FO data and estimating groundwater changes

We used GRACE/GRACE-FO monthly gridded datasets which represent the total water storage (TWS) anomalies relative to a mean of 2004–2009. This data was released by JPL (Jet Propulsion Laboratory) which has undergone the Mascon approach (RL06.1Mv03) and CRI (Coastal Resolution Improvement) to minimize the leakage errors of land-ocean pixels. It is accessible in GEE from April 2002 to the present, with a spatial resolution of around 55 km. There is no data available on the scaling factor in GEE. We obtained the CLM4.SCALE\_FACTOR.JPL.MSCNv03CRI scaling factor file from PODAAC ([https://podaac.jpl.nasa.gov/dataset/TELLUS\\_GRAC-GRFO\\_MASCON\\_CRI\\_GRID\\_RL06.1\\_V3](https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC-GRFO_MASCON_CRI_GRID_RL06.1_V3)) and subsequently uploaded it to the GEE assets folder. A scaling factor was applied to restore the original signals in GRACE data that were lost in the pre-processing. TWS anomalies represent contributions of groundwater storage (GWS), soil moisture (SMS), canopy water storage (CWS), snow water storage (SnWS), and surface water storage (SWS) anomalies (Giroto and Rodell, 2019; Li et al., 2019a, 2019b), and can be expressed as follows:

$$\text{TWS anomalies} = (\text{GWS} + \text{SM} + \text{CWS} + \text{SnWS} + \text{SWS}) \text{ anomalies} \quad (1)$$

Since snow, canopy, and surface water storage are negligible in the study area (Fig. S2), the TWS can be expressed as follows:

$$\text{TWS anomalies} = \text{GWS anomalies} + \text{SMS anomalies} \quad (2)$$

To derive GWS for groundwater quantity estimation, we subtract SM anomalies from the GRACE TWS anomalies. This assumption is consistent with documented studies in Saudi Arabia (e.g., Alshehri and Mohamed, 2023a, 2023b, 2023c; Mohamed et al., 2022), which have employed similar methodologies.

$$\text{GWS anomalies} = \text{TWS anomalies} - \text{SMS anomalies} \quad (3)$$

Since SM monitoring data is unavailable for this region, we used the data from ERA5-Land (Muñoz Sabater et al., 2021), a reanalysis dataset that combines model and observational data globally to provide land and atmospheric variables. ERA5-Land provides volumetric soil moisture data up to 280 cm depth across four different SM layers at a spatial resolution of ~ 11 km. We converted the ERA5-Land SMS into its anomalies by subtracting its mean of 2004–2009. We upscaled the ERA5-Land SMS anomalies to a 55 km resolution and subsequently subtracted it from the GRACE TWS anomalies to estimate GWS anomalies. Monthly GWS dynamics were hence estimated from 2002 to 2023.

TWS anomalies from GRACE and GRACE-FO datasets contain periodic gaps of 31 months from 2002 to 2023. To maintain the continuity of estimates of GWS anomalies, we applied a data-driven gap-filling approach to reconstruct the missing data of TWS anomalies for 31 months. We used the STL (Seasonal Trend Loess) statistical method (Cleveland et al., 1990) to estimate TWSA

anomalies for missing months using existing data of GRACE and GRACE-FO. We selected STL because of its efficient performance documented in the literature (Khorrami et al., 2023; Ali et al., 2024) as well as simplified algorithms in Rstudio (<https://cran.r-project.org/web/packages/stlplus/index.html>). We passed monthly TWS anomalies from the STL algorithm which estimates trend (T), seasonality (S), and residual (R) in the time series (t). STL was fitted with 1 degree of the locally-fitted polynomial in trend and seasonal extraction.

$$\text{TWS}_t \text{ anomalies} = T_t + S_t + R_t \quad (4)$$

We trained the STL algorithms on 80 % of TWS anomalies and kept 20 % of the data to ensure the validity of estimated results for missing months. After performing the testing evaluations, we used the trained STL to estimate the trend for the missing months and subsequently combined it with long-term mean seasonal and residual components for the reconstructing TWS anomalies as follows:

$$\text{TWS}_t \text{ anomalies} = y(T_t) + \text{mean}(S_t + R_t) \quad (5)$$

### 2.3. Input variables/predictors for spatial downscaling

The main input datasets used in this study include actual evapotranspiration (ET), population density, human modification index (HMI), surface elevation, land cover, land surface temperature (LST), mean air temperature, precipitation, and soil moisture (SM). These variables were selected as predictors to downscale the GRACE data, based on their ability to capture groundwater storage variability and their accessibility in the GEE.

Elevation data was extracted from the Global 30 Arc-Second Elevation (GTOPO30) at 30-arc seconds (~ 1 kilometer) provided by the USGS Earth Resources Observation and Science (EROS) Center. Annual population density data was obtained from LandScan at 1-km resolution from 2002 to 2022. It is developed using advanced spatial modeling techniques and geospatial data sources (Sims et al., 2023). Land cover data was derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) MCD12Q1 Version 6.1 dataset, available annually at a ~ 0.5 km resolution and developed from combined Terra and Aqua sensors (<https://lpdaac.usgs.gov/products/mcd12q1v061/>). We used monthly LST composites of day and night data from MOD21C3.061 at ~ 1 km (<https://lpdaac.usgs.gov/products/mod21c3v061/>). The human modification (HM) data at 1 km resolution was taken from a previously published study, which measures changes in terrestrial lands (Kennedy et al., 2019).

Monthly precipitation, mean air temperature, and actual evapotranspiration data were obtained from TerraClimate at ~ 5 km resolution. TerraClimate estimates are based on combining climatically aided interpolation, high-resolution climatological data from WorldClim, time-varying data from CRU Ts4.0, and the Japanese 55-year Reanalysis (JRA55) (Abatzoglou et al., 2018). Monthly SM data was based on ERA5-Land estimates at ~ 11 km resolution (Muñoz Sabater, 2019). We resampled TerraClimate and ERA5-Land variables to a 1 km resolution using a bilinear resampling approach in GEE. Since elevation, population density, human modification, and land cover datasets are not available in monthly time steps, we used the long-term average of these variables from 2002 to 2023 to capture the impacts of anthropogenic factors in GWS variability. Temperature, precipitation, evapotranspiration, SM, and LST were used as time-varying variables with monthly inputs in the downscaling process.

### 2.4. GEE, machine learning model, and spatial downscaling

#### 2.4.1. Google earth engine (GEE)

GEE is a cloud-based geospatial analysis platform that enables scalable processing of Earth observation datasets without the need for local computing resources. We executed all modeling and visualization processes in the GEE code editor environment. GRACE/GRACE-FO TWS and GWS data, and high-resolution predictors were accessed and pre-processed on the GEE platform to prepare them for the downscaling framework (Gorelick et al., 2017). We performed downscaling using three ML packages available in GEE that are smileRF (<https://developers.google.com/earth-engine/apidocs/ee-classifier-smilerandomforest>), smileCART (<https://developers.google.com/earth-engine/apidocs/ee-classifier-smilecart>), and smileGTB (<https://developers.google.com/earth-engine/apidocs/ee-classifier-smilegradienttreeboost>).

#### 2.4.2. Description of ML base models

RF model is an ensemble of decision trees (i.e.,  $T_1, T_2, \dots$ , and  $T_n$ ) that independently predict the target variable (GRACE GWS). It randomly samples data, constructs multiple decision trees, and generates a prediction from each tree. This model effectively handles non-linear and complex relationships among variables. The final prediction is the average of all tree predictions. The RF prediction  $\widehat{y}_{RF}$  for a given input X (which includes predictor variables such as soil moisture, precipitation, etc.) is expressed as follows:

$$\widehat{y}_{RF}(X) = \frac{1}{n} \sum_{i=1}^n T_i(X) \quad (6)$$

where  $T_i(X)$  is the prediction of the ith tree for input X, and n is the total number of trees in the forest.

GTB is an iterative method in which each new tree  $T_k$  is trained to correct the residuals (errors) from the previous iteration. GTB updates residuals in each tree before moving to the next. However, instead of averaging, GTB makes the final prediction as a sum of the

predictions from all trees, weighted by a learning rate  $\eta$ . The GTB prediction  $\widehat{y}_{GTB}$  after K iterations for input X is expressed as follows:

$$\widehat{y}_{GTB}(X) = \sum_{k=1}^K \eta \cdot T_k(X) \quad (7)$$

where,  $T_k(X)$  is the prediction of the kth tree for input X, K is the number of boosting iterations, and  $\eta$  is the learning rate, which controls the contribution of each tree to the final prediction.

The CART algorithm is a decision tree model that recursively splits the data based on predictor variables to minimize variance (or impurity) at each node. The prediction of the CART model is based on a single decision tree T, which directly predicts the target variable for a given input X. The CART prediction  $\widehat{y}_{CART}$  for input X is expressed as follows:

$$\widehat{y}_{CART}(X) = T(X) \quad (8)$$

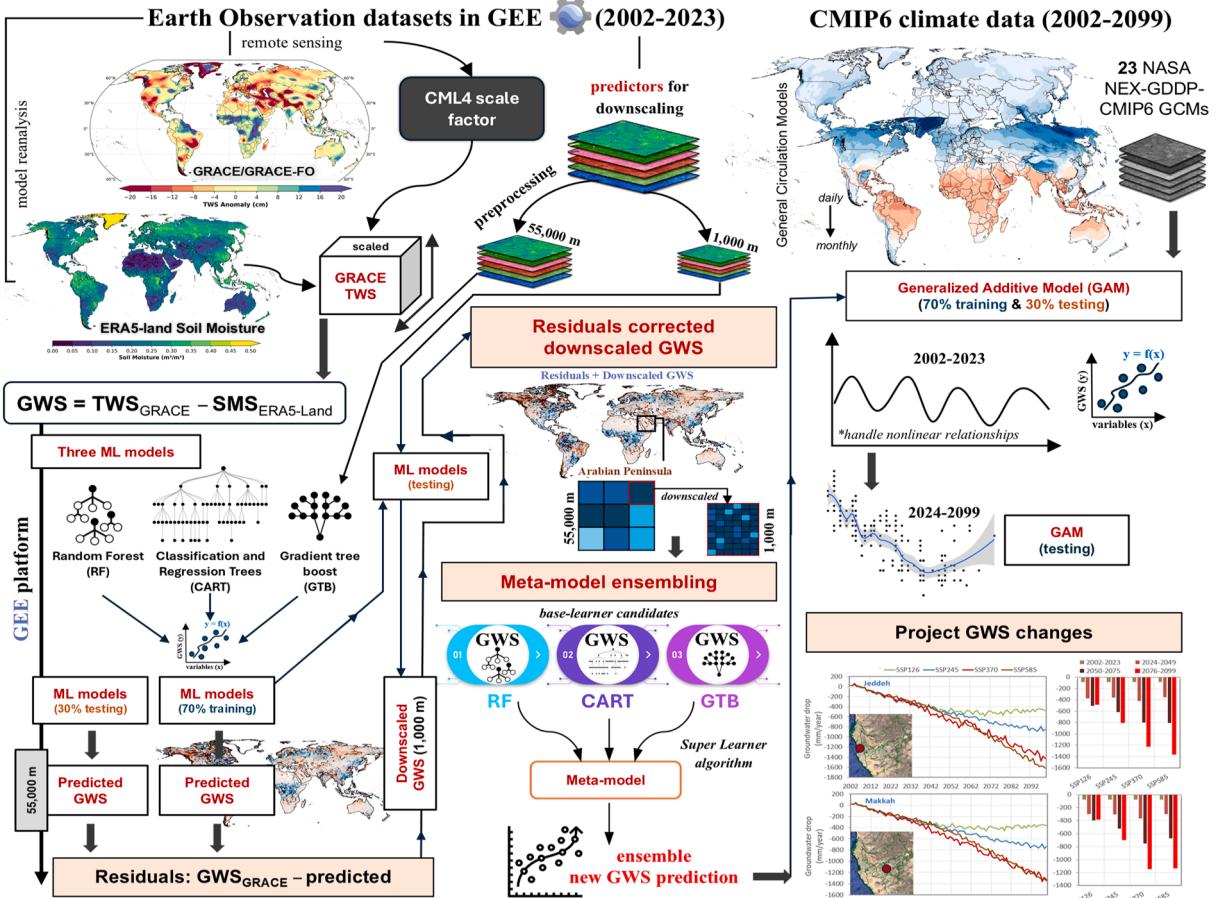
#### 2.4.3. Spatial downscaling

The spatial downscaling process can be mathematically represented as  $Y = f(X_i) + \epsilon$ , where Y is the downscaled GWS data,  $X_i$  represents predictor variables, and  $\epsilon$  is the predictive error that's not explained by the model reference to the observed GRACE GWS anomalies. Our downscaling approach involved the following steps:

**Training and testing at a coarse scale (55 km):** In the first step, data of all covariates predictors (ELV, LC, PD, NDVI, ET, LST, PRECIP, TEMP, SM) were aggregated to the resolution of GRACE/GRACE-FO ( $\sim 55$  km) which is divided into 70/30 train-test using “randomColumn()” function with a fixed seed in GEE. ML modes were trained at 70% of GRACE grids to establish a correlation function between GRACE GWS anomalies and covariates predictors at a coarse resolution ( $\sim 55$  km) using the following equations.

$$\text{GWS anomalies}_{(\text{training}70\%, 55 \text{ km})} = f(\text{ELV}, \text{LC}, \text{PD}, \text{NDVI}, \text{ET}, \text{LST}, \text{PRECIP}, \text{TEMP}, \text{SM})_{(55 \text{ km})} \quad (9)$$

where ELV, LC, PD, NDVI, ET, LST, PRECIP, TEMP, and SM represent elevation, land cover, population density, normalized difference



**Fig. 2.** Methodological framework comprising GEE (Google Earth Engine), ML (machine learning), spatial downscaling, and climate projections of groundwater changes.

vegetation index, actual evapotranspiration, land surface temperature, precipitation, air temperature, and SM, respectively. Trained ML models were further used with new 30% data of covariates predictors (ELV, LC, PD, NDVI, ET, LST, PRECIP, TEMP, SM) to estimate the GWS anomalies at testing grids.

$$\text{GWS anomalies}_{(\text{testing}30\%, 55 \text{ km})} = f(\text{ELV}, \text{LC}, \text{PD}, \text{NDVI}, \text{ET}, \text{LST}, \text{PRECIP}, \text{TEMP}, \text{SM})_{(55 \text{ km})} \quad (10)$$

GWS anomalies at testing grids were compared with the observed GRACE GWS anomalies to ensure the training accuracy of ML models. We calculated the  $R^2$ , root mean square error (RMSE: mm), and bias (%) matrices to evaluate the training and testing performance of ML models.

**Residuals estimations (1 km):** The model's predicted GWS anomalies at the training and testing grids were compared with the observed GRACE GWS anomalies to calculate the training and testing residuals. Residuals obtained from both training and testing grids were spatially merged/combined to generate 55 km residual map. Residuals at 55 km resolution were interpolated to 1 km using a bilinear approach adopting a previously published method (Wang et al., 2024; Arshad et al., 2024).

**Testing at high-resolution and downscaling estimates (1 km):** We applied regional downscaling which works on the assumption that the statistical relationship developed between the GRACE GWS anomalies and predictor variables at 55 km resolution can be applied with 1 km predictors. Thus, we tested the ML models with high-resolution (1 km) covariates predictors (ELV, LC, PD, NDVI, ET, LST, PRECIP, TEMP, SM). ML models at this stage used the coefficients and non-linear transformation function from the training stage with high-resolution (1 km) covariates predictors to make a new prediction of high-resolution estimates of GWS anomalies following the established regional downscaling approach (Yin et al., 2022a, 2022b):

$$\text{GWS anomalies}_{(1\text{km})} = f(\text{ELV}, \text{LC}, \text{PD}, \text{NDVI}, \text{ET}, \text{LST}, \text{PRECIP}, \text{TEMP}, \text{SM})_{(1\text{km})} \quad (11)$$

**Residual correction of downscaled estimates (1 km):** Since the model output at this stage does not inherently include error estimates. Following the previously established methods (Wang et al., 2024; Arshad et al., 2024), a residual correction was performed by adding 1 km residuals from training and testing with the high-resolution estimates of *GWS anomalies* (1 km).

In the regional-scale downscaling approach, ML models were spatially trained using GRACE GWS pixels within the study region for a single time step. For example, we selected a specific time point (e.g., January 2003) and trained models using GRACE GWS anomalies at coarse resolution (55 km) alongside covariates predictors like LST, SM, PD, LC, precipitation, etc. The trained function ( $Y = f(X_i)$ ) was then applied to the 1-km predictors. We trained the models spatially for individual time steps (each month) and the process was repeated for all 260-time steps (April 2002–December 2023), generating 1 km monthly GWS anomalies for the entire study region. We used a time-step-based regional-scale downscaling approach over modeling individual grid cell time series to effectively capture the spatial heterogeneity of GWS dynamics and its relationship with environmental variables. Additionally, few predictor variables are static (e.g., elevation, land cover) as well as temporally coarse (e.g., annual population density), which limits their applicability in developing model training with monthly time series over a single grid. This approach has been widely used by several studies (Sabzehee et al., 2023; Ali et al., 2024; Arshad et al., 2024; Roy et al., 2024; Wang et al., 2024). The methodological procedures and downscaling framework are detailed in Fig. 2, showing the integration of GEE and ML algorithms.

#### 2.4.4. Implementation and hyperparameter tuning in GEE

Skill evaluation of the model's predictive performance and parameter tuning are important steps of downscaling. In GEE, we used “randomColumn()” function with a fixed seed to generate a 70/30 train–test split of spatial samples. A grid-based hyperparameter tuning approach was performed to choose the best combination of hyperparameters. Before performing tuning, we specified pre-defined ranges of parameters from published studies and recommended values. The hyperparameters for RF involved the number of trees (1–500), variables per split (1–7), maximum number of nodes (5–30), and minimum leaf population (1–5). In the case of GTB tuning involves the number of trees (1–500), learning rate or shrinkage (0.1–0.3), sampling rate (0.5–1.0), and maximum nodes (5–30). For CART, we tested maximum nodes (5 – 200) and leaf sizes (1–5). Models were trained using 70 % data with predefined ranges of hyperparameters and skilled performance was evaluated using root mean square error (RMSE) against the 30 % test data. Currently, GEE lacks the configuration of a K-fold cross-validation approach. Therefore, we carried K-fold manual grid search across the predefined ranges of parameters using a hold-out validation approach. By repeating training and testing over multiple spatial samples, we identified the best combination of hyperparameters that yield lower RMSE against the observations on test data.

#### 2.5. Stacking and meta-modeling ensembling

We constructed a meta-model by integrating predictions of three base models RF, CART, and GTB to enhance the predictive accuracy of downscaled GWS anomalies. The output base models were used as input features in the meta-model developed using the “Super Learner- SL” algorithm (Van der Laan et al., 2007). SL algorithm integrates base learners in a cross-validated (CV) framework to retrieve optimized prediction. We used the CV.SuperLearner() function in Rstudio to perform the stacking process. We selected the non-negative least squares (method.NNLS) method in CV.SuperLearner() function to estimate the coefficients for the super learner that combine the relative strengths of the base models while reducing overfitting in the CV framework. Super Learner dynamically adjusts model weights based on performance, providing a more accurate ensemble output. We used SL.library which facilitates computation of V-fold cross-validation, enabling the meta-learner to reduce overfitting and retrieve optimal performance from base models. More details about CV.SuperLearner() and SL.library can be found at <https://www.rdocumentation.org/packages/SuperLearner/versions/2.0-29/topics/SL.library>.

## 2.6. Potential drivers of groundwater changes

We examined potential linkages between climate and human impacts on GWS variability, assessing both large-scale and local-scale footprints. The impacts of the large-scale mode of variability for the coupled atmospheric-ocean systems affecting groundwater changes were assessed using El Niño-Southern Oscillation (ENSO), Indian Dipole Ocean (IOD), and Pacific Decadal Oscillation (PDO) indices provided by NOAA (National Oceanic and Atmospheric Administration). We collected monthly data on these variables from 2002 to 2023. The groundwater loss/gain was calculated during the composite anomalies of cold (negative anomalies) and warm (positive anomalies) phases of the large-scale mode of climate variability. Local-scale footprints were assessed using precipitation, temperature, and soil moisture changes. Data of these variables are based on the ERA5-land product. Further, the local-scale footprint of groundwater storage variability was also interpreted based on the population density dynamics from 2002 to 2023.

## 2.7. Projected changes in groundwater

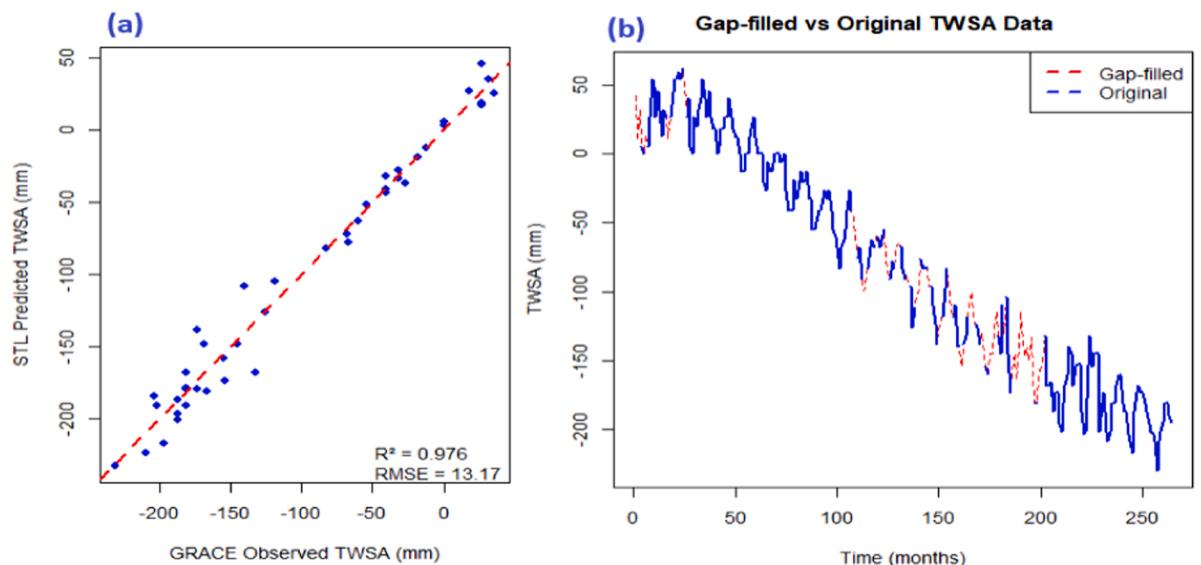
To project future groundwater changes, we incorporated CMIP6 climate data as predictors, including mean air temperature, precipitation, and a heatwave index (calculated as the number of days with temperatures exceeding 35°C). We utilized ensemble outputs from 23 General Circulation Models (GCMs) provided in the most recent release of CMIP6. These models were selected to examine the projected climate impacts through 2100. The GCM data used were based on variant labels of r1i1p1f1 and included simulations under various SSP GHG emission scenarios (Eyring et al., 2016). Specifically, we selected four SSP scenarios, i.e., SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, representing very low (or sustainability), low, medium (or medium-high), and high-end GHG emissions, respectively, to assess the potential effects of climate change on groundwater resources. We utilized the bias-corrected, high-resolution CMIP6 climate data at  $0.25^\circ \times 0.25^\circ$  (approximately  $25 \times 25$  km) from the NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP-CMIP6) (<https://www.ncdc.noaa.gov/services/data-collections/land-based-products/nex-gddp-cmip6>), with daily data accessed via GEE. The datasets have undergone the Bias-Correction Spatial Disaggregation (BCSD) method, a widely adopted trend-preserving statistical downscaling technique that generates precise, high-resolution data from GCMs (Tapas et al., 2024; Lakshmi, 2024b; Do et al., 2024b; Nguyen et al., 2024).

The Generalized Additive Model (GAM) (Wood, 2020) is employed to project GWS changes using CMIP6 predictors (temperature, precipitation, and heatwaves) and monthly downscaled GWS from 2002 to 2023. The model was trained using downscaled GWS data from 2002 to 2023. We used 70 % of the GWS data in training GAM and then 30 % as testing data to evaluate the model performance with selected predictors. Finally, after model testing evaluation, the trained algorithms were applied with CMIP6 predictors for 2024–2099 to quantify potential changes in groundwater storage for the near-future (2024–2049), mid-future (2050–2075), and far-future (2076–2099) across four sub-regions, including Makkah, Jeddah, Al Jumum, and Bahrah.

We selected the GAM approach because of its capability to capture and handle nonlinear relationships and efficiently deal with complex environmental variables (Murphy et al., 2019). The GAM model used in this study is expressed as follows:

$$GWS = \beta_0 + s(PR, k) + s(TEMP, k) + s(NHWD, k) + \epsilon \quad (12)$$

where GWS is monthly downscaled GWS data from 2002 to 2023. PR, TEMP, and NHWD represent precipitation, temperature, and



**Fig. 3.** (a) Numerical evaluations and comparison of STL-based predicted TWSA with GRACE-original TWSA on testing data of randomly selected 45 months from 2002 to 2023, (b) regional time series of STL-based reconstructed (gap-filled) and GRACE-original TWSA from 2002 to 2023.

number of heatwave days, respectively.  $\beta_0$  is the model's constant (intercept).  $s(x, k)$  represents the spline function (Wood et al., 2016) applied to each predictor.  $k$  is a smoothing parameter that constrains the degrees of freedom to prevent overfitting. We performed analysis using the “mgcv” library in the RStudio.

## 2.8. Comparison of GRACE-based GWS anomalies with publically accessible datasets

Due to the absence of groundwater data from monitoring wells in the study region. We compared our findings with the published literature and the groundwater storage anomalies simulated by the Global Hydrological Model (WaterGAP v2.2d) from 2002 to 2019. WaterGAP estimates global surface water and human water consumption (Müller Schmied et al., 2021) (<https://doi.pangaea.de/10.1594/PANGAEA.948461?format=html#download>). WaterGAP models the effects of human activities on global water resources and has been employed alongside GRACE to calculate non-anthropogenic groundwater variations (Li et al., 2019a, 2019b). Additionally, we also compared our results with GLWS 2.0 which provides global total water storage anomalies including groundwater and soil moisture at  $0.5^\circ \times 0.5^\circ$  spatial resolution from 2003 to 2019 (Eicker et al., 2014; Gerdener et al., 2023). More details about the GLWS 2.0 data are available at <https://doi.pangaea.de/10.1594/PANGAEA.954742?format=html#download>

## 3. Results

### 3.1. Reconstructed TWS anomalies for bridging the gap between GRACE and GRACE-FO

Fig. 3 presents the numerical analysis of reconstructed TWS anomalies for 31 periodic months in GRACE and GRACE-FO datasets. The results from testing data over 45 months (Fig. 3a) demonstrated that the STL algorithm effectively estimates the hidden TWS anomalies, achieving a higher  $R^2$  (0.97) and a smaller RMSE ( $\sim 13$  mm) in comparison to the observed GRACE and GRACE-FO TWSA anomalies. The temporal variations and time series of gap-filled TWS anomalies exhibit strong agreement with observed GRACE/GRACE-FO TWS anomalies while preserving the seasonality and overall trend of the data (Fig. 3b). These findings underscore the

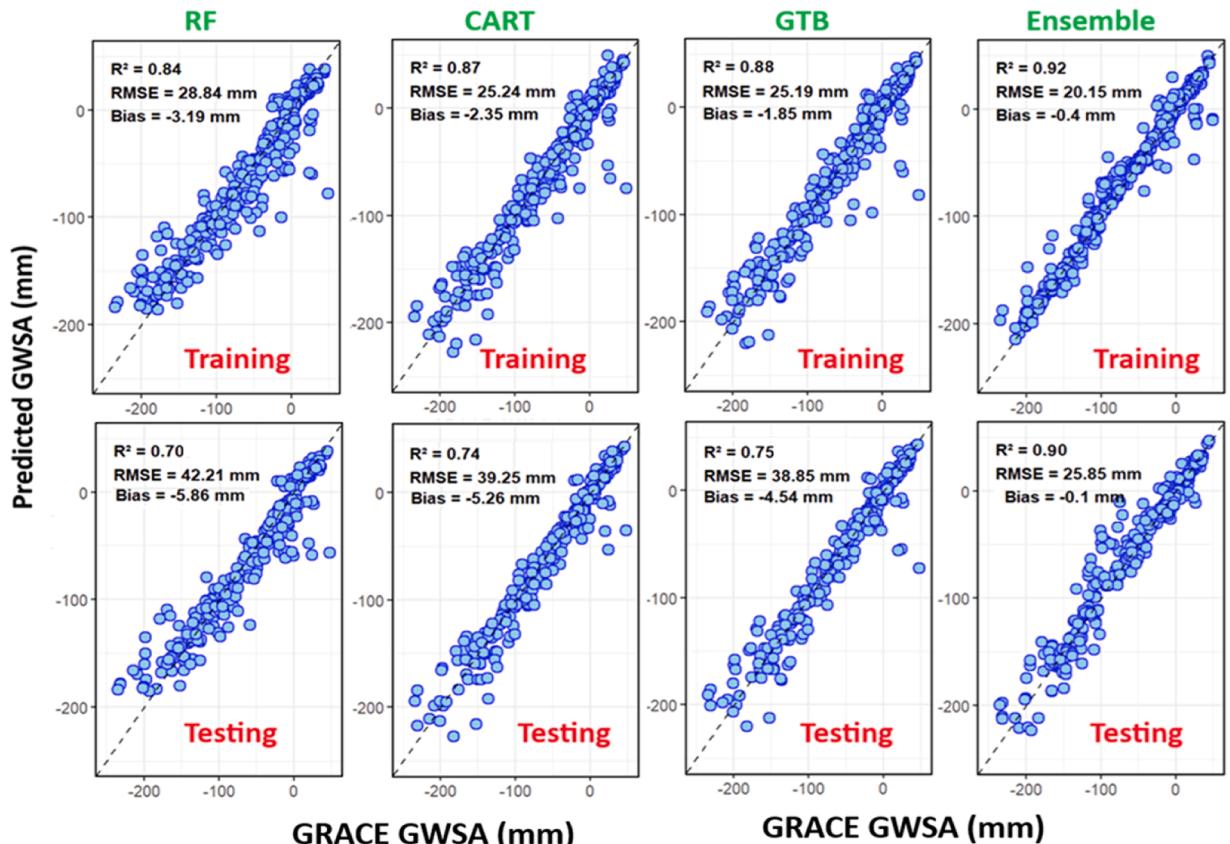
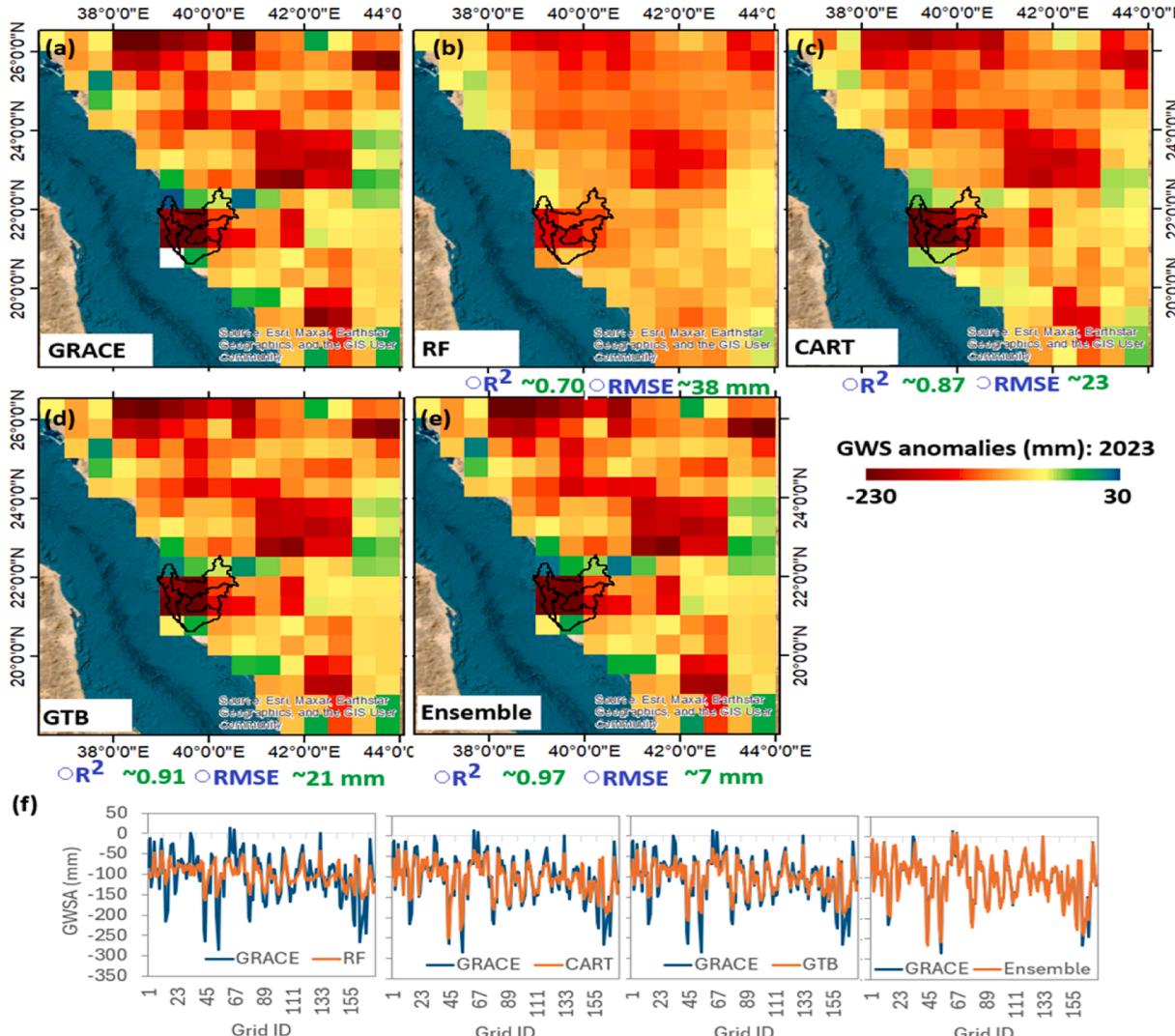


Fig. 4. Comparison between model-predicted GWS and GRACE GWS anomalies for training (70 %) and testing (30 %) data. Results are compared across three ML models and their ensemble prediction is obtained from a meta-model ensembling framework. Blue dots represent GWS anomalies over the GRACE grid for representative years (2002, 2005, 2008, 2011, 2014, 2017, 2020 and 2023). Bias is calculated as predicted minus observed GRACE GWS anomalies therefore negative values indicated underestimations relative to observed GRACE GWS anomalies data.

reliability of the reconstructed dataset and ensure continuity in the estimations of GWS anomalies from GRACE and GRACE-FO data. Previous work has also recorded the reliable efficacy of the STL method in reconstructing data to bridge the gap between GRACE and GRACE-FO (Arshad et al., 2024; Khorrami et al., 2023; Ali et al., 2024).

### 3.2. Hyperparameters tuning and predictive skills of ML models

**Fig. S1–S3** demonstrates the predictive capabilities of RF, GTB, and CART with various hyperparameter combinations. The predictive capabilities of the RF model improved, as seen by a reduction in the Root Mean Square Error (RMSE) from ~ 48 mm to 37.5 mm, with an increase in the number of trees from 1 to 450 during the training process. A significant reduction in RMSE was observed as the number of trees increased from 1 to 55, however additional increments do not adequately affect the model's predictive skills. Upon further comparison of tree performance with various leaf size combinations, we noted that the model exhibited superior performance with a leaf size of 5. VariablePerSplit yields highly uncertain outputs when choosing 1. A significant reduction in RMSE error was observed with an increase of VariablePerSplit to 7. Increasing the maxnode from 5 to 30 significantly impacts the prediction skills of the RF model. Overall, the RF performs superior with 60 trees, a leaf size of 7 VariablePerSplit, and a maximum of 30 nodes (**Fig. S1**). Augmenting the number of trees from 1 to 55 enhances the predictive skills of the GTB. While additional increments in the number of trees do not influence model accuracy. Changing the learning rate or shrinkage from 0.1 to 0.3 does not influence the model's accuracy, indicating its reduced sensitivity during training. The sampling rate somewhat influences model accuracy, with a sampling rate of 1 resulting in a slightly lower RMSE. In comparison to RF, the maxnodes exhibited reduced sensitivity in GTB. GTB

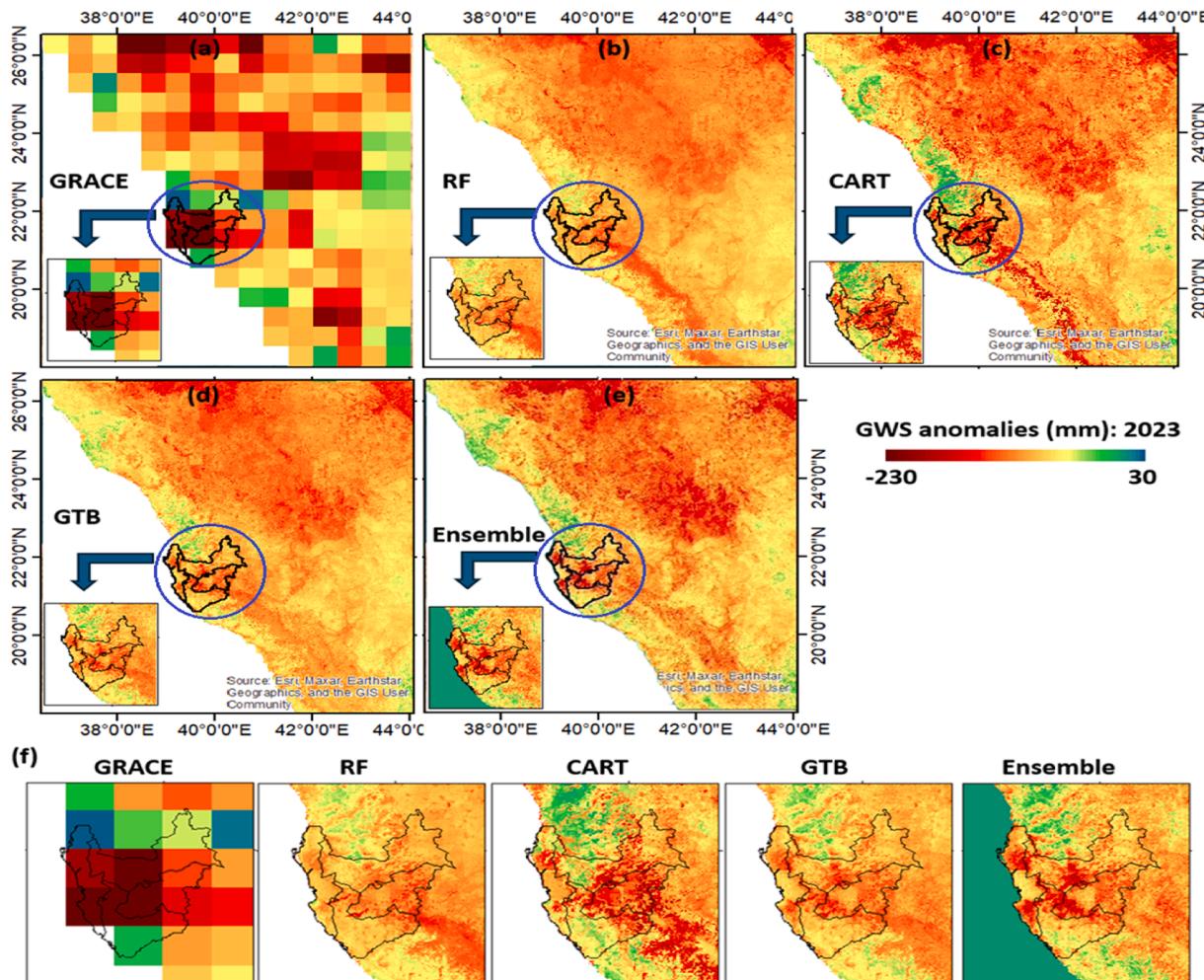


**Fig. 5.** Comparison of RF, CART, GTB, and ensembles predicted GWS anomalies with GRACE GWS anomalies at 55 km resolution for the year 2023: (a–e) spatial patterns and (f) profile of predicted GWS anomalies across 174 grids.

achieves superior accuracy with a maximum of 15 nodes, but each increase in nodes beyond this point slightly reduces model accuracy (Fig. S2). The CART algorithms were optimized by adjusting maxnodes and leaf size parameters. As the number of leaf nodes increases from 1 to 3, the model demonstrates improved accuracy. Further increases in leaf size diminish model performance and result in a greater RMSE. The performance of CART was optimal at a leaf size of 3, equivalent to a maximum of 99 nodes (Fig. S3).

### 3.3. Statistical evaluations of predicted and downscaled GWS anomalies

The optimized parameters were subsequently employed to rebuild the machine learning models for the prediction and downscaling of GWS anomalies. When comparing the predictive performance among the three ML models used (Fig. 4), the CART predicted GWS shows moderate accuracy based on  $R^2$  (0.87) and RMSE (25.24 mm) against the original GRACE GWS data. The GTB model achieves slightly better performance ( $R^2 = 0.88$  and RMSE = 25.19 mm). Although GTB demonstrated comparable performance in terms of  $R^2$  and RMSE in comparison to CART. In terms of bias, GTB exhibited superior performance with the lowest average bias of  $-1.85$  mm, in contrast to CART at  $-2.35$  mm and RF at  $-3.19$  mm. The RF model achieves slightly lower accuracy ( $R^2 = 0.84$  and RMSE = 28.84 mm). During the testing phase, all models demonstrated a reduction in predictive accuracy, exhibiting increased bias and RMSE in comparison to the observed GRACE GWSA. The overall  $R^2$  decreased to a range of 0.70–0.75, accompanied by an increase in RMSE to a range of 38–42 mm across all models. The decline in predictive capabilities indicates that these models fail to adequately capture the heterogeneity and spatial variability in GWS anomalies with hidden data, despite their superior performance on training samples. Combining predictions of RF, CART, and GTB into the meta-model, the new ensemble prediction of GWS anomalies showed a significant improvement compared to the GRACE-observed GWS anomalies. The ensemble mean of all three ML models maintains a high  $R^2$  (training = 0.92; testing = 0.90) while reducing RMSE (training = 20.15 mm; testing = 25.85 mm). Regardless of higher bias at the testing stage, the meta-model significantly reduced the model-specific bias to  $-0.1$  mm and improved the predictive skills of the

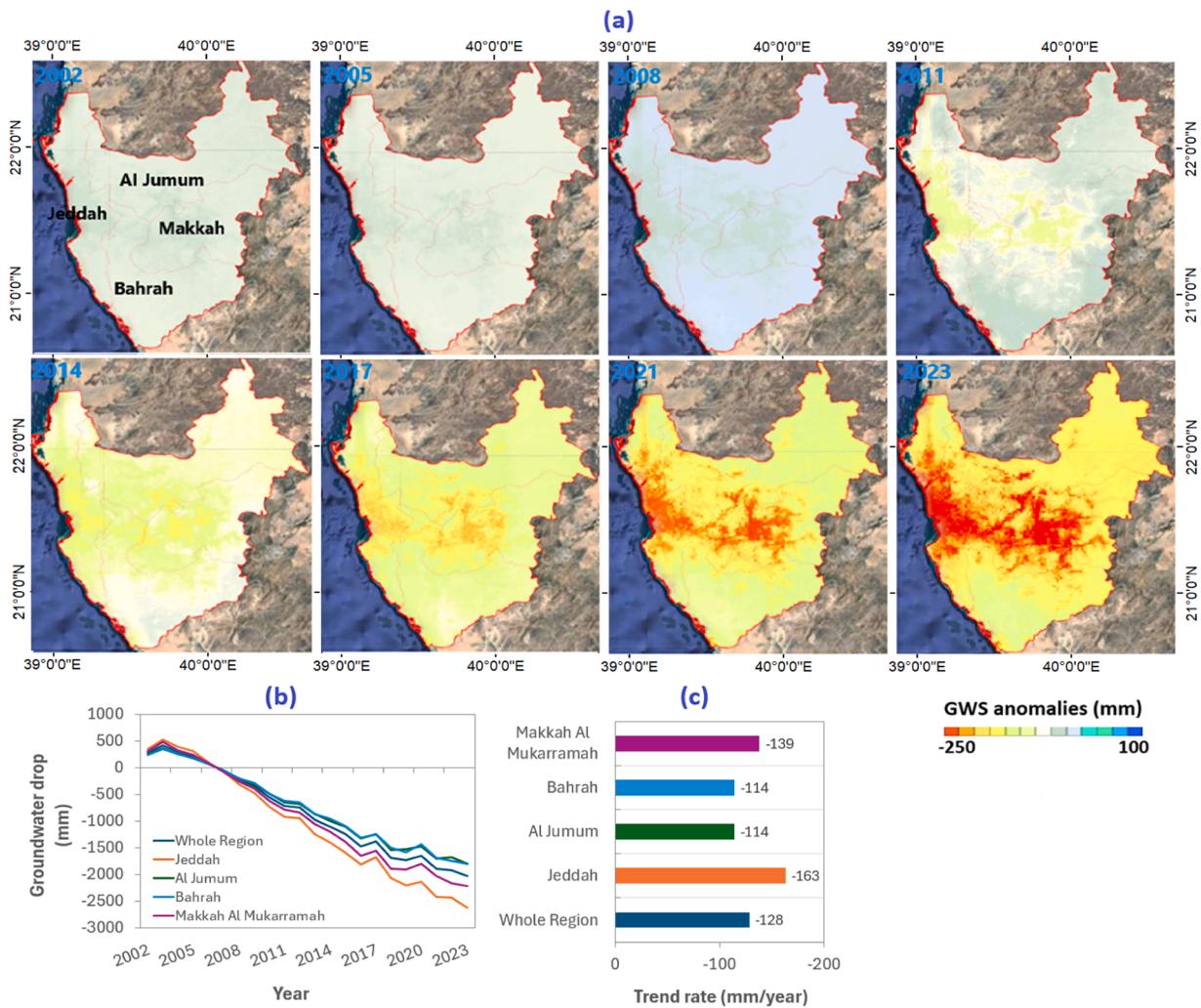


**Fig. 6.** Spatial patterns of GWS anomalies obtained from GRACE and ML-downscaled results for 2023 across (a–e) comparison across a larger extent of the Saudi Arabia and (f) snap view of the results for the target study sub-regions.

modeling framework. The tight clustering of points along the 1:1 ratio line indicates the reliability and accuracy of the multi-model ensemble approach in capturing GWS variability across the study region (Fig. 4).

Fig. 5a–e provided the numerical evaluation of the model's specific spatial patterns of GWS anomalies and their ensemble estimates compared to the GRACE-observed GWS anomalies for example year 2023. RF markedly underestimates the GWS anomalies in comparison to the GRACE-based GWS anomalies, especially over the majority of grids with severe GWS decline (Fig. 5a–b). The overall performance of the RF model was recorded as  $R^2$  0.70 and RMSE 38 mm (Fig. 5b). Further observing the profile of GWA anomalies across 174 grids, it is evident that RF does not adequately capture the signals of higher negative values, resulting in a smooth pattern of GWS anomalies (Fig. 5f). The GWS anomalies predicted by CART and GTB exhibit a consistent pattern with the GRACE-derived GWS anomalies, demonstrating comparatively higher predictive skills shown by a  $R^2$  ranging from 0.87 to 0.91 and an RMSE of 21–23 mm (Fig. 5c–d). These models still struggle to significantly capture the negative anomalies of GWS over several grids (Fig. 5f). The spatial map of ensemble estimates of GWS anomalies (Fig. 5e) demonstrated that the ensemble approach mitigates underestimation effects and enhances consistency with GRACE-based estimates. Additionally, pixels corresponding to higher GWS anomalies that were not well captured by model-specific predictions were significantly estimated by ensemble prediction with excellent predictive skills demonstrated by higher  $R^2$  of 0.97 and lower RMSE of ~ 7 mm.

Compared with the original GRACE data (55 km), the downscaled data (1 km) reveals a significant improvement in capturing detailed spatial patterns and heterogeneity of GWS changes (Fig. 6a–e). The original GRACE data indicates hotspots of GWS decline of approximately – 230 mm in coastal study regions (areas marked with blue circles). However, the coarse GRACE grids shared pixels across sub-regions, resulting in a uniform GWS decline pattern across urban and peri-urban areas and hence making it challenging to differentiate GWS variations in specific locales (Fig. 6a). While the downscaled results from all ML models capture the fine-scale



**Fig. 7.** (a) Spatial maps indicated changes in annual average GWS anomalies from 2002 to 2023, (b) temporal changes in GWS across different sub-regions, and (c) trend rates of annual GWS anomalies from 2002 to 2023 across the entire study region and its sub-regions. Annual GWS anomalies in plot “b” indicated the summation of monthly GWS anomalies to represent the total annual loss/gain.

variability that was previously obscured in the coarse resolution GRACE data (Fig. 6b–e). When combined with population patterns (Fig. 1d), the downscaled data reveals a higher GWS decline (−200 to −230 mm) in areas with higher population densities, particularly in Makkah and Jeddah, while relatively less decline is observed in surrounding areas. The GTB, CART, and ensemble estimates show spatial patterns consistent with GRACE trends, while the RF model tends to underestimate the extent of GWS decline, especially in regions where GRACE data indicated a more pronounced drop (Fig. 6f).

### 3.4. Spatiotemporal variations of GWS across study focus region

GWS anomalies were derived by subtracting soil moisture storage anomalies from GRACE-based terrestrial water storage anomalies, assuming that canopy surface water and glaciers/snow water storage contributions are negligible. Fig. S2 illustrates the temporal changes and trend rate of snow, soil moisture, canopy and terrestrial water storage between 2002 and 2023. Snow water storage showed a flat trend (0.00 mm/year), verifying this assumption. Canopy water storage remains consistently low (< 0.01 mm/month), with a near-zero trend rate (0.00017 mm/year). Soil moisture storage exhibits a clear declining trend rate (−36.25 mm/year) between 2002 and 2023. Terrestrial Water Storage declines more sharply (−136.77 mm/year), indicating groundwater depletion is the dominant driver. Our analysis approach and assumption is consistent with documented studies conducted in various regions of Saudi Arabia (e.g., Alshehri and Mohamed, 2023a, 2023b, 2023c; Mohamed et al., 2022).

Annual analysis of GWS anomalies from 2002 to 2023 reveals a pronounced progression of drop in groundwater storage across the study area (Fig. 7). Groundwater storage variations remained relatively stable between 2002 and 2008, with initial signs of depletion emerging in 2011. The period from 2014 to 2017 marks an acceleration in GWS losses, characterized by the emergence of significant depletion zones in the central region. In 2021, the widespread depletion hotspots become evident, particularly in the metropolitan areas of Makkah and Jeddah, with storage losses intensifying through 2023 and reaching up to −250 mm across large portions of the region (Fig. 7a). Additionally, a recent study by Esper et al. (2024) reported that the summer of 2023 was the warmest in the past 2000 years, surpassing previous records due to high global temperatures and an increase in extreme weather events. This unprecedented heat has intensified pressures on groundwater resources, with many regions experiencing heightened depletion due to increased evapotranspiration rates and prolonged dry conditions, further straining already stressed aquifers.

Quantitative analyses across sub-regions reveal varying depletion rates, with Jeddah experiencing the highest rate at −165 mm/year, followed by Makkah at −140 mm/year. The temporal evolution shows an initial positive storage-changing period until 2004, followed by a consistently declining trend across all regions, with the rate of decline accelerating particularly after 2014. While Bahrah and Al Jumum show comparatively lower decline rates at −110 mm/year, their depletion remains significant and contributes substantially to the region's overall groundwater losses (Fig. 7b–c).

The volumetric assessment of groundwater losses from 2002 to 2023 reveals varying impacts across sub-regions, reflecting both area size and GWS depletion rates (Table 1). Specifically, Al Jumum, the largest sub-region with an area of 6326.70 km<sup>2</sup>, experiences the highest total groundwater loss at 15.31 km<sup>3</sup>, despite a relatively moderate GWS decline rate of −0.11 m/year. Besides, Jeddah, with a smaller area of 3307.73 km<sup>2</sup>, records a similar depletion rate of −0.16 m/year, resulting in a volumetric loss of 11.64 km<sup>3</sup>. Bahrah and Makkah, covering areas of 4704.39 km<sup>2</sup> and 3862.53 km<sup>2</sup>, respectively, also show comparable total losses of 11.38 km<sup>3</sup> and 11.90 km<sup>3</sup>, each with a GWS decline rate of approximately −0.11 m/year in Bahrah and 0.14 m/year in Makkah. These results indicate the combined effect of spatial extent and depletion rates on total volumetric groundwater losses, indicating that even regions with moderate depletion rates can experience significant groundwater losses over large areas.

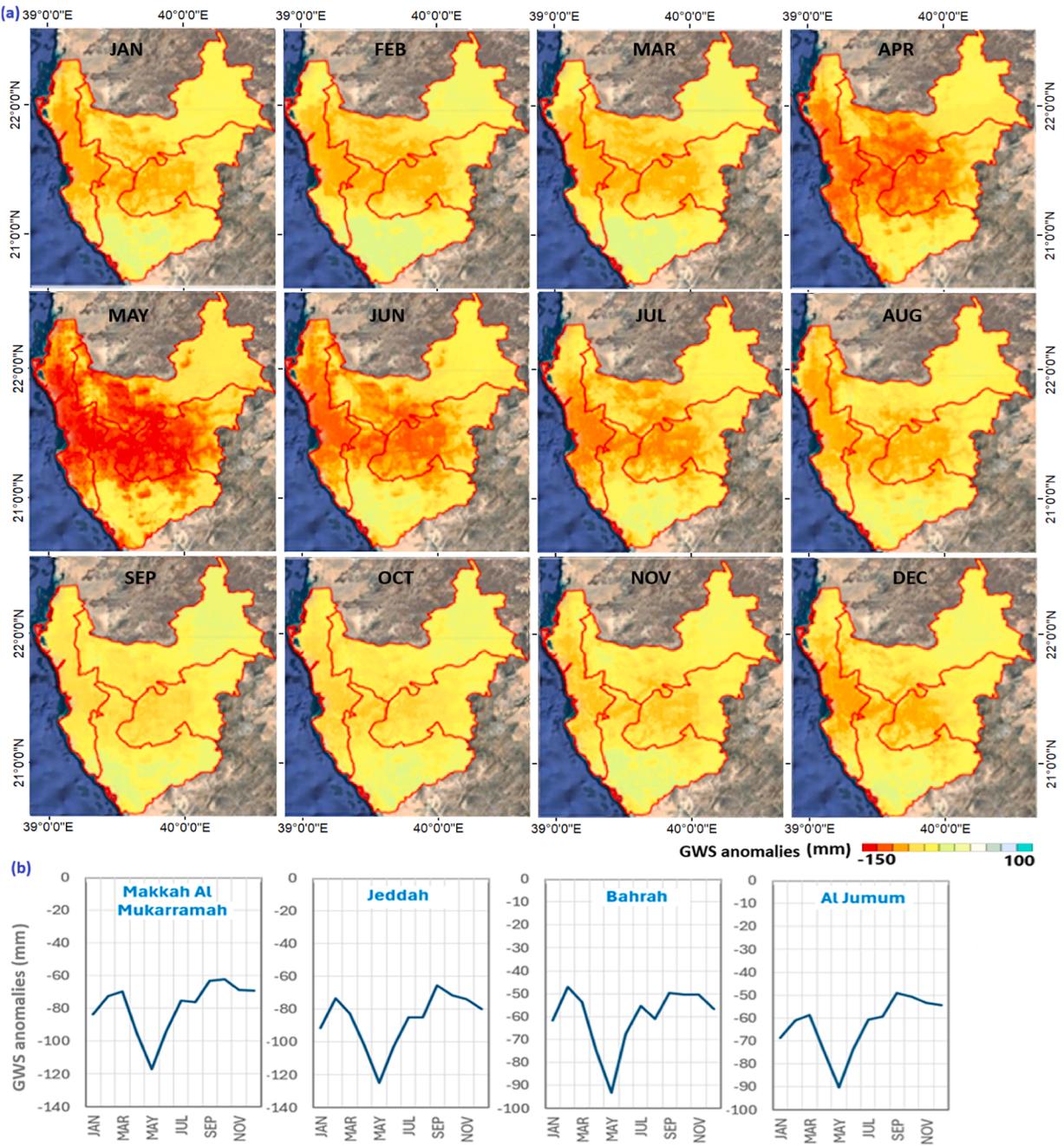
Monthly analyses show different seasonal patterns in GWS depletion, with the most severe losses occurring from April to July (Fig. 8a) in which May is found to have the largest depletion. Specifically, during these months, Jeddah and Makkah experience a maximum GWS decline rate of up to −150 mm, with the most intense drop visible in May. The monthly temporal changes (Fig. 8b) show that Makkah experiences the peak GWS decline at −118 mm in May, while Jeddah records a maximum decline rate of −125 mm during the same period. Bahrah and Al Jumum exhibit synchronized but comparatively moderate peak losses of approximately −90 mm in May, indicating a regional pattern of seasonal depletion. The seasonal cycle shows a recovery period from September through November, though the overall trend remains negative throughout the years. This pattern suggests a strong influence of seasonal factors, likely linked to increased groundwater extraction during warmer months and reduced withdrawal during cooler periods. Seasonal recovery may also be related to groundwater replenishing from precipitation recharge (Ahmed et al., 2018). The spatial distribution of these seasonal changes (Fig. 8a) shows that the central regions consistently experience more severe depletion compared to peripheral areas, regardless of season, indicating persistent stress on groundwater resources in these regions.

**Table 1**

Volumetric loss in groundwater storage in different sub-regions from 2002 to 2023.

| Sub-regions | Total area (km <sup>2</sup> ) | GWS drop (m/year) | Volumetric groundwater loss (km <sup>3</sup> ) |
|-------------|-------------------------------|-------------------|--|
| Jeddah      | 3307.73                       | −0.16             | −11.64   |
| Al Jumum    | 6326.70                       | −0.11             | −15.31   |
| Bahrah      | 4704.39                       | −0.11             | −11.38   |
| Makkah      | 3862.53                       | −0.14             | −11.90   |

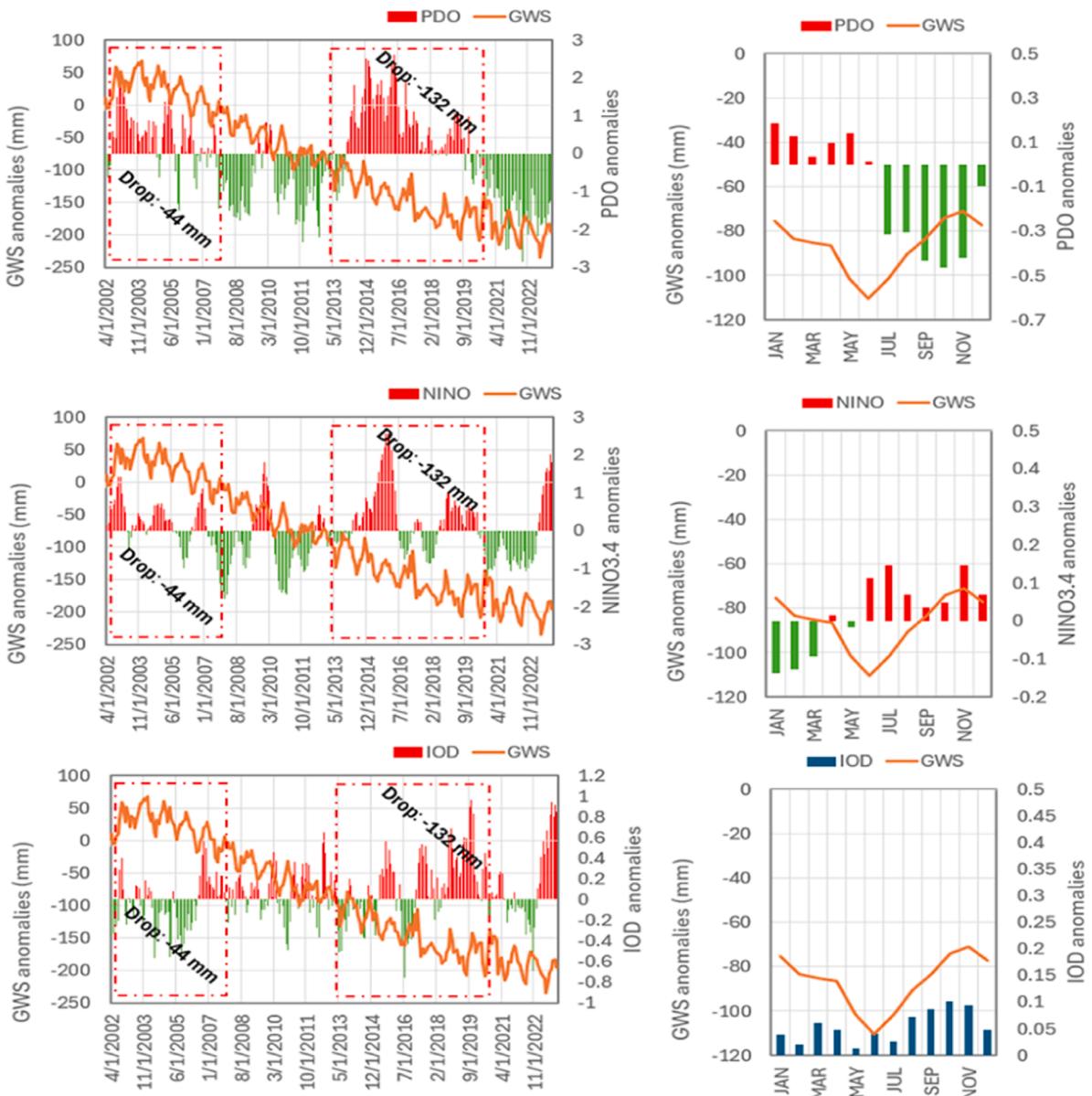
Volumetric loss = GWS drop (km/year) × 22 years × total area.



**Fig. 8.** (a) Spatial and temporal changes in monthly GWS anomalies from 2002 and 2023. (b) Monthly temporal changes in groundwater anomalies from 2002 to 2023 over sub-regions.

### 3.5. Climatic and anthropogenic impacts on GWS variability

The analysis of large-scale climate oscillations reveals complex interactions with GWS anomaly patterns (Fig. 9). Specifically, the PDO exhibits two different periods of strong positive anomalies: 2002–2007 and 2016–2022, associated with groundwater declines of  $-44$  mm and  $-132$  mm, respectively. The later period shows a significantly more severe impact on groundwater depletion, suggesting an intensification of climate-driven effects. The NINO3.4 indices pronounced groundwater storage declines during El Niño events, with the most significant effects observed from 2012 to 2016. Notably, the IOD shows synchronous behavior during this period, with positive phases corresponding to groundwater depletion exceeding  $-132$  mm, indicating potential compound effects of multiple climate oscillations on regional water resources. Guo et al., (2021) also reported a strong connection between global water storage

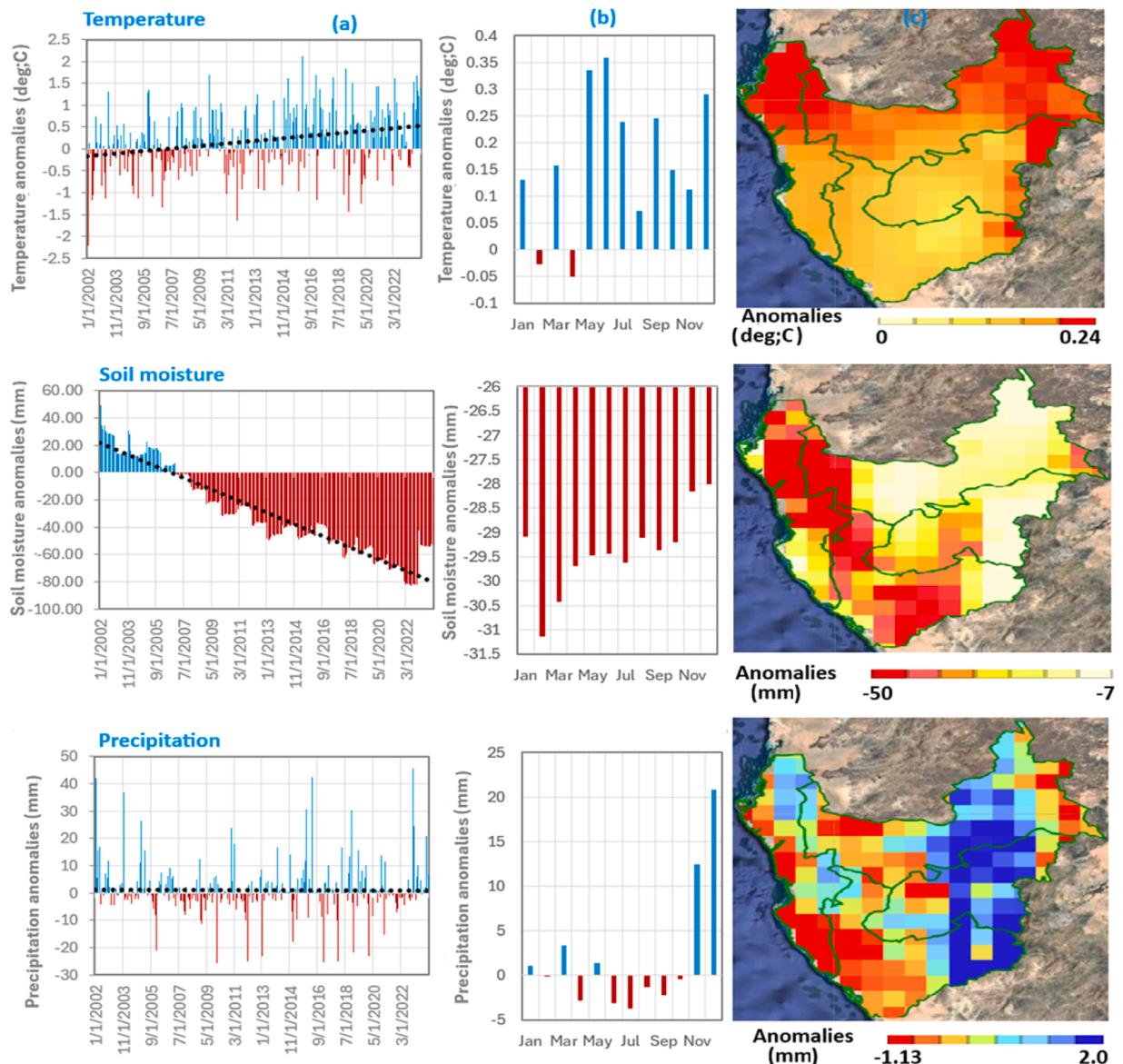


**Fig. 9.** Monthly changes in GWS anomalies and their interactions with large-scale circulation patterns (El NINO, PDO, and IOD).

variability the large-scale climate patterns.

On the other hand, local climate variables exhibited distinctive patterns of environmental change throughout the study period (Fig. 10). Specifically, temperature anomalies show a persistent warming trend, with monthly analysis revealing peak positive anomalies of up to  $0.35^{\circ}\text{C}$  during May and June. The spatial distribution of temperature anomalies indicates higher intensity in the northern regions, reaching up to  $0.24^{\circ}\text{C}$ . SM conditions dry significantly, shifting from positive anomalies (up to 20 mm) in the early 2000s to severe deficits (reaching  $-80 \text{ mm}$ ) by 2022. This decline becomes particularly pronounced after 2016, with the Jeddah and Bahrah regions experiencing the most severe deficits, up to  $-50 \text{ mm/month}$ . Besides, precipitation patterns exhibit high temporal variability, with monthly analysis indicating significant decreases from May to September. Additionally, spatial analyses of precipitation reveal predominantly negative anomalies ( $-1.13 \text{ mm}$ ) in the Jeddah and Bahrah areas, while some eastern regions show positive anomalies of up to  $2.0 \text{ mm}$ .

Furthermore, when examining human-induced impacts, we found associations between demographic and urban changes and groundwater depletion from 2002 to 2023 (See Fig. S5). Specifically, population density increases substantially in metropolitan areas like Makkah and Jeddah, with some localities experiencing dramatic growth, exceeding 185,000 inhabitants. Besides, a steady rise in population density is observed, from approximately 200 people/ $\text{km}^2$  in 2001 to over 300 people/ $\text{km}^2$  by 2023 in these regions. Meanwhile, developed area expansion is significant, increasing from approximately  $600 \text{ km}^2$  in 2002 to over  $650 \text{ km}^2$  by 2023, with a



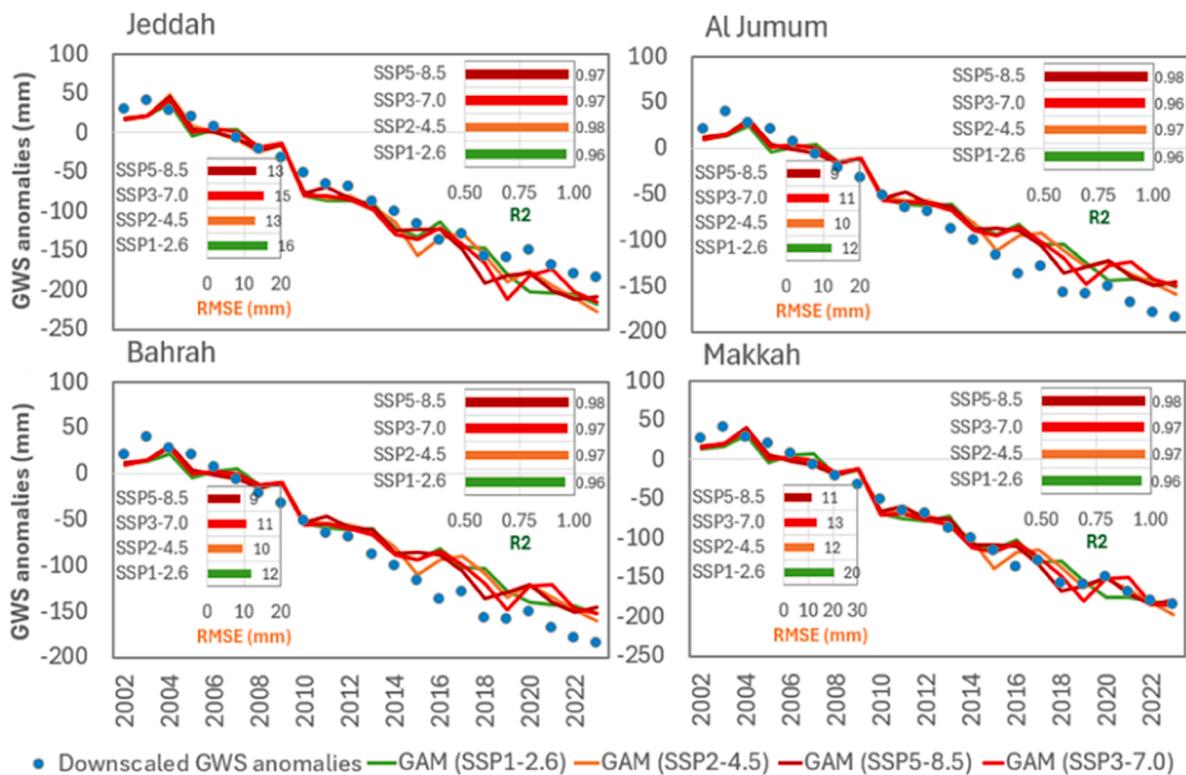
**Fig. 10.** Spatial and temporal changes in soil moisture (SM), precipitation, and temperature between 2002 and 2023: (a) monthly time series of soil moisture (SM), precipitation, and temperature anomalies, (b) seasonal changes across different months and (c) spatial maps of long-term average soil moisture (SM), precipitation, and temperature anomalies from 2002 to 2023.

notable acceleration after 2010 as annual expansion rates escalate. Notably, the spatial pattern of these anthropogenic factors aligns with areas experiencing severe groundwater depletion (Fig. 7) in which coastal regions reveal more urbanization (Fig. S5).

### 3.6. Projected GWS variations under future climate warmings

Before using GAM for predicting future GWS changes with CMIP6 predictors (temperature, precipitation, and heatwaves) and monthly downscaled GWS, we assessed the predictive accuracy of GAM using the testing data (Fig. S6). Trained GAM when applied to testing data (30 %) showed strong agreement of predicted GWS anomalies with observed GWSA for all emission scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) and four representative sites (Makkah, Jeddah, Al Jumum and Bahrah). The coefficient of determination ( $R^2$ ) for four representative sites ranged from 0.87 to 0.89 under SSP1-2.6, reached 0.98 under SSP2-4.5, ranged from 0.94 to 0.95 under SSP3-7.0, and ranged from 0.91 to 0.93 under SSP5-8.5, indicating robust model performance across varying future climate scenarios (Fig. S6).

Fig. 11 indicates the comparison between downscaled GWS (observed) and GAM-predicted GWS for whole data (training and testing) from 2002 and 2023. Overall, GAM accurately predicts GWS, exhibiting high  $R^2$  values between 0.96 and 0.98 across all

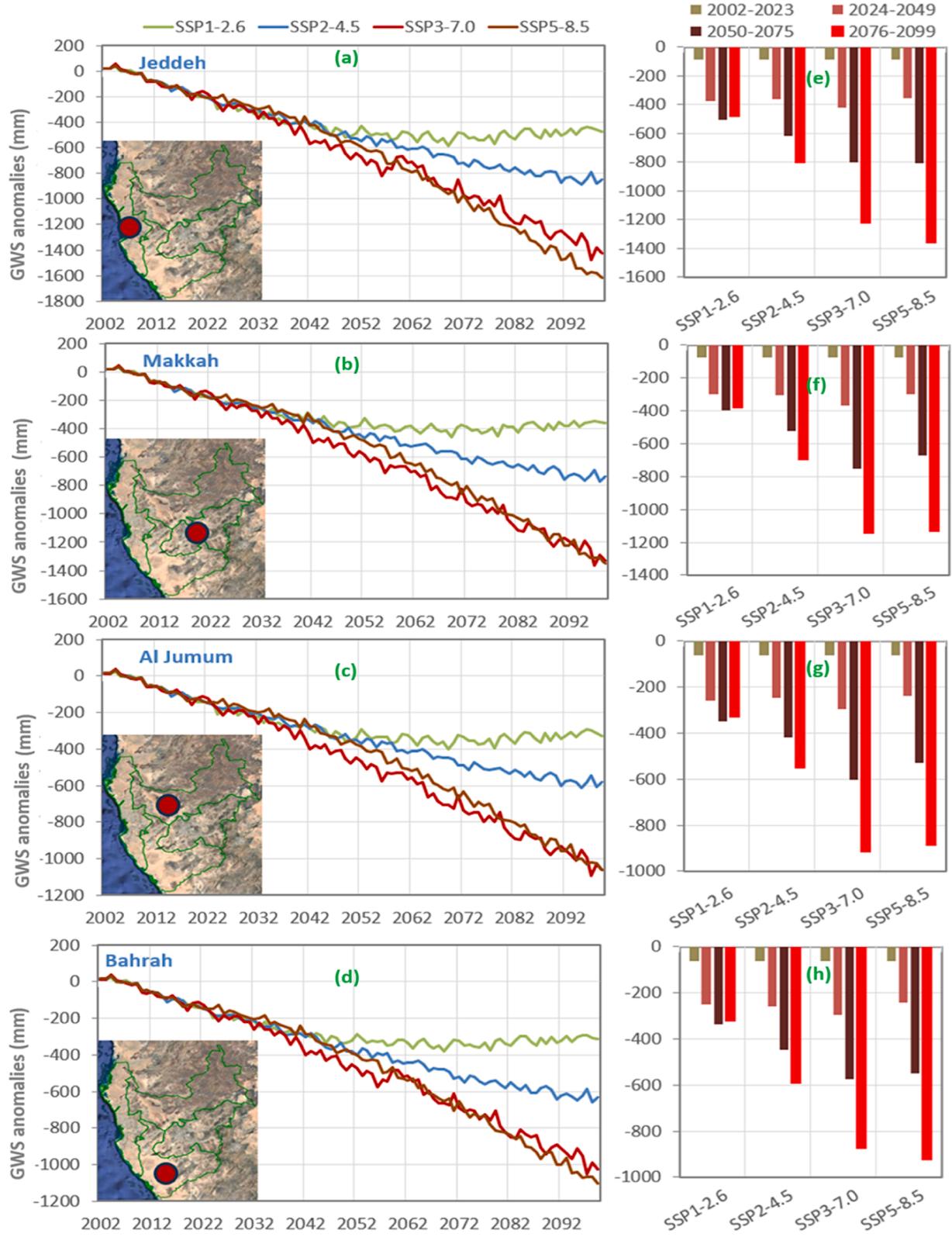


**Fig. 11.** Comparison between annual average downscaled GWS and GAM predicted GWS between 2002 and 2023 under three emission scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) for four representative sites (Makkah, Jeddah, Al Jumum and Bahrah). GAM predicted GWS anomalies results here are based on both training and testing predictions.

representative regions. Notably, GAM predicts the GWS more effectively under high-emission scenarios, exhibiting higher R<sup>2</sup> values [Makkha = 0.98, Jeddah = 0.97, Al Jumum = 0.98, and Bahrah = 0.98] and lower RMSE values [Makkha = 11 mm, Jeddah = 13 mm, Al Jumum = 9 mm, and Bahrah = 9 mm] in comparison to low emission scenarios. Temporal variations of GWS anomalies predicted by the GAM model were consistent with observed GWS anomalies from 2003 to 2023, revealing that GAM captures the anthropogenic-induced GWS declining trend reasonably and can adequately interpret the future variability in GWS with the chosen CMIP6 predictor variables (Fig. 11).

Analysis of future groundwater storage (GWS) projections under different SSP scenarios reveals increasingly severe depletion patterns through 2099, with marked variations across regions and scenarios (Fig. 12). The temporal evolution of GWS variations shows distinct trajectories across the four SSP scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5), with the most severe impacts projected under SSP5-8.5 (high emissions scenario) (Fig. 12a–d). Regional GWS projections for 2076–2099 indicate severe declines across the study area. In Makkah, GWS decline is projected to increase from – 77.87 mm (2002–2023) to – 1132.39 mm under SSP5-8.5 by 2076–2099, with a critical transition period between 2040 and 2060 where scenario outcomes significantly diverge, although SSP1-2.6 suggests stabilization around – 400 mm. Despite historically lower GWS decline (-63.17 mm from 2002 to 2023), Al Jumum is projected to have intensified GWS decline of – 888.15 mm under SSP5-8.5 by 2076–2099, following a more gradual GWS decline than other regions, with SSP1-2.6 suggesting possible stabilization around – 350 mm. In contrast, Bahrah shows the most severe intensification, with SSP5-8.5 projected GWS decline reaching – 1362.87 mm by 2076–2099 and exhibiting particularly sharp GWS declines after 2060. The low and moderate GHG scenarios (SSPs 2-4.5 and 3-7.0) show a progressive worsening, situated between the extremes of SSP5-8.5 and the very low GHG emission SSP1-2.6, which projects stabilization around – 400 mm (Fig. 12a–d). Furthermore, the temporal analysis reveals three different periods of GWS decline. The period from 2024 to 2049 marks the initial divergence of scenarios, characterized by a moderate decline across all SSPs. From 2050 to 2075, a notable acceleration in GWS drops is found, particularly pronounced under SSP3-7.0 and SSP5-8.5. The final period, 2076–2099, shows the greatest divergence among scenarios, with SSP5-8.5 demonstrating the most severe GWS decline across all regions (Fig. 12e–h).

The historic (2002–2023) and projected (2024–2099) trend rates of GWS anomalies in different sub-periods under different warming scenarios are provided in Table S2. At the regional level, the average GWS trend improves from – 125.06 mm/year (2002–2023) to + 30.11 mm/year under SSP1-2.6 by 2076–2099, confirming the scenario's sustainability. In Al Jumum and Bahrah, historically lower depletion rates transition into sharp future losses under high emissions, with Al Jumum reaching – 209.63 mm/year and Bahrah – 224.65 mm/year under SSP3-7.0 by 2076–2099. However, under SSP1-2.6, both locations show signs of stabilization or recovery with positive trends in the final sub-period (e.g., +25.61 mm/year in Al Jumum, +22.03 mm/year in Bahrah). In contrast,



(caption on next page)

**Fig. 12.** Projected changes in average annual GWS anomalies (mm) indicate irreversible groundwater drop until 2099 under different emission scenarios across four selected regions. (a-d) time series of annual average GWS anomalies variations from 2002 to 2099 and (e-h) average GWS anomalies for different sub-periods including historic (2002–2023), and future (2024–2029, 2050–2075 and 2076–2099) across four selected regions.

under SSP5-8.5, GWS projected to decline at a rate of  $-216.12 \text{ mm/year}$  by 2076–2099, indicating high vulnerability under continued emissions growth (Table S2).

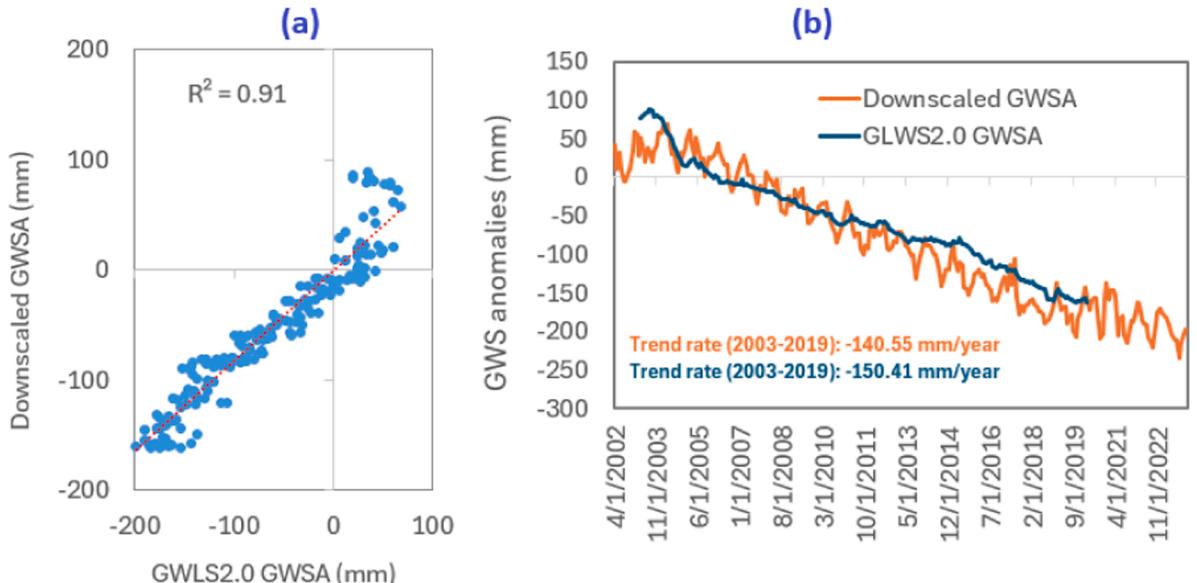
Our projections are set against a backdrop of significant changes in regional drought patterns and climate variations. We used the Standardized Precipitation Evapotranspiration Index (SPEI) across multiple time scales (1-, 6-, 12-, 18-, and 24 months), revealing a consistent trend toward more severe and prolonged drought conditions, particularly after 2013 (see Fig. S7 in Supplementary material). Specifically, the most severe drought occurs between 2014 and 2017, with SPEI values dropping below  $-3$ , indicating extreme drought conditions (see Table S1). Additionally, monthly SPEI patterns show heightened drought severity during summer months (July–September), with SPEI-1 exhibiting the highest variability and SPEI-24 indicating more sustained drought conditions. These drought patterns directly impact GWS, as reduced precipitation and elevated evapotranspiration, particularly during drought periods, lead to lower groundwater recharge and heightened GWS depletion.

On the other hand, these shifts in drought patterns align with projected climate variations under SSP scenarios through 2099, suggesting compounding effects on groundwater resources. Specifically, temperature projections reveal a consistent upward trend across all scenarios, with SSP5-8.5 predicting the most severe increase, from a historical average of  $27.1^\circ\text{C}$  (2002–2023) to  $31.68^\circ\text{C}$  by 2076–2099, while SSP1-2.6 suggests a more moderate rise to  $28.1^\circ\text{C}$  (Fig. S8). This warming trend is accompanied by a sharp increase in heat stress, measured by days exceeding  $35^\circ\text{C}$  annually. Under SSP5-8.5, these heat days are projected to rise from the historical average of 30 days per year to  $\sim 139$  days per year by 2076–2099, while SSP1-2.6 projects a more modest increase to  $\sim 47$  days per year. Precipitation patterns show greater variability and uncertainty across scenarios, with all SSPs suggesting slight increases in annual precipitation. Notably, the 2050–2075 period sees peak precipitation levels, reaching  $113.7 \text{ mm}$  under SSP5-8.5, followed by a slight decline in 2076–2099. However, this projected increase in precipitation appears insufficient to counterbalance the elevated evapotranspiration rates and growing water demand, as indicated by the continued decline in GWS across all scenarios (Fig. 12).

#### 4. Discussion

##### 4.1. Observational data is a major barrier to validation and effective groundwater management

Previous studies have also utilized GRACE data to study terrestrial and groundwater changes in the Saudi Arabia (Othman et al., 2019; Alshehri and Mohamed, 2023a, 2023b, 2023c; Usman et al., 2024). Due to the unavailability of groundwater monitoring in the region, these studies also lack validation which is also a key limitation in our work. Alternatively, we compared our findings with reference data and published studies. We noted our results are well aligned with the values of GWS changes that had been previously published in Saudi Arabia (Othman et al., 2019; Mohamed et al., 2022; Alshehri and Mohamed, 2023a, 2023b, 2023c; Usman et al.,



**Fig. 13.** Comparison of downscaled groundwater storage anomalies (GWSA) against GLWS2.0. (a) Scatter plot comparing downscaled GWSA with GLWS2.0 GWSA across the study region for the period 2003–2019 and (b) time series of monthly GWS anomalies (mm) from the downscaled dataset and GLWS2.0, showing temporal patterns and trend rate.

2024). These studies also reported the groundwater drawn down in the last two decades, particularly across Makkah and other highly populated regions. Mohamed et al. (2022) and Alshehri and Mohamed (2023a, 2023b, 2023c) have also reported a declining rate of GWS anomalies ranging from – 6 to – 13 mm/year based on monthly annual average data (or – 72 mm/year – 156 mm/year based on annual total GWS anomalies) in Saudi Arabia which is quite close to our reported results. Most of the previous work mainly reported regional groundwater changes in the Saudi Arabia. The key component of this work was to remove the barrier to data availability and effective groundwater management by providing a framework and tool that can help to understand the local scale groundwater changes in data-scarcity regions using remote sensing datasets. Many researchers across the World have also acknowledged the reasonable performance of GRACE reference to the observational data (Rodell et al., 2007; Bhanga et al., 2016; Rateb et al., 2020; Arshad et al., 2024), which supports our assumption that it can be used to monitor groundwater changes in data-scarcity areas like Saudi Arabia with a reasonable level of trust. Our results of downscaled GWS variations showed a strong agreement with GLWS2.0 GWS anomalies with an  $R^2$  of 0.91 (Fig. 13a). The temporal trends are consistent, with the downscaled GWSA showing a trend rate of – 140.55 mm/year and the GLWS2.0 product indicating a similar trend of – 150.41 mm/year during the 2003–2019 period (Fig. 13b). Our results of trend rates of GWS anomalies (Fig. S9) and volumetric loss in groundwater storage (Table 2) were also closely matched with the simulated output of WaterGAP v2.2d. WaterGAP v2.2d results indicated a drop in GWS across all selected sub-regions with an overall regional declining rate of – 144 mm/year (Fig. S9) which is comparable to GRACE-based GWS results (-125 mm/year) (Fig. 7). Similarly, the volumetric loss of GRACE-based GWS was found to – 12.84 km<sup>3</sup> (Table 1) which is also satisfactory when compared with the WaterGAP v2.2d results (-10.83 km<sup>3</sup>) between 2002 and 2019. The slightly underestimated volumetric loss of GWS from WaterGAP v2.2d could also be a reason that the simulated results were only considered up to 2019. Comparisons with these publically accessible datasets enhance the credibility of the downscaled GWSA estimates, especially in data-sparse regions. In addition to these efforts, we still acknowledge the limitations of comparing data with observational information, and we hope that future work will improve the validation framework and collect more data to advance the methodology.

#### 4.2. Responses of covariates variables in ML and correlations with GWS anomalies

We found significant relationships for several predictors with GWS changes over this region (Table 3). Among them, elevation exhibits a consistently strong correlation with GWS changes, with Pearson coefficients ( $r$ ) as high as 0.99 (2008) and 0.87 (2011). This strong statistical association is consistent with established hydrogeological principles, where elevation influences hydraulic gradients and groundwater flow paths, thereby impacting storage and its variability (Toth, 1963). Meanwhile, LST demonstrates a notable correlation, particularly in more recent years, with  $r$  values reaching 0.76 (2020) and 0.68 (2023). This statistically significant relationship can be attributed to LST's critical role in modulating evapotranspiration rates and surface energy balance, processes that directly affect groundwater recharge and depletion dynamics (Anderson et al., 2012). Thus, these observed variations in the strength of these correlations across different years underscore the dynamic interplay of factors influencing GWS within our study.

We noted that the response variables in each model differed during training (Fig. S10). In the case of RF, elevation indicated the highest feature importance at 15 %, followed by land surface temperature at 14 %, human modification at 13 %, precipitation at 13 %, and soil moisture at 13 %. In the context of CART, feature relevance was ranked as follows: elevation (21 %), land surface temperature (19 %), NDVI (16 %), and precipitation (14 %). Despite the moderate correlation of soil moisture with GWS anomalies (referenced in Table 3), it demonstrates a higher feature importance of 55 % in GTB, underscoring its critical role in sustaining the predictive accuracy of GTB-based GWS anomalies. Therefore, it is important to emphasize that correlation itself is not a direct indicator when considering covariates in the predictive framework. Overall, elevation, soil moisture, and land surface temperature variables were identified as the most important features in the predictive framework. The different relative contributions of covariates in each model could be associated with model-specific learning algorithms, spatial heterogeneity, and hydrological relevance response.

#### 4.3. Implications of methodological advances in multi-model ensemble downscaling

In general, we found several significant implications for groundwater monitoring methodology and practice when integrating various machine learning algorithms with GEE for GRACE data downscaling. Specifically, compared to single-model applications, our ensemble approach—incorporating the RF, CART, and GTB algorithms—performed better and demonstrated the usefulness of ensemble methods for reducing model-specific biases and enhancing prediction accuracy. This approach supports and validates recent findings by Wang et al. (2023) on the advantages of ensemble methods. More importantly, GEE in this study has proved effective in addressing long-standing computational scaling issues noted by Amani et al. (2020), thereby increasing the affordability and accessibility of high-resolution groundwater monitoring for global water resource management.

The implications of high-resolution GWS data are particularly evident for urban water management. Our approach has revealed

**Table 2**

Volumetric loss in groundwater storage in different sub-regions from 2002 to 2019 based on WaterGAP v2.2d.

| Sub-regions | Total area (km <sup>2</sup> ) | GWS drop (m/year) | Volumetric groundwater loss (km <sup>3</sup> ) |
|-------------|-------------------------------|-------------------|--|
| Jeddah      | 3307.73                       | – 0.23            | – 13.69  |
| Al Jumum    | 6326.70                       | – 0.12            | – 13.67  |
| Bahrah      | 4704.39                       | – 0.09            | – 7.62   |
| Makkah      | 3862.53                       | – 0.12            | – 8.34   |

**Table 3**

Absolute values of Pearson correlation of selected predictors with GWS changes.

|                          | 2002 | 2005 | 2008 | 2011 | 2014 | 2017 | 2020 | 2023 |
|--------------------------|------|------|------|------|------|------|------|------|
| Elevation                | 0.82 | 0.88 | 0.99 | 0.87 | 0.63 | 0.54 | 0.40 | 0.53 |
| Evapotranspiration       | 0.38 | 0.45 | 0.47 | 0.49 | 0.42 | 0.41 | 0.37 | 0.30 |
| Human modification index | 0.14 | 0.23 | 0.25 | 0.29 | 0.31 | 0.37 | 0.23 | 0.12 |
| LST                      | 0.44 | 0.48 | 0.44 | 0.51 | 0.54 | 0.51 | 0.76 | 0.68 |
| NDVI                     | 0.19 | 0.27 | 0.26 | 0.22 | 0.19 | 0.24 | 0.20 | 0.14 |
| Population density       | 0.32 | 0.32 | 0.37 | 0.33 | 0.33 | 0.40 | 0.34 | 0.29 |
| Precipitation            | 0.19 | 0.39 | 0.47 | 0.48 | 0.53 | 0.24 | 0.42 | 0.42 |
| Soil moisture            | 0.34 | 0.36 | 0.34 | 0.37 | 0.32 | 0.31 | 0.41 | 0.44 |

previously undetected groundwater depletion trends in the highly urbanized regions of Makkah and Jeddah, thereby extending the findings of [Chen et al. \(2019\)](#) and [Ali et al. \(2022\)](#) on the need for high-resolution monitoring. Additionally, the spatial heterogeneity of groundwater changes was effectively captured by the multi-model ensemble technique, which proved very useful for monitoring and evaluating urban water dynamics, where extraction patterns may vary significantly across the region. Our framework showed promising results that demonstrated its ability to integrate multiple environmental and anthropogenic predictors from GEE into the downscaling framework, offering more insights into groundwater dynamics. Building on and expanding the work of [Liu et al. \(2020\)](#) and [Luo et al. \(2023\)](#), the successful implementation of this multi-variable framework has important implications for the advancement of ML applications in environmental monitoring. Notably, the framework's capacity to simultaneously analyze land use, population density, and climate variables has enabled unprecedented insights into the interplay between natural and anthropogenic influences on groundwater storage changes, with important implications for both research methodology and policy development.

Our approach proved an ability to obtain the high accuracy of GRACE data while significantly enhancing spatial resolution that has noticeable implications for groundwater monitoring, particularly in regions with complex urban-rural divisions. This development offers new opportunities for integrated water resource management and addresses a significant gap in present monitoring capabilities. While the framework's success in capturing fine-scale variations reinforces findings by [Roy et al. \(2024\)](#) on multi-model effectiveness, it also extends their work by demonstrating practical applications in complex urban environments. However, these methodological advances come with important caveats regarding extreme value predictions, especially in areas experiencing rapid depletion. This limitation has key implications for monitoring high stress zones and underscores the need for complementary validation approaches.

The implications of these methodological advances are the technical and practical aspects of GRACE multi-variable downscaling applications for water resources management ([Xu et al., 2023](#)). While demonstrating this approach is effective in computational efficiency, our work shows that it is promising and reproducible for other studies. Through analysis of the western regions of Saudi Arabia, we highlighted its effectiveness to use for similar regions with high risks in terms of groundwater over-extraction and depletion. This thus builds a strong foundation for capturing complex patterns of groundwater spatial variations, offering a more useful approach for informed decision-making in groundwater monitoring.

#### 4.4. Drivers of groundwater depletion

We explored the connections of GWS changes with local climate anomalies, atmospheric circulation patterns, and anthropogenic factors. The significant association between positive anomalies (warm phases) of atmospheric circulation patterns (i.e., ENSO, IOD, PDO) and the expedited groundwater decline corresponds with worldwide findings by [Gleeson et al. \(2019\)](#) concerning the influence of extensive climatic patterns on regional water resources. Specifically, the substantial decrease in GWS during the warm phase (2012–2016) was found to be significantly impacted by climate variability. This even becomes more intensified under the combined impacts of climate oscillations and human factors, which was indicated by the high GWS reduction during peak religious tourism and climate stress (see [Section 3.3](#)).

Further, GWS decline was found to be associated with local climate anomalies (air temperature, precipitation, and soil moisture). Seasonal and short-term changes in GWS variability can be related to natural variability in local climate anomalies ([Ahmed et al., 2019](#)). The declining soil moisture compounded with rising temperatures would be another factor exacerbating the groundwater decline rate. These findings confirm recent research by [Bhattarai et al. \(2023\)](#) and [Jasechko et al. \(2024\)](#) on the compounding effects of climate change on groundwater resources in arid regions. Dry climate and anthropogenic warming are expected to increase the flash drought which would have certain implications for groundwater resources ([Mishra et al., 2021; Neelam et al., 2024](#)). April and July exhibited a significant drop in groundwater storage, particularly in Makkah, where heightened water demands due to national and internal tourism exacerbated an already strained aquifer. In addition, rapid urbanization is also heightening the vulnerabilities to groundwater resources, as reported by [Arshad et al. \(2024\)](#) and [Mohamed et al. \(2024\)](#) regarding the impact of urban expansion on groundwater systems in Saudi Arabia.

#### 4.5. Regional water security context and religious tourism

The western coastal region of Saudi Arabia exemplifies a distinctive case study in which fast urbanization converges with religious tourism, revealing intricate patterns of groundwater stress. Our high-resolution research indicated that Makkah undergoes groundwater depletion rates of ~130 mm/year, with maxima coinciding with significant religious events. Muslims around the world travel to

Makkah to perform Hajj and Umrah. Results correspond with the observations of [Abonomi et al. \(2022\)](#) concerning the environmental impacts of tourism and corroborate the concerns expressed by [Tabash et al. \(2023\)](#) on the water security consequences of tourism expansion under Saudi Vision 2030. The interplay of population expansion and intermittent increases in visitor traffic poses distinct issues for water resource management. In Makkah, depletion rates escalate during peak pilgrimage months, especially during the summer Hajj seasons, when natural water stress is pronounced. This pattern agrees with [Ascoura's \(2013\)](#) results regarding the correlation between religious tourism and urban resource strain.

The higher groundwater depletion rate in Jeddah (-165 mm/year) may be related to its primary role as a main entry point for pilgrims, thus increasing the burden on the city's water supplies. Continuous groundwater declines in the last two decades could be attributed to insufficient management practices. It comes from the lack of monitoring data that hinders the understanding of how urban growth, tourism, and climate change may have impacts on groundwater availability and use. This highlights the broader regional water security issues in the Middle East, where major water management difficulties are increased by population growth and climate change.

The projected climate change and its implications on groundwater resources highlight future water security and management. Specifically, under a high-end scenario (i.e., SSP5-8.5) GWS would significantly drop between 2024 and 2099, raising concerns about future regional water security throughout the twenty Century. Our results on projected GWS are also aligned with global perspectives and findings reported by [Famiglietti \(2014\)](#) and [Bierkens and Wada \(2019\)](#). The heatwave days are projected to increase in the future, particularly for developed regions in Makkah and Jeddah. It raises important concerns for developing mitigation and adaptation plans to minimize the risk of environmental changes and maintain regional long-term water security. This is particularly evident in the low-end (SSP-1.2.6) scenario which indicates that if the emission level keeps lower, it can considerably reduce the groundwater risk.

High-resolution data from a multi-model approach assists in implementing efficient water management plans by pinpointing the regions of undergone higher groundwater depletion. The results imply that a sustainable strategy is needed for groundwater management in these religious tourism destinations, focusing on maintaining a safe level of urban water usage and increasing demand due to tourism. This involves creating seasonal water management strategies for Makkah and other religious tourism destinations that account for pilgrim influxes while preserving sustainable baseline extraction rates. With our improved spatial resolution data, the deployment of smart water monitoring tools can assist in identifying areas of significant depletion and determining the best time for interventions. Alternative water sources might also reduce the stress on groundwater supplies, such as treated wastewater for non-potable usage during seasons of high tourism visits.

Finally, considering the robust relationships we found between climate indices and groundwater depletion rates, including climate dynamics in water management planning is an important strategy. This aligns with recent suggestions for climate-adaptive water management found by [Gleeson et al. \(2019\)](#) and [Bhattarai et al. \(2023\)](#). Additionally, our work showcases a promising approach to integrate this high-resolution GRACE data for regional groundwater monitoring that would significantly help to establish drought early warning systems. This initiative would then play a critical role in tackling climate-related issues, particularly for regions with high-frequent depletion periods, helping to preserve the sustainability of regional groundwater resources.

## 5. Concluding remarks

This study introduced a multi-model ensemble approach for downscaling GRACE data to a high resolution ( $1 \text{ km}^2$ ) with the main goal of assessing the local trends of GWS changes across the western coastal region of Saudi Arabia. We incorporated projected future climate variables from the CMIP6 project, alongside the  $1\text{-km}^2$  downscaled GRACE data, to project GWS changes from 2024 to 2099. The multi-model ensemble ML downscaling showcased its remarkable ability to enhance the spatial resolution of GRACE data from  $55 \text{ km}^2$  to  $1 \text{ km}^2$  resolution. The multi-model ensemble predicted GWS showed remarkable consistency with the spatial pattern of GRACE GWS, achieving high accuracy ( $R^2 = 0.92$  and RMSE = 20 mm) compared to the individual model ( $R^2 = 0.84\text{--}0.88$  and RMSE = 25–28 mm). The downscaled GWS data presents an opportunity to reveal previously hidden vulnerabilities to groundwater across the region, which would be challenging to identify with the coarser resolution GRACE GWS data, particularly in urban regions where depletion rates reach at least – 130 mm/year. Besides, new insights into the temporal dynamics of groundwater stress in arid regions were obtained by identifying seasonal patterns of groundwater depletion and their intricate interactions with regional environmental changes, large-scale climate oscillations, and anthropogenic factors (i.e., urbanization and religious tourism). Looking into the future projected groundwater storage indicates an irreversible decline throughout the 21st Century. Under the high-end (SSP5-8.5) GHG scenario, the rate of decline would be higher, with projected losses in metropolitan areas exceeding – 1130 mm. However, if appropriate management and mitigation strategies are implemented, the reduction in GWS could be lessened under the low-emission scenarios (SSP1-2.6).

Communities all over the world will have easier access to high-resolution groundwater monitoring with the suggested framework within GEE, which creates a computationally effective and repeatable method for processing massive satellite data. Our work provides the necessary scientific foundation that could be used to reduce future water-related challenges, particularly driven by the changing climate and overpopulation.

## CRediT authorship contribution statement

**Arfan Arshad** conceived the ideas, performed the experiments, analyzed the results, and prepared the first draft of the manuscript under the supervision of **Ali Mirchi** and **Amir AghaKouchak**. **Thanh Nhan Duc Tran**, **Muhammad Shafeeqe**, and **Ali Mirchi** assisted to improve the first draft of the manuscript. **Cenlin He**, **Md Masudur Rahman**, **Zaichen Xiang** and **Jessica Besnier** provided

oversight, and helped with technical review and editing the manuscript.

### 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.

### Acknowledgements

The authors acknowledge funding from the NSF AccelNet Program (Award no. OISE 2114701). We appreciate two anonymous reviewers whose comments helped improve this work. Any opinions, findings, conclusions, or recommendations expressed in this publication are solely those of the authors.

### Competing interests

The authors declare no competing interests.

### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ejrh.2025.102552](https://doi.org/10.1016/j.ejrh.2025.102552).

### Data availability

Data will be made available on request.

### References

- Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958 to 2015. *Sci. Data* 5, 170191. <https://doi.org/10.1038/sdata.2017.191>.
- Abdella, F.I., El-Sofany, W.I., Mansour, D., 2024. Water scarcity in the Kingdom of Saudi Arabia. *Environ. Sci. Pollut. Res.* 1–12.
- Abonomi, A., De Lacy, T., Pyke, J., 2022. Environmental impact of the hajj. *Int. J. Relig. Tour. Pilgr.* 10 (1), 12.
- Agarwal, V., Akyilmaz, O., Shum, C.K., Feng, W., Yang, T.Y., Forootan, E., Uz, M., 2023. Machine learning based downscaling of GRACE-estimated groundwater in Central Valley, California. *Sci. Total Environ.* 865, 161138.
- Ahmed, M., 2020. Sustainable management scenarios for northern Africa's fossil aquifer systems. *J. Hydrol.* 589, 125196.
- Ahmed, M., Abdelmohsen, K., 2018. Quantifying modern recharge and depletion rates of the Nubian Aquifer in Egypt. *Surv. Geophys.* 39, 729–751.
- Ahmed, M., Aqnoy, M., El Messari, J.S., 2021. Sustainability of Morocco's groundwater resources in response to natural and anthropogenic forces. *J. Hydrol.* 603, 126866.
- Ahmed, M., Wiese, D.N., 2019. Short-term trends in Africa's freshwater resources: rates and drivers. *Sci. Total Environ.* 695, 133843.
- Akbar, M.U., Mirchi, A., Arshad, A., Alian, S., Mehata, M., Taghvaeian, S., Wagner, K., 2025. Multi-model ensemble mapping of irrigated areas using remote sensing, machine learning, and ground truth data. *Agric. Water Manag.* 312, 109416.
- Ali, S., Liu, D., Fu, Q., Cheema, M.J.M., Pal, S.C., Arshad, A., Zhang, L., 2022. Constructing high-resolution groundwater drought at spatio-temporal scale using GRACE satellite data based on machine learning in the Indus Basin. *J. Hydrol.* 612, 128295.
- Ali, S., Ran, J., Luan, Y., Khorrami, B., Xiao, Y., Tangdamrongsub, N., 2024. The GWR model-based regional downscaling of GRACE/GRACE-FO derived groundwater storage to investigate local-scale variations in the North China Plain. *Sci. Total Environ.* 908, 168239.
- Almazroui, M., Islam, M.N., Jones, P.D., 2013. Urbanization effects on the air temperature rise in Saudi Arabia. *Clim. Change* 120, 109–122. <https://doi.org/10.1007/s10584-013-0796-2>.
- Alshehri, F., Mohamed, A., 2023a. Analysis of groundwater storage fluctuations using GRACE and remote sensing data in Wadi As-Sirhan, Northern Saudi Arabia. *Water* 15 (2), 282. <https://doi.org/10.3390/w15020282>.
- Alshehri, F., Mohamed, A., 2023b. Analysis of groundwater storage fluctuations using GRACE and remote sensing data in Wadi As-Sirhan, Northern Saudi Arabia. *Water* 15 (2), 282.
- Alshehri, F., Mohamed, A., 2023c. Investigation of groundwater potential using gravity data in Wadi Fatimah and its surroundings, Western Saudi Arabia. *Front. Earth Sci.* 11, 1225992.
- Altayyar, M.O., Ali, S., Larson, A.E., Boving, T., Thiem, L., Akanda, A.S., 2024. Quantifying groundwater depletion in Arabian Peninsula transboundary aquifer systems: understanding natural and anthropogenic drivers. *Groundw. Sustain. Dev.* 26, 101293.
- Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., Brisco, B., 2020. Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13, 5326–5350.
- Anderson, M.C., Allen, R.G., Morse, A., Kustas, W.P., 2012. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. *Remote Sens. Environ.* 122, 50–65. <https://doi.org/10.1016/j.rse.2011.08.025>.
- Arshad, A., Mirchi, A., Taghvaeian, S., Aghakouchak, A., 2024. Downscaled-GRACE data reveal anthropogenic and climate-induced water storage decline across the Indus Basin. *Water Resour. Res.* 60 (7). <https://doi.org/10.1029/2023WR035882> (e2023WR035882).
- Ascoura, I.E., 2013. Impact of pilgrimage (Hajj) on the urban growth of the Mecca. *J. Educ. Soc. Res.* 3 (2), 255–263.
- Bhanja, S.N., Mukherjee, A., Saha, D., Velicogna, I., Famiglietti, J.S., 2016. Validation of GRACE based groundwater storage anomaly using in-situ groundwater level measurements in India. *J. Hydrol.* 543, 729–738.
- Bhattarai, N., Lobell, D.B., Balwinder-Singh, Fishman, R., Kustas, W.P., Pokhrel, Y., Jain, M., 2023. Warming temperatures exacerbate groundwater depletion rates in India. *Sci. Adv.* 9 (35), eadi1401. <https://doi.org/10.1126/sciadv.ad1401>.
- Bierkens, M.F., Wada, Y., 2019. Non-renewable groundwater use and groundwater depletion: a review. *Environ. Res. Lett.* 14 (6), 063002. <https://doi.org/10.1088/1748-9326/ab1a5f>.

- Bilbao-Barrenetxea, N., Martínez-España, R., Jimeno-Sáez, P., Faria, S.H., Senent-Aparicio, J., 2024. Multi-model ensemble machine learning approaches to project climatic scenarios in a River Basin in the pyrenees. *Earth Syst. Environ.* 8, 1–19.
- Chen, C., Chen, Q., Qin, B., Zhao, S., Duan, Z., 2020. Comparison of different methods for spatial downscaling of GPM IMERG V06B satellite precipitation product over a typical arid to semi-arid area. *Front. Earth Sci.* 8, 536337. <https://doi.org/10.3389/feart.2020.536337>.
- Chen, Q., Miao, F., Wang, H., Xu, Z.X., Tang, Z., Yang, L., Qi, S., 2020. Downscaling of satellite remote sensing soil moisture products over the Tibetan Plateau based on the random forest algorithm: preliminary results. *Earth Space Sci.* 7 (6). <https://doi.org/10.1029/2020EA001265> (e2020EA001265).
- 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. *Sci. Total Environ.* 649, 372–387. <https://doi.org/10.1016/j.scitotenv.2018.08.352>.
- Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. STL: A seasonal-trend decomposition. *J. Off. Stat.* 6 (1), 3–73.
- Cotterman, K.A., Kendall, A.D., Basso, B., Hyndman, D.W., 2018. Groundwater depletion and climate change: future prospects of crop production in the Central High Plains Aquifer. *Clim. Change* 146, 187–200. <https://doi.org/10.1007/s10584-017-1947-7>.
- Do, S.K., Akhtar, F., Goffin, B., Aryal, A., Lipscomb, M., Lakshmi, V., 2024a. Assessing terrestrial water storage variations in Afghanistan using GRACE and FLDAS-Central Asia data. *J. Hydrol. Reg. Stud.* 55, 101906. <https://doi.org/10.1016/j.ejrh.2024.101906>.
- Do, S.K., Nguyen, B.Q., Tran, V.N., Grodzka-Lukaszewska, M., Sinićyn, G., Lakshmi, V., 2024b. Investigating the future flood and drought shifts in the transboundary Srepok river basin using CMIP6 projections. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* <https://doi.org/10.1109/jstars.2024.3380514>.
- Dou, J., et al., 2020. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* 17, 641–658.
- Eicker, A., Schumacher, M., Kusche, J., Döll, P., Schmied, H.M., 2014. Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an ensemble Kalman filter: First results. *Surv. Geophys.* 35 (6), 1285–1309.
- Esper, J., Torbenson, M., Büntgen, U., 2024. 2023 summer warmth unparalleled over the past 2000 years. *Nature* 631, 94–97. <https://doi.org/10.1038/s41586-024-07512-y>.
- Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9, 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>.
- Famiglietti, J.S., 2014. The global groundwater crisis. *Nat. Clim. Change* 4 (11), 945–948. <https://doi.org/10.1038/nclimate2425>.
- Gazzawe, F., Albahee, M., 2024. Reducing traffic congestion in makkah during hajj through the use of ai technology. *Heliyon* 10 (1), e23304. <https://doi.org/10.1016/j.heliyon.2023.e23304>.
- Gerdener, H., Kusche, J., Schulze, K., Döll, P., Klos, A., 2023. The global land water storage data set release 2 (GLWS2.0) derived via assimilating GRACE and GRACE-FO data into a global hydrological model. *J. Geod.* 97 (7), 73.
- Girotto, M., Rodell, M., 2019. Terrestrial water storage. Extreme hydroclimatic events and multivariate hazards in a changing environment. Elsevier, pp. 41–64.
- Gleeson, T., Cuthbert, M., Ferguson, G., Perrone, D., 2019. Global groundwater sustainability, resources, and systems in the anthropocene. *Annu. Rev. Earth Planet. Sci.* 48 (1), 431–463. <https://doi.org/10.1146/annurev-earth-071719-055251>.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. (<https://doi.org/10.1016/j.rse.2017.06.031>).
- Gou, J., Soja, B., 2024. Global high-resolution total water storage anomalies from self-supervised data assimilation using deep learning algorithms. *Nat. Water* 2 (2), 139–150.
- Guo, L., Li, T., Chen, D., Liu, J., He, B., Zhang, Y., 2021. Links between global terrestrial water storage and large-scale modes of climatic variability. *J. Hydrol.* 598, 126419.
- Hossein Abadi, L., Aghighi, H., Matkan, A., Shakiba, A., 2023. Downscaling and evaluation of evapotranspiration using remotely sensed data and machine learning algorithms (study area: Moghan Plain, Iran). *ISPRS Ann. Photogramm. Remote Sens. Inf. Sci.* 10, 295–300. <https://doi.org/10.5194/isprs-annals-X-4-W1-2022-295-2023>.
- Ibrahim, A., Wayayok, A., Shafri, H.Z.M., Toridi, N.M., 2024. Remote sensing technologies for unlocking new groundwater insights: a comprehensive review. *J. Hydrol.* X 23, 100175. <https://doi.org/10.1016/j.hydora.2024.100175>.
- Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsuddoha, M., Taylor, R.G., Kirchner, J.W., 2024. Rapid groundwater decline and some cases of recovery in aquifers globally. *Nature* 625 (7996), 715–721. <https://doi.org/10.1038/s41586-023-06879-8>.
- Jawadi, H.A., Farahmand, A., Fensham, R., Patel, N., 2024. Evaluating groundwater storage variations in Afghanistan using GRACE, GLDAS, and in-situ measurements. *Model. Earth Syst. Environ.* 1–17. <https://doi.org/10.1007/s40808-024-02084-2>.
- Jódar, J., Urrutia, J., Herrera, C., Custodio, E., Martos-Rosillo, S., Lambán, L.J., 2024. The catastrophic effects of groundwater intensive exploitation and Megadrought on aquifers in Central Chile: global change impact projections in water resources based on groundwater balance modeling. *Sci. Total Environ.* 914, 169651. <https://doi.org/10.1016/j.scitotenv.2023.169651>.
- Kennedy, C.M., Oakleaf, J.R., Theobald, D.M., Bauch-Murdo, S., Kiesecker, J., 2019. Managing the middle: a shift in conservation priorities based on the global human modification gradient. *Glob. Change Biol.* 1–16. <https://doi.org/10.1111/gcb.14549>.
- Khorrami, B., Ali, S., Sahin, O.G., Gunduz, O., 2023. Model-coupled GRACE-based analysis of hydrological dynamics of drying Lake Urmia and its basin. *Hydroclim. Process.* 37 (5), e14893.
- Khosravi, Y., Ouarda, T.B., Homayouni, S., 2025. Developing an ensemble machine learning framework for enhanced climate projections using CMIP6 data in the Middle East. *npj Clim. Atmos. Sci.* 8 (1), 1–24.
- Kim, Jae-Seung, Ki-Weon Seo, Byeong-Hoon Kim, Ryu, Dongryeol, Chen, Jianli, Clark Wilson, 2024b. High-resolution terrestrial water storage estimates from GRACE and land surface models. *Water Resour. Res.* 60 (2) (e2023WR035483).
- Kim, J.-S., Seo, K.-W., Kim, B.-H., Ryu, D., Chen, J., Wilson, C., 2024a. High-resolution terrestrial water storage estimates from GRACE and land surface models. *Water Resour. Res.* 60 (2), e2023WR035483. <https://doi.org/10.1029/2023WR035483>.
- Lakshmi, V., 2024b. Visualization-driven hydrologic assessment using gridded precipitation products. *Hydroclim. Process.* 38 (10), e15286. <https://doi.org/10.1002/hyp.15286>.
- Li, W., Ni, L., Li, Z.L., Duan, S.B., Wu, H., 2019b. Evaluation of machine learning algorithms in spatial downscaling of MODIS land surface temperature. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12 (7), 2299–2307. <https://doi.org/10.1109/jstars.2019.2896923>.
- Li, B., Rodell, M., Sheffield, J., Wood, E., Sutanudjaja, E., 2019a. Long-term, non-anthropogenic groundwater storage changes simulated by three global-scale hydrological models. *Sci. Rep.* 9 (1), 10746.
- Liu, P.W., Famiglietti, J.S., Purdy, A.J., Adams, K.H., McEvoy, A.L., Reager, J.T., Rodell, M., 2022. Groundwater depletion in California's Central Valley accelerates during megadrought. *Nat. Commun.* 13 (1), 7825. <https://doi.org/10.1038/s41467-022-35582-x>.
- Liu, Y., Jing, W., Wang, Q., Xia, X., 2020. Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms. *Adv. Water Resour.* 141, 103601. <https://doi.org/10.1016/j.advwatres.2020.103601>.
- Luo, Q., Liang, Y., Guo, Y., Liang, X., Ren, C., Yue, W., Jiang, X., 2023. Enhancing spatial resolution of GNSS-R soil moisture retrieval through XGBoost algorithm-based downscaling approach: a case study in the Southern United States. *Remote Sens.* 15 (18), 4576. <https://doi.org/10.3390/rs15184576>.
- MacAllister, D.J., Krishan, G., Basharat, M., Cuba, D., MacDonald, A.M., 2022. A century of groundwater accumulation in Pakistan and northwest India. *Nat. Geosci.* 15 (5), 390–396. <https://doi.org/10.1038/s41561-022-00926-1>.
- Marshall, S.R., Tran, T.N.D., Tapas, M.R., Nguyen, B.Q., 2025. Integrating artificial intelligence and machine learning in hydrological modeling for sustainable resource management. *Int. J. River Basin Manag.* 1–17. <https://doi.org/10.1080/15715124.2025.2478280>.
- Mir, M.A., Ashraf, M.W., 2023. The challenges and potential strategies of Saudi Arabia's water resources: a review in analytical way. *Environ. Nanotechnol. Monit. Manag.* 20, 100855. <https://doi.org/10.1016/j.enmm.2023.100855>.
- Mishra, V., Aadhar, S., Mahto, S.S., 2021. Anthropogenic warming and intraseasonal summer monsoon variability amplify the risk of future flash droughts in India. *Npj Clim. Atmos. Sci.* 4 (1), 1.

- Mishra, V., Dangar, S., Tiwari, V.M., Lall, U., Wada, Y., 2024. Summer monsoon drying accelerates India's groundwater depletion under climate change. *Earth's Future* 12 (8). <https://doi.org/10.1029/2024EF004516> (e2024EF004516).
- Mohamed, A., Abdelrahman, K., Abdelrady, A., 2022. Application of time-variable gravity to groundwater storage fluctuations in Saudi Arabia. *Front. Earth Sci.* 10, 873352.
- Mohamed, A., Alarifi, S.S., Mohammed, M.A., 2024. Geophysical monitoring of the groundwater resources in the southern Arabian peninsula using satellite gravity data. *Alex. Eng. J.* 86, 311–326. <https://doi.org/10.1016/j.aej.2023.11.060>.
- Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Döll, P., 2021. The global water resources and use model WaterGAP v2. 2d: model description and evaluation. *Geosci. Model Dev.* 14 (2), 1037–1079.
- Muñoz Sabater, J., 2019. ERA5-Land monthly averaged data from 1981 to present. Copernic. Clim. Change Serv. (C3S) Clim. Data Store (CDS). <https://doi.org/10.24381/cds.68d2bb30>.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Thépaut, J.N., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth syst. sci. data* 13 (9), 4349–4383.
- Murphy, R.R., Perry, E., Harcum, J., Keisman, J., 2019. A generalized additive model approach to evaluating water quality: Chesapeake Bay case study. *Environ. Model. Softw.* 118, 1–13. <https://doi.org/10.1016/j.envsoft.2019.03.027>.
- Neelam, M., Hain, C., 2024. Global flash droughts characteristics: onset, duration, and extent at watershed scales. *Geophys. Res. Lett.* 51 (10) (e2024GL109657).
- Nguyen, B.Q., Van Binh, D., Tran, T.N.D., Kantoush, S.A., Sumi, T., 2024. Response of streamflow and sediment variability to cascade dam development and climate change in the Sai Gon Dong Nai River basin. *Clim. Dyn.* 1–21. <https://doi.org/10.1007/s00382-024-07319-7>.
- Othman, A., Abotalib, A.Z., 2019. Land subsidence triggered by groundwater withdrawal under hyper-arid conditions: case study from Central Saudi Arabia. *Environ. Earth Sci.* 78 (7), 243.
- Rajaeian, S., Ketabchi, H., Ebadi, T., 2024. Investigation on quantitative and qualitative changes of groundwater resources using MODFLOW and MT3DMS: a case study of Hashtgerd aquifer, Iran. *Environ. Dev. Sustain.* 26 (2), 4679–4704. <https://doi.org/10.1007/s10668-022-02904-4>.
- Rateb, A., Scanlon, B.R., Pool, D.R., Sun, A., Zhang, Z., Chen, J., Zell, W., 2020. Comparison of groundwater storage changes from GRACE satellites with monitoring and modeling of major US aquifers. *Water Resour. Res.* 56 (12) (e2020WR027556).
- Rice, J.S., Emanuel, R.E., 2017. How are streamflow responses to the El Niño Southern Oscillation affected by watershed characteristics? *Water Resour. Res.* 53, 4393–4406. Return to ref 60 in article.
- Rodell, M., Chen, J., Kato, H., Famiglietti, J.S., Nigro, J., Wilson, C.R., 2007. Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol. J.* 15, 159–166.
- Roy, D., Das, B., Rathore, P., Santra, P., Deb, S., Bhattacharya, B.K., Chakraborty, D., 2024. Downscaling Land Surface Temperature at Farm-Scale Using Multi-Model Approaches over Diverse Agro Climatic Zones. 2024010413. (<https://doi.org/10.20944/preprints202401.0413.v1>).
- Sabzehei, F., Amiri-Simkooei, A.R., Iran-Pour, S., Vishwakarma, B.D., Kerachian, R., 2023. Enhancing spatial resolution of GRACE-derived groundwater storage anomalies in Urmia catchment using machine learning downscaling methods. *J. Environ. Manag.* 330, 117180.
- Senanayake, I.P., Pathira Arachchilage, K.R., Yeo, I.Y., Khaki, M., Han, S.C., Dahlhaus, P.G., 2024. Spatial downscaling of satellite-based soil moisture products using machine learning techniques: a review. *Remote Sens.* 16 (12), 2067. <https://doi.org/10.3390/rs16122067>.
- Sims, K., Reith, A., Bright, E., Kaufman, J., Pyle, J., Epting, J., Gonzales, J., Adams, D., Powell, E., Urban, M., Rose, A., 2023. LandScan Global 2022 [Data Set]. Oak Ridge National Laboratory. <https://doi.org/10.48690/1529167>.
- Sun, W., Trevor, B., 2018. A stacking ensemble learning framework for annual river ice breakup dates. *J. Hydrol.* 561, 636–650.
- Tabash, M., Farooq, U., Refaei, G., Al-Faryan, M., Athamena, B., 2023. Impact of religious tourism on the economic development, energy consumption and environmental degradation: evidence from the Kingdom of Saudi Arabia. *Tour. Rev.* <https://doi.org/10.1108/tr-07-2022-0347>.
- Tapas, M.R., Do, S.K., Etheridge, R., Lakshmi, V., 2024. Investigating the impacts of climate change on hydroclimatic extremes in the Tar-Pamlico River basin, North Carolina. *J. Environ. Manag.* 363, 121375. <https://doi.org/10.1016/j.jenvman.2024.121375>.
- Tapley, B.D., Bettadpur, S., Watkins, M., Reigber, C., 2004. The gravity recovery and climate experiment: mission overview and early results. *Geophys. Res. Lett.* 31 (9). <https://doi.org/10.1029/2004GL019920>.
- Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., Treidel, H., 2013. Ground water and climate change. *Nat. Clim. Change* 3 (4), 322–329. <https://doi.org/10.1038/nclimate1744>.
- Toth, J., 1963. A theoretical analysis of groundwater flow in small drainage basins. *J. Geophys. Res.* 68 (16), 4795–4812. <https://doi.org/10.1029/JZ068i016p04795>.
- Tran, T.N.D., Nguyen, B.Q., Grodzka-Lukaszewska, M., Sinicyn, G., Lakshmi, V., 2023. The role of reservoirs under the impacts of climate change on the Srepok River basin, Central Highlands of Vietnam. *Front. Environ. Sci.* 11, 1304845. <https://doi.org/10.3389/fenvs.2023.1304845>.
- Usman, M., Heki, K., 2024. Satellite gravimetry observations on the state of groundwater level variability in the Arabian Peninsula Region and the associated socio-economic sustainability challenges. *Groundw. Sustain. Dev.* 26, 101270.
- Uz, M., Akyilmaz, O., Shum, C.K., 2024. Deep learning-aided temporal downscaling of GRACE-derived terrestrial water storage anomalies across the Contiguous United States. *J. Hydrol.* 132194.
- Van der Laan, M.J., Polley, E.C., Hubbard, A.E., 2007. Super learner. *Stat. Appl. Genet. Mol. Biol.* 6 (1).
- Wang, Y., Li, C., Cui, Y., Cui, Y., Xu, Y., Hora, T., Long, D., 2024. Spatial downscaling of GRACE-derived groundwater storage changes across diverse climates and human interventions with Random Forests. *J. Hydrol.* 640, 131708.
- Wang, T., Zhang, Y., Zhi, X., Ji, Y., 2023. Multi-model ensemble forecasts of surface air temperatures in Henan Province based on machine learning. *Atmosphere* 14 (3), 520. <https://doi.org/10.3390/atmos14030520>.
- Wood, S.N., 2020. Inference and computation with generalized additive models and their extensions. *Test* 29 (2), 307–339. <https://doi.org/10.1007/s11749-020-00711-5>.
- Wood, S.N., Pya, N., Saefken, B., 2016. Smoothing parameter and model selection for general smooth models (with discussion). *J. Am. Stat. Assoc.* 111, 1548–1575. <https://doi.org/10.1080/01621459.2016.1180986>.
- Xu, L., Ferris, D., Huggins, X., Wong, J.S., Mohan, C., Sadri, S., Famiglietti, J.S., 2023. From coarse resolution to practical solution: GRACE as a science communication and policymaking tool for sustainable groundwater management. *J. Hydrol.* 623, 129845.
- Yang, B., Liu, H., Kang, E.L., Shu, S., Xu, M., Wu, B., Yu, B., 2021. Spatio-temporal Cokriging method for assimilating and downscaling multi-scale remote sensing data. *Remote Sens. Environ.* 255, 112190. <https://doi.org/10.1016/j.rse.2020.112190>.
- Yezli, S., Yasin, Y.M., Awam, A.H., Attar, A.A., Al-Jahdali, E.A., Alotaibi, B.M., 2017. Umrah: An opportunity for mass gatherings health research. *Saudi Med. J.* 38 (8), 868.
- Yin, W., Zhang, G., Han, S.C., Yeo, I.Y., Zhang, M., 2022a. Improving the resolution of GRACE-based water storage estimates based on machine learning downscaling schemes. *J. Hydrol.* 613, 128447. <https://doi.org/10.1016/j.jhydrol.2022.128447>.
- Yin, W., Zhang, G., Liu, F., Zhang, D., Zhang, X., Chen, S., 2022b. Improving the spatial resolution of GRACE-based groundwater storage estimates using a machine learning algorithm and hydrological model. *Hydrogeol. J.* 30 (3), 947–963.
- Zhai, B., Chen, J., 2018. Development of a stacked ensemble model for forecasting and analyzing daily average PM2. 5 concentrations in Beijing, China. *Sci. Total Environ.* 635, 644–658.