scientific reports

Check for updates

OPEN

Assessing climate and land use impacts on surface water yield using remote sensing and machine learning

Amanuel Kumsa Bojer[1,2]✉, Muluneh Woldetsadik Abshare[1]✉, Fitsum Mesfin[2] & Ayad M. Fadhil Al-Quraishi[3]

Climate and land use changes are critical factors affecting watershed water yields, with significant implications for water resources at both local and regional levels. This study examined the combined effects of temporal and spatial climate variability and land use/land cover (LULC) changes on surface water yield and availability in the Gilgel Gibe watershed, Ethiopia, from 1993 to 2023. Utilizing the Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST) hydrological models, machine learning, and remote sensing techniques, this study assessed variations in water resources and their impacts on basin water yield. This study utilized Landsat (30 m), MODIS (500 m–1 km), and 4 km resolution climate datasets from the United States Geological Survey (USGS) and NASA POWER for large-scale climate and land-use analyses from 1993 to 2023. An ensemble of machine learning models, including Random Forest (RF), Support Vector Machine (SVM), and XGBoost (XGB), were used to evaluate the effects of climate variability and land use on annual water yield. The study revealed significant land cover changes over a 30-year period. Shrubland decreased from 1,108.37 km^2 (21.54%) in 1993 to 295.22 km^2 (5.74%) in 2023. Grasslands and wetlands also showed declining trends. In contrast, water bodies increased from 12.51 km^2 (0.24%) to 41.57 km^2 (0.81%), primarily due to the construction of the Gilgel Gibe hydroelectric dam, and forested areas slightly decreased from 626.73 km^2 (12.18%) to 534.18 km^2 (10.38%). The surface runoff decreased to 15.78% in 2021 and 15.28% in 2022, whereas the water yield dropped from 1.22% in 1993 to 0.83% by 2023. This study also showed a reduction in lateral flow and higher evapotranspiration levels in 2000 and 2017. The decrease in runoff can be attributed to the loss of wetlands and grasslands, reduced precipitation, and regulatory effects of hydropower operations. In contrast, elevated evapotranspiration levels were primarily attributed to temperature extremes, vegetation stress, and potential increases in irrigation practices. These findings underscore the importance of climatic elements in regulating river discharge and the necessity for smart land use planning to prevent negative environmental consequences on water resources.

Keywords Water yield model, Climate change, LULC change, Evapotranspiration, Integrated impact analysis, Gilgel gibe watershed, Machine learning & remote sensing

Climate and LULC changes are increasingly recognized as significant threats to the sustainability of water resources, particularly within African watersheds[1]. These changes can lead to altered rainfall patterns, increased evapotranspiration, and changes in surface and groundwater availability, thereby affecting human populations and ecosystems. Watersheds in sub-Saharan Africa have also received limited research attention in terms of LULC changes, particularly in response to increasing anthropogenic pressure and disturbances[2]. Climate projections indicate that these anthropogenic drivers will significantly alter hydrological patterns by 2050, further exacerbating water resource challenges[3]. The combined effects of climate change and LULC change are expected to substantially modify seasonal streamflow patterns and overall water availability, presenting significant challenges in achieving Sustainable Development Goals (SDGs) for water resources in the foreseeable future[4–7]. Water yield, a crucial indicator of water resource availability within watersheds, is significantly influenced

[1]Department of Geography and Environmental Studies, Addis Ababa University, P.O. Box 1176, Addis Ababa, Ethiopia. [2]Ethiopian Artificial Intelligence Institute, PO Box 40782, Addis Ababa, Ethiopia. [3]Petroleum and Mining Engineering Department, Tishk International University, Erbil, Iraq. ✉email: amanuel.kumsa@aau.edu.et; mulunehabshare@gmail.com

by climate and LULC changes[8–10]. It is essential to understand how these factors impact water yield, as their effects differ based on the unique geographic, climatic, and socioeconomic characteristics of each watershed[11]. However, there are still gaps in our understanding of the diverse impacts across various watersheds, especially concerning how watershed-specific factors influence the effects of these environmental changes[12–15]. In addition, understanding the long-term influence of LULC changes on water resources is essential for evaluating the current and future impacts on hydrological systems.

Moreover, numerous studies have examined the effects of climate and land use at different scales, and there is a significant gap in the literature regarding the simultaneous analysis of similarities and differences in their influence on water yields across multiple watersheds. Previous studies have noted that geographic and climatic characteristics play a vital role in shaping water yield responses to climate and land use changes[16–20]. Nevertheless, the current literature lacks a comprehensive analysis that concurrently examines the effects of climate change and LULC across multiple watersheds with varied characteristics[21,22]. This gap may be due to the limitations of models such as InVEST, which, although effective in calculating watershed water supply, often fails to sufficiently capture factors specific to each watershed. Furthermore, inconsistencies in data and standard concerns may hinder accurate water yield assessments.

Water yield is an important measure of ecosystem health and sustainable development[23,24]. To better understand watershed dynamics, it is essential to analyze the spatial and temporal variations in water yield. The Gilgel Gibe watershed is particularly significant, not only for its role in powering Ethiopia's Gibe I, II, and III hydroelectric systems, but also for serving as a key water source for Jimma Town. This watershed faces significant environmental pressure owing to human activity, which affects its ability to provide essential ecosystem services. The main factors contributing to these pressures were changes in land use and climate. Prior studies using the InVEST model have assessed the watershed water production capacity and explored the effects of climate and land-use changes on water yield[25–27]. Despite extensive research, there is a significant gap in the literature regarding the simultaneous analysis of similarities and differences in the influence of climate and LUC on water yields across multiple watersheds. This inconsistency can be attributed to several factors. First, while the InVEST model is commonly utilized for estimating watershed water yield, it lacks a tailored approach for deriving watershed-specific parameters for various watersheds. Second, the collection and processing of extensive data from watersheds is essential because ensuring data quality directly affects the accuracy of water yield assessments. Various data types often require transformation into a uniform format that aligns with modeling requirements.

This study uniquely investigated the combined effects of climate and LULC change on water yield in the Gilgel Gibe watershed, a region of substantial ecological, agricultural, and ecosystem value. The combined effects of these factors have significant impacts on regional water resource availability and biodiversity conservation, as well as vital ramifications for ecological stability and sustainable management. How do changes in LULC, specifically deforestation and agricultural expansion, affect the water yield trends over time in the region? To what extent do climate variability factors, such as changes in rainfall and temperature patterns, amplify the impacts of LULC changes on water yield? What are the integrated impacts of altered water yield on regional water resources and biodiversity conservation, and what are the broader implications for ecological sustainability? This watershed plays a critical role in downstream irrigation, municipal water supply, and biodiversity conservation, making it an ideal case study to explore how these changes can jeopardize water resource sustainability. By utilizing integrated modeling frameworks and advanced approaches, including machine learning and remote sensing, this study aims to bridge critical knowledge gaps regarding watershed hydrology. The diverse land cover of the Gilgel Gibe watershed, encompassing forests, shrublands, grasslands, agricultural areas, and wetlands, is essential for examining the varying effects of LULC changes on hydrological processes.

Additionally, the diverse geographic and climatic conditions within the watershed, including fluctuating rainfall patterns and elevation gradients, further enhance its suitability as a model for studying the spatial variability in water yield and availability. The presence of hydropower projects in the area also ensures the availability of substantial hydrological data, including river flow, reservoir levels, and meteorological data, which are essential for this type of analysis. Recent studies by[28,29] highlighted the significance of integrating these advanced methodologies to reveal the spatiotemporal patterns of LULC change impacts on the water availability. Moreover, integrating climate projections into these models enables the exploration of potential future scenarios, providing insights into the risks of water scarcity and the socioeconomic impacts of diminishing water availability[30,31]. Water yield declines in the Gilgel Gibe watershed might have a substantial impact on hydropower production, which is essential to Ethiopia's energy supply. Reduced base flow and higher sedimentation are expected to reduce reservoir storage capacity and turbine efficiency, potentially resulting in significant revenue losses of millions of dollars each year. For example, in May 2019, the Ethiopian government and the Ministry of Water, Irrigation, and Electricity (MoWIE) estimated a 476-megawatt power shortage caused by climate change impacts on the Omo-Gibe dams in the Omo-Gibe River basin[32]. On the agricultural front, reduced water availability may limit irrigation capacity, resulting in significant yield decreases for staple crops, such as maize and teff, which are critical for local people's survival. The Gibe River watershed includes three cascading dams, Gilgel Gibe I, Gilgel Gibe II, and Gilgel Gibe III, that support irrigation projects on 175,000 hectares of agricultural land[33].

By integrating climate and land use scenarios, this study aims to provide practical insights for policymakers and stakeholders to facilitate more sustainable water management in response to changing environmental conditions. The primary objectives of this study were as follows: (1) to evaluate temporal variations in surface water yield under climate variability, (2) to analyze the spatial distribution of LULC change and its direct impact on water availability, and (3) to develop predictive models that integrate climate and land use projections to forecast future water yield fluctuations. Furthermore, the research findings provide essential hydrological insights, presenting transferable methods and lessons that may be applied to model-driven watershed studies, both locally and worldwide. These findings are significant for designing efficient water management strategies

in the Gilgel Gibe watershed and understanding how environmental changes affect water yields in this critical region. Incorporating scenario-based analyses and methodologies from similar research could improve the study's results, giving policymakers solid facts to address socioeconomic concerns and encourage sustainable water resource management practices.

Materials and methods

Study area

The Gilgel Gibe watershed is located 272 km southwest of Addis Ababa, in the Oromia region of Ethiopia (Fig. 1). The watershed lies between 7°19′N and 8°4′N latitude and 36°37′E and 37°53′E longitude. The Wabe River, which drains much of the West Gurage land, contributes to the Omo River, forming part of the Omo-Gibe River Basin. This basin encompasses both the Gibe and Omo Rivers, which drain the upper and lower parts of the watershed, respectively. These watersheds have a total catchment area of 5,105.22 km^2, of which the Gibe watershed comprises 961.21 km^2 (18.83%). This study assessed the potential impacts of climate and LULC changes on water yield in the Gilgel Gibe watershed, which was selected based on several criteria. First, the availability of reliable runoff data at the watershed outlets is a major consideration. Second, preference was given to watersheds in the upper reaches of river systems to minimize external influences from adjacent watersheds. Third, the watershed's diverse topography, climatic conditions, and socio-economic activities, especially varying precipitation patterns, were factors that made it suitable for comprehensive analysis. These considerations ensure the ability of the study to accurately assess the watershed response to potential future climate and land-use changes.

The Gilgel Gibe watershed exhibits a contrasting elevation range, varying from 1,100 m above the mean sea level (m.a.s.l.) in the lowlands to 3,341 m.a.s.l. in the highlands. Such variation could have significantly influenced precipitation patterns, with annual rainfall ranging from 990 mm in the southern lowlands to 2,500 mm in the highlands. The average annual rainfall across the basin was about 1,508 mm. The mean annual temperature within the basin varies with regional differences, ranging from less than 7.53 °C in the western highlands to approximately 34 °C in the southern lowlands[34,35]. Highland areas are distinguished by their steep slopes and rugged, dissected hills, whereas lowland regions are characterized by smoother, more gently undulating terrain. The primary soil classifications observed within the study area contained predominantly Nitosols (46.72%), followed by Vertisols (27.75%), and Luvisols (18.31%). Additionally, minor occurrences of Leptosols (3.32%), Cambisols (2.4%), and Xerosols (0.61%) were also recognized[36].

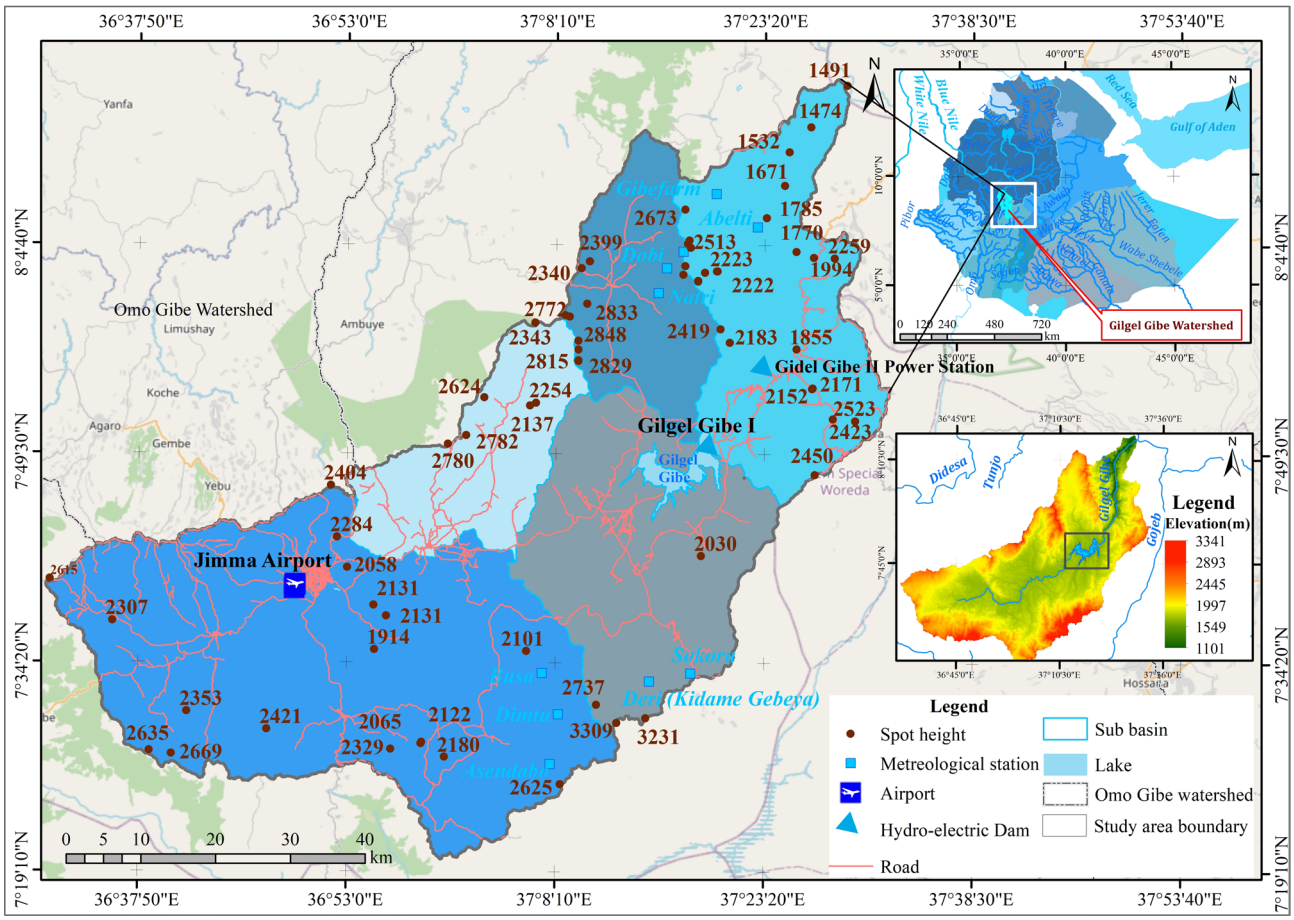

Fig. 1. Location map of the study area.Source: The maps were created using ArcGIS software (version 10.8).

Data source and preprocessing

Data preprocessing plays an essential role in LULC classification and has a substantial impact on the accuracy of classification results[37]. This involves several procedures performed before LULC classification to enhance the data quality and optimize the accuracy of the classification outputs[38]. Earth Observation Data (EOD), particularly satellite imagery, plays an important role in the wide use of Landsat data because of their availability through platforms such as the United States Geological Survey (USGS) and Google Earth Engine (GEE). In this study, Landsat-5, Landsat-7, and Landsat-8&9 surface reflectance Tier 1 data were selected for their high quality and accuracy. The use of multitemporal Landsat data presents several challenges, particularly owing to cloud cover and the occurrence of missing data. Cloud cover significantly affects the quality of satellite images. To ensure the reliability of the data, atmospheric corrections such as geometric, radiometric, and noise removal were applied using the Landsat-8–9 Surface Reflectance Code (LASRC). LaSRC algorithm enables greater flexibility in processing Landsat data compared with using pre-processed Level-2 products. This flexibility allowed us to address the unique environmental conditions in the Gilgel Gibe watershed, such as variations in atmospheric composition, topography, and seasonal dynamics, ensuring more accurate reflectance values.

Landsat Surface Reflectance (Level-2) is reliable for general applications and may not always be consistent for all periods or regions, particularly in areas with complex terrain and microclimates. Therefore, the use of LaSRC preprocessing techniques in this study is essential to effectively address atmospheric issues and ensure high-quality data tailored to the specific environmental characteristics of the study area. Moreover, a combination of software tools Arc Geographic Information System (ArcGIS Desktop 10.8), ERDAS Imagine Hexagon 2015, and Environment for Visualizing Images (ENVI) Exelis 5.0 was used for preprocessing tasks, including image registration, resampling, mosaicking, image classification, and extracting relevant features from the geospatial data[39]. Additionally, Invest 3.13.0 played an essential role in integrating the biophysical parameters and hydrological data, which provided a valuable context for our analysis. It provides spatially explicit assessments that quantify the impacts of land use and climate change on ecosystem services, including water yield and sediment retention, to inform sustainable resource management decisions. This integration allowed for more precise modeling of the interactions between land use, climate change, and water yield. Furthermore, advanced machine learning models were employed to model the combined effects of LULC changes and climate variability on water yield within the Gilgel Gibe watershed. These models, including Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), provide a sophisticated framework for analyzing complex interactions and forecasting future hydrological trends based on historical data and projected scenarios. The land cover analysis was conducted for the years 1993, 2001, 2009, 2017, and 2023, providing a comprehensive overview of the changes in the study area over time (Table 1).

A comprehensive analysis was conducted using satellite imagery from various Landsat missions to assess the integrated impact of climate change and LULC changes on the Gilgel Gibe water yield. Specifically, data from the Operational Land Imager (OLI) on Landsat 8–9, the Enhanced Thematic Mapper Plus (ETM +) on Landsat 7, and the Thematic Mapper (TM) on Landsat 5 were utilized. These datasets provide comprehensive insights into temporal and spatial changes in land cover over time. Each LULC class was systematically classified according to the description provided in Table 2. LULC classifications describe separate land surface types based on human activities and natural vegetation. Each class plays a crucial role in ecosystem function, and changes in these classes owing to land cover or climate change can significantly impact hydrological processes and ecosystem services. This study employed a supervised maximum likelihood classification method to analyze LULC in the Gilgel Gibe Watershed. This method is widely used and recognized in remote sensing because of its ability to classify satellite imagery based on predefined training samples. Supervised classification involves selecting representative sample areas within the imagery that are known to correspond to specific land cover types, such as forests, shrubland, grassland, wetland water bodies, croplands, and urban areas. These training samples are used to train a classifier algorithm, such as the maximum likelihood, which calculates the probability of each pixel in the image belonging to each predefined class.

Data types

This study used hydroclimatic data from multiple sources. Specifically, data on minimum and maximum temperatures as well as rainfall from 1993 to 2023 were acquired from the Ethiopian Meteorological Institute (EMI) and National Aeronautics and Space Administration (NASA) power access (https://power.larc.nasa.gov/data-access-viewer/). This hydroclimate dataset was presented in a gridded format with a resolution of 4 km × 4 km. It combines observed station data from the national network managed by EMI with satellite-based rainfall and temperature estimates from reliable sources such as the National Oceanic and Atmospheric Administration

Year	sensor	Data type	Resolution (m)	Bands	Path/Row	Sensor type	Wavelengths (µm)	Preprocessing
1993	Landsat 5 TM	Surface reflectance tier 1	30	7	169/054, 169/055, 170/054, 170/055	Multispectral	0.45—12.5	Atmospheric correction: LASRC
2001	Landsat 7 ETM +	Surface reflectance tier 1	30	8	169/054, 169/055, 170/054, 170/055	Multispectral	0.43—2.13	Atmospheric correction: LASRC
2009	Landsat 5 TM	Surface reflectance tier 1	30	7	169/054, 169/055, 170/054, 170/055	Multispectral	0.45—12.5	Atmospheric correction: LASRC
2017	Landsat 8 OLI	Surface reflectance tier 1	30	9	169/054, 169/055, 170/054, 170/055	Multispectral	0.43—2.30	Atmospheric correction: LASRC
2023	Landsat 8 OLI	Surface reflectance tier 1	30	9	169/054, 169/055, 170/054, 170/055	Multispectral	0.43—2.30	Atmospheric correction: LASRC

Table 1. Description of satellite image data.

LULC Class	Description
Cultivated land	Cultivated lands, small-scale irrigation areas, and scattered rural settlements were grouped into one category due to the challenge of distinguishing between rural settlements and fragmented cropland
Forestland	Areas with dense forests and closed canopies, such as eucalyptus plantations within homesteads and cultivated lands
Grazing land	Areas are mainly covered with grasses, serving as grazing fields and sources of animal feed
Built-up area	Regions are characterized by entirely artificial surfaces, encompassing residential zones such as emerging rural towns, villages, housing developments, and various other built structures
Shrubland	Land covered mainly by shrub-bushland, with scattered small trees and low-growing grass
Waterbody	Areas covered by natural and man-made water bodies, such as lakes, rivers, ponds, and reservoirs
Wetland	Wetlands and marshy areas that are seasonally or permanently saturated with water provide important ecological functions

Table 2. Categories of LULC changes in the study area.

(NOAA), the Coupled Model Intercomparison Project Phase 6 (CMIP6), and the US NASA[40–42]. This gridded dataset serves as high-quality data, filling the spatial and temporal gaps in Ethiopia's national observation network[43].

Additionally, this acquired precipitation, LULC, potential evapotranspiration (PET), soil moisture, and watershed boundary data from the Google Earth Engine and the Ethiopian Space Science and Geospatial Institute (SSGI). Furthermore, a biophysical table containing attributes for each LULC type, along with the seasonality factor (Z), was obtained. The biophysical table aided the model in differentiating between wet and dry season water yield, which is essential for managing water resources effectively. Daily rainfall and temperature data from four meteorological stations within and surrounding the Gilgel Gibe watershed were obtained from EMI for 30 years, from 1993 to 2023. To address spatial climate variability in the Gilgel Gibe watershed, this study integrated station data with gridded datasets (CHIRPS and NASA) and applied spatial interpolation techniques to estimate climatic variables across the watershed, thereby minimizing the biases associated with station density. To generate a continuous spatial layer of annual average precipitation, this study utilized ordinary Kriging interpolation in ArcGIS. Kriging was selected to analyze rainfall changes due to its advanced spatial interpolation capabilities and ability to account for precipitation data variability and spatial autocorrelation. This method provides more accurate rainfall estimations, particularly in areas with complex topographies and diverse microclimates across the Gilgel Gibe watershed. In this study, the effectiveness of the Kriging, multivariate linear interpolation (MvLI), and elevation-based interpolation (SbLI) techniques in generating grid-based rainfall datasets for the Gilgel Gibe watershed was evaluated. Based on this assumption, ordinary Kriging was deemed to be the most suitable technique for this study. Figure 2 was created using CHIRPS satellite imagery and DEM data obtained from USGS Earth Explorer (https://earthexplorer.usgs.gov) and NOAA, processed and analyzed using ArcGIS 10.8

Hydrogeological data

This study extensively examined the water yield of the Gilgel Gibe Watershed by integrating hydrogeological, remote sensing, and climate data. These data are essential for understanding geological properties, groundwater flow patterns, and aquifer characteristics that significantly influence surface water availability and hydrological processes[44]. It also accounts for the buffering effect of groundwater on surface water fluctuations caused by climate change and LULC dynamics[45]. This integrated method enables a holistic evaluation of future water availability by considering surface and subsurface water resources[46]. In addition, it enhances the predictive accuracy of machine learning and remote sensing models by providing a contextual understanding of the interactions between climate variables, LULC changes, and surface water yield. In the water yield models, essential variables such as precipitation, vegetation, runoff coefficients, and evapotranspiration were calibrated to maximize the prediction accuracy. Sensitivity analysis was used to determine the most influential parameters, while observed hydrological data over time enhanced the model performance across different climatic conditions and flow regimes. Previous studies have emphasized the importance of incorporating hydrogeological information into watershed-scale hydrological modeling, emphasizing its role in improving model reliability and robustness[47]. By combining hydrogeological data with advanced analytical techniques, this study intends to improve our understanding of the complex relationship between environmental factors shaping water resource dynamics in the Gilgel Gibe watershed (Fig. 3).

Climate trend analysis

Previous research has studied temperature and precipitation patterns in Ethiopia at several geographical and temporal scales. This study improves watershed-level analysis by focusing specifically on precipitation, temperature, and potential evapotranspiration (PET) to examine their effects on surface water recharge. Climate models and trend analysis tests, along with WorldClim data, were utilized to identify trends in time-series climatic variables. The Mann–Kendall test and Sen's slope estimator were used to detect and quantify trends in annual rainfall, maximum and minimum temperatures, average yearly temperature, and potential evapotranspiration (PET) at a 95% confidence level[48,49]. These methods, combined with data from the WorldClim database, ensured a robust and comprehensive analysis of the climate trends in the Gilgel Gibe watershed. Based on daily climate data from 1993 to 2023, consistent annual trends were observed for solar radiation, maximum and minimum temperatures, and wind speed, indicating stability over this period. In particular, precipitation showed a significant anomaly in 2000, deviating from typical patterns (Fig. 4a). The analysis of yearly variations (Fig. 4b)

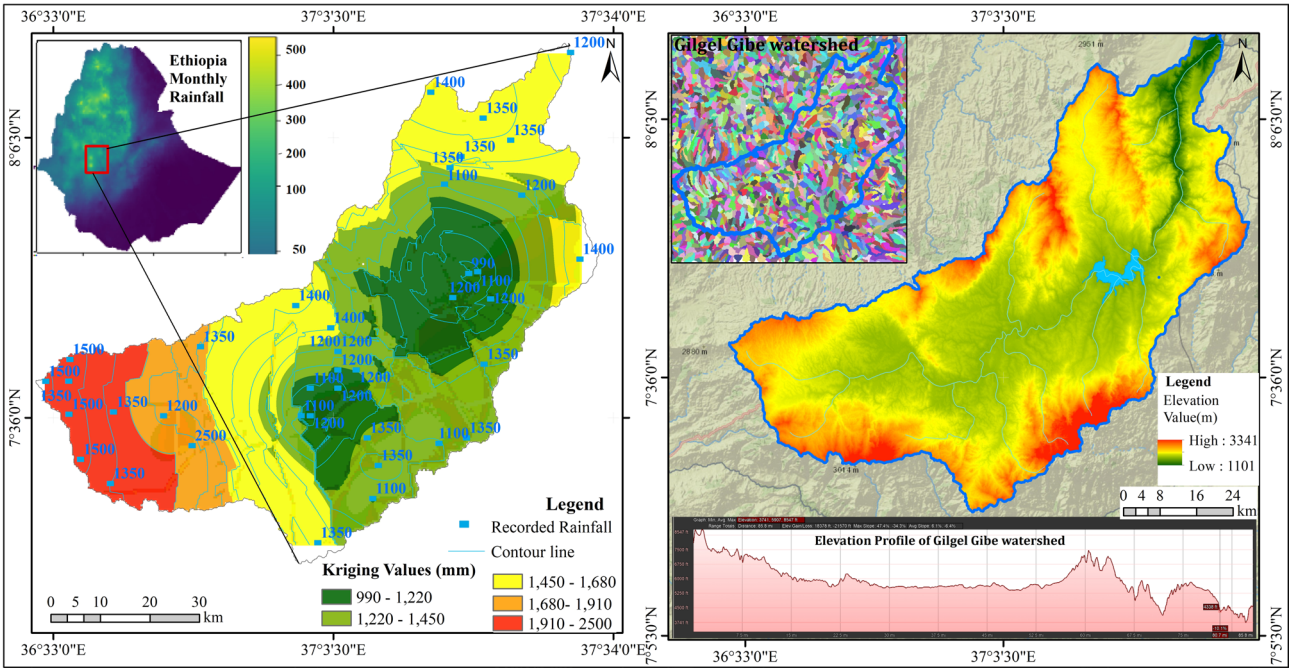

Fig. 2. Spatially interpolated rainfall distribution in the gilgel gibe watershed.Source: The maps were created using ArcGIS software (version 10.8).

highlighted distinct trends across all variables, reflecting the dynamic nature of these climatic factors influenced by natural variability and anthropogenic impacts. Such variability emphasizes the importance of continuous monitoring and analysis for a better understanding of long-term climatic patterns and their impact on Gilgel Gibe's water resources and to inform water yield management and planning activities. Analysis of climatic data trends for the Gilgel Gibe Basin revealed prominent variability in climate variables, reflecting significant climatic changes that affect water yield. Changing precipitation and temperature patterns directly influence the basin's water yield, with changes in precipitation leading to periods of excessive runoff or reduced water availability, thereby altering the water balance.

Methods

This study used advanced remote sensing techniques and water yield modeling to assess the integrated impacts of climate and LULC changes on surface water yield in the Gilgel Gibe Watershed. These methodologies are systematically categorized according to their structural composition and predictive approaches, creating a comprehensive framework for assessing environmental changes[50,51]. By integrating various datasets and sophisticated modeling techniques, this study aimed to elucidate the complex interactions among climate variables, LULC changes, and their impacts on regional water resources (Fig. 5). The study used RF, a robust ensemble learning algorithm that enhances prediction accuracy by aggregating outputs from multiple decision trees trained on random data subsets[52–55]. Additionally, an SVM is used to model the complex relationships between climate factors, land-use patterns, and surface water yield, leveraging historical data to forecast future trends while accommodating additional spatial and temporal variables[56,57]. Furthermore, this study encompasses XGBoost and LightGBM, both of which are recognized for their efficiency and accuracy in processing large datasets. XGBoost iteratively constructs decision trees to refine predictions, whereas LightGBM applies novel strategies to enhance the training efficiency[58–60]. This provides a realistic and robust assessment of the models'performance on time-dependent datasets. The models were executed on an MITRO V15 device equipped with a 13 th Gen Intel(R) Core (TM) i7-13700H processor operating at 2400 MHz, with 10 cores and 16 logical processors. The implementation utilized Scikit-Learn and the Keras API within TensorFlow (version 2.10.0), offering a powerful and optimized framework for model training and testing. Through the integration of these advanced algorithms, this study aims to develop robust predictive models that illustrate the complex interaction between climate variables, land-use dynamics, and surface water yield, thereby providing essential insights for sustainable water management (Fig. 5).

Hyperparameter optimization for ML models

In this study, hyperparameter optimization was vital for enhancing the accuracy and generalization of machine-learning models for predicting water yield under diverse climate and land-use scenarios. The hyperparameters of the four machine learning models, Random Forest Regressor, XGB Regressor, LGBM Regressor, and SVR were optimized using BayesianSearchCV with threefold cross-validation. For the Random Forest Regressor, the number of trees (n_estimator) was tuned between 50 and 500, while the maximum tree depth (max_depth) was adjusted between 5 and 20. These settings were evaluated over 50 iterations using threefold cross-validation.

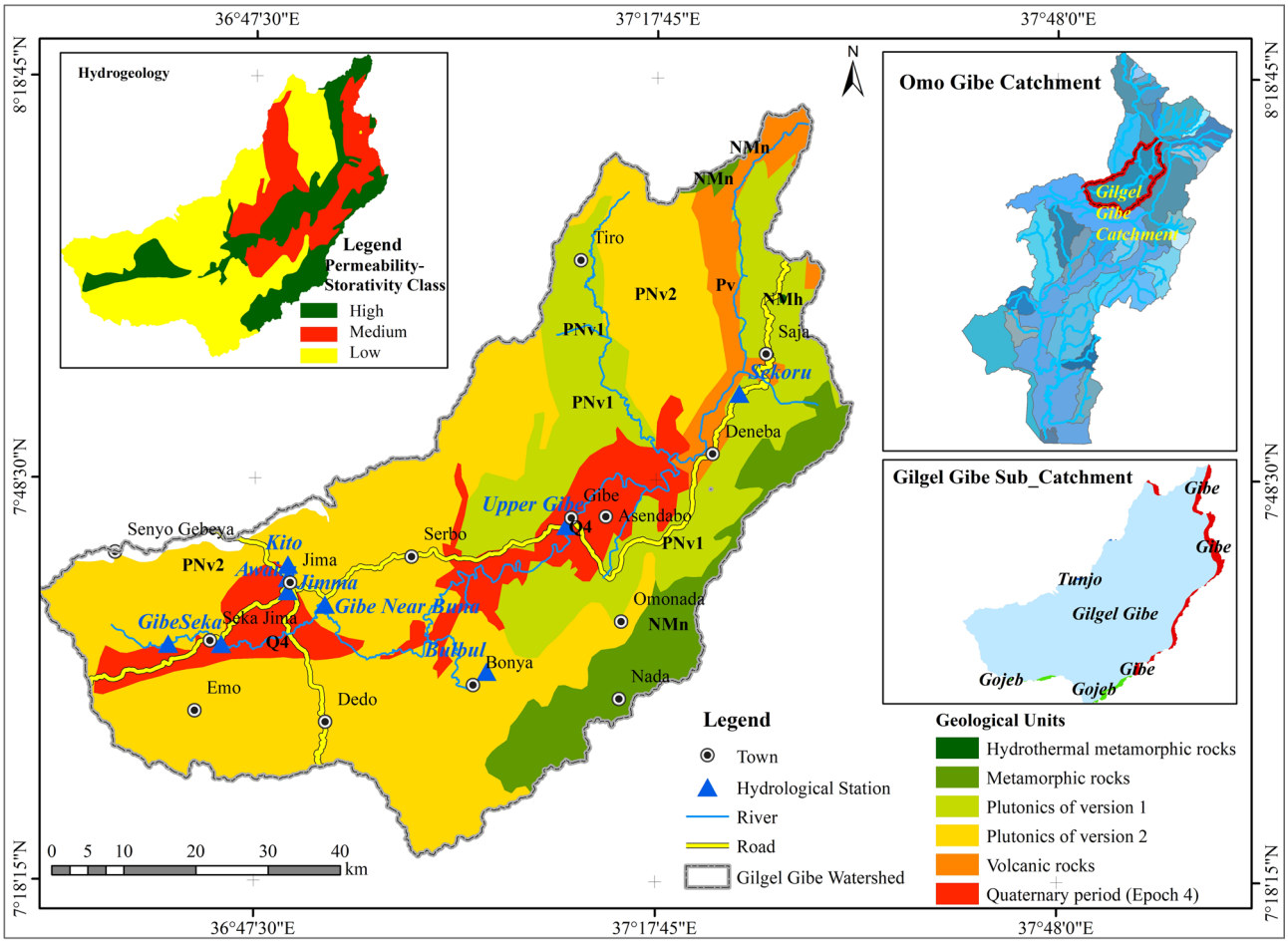

Fig. 3. Hydrogeological maps of the gilgel gibe catchment.Source: The maps were created using ArcGIS software (version 10.8).

The XGB Regressor was optimized for n_estimator, max_depth, and the learning rate (ranging from 0.01 to 0.1), also employing 50 iterations and threefold cross-validation. Similarly, the LGBM Regressor's n_estimator, max_depth, and learning rate (ranging from 0.01 to 0.1) were tuned using the same procedure. For the SVR, the regularization parameter (C) was adjusted between 0.1 and 1000, the epsilon value ranged from 0.01 to 0.1, and the kernel was selected from options'linear'and'rbf.'This systematic optimization ensured that each model was finely tuned for high predictive accuracy while mitigating overfitting. For time-series data, walk-forward cross-validation was implemented to evaluate models sequentially over yearly increments. This approach incrementally expands the training dataset with each iteration, testing the model on the next time period. Such a method simulates real-world prediction scenarios and ensures that predictions are based only on past data, avoiding data leakage. This provides a realistic and robust assessment of the models'performance on time-dependent datasets. Models were trained and tested using a rolling-forward approach. The initial training period was 1993–2016, with testing performed in 2017. This procedure was repeated iteratively, with training periods of 1993–2017 (tested on 2018), 1993–2018 (tested on 2019), 1993–2019 (tested on 2020), and 1993–2020 (tested on 2021). Independent cross-validation was then conducted using climate data (1993–2017 training, 2018–2022 validation) and hydrological data (1995–2016 training, 2017–2021 validation). To avoid overfitting and assess the generalization capabilities of each model, an 85/15 data split was employed, where 85% of the data were used for training and 15% for testing. Cross-validation using a k-fold approach was applied, which aided in preventing overfitting and ensured the robustness of each model. Parameter tuning was conducted using BayesianSearchCV to optimize hyperparameters specific to each model: depth, minimum samples split, and estimators for Random Forest; kernel type and regularization for SVM; learning rate, maximum depth, and estimators for XGBoost and LightGBM. Validation metrics included mean absolute error (MAE), root mean square error (RMSE), Nash–Sutcliffe efficiency (NSE), Willmott's Index of Agreement (d), and R-squared (R^2) values to comprehensively evaluate the model performance.

InVEST water yield model

The InVEST model was used to determine the water yield in the Gilgel Gibe watershed. This model operates on a spatially explicit framework utilizing the cell size of the Digital Elevation Model (DEM) to assess water yield dynamics. The InVEST Water Yield model was selected based on its capacity to estimate water yields

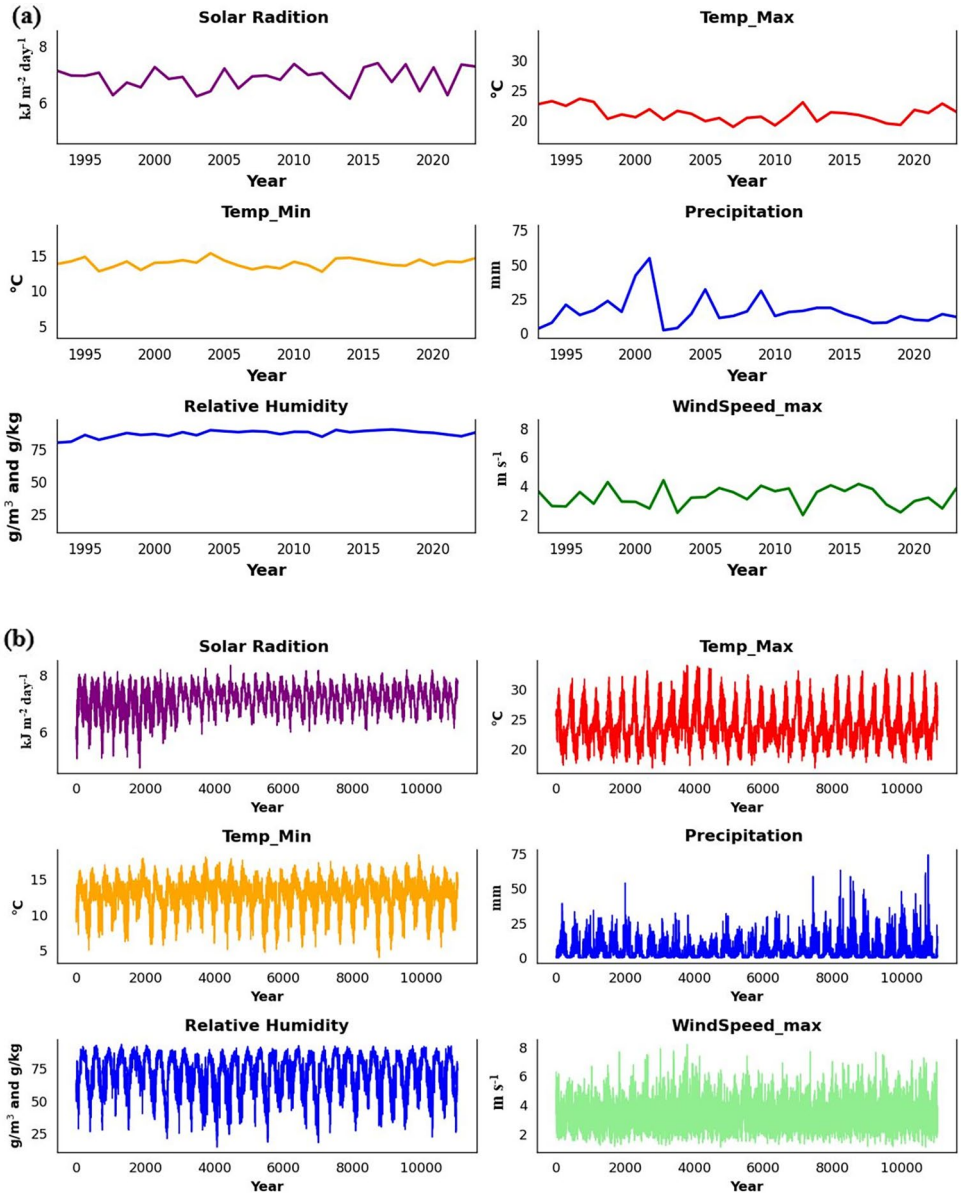

Fig. 4. **a** Daily time series of key climate variables in the gilgel gibe watershed. **b** Plots of daily climate data points over years.

across 10 representative watersheds within the Gilgel Gibe watershed, as documented in previous studies[61,62]. The InVEST model was used for water yield simulations in this study because of its distinctive advantages in simulating ecosystem services in complex hydrological settings[63–65]. The InVEST model is widely used for assessing ecosystem services; however, it has significant limitations, particularly the assumption of uniform soil properties over large geographic areas. This generalization can lead to inaccuracies when modeling hydrological processes in diverse landscapes, such as the Gilgel Gibe watershed, where soil characteristics, such as texture, permeability, organic matter content, and water-holding capacity, vary significantly. To address these limitations, this study integrates the InVEST model with high-resolution, localized soil datasets collected by the Ethiopian Agricultural Transformation Institute (ATI).

Additionally, complementary tools, including hybrid machine learning models, are employed to enhance the predictive accuracy in regions with insufficient data. These models combine physical principles with data-driven techniques. The integration of remote sensing, GIS, and advanced climate models further improves the capacity of the model to evaluate the impacts of LULC changes on runoff, evapotranspiration, and groundwater recharge. These methods aim to enhance the accuracy and reliability of hydrological modeling, providing a comprehensive understanding of the interactions between water, land, and climate. This model, which is widely used for ecosystem service valuation and land-use planning, requires calibration using various input parameters, some of which are either unknown or subject to significant variability. To manage uncertainty during calibration,

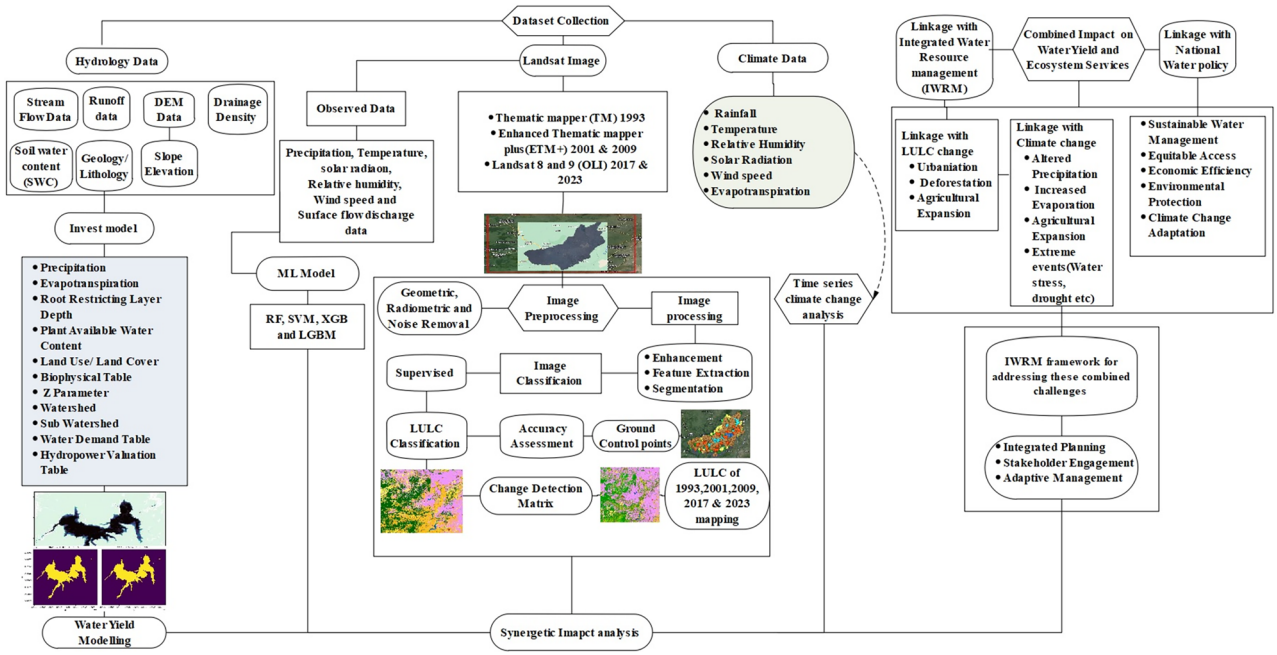

Fig. 5. Methodological flowchart of the study.

sensitivity analysis was employed to evaluate the influence of key input variables, such as LULC, slope, and precipitation, on the model results.

Furthermore, uncertainty analysis and error propagation were used to quantify the overall uncertainty in model predictions by accounting for the variability in each input and evaluating the cumulative effect on the model's results. To further validate the model, it was calibrated against independent data sources by comparing model predictions with independent datasets derived from field observations and remote sensing products. InVEST is suitable for integrating land use and climatic factors into hydrological outputs, which aligns with perfectly the study's objective of understanding how LULC changes and climate variability influence water yield. The model used remote sensing and GIS data to generate quantitative water yield outputs at the watershed scale, which were then compared with the observed runoff data for calibration and validation. Based on the Budyko curve and the average annual precipitation, the annual water yield (WY) for each location (pixel) was calculated using Eq. 1, as outlined by[66,67].

$$Y(x) = \left(1 - \frac{rAET(x)}{P(x)}\right) * P(x) \tag{1}$$

where AET(x) and P(x) represent annual actual evapotranspiration (mm) and precipitation (mm) at pixel x, respectively. This study employs the Budyko equation to estimate actual evapotranspiration (AET) at each pixel or location within the landscape Eq. 2[68–70].

$$AET(x)/\ P(x) = 1 + (PET(x)/\ P(x)) - \{1 + [(PET(x)/\ P(x)) * \omega]\}\hat{}(1/\omega) \tag{2}$$

where AET(x) represents the annual actual evapotranspiration (mm) and P(x) is the precipitation (mm) at pixel x. Additionally, a non-physical parameter (ω) Is incorporated to account for the partitioning of precipitation between runoff and evapotranspiration. Watersheds with elevated ω values demonstrated an increased ability to transform precipitation into evapotranspiration, whereas those with lower values exhibited a reduced capacity for this process.

$$PET(x) = Kc(lx) * ET0(x) \tag{3}$$

The PET for each pixel was calculated using Eq. 3. This equation incorporates a vegetation coefficient (Kc(lx)) that varies based on LULC type and ranges from 0 to 1.5. In addition, the reference evapotranspiration (ET0(x)) was included, which was computed using a modified version of[71,72] Azzam et al. (2023) and Hargreaves and Samani (1985).

$$ET0 = 0.0013 \times 0.408 \times RA \times (Tav + 17) \times (TD - 0.0123P)\hat{}0.76 \tag{4}$$

In Eq. 4, RA represents solar radiation, Tav is the average of the daily maximum and minimum temperatures, TD is the temperature difference between the daily maximum and minimum, and P is the monthly precipitation (Li et al., 2022).

$$\omega\,(x) = \mathrm{ZAWC}(x)/\,\mathrm{P}(x) + 1.25 \quad (5)$$

The ω parameter is an experimental value requiring further investigation for site-specific calibration. The InVEST model uses a computational formula for ω (Eq. 5), based on global data[73–75]. The Z parameter describes the spatial distribution of precipitation events. Areas with more precipitation events (for the same total annual precipitation) typically have higher Z values[76]. Budyko's theory states that a higher Z value reduces the sensitivity of the model to this parameter[77]. Table 3 presents an in-depth evaluation of the water yield in the Gilgel Gibe Watershed, emphasizing the effects of LULC variations and climate variability.

The sensitivity ranges for various input data reflect the inherent uncertainties and variability. LULC classification values (± 20–30%) account for misclassification and changes over time, whereas precipitation data (± 10%) capture uncertainties from satellite and ground-based measurements. Temperature data (± 1 °C to ± 2 °C) allow for minor fluctuations, and watershed and sub-watershed boundaries depend on the vector data and DEM resolution. The evapotranspiration range (± 10%) reflects climatic variability and soil moisture measurements (± 10%) address spatial and temporal data uncertainties. The ± 10% Z factor was selected to address potential variability in slope steepness and account for uncertainties in terrain data, which limitations may influence in the resolution or quality of topographic datasets. By applying a ± 10% range, the calibration process ensures that the model accurately depicts the impact of terrain on hydrological outcomes without overfitting localized data that may not capture broader trends. The Kc factor (± 10%) accounts for crop-specific variations, and the discharge data may have uncertainties owing to measurement errors or gaps. These ranges ensure that the model calibration accommodates data variability.

Spatially explicit modeling of water yield

This study employed a spatially explicit model to assess water yield dynamics within the study area. The model operates on a gridded map where each pixel represents a specific location. This method allows the integration of spatial heterogeneity in key water yield drivers such as soil type, precipitation, and vegetation type[78,79]. The model calculates the average annual water yield (Yxj) for each grid cell (usually 300 m x 300 m) using factors such as the average annual precipitation (Px), annual reference evapotranspiration, soil depth, plant available water content (PAWC), root depth, and land-use characteristics. This equation-based calculation enabled the estimation of the water yield across the entire study area, accounting for spatial variations in these driving factors. The modeling process was complicated by uncertainties in the high-resolution input data, including LULC classification variability, which was addressed through quality control and cross-validation. The uncertainty in the precipitation data was decreased by combining satellite observations with ground-based measurements and spatial interpolation, whereas temperature variations were handled by integrating historical and satellite-derived data. The elevation data resolution was improved for a more precise watershed boundary delineation, and evapotranspiration and soil moisture uncertainties were minimized using multi-source data and interpolation. Seasonal adjustments and literature-based estimates were used to account for the Kc factor's variability. Finally, discharge data gaps and seasonal variations were addressed by cross-referencing with the model predictions, ensuring that the model effectively handled the uncertainties in each variable to provide reliable results.

Climate and LULC change scenarios

Scenario analysis is a powerful tool used to forecast future outcomes based on projected conditions, assuming that current observed phenomena or trends continue. Understanding the association between climate change and LULC change is essential for predicting and managing water resources. The water yield of the Gilgel Gibe Watershed was estimated using the InVEST water yield module for 1993, 2001, 2009, 2017, and 2023. Three distinct scenarios were developed using scenario analysis methodologies to examine the effects of climate change and LULC on water yield. The scenarios considered a range of variables, including probable temperature changes, precipitation patterns, and LULC change trajectories. For instance, scenarios might include variations in urban expansion, deforestation rates, and agricultural practices, along with different climate projections, such as increased rainfall variability or higher average temperatures. Combining LULC changes with climate change projections provides a more precise and holistic understanding of the water resource dynamics, as noted by[80].

Data	Format	Sources	Range of sensitivity
LULC	Raster	United States Geological Survey (USGS)	± 20–30%
Precipitation	Gridded raster/Numeric	Ethiopian Metrological Institute (EMI) https://data.chc.ucsb.edu/products/CHIRPS-2.0/	± 10%
Temperature	Gridded raster/Numeric	Ethiopian Metrological Institute (EMI)	± 1 °C to ± 2 °C
Watershed	Vector	Ethio_GIS (https://www.ethiogis-mapserver.org/)	n/a
Sub-watershed	Vector	Digital Elevation Model (DEM)	n/a
Potential evapotranspiration	Raster	Hargreaves and Allen (2003) and google erath engine (https://code.earthengine.google.com/)	± 10%
Soil moisture	Raster	Google earth engine	± 10%
Z Factor	Constant	Sharp et al. (2018)	± 10%
Kc Factor	Numeric	Allen et al. (1998)	± 10%
Discharge data	Numeric	Ministry of Water and Energy (MoWE)	n/a

Table 3. Input data for the InVEST water yield model.

Similarly[81], emphasized the necessity of scenario analysis to anticipate and mitigate potential adverse effects on water yield and ecosystem services. Finally, the scenario analysis of climate and LULC change in the Gilgel Gibe watershed underlines the potential impacts on water yield and the significance of integrated management methods that encompass land-use policies and climate adaptation measures.

Evapotranspiration data

This study used Google Earth Engine satellite data, including the Global Land Data Assimilation System (GLDAS), Moderate Resolution Imaging Spectroradiometer (MODIS), and Penman–Monteith, to assess the impact of evapotranspiration on the water yield of the Gilgel Gibe Watershed. The FAO-56 Penman–Monteith (PM) equation is widely recognized as the benchmark method for estimating reference evapotranspiration (ET_0), ensuring consistency and comparability across diverse geographic regions. MODIS and GLDAS datasets were selected for their extensive spatial coverage, high temporal resolution, and suitability for the hydrological processes of the study area. MODIS provides vital data on vegetation indices, land cover, and temperature, essential for understanding evapotranspiration (ET) dynamics (Heck et al., 2019). The daily evapotranspiration data series were obtained by temporally interpolating values adjacent to MODIS 16-day Terra climate products, yielding data that corresponded to satellite overpass times (Singh et al., 2024). Various datasets were used for the water yield analysis, including evapotranspiration (ET), Latent Heat Flux (LE), Potential ET (PET), Potential LE (PLE), Latent Heat layers (LE and PLE), and annual quality control (ET_QC). The resolution of the MODIS products, ranging from 1 km to 500 m, was selected as the final resolution for ET calculations[82]. The GLDAS air temperature products were downscaled to a 1 km resolution using a series of steps: (a) For each GLDAS grid (0.25° ×0.25°), the air temperature was considered as the temperature at the grid's average elevation. (b) The sea level temperature data were then resized to a resolution of 1 km using bilinear interpolation. (c) The air temperature at the original elevation was calculated from the resized data, effectively reversing step (a).

In addition, GLDAS offers key meteorological variables such as precipitation, temperature, and solar radiation, which are ideal for modeling water balance and ET[83]. GLDAS data are particularly valuable in areas with limited ground-based measurements. The daily ET estimates are derived through a process that includes (1) data assimilation, (2) model calibration, (3) potential evapotranspiration (PET) calculation using atmospheric variables, and (4) conversion of PET to actual ET by considering soil moisture and vegetation. These steps ensured that the GLDAS provided accurate and reliable ET estimates for hydrological assessments (Table 4). The Penman–Monteith equation is a widely recognized method for estimating evapotranspiration (ET), which refers to the water transfer from the land surface to the atmosphere[84]. This process occurs through evaporation from the soil and water bodies, as well as transpiration from vegetation. However, its accuracy is highly dependent on the availability of high-quality meteorological data, including solar radiation, wind speed, temperature, and humidity. In data-scarce regions, the lack of reliable meteorological inputs can introduce significant errors, leading to misclassifications in ET estimations. To enhance the accuracy of this study, we incorporated high-resolution spatial datasets along with ground-based meteorological observations. This integration provides a more robust assessment of ET_0 variability and mitigates the uncertainties associated with missing or low-quality data. However, the PM equation is computationally demanding and requires specialized meteorological stations and remote sensing datasets for precise parameter estimation. Given these challenges, the application of the Penman–Monteith equation necessitates careful data handling and methodological refinement to ensure reliability. An accurate estimation of ET_0 depends on precise incident solar radiation (SRD) measurements, either directly obtained from meteorological stations or derived through gap-filling techniques[85]. The selection of the PM method in this study was validated by comparing its ET_0 estimates with MODIS-derived ET data and ground-based observations, thereby confirming its suitability for assessing regional evapotranspiration patterns. The FAO Penman–Monteith method for calculating reference evapotranspiration (ETo) is expressed as follows (Eq. 6):

$$ET = \frac{0.408\Delta\,(Rn - G) + \frac{\gamma 900}{T+273} u2(es - ea)}{\Delta + \gamma(1 + 0.34u2)} \tag{6}$$

where ET is evapotranspiration (mm/day), Δ is the slope of the saturation vapor pressure curve (kPa/°C), Rn is the net radiation (MJ/m^2/day), G is the soil heat flux density (MJ/m^2/day), γ is the psychrometric constant (kPa/°C), T is the mean air temperature (°C), u_2 is the wind speed at 2 m height (m/s), es is the saturation vapor pressure (kPa), and ea is the actual vapor pressure (kPa).

Model scenario settings

This study developed detailed and immersive simulated scenarios, advancing beyond the mere comparisons found in the literature. Previous studies on ecosystem service simulations typically created two to four development

ID	Parameters	Spatial resolution	Temporal resolution
1	GLDAS	0.25°* 0.25°	1 day
2	MODIS	1 km	16-day composite
3	MODIS	500 m	8-day composite
4	Penman–monteith	0.1°*0.1°	1 month

Table 4. Satellite data used in ET calculations.

scenarios, such as economic priority, ecological protection, and sustainable development, by limiting land-use expansion[86–88]. In contrast, this study adopted a novel approach that integrates historical climate and land-use data to create diverse scenarios that assess the relative impacts of climate change and LULC change on water yield services. Table 5 outlines the three main types of scenarios explored in this study. Initially, the actual water yield conditions in the Gilgel Gibe River Basin were computed using the InVEST model's water yield module for 1993, 2009, 2019, and 2023. In the climate change scenarios, LULC change was held constant to isolate the effects of climate variation on ecosystem water services. This study segmented the period from 1993 to 2023 into five intervals (1993–2009, 2009–2023, 2009–2019, 2019–2023, and 1993–2023) corresponding to scenarios 1 to 3. For instance, scenario 1 utilized climate data from 1993 (multi-year average from 1996 to 2006), whereas the LULC data remained unchanged from 1993. This allowed for the assessment of climate change impacts on water yield compared to the actual conditions in 1993. Similarly, the LULC change scenarios kept climate elements constant and divided the study period into scenarios 2–3 to assess the impact of LULC changes on regional water yield over three time periods. Comparing these ten sub-scenarios with actual conditions revealed the specific contributions of climate change and LULC change to water yield across different phases.

Accuracy assessment

Assessing accuracy is crucial for determining how precisely maps can be used for accurate and effective data usage. Inadequate precision levels often indicate poor satellite data classification for the LULC categories[89,90]. The classification accuracy was assessed using precision metrics derived from the error matrices. An error matrix was created using the validation points and classified data. The Kappa coefficient (KC) was used to evaluate classification accuracy. Per-class evaluation improves the Kappa coefficient by resolving the accuracy limitation caused by the amount of correctly identified cases in each class, often known as user accuracy[91]. In LULC classification, where some classes may be under or overrepresented, the Kappa coefficient is essential for managing class imbalance and introducing bias into accuracy assessments[92]. It corrects for such imbalances by considering classification consistency across all classes rather than just overall accuracy, which is especially essential for rare or fragmented land classes that may distort conventional measures[93]. A confusion matrix was used to examine metrics such as overall accuracy (OA), producer accuracy (PA), and consumer accuracy (CA)[94]. Using a random sampling process, 700, 936, 679, 800, and 750 Ground Control Points (GCP) were collected for 1993, 2001, 2009, 2017, and 2023, respectively. The Kappa coefficient was calculated using the following formula:

$$\mathrm{K}\hat{} = \frac{\mathrm{n}\sum_{\mathrm{i}=1}^{\mathrm{r}} \mathrm{xii} - \sum_{\mathrm{i}=1}^{\mathrm{r}} (\mathrm{xi} + . \ \mathrm{x} + \mathrm{i})}{\mathrm{N}^2 - \sum_{\mathrm{i}=1}^{\mathrm{r}} (\mathrm{xi} + . \ \mathrm{x} + \mathrm{i})} \tag{7}$$

In the error matrix, $K\hat{}$ represents the kappa coefficient, r denotes the number of rows, Xii indicates the number of observations on the main diagonal, $Xi+$ is the total number of observations in row i, $X+i$ is the total number of observations in column i, and N is the total number of observations within the matrix.

Change detection

Change detection in LULC is crucial for understanding the ecosystem of a watershed and its impact on water yield. Change detection matrices, which provide a structured approach for quantifying and interpreting LULC changes over time, were used in this process[95]. The Gilgel Gibe watershed in Ethiopia has experienced significant LULC changes over the years due to factors such as agricultural expansion, urbanization, unsustainable water use, and climate change[96,97]. This study used a change detection matrix based on remote sensing data and geospatial analysis to analyze LULC changes and their impact on water yield. The change detection matrix identifies transitions between different land cover types, allowing for a detailed understanding of the landscape dynamics[98,99]. The following equations were used to quantify changes in LULC:

$$\text{Loss} = \text{Row total} - \text{diagonals of each class unchanged} \tag{8}$$

$$\text{Net change} = \text{Gain} - \text{Loss} \tag{9}$$

$$\text{Gain} = \text{Column total} - \text{diagonals of each class(unchanged)} \tag{10}$$

$$\text{Net persistence(NP)} = \frac{\text{Net change}}{\text{diagonals of each class(unchanged)}} \tag{11}$$

Model calibration and validation

The model was calibrated using data from January 1, 1993, to December 31, 2023. The prediction accuracy of the model was validated by segmenting the study period. Calibration data were used from 1993 to 2018, whereas

Actual scenarios	Climate change scenarios	Land use type change scenarios	Hydrological discharge scenarios
	1993	**2009**	**2023**
Climate change	1993	2009	2023
Land use type	1993	2009	2023
Hydrological flow	1995	2019	2023

Table 5. Scenarios for climate change, LULC change, and surface flow.

validation data were set aside from 2019 to 2023. This division enabled us to assess the model's capacity to generalize novel, previously unseen data. This involved minimizing the differences between the simulated and observed streamflow data at the three stream gauges in the model area. Calibration was accomplished by fine-tuning the essential parameters to eliminate variances in various streamflow characteristics, such as volume error, mean flow, minimum flow, maximum flow, runoff, and seasonal volume (summer and winter). The performance of the InVEST model is determined by parameter Z, which is an empirical constant representing regional precipitation patterns and hydrogeological properties[100,101]. The best Z value is often established by comparing the simulated water yield of the model to the observed runoff data[102]. This study evaluates the accuracy of the InVEST model by analyzing runoff data from three hydrological stations in the Gilgel Gibe Basin: Gilgel Gibe, Bulbul, and Bidru Awana, from 1995 to 2023. The Ethiopian Ministry of Water and Energy published data obtained from the monitoring systems at each station. Five statistical metrics were used to evaluate model performance: Nash–Sutcliffe Efficiency (NSE), Willmott's Index of Agreement (d), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2)[103–106]. The NSE index, which ranges from negative infinity to one, assesses the prediction accuracy; a value of one indicates a perfect match between the observed and simulated data. The NSE can be defined mathematically as

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n} \left(yi - \widehat{yi}\right)^2}{\sum_{i=1}^{n} \left(yi - \widehat{yi}\right)^2} \tag{12}$$

Willmott's Index of Agreement (d) measures the alignment between model predictions and observed data, ranging from 0 to 1, where 1 signifies perfect agreement. Higher numbers imply better model performance. The calculation is performed using the following formula, where n represents the number of observations, yi denotes the actual value of the i-th observation, and ŷ i indicates the predicted value for the i-th observation:

$$\mathrm{d} = 1 - \frac{\sum_{i=1}^{n} \left(yi - \widehat{yi}\right)^2}{\sum_{i=1}^{n} \left(yi - \widehat{yi}\right) + \left(yi - \widehat{yi}\right)^2} \tag{13}$$

where yi represents the observed values, $\widehat{yi}$ denotes the simulated values, and $\widehat{yi}$ bar $|\ \widehat{yi}\ |$ Is the mean of the observed values. The Root Mean Square Error (RMSE) measures the average magnitude of errors between expected and observed values. This method calculates the standard deviation of the residuals, which reflects the degree of divergence from the true values. The formula for RMSE is given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}} \sum_{i=1}^{n} \left(yi - \widehat{yi}\right)^2} \tag{14}$$

where: yi represents the observed values, $\widehat{y}$ i Denotes the predicted values and n is the number of observations. Mean Absolute Error (MAE) measures the average size of mistakes in a dataset by computing the mean of the absolute differences between observed and anticipated values. This metric measures the average magnitude of the residuals and provides insights into the correctness of the model. The MAE was calculated using the following formula:

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\ yi - yi\ | \tag{15}$$

The Coefficient of Determination (R^2) indicates the variance in the dependent variable that can be explained by a linear regression model. It is scale-independent, meaning that it is not affected by the size of the dependent variable. R^2 ranges from 0 to 1, where a value of 1 indicates a perfect fit.

$$\mathrm{R}^2 = 1 - \frac{\sum \left(\mathrm{yi} - \widehat{\mathrm{yi}}\right)^2}{\sum \left(\mathrm{yi} - \widehat{\mathrm{yi}}\right)^2} \tag{16}$$

Linear regression analysis was conducted to evaluate the accuracy of the simulated water yields from the InVEST and ML models. The results shown in Fig. 6 reveal a strong correlation between the simulated and observed datasets. This strong correlation validates the efficacy of the climate and hydrological models in accurately simulating the water yield and is closely aligned with the observed runoff patterns.

Results

LULC change analysis

The results from our remote sensing classifier provide insightful predictions for LULC classes across five distinct periods: 1993, 2001, 2009, 2017, and 2023 (Table 6). These results underscored the dynamic nature of the various LULC categories over time. For instance, shrubland, forest, grassland, and wetland covers showed notable fluctuations. Shrubland decreased from 1108.37 km^2 (21.54%) in 1993 to 295.22 km^2 (5.74%) in 2023, reflecting a significant coverage area loss of 849.15 km^2 between these two periods. The extent of water bodies also changed by increasing from 12.51 km^2 (0.24%) in 1993 to 41.57 km^2 (0.81%) in 2023, largely due to the construction of

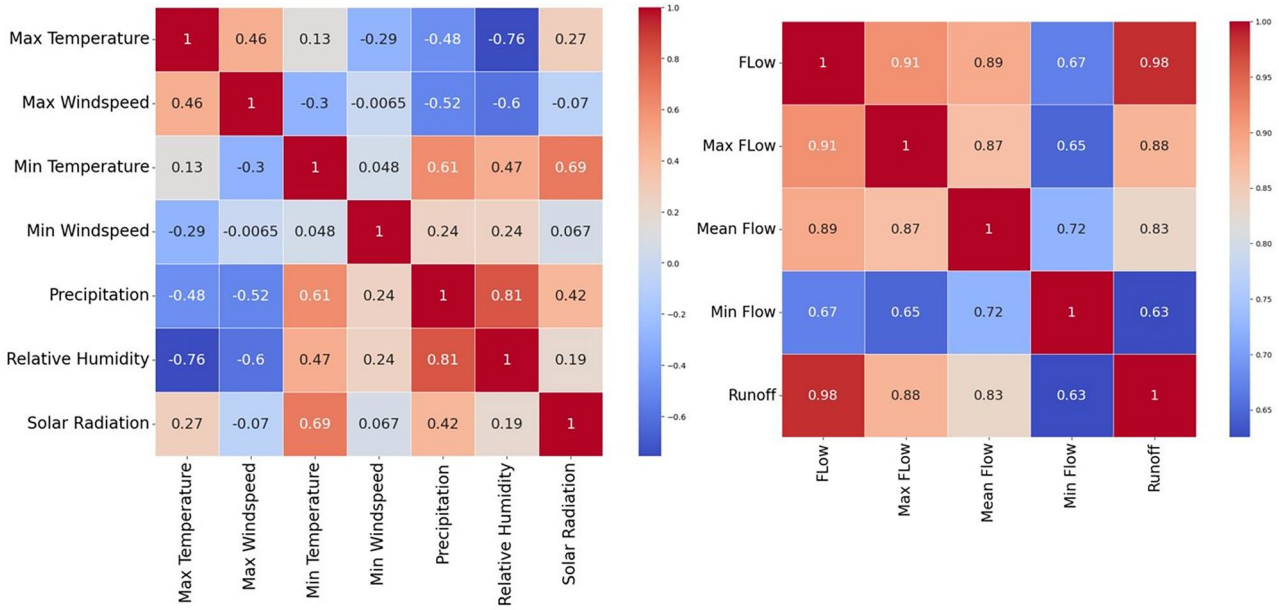

Fig.6. Correlation matrix for various features in the datasets.

ID	LULC class	Year 1993		Year 2001		Year 2009		Year 2017		Year 2023	
		Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)
1	Built-up area	31.02	0.60	85.54	1.66	149.92	2.91	181.53	3.53	150.78	2.93
2	Cultivated land	1969.58	38.28	2697.97	52.44	3310.10	64.34	3878.51	75.39	3636.17	70.68
3	Forest	626.73	12.18	1021.80	19.86	630.13	12.25	360.84	7.01	534.18	10.38
4	Grassland	941.67	18.30	104.40	2.03	249.15	4.84	94.79	1.84	184.63	3.59
5	Shrubland	1108.37	21.54	815.10	15.84	446.44	8.68	479.41	9.32	295.22	5.74
6	Waterbody	12.51	0.24	378.84	7.36	310.10	6.03	98.46	1.91	41.57	0.81
7	Wetland	454.98	8.84	41.19	0.80	49.02	0.95	51.31	1.00	302.35	5.88
Total		**5144.85**	**100.00**	**5144.84**	**100.00**	**5144.86**	**100.00**	**5144.85**	**100.00**	**5144.90**	**100.00**

Table 6. LULC changes from 1993 to 2023.

the Gilgel Gibe hydroelectric dam starting in 2003. This study showed that the Gilgel Gibe Dam has a significant impact on LULC, resulting in both direct and indirect environmental changes. Since its construction in 2003, the dam has altered local hydrology, comparing data from 1993, leading to changes in vegetation, soil properties, and land use in the surrounding areas. These changes are primarily driven by the creation of the reservoir, altered water availability for agriculture, and dynamic socioeconomic activities. The dam has caused land inundation, converting terrestrial areas into water bodies, which has destroyed natural habitats, such as forests, grasslands, and wetlands, leading to habitat loss and fragmentation.

In addition, it has contributed to the expansion of irrigated agriculture. The forested areas declined from 626.73 km^2 (12.18%) in 1993 to 534.18 km^2 (10.38%) in 2023. The main reason for forest conversion to agricultural land in the Gilgel Gibe watershed is the rapid population growth and increasing demand for arable land. Local households depend on agriculture as their primary source of livelihood, clear forests to meet their subsistence, and commercial farming needs, further exacerbated by limited access to alternative income sources. Government initiatives such as villagization and large-scale farming projects also contribute to agricultural land expansion. Additionally, the fertile soils in the watershed make them highly attractive for agriculture, leading to further encroachment on forests. To address the loss of ecosystems, stakeholders should develop strategies that balance agricultural development with forest conservation, and promote sustainable land use and ecosystem health. Similarly, grasslands and wetlands showed a dramatic reduction, with grasslands shrinking from 941.67 km^2 (18.30%) in 1993 to 184.63 km^2 (3.59%) in 2023 and wetlands decreased from 454.98 km^2 (8.84%) to 302.35 km^2 (5.88%) over the same period. Grasslands and wetlands are vital ecosystems that support biodiversity, regulate the climate, and provide essential ecological services. These changes can be attributed to complex interactions between natural and anthropogenic factors. Climate cycles and long-term trends may have contributed to the reductions in grassland, wetland, forest, and shrubland areas. Shrinking grasslands fragment habitats, reduce food sources, and breeding grounds for species such as herbivores, predators, and pollinators. Similarly, wetland loss disrupts critical breeding and migration routes for birds, fish, and amphibians, leading to

a population decline in the study area. Built-up areas have expanded consistently, driven by increasing demand for housing and infrastructure development. Urban areas expanded consistently from 31.02 km^2 (0.60%) in 1993 to 150.78 km^2 (2.93%) in 2023. This growth reflects the progression of urban development shaped by resource availability and prevailing environmental conditions. Conversely, from 2017 to 2023, the urban landscape of the study area has changed significantly owing to government policies targeting illegal housing. A demolition policy implemented in 2022 and continuing to 2023 reduced built-up areas by removing illegal houses and settlements. This effort addressed the unregulated urban expansion and promoted urban greenery. Cultivated land also showed a substantial increase, from 1969.58 km^2 (38.28%) in 1993 to 3636.17 km^2 (70.68%) in 2023. This unprecedented expansion was driven by population growth, improved agricultural practices, and favorable irrigation conditions. These results highlight the critical need to implement sustainable land management strategies that harmonize developmental goals with environmental conservation. Targeted policies and conservation efforts are essential to mitigate adverse effects and promote sustainable development in the Gilgel Gibe watershed (Fig. 7).

Generally, in the study area, land use changes, particularly agricultural expansion, significantly affect water yield. Clearing vegetation for crops reduces soil permeability and increases runoff, thereby disrupting water infiltration, groundwater recharge, and streamflow. For example, large-scale farming near the Gilgel Gibe Dam has led to higher runoff during rainfall. The conversion of wetlands and grasslands also exacerbates this by removing natural water regulation, leading to lower water yields in dry periods and increased flooding during rainy periods. To mitigate these effects, sustainable land management practices like agroforestry, water-efficient irrigation, and soil conservation should be prioritized.

Accuracy assessment

The accuracy assessment of LULC classification for 1993, 2001, 2009, 2017, and 2023 shows a significant and consistent improvement in predictive performance over time. The overall accuracy of the LULC classification progressively improved, reaching 94.94% in 1993, 95.95% in 2001, 95.44% in 2009, 97.22% in 2017, and 96.71% in 2023 (Table 7). Correspondingly, The Kappa coefficient, indicating the agreement between predicted and actual categories, ranged from 0.94 to 0.96 across the study period. This metric is highly useful in remote-sensing-based LULC classification, particularly when ground truth data are acquired on handheld GPS devices. These findings demonstrate the excellent precision of our LULC classification system, which exceeds the widely accepted accuracy threshold in land cover investigations. Particularly, in 2017, the classification performed admirably, with an overall accuracy of 97.22% and a Kappa value of 0.96, indicating a high agreement between predicted

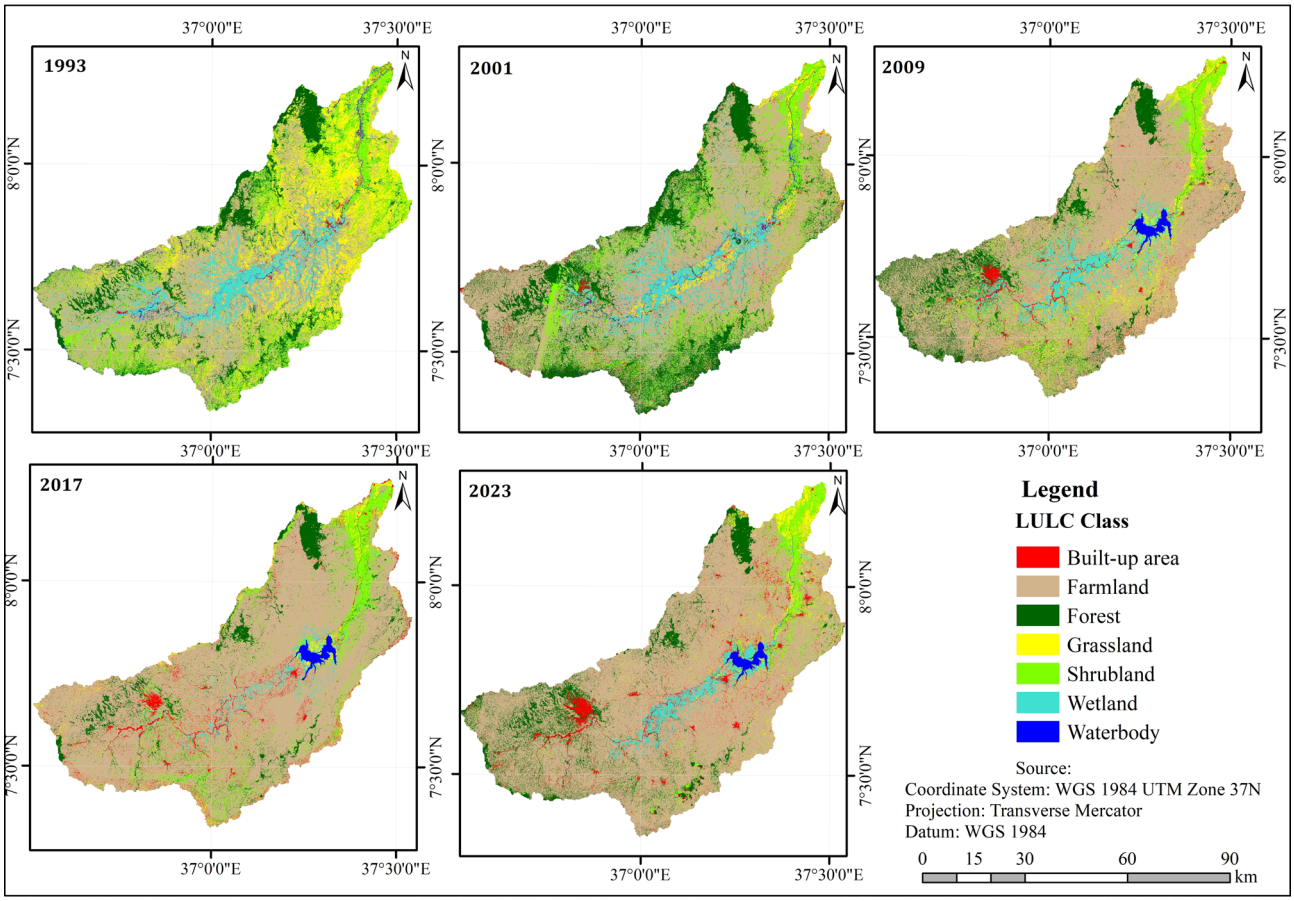

Fig. 7. LULC map of 1993 to 2023 *Source: The maps were created using ArcGIS software (version 10.8).*

and actual land cover types. In 2023, the kappa coefficient showed a slight decline to 0.9615 compared with previous years. This decrease can be attributed to rapid and fragmented LULC changes, including agricultural expansion and climate variability, within the study area. Agricultural expansion, characterized by the conversion of natural ecosystems into agricultural land, results in increasingly dynamic and unpredictable LULC patterns. This transition presents challenges in accurately classifying land cover, particularly when agricultural plots are small and dispersed or when there are shifts in crop types that may not be easily detected by classification models. The results of our classification method can be attributed to the use of advanced classification algorithms, high-resolution satellite imagery, and extensive field data. Finally, this study indicated that high-resolution satellite imagery combined with comprehensive ground-truth data can improve the precision of LULC mapping.

Change detection matrix

A change detection matrix was developed to analyze LULC changes in the Gilgel Gibe watershed between 1993 and 2023. The analysis revealed significant transitions, with over two-thirds (65.9% or 3390.14 km^2) of the watershed experiencing a change in LULC classes during this period (Table 8). The most significant change was from forest to agricultural land, accounting for 51.33% of the transformation, mainly in areas far from urban centers. Forests also transitioned into shrubland (114.55 km^2) and grassland (14.32 km^2), threatening biodiversity and ecosystem services Interestingly, about 5.98 km^2 and 321.55 km^2 of Cultivated land were converted to water bodies and forests, respectively, largely due to the construction of Gilgel Gibe I hydroelectric power dam and the reservoir created and government initiatives promoting private and community forests, particularly for coffee growers.The conversion of forests into agricultural lands significantly disrupts natural water cycles, worsening water scarcity. These changes affect the water yield by reducing groundwater recharge, increasing surface runoff, and contributing to soil erosion, ultimately leading to lower water quality and availability. For example, the absence of tree cover increases evaporation, soil moisture diminishes rapidly, and surface runoff intensifies, further reducing the ability of land to replenish groundwater recharge. Comparatively, more agricultural land transitioned to built-up areas (14.42 km^2), with minimal reverse transitions. These findings align with previous studies that highlighted the expansion of urban areas into natural landscapes, necessitating sustainable urban management and ecosystem-based adaptation strategies.

The results of the image difference and LULC change analysis for the Gilgel Gibe watershed from 1993 to 2023, as shown in Fig. 8, indicate substantial changes across various classes. The largest change was observed adversely, affecting 3484.62 km^2, which constitutes 67.73% of the total area. By contrast, there was an increase of 615.49 km^2 (11.96%). Additionally, there were minor decreases and increases, impacting 42.58 km^2 (0.83%) and 736.40 km^2 (14.31%), respectively. Notably, 266.04 km^2 (5.17%) of the study area remained unchanged. Overall, changes that accounted for a total area of 5145.14 km^2 indicated significant land cover dynamics in the watershed during the study period. This analysis highlights significant transitions in land cover types in the Gilgel Gibe watershed, emphasizing the need for sustainable land management practices. The observed dynamics, indicating that water yield changes are more attributable to loss mechanisms (68%), indicate significant implications of increasing cropland at the expense of decreasing green cover, such as shrubs and forest areas, on basin water yield. As cropland expands, it typically replaces areas of natural vegetation, which has substantial effects on the hydrological cycle. Reduced green cover diminishes the land's ability to absorb and retain rainfall, leading to increased surface runoff and decreased groundwater recharge. This change in land use can exacerbate the impacts of observed changes in rainfall, evapotranspiration, and temperature. For instance, increased cropland often correlates with higher levels of evapotranspiration due to intensified agricultural activities, potentially leading to reduced water availability in the basin. Concurrently, decreased green cover lowers the transpiration rates and disrupts local climate regulation, which can further alter temperature patterns and precipitation dynamics. This reduction in natural vegetation also diminishes the basin's capacity to stabilize water yields, making it more susceptible to extreme weather events and variability in rainfall. Overall, the LULC dynamics marked by increased cropland and decreased green cover can thus have profound implications for water yield, exacerbating the effects of changing climatic variables and reducing the basin's resilience to hydrological fluctuations.

	1993		2001		2009		2017		2023	
LULC class	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)
Forest	96.67%	90.63%	96.36%	94.64%	96.36%	96.36%	98.18%	98.18%	96.36%	96.36%
Shrubland	95.71%	90.63%	95.00%	91.94%	93.33%	98.25%	95.00%	98.28%	95.00%	96.61%
Waterbody	94.00%	92.16%	94.00%	94.00%	98.00%	90.74%	96.00%	97.96%	100.00%	94.34%
Wetland	96.67%	98.31%	94.00%	97.92%	90.00%	95.74%	98.00%	96.08%	94.00%	95.92%
Cultivated land	96.36%	94.64%	98.57%	95.83%	98.57%	93.24%	100.00%	95.89%	97.14%	100.00%
Built-up area	94.00%	97.92%	96.00%	100.00%	96.00%	97.96%	96.00%	97.96%	98.00%	96.08%
Grassland	90.00%	100.00%	96.67%	98.31%	95.00%	96.61%	96.67%	96.67%	96.67%	96.67%
Overall accuracy	94.94%		95.95%		95.44%		97.22%		96.71%	
Overall, Kappa statistics	0.9407		0.9526		0.9467		0.9674		0.9615	

Table 7. Presents the accuracy assessment results for the classified images from 1993 to 2023.

	Initial state 1993															
	LULC class	WB		BUA		SHL		F		FL		GL		WL		Row Total	Loss
		Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)	Area (km^2)	Area (%)		
Final state 2023	WB	**1.53**	12.28	2.24	7.22	2.49	0.22	2.04	0.33	15.79	0.80	2.95	0.31	14.52	3.19	41.56	40.03
	BUA	0.62	4.98	**7.92**	25.54	27.51	2.48	5.76	0.92	60.28	3.06	24.58	2.61	24.00	5.28	150.67	142.75
	SHL	2.16	17.34	0.86	2.77	**169.43**	15.29	33.86	5.41	42.07	2.14	39.78	4.23	6.92	1.52	295.08	125.65
	F	1.24	9.95	1.73	5.58	114.55	10.34	**257.33**	41.08	130.89	6.65	14.32	1.52	13.78	3.03	533.84	276.51
	FL	5.98	47.99	14.42	46.50	745.71	67.31	321.55	51.33	**1609.58**	81.74	757.14	80.44	180.66	39.71	3635.04	2025.46
	GL	0.37	2.97	1.57	5.06	44.16	3.99	4.26	0.68	34.52	1.75	**88.27**	9.38	11.33	2.49	184.48	96.21
	WL	0.56	4.49	2.27	7.32	3.97	0.36	1.62	0.26	75.94	3.86	14.23	1.51	**203.75**	44.78	302.34	98.59
	Class total	**12.46**	**100**	**31.01**	**100**	**1107.82**	**100**	**626.42**	**100**	**1969.07**	**100**	**941.27**	**100**	**454.96**	**100**	**5143.01**	
	Gain	10.93		23.09		938.39		369.09		359.49		853		251.21			
	Net change	−29.1		−119.66		812.74		92.58		−1665.97		756.79		152.62		**3390.14**	
	NP	19.01		15.11		4.79		0.36		0.84		0.80		0.33			

Table 8. Change detection matrix of 1993 to 2023.

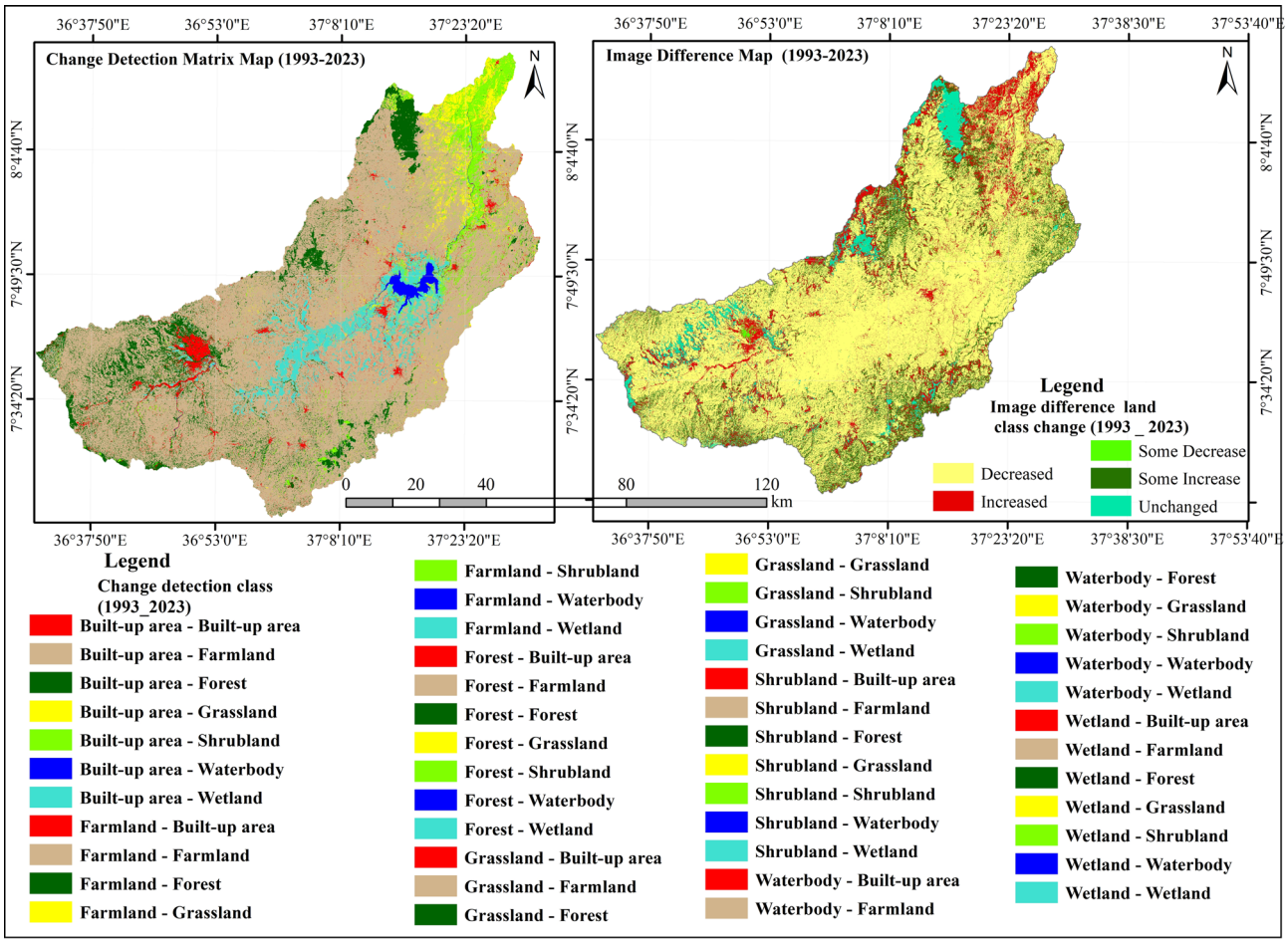

Fig. 8. Change matrix and image difference map of 1993 to 2023 *Source: The maps were created using ArcGIS software (version 10.8).*

Climate change scenario analysis

Figure 9 shows the annual variability in rainfall, revealing notable fluctuations in monthly precipitation levels across different years. In 1993, the mean monthly maximum rainfall reached 87.54 mm, the highest recorded during the study period, while the minimum was 20.26 mm. The elevated rainfall levels observed in this study are likely driven by strong monsoonal activity and positive climate anomalies, such as the Indian Ocean Dipole (IOD) or El Niño-Southern Oscillation (ENSO). Similar research highlights the influence of atmospheric circulation patterns, such as the Intertropical Convergence Zone (ITCZ), in generating high rainfall variability[107]. The year 2009 also saw substantial rainfall, with mean monthly values ranging from a peak of 68.54 mm to a low of 26.97 mm. These values suggest enhanced precipitation, possibly linked to global climatic phenomena, such as La Niña events, which are known to bring wetter-than-average conditions to parts of East Africa. The year 2017 was particularly significant because of the severe precipitation conditions. During this year, the mean monthly rainfall ranged from 12.64 mm at its highest to 7.24 mm at its lowest, coinciding with a major drought affecting much of Ethiopia. Severe drought coinciding with low rainfall, attributed to Indian Ocean warming and altered atmospheric circulation, affects Ethiopia, causing widespread agricultural losses, water scarcity, and food insecurity[108]. In 2001 and 2023, there were notable variations in the mean monthly rainfall. In 2001, rainfall ranged from a high of 26.54 mm to a low of 9.28 mm. In 2023, the maximum was slightly lower at 16.84 mm, and the minimum was 7.95 mm. These variations illustrate the influence of climatic conditions on rainfall patterns over time. Rainfall variability during these years likely reflects regional land use changes and moderate climatic disturbances. Declining trends have been attributed to urbanization and global warming-driven changes in precipitation patterns[109].

Figure 10 shows the seasonal temperature and precipitation patterns. Temperatures are highest from March to June (median ~21 °C in May) and lowest in December and January (~ 18 °C), with constant variations. Precipitation is low from December to February and increases significantly from June to September, peaking in July and August with medians over 2,000 mm and a maximum exceeding 3,500 mm. These patterns emphasize various climatic cycles that are important for agricultural and water resource management.

The correlation analyses presented in Table 9 further highlight the performance of this model. For the period from 1993 to 2001, the correlation value was 0.9957, indicating a high agreement between the model simulations and observed data, reflecting the model's reliability during this period. The correlation improved to 0.9986 from

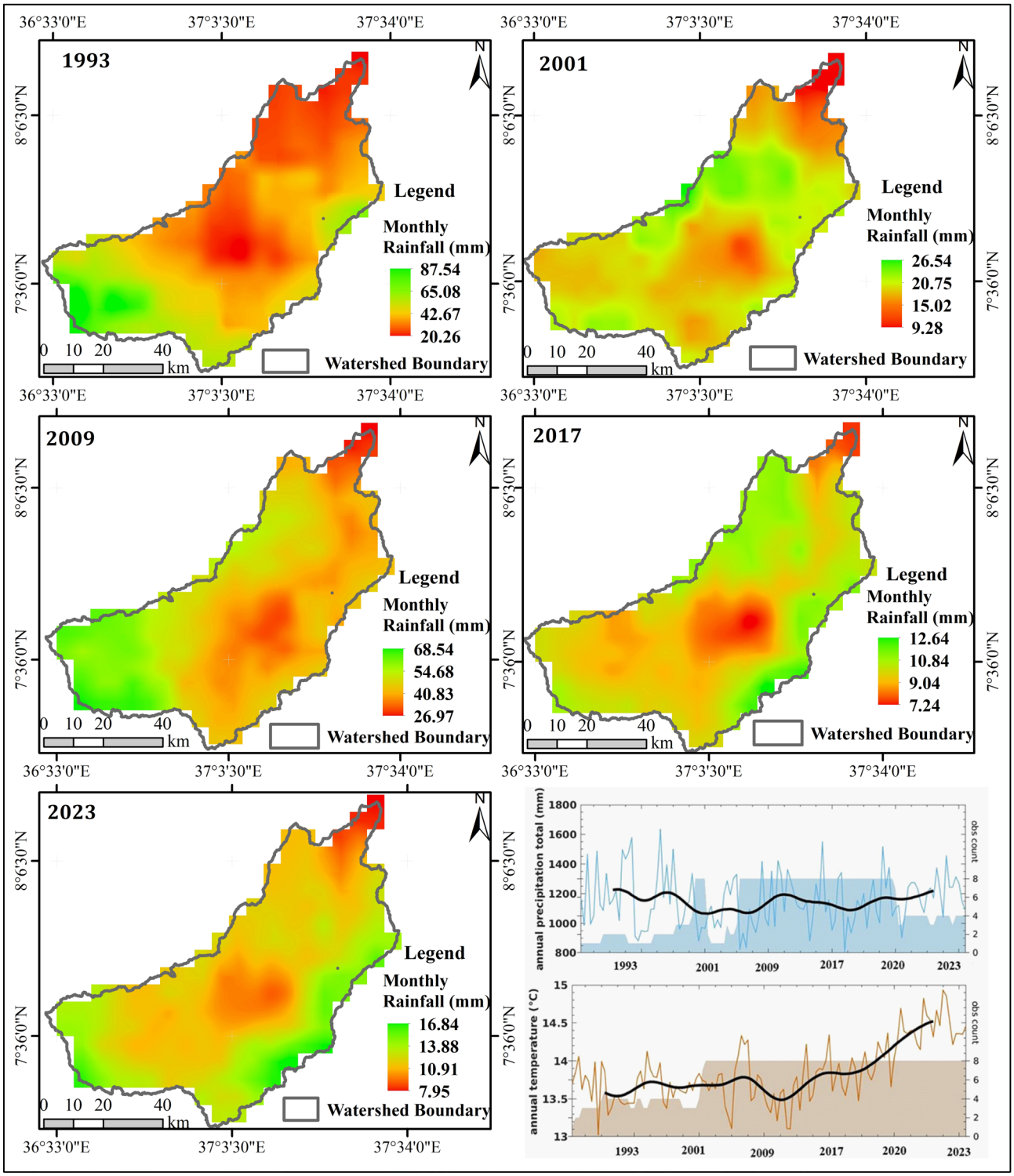

Fig. 9. Spatial distribution of precipitation and temperature variations and their influence on surface water yield in the gilgel gibe catchment.

2001 to 2009, suggesting enhanced model accuracy and performance, which is likely due to improved calibration or data quality. From 2009 to 2017, the correlation remained high at 0.9983, demonstrating consistent model performance. The most recent period, from 2017 to 2023, showed the highest correlation value of 0.99997, indicating an almost perfect alignment between the simulated and observed data, possibly due to advancements in modeling techniques or data quality. Overall, the correlation value for the entire study period from 1993 to 2023 was 0.9959, highlighting the model's strong performance and ability to accurately simulate water yield dynamics under diverse climatic and land-use conditions.

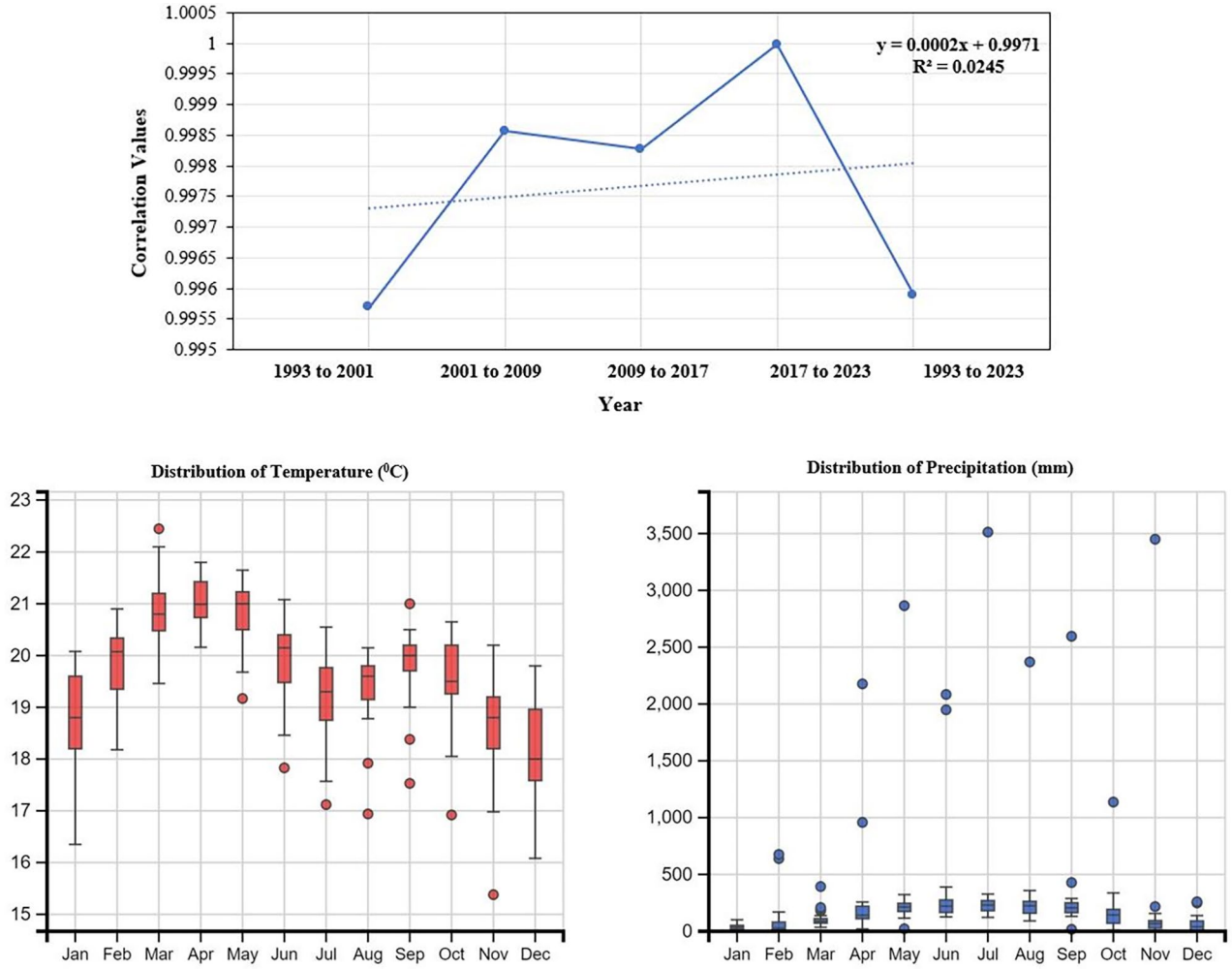

Fig.10. Correlation matrix of climate variables from 1993 to 2023.

Years compared	Correlation values
1993 to 2001	0.99568724
2001 to 2009	0.998560882
2009 to 2017	0.998274231
2017 to 2023	0.999970849
1993 to 2023	0.995892826

Table 9. Correlation analysis of long-term climate data.

Comparative evaluation of IDW and kriging interpolation models

Spatial interpolation was used in this study to estimate climate variables (precipitation and temperature) across ungauged locations in the Gilgel Gibe watershed for water yield modeling. This study compared two widely used methods: Inverse Distance Weighting (IDW), which is deterministic, and distance-based weighting. Ordinary Kriging: Geostatistical method incorporating spatial autocorrelation (Table 10).

The comparative analysis of annual mean precipitation interpolation methods demonstrated that Ordinary Kriging with an exponential variogram model (including the nugget effect) outperformed IDW (power parameter p = 2) across all validation metrics. Kriging achieved a lower RMSE (89.7 mm/yr vs. 112.3 mm/yr) and MAE (68.4 mm/yr vs. 87.6 mm/yr), representing a 20.1% and 21.9% improvement in accuracy, respectively. Additionally, Kriging exhibited a stronger correlation with the observed data ($R^2 = 0.81$ vs. 0.72), confirming its superior ability to capture spatial precipitation patterns, particularly in regions with complex terrain or sparse station coverage. These findings align with those of previous studies in similar climatic regions, reinforcing Kriging as the preferred method for hydrological applications requiring high-precision precipitation estimates.

Evaluation of ML algorithms for climate data analysis

The climate data for the Gilgel Gibe watershed indicate a significant temperature range, with maximum values reaching 34 °C and minimum values decreasing to 14 °C. Precipitation data revealed that the region experiences monthly rainfall ranging from 0 to 50 mm. Solar irradiance levels in the area vary between 6 and 7.75 kilojoules per square meter (kJ/m^2), while wind speeds fluctuate between 0 and 8 m per second. The relative humidity in the watershed ranges from 40 to 90%, reflecting the region's role as a condensation point for air masses that condense into liquid water at constant atmospheric pressure. This climatic variability throughout the year has a notable impact on water yield in the watershed (Fig. 11a, b, and c).

The climate model performance was assessed using data from two precipitation stations (Gilgel Gibe and Jimma) and two temperature stations (Omo Nada and Deneba). The performance of the SVM algorithm was notably high across the analyzed climate stations. For the Gilgel Gibe station, the SVM model achieved the following performance metrics: MAE = 0.36, RMSE = 0.45, R^2 = 0.81, NSE = 0.81, and Willmott's Index of Agreement (d) = 0.94. These results indicate that the SVM model effectively minimized the prediction errors and was closely aligned with the observed data, demonstrating a robust capacity for capturing complex climatic patterns. Similarly, at Jimma station, the SVM model performed with MAE = 0.32, RMSE = 0.44, R^2 = 0.74, NSE = 0.74, and d = 0.91. Table 11 and Fig. 12 provide a comprehensive overview of the model performance, showing that all models exhibited strong fits to the observed data, as evidenced by metrics approaching 1. This alignment indicates that the models are highly accurate with minimal inconsistencies between the predicted and actual values. Finally, the analysis underscores the robustness of our climate models, revealing significant correlations between climatic variation and water resources in the Gilgel Gibe watershed. The results confirm the accuracy and reliability of the model in assessing climate change impacts on water yield, offering a solid foundation for effective water resource management and climate adaptation strategies.

Generally, the Gilgel Gibe watershed is a crucial hydrological resource in Ethiopia and faces water yield challenges driven by climate change, land use dynamics (deforestation, agricultural expansion, and urbanization), and socio-economic pressures. These activities disrupt the hydrological cycles, leading to increased surface runoff, reduced infiltration, and diminished groundwater recharge. Over the past 30 years, the study area has experienced a loss of 276.51 km^2 (5.36%) of forest cover. The removal of forest vegetation leads to increased surface runoff and sediment yield, ultimately reducing water availability. Although deforestation may initially lead to a temporary increase in water yield, it disrupts the hydrological cycle by accelerating soil degradation and reducing groundwater recharge capacity. Findings from the Gilgel Gibe sub-basin indicate that deforestation has resulted in a 15% increase in annual runoff while causing a 20% reduction in dry-season baseflow due to diminished groundwater recharge[110]. This study confirmed that deforestation is a primary factor driving LULC fluctuations within the Gilgel Gibe Catchment. Forests are being cleared for agriculture, construction, firewood, charcoal production, and timber extraction; however, replanting and conservation efforts remain inadequate. Consequently, the region faces multiple environmental challenges, including increased soil erosion, water scarcity, heightened flood risks, rising temperatures, and more frequent droughts. Ethiopia's climate, characterized by alternating wet and dry seasons influenced by the Inter-Tropical Convergence Zone (ITCZ), exacerbates these challenges, resulting in erosion, sedimentation, and altered water yields. Similar trends have been observed in other studies, such as[111], who explored topography-based methods for downscaling sub-grid precipitation in Earth system models, as well as in watersheds such as the Awash River Basin[112], which revealed reduced base flow and increased runoff due to land cover change. To address these issues, actionable recommendations include strengthening climate-resilient infrastructure (check dams, terraces, reservoirs), as demonstrated in the Upper Tana Basin[113]; implementing community-based water use planning, such as water-efficient irrigation training; incorporating ecosystem-based approaches such as reforestation and controlled grazing[114] and integrating local actions with national (Ethiopia's Growth and Transformation Plan (GTP) and global (Sustainable Development Goals (SDG) policies, particularly within frameworks such as the Nile Basin Initiative. These strategies, which focus on infrastructure, community engagement, and ecosystem preservation, can inform broader sustainable water management across Ethiopia and similar regions.

Machine learning-based streamflow modeling

In this study, four machine learning models, RF, SVM, XGB, and LGBM, were used to simulate streamflow. The daily streamflow predictions of these models are illustrated in Fig. 13 (a, b, and c), offering a comprehensive view of the simulated streamflow patterns. The dataset was partitioned into two phases: the training phase, which encompassed January 1995 to December 2019 (85% of the data from January 1995 to) and the testing phase, which spanned January 2020 to December 2022 (15% of the data). The effectiveness of RF, SVM, XGB, and LGBM in streamflow simulations was evaluated using several statistical indicators. The RF model showed the highest accuracy in predicting observed discharge values compared to the SVM and XGB models, as shown in the regression graphs and the GitHub repository (https://github.com/amanuel2025/Climate-and-Hydrology-Data-Analysis-using-ML). This superior performance can be attributed to RF's capability of RF to capture complex

Metric	IDW Power parameter (p = 2)	Ordinary kriging variogram model (Exponential), nugget effect
RMSE (mm/yr)	112.3	89.7
MAE (mm/yr)	87.6	68.4
R^2	0.72	0.81

Table 10. Comparison of annual mean precipitation interpolation (IDW and Kriging) metrics.

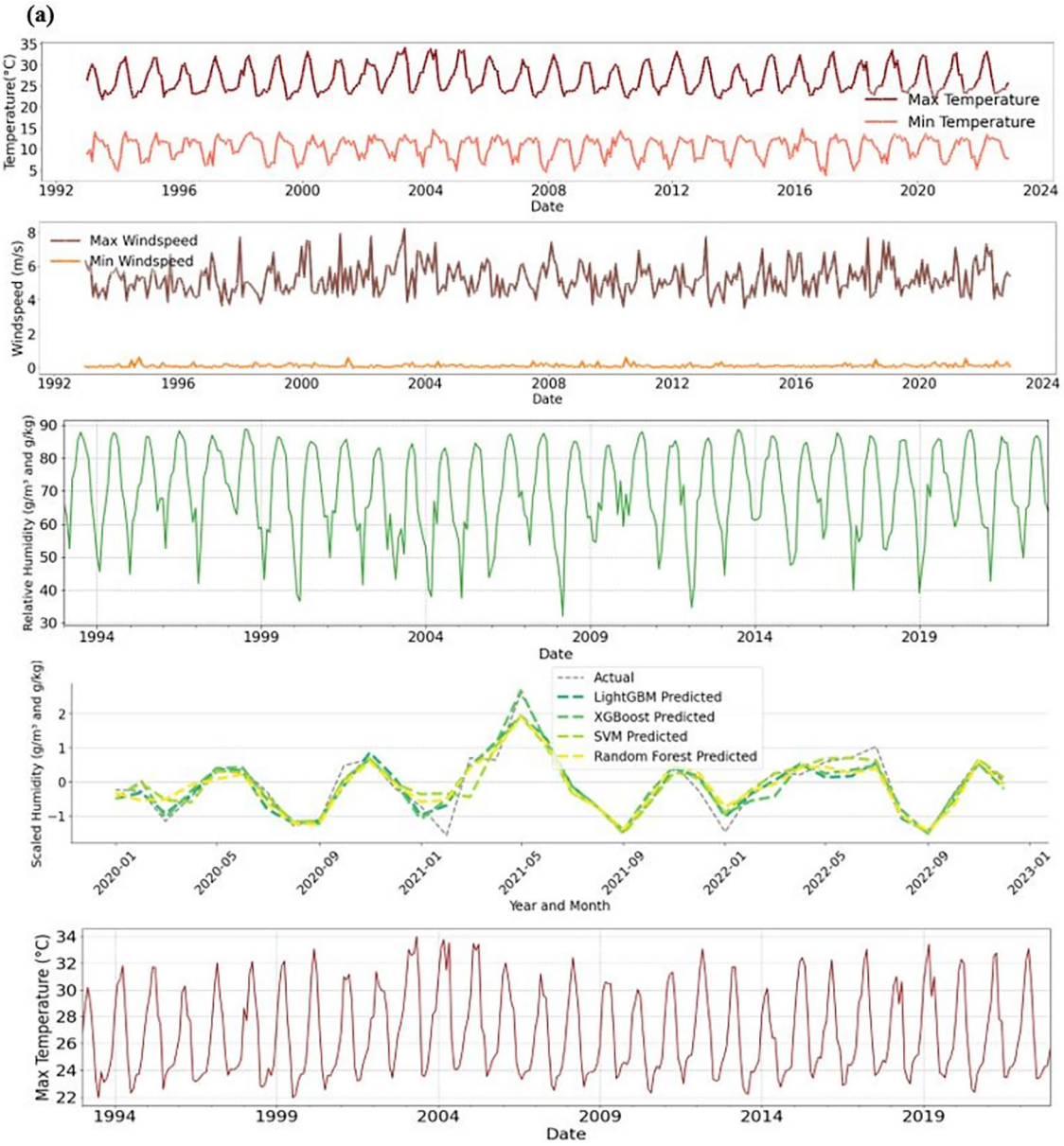

Fig. 11. (**a**, **b**, and **c**) Comparison of observed and predicted values for climatic variables using different machine learning algorithms.

nonlinear relationships in the data, making it particularly adept at handling complex streamflow dynamics. Both the RF and XGB models excel in processing sequential data, whereas SVM and LGBM are better suited for dealing with fuzzy and uncertain rule-based systems.

Fig. 13 (a, b, and c) present the analysis of the RF, SVM, XGB, and LGBM models, examining the relationship between the five input parameters (daily mean, daily flow, discharge, and runoff) and the output parameter (mean daily discharge). The model performance was evaluated using five statistical indicators: RMSE, MAE, R^2, NSE, and Willmott's Index of Agreement (d). During training, the RF model achieved MAE, RMSE, R^2, NSE, and d values of 0.38, 0.61, 0.76, 0.76, and 0.93, respectively, indicating an average deviation between the predicted and observed discharges. During testing, these values improved slightly, with MAE, RMSE, R^2, NSE, and WIA values of 0.25, 0.34, 0.82, 0.82, and 0.94, respectively, showing a minor increase in prediction error compared to the training phase. The XGB model showed a moderate correlation between predicted and observed discharge values, with R^2 values of 0.72 during training and 0.80 during testing. This indicated that the model explained 72% and 80% of the discharge variance in the training and testing phases, respectively. These results highlight the effectiveness of the model in predicting streamflow dynamics using daily flow, maximum and minimum discharge, and runoff data from Gilgel Gibe water yield (Table 12 and Fig. 14).

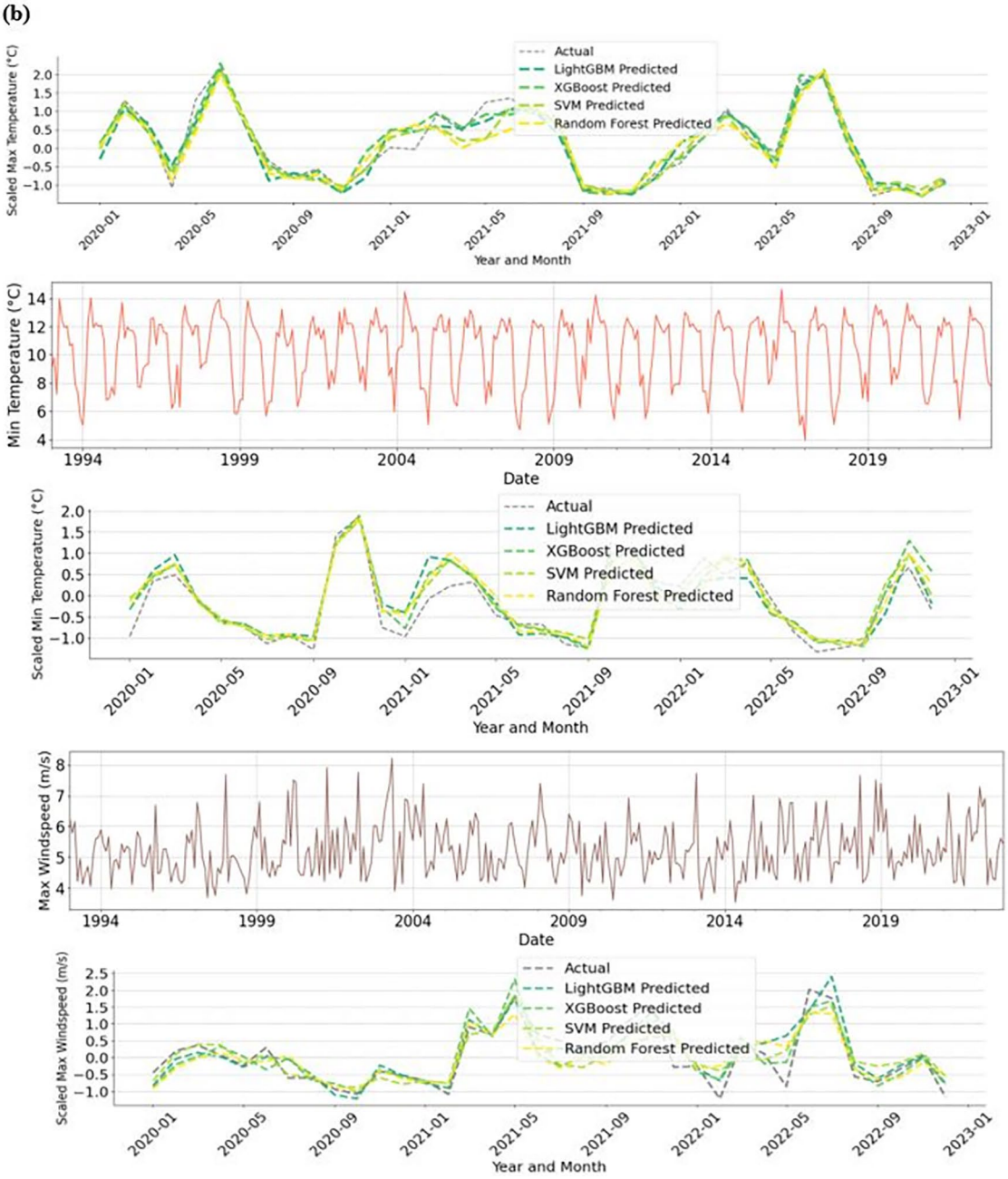

Fig. 11. (continued)

Machine learning cross validation results for climate and hydrological data

The assessment of climate and land use changes on water yield within the Gilgel Gibe watershed necessitates the integration of diverse spatial and non-spatial datasets to identify key influencing factors. Although the selection of cross-validation methods may appear secondary when a model exhibits satisfactory performance on a given dataset, our findings underscore the profound impact that cross-validation strategies have on the model's capability to accurately evaluate climate and land use change effects on water yield. To enhance the model's generalizability and mitigate overfitting, hyperparameter optimization was conducted using BayesianSearchCV with the hyperparameter'c'configured to 3. Furthermore, a walk-forward cross-validation approach was employed to incrementally validate the model's performance. Initially, the dataset was partitioned into 90% for training and 10% for testing. The validation process followed a sequential progression, where the model was trained on the 1993–2016 data and tested on 2017, then retrained on 1994–2017 and tested on 2018, continuing this pattern incrementally up to 2021 for climate. This rigorous approach facilitated a more reliable evaluation of the model's predictive accuracy over varying temporal spans. The evaluation of model performance was conducted using a comprehensive suite of statistical metrics, including MAE, RMSE, R^2, NSE, and d, as detailed in Table 13 &Table 14. The cross-validation results of SVM model demonstrated strong performance

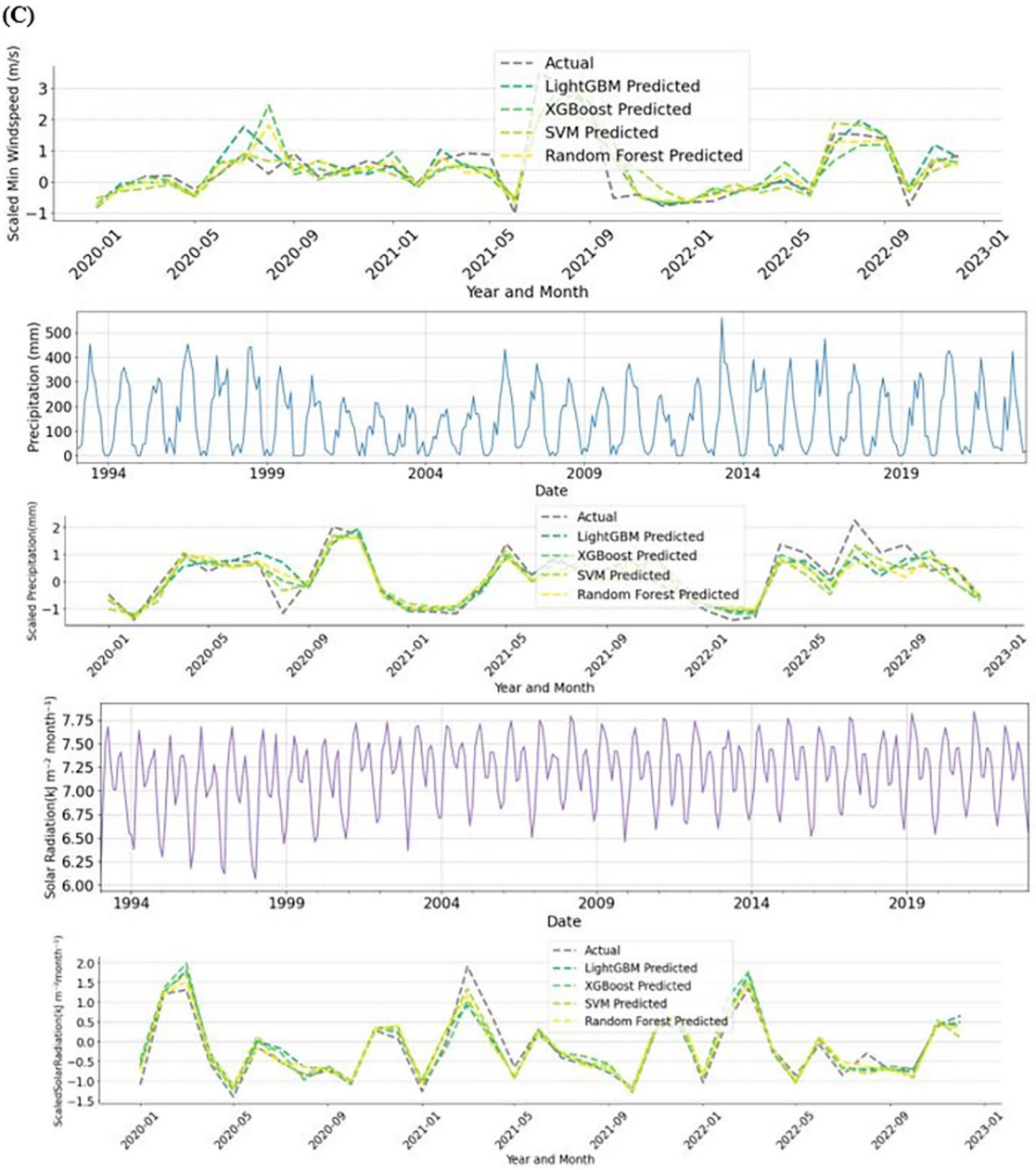

Fig. 11. (continued)

in predicting streamflow components. Mean streamflow predictions achieved consistently high accuracy during training ($R^2 > 0.99$; 2018: 0.995, 2021: 0.994). Maximum streamflow also exhibited excellent performance ($R^2 > 0.97$ for most years, 2018: 0.980, 2017: 0.974). Stream runoff predictions yielded R^2 values between 0.82 (2021) and 0.99 (2018). Minimum streamflow prediction accuracy varied, with high R^2 values in 2018 (0.986) and 2019 (0.995) but a lower value in 2020 (0.733). Overall streamflow prediction accuracy also varied interannually, with higher R^2 values in 2018 (0.956) and 2019 (0.895) and a lower value in 2021 (0.708). These results suggest that while the SVM model performs exceptionally for mean and maximum streamflow, minimum streamflow and streamflow for some periods require further refinement. Overall, the SVM model shows strong potential for accurate water yield predictions, particularly for critical metrics such as mean and maximum streamflow, providing reliable insights for hydrological management in the Gilgel Gibe watershed. These metrics provide a robust framework for quantifying predictive accuracy and model reliability in capturing the underlying patterns of climate and hydrological data.

The cross-validation results of the SVM model for climate data in the Gilgel Gibe watershed (Table 13) revealed variable performance across different climate factors. For Maximum Wind Speed, the model achieved R^2 values ranging from 0.60 to 0.95, with the highest performance observed in 2018 ($R^2 = 0.95$). The model consistently demonstrated strong performance for Minimum Temperature, with R^2 values spanning from 0.91 in 2021 to 0.97 in 2019 and 2022. Maximum Temperature displayed particularly impressive results, with R^2 values

Gauging station	ML model	MAE	RMSE	R^2	NSE	d
Gilgel gibe	RF	0.38	0.5	0.76	0.76	0.92
	SVM	0.36	0.45	0.81	0.81	0.94
	XGB	0.38	0.47	0.79	0.79	0.93
	LGBM	0.37	0.54	0.72	0.72	0.91
Jimma	RF	0.35	0.51	0.66	0.66	0.88
	SVM	0.32	0.44	0.74	0.74	0.91
	XGB	0.27	0.38	0.81	0.81	0.94
	LGBM	0.36	0.46	0.72	0.72	0.9
Omo nada	RF	0.24	0.33	0.9	0.9	0.97
	SVM	0.2	0.28	0.93	0.93	0.98
	XGB	0.18	0.21	0.96	0.96	0.99
	LGBM	0.21	0.28	0.93	0.93	0.98
Deneba	RF	0.26	0.38	0.86	0.86	0.96
	SVM	0.22	0.3	0.91	0.91	0.98
	XGB	0.18	0.23	0.95	0.95	0.99
	LGBM	0.24	0.31	0.91	0.91	0.98

Table 11. Comparison of machine learning model performance for monthly climate data in the gilgel gibe watershed.

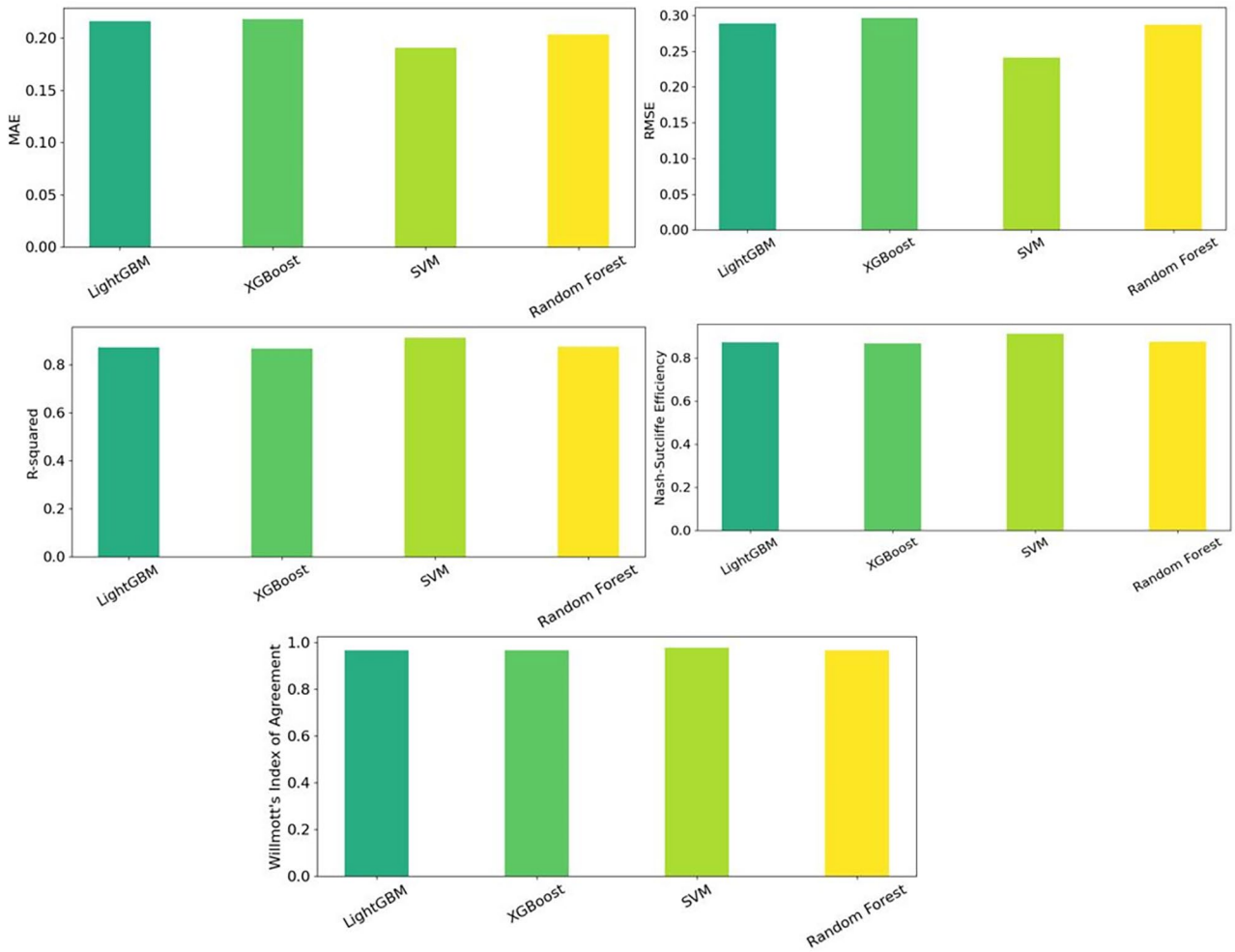

Fig.12. Comparison of performance metrics for RF, SVM, XGB, and LGBM on climate data.

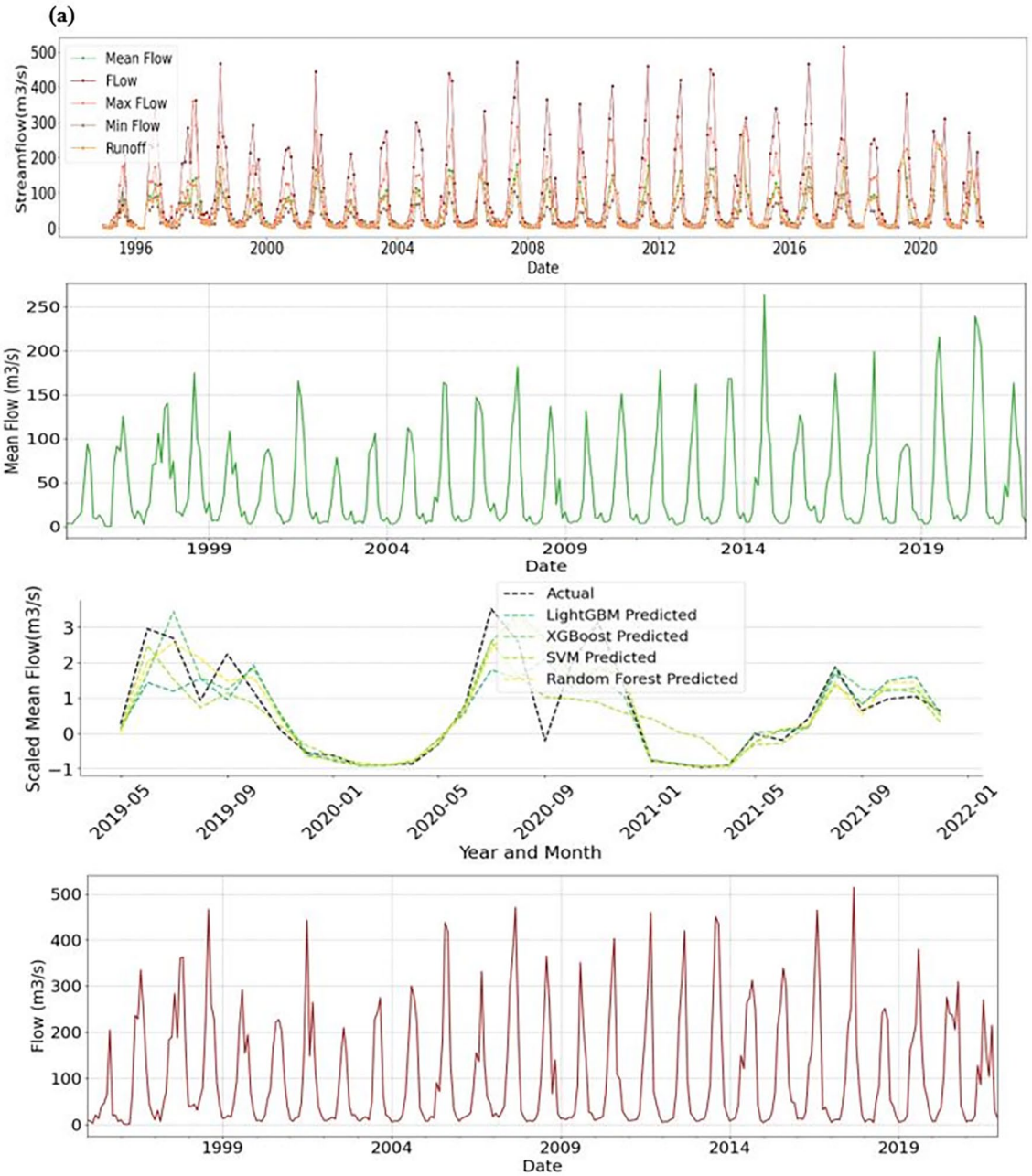

Fig.13. (**a**, **b**& **c**) of Observed and predicted streamflow for RF, SVM, XGB, and LGBM models during training and testing phases.

nearing 0.99 in both 2020 and 2022. Solar Radiation maintained high accuracy across all years, with R^2 values exceeding 0.97, peaking at 0.99 in 2020. Precipitation also showed strong predictive performance, particularly in 2020 ($R^2 = 0.97$). Lastly, Relative Humidity exhibited a range of R^2 values from 0.89 in 2018 to 0.96 in 2022. Overall, the SVM model demonstrated robust performance in predicting climate factors, with consistently high R^2 values for most variables.

Evaluation and validation of the invest model

The InVEST model estimates water yield using Budyko's hydrological framework, and its accuracy is highly dependent on proper calibration. The Z calibration constant is a key parameter in the model that controls the partitioning of precipitation between evapotranspiration (ET) and water yield. A higher Z value (> 10%) leads to an overestimation of ET, reducing water yield predictions, whereas a lower Z value (< 10%) increases the estimated water yield but may underestimate ET losses. In this study, an initial Z value of 10 was used, which was aligned with the global average for mixed land cover (Table 15).

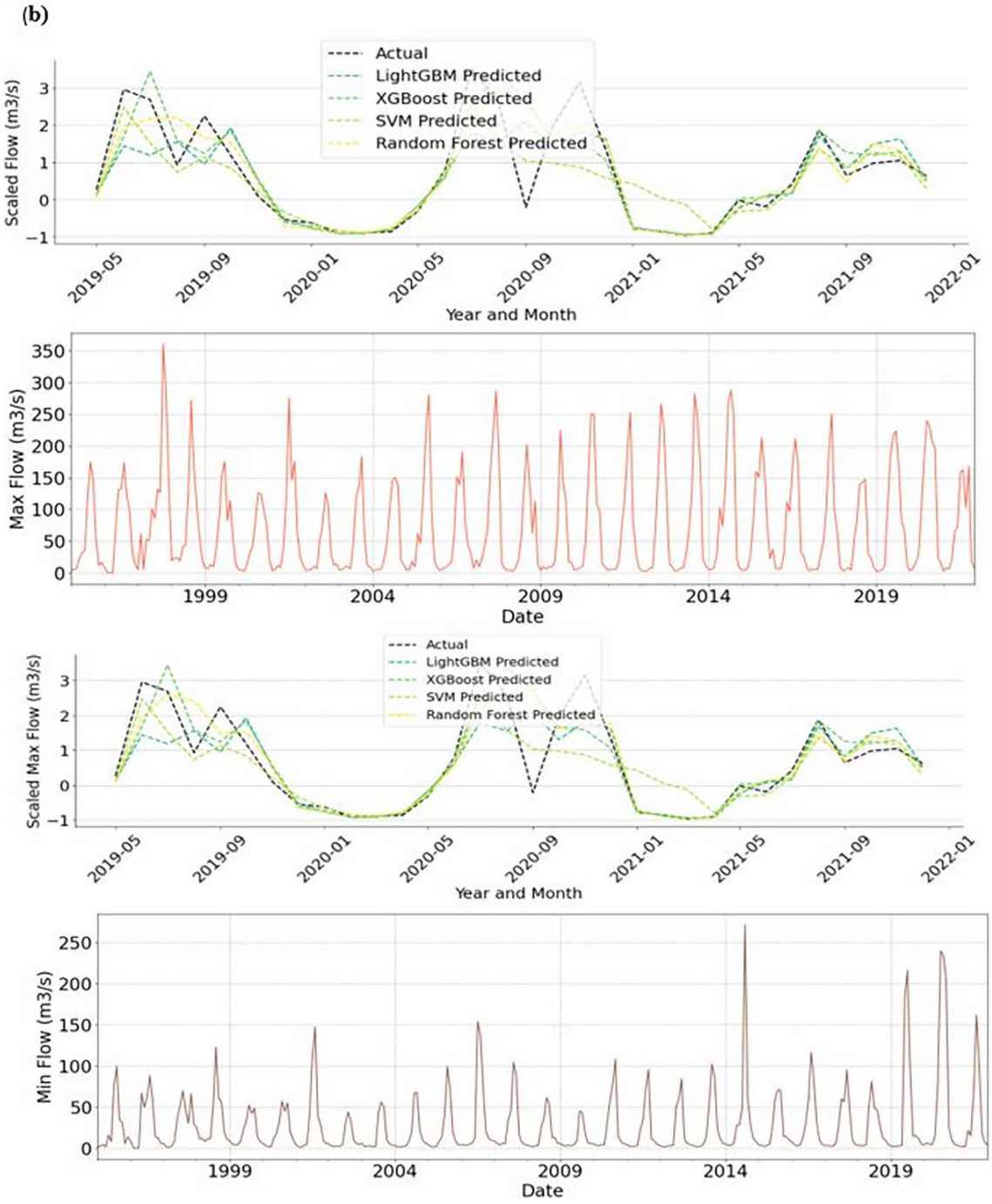

Fig. 13. (continued)

The calibration of the Z factor in the InVEST model significantly influences the accuracy of water yield estimation. The results show that adjusting Z improves model performance by reducing runoff overestimation in tropical regions and refining the balance between evapotranspiration (ET) and water yield. For the Gilgel Gibe Watershed, a Z value of 11.2 resulted in NSE improvement from 0.59 to 0.70 and $R^2 = 0.65$, with a + 5.8% bias, suggesting a better representation of mixed agricultural and forested land cover. When the Z value was set to 12, NSE improved to 0.68, R^2 increased to 0.71, and bias reached + 8.2%, indicating a more stable estimate but slightly higher error in dry-season flow estimation. However, the initial Z value of 10, which aligns with the global average for mixed land cover, showed weaker performance (NSE = 0.55, $R^2 = 0.58$, and bias = + 15%), demonstrating overestimated runoff in tropical environments. This highlights the necessity of local calibration to refine hydrological predictions. The machine learning-based optimization of Z was used for these purposes to reduce bias to + 3.1%, demonstrating its potential in fine-tuning water yield models for regional hydrology.

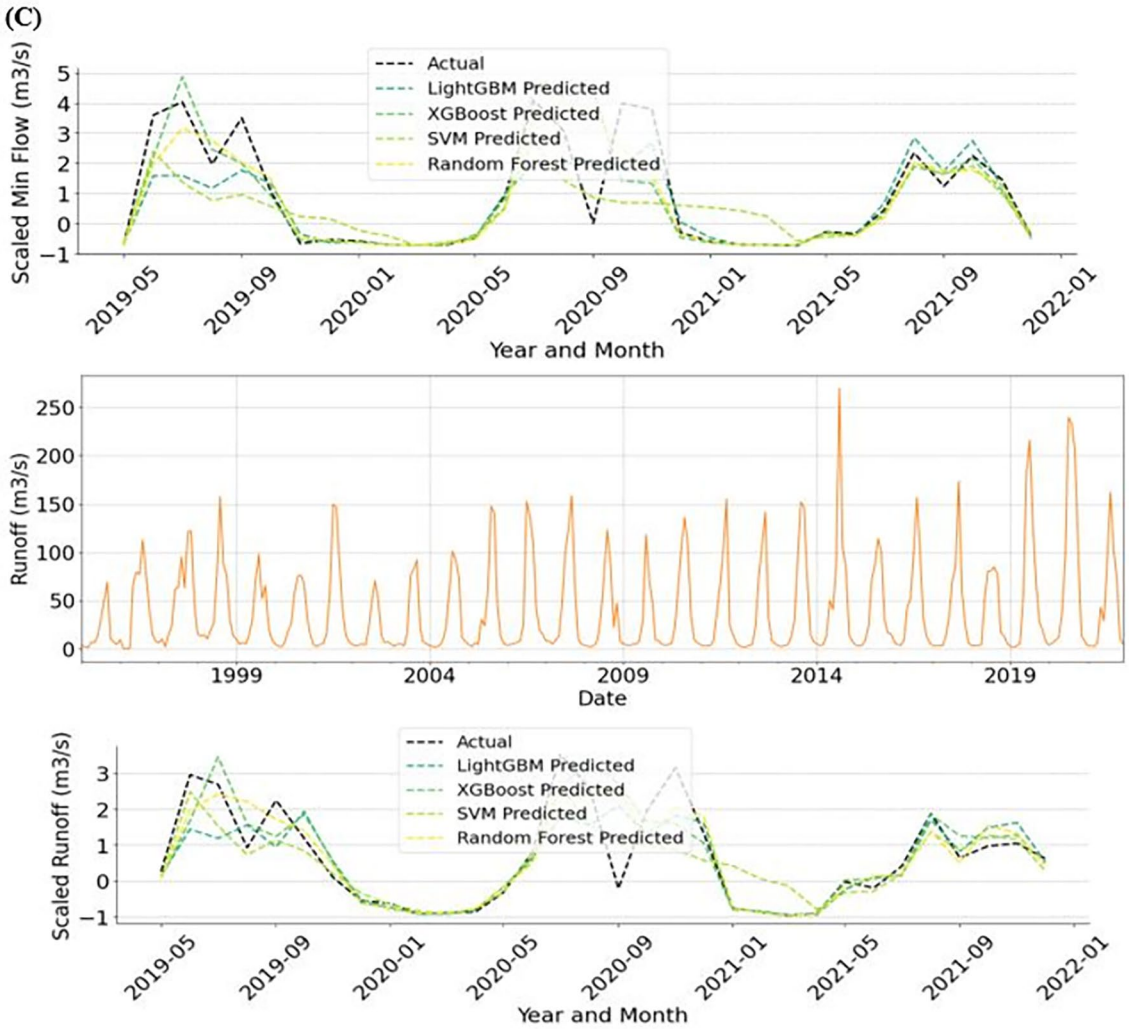

Fig. 13. (continued)

Gauging station	ML model	MAE	RMSE	R^2	NSE	d
Gilgel gibe	RF	0.38	0.61	0.76	0.76	0.93
	SVM	0.43	0.61	0.76	0.76	0.92
	XGB	0.35	0.66	0.72	0.72	0.92
	LGBM	0.43	0.68	0.7	0.7	0.9
Bulbul	RF	0.19	0.29	0.8	0.8	0.95
	SVM	0.13	0.17	0.93	0.93	0.98
	XGB	0.15	0.24	0.86	0.86	0.96
	LGBM	0.16	0.25	0.84	0.84	0.96
Bidru awana	RF	0.45	0.95	0.68	0.68	0.88
	SVM	0.55	1.19	0.49	0.49	0.77
	XGB	0.42	1.01	0.63	0.63	0.87
	LGBM	0.46	0.98	0.66	0.66	0.87

Table 12. Evaluation of machine learning models for daily streamflow simulation in the gilgel gibe watershed.

Evapotranspiration impact

Evapotranspiration (ET) is key to understanding global hydrological budgets and water yield modeling. Its precise estimation is essential for comprehending the water balance and formulating effective water resource management plans for the Gilgel Gibe Watershed. However, regional-scale estimations using meteorological

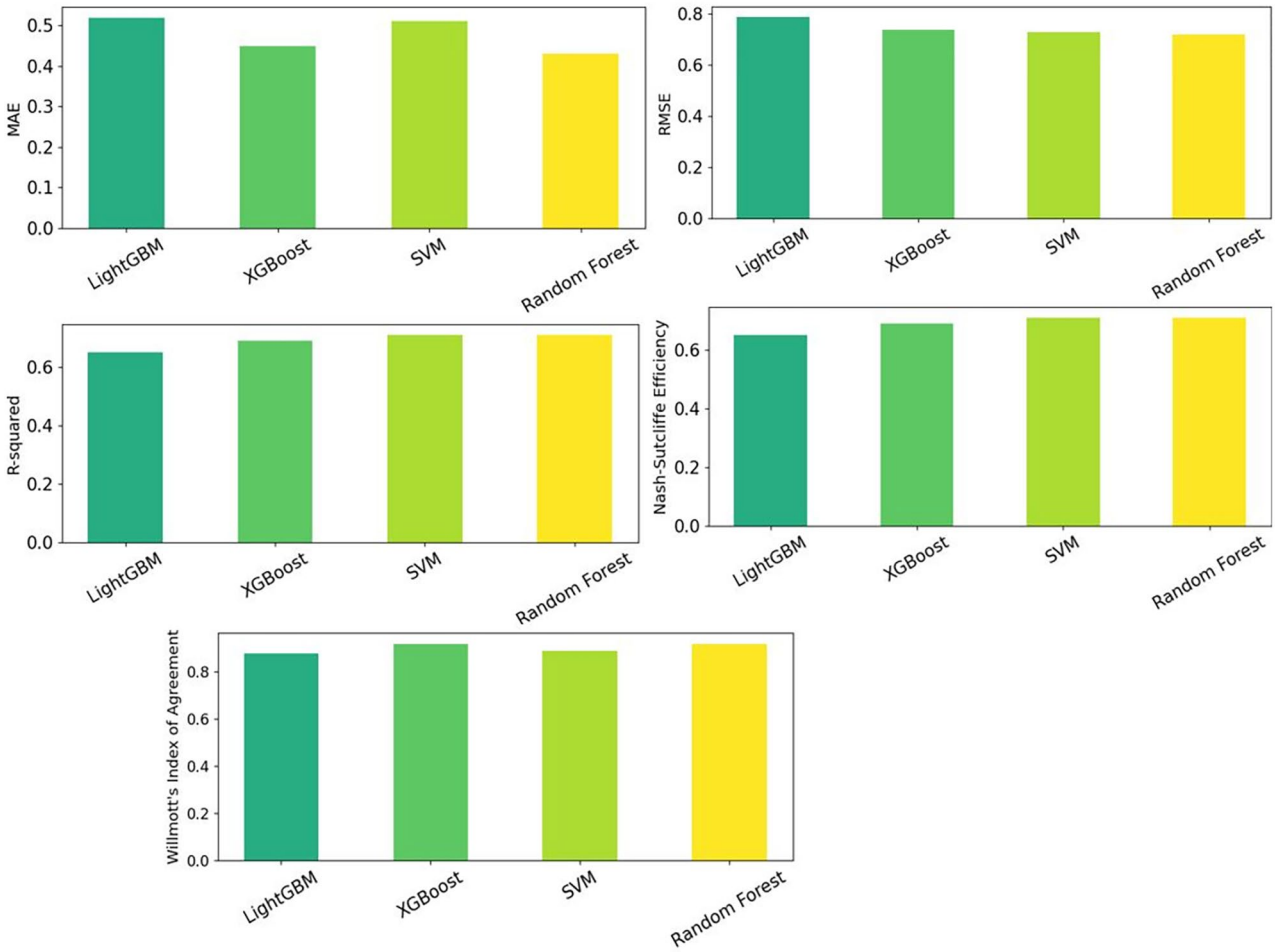

Fig.14. Statistical evaluation of RF, SVM, XGB, and LGBM models on hydrological data.

data from the GLDAS, MODIS 1 km and 500 m, and the Penman–Monteith method, an equation used to compute water evaporation from vegetated surfaces, are less commonly employed (Fig. 15a.,15b and Fig. 16).

Analysis of potential evapotranspiration (PET) trends from 2000 to 2023 revealed varied spatial distributions without a distinct overall pattern. On a daily scale, PET trends (Fig. 15a) ranged from 1 to 4 mm, except for MODIS 500 m, which showed a slightly decreasing trend close to 0.9 mm. Annually (Fig. 15b), all evapotranspiration (ET) graphs consistently show increasing PET trends, ranging from 100 mm to 350–400 mm. Figures 15a and b illustrate the magnitudes of these detected PET trends across various meteorological stations on daily and annual scales, respectively. During summer, a seasonal increase in PET was evident, ranging from 1 to 4 mm. The Penman–Monteith-Leuning method for ET components highlights distinct behaviors in different evapotranspiration processes. Canopy Evaporation (Ec) showed values greater than 1.5 mm, reaching up to 3 mm. This represents the evaporation of water from wet canopy surfaces after rainfall, where water on the leaves and stems evaporates back into the atmosphere.

Interception Evaporation (Ei) ranges from 0 mm to 0.5 mm, indicating the water intercepted by the canopy and evaporated without reaching the ground, including the water caught on leaves and stems during rainfall. Soil Evaporation (Es) values ranged from 1 to 2 mm throughout the study period, highlighting the evaporation of water directly from the soil surface. This process is influenced by the soil moisture content, soil type, temperature, and other environmental factors. These findings are consistent with those of previous studies on evapotranspiration, such as[115], which indicated that increasing temperatures and changing precipitation patterns strongly affected PET trends. The consistent increase in PET, particularly in summer, highlights the impact of climate change on water availability. The varying behavior of the canopy, interception, and evaporation of ET components offer valuable insights into the ecosystem's water balance, stressing the need for adaptive water management strategies in response to these changes.

The spatial trend of evapotranspiration, as depicted in Fig. 16, showed that the highest ET was recorded in the years 2000 and 2017, with values of 357.34 mm and 355.24 mm, respectively. Ethiopia faces several significant natural hazards, including droughts, with different spatial and temporal patterns. In 1999/2000, the country experienced a short-season rainfall failure that affected approximately 10 million people. In 2017, spring rain failed, while the main rainy season brought heavy rainfall and flooding across the country. By June, the number of districts needing emergency food aid rose from 192 to 228, with approximately 461 districts facing severe

Stream flow						
Hyperparameter ML model	**Validation periods**	**MAE**	**RMSE**	**R^2**	**NSE**	**d**
SVM	2017	0.160116	0.235198	0.812101	0.812101	0.949788
	2018	0.133798	0.199724	0.956126	0.956126	0.988273
	2019	0.076868	0.121076	0.894572	0.894572	0.970449
	2020	0.104187	0.149996	0.820345	0.820345	0.952220
	2021	0.076874	0.109470	0.707951	0.707951	0.901344
Maximum stream flow						
SVM	2017	0.038059	0.071391	0.974375	0.974375	0.993071
	2018	0.061420	0.101327	0.979960	0.979960	0.994856
	2019	0.080035	0.140080	0.916541	0.916541	0.975710
	2020	0.111268	0.160798	0.804183	0.804183	0.958858
	2021	0.094128	0.123544	0.401252	0.401252	0.836768
Minimum stream flow						
SVM	2017	0.311365	0.503064	0.766623	0.766623	0.935335
	2018	0.082203	0.120666	0.985826	0.985826	0.996488
	2019	0.044067	0.057217	0.995026	0.995026	0.998781
	2020	0.187347	0.291490	0.733111	0.733111	0.938819
	2021	0.083076	0.110742	0.776238	0.776238	0.927561
Mean stream flow						
SVM	2017	0.035880	0.069504	0.990127	0.990127	0.997519
	2018	0.045885	0.060779	0.995246	0.995246	0.998826
	2019	0.067674	0.109928	0.978119	0.978119	0.994382
	2020	0.082538	0.121026	0.947999	0.947999	0.987198
	2021	0.034554	0.046590	0.994821	0.994821	0.998724
Stream runoff						
SVM	2017	0.096648	0.138755	0.950993	0.950993	0.986841
	2018	0.060236	0.083648	0.990306	0.990306	0.997562
	2019	0.091977	0.137801	0.880007	0.880007	0.967510
	2020	0.065171	0.099424	0.941039	0.941039	0.983530
	2021	0.061919	0.091692	0.818511	0.818511	0.939771

Table 13. Best-performing ML models across different training periods for hydrological data prediction/validation.

difficulties. The Gilgel Gibe watershed was one of the sites affected by this drought in 2000 and 2017. Owing to inadequate spring rains, the number of individuals requiring food assistance has increased to 7.8 million. Additionally, linked to the El Niño weather phenomenon, the 2015–2017 drought was severe and impacted large areas of Ethiopia. This has resulted in widespread crop failure, livestock deaths, and acute food insecurity, affecting millions of people across the country[116,117].

In addition, our study indicated that high ET values also reflect high water demand for crops, which is essential for maximizing agricultural yields. Ecosystems are intrinsically linked to water cycles. Rising ET values signify healthy and productive ecosystems with active vegetation growth. However, these ET values are influenced by and contribute to climate change. Increased temperature and altered precipitation patterns can elevate ET rates, creating response mechanisms that exacerbate warming. The Penman–Monteith equation proved to be the most effective model for estimating ET in the region. This model incorporates factors such as temperature, humidity, wind speed, and solar radiation. The results showed consistently high ET values in the Gilgel Gibe watershed, which was evident through both periodical and spatial variations in its distribution. Periodical variation of the changes in ET values observed over different time periods, reflecting seasonal or annual fluctuations driven by factors such as changes in vegetation, rainfall patterns, and temperature. Spatial variation denotes the differences in ET values across different geographic locations within the watershed, influenced by variations in land cover, topography, and soil characteristics.

In addition, this study investigated how climate change and land use affect the water yield. Correlation analysis showed strong associations between climate factors and water yield, while regression quantified land-use impacts. Machine learning models, including RF, SVM, EGBM, and LGBM, were trained on historical data and cross-validated to analyze future water yield under various scenarios. The combined effect of these periodical and spatial variations provides a comprehensive understanding of the dynamics of evapotranspiration in the Gilgel Gibe watershed. In summary, high evapotranspiration values have profound implications across various domains, from water resource management and agriculture to ecosystem health and climate regulation. Understanding and managing these high ET rates through advanced measurement, modeling techniques, and

Maximum wind speed						
Hyperparameter ML model	**Validation periods**	**MAE**	**RMSE**	**R^2**	**NSE**	**d**
SVM	2018	0.218408	0.294760	0.947563	0.947563	0.986990
	2019	0.170316	0.221140	0.598750	0.598750	0.829220
	2020	0.133207	0.159099	0.877917	0.877917	0.964957
	2021	0.541286	0.652379	0.797931	0.797931	0.922045
	2022	0.151441	0.185899	0.950845	0.950845	0.987379
Maximum wind speed						
SVM	2018	0.325928	0.386116	0.928420	0.928420	0.979493
	2019	0.188751	0.289865	0.842745	0.842745	0.954945
	2020	0.112529	0.153765	0.885784	0.885784	0.972063
	2021	0.236564	0.332412	0.838126	0.838126	0.957687
	2022	0.325928	0.386116	0.928420	0.928420	0.979493
Minimum temperature						
SVM	2018	0.225324	0.279434	0.908737	0.908737	0.978468
	2019	0.154623	0.181111	0.965496	0.965496	0.991379
	2020	0.133424	0.205834	0.956021	0.956021	0.989015
	2021	0.180752	0.225089	0.917912	0.917912	0.977710
	2022	0.093249	0.138830	0.968224	0.968224	0.991379
Maximum temperature						
SVM	2018	0.148155	0.207708	0.962677	0.962677	0.990076
	2019	0.173054	0.232420	0.941218	0.941218	0.985846
	2020	0.042402	0.070037	0.995414	0.995414	0.998833
	2021	0.110115	0.136153	0.978346	0.978346	0.994509
	2022	0.060151	0.078122	0.994633	0.994633	0.998665
Solar radiation						
SVM	2018	0.093126	0.125540	0.975976	0.975976	0.994345
	2019	0.077390	0.091347	0.988927	0.988927	0.997175
	2020	0.057939	0.084737	0.987749	0.987749	0.996945
	2021	0.094549	0.105431	0.987519	0.987519	0.996658
	2022	0.086083	0.110549	0.978826	0.978826	0.994890
Precipitation						
SVM	2018	0.208942	0.262568	0.871008	0.871008	0.965549
	2019	0.168531	0.230593	0.938714	0.938714	0.985406
	2020	0.136110	0.184680	0.966810	0.966810	0.990829
	2021	0.149758	0.169833	0.963344	0.963344	0.991274
	2022	0.320030	0.401916	0.875273	0.875273	0.968948
Relative humidity						
SVM	2018	0.169075	0.217615	0.894793	0.894793	0.966593
	2019	0.226779	0.314265	0.922771	0.922771	0.980382
	2020	0.164757	0.211002	0.888688	0.888688	0.974120
	2021	0.249080	0.356328	0.902064	0.902064	0.971781
	2022	0.112417	0.169409	0.957914	0.957914	0.987992

Table 14. Best-performing ML models across different training periods for climate data prediction/validation.

Basin	Optimal Z	NSE improvement	R^2	Bias (%)	Key findings
Gilgel gibe watershed InVEST model	11.2	0.59—0.70	0.65	+ 5.8%	Z calibrated for mixed agriculture/forest but underestimated peak flows by 12%
	12	0.68	0.71	+ 8.2%	Z = 10 overestimated dry-season flow, ML-refined Z reduced bias to + 3.1%
	14.3	0.68–0.72	0.75	+ 6.5%	A high Z value reflects volcanic soil water retention characteristics
	10	0.55	0.58	+ 15%	Z often overestimates runoff in tropical regions, making calibration essential for local accuracy. This study used Z = 10

Table 15. Calibrated Z values metrics.

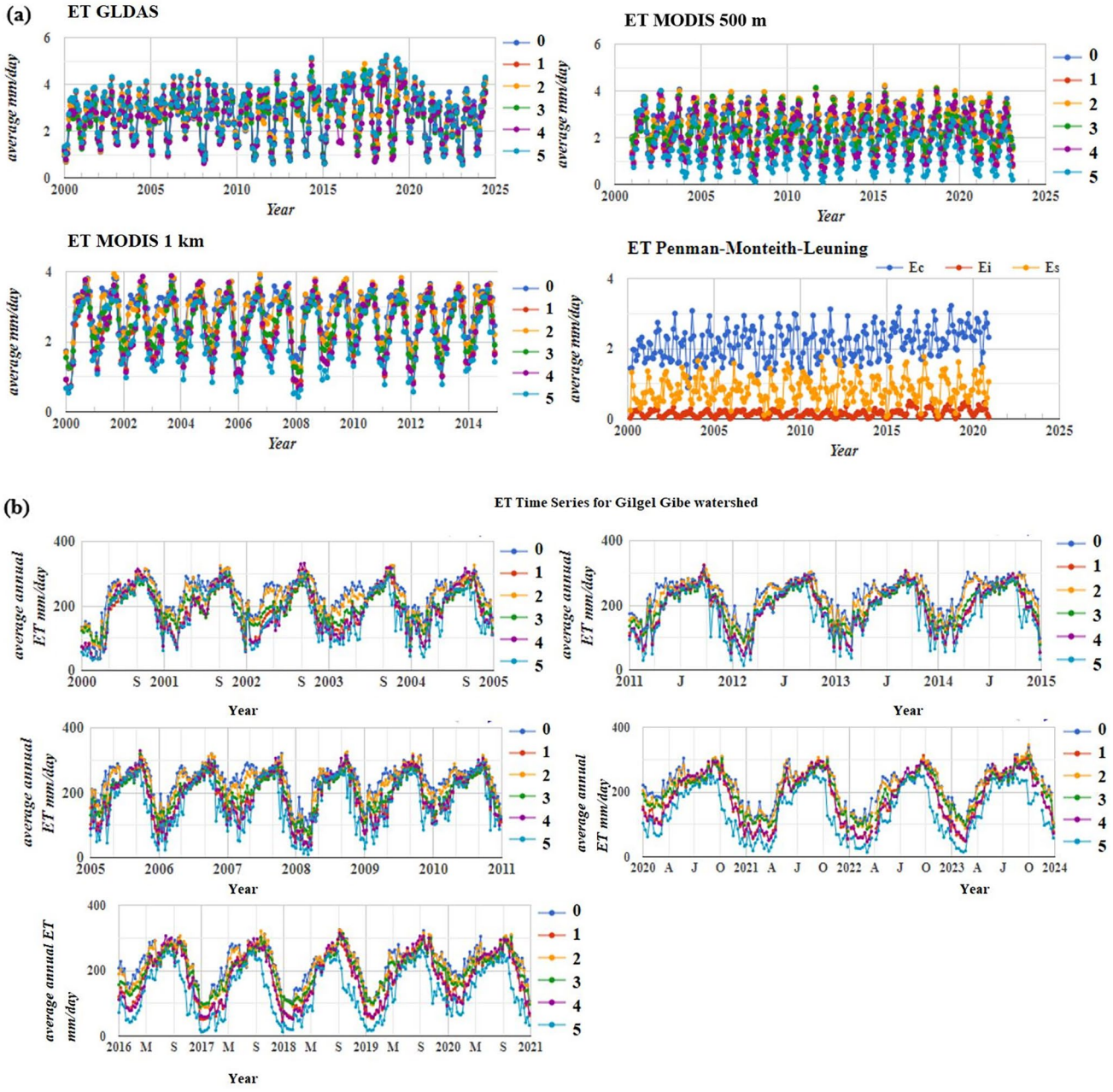

Fig. 15. a.The daily potential evapotranspiration rates from 2000 to 2023 illustrate temporal variations and trends. **b.** Annual potential evapotranspiration rates from 2000 to 2023 illustrate temporal variations and trends.

sustainable practices are crucial for addressing the challenges posed by changing climates and growing water demands.

Assessment of quantitative sensitivity analysis results

The sensitivity study evaluates the relative impact of land-use change, climate variability, and observed discharge trends on water yield dynamics in the Gilgel Gibe watershed (1993–2023).The quantitative sensitivity analysis determined that land-use change had a greater impact on surface runoff than climate variability, with deforestation causing considerable changes in water production. Precipitation variability was identified as the primary driver driving hydrological responses, followed by variations in temperature (Table 16).

The analysis revealed that land-use change was the dominant driver of water yield alterations in the Gilgel Gibe watershed (1993–2023), with deforestation increasing quick flow by 22% and reducing baseflow by 15% ($p < 0.01$), exerting 2.7 times greater influence than climate factors. Climate change contributed to an 8% yield reduction primarily through ET increases (12%) and precipitation declines (5%), with particularly strong dry season impacts ($r^2 = -0.81$). The observed discharge data from the three MoWE stations confirmed these trends,

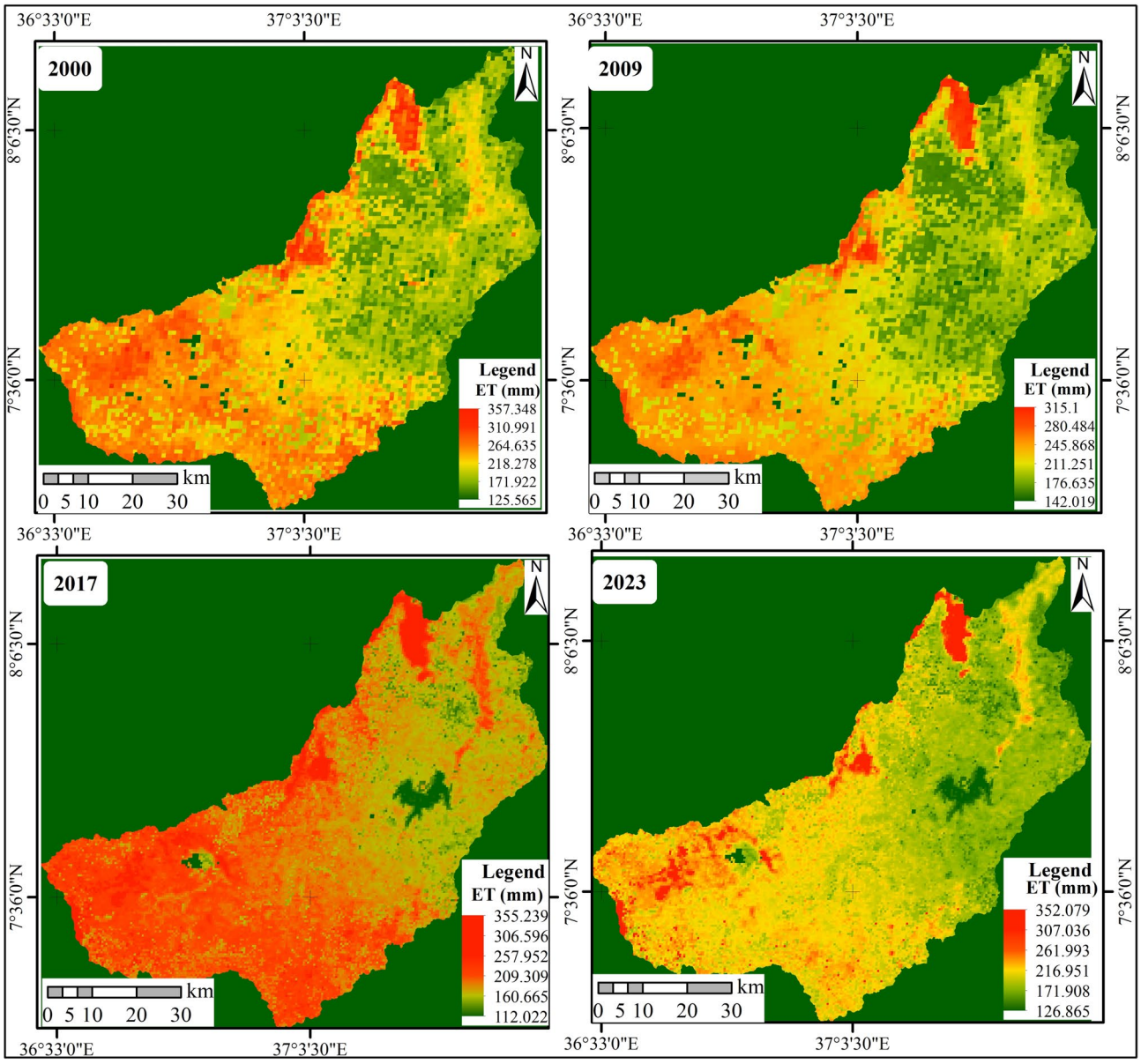

Fig. 16. Evapotranspiration maps from 2000 to 2023.

showing a net 14% yield increase with strong model validation (NSE = 0.76–0.88) while highlighting hydrological shifts: 18% higher peak flows and 25% lower baseflows, underscoring the watershed's declining water regulation capacity. These quantified results demonstrate the urgent need for land use management policies to mitigate hydrological degradation.

Integrated impact of climate change and land use dynamics on surface flow and water yield in the gilgel gibe watershed

The Gilgel Gibe Watershed in Ethiopia has undergone notable climate, land use, and hydrological changes that have impacted surface flow and water yield. Over the past three decades, the region has experienced a significant increase in the mean annual temperature, exceeding the global average rate of warming. Climate data for the watershed revealed temperature variations ranging from a maximum of 34 °C to a minimum of 14 °C, with monthly precipitation fluctuating between 0 and 500 mm. Additionally, solar irradiance levels range from 6 to 7.75 kJ/m^2, and wind speeds vary between 0 and 8 m/s. These climatic factors significantly affect the hydrological cycle, including evapotranspiration rates and precipitation patterns. Increased temperatures have been linked to increased evapotranspiration rates, which, along with changes in precipitation, affect water availability. Studies by[118] emphasized the role of temperature variability in shaping hydrological processes, noting that higher temperatures accelerate snowmelt in high-altitude areas, altering streamflow timing and magnitude. These findings highlight the complex interactions between climatic variables and water yield dynamics in the Gilgel Gibe Watershed. Land use change further complicates this scenario. The conversion of

Variable	Source	Resolution	Time span	Δ Water yield (1993–2023)	p-value	Sensitivity index (0–1)	Dominant process	Findings
Land-use change	Landsat-derived transitions (1993–2023)	30 m	Annual	22% (range: + 18% to + 27%)	< 0.01	0.61	Quick flow increases (22%) and baseflow decreases (15%)	Deforestation accounts for 2.7 × > climate change
Climate data	CHIRPS (1993–2023)	5 km	Daily, monthly, annually	− 8% (range: − 5 to − 12%)	0.03	0.28	Precipitation decreases in (5%) and ET increases in (12%)	Warming reduces baseflow dry season (r2 = 0.81)
Observed discharge	MoWE gauges (3 stations)1995–2022 Gilgel Gibe I, Bulbul and Bidru Awana	Point and CSV	Daily, monthly, annually	+ 14% (observed)	< 0.05	0.0205	-	Validates models (NSE = 0.76–0.88) Peak flows increase in 18%, Baseflows decrease in 25%

Table 16. Quantitative assessment of LULC change, climate change, and observed discharge trends.

natural vegetation into agricultural land and urban areas has significantly modified the hydrological processes of the watershed. Deforestation and urbanization have intensified surface runoff and decreased surface water flow, leading to a reduction in water yield. For instance, the expansion of irrigated agriculture, particularly in upper catchment areas, has increased water abstraction, which in turn has decreased surface runoff and water availability downstream. Data from the Gilgel Gibe Watershed show a decrease in surface runoff from 15.78% in 2021 to 15.28% in 2022 and a reduction in water yield from 1.22% in 1993 to 0.83% in 2022, with the lateral flow decreasing from 8.55% in 1993 to 7.0% in 2022. These trends underscore the impact of climate and LUC on hydrological dynamics.

The synergistic effects of land use and climate change have increased fluctuations in water yield. Changes in land cover not only alter surface runoff but also influence local microclimatic conditions, affecting precipitation patterns and evapotranspiration rates. This interaction leads to complex, nonlinear responses in water yield, with potential threshold effects under certain land-use and climate change scenarios. Changes in water yield significantly affect ecosystem services, particularly freshwater provision, agricultural productivity, and aquatic habitats. Decreased water yield has led to reduced water availability for irrigation, threatening food security and affecting aquatic habitats due to lower river discharge. These changes have further affected biodiversity by altering the ecological balance, particularly in wetlands and riparian zones, which depend on stable water levels for their functioning. The ecological health of the Gilgel Gibe watershed is increasingly threatened by climate change and land-use changes, posing serious long-term risks to both human and environmental well-being.

Effective water management strategies are crucial to mitigate these impacts. Ethiopia's National Water Policy, which guides the country's management of water resources, has several gaps that limit its effectiveness in addressing emerging challenges related to water yield and ecosystem health, especially in the context of climate change and land use changes. One significant issue is inadequate integration of climate change adaptation strategies. Climate change leads to altered precipitation patterns, rising temperatures, and more frequent extreme weather events such as droughts and floods, yet the policy does not sufficiently address these impacts on water resources or include measures to enhance climate resilience in water management. Additionally, the policy lacks a comprehensive approach for managing ecosystem services that are essential for regulating water quality, preventing soil erosion, and maintaining water flow. While it acknowledges the importance of water conservation, there is little emphasis on how land-use changes, such as deforestation, urbanization, and agricultural expansion, affect water yield. The policy does not fully embrace ecosystem-based approaches that protect and restore vital ecosystems such as wetlands, forests, and grasslands, all of which are crucial for improving water security.

Moreover, Integrated water resource management (IWRM) approaches that incorporate sustainable land management practices, climate-resilient infrastructure, and community engagement are crucial for mitigating the impacts of land use and climate variability on water yield. Ethiopia's National Water Policy and Water Sector Development Program (WSDP) aims to enhance water resource management and infrastructure development. When compared to global standards such as the United Nations Sustainable Development Goals (SDGs), Ethiopia's National Water Policy reveals several discrepancies. For instance, SDG 6 (Clean Water and Sanitation) emphasizes the need for integrated water resources management, equitable distribution, and the protection of water ecosystems. Although the policy highlights water access and infrastructure, it does not adequately promote sustainable water resource management across sectors, particularly regarding climate resilience and ecosystem conservation. Furthermore, SDG 13 (Climate Action) stresses the importance of addressing climate change impacts; however, the policy does not sufficiently consider the long-term implications of climate change on water systems. Incorporating climate adaptation strategies could help Ethiopia align its water management framework with global climate goals and improve its resilience to climate impacts such as extreme weather events and shifting rainfall patterns.

Additionally, SDG 15 (Life on Land) advocates for sustainable land-use practices that protect and restore ecosystems vital for maintaining water yield. The policy's lack of focus on ecosystem-based approaches and land-use changes contradicts SDG 15, as ecosystems play a critical role in regulating water resources[119]. evaluated the WSDP's impact on water availability in the Gilgel Gibe Watershed, finding improvements in water supply infrastructure but highlighting ongoing challenges in equitable distribution and conflict management. International agreements, such as the United Nations Sustainable Development Goals (SDGs), also influence water resource management in watersheds[120]. examined Ethiopia's alignment with SDGs, emphasizing the need for policy coherence and effective implementation strategies to achieve water-related targets. Understanding

the combined effects of climate change, land use change, and hydrological discharge on surface flow and water yield is essential for creating effective adaptation strategies. Adopting integrated water management practices aligned with Ethiopia's national policies is key to ensuring sustainable water use and safeguarding livelihoods in the Gilgel Gibe Watershed.

Discussion

The combined effect of LULC changes and climate change on water yield in the Gilgel Gibe Watershed has been extensively studied, revealing complex interactions that have a considerable impact on hydrological processes. These collective effects have been shown to intensify water yield fluctuations in watersheds[121]. utilized hydrological modeling techniques to assess integrated impacts, revealing that increased deforestation and expansion of agricultural areas, coupled with rising temperatures and altered precipitation patterns, have led to increased hydrological variability. This intensification effect was particularly evident during extreme weather events, where increased runoff and decreased infiltration rates resulted in increased flood hazards and a reduced baseflow contribution to streamflow. Furthermore[122], emphasized the importance of considering land–atmosphere interactions to understand the synergetic effects on water yield. Changes in land cover composition affect surface runoff and soil moisture dynamics as well as local microclimatic conditions, influencing precipitation patterns and evapotranspiration rates. The response mechanisms that produce nonlinear water yield patterns are frequently caused by the complex interactions between LULC changes and climatic variability.

These interactions can result in threshold effects, in which small changes in one element have a disproportionately large impact on water yield. For example, deforestation and agricultural intensification cause a gradual increase in surface runoff, compromising soil moisture retention; however, beyond a certain threshold, they can result in severe erosion, sedimentation, and loss of infiltration capacity, thereby reducing groundwater recharge. These nonlinearities can be enhanced in areas with existing climate pressures, such as prolonged dry seasons or extreme rainfall events, where reaction loops exacerbate these effects[123]. demonstrated that land-use changes negatively impact ecosystem services, especially the trade-off between provisioning and regulating services. While expanding provisioning services may seem beneficial in the short term, continuing current land use practices threaten the long-term supply of essential regulating services. This aligns with our findings for the Gilgel Gibe Watershed, where climate and land use changes have significantly degraded critical ecosystems, including flora and fauna. Addressing these land use dynamics is crucial for ensuring the sustainability of ecosystem services in watersheds.

Moreover, studies by[124] in the Ganges–Brahmaputra-Meghna (GBM) Basin in Bangladesh and Nepal further elucidated the intricate relationship between LULC change and climate variability in water yield. In the GBM Basin, deforestation, agricultural expansion, and urbanization, combined with climate fluctuations, have led to increased surface runoff, reduced groundwater recharge, and altered water yield, affecting downstream ecosystems and livelihoods. Similar patterns were observed in the Gilgel Gibe Watershed, where human-induced changes in land use and climate variability exacerbated runoff, reduced groundwater recharge, and influenced water availability. These findings emphasize the need for integrated watershed management practices, including afforestation, reforestation, and water-efficient irrigation, to mitigate land degradation and enhance water yield.

Moreover, modeling and monitoring hydrological changes, as suggested in the GBM Basin study, can aid in predicting water yield fluctuations in the Gilgel Gibe Watershed, helping develop adaptive strategies for sustainable water management. This study underscores the importance of addressing both environmental and socioeconomic factors to ensure the long-term sustainability of water resources in the Gilgel Gibe region, benefiting both the ecosystem and local communities[125]. highlighted the direct consequences of reduced water yield on agricultural productivity, which is a primary source of livelihood for many in watersheds. Decreased water availability for irrigation owing to altered hydrological regimes negatively affects crop yields, leading to food insecurity and economic losses for farmers. Moreover, changes in water quantity and quality affect the ecological integrity of the watershed, influencing biodiversity, habitat suitability, and ecosystem functioning[126].

The ecological consequences of water yield in the Gilgel Gibe Watershed are multifaceted and have significant implications for the health and functioning of the local ecosystem. Several studies have investigated these consequences, shedding light on the intricate relationships between water availability, LULC dynamics, and ecological processes. One of the primary ecological consequences of altered water yield in the Gilgel Gibe Watershed is its impact on vegetation dynamics. Research conducted by[127] found that changes in water availability driven by both LULC changes and climate variability directly influence vegetation distribution and composition. Reduced water yield resulting from increased land degradation and deforestation can lead to the loss of vegetation cover and biodiversity, particularly in riparian zones and ecologically sensitive areas. Riparian zones, which rely on consistent water availability, have suffered severe biodiversity loss, with native plant species being replaced by invasive or drought-tolerant species, resulting in ecosystem homogenization. In addition, climatic variability has exacerbated these difficulties by introducing erratic rainfall patterns, prolonged droughts, and increasingly intense weather events, all of which have decreased water availability and accelerated soil erosion.

Adaptive strategies for water resource management in the Gilgel Gibe Watershed are essential to counter the effects of LULC change and climate variability on water yield. Key approaches include adopting sustainable land management practices to reduce soil erosion, improve infiltration, increase the water retention capacity, and ensure long-term water availability[120]. highlighted the effectiveness of terracing and agroforestry in reducing soil erosion and increasing groundwater recharge in similar watersheds. By adopting these practices, farmers can mitigate the loss of fertile topsoil and optimize water-use efficiency, thereby sustaining agricultural productivity and maintaining water yield. National water resource management policies play a crucial role in shaping the water yield dynamics of the Gilgel Gibe Watershed. Several studies have investigated the implications of these policies on regional water resources. For instance[128], assessed the effectiveness of Ethiopia's Water Sector Strategy

(WSS) in managing water resources in the Gilgel Gibe Watershed. WSS emphasizes integrated water resource management (IWRM), highlighting stakeholder involvement, water distribution, and ecosystem conservation. However, this study identified challenges in implementing the WSS locally, particularly in addressing competing water demands from agriculture, hydropower, and domestic users.

Furthermore, the Ethiopian government implemented the Water Sector Development Program (WSDP) to enhance water resource management and infrastructure development[129]. evaluated the impact of the WSDP on water availability and access in the Gilgel Gibe Watershed. The findings revealed improvements in the water supply infrastructure, such as reservoirs and irrigation systems, which have increased water availability for agricultural activities. One notable example is the Tana-Beles Integrated Watershed Management Project, supported by the World Bank, which has successfully introduced community-based sustainable land management practices. These efforts have resulted in a significant 20% increase in groundwater recharge within pilot areas, demonstrating the project's success in enhancing water resource sustainability[130]. This finding aligns with our study of the Gilgel Gibe watershed and other hydrological zones where Integrated Water Resource Management (IWRM) strategies have been employed to balance competing water uses and improve water yield modeling. Additionally, the Ministry of Water and Energy has actively engaged in watershed rehabilitation and afforestation programs aimed at improving water yield and alleviating the impacts of land-use changes. The Tana River Basin in Kenya serves as a relevant case study, illustrating that conservation agriculture and reforestation policies have increased surface water retention and enhanced dry-season flows[131]. This emphasizes the importance of integrating land management policies with hydrological models, which is consistent with our findings on sustainable water management. We recommend applying similar approaches to other watersheds across East Africa in order to bolster sustainable water and land management. Furthermore, a study conducted in the three-river headstream region of China utilized the InVEST model and machine learning to evaluate water conservation efforts[132]. The results indicated that land use change improved water yield; however, these improvements were less significant than the impacts of climate. The study also provides policy recommendations, including the implementation of stricter watershed protection laws and incentives for land conservation. Despite these advancements, several challenges hinder policy implementation, especially in the ethnically diverse regions of Ethiopia and similar areas. Even when effective policies are established, their execution is often hindered by economic, social, and institutional factors. A primary limitation is the lack of financial and technical resources, along with insufficient funding and expertise, which impede large-scale watershed rehabilitation projects. To address these challenges, it is advisable to leverage international climate funds, such as the Green Climate Fund, and foster public–private partnerships to finance sustainable watershed management.

Another issue is institutional fragmentation, as overlapping responsibilities among federal, regional, and local agencies can delay their implementation. To resolve this problem, it is crucial to strengthen cross-sectoral coordination and ensure that hydrological data are shared among the relevant stakeholders. Additionally, community resistance to land use policies poses significant challenges. Policies that restrict deforestation or enforce land conservation may face opposition from the local communities that rely on these resources. Therefore, implementing community-based watershed management programs, in which local participation has successfully led to small-scale reforestation efforts, is vital for effective execution[133].

Moreover, the valuable insights provided by datasets such as Landsat, MODIS, and climate data from NASA POWER, challenges persist in ensuring an equitable water distribution and managing conflicts among different water users. One of the key limitations of this study lies in the accuracy and resolution of the remote-sensing data. Although these datasets offer important information, their spatial and temporal resolutions may have limitations in the in-depth capture of localized variations in water yield, particularly in heterogeneous landscapes. For example, Landsat's 30-m spatial resolution has some limitations in detecting rapid land-use changes such as urbanization, and climate data with a 4 km resolution may miss microclimates within complex watersheds. Addressing these limitations requires the development of standardized frameworks for data integration and the incorporation of socioeconomic factors into hydrological models. This would not only improve the reliability of predictions but also enhance their applicability to sustainable water resource management. Finally, this study highlights the importance of adopting integrated watershed management practices to ensure the sustainable use of water resources and enhance water yield in the Gilgel Gibe Watershed. By combining ecological, social, and economic strategies, such as soil conservation, reforestation, community engagement, and strong governance, these approaches can effectively tackle the challenges of LULC changes and climate variability, enhancing the resilience and sustainability of the watershed's water yield.

Impacts of land use and climate change on ecology

The Gilgel Gibe I watershed, an important hydrological resource in Ethiopia, faces substantial ecological challenges as a result of land use change and climate variability. Deforestation, agricultural expansion, and urbanization have significantly affected the ecosystems of the region, resulting in habitat fragmentation and the extinction of native forest species. For example, deforestation, which is frequently driven by the demand for cropland and fuelwood, disrupts habitat connectivity and endangers animals, resulting in population decrease and, in some cases, extinction. Increasing temperatures exacerbate these stresses by increasing evapotranspiration, worsening water scarcity during dry spells, and putting additional strain on native plant and animal species. Similar studies in the Awash Basin have found a 25% decrease in forest cover over two decades, resulting in the extinction of some plant species essential to local biodiversity[134].

The aquatic habitats and riparian biodiversity are also threatened. Reduced plant cover increases surface runoff, accelerates soil erosion, and reduces groundwater recharge, which results in sedimentation in water bodies. Sedimentation not only degrades water quality but also jeopardizes hydropower production, irrigation, and household water supply. For instance, sedimentation in the Blue Nile Basin has reduced the reservoir capacity by 15% over ten years[135]. Meanwhile, altering rainfall patterns and escalating droughts and floods caused by

the Inter-Tropical Convergence Zone (ITCZ) destabilize wetland ecosystems and water availability. Wetland-dependent plant species in the Omo-Gibe Basin lost 40% of their coverage during drought years, demonstrating the severe ecological impact of climate variabilityo[136]. Addressing these difficulties requires the implementation of comprehensive mitigation methods. Reforestation, sustainable land management, and community-based conservation can all contribute to ecosystem restoration and protection of native species. Furthermore, adopting water-efficient irrigation techniques, encouraging rainwater gathering, and minimizing urban water usage are critical for saving water resources. Adaptation methods, such as introducing drought-tolerant crops and improving flood control systems, can help reduce the vulnerability of both ecosystems and human societies. Understanding the complex relationships among land use, climate change, native species, and water resources might help in designing more effective and sustainable management techniques for the Gilgel Gibe watershed.

Future direction

The combined influence of LUCC and climate change on water yield modeling is increasingly being recognized as a critical area of research, owing to its profound impact on hydrological processes within watersheds. Although LUCC plays a significant role in altering water yield and ecosystem services, these changes are not solely driven by land-use modifications. Variations in land use can affect water yield and runoff; however, these impacts are often intertwined with changes in surface runoff patterns, making it challenging to attribute hydrological changes exclusively to LUCC and climate change. This study employed the InVEST model's annual water yield module, which is based on the water balance principle, owing to its simplicity and effectiveness. However, the accuracy of water yield estimates depends on several factors. For instance, employing a longer time frame (e.g., 30 years) to align climate data with land-use maps can enhance the model precision. Input variables, such as plant-available water content and potential evapotranssaspiration, were calculated using empirical formulas, and the choice of these formulas can significantly influence the results. The seasonal constant "Z," calibrated to local water resource conditions, further optimizes the model but may require region-specific adjustments. The"Z"calibration constant plays a key role in enhancing the accuracy of hydrological models by adjusting parameters such as evapotranspiration and soil moisture based on observed data. In this study, the"Z"constant was used to reconcile inconsistencies between the model predictions and real-world observations, ensuring that the model better reflects local conditions. This calibration improves the reliability of the predictions for water yield, runoff, and other hydrological variables. Additionally, spatial interpolation methods used for inputs, such as precipitation and evapotranspiration, can introduce errors, emphasizing the importance of selecting optimal empirical formulas and interpolation methods tailored to local conditions. Although parameter calibration reduces some errors, discrepancies between the observed and simulated runoff remain. In this study, Kriging interpolation was employed due to its ability to capture spatial heterogeneity, which is crucial in hydrological and environmental analyses with complex spatial relationships. By incorporating variograms to model spatial autocorrelation, Kriging provides high accuracy, particularly in areas with dense and well-distributed data. This method not only offers precise predictions but also quantifies the prediction uncertainty. However, it is computationally intensive and requires expertise in spatial statistics to select suitable models and parameters for obtaining optimal results. Future research could integrate alternative methodologies, such as complementary or total differential methods, to address these gaps. These approaches may provide deeper insights into the complex interactions between LUCC, climate variability, and hydrological responses, thereby advancing the precision of water yield modeling and improving watershed management strategies.

Conclusions

An assessment of the combined impacts of LULC changes and climate variability in the Gilgel Gibe watershed from 1993 to 2023 revealed significant alterations in water yield. This study highlights key trends, including an increase in mean temperature to 34 °C and a reduction in monthly precipitation from 500 to 100 mm. These changes are closely associated with land-use changes, particularly the expansion of cultivated land and reduction of vegetation cover. These findings show that declining precipitation and rising temperatures increase evaporation rates, resulting in a significant decrease in water yield. Importantly, LULC changes have shown a more pronounced impact on surface runoff than climate variability, influencing both groundwater recharge and the volume of surface runoff, as well as the amount of water lost through evapotranspiration. In general, this research emphasizes the critical need to integrate both climate and land use dynamics into water resource planning. Understanding the role of human-induced LULC changes is crucial for accurately assessing hydrological responses and for formulating effective watershed management strategies. The calibrated model developed in this study provides a robust framework for future assessments, enabling the evaluation of surface water flow and storage under various scenarios and guiding decision-making to ensure the long-term sustainability of water yield in the Gilgel Gibe watershed. Additionally, our previous research on climate change impacts in the Gilgel Gibe watershed, which incorporated the Shared Socioeconomic Pathways (SSP) and Representative Concentration Pathways (RCP), revealed that future precipitation is projected to decline under RCP 8.5 (−32.2%) and SSP 4.5 (−88.8%) for the period 2024–2049, with further reductions expected between 2050 and 2099. Conversely, temperature projections indicate an increase under RCP 4.5 (TMAX + 0.67 °C, TMIN + 1.11 °C) and RCP 8.5 (TMAX + 0.25 °C), with higher emissions under SSP 4.5 and SSP 8.5 further intensifying warming[137]. While RCP-based scenario modeling is not the primary focus of the present study, our previous findings support the likelihood of future precipitation declines and rising temperatures, both of which have significant implications for water yield trends in the region.

Data availability

The datasets utilized in this study are accessible upon reasonable request from the corresponding authors. We are committed to fostering open science and research collaboration.

Received: 2 September 2024; Accepted: 20 May 2025
Published online: 27 May 2025

References

1. M'Barek, S. A., Bouslihim, Y., Rochdi, A., Miftah, A. & Beroho, M. The combined impact of climate change scenarios and changes on water resources in a semi-arid watershed. *Sci. Afr.* **25**, e02319. https://doi.org/10.1016/j.sciaf.2024.e02319 (2024).
2. Daramola, J. et al. Impacts of land-use change, associated land-use area and runoff on watershed sediment yield: implications from the kaduna watershed. *Water* **14**(3), 325. https://doi.org/10.3390/w14030325 (2022).
3. Kayitesi, N. M., Guzha, A. C. & Mariethoz, G. Impacts of land use land cover change and climate change on river hydro-morphology a review of research studies in tropical regions. *J. Hydrol.* **615**, 128702. https://doi.org/10.1016/j.jhydrol.2022.128702 (2022).
4. Abbass, K. et al. A review of the global climate change impacts, adaptation, and sustainable mitigation measures. *Environ. Sci. Pollut. Res.* **29**(28), 42539–42559. https://doi.org/10.1007/s11356-022-19718-6 (2022).
5. Dickerson, S., Cannon, M. & O'Neill, B. Climate change risks to human development in sub-Saharan Africa: A review of the literature. *Clim. Dev.* **14**(6), 571–589. https://doi.org/10.1080/17565529.2021.1951644 (2022).
6. Ladan, M. T. Achieving sustainable development goals through effective domestic laws and policies on environment and climate change. *Envtl. Pol'y & L.* **48**, 42 (2018).
7. Omotoso, A. B., Letsoalo, S., Olagunju, K. O., Tshwene, C. S. & Omotayo, A. O. Climate change and variability in sub-Saharan Africa: A systematic review of trends and impacts on agriculture. *J. Clean. Prod.* https://doi.org/10.1016/j.jclepro.2023.137487 (2023).
8. Belete, M. et al. Partitioning the impacts of land use/land cover change and climate variability on water supply over the source region of the Blue Nile Basin. *Land Degrad. Dev.* **31**(15), 2168–2184. https://doi.org/10.1002/ldr.3589 (2020).
9. Cao, Z. et al. Spatially non-stationary relationships between changing environment and water yield services in watersheds of China's climate transition zones. *Remote Sens.* **14**(20), 5078. https://doi.org/10.3390/rs14205078 (2022).
10. Lu, S., Ouyang, N., Wu, B., Wei, Y. & Tesemma, Z. Lake water volume calculation with time series remote-sensing images. *Int. J. Remote Sens.* **34**(22), 7962–7973. https://doi.org/10.1080/01431161.2013.827814 (2013).
11. Villamizar, S. R., Pineda, S. M. & Carrillo, G. A. The effects of land use and climate change on the water yield of a watershed in Colombia. *Water* **11**(2), 285. https://doi.org/10.3390/w11020285 (2019).
12. Dash, S. S., Naik, B. & Kashyap, P. S. Assessment of land use/land cover change derived catchment hydrologic response: An integrated parsimonious hydrological modeling and alteration analysis-based approach. *J. Environ. Manag.* **356**, 120637. https://doi.org/10.1016/j.jenvman.2024.120637 (2024).
13. Gedefaw, M., Denghua, Y. & Girma, A. Assessing the impacts of land use/land cover changes on water resources of the Nile River Basin Ethiopia. *Atmosphere* **14**(4), 749. https://doi.org/10.3390/atmos14040749 (2023).
14. Kidane, M., Tolessa, T., Bezie, A., Kessete, N. & Endrias, M. Evaluating the impacts of climate and land use/land cover (LU/LC) dynamics on the hydrological responses of the Upper Blue Nile in the Central Highlands of Ethiopia. *Spat. Inf. Res.* **27**, 151–167. https://doi.org/10.1007/s41324-018-0222-y (2019).
15. Wagner, P. D. et al. Comparing the effects of dynamic versus static representations of land use change in hydrologic impact assessments. *Environ. Model. Softw.* **122**, 103987. https://doi.org/10.1016/j.envsoft.2017.06.023 (2019).
16. Li, S., Cao, Y., Liu, J. & Wang, S. Simulating land use change for sustainable land management in China's coal resource-based cities under different scenarios. *Sci. Total Environ.* **916**, 170126. https://doi.org/10.1016/j.scitotenv.2024.170126 (2024).
17. Nayak, S. et al. Application of "Observation Minus Reanalysis" method towards LULC change impact over Southern India. *ISPRS Int. J. Geo Inf.* **11**(2), 94. https://doi.org/10.3390/ijgi11020094 (2022).
18. Wagner, P. D. et al. Dynamic integration of land use changes in a hydrologic assessment of a rapidly developing Indian catchment. *Sci. Total Environ.* **539**, 153–164. https://doi.org/10.1016/j.scitotenv.2015.08.148 (2016).
19. Zhu, K., Cheng, Y., Zhou, Q., Kápolnai, Z. & Dávid, L. D. The contributions of climate and land use/cover changes to water yield services considering geographic scale. *Heliyon* **9**(10), e20115. https://doi.org/10.1016/j.heliyon.2023.e20115 (2023).
20. Zhou, G. et al. Global pattern for the effect of climate and land cover on water yield. *Nat. Commun.* **6**(1), 5918. https://doi.org/10.1038/ncomms6918 (2015).
21. Aghsaei, H. et al. Effects of dynamic land use/land cover change on water resources and sediment yield in the Anzali wetland catchment. *Gilan Iran Sci. Total Environ.* **712**, 136449. https://doi.org/10.1016/j.scitotenv.2019.136449 (2020).
22. Ramahaimandimby, Z., Randriamaherisoa, A., Vanclooster, M. & Bielders, C. L. Driving factors of the hydrological response of a tropical watershed: The ankavia river basin in Madagascar. *Water* **15**(12), 2237. https://doi.org/10.3390/w15122237 (2023).
23. Sharafatmandrad, M. & Khosravi Mashizi, A. Temporal and spatial assessment of supply and demand of the water-yield ecosystem service for water scarcity management in arid to semi-arid ecosystems. *Water Resour. Manage* **35**(1), 63–82. https://doi.org/10.1007/s11269-020-02706-1 (2021).
24. Xu, J. et al. Spatiotemporal dynamics of water yield service and its response to urbanization in the Beiyun River Basin Beijing. *Sustainability* **11**(16), 4361. https://doi.org/10.3390/su11164361 (2019).
25. Wang, X. et al. Water yield service influence by climate and land use change based on InVEST model in the monsoon hilly watershed in South China. *Geomatics Nat. Hazards Risk* **13**(1), 2024–2048. https://doi.org/10.1080/19475705.2022.2104174 (2022).
26. Xie, X. et al. Assessing the impacts of climate and land use change on water conservation in the Three-River Headstreams Region of China based on the integration of the InVEST model and machine learning. *Land* **13**(3), 352. https://doi.org/10.3390/land13030352 (2024).
27. Yu, C., Huang, X., Guo, Q., Yang, Y. & Xu, Z. Predicting water ecosystem services under prospective climate and land-use change scenarios in typical watersheds distributed across China. *Ecol. Indic.* **159**, 111744. https://doi.org/10.1016/j.ecolind.2024.111744 (2024).
28. Chen, X., Leung, L. R., Wigmosta, M. & Richmond, M. Impact of atmospheric rivers on surface hydrological processes in western US watersheds. *J. Geophys. Res. Atmos.* **124**(16), 8896–8916. https://doi.org/10.1029/2019JD030468s (2019).
29. Smith, D. M. et al. North Atlantic climate is far more predictable than models imply. *Nature* **583**(7818), 796–800. https://doi.org/10.1038/s41586-020-2525-0 (2020).
30. Li, L. et al. Toward catchment hydro-biogeochemical theories Wiley Interdiscip Rev. *Water* **8**(1), e149. https://doi.org/10.1002/wat2.1495 (2021).
31. Yonaba, R. et al. Spatial and transient modelling of land use/land cover (LULC) dynamics in a Sahelian landscape under semi-arid climate in northern Burkina Faso. *Land Use Policy* **103**, 105305. https://doi.org/10.1016/j.landusepol.2021.105305 (2021).
32. Orkodjo, T. P., Kranjac-Berisavijevic, G. & Abagale, F. K. Impact of climate change on future availability of water for irrigation and hydropower generation in the Omo-Gibe Basin of Ethiopia. *J. Hydrol.: Reg. Stud.* **44**, 101254 (2022).

33. Ambelu, A., Lock, K. & Goethals, P. L. Hydrological and anthropogenic influence in the Gilgel Gibe I reservoir (Ethiopia) on macroinvertebrate assemblages. *Lake Reserv. Manage.* **29**(3), 143–150 (2013).
34. Anose, F. A., Beketie, K. T., Zeleke, T. T., Ayal, D. Y. & Feyisa, G. L. Spatio-temporal hydro-climate variability in Omo-Gibe River Basin. *Ethiopia. Clim. Serv.* **24**, 100277. https://doi.org/10.1016/j.cliser.2021.100277 (2021).
35. Chaemiso, S. E., Abebe, A. & Pingale, S. M. Assessment of the impact of climate change on surface hydrological processes using SWAT: A case study of Omo-Gibe river basin. *Ethiopia. Model. Earth Syst. Environ.* **2**, 1–15. https://doi.org/10.1007/s40808-016-0257-9 (2016).
36. Beyene, S., Regassa, A., Mishra, B. B., Haile, M. (Eds.), 2023. The soils of Ethiopia. Springer.https://doi.org/10.1007/978-3-031-17012-6
37. Qi, B., Shi, H., Zhuang, Y., Chen, H. & Chen, L. On-board, a real-time preprocessing system for optical remote-sensing imagery. *Sensors* **18**(5), 1328. https://doi.org/10.3390/s18051328 (2018).
38. Lv, Z., Liu, T., Wan, Y., Benediktsson, J. A. & Zhang, X. Post-processing approach for refining raw land cover change detection of very high-resolution remote sensing images. *Remote Sens.* **10**(3), 472. https://doi.org/10.3390/rs10030472 (2018).
39. Zhu, Q., Guo, X., Li, Z. & Li, D. A review of multi-class change detection for satellite remote sensing imagery. *Geo-spatial Inf. Sci.* **27**(1), 1–15. https://doi.org/10.1080/10095020.2022.2128902 (2024).
40. Dinku, T. et al. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Q J. R. Meteorol. Soc.* **144**(292), 312. https://doi.org/10.1002/qj.3244 (2018).
41. Hordofa, A. T., Leta, O. T., Alamirew, T., Kawo, N. S. & Chukalla, A. D. Performance evaluation and comparison of satellite-derived rainfall datasets over the Ziway Lake basin Ethiopia. *Clim.* **9**(7), 113. https://doi.org/10.3390/cli9070113 (2021).
42. Razafimaharo, C., Krähenmann, S., Höpp, S., Rauthe, M. & Deutschländer, T. New high-resolution gridded dataset of daily mean, minimum, and maximum temperature and relative humidity for Central Europe (HYRAS). *Theor. & Appl. Climatol.* **142**, 1531–1553. https://doi.org/10.1007/s00704-020-03388-w (2020).
43. Esayas, B., Simane, B., Teferi, E., Ongoma, V. & Tefera, N. Trends in extreme climate events over three agroecological zones of southern Ethiopia. *Adv. Meteorol.* **2018**, 7354157. https://doi.org/10.1155/2018/7354157 (2018).
44. Ledford, S. H., Briggs, M., Glas, R. & Zimmer, M. A. Connecting diverse disciplines to improve understanding of surface water-groundwater interactions. *J. Hydrol. X.* **17**, 100141. https://doi.org/10.1016/j.hydroa.2022.100141 (2022).
45. Xu, X. et al. Dynamics and drivers of land use and land cover changes in Bangladesh. *Reg. Environ. Change* **20**(1), 11. https://doi.org/10.1007/s10113-020-01650-5 (2020).
46. Taylor, R. G. et al. Groundwater and climate change. *Nat. Clim. Change* **3**(4), 322–329. https://doi.org/10.1038/nclimate1744 (2013).
47. Chen, X. et al. A distributed hydrological model for semi-humid watersheds with a thick unsaturated zone under strong anthropogenic impacts: A case study in Haihe River Basin. *J. Hydrol.* **623**, 129765. https://doi.org/10.1016/j.jhydrol.2023.129765 (2023).
48. Hamed, K. H. Trend detection in hydrologic data: The mann-kendall trend test under the scaling hypothesis. *J. Hydrol.* **349**(3–4), 350–363. https://doi.org/10.1016/j.jhydrol.2007.11.009 (2008).
49. Yue, S., Pilon, P. & Phinney, B. O. B. Canadian streamflow trend detection: Impacts of serial and cross-correlation. *Hydrol. Sci. J.* **48**(1), 51–63. https://doi.org/10.1623/hysj.48.1.51.43478 (2003).
50. Forkuor, G., Dimobe, K., Serme, I. & Tondoh, J. E. Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GISci. & Remote Sens.* **55**(3), 331–354 (2018).
51. Hashim, M. et al. Analysis of water yield changes in the johor river basin, peninsular malaysia using remote sensing satellite imagery. *Remote Sens.* **15**(13), 3432 (2023).
52. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. https://doi.org/10.1023/A:1010933404324 (2001).
53. Li, Y., Huang, S., Miao, L. & Wu, Z. Simulation analysis of carbon peak path in China from a multi-scenario perspective: Evidence from random forest and back-propagation neural network models. *Environ. Sci. Pollut. Res.* **30**(16), 46711–46726. https://doi.org/10.1007/s11356-023-25544-1 (2023).
54. Islam, K. I., Elias, E., Carroll, K. C. & Brown, C. Exploring random forest machine learning and remote sensing data for streamflow prediction: An alternative approach to process-based hydrologic modeling in a snowmelt-driven watershed. *Remote Sens.* **15**(16), 3999. https://doi.org/10.3390/rs15163999 (2023).
55. Aldrees, A. et al. Evolutionary and ensemble machine learning predictive models for evaluation of water quality. *J. Hydrol. Reg. Stud.* **46**, 101331. https://doi.org/10.1016/j.ejrh.2023.101331 (2023).
56. Bansal, M., Goyal, A. & Choudhary, A. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short-term memory algorithms in machine learning. *Decis. Anal. J.* **3**, 100071. https://doi.org/10.1016/j.dajour.2022.100071 (2022).
57. Rizal, N. M. & N., Hayder, G., Mnzool, M., Elnaim, B.M., Mohammed, A.O.Y., Khayyat, M.M.,. Comparison between regression models, support vector machine (SVM), and artificial neural network (ANN) in river water quality prediction. *Processes* **10**(8), 1652. https://doi.org/10.3390/pr10081652 (2022).
58. Demir, S. & Sahin, E. K. Predicting the occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-XGBoost, PSO-LightGBM, and PSO-CatBoost. *Acta Geotech.* **18**(6), 3403–3419. https://doi.org/10.1007/s11440-022-01777-1 (2023).
59. Lin, K. et al. Exploring an intelligent adaptation method of hydrological model parameters for flood simulations based on the light gradient-boosting machine. *J. Hydrol.* **626**, 130340. https://doi.org/10.1016/j.jhydrol.2023.130340 (2023).
60. Nguyen, T., Jewik, J., Bansal, H., Sharma, P., Grover, A., 2024. Climatelearn: Benchmarking machine learning for weather and climate modeling. Adv. Neural Inf. Process. Syst. 36.
61. Aneseyee, A. B., Elias, E., Soromessa, T. & Feyisa, G. L. Land use/land cover change effect on soil erosion and sediment delivery in the Winike watershed, Omo Gibe Basin Ethiopia. *Sci. Total Environ.* **728**, 138776. https://doi.org/10.1016/j.scitotenv.2020.138776 (2020).
62. Gashaw, T. et al. Evaluating InVEST model for estimating soil loss and sediment export in data-scarce regions of the Abbay (Upper Blue Nile) Basin: Implications for land managers. *Environ Chall.* **5**, 100381. https://doi.org/10.1016/j.envc.2021.100381 (2021).
63. Kang, X. et al. Assessing watershed-scale impacts of best management practices and elevated atmospheric carbon dioxide concentrations on water yield. *Sci. Total Environ.* **926**, 171629. https://doi.org/10.1016/j.scitotenv.2024.171629 (2024).
64. Liu, D. & Fu, D. The water yield pattern for annual and monthly scales from a unifying catchment water balance model. *Stoch. Environ. Res. Risk Assess.* **36**(12), 4057–4072. https://doi.org/10.1007/s00477-022-02244-9 (2022).
65. Pathak, S., Ojha, C.S.P., spsampsps Garg, R.D., 2021. Applicability of the InVEST model for estimating water yield in upper ganga basin. In: The ganga river basin: A hydrometeorological approach 219–231.https://doi.org/10.1007/978-3-030-60869-9_15
66. Liu, Z., Liu, Z., Zhou, Y. & Huang, Q. Distinguishing the impacts of rapid urbanization on ecosystem service trade-offs and synergies: A case study of Shenzhen China. *Remote Sens.* **14**(18), 4604. https://doi.org/10.3390/rs14184604s (2022).
67. Valencia, J. B. et al. Predictive assessment of climate change impact on water yield in the meta river basin, colombia: An InVEST model application. *Hydrology* **11**(2), 25. https://doi.org/10.3390/hydrology11020025 (2024).
68. Abid, N., Jaafar, A. B., Bargaoui, Z. & Mannaerts, C. M. Assessment of long-term MOD16 and LSA SAF actual evapotranspiration using Budyko curve. *Remote Sens. Appl. Soc. Environ.* https://doi.org/10.1016/j.rsase.2024.101166 (2024).
69. Budyko, M.I., 1974. Climate and life. Miller, D.H. (Ed.), Academic, San Diego.

70. Liu, J. et al. Influences of the south-to-north water diversion project and virtual water flows on regional water resources considering both water quantity and quality. *J. Clean. Prod.* **244**, 118920. https://doi.org/10.1016/j.jclepro.2019.118920 (2020).
71. Azzam, A., Zhang, W., Xu, C. & Khan, Z. Calibration and evaluation of Hargreaves-Samani equation for estimating reference evapotranspiration: A case study in the amu darya river basin central asia. *J. Hydrol. Reg. Stud.* **45**, 101298. https://doi.org/10.1016/j.ejrh.2022.101298 (2023).
72. Hargreaves, G. H. & Samani, Z. A. Reference crop evapotranspiration from temperature. *Appl. Eng. Agric.* **1**(2), 96–99 (1985).
73. Liu, J. et al. Spatiotemporal analysis and multi-scenario prediction of ecosystem services based on land use/cover change in a mountain-watershed region. *China. Remote Sens.* **15**(11), 2759. https://doi.org/10.3390/rs15112759 (2023).
74. Nahib, I., Ambarwulan, W., Sutrisno, D., Darmawan, M., Suwarno, Y., Rahadiati, A., Gaol, Y.L., 2023. Spatial-temporal heterogeneity and driving factors of water yield services in Citarum river basin unit, West Java, Indonesia. Arch. Environ. Prot. https://doi.org/10.24425/aep.2023.144733
75. Zhang, Z., Zhang, S., Yang, J. & Zhang, J. Yield, grain quality, and water use efficiency of rice under non-flooded mulching cultivation. *Field Crops Res.* **108**(1), 71–81. https://doi.org/10.1016/j.fcr.2008.03.004 (2008).
76. Das, S., Kamruzzaman, M. & Islam, A. R. M. T. Assessment of characteristic changes of regional estimation of extreme rainfall under climate change: A case study in a tropical monsoon region with the climate projections from CMIP6 model. *J. Hydrol.* **610**, 128002. https://doi.org/10.1016/j.jhydrol.2022.128002 (2022).
77. Wang, C., Wang, S., Fu, B. & Zhang, L. Advances in hydrological modeling with the Budyko framework: A review. *Prog. Phys. Geogr.* **40**(3), 409–430. https://doi.org/10.1177/0309133315620997 (2016).
78. Cenobio-Cruz, O. et al. Improvement of low flows simulation in the SASER hydrological modeling chain. *J. Hydrol. X* **18**, 100147. https://doi.org/10.1016/j.hydroa.2022.100147 (2023).
79. Niu, S. et al. Water-use efficiency in response to climate change: from leaf to ecosystem in a temperate steppe. *Glob. Change Biol.* **17**(2), 1073–1082. https://doi.org/10.1111/j.1365-2486.2010.02280.x (2011).
80. Sun, Y., Liu, N., Shang, J. & Zhang, J. Sustainable utilization of water resources in China: A system dynamics model. *J. Clean. Prod.* **142**(613), 625. https://doi.org/10.1016/j.agwat.2023.108388 (2017).
81. Li, Z. et al. A comparative analysis of changes in temperature and precipitation extremes since 1960 between China and Greece. *Atmosphere* **13**(11), 1824. https://doi.org/10.3390/atmos13111824 (2022).
82. Volk, J. M. et al. Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. *Nat. Water* **2**(2), 193–205. https://doi.org/10.3390/w16030422 (2024).
83. Zhao, Y., Lu, Z. & Wei, Y. An assessment of global precipitation and evapotranspiration products for regional applications. *Remote Sens.* **11**(9), 1077. https://doi.org/10.3390/rs11091077 (2019).
84. Ippolito, M., Caro, D., Cannarozzo, D. M., Provenzano, G. & Ciraolo, G. Evaluation of daily crop reference evapotranspiration and sensitivity analysis of FAO Penman-Monteith equation using ERA5-Land reanalysis database in Sicily Italy. *Agric. Water Manage.* **295**, 108732. https://doi.org/10.1016/j.agwat.2024.108732 (2024).
85. Sabino, M. & de Souza, A. P. Global sensitivity of penman-monteith reference evapotranspiration to climatic variables in mato grosso Brazil. *Earth* **4**(3), 714–727 (2023).
86. Chanapathi, T., Thatikonda, S. & Raghavan, S. Analysis of rainfall extremes and water yield of Krishna River basin under future climate scenarios. *J. Hydrol. Reg. Stud.* **19**(287), 306. https://doi.org/10.1016/j.ejrh.2018.10.004 (2018).
87. Houet, T. et al. Combining narratives and modeling approaches to simulate fine-scale and long-term urban growth scenarios for climate adaptation. *Environ. Model. Softw.* **86**, 1–13. https://doi.org/10.1016/j.envsoft.2016.09.010 (2016).
88. Zhu, G. et al. The isotopes of precipitation have climate change signal in arid Central Asia. *Glob. & Planetary Change* **225**, 104103 (2023).
89. Tadese, M., Kumar, L., Koech, R. & Kogo, B. K. Mapping of land-use/land-cover changes and its dynamics in awash river basin using remote sensing and GIS. *Remote Sen. Appl.: Soc. & Environ.* **19**, 100352. https://doi.org/10.1016/j.rsase.2020.100352 (2020).
90. Yonaba, R. et al. A dynamic land use/land cover input helps in picturing the Sahelian paradox: Assessing variability and attribution of changes in surface runoff in a Sahelian watershed. *Sci. Total Environ.* **757**, 143792. https://doi.org/10.1016/j.scitotenv.2020.143792 (2021).
91. McCarty, D. A., Kim, H. W. & Lee, H. K. Evaluation of light gradient boosted machine learning technique in large scale land use and land cover classification. *Environments* **7**(10), 84. https://doi.org/10.3390/environments7100084 (2020).
92. Hussain, K. et al. Analysing LULC transformations using remote sensing data: insights from a multilayer perceptron neural network approach. *Ann. GIS* https://doi.org/10.1080/19475683.2024.2343399 (2024).
93. Roy, S. K. et al. Dynamic assessment and prediction of land use alterations influence on ecosystem service value: A pathway to environmental sustainability. *Environ. & Sustain. Indic.* **21**, 100319. https://doi.org/10.1016/j.indic.2023.100319 (2024).
94. Islami, F. A., Tarigan, S. D., Wahjunie, E. D. & Dasanto, B. D. Accuracy assessment of land use change analysis using google earth in sadar watershed mojokerto regency. *In: IOP Conf. Ser.: Earth Environ.* **950**(1), 012091. https://doi.org/10.1088/1755-1315/950/1/012091 (2022).
95. Jensen, O. C. & Azofeifa, G. A. S. Satellite-derived ecosystems classification: Image segmentation by ecological region for improved classification accuracy, a boreal case study. *Int. J. Remote Sens.* **27**(2), 233–251. https://doi.org/10.1080/01431160500181812 (2006).
96. Belay, T., Melese, T. & Senamaw, A. Impacts of land use and land cover change on ecosystem service values in the Afro-alpine area of guna mountain Northwest Ethiopia. *Heliyon* **8**(12), e12246. https://doi.org/10.1016/j.heliyon.2022.e12246 (2022).
97. Gessesse, B., Bewket, W., Bräuning, A., 2019. Land use and land cover transformation and its implication on land degradation: The case of Modjo watershed, Ethiopia. In: Extreme Hydrology and Climate Variability. Elsevier 503–518. https://doi.org/10.1016/B978-0-12-815998-9.00038-5
98. Barman, S., Singh, W. R., Kalita, B. & Tyagi, J. A combined impact assessment of climate and land use/land cover change in an Eastern Himalayan watershed in northeast India. *Environ. Monit. Assess.* **196**(3), 294 (2024).
99. Lambin, E. F., Geist, H. J. & Lepers, E. Dynamics of land-use and land-cover change in tropical regions. *Annu. Rev. Environ. Resour.* **28**(1), 205–241. https://doi.org/10.1146/annurev.energy.28.050302.105459 (2003).
100. Ding, Q., Chen, Y., Bu, L. & Ye, Y. Multi-scenario analysis of habitat quality in the Yellow River delta by coupling FLUS with InVEST model Int. *J. Environ. Res. Public Health* **18**(5), 2389. https://doi.org/10.3390/ijerph18052389 (2021).
101. Li, X. et al. Untangling the effects of climate change and land use/cover change on spatiotemporal variation of evapotranspiration over China. *J. Hydrol.* **612**, 128189. https://doi.org/10.1016/j.jhydrol.2022.128189 (2022).
102. Almeida, B. & Cabral, P. Water yield modeling, sensitivity analysis, and validation: A study for Portugal. *ISPRS Int. J. Geo-Inf.* **10**(8), 494. https://doi.org/10.3390/ijgi10080494 (2021).
103. Chicco, D., Warrens, M. J. & Jurman, G. The coefficient of determination R-squared is more informative than SMAPE MAE, MAPE, MSE, and RMSE in regression analysis evaluation. *PeerJ Comput Sci.* **7**, e623. https://doi.org/10.7717/peerj-cs.623 (2021).
104. Kouadri, S., Elbeltagi, A., Islam, A. R. M. T. & Kateb, S. Performance of machine learning methods in predicting water quality index based on irregular data set: Application on Illizi region (Algerian southeast). *Appl. Water Sci.* **11**(12), 190. https://doi.org/10.1007/s13201-021-01528-9 (2021).
105. Hussain, D. & Khan, A. A. Machine learning techniques for monthly river flow forecasting of Hunza River Pakistan. *Earth Sci. Inform.* **13**(3), 939–949. https://doi.org/10.1007/s12145-020-00450-z (2020).

106. Asadollah, S. B. H. S., Sharafati, A., Motta, D. & Yaseen, Z. M. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *J. Environ. Chem. Eng.* **9**(1), 104599. https://doi.org/10.1016/j.jece.2020.104599 (2021).
107. Dasari, H. P. et al. Long-term changes in the Arabian Peninsula rainfall and their relationship with the ENSO signals in the tropical Indo-Pacific. *Clim. Dyn.* https://doi.org/10.1007/s00382-021-06062-7 (2021).
108. Kogo, B. K., Kumar, L. & Koech, R. Climate change and variability in Kenya: A review of impacts on agriculture and food security. *Environ., Dev. & Sustain.* **23**(1), 23–43. https://doi.org/10.1007/s10668-020-00589-1 (2021).
109. Mortoja, M. G. & Yigitcanlar, T. Local drivers of anthropogenic climate change: Quantifying the impact through a remote sensing approach in Brisbane. *Remote Sens.* **12**(14), 2270. https://doi.org/10.3390/rs12142270 (2020).
110. Tilahun, Z. A., Bizuneh, Y. K. & Mekonnen, A. G. The combined effects of LULC changes and climate change on hydrological processes of gilgel gibe catchment, southwest Ethiopia. *Cogent Soc. Sci.* **11**(1), 2473644 (2025).
111. Tesfa, T. K., Leung, L. R. & Ghan, S. J. Exploring topography-based methods for downscaling subgrid precipitation for use in earth system models. *J. Geophys. Res.: Atmos.* **125**(5), e2019JD031456. https://doi.org/10.1029/2019JD031456 (2020).
112. Daba, B. I., Demissie, T. A. & Tufa, F. G. Sediment yield modeling in awash melkasa dam watershed, upper awash River basin Ethiopia. *Acta Geophys.* **71**(5), 2287–2306. https://doi.org/10.1007/s11600-022-00972-8 (2023).
113. Nyingi, R. W., Mwangi, J. K., Karimi, P. & Kiptala, J. K. Reliability of stream flow in inter-basin water transfer under different climatic conditions using remote sensing in the Upper Tana basin. *Phys. & Chem. Earth, Parts A/B/C* **134**, 103527. https://doi.org/10.1016/j.pce.2023.103527 (2024).
114. Berihun, M. L. et al. Exploring land use/land cover changes, drivers and their implications in contrasting agroecological environments of Ethiopia. *Land Use Policy* **87**, 104052. https://doi.org/10.1016/j.landusepol.2019.104052 (2019).
115. Sukanya, S., & Joseph, S., 2023. Climate change impacts on water resources: An overview. In: Visualization Techniques for Climate Change with Machine Learning and Artificial Intelligence, https://doi.org/10.1016/B978-0-323-99714-0.00008-X
116. Mera, G. A. Drought and its impacts in Ethiopia weather clim. *Extremes* **22**(24), 35. https://doi.org/10.1016/j.wace.2018.10.002 (2018).
117. United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA), 2017. Ethiopia: Humanitarian Needs Overview. Retrieved from https://reliefweb.int/report/ethiopia/ethiopia-humanitarian-needs-overview
118. Tilahun, Z. A., Bizuneh, Y. K. & Mekonnen, A. G. The impacts of climate change on hydrological processes of gilgel gibe catchment, southwest Ethiopia. *PLoS ONE* **18**(6), e0287314. https://doi.org/10.1371/journal.pone.0287314 (2023).
119. Arega, K. A. et al. Analysis of spatio-temporal variability of groundwater storage in Ethiopia using Gravity Recovery and Climate Experiment (GRACE) data. *Environ. Earth Sci.* **83**(7), 206. https://doi.org/10.1007/s12665-024-11508-2 (2024).
120. Tesfaye, B., Lengoiboni, M., Zevenbergen, J. & Simane, B. Mapping land use land cover changes and their determinants in the context of a massive free labour mobilization campaign: Evidence from South Wollo Ethiopia. *Remote Sens.* **13**(24), 5078. https://doi.org/10.3390/rs13245078 (2021).
121. Gebrehiwot, K., Desalegn, T., Woldu, Z., Demissew, S. & Teferi, E. Soil organic carbon stock in abune yosef afroalpine and sub-afroalpine vegetation, northern Ethiopia. *Ecol. Process.* **7**(1), 1–9. https://doi.org/10.1186/s13717-018-0117-9 (2018).
122. Dessu, T., Korecha, D., Hunde, D. & Worku, A. Long-term land use land cover change in urban centers of southwest Ethiopia from a climate change perspective. *Front. Clim.* **2**, 577169. https://doi.org/10.3389/fclim.2020.577169 (2020).
123. Teshome, D. S., Tolessa, T., Gemeda, D. O., Taddese, H. & You, S. Impact of land use/land cover (LU/LC) changes on ecosystem service values in muger sub-basin, upper blue nile basin. *Ethiopia. Environ. Chall.* **17**, 101041. https://doi.org/10.1016/j.envc.2024.101041 (2024).
124. Rahman, M. M. & Szabó, G. Impact of land use and land cover changes on urban ecosystem service value in dhaka Bangladesh. *Land* **10**(8), 793. https://doi.org/10.3390/land10080793 (2021).
125. Tesfaye, S. et al. Observed and model simulated twenty-first-century hydro-climatic change of Northern Ethiopia. *J. Hydrol. Reg. Stud.* **22**, 100595. https://doi.org/10.1016/j.ejrh.2019.100595 (2019).
126. Gashaw, T., Tulu, T., Argaw, M. & Worqlul, A. W. Modeling the impacts of land use–land cover changes on soil erosion and sediment yield in the andassa watershed, upper blue nile basin. *Ethiopia. Environ. Earth Sci.* **78**, 1–22. https://doi.org/10.1007/s12665-019-8726-x (2019).
127. Alemayehu, S. et al. Evaluating land suitability and potential climate change impacts on alfalfa (Medicago sativa) production in Ethiopia. *Atmosphere* **11**(10), 1124. https://doi.org/10.3390/atmos11101124 (2020).
128. Abebe, G., Getachew, D. & Ewunetu, A. Analyzing land use/land cover changes and its dynamics using remote sensing and GIS in gubalafito district Northeastern Ethiopia. *SN Appl. Sci.* **4**(1), 30. https://doi.org/10.1007/s42452-021-04915-8 (2022).
129. Berhanu, Y., Dalle, G., Sintayehu, D. W., Kelboro, G. & Nigussie, A. Land use/land cover dynamics driven changes in woody species diversity and ecosystem services value in tropical rainforest frontier: A 20-year history. *Heliyon* **9**(2), e13711. https://doi.org/10.1016/j.heliyon.2023.e13711 (2023).
130. Taye, M. T., Assefa, T. T., Ebrahim, G. Y., Lautze, J. & Seid, A. H. Integrated water storage assessment in the tana-beles sub-basin Ethiopia. *Sustain. Water Res. Manage.* **11**(2), 1–20 (2025).
131. Langat, P. K., Kumar, L. & Koech, R. Temporal variability and trends of rainfall and streamflow in tana river basin. *Kenya. Sustain.* **9**(11), 1963 (2017).
132. Xie, X. et al. Assessing the impacts of climate and land use change on water conservation in the three-river headstreams region of China Based on the integration of the InVEST model and machine learning. *Land* **13**(3), 352 (2024).
133. Mekuria, W., Hagos, F., Nigussie, L., Admasu, W., & Bitew, A. 2024. Impacts of landscape and household climate-smart water management practices in the Awash River Basin, Ethiopia.
134. Alehu, B. A., Desta, H. B. & Daba, B. I. Assessment of climate change impact on hydro-climatic variables and its trends over gidabo watershed. *Model. Earth Syst. & Environ.* **8**(3), 3769–3791 (2022).
135. Gao, X. J. et al. Future changes in thermal comfort conditions over China based on multi-RegCM4 simulations. *Atmos. & Ocean. Sci. Lett.* **11**(4), 291–299 (2018).
136. Toma, M. B., Ulsido, M. D., Bitew, A. M. & Meja, M. F. Downscaling, projection, and analysis of expected future climate change in a watershed of omo-gibe basin of Ethiopia. *Front. Water* **6**, 1444638 (2024).
137. Bojer, A. K., Woldetsadik, M. & Biru, B. H. Machine learning and CORDEX-Africa regional model for assessing the impact of climate change on the gilgel gibe watershed Ethiopia. *J. Environ. Manage.* **363**, 121394 (2024).

Acknowledgements

The authors sincerely thank the Earth Observatory Dataset (USGS), Addis Ababa University, Ethiopian Space Science and Geospatial Institute (ESSGI), Ministry of Water and Energy (MoWE), Mr. Olkeba Bashe, and the Ethiopian Meteorological Institute for their generous provision of critical datasets, which were vital to the success of this study.

Author contributions

Amanuel Kumsa Bojer: conceived the study, collected and processed data, developed remote sensing models, drafted the manuscript, conducted statistical analyses, and ensured data integrity. Dr. Muluneh Woldetsadik:

provided expert guidance and supervision throughout the research process. Fitsum Mesfin: made substantial contributions to machine learning model development and implementation. Professor Ayad M. Fadhil Al-Qurai-shi: supervision, quality edition, and language correction.The authors affirm that this manuscript is original and has not been published or submitted for consideration elsewhere. They secured all the required permissions for any copyrighted material included. Each author has contributed significantly to the research and has approved the final version of the manuscript.

Funding
This study was conducted independently without external financial support.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-03493-8.

Correspondence and requests for materials should be addressed to A.K.B. or M.W.A.