Surface water dynamics and driving factors in the Huangshui River Basin of the Xining-Haidong Corridor in China from 2000 to 2024 based on Google Earth Engine

CHENG Ruixuan¹, QIN Ke^{1*}, YIN Sijiang²

2026-07-16

ChinaRxiv

Abstract

Sustaining surface water in arid mountain-valley corridors has become increasingly difficult under a warming-wetting climate and urbanization. Focusing on the Huangshui River Basin in the Xining-Haidong Corridor of China, this study used Google Earth Engine (GEE) to process Landsat 5/7/8/9 surface reflectance imagery to reconstruct open-water dynamics within the riparian corridor. A three-expert ensemble voting framework integrated spectral water indices, Dynamic Surface Water Extent rules, and a Random Forest classifier, and the extraction results were validated using 1200 manually interpreted points. Trend analysis, Random Forest attribution, and GeoDetector were then applied to assess temporal changes and the combined effects of climate, topography, and human activity. The extraction achieved an overall accuracy of 96.17% and a Kappa coefficient of 0.923. The mapped open-water area increased from 74.33 km2 in 2000 to 121.67 km2 in 2024, corresponding to a net gain of 47.34 km2 (63.68%). The Mann-Kendall test indicated a significant upward trend (Z=6.66; P<0.001), with a Theil-Sen slope of 1.45 km2/a. Sequential Mann-Kendall analysis identified no robust year of abrupt change, although the increasing trend became significant after 2007. Attribution results showed that topography remained the dominant spatial regulator of water persistence, as low-elevation valley floors concentrate both runoff accumulation and human land use. Population density (PD) and nighttime light (NTL) signals were stronger in urbanized reaches, where high impervious-surface values and mapped water co-occurred around managed water environments, including regulated channels, impoundments, and reservoir storage. Overlay analysis of barriers and reservoirs further suggested that engineering regulation may account for part of the persistent water patches along the corridor. These findings reveal a coupled mechanism involving climate, topography, and human regulation in shaping surface water change in a water-limited plateau river corridor, and provide evidence for water-resource management and ecological restoration in the upper Yellow River Basin.

Full Text

J Arid Land (2026) 18(7): 1192–1212
doi: 10.1016/j.aridl.2026.07.002; CSTR: 32276.14.JAL.20250668

Surface water dynamics and driving factors in the Huangshui River Basin of the Xining-Haidong Corridor in China from 2000 to 2024 based on Google Earth Engine

CHENG Ruixuan1, QIN Ke1*, YIN Sijiang2

1 College of Horticulture, China Agricultural University, Beijing 100083, China;

2 Beijing Academy of Forestry and Landscape Architecture, Beijing 100102, China

Abstract: Sustaining surface water in arid mountain-valley corridors has become increasingly difficult under a warming-wetting climate and urbanization. Focusing on the Huangshui River Basin in the Xining-Haidong Corridor of China, this study used Google Earth Engine (GEE) to process Landsat 5/7/8/9 surface reflectance imagery to reconstruct open-water dynamics within the riparian corridor. A three-expert ensemble voting framework integrated spectral water indices, Dynamic Surface Water Extent rules, and a Random Forest classifier, and the extraction results were validated using 1200 manually interpreted points. Trend analysis, Random Forest attribution, and GeoDetector were then applied to assess temporal changes and the combined effects of climate, topography, and human activity. The extraction achieved an overall accuracy of 96.17% and a Kappa coefficient of 0.923. The mapped open-water area increased from 74.33 km^2 in 2000 to 121.67 km^2 in 2024, corresponding to a net gain of 47.34 km^2 (63.68%). The Mann-Kendall test indicated a significant upward trend ($Z = 6.66; P < 0.001$), with a Theil-Sen slope of 1.45 km^2/a. Sequential Mann-Kendall analysis identified no robust year of abrupt change, although the increasing trend became significant after 2007. Attribution results showed that topography remained the dominant spatial regulator of water persistence, as low-elevation valley floors concentrate both runoff accumulation and human land use. Population density (PD) and nighttime light (NTL) signals were stronger in urbanized reaches, where high impervious-surface values and mapped water co-occurred around managed water environments, including regulated channels, impoundments, and reservoir storage. Overlay analysis of barriers and reservoirs further suggested that engineering regulation may account for part of the persistent water patches along the corridor. These findings reveal a coupled mechanism involving climate, topography, and human regulation in shaping surface water change in a water-limited plateau river corridor, and provide evidence for water-resource management and ecological restoration in the upper Yellow River Basin.

Keywords: Landsat; Google Earth Engine (GEE); surface water dynamics; riparian corridor; river regulation; GeoDetector

Citation: CHENG Ruixuan, QIN Ke, YIN Sijiang. 2026. Surface water dynamics and driving factors in the Huangshui River Basin of the Xining-Haidong Corridor in China from 2000 to 2024 based on Google Earth Engine. Journal of Arid Land, 18(7): 1192–1212. https://doi.org/10.1016/j.aridl.2026.07.002; https://cstr.cn/32276.14.JAL.20250668

---

*Corresponding author: QIN Ke (E-mail: qinke@cau.edu.cn)
Received 2025-12-31; revised 2026-05-13; accepted 2026-07-01

http://jal.xjegi.com; https://www.keaipublishing.com/en/journals/journal-of-arid-land/

This version posted 2026-07-16.

1 Introduction

Surface water is one of the most sensitive components of arid and semi-arid environments because it responds rapidly to climate variability, land use change, and water regulation (Evans and Geerken, 2004; Wessels et al., 2007; Veldkamp et al., 2017). It provides ecological refugia, supports agricultural production, and contributes to urban water security, particularly in regions where runoff is spatially concentrated and human water demand is high (Viviroli et al., 2007, 2020). Global Landsat-based assessments provide the broader remote-sensing context by documenting changes in surface water extent, occurrence, and seasonality, while full time-series mapping has improved the detection of long-term open-water dynamics (Pekel et al., 2016; Pickens et al., 2020). In mountain and plateau regions, warming, changes in precipitation, snow and glacier dynamics, and human water use jointly alter runoff seasonality and the distribution of open-water (Konapala et al., 2020; Ran et al., 2023; Sui et al., 2023; Xiao et al., 2024). On the Qinghai-Xizang Plateau, China recent studies have generally reported increasing surface water in many interior basins, but they have also highlighted pronounced regional heterogeneity rather than a uniform response across the plateau (Immerzeel et al., 2020; Ran et al., 2023; Sui et al., 2023; Xiao et al., 2024).

The Huangshui River Basin in the Xining-Haidong Corridor, China differs fundamentally from the lake-dominated interior basins of the plateau. From a physical perspective, it is an arid mountain-valley system in which runoff is generated in the surrounding highlands, whereas water storage and use are concentrated on narrow valley floors. From a socio-economic perspective, it is a densely populated and intensively developed corridor in eastern Qinghai Province, China where cities, transportation infrastructure, and cultivated land are compressed into limited habitable space (Li and Xin, 2024; Liu et al., 2024; Xu et al., 2024). Previous basin- and valley-scale studies have shown that human activities in the Huangshui system affect runoff and ecological flow and are closely associated with land development, vegetation change, and cropland-related ecological pressure (Wei et al., 2021; Fan et al., 2024; Li and Xin, 2024; Liu et al., 2024; Xu et al., 2024). This concentration of water storage and use, settlement, cropland, infrastructure, and managed channels on the valley floor indicates that surface water in the Huangshui River Basin is shaped by the climatic background, intensive human occupation, and water regulation. The basin therefore provides an appropriate setting for examining how scarce water resources are reorganized within a heavily utilized plateau river corridor.

However, this type of corridor system remains less well understood than the broader pattern observed across the plateau. Remote-sensing studies of surface water change on the Qinghai-Xizang Plateau have mainly emphasized broad climatic controls and lake dynamics, leaving topographically confined river-corridor basins with intensive human disturbance less resolved (Pekel et al., 2016; Pickens et al., 2020; Ran et al., 2023; Sui et al., 2023). In arid inland basins and urban rivers, urbanization, channel engineering, and regulated water storage can redirect runoff, alter water residence time, increase evaporative loss, and modify groundwater-river interactions, particularly in developed valley reaches (Heritage and Entwistle, 2020; Mao et al., 2021; Li et al., 2023; Oswald et al., 2023; Banner et al., 2024). The Huangshui River Basin and the Hehuang Valley provide a representative setting for these coupled processes, and recent studies have linked human activities to runoff change, ecological flow alteration, ecohydrological regulation, land development, vegetation change, and farmland-related ecological pressure in this region (Wei et al., 2021; Fan et al., 2024; Li and Xin, 2024; Liu et al., 2024; Xu et al., 2024). At the same time, 30 m Landsat observations mainly resolve the main stem, wider tributaries, reservoirs, and other medium-to-large open-water bodies; narrow channels, irrigation ditches, and small wetlands are more likely to be omitted or mixed with adjacent land-cover types (Dong et al., 2022; Vanderhoof et al., 2023). This scale constraint defines the scope of the present analysis, which focuses on mapped open-water dynamics and their climatic, topographic, and human-associated controls within the detectable river-corridor water system.

Against this background, this study used Google Earth Engine (GEE) and long-term optical remote-sensing products from 2000 to 2024 to reconstruct surface water dynamics in the Huangshui River Basin and its riparian corridor. The water-mapping framework integrated Landsat-based

spectral water indices, Dynamic Surface Water Extent rules, and machine-learning methods to reduce confusion caused by mountain shadows, built-up land, and mixed wetland pixels (Breiman, 2001; Xu, 2006; Feyisa et al., 2014; Gorelick et al., 2017; Jones, 2019). Trend detection, Random Forest attribution, and GeoDetector were then applied to examine temporal changes, the contributions of climatic and anthropogenic factors, and spatially stratified heterogeneity (Mann, 1945; Sen, 1968; Breiman, 2001; Wang et al., 2010, 2016). Specifically, this study addressed three questions: (1) how mapped surface water has changed over the past 25 a in this arid and highly developed corridor? (2) how climatic conditions, topography, and human activities have jointly shaped its spatial pattern? And (3) how river regulation, independent temporal-consistency checking of the long-term series, and the detection limit of 30 m imagery constrain the interpretation of mapped water change. By focusing on a water-limited mountain corridor within the plateau, this study provides a region-specific understanding of surface water dynamics and a practical basis for water management in the Xining-Haidong Corridor.

2 Materials and methods

2.1 Study area

The study area is the 10 km riparian zone within the Huangshui River Basin (35°38$ ‘43’ ‘–37°22’ 44’ ‘N, 100°58’ 14’ ‘–103°04’ 22’ ’ $E) in the Xining-Haidong Corridor of eastern Qinghai Province, China. The Huangshui River is an important tributary of the upper Yellow River and the principal river corridor of the Hehuang Valley. The basin lies between the Qinghai-Xizang Plateau and the Loess Plateau and shows strong topographic relief, vertical environmental gradients, and a mountain-runoff-generation and valley-concentration pattern.

The region has a plateau continental semi-arid climate, with warm-season precipitation and high evapotranspiration demand. Unlike the sparsely populated lake regions in the plateau interior, the Xining-Haidong Corridor is a densely used valley where settlement, agriculture, transport infrastructure, and industry are concentrated in limited lowland space (Li and Xin, 2024). Surface water is therefore sensitive to both hydrological fluctuation and artificial regulation.

To reflect this coupled setting, the analysis used a riparian ecological corridor rather than only the administrative boundary (Fig. 1). The corridor retained the Xining-Haidong management context while focusing on the main valley zone where human-water interactions are most intensive. This design supports analysis of climatic influences, urban expansion, agricultural use, and river regulation within the most heavily used part of the basin.

2.2 Data sources and preprocessing

A 2000–2024 geospatial dataset was built from surface water, climate, topography, vegetation, land-cover, population, nighttime light (NTL), and engineering-context data (Table 1). Most spatial processing was conducted on GEE (https://earthengine.google.com), while statistical analysis and figure preparation were completed in Python v.2.7. All layers were harmonized to the World Geodetic System 1984 (WGS-84) for extraction and then summarized to a common grid for modelling (Gorelick et al., 2017).

2.2.1 Surface water extraction Surface water was mapped from Landsat 5/7/8/9 Collection 2 Level-2 Surface Reflectance products for 2000–2024. Images were restricted to 1 May–30 September to reduce frozen-season and snow effects. The quality assessment pixel (QA_PIXEL) band was used to remove high-confidence cloud, cloud shadow, and snow or ice pixels. Reflectance values were converted with the product scale and offset factors, and Thematic Mapper/Enhanced Thematic Mapper Plus (TM/ETM+) and Operational Land Imager/Operational Land Imager-2 (OLI/OLI-2) bands were mapped to a common spectral set before annual warm-season compositing.

The final binary water mask was generated with a three-expert ensemble. The first expert used the modified normalized difference water index (MNDWI), enhanced-modified normalized difference water index (E-MNDWI), and automated water extraction index for no-shadow

conditions ($AWEI_{nsh}$) to enhance open-water under complex terrain and urban backgrounds (Xu, 2006; Feyisa et al., 2014). The second expert followed a Dynamic Surface Water Extent (DSWE)-like rule set using MNDWI, normalized difference vegetation index (NDVI), near-infrared (NIR), and shortwave-infrared (SWIR) tests to identify open-water and mixed wetland pixels (Jones, 2019). The third expert was a supervised Random Forest classifier trained from visually interpreted high-resolution samples using spectral bands, water indices, elevation, and slope as predictors (Breiman, 2001). Pixels receiving at least two water votes were classified as water.

Visible labels in Fig. 1: Xining City; Haidong City; Huangshui River; N; Legend; River; Administrative boundary; Study area (10 km from the river); Altitude (m); High: 4854; Low: 1659; 0; 40 km.

Fig. 1 Elevation and river networks of the study area. The map shows the 10 km riparian study area along the major river network across Xining City and Haidong City.

The index equations were defined as follows:

$$\mathrm{MNDWI} = \frac{\rho_{\mathrm{Green}} - \rho_{\mathrm{SWIR1}}}{\rho_{\mathrm{Green}} + \rho_{\mathrm{SWIR1}}}, \tag{1}$$

$$\text{E-MNDWI} = \frac{\rho_{\mathrm{Green}} - \rho_{\mathrm{SWIR2}}}{\rho_{\mathrm{Green}} + \rho_{\mathrm{SWIR2}}}, \tag{2}$$

$$\mathrm{AWEI}_{nsh} = 4(\rho_{\mathrm{Green}} - \rho_{\mathrm{SWIR1}}) - (0.25\rho_{\mathrm{NIR}} + 2.75\rho_{\mathrm{SWIR2}}), \tag{3}$$

where ρ_{Green}, ρ_{SWIR1}, ρ_{SWIR2}, and ρ_{NIR} denote Landsat surface reflectance in the green, shortwave-infrared 1, shortwave-infrared 2, and near-infrared bands, respectively; MNDWI, E-MNDWI, and $AWEI_{nsh}$ are dimensionless indices.

Post-processing removed likely pseudo-water pixels using a slope threshold (>15.00°), a dimensionless blue-band surface-reflectance threshold (<0.25), and a minimum patch size of 10 pixels. Annual water areas and representative-year maps were generated only from these Landsat ensemble masks. The Joint Research Centre (JRC) Global Surface Water Yearly History product was used only as an independent temporal-consistency check for 2000–2021 and did not enter the area calculation or hydrological-barrier mapping and overlay (Pekel et al., 2016).

2.2.2 Driving factors and gridded panel

Complete annual records of dams, sluices, reservoirs, and operating rules were unavailable for 2000–2024. River-engineering variables were therefore not used as separate annual predictors, and dam, reservoir, and river-barrier records were retained only as spatial context for regulated reaches.

Table 1 Main datasets, variables, and analytical roles in the analysis

Main variable	Resolution	Period	Dataset/GEE collection ID
Annual SWA (km^2)	30 m	2000–2024	Landsat Collection 2 Level-2 SR:LANDSAT/LT05/C02/T1_L2;LANDSAT
JRC annual water class	30 m	2000–2021	JRC Global Surface Water Yearly History:JRC/GSW1_4/YearlyHistory.
Grid-cell SWF	1 km grid	2000–2024	Calculated from the Landsat ensemble mask.
Annual total precipitation (mm)	0.05°	2000–2024	CHIRPS Daily: UCSB-CHG/CHIRPS/DAILY.
Annual mean 2-m air temperature (°C)	0.10°	2000–2024	ERA5-Land Monthly Aggregated:ECMWF/ERA5_LAND/MONTHLY_A
Annual mean RH (%)	0.10°	2000–2024	ERA5-Land Monthly Aggregated:ECMWF/ERA5_LAND/MONTHLY_A
Daytime LST (°C)	1 km	2000–2024	MODIS/Terra Land Surface Temperature Daily:MODIS/061/MOD11A1.
Peak growing-season (May–September) NDVI	250 m	2000–2024	MODIS/Terra Vegetation Indices 16-Day:MODIS/061/MOD13Q1.
Elevation (m)	30 m		NASADEM: NASA/NASADEM_HGT/001.

Main variable	Resolution	Period	Dataset/GEE collection ID
Slope (°)	30 m		Calculated from NASADEM.
ISF	500 m/10 m source data; aggregated to 1 km	2000–2024	MODIS MCD12Q1 before 2016:MODIS/006/MCD12Q1; Dynamic World from 2016: GOOGLE/DYNAMICWORLD/V1.
CF	500 m/10 m source data; aggregated to 1 km	2000–2024	MODIS MCD12Q1 before 2016:MODIS/006/MCD12Q1; Dynamic World from 2016: GOOGLE/DYNAMICWORLD/V1.
PD (persons/km^2)	About 100 m; aggregated to 1 km	2000–2024; 2020 layer carried forward for 2021–2024	WorldPop: World-Pop/GP/100m/pop.
NTL	DMSP-OLS/VIIRS source data; aggregated to 1 km	2000–2024; 2022 layer carried forward for 2023–2024	NOAA/DMSP-OLS/NIGHTTIME_LIGHTS;NOAA/VIIRS/ and VCMSLCFG.
HAI	1 km grid	2000–2024	Calculated in this study; arithmetic mean of ISF and min-max-normalized PD.
Dam/barrier location	Vector	Data-available period	Global Dam Watch barriers: projects/sat-io/opendatasets/GDW/GDW_BARRIERS_V
Reservoir extent	Vector	Data-available period	Global Dam Watch reservoirs: projects/sat-io/opendatasets/GDW/GDW_RESERVOIRS

Main variable	Resolution	Period	Dataset/GEE collection ID
Major river network	Vector		HydroSHEDS free-flowing rivers:WWF/HydroSHEDS/v1/FreeFlowingRi

Note: GEE, Google Earth Engine; JRC, Joint Research Centre; CHIRPS, Climate Hazards Group InfraRed Precipitation with Station data; ERA5-Land, fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis for land; MODIS, Moderate Resolution Imaging Spectroradiometer; NASADEM, National Aeronautics and Space Administration (NASA)-digital elevation model; NOAA, National Oceanic and Atmospheric Administration; DMSP-OLS, Defense Meteorological Satellite Program-Operational Linescan System; VIIRS, Visible Infrared Imaging Radiometer Suite; HydroSHEDS, Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales; GDW, Global Dam Watch. SWA, surface water area; SWF, surface water fraction; RH, relative humidity; LST, land surface temperature; NDVI, normalized difference vegetation index; ISF, impervious-surface fraction; CF, cropland fraction; PD, population density; NTL, nighttime light; HAI, composite human-activity index. HAI is a study-specific proxy rather than a standardized product variable.

A 1 km$\times$1 km grid in Universal Transverse Mercator (UTM) Zone 47N (European Petroleum Survey Group (EPSG) code: 32647) was used as the modelling unit. For 30 m classification layers, grid-cell fractions were calculated from pixel areas; continuous variables were summarized as grid means after harmonization. Invalid grids and outliers were removed, producing a 25-a grid-year panel with 31,164 records.

2.3 Methodology

The overall workflow is summarized in Figure 2. The first stage constructs the 10 km riparian ecological corridor and harmonizes Landsat, terrain, climate, vegetation, land-cover, population, and NTL datasets in GEE. The second stage applies the Landsat ensemble water mapping scheme, aggregates water and driver variables to 1 km grids, and then combines accuracy assessment, trend diagnostics, principal component analysis (PCA) diagnostics, Random Forest residual diagnostics, and GeoDetector analysis to evaluate map reliability, temporal change, spatial heterogeneity, and interacting drivers.

Phase I Data preparation and surface water mapping

- **Landsat imagery**
  - Collection 2 Level-2 surface reflectance
  - Landsat 5/7/8/9, 2000–2024
  - May–September scenes

- **Natural-environment data**
  - CHIRPS precipitation
  - ERA5-Land air temperature, RH,
  - MODIS daytime LST, NDVI
  - NASADEM elevation, and slope
- **Human-activity data**
  - MODIS/Dynamic World land-cover
  - WorldPop PD
  - DMSP-OLS/NPP-VIIRS lights
  - Global Dam Watch barriers
- **Ensemble water mapping**
  - QA cloud/snow masking
  - Reflectance, band harmonization
  - Annual warm-season composites
  - MNDWI, E-MNDWI, AWEI_{nsh}
  - DSWE-like rules+Random Forest
  - Voting+terrain/patch filters
- **Natural-predictor preparation**
  - Annual totals, means and peaks
  - Reprojection and unit harmonizing,
  - Grid-cell summarization
- **Human-indicator construction**
  - ISF and CF from land-cover
  - PD and NTL harmonization
  - HAI=mean(ISF/normalized PD)
  - Engineering layers as context
- **10 km riparian study zone and 1 km$×$1 km analysis grid**
  - Annual SWA, grid-cell SWF and candidate predictors; 25-a grid-year panel (18,170 records)

Phase II Statistical analysis and attribution

- **Validation and temporal analysis**
  - 1200 independent reference points
  - JRC consistency check (2000–2021)
  - Theil-Sen slope, Mann-Kendall test
  - Sequential UF/UB statistics
- **Multivariate and Random Forest diagnostics**
  - Z-score, CV,
  - PCA
  - Climate-baseline, residual SWF
  - Random Forest importance
- **Spatial heterogeneity and context**
  - GeoDetector factor detector
  - GeoDetector interaction detector
  - Dam/reservoir overlay
  - Scale limits and uncertainty

- **Integrated interpretation**
  - Long-term mapped water dynamics and associated controls; factor interactions; and regulation context and uncertainty

Fig. 2 Analytical framework of this study. CHIRPS, Climate Hazards Group InfraRed Precipitation with Station data; ERA5-Land, fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis for land; RH, relative humidity; MODIS, Moderate Resolution Imaging Spectroradiometer; LST, land surface temperature; NDVI, normalized difference vegetation index; NASADEM, National Aeronautics and Space Administration (NASA)-digital elevation model; DMSP-OLS, Defense Meteorological Satellite Program-Operational Linescan System; NPP-VIIRS, National Polar-orbiting Partnership-Visible Infrared Imaging Radiometer Suite; QA, quality assessment; MNDWI, modified normalized difference water index; E-MNDWI, enhanced-modified normalized difference water index; AWEI$_{nsh}$, automated water extraction index for no-shadow conditions; DSWE, Dynamic Surface Water Extent; ISF, impervious-surface fraction; CF, cropland fraction; PD, population density; NTL, nighttime light; HAI, composite human-activity index; JRC, Joint Research Centre; UF and UB, forward and backward sequential Mann-Kendall statistics, respectively; SWA, surface water area; SWF, surface water fraction; CV, coefficient of variation; PCA, principal component analysis.

2.3.1 Long-term trend detection

Monotonic trends were evaluated with the Theil-Sen slope estimator and the Mann-Kendall test, which are suitable for non-normal environmental time series (Mann, 1945; Sen, 1968). The Theil-Sen slope was calculated as:

$$\beta = \text{Median}\left(\frac{x_j - x_i}{j - i}\right) \quad (i < j), \tag{4}$$

where x_i and x_j are observations in years i and j, respectively; and β is the Theil-Sen slope in the unit of the analysed variable per year. A positive β indicates an increasing trend. Sequential Mann-Kendall statistics were also used to identify possible mutation years.

2.3.2 Random Forest attribution

To partition the climate-associated component of mapped water-area variation from residual non-climatic variation, we used a RESTREND-inspired framework in which Random Forest replaced the original linear regression step (Evans and Geerken, 2004; Wessels et al., 2007; Tyralis et al., 2019). Climate variables, including annual total precipitation, annual mean 2 m air temperature, daytime land surface temperature (LST), and annual mean relative humidity (RH), were used to predict the observed surface water fraction (SWF$_{obs}$). The residual surface water fraction (SWF$_{res}$) was calculated as:

$$\mathrm{SWF}_{res} = \mathrm{SWF}_{obs} - \mathrm{SWF}_{clim}, \quad (5)$$

where SWF_{obs}, SWF_{clim}, and SWF_{res} are the observed, climate-predicted, and residual surface water fractions, respectively. All three terms are dimensionless. Trends in SWF_{res} describe residual deviations from the climate-predicted baseline. These residual deviations were evaluated together with land use, population, NTL, and regulation-context information, with engineering regulation retained as spatial context because annual operation records were unavailable.

2.3.3 GeoDetector analysis

Spatial heterogeneity and factor interactions were analysed with GeoDetector (Wang et al., 2010, 2016). The Factor Detector quantified the explanatory power of each driver using the q-statistic:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2}, \quad (6)$$

where q is the dimensionless explanatory power; h is a stratum; L is the number of strata; N_h and N are the numbers of units in stratum h and in the whole region, respectively; and σ^2 is the corresponding variances. A larger q indicates stronger spatial explanatory power. The Interaction Detector compared $q(X1 \cap X2)$ with $q(X1)$ and $q(X2)$ to determine whether two factors produced bi-factor or nonlinear enhancement.

2.3.4 Accuracy assessment

The accuracy of the surface water extraction was assessed with an independent point-based validation. High-resolution historical imagery from Google Earth Pro was used as reference data. A total of 1200 validation points were manually interpreted for three validation years (2013, 2018, and 2024), including 600 water points and 600 non-water points. These years were selected because high-resolution reference imagery was sufficiently available for independent visual interpretation. This validation was independent of the JRC temporal-consistency check. The Kappa coefficient was used to evaluate agreement beyond chance and was calculated as:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \quad (7)$$

where κ is the dimensionless Kappa coefficient; p_o is the observed agreement; and p_e is the expected agreement by chance.

2.3.5 PCA

PCA was used as a diagnostic analysis to summarize covariation among the candidate driving variables and to identify dominant multivariate gradients. Before PCA, all candidate driving

variables were standardized to Z-scores so that variables with different units could be compared on the same scale. PCA was performed on the standardized driver matrix, and the first two principal components were retained for visual interpretation in a biplot. The PCA biplot was interpreted as a loading-space summary of covariation among variables rather than as independent causal evidence; therefore, composite human-activity index (HAI) was treated as a derived covariate with expected covariance with impervious-surface fraction (ISF) and population density (PD).

3 Results

3.1 Accuracy assessment of surface water extraction

As shown in the confusion matrix (Table 2), the extraction achieved an overall accuracy of 96.17% and a Kappa coefficient of 0.923. For the water class, user' s accuracy and producer' s accuracy were 97.42% and 94.83%, respectively. These results indicate that the combination of physical constraints and ensemble classification effectively reduced misclassification caused by mountain shadows and built-up surfaces, and yielded a reliable 25-a surface water dataset.

Table 2 Confusion matrix for surface water extraction accuracy assessment

Classification	Reference: water point	Reference: non-water point	Row total	User' s accuracy
Classified: water point	569	15	584	97.42%
Classified: non-water point	31	585	616	95.00%
Column total	600	600	1200	-
Producer' s accuracy	94.83%	97.50%	-	Overall accuracy96.17%Kappa coefficient:0.923

Note: Reference water/non-water denotes the manually interpreted class from high resolution Google Earth Pro imagery; and classified water/non-water denotes the class produced by the Landsat ensemble extraction. "-" means no value.

3.2 Temporal evolution characteristics of surface water and its driving factors

Figure 3 provides the spatial context for the two representative years before the full annual time series is examined. Mapped open-water is concentrated along the Huangshui River main stem and wider tributary valleys, while candidate hydrological barriers are clustered along the same regulated reaches. The comparison between 2000 and 2024 is used to illustrate the spatial distribution of mapped water and engineering context; quantitative interannual changes are evaluated using the annual statistics in Figure 4.

Figure 4 summarizes the interannual changes in corridor-scale surface water area (SWA) and its potential driving factors during 2000–2024. Based on the output of the three-expert ensemble voting framework, the mapped SWA increased from 74.33 km^2 in 2000 to 121.67 km^2 in 2024, representing a net gain of 47.34 km^2 (63.68%) (Fig. 3). The Mann-Kendall test and Sen' s slope analysis confirmed a significant upward trend ($Z = 6.66,\ P < 0.001$; Sen' s slope=1.45 km^2/a). Sequential Mann-Kendall analysis identified no robust abrupt change year within the 95% confidence interval, although the forward sequential Mann-Kendall statistic exceeded the significance threshold after 2007, indicating a sustained intensification of the increasing trend rather than a single discrete shift. The level of mapped open-water after 2021 was notably higher than that during 2000–2021. Because the reported time series was derived from the ensemble output based on Landsat imagery, this increase does not reflect a boundary between different source products. One plausible explanation is the increase in valid observation density after Landsat 9

(a) 2000 (b) 2024

SWA=74.33 km^2

Legend
Mapped surface water
Barrier inside corridor
Connected outer barrier

SWA=121.67 km^2

N

0 40 km

Fig. 3 Spatial distribution of mapped surface water and candidate hydrological barriers in the two representative years 2000 (a) and 2024 (b)

became operational in 2022, together with relatively wet warm seasons in several recent years. To assess the potential influence of any residual classifier sensitivity associated with this change in observation density, we cross-validated the ensemble series against the JRC Global Surface Water Yearly History product over the overlapping period from 2000 to 2021. The JRC product also showed a significant increase, from 74.33 to 106.50 km^2 ($Z = 6.09$, $P < 0.001$; Sen's slope = 1.18 km^2/a), supporting the long-term increasing tendency while indicating that the absolute magnitude of mapped surface water in the most recent years is affected by increased observation density. Elevation and slope are static topographic variables and therefore were not plotted as annual trend panels in Figures 4 and 5; they were retained for classification support, spatial heterogeneity assessment, PCA, Random Forest modelling, and GeoDetector analysis. For the dynamic potential drivers, precipitation and air temperature showed increasing trends. PD and NTL also showed higher proxy-layer values in the later period, whereas cropland fraction (CF), ISF, HAI, and LST declined. NDVI showed no clear linear trend, and RH fluctuated with a weak decreasing tendency.

Figure 5 shows the standardized spatiotemporal fingerprints (Z-scores) of SWF and its drivers from 2000 to 2024. SWF showed negative anomalies in the early 2000s, approached near-normal conditions during 2007–2011, and remained predominantly positive after 2012, with the highest anomalies occurring around 2022–2023. PD and NTL shifted from negative to positive anomalies during the middle to late study period, whereas ISF and HAI changed from positive anomalies in the early years to negative anomalies in the later years. CF became persistently negative after approximately 2014, especially from 2015 onward. Climatic signals were more episodic: precipitation alternated between dry and wet years. The air temperature coefficient became positive after 2021, whereas the RH and LST coefficients showed substantial interannual fluctuations, and NDVI anomalies fluctuated without a clear monotonic shift; the later period was characterized by greater water persistence and stronger population and NTL signals, whereas ISF and HAI did not show comparable strengthening.

3.3 Analysis of spatiotemporal heterogeneity of driving factors

Figure 6 ranks the variability of SWF and its potential drivers using the coefficient of variation (CV). ISF ($CV = 3.81$) and PD ($CV = 3.40$) showed the highest variability, both exceeding the $CV = 1.00$ reference line and indicating highly uneven grid-level distributions. SWF ($CV = 1.72$) and CF ($CV = 1.64$) also exhibited substantial variability, whereas NTL ($CV = 1.16$) was only slightly above the reference line. Slope ($CV = 0.90$) showed lower variability. Air temperature ($CV = 0.39$), HAI ($CV = 0.36$), NDVI ($CV = 0.34$), LST ($CV = 0.18$), precipitation ($CV = 0.14$), elevation ($CV = 0.13$), and RH ($CV = 0.11$) remained below 0.40. Overall, the CV ranking highlighted stronger heterogeneity in ISF, PD, SWF, and CF, while most climate, terrain, vegetation, and HAI variables showed lower variability.

Fig. 4. Interannual trends of SWA and driving factors in the study area during 2000–2024. (a) annual SWA; (b) annual total precipitation; (c) annual mean 2-m air temperature; (d) RH; (e) LST; (f) NDVI; (g) CF; (h) PD; (i) NTL; (j)

ISF; (k) HAI; (l) Pearson' s correlation coefficient (r) between SWA and each driver.

Visible panel annotations: (a) SWA, $R^2 = 0.912,\ P < 0.001$; (b) precipitation, $R^2 = 0.246,\ P = 0.012$; (c) air temperature, $R^2 = 0.264,\ P = 0.009$; (d) RH, $R^2 = 0.057,\ P = 0.249$; (e) LST, $R^2 = 0.171,\ P = 0.040$; (f) NDVI, $R^2 = 0.004,\ P = 0.765$; (g) CF, $R^2 = 0.759,\ P < 0.001$; (h) PD, $R^2 = 0.850,\ P < 0.001$; (i) NTL, $R^2 = 0.865,\ P < 0.001$; (j) ISF, $R^2 = 0.469,\ P < 0.001$; (k) HAI, $R^2 = 0.662,\ P < 0.001$. In panel (l), Pearson' s r values with SWA are shown as: NTL 0.89, PD 0.83, air temperature 0.63, precipitation 0.51, NDVI −0.03, RH −0.32, LST −0.46, ISF −0.73, CF −0.75, and HAI −0.77.

3.4 Multidimensional collinearity and PCA of driving factors

This PCA biplot illustrates the relationships among the driving factors based on the first two principal components. The first principal component (PC1) accounts for 46.20% of the total variance, whereas the second principal component (PC2) accounts for 23.50%, yielding a cumulative explained variance of 69.70% (Fig. 7). PC1 mainly represents a thermal-urbanization gradient. LST and temperature exhibit the strongest positive loadings on this axis and are closely aligned, indicating a strong positive association. Human activity indicators, including HAI, NTL, PD, and ISF, also project onto the positive side of PC1, suggesting that more intense anthropogenic activity co-varies with the thermal environment along this dominant gradient. In contrast, elevation and RH, together with precipitation and NDVI to a lesser extent, project onto the negative side of PC1, indicating an inverse relationship with the thermal-urbanization dimension.

Fig. 5 Spatiotemporal fingerprint (Z-score) of grid-cell SWF and driving factors during 2000–2024. Each cell represents the annual Z-score; red and blue denote positive and negative anomalies relative to the multi-year mean.

Fig. 6 Spatiotemporal heterogeneity ranking of SWF and driving factors quantified by the CV. The red dashed line denotes CV=1.00; variables above this line show high heterogeneity, whereas variables below this line indicate more stable spatiotemporal patterns.

PC2 reflects a gradient contrasting cropland and vegetation conditions with terrain constraints. CF exhibits the strongest positive loading on PC2, and NDVI also shows a clear positive contribution, indicating close associations with agricultural and vegetation conditions. By contrast, slope loads strongly in the negative direction, highlighting the constraining effect of topography on land use and vegetation distribution. Overall, the first two components capture the driving factors in terms of two dominant dimensions: thermal and anthropogenic intensification, and land-cover structure and terrain constraints.

Fig. 7 PCA biplot of candidate driving variables for grid-cell SWF. Arrows represent variable loadings on first principal component (PC1) and second principal component (PC2). Arrow direction indicates the sign of the loading, and arrow length indicates loading magnitude. Red arrows denote the human-activity proxy variables (PD, NTL, ISF and HAI) and blue arrows denote the hydroclimatic, terrain, and vegetation variables; the colors are used only to distinguish these analytical groups. Smaller angles between arrows indicate stronger positive correlations, angles close to 180.00° indicate negative correlations, and near-right angles indicate weak correlations. The dashed circle represents the unit circle.

3.5 Non-linear response analysis of driving factors based on Random Forest

To further examine the nonlinear contribution and direction of influence of each driving factor on surface water distribution, we constructed a Random Forest regression model. Signed feature importance was used to interpret the direction and relative magnitude of the model response, and residual trend analysis was used as a complementary diagnostic.

Figure 8 shows the Random Forest importance of all candidate drivers for the study area. The importance is non-negative and reflects each variable' s contribution to predicting the water indicator rather than the direction of its effect. NDVI has the highest importance, followed by elevation, LST, and slope. In this mountain-valley setting, the high importance of NDVI is consistent with a spatial pattern in which peak-season vegetation cover is greatest on vegetated hillslopes, where open-water is unlikely to accumulate, whereas persistent water is concentrated along the sparsely vegetated valley floor and the main channel. The same spatial logic applies to elevation and slope, which represent topographic positions that are generally unfavorable for open-water accumulation in this corridor. These results indicate that the model attributes much of the spatial variability in the water indicator to constraints imposed by vegetation, terrain, and thermal conditions within the mountain-valley system. Air temperature, ISF, and precipitation have comparatively small importance, and the remaining human-activity indicators, including PD, HAI, NTL, and CF, cluster near zero. This pattern suggests that these variables provided limited incremental explanatory power once other covarying predictors were included, consistent with the strong collinearity among urbanization-related variables, such as PD, NTL, and ISF.

The early-late radar comparison reveals a clear structural shift between the early (2000–2002) and late (2022–2024) periods (Fig. 8b). In normalized terms, SWF was higher in the late period, together with air temperature, precipitation, PD, and NTL, whereas CF, RH, LST, ISF, and HAI were higher in the early period. This contrast indicates that the recent increase in mapped surface

water coincided with regional warming and with intensifying population and NTL signals, but not with a sustained rise in ISF or HAI. The two panels are therefore mutually consistent: vegetation, terrain, and thermal gradients carried most of the predictive information (Fig. 8a), while the temporal intensification of human activity was largely confined to the population and NTL signals rather than extending to ISF or HAI.

Fig. 8 Random Forest importance (a) and early-late min-max-normalized profiles (b) of SWF and candidate driving variables. In Figure 8a, the importance is non-negative and bar length represents the magnitude of the importance, which does not indicate the direction of the effect. Values in Figure 8b were normalized to 0–1 across the annual series before averaging within each period.

To further evaluate whether the Random Forest model captured nonlinear relationships in heterogeneous spatial data without systematic bias, we conducted a binned residual trend analysis for each driving factor. The residual, defined as the difference between observed and model predicted values, provides a straightforward diagnostic of model bias.

As shown in Figure 9, residuals for the candidate drivers are generally centered around the zero baseline. The local mean curves derived from quantile binning are nearly horizontal and remain close to zero, indicating no clear systematic drift across the observed variable ranges. This pattern suggests that no strong single-factor residual structure remains after model fitting. This pattern is also observed for topographic and climatic factors with strong spatial heterogeneity, such as elevation and precipitation.

These results suggest that the Random Forest model adequately captured the main nonlinear relationships between surface water and the elevation and hydrothermal gradients. Human activity variables, including NTL, ISF, and PD, are strongly right-skewed in space, with weak disturbance across most areas but much stronger disturbance in a few core valley zones. Even in these high-value areas, however, the binned residual means remain close to zero, indicating no clear tendency toward either underestimation or overestimation. Overall, the residual diagnostics support the adequacy of the model for the present attribution analysis.

3.6 Spatiotemporal evolution of driving mechanisms and factor interaction detection

To explore the temporal non-stationarity of factor explanatory power and the interactive enhancement among drivers, we used the factor detector and interaction detector modules of GeoDetector.

3.6.1 Temporal evolution of driving forces

To investigate temporal variation in explanatory power, we calculated annual q-values for each driving factor from 2000 to 2024. As shown in Figure 10, the spatial determinants of surface water were not static but changed over time.

NDVI generally retained relatively high explanatory power throughout the study period, indicating that vegetation-related spatial gradients remained closely associated with the surface

water pattern. Elevation was also important in the early years but weakened after the middle of the study period. Precipitation showed pulse-like variability, with higher q-values in years such as 2008, 2010, 2012, and 2023, suggesting sensitivity to interannual hydroclimatic variability. Human activity indicators, including NTL, ISF, and PD, generally showed lower basin-scale explanatory power than NDVI, although they exceeded 0.10 in several years and may still exert localized effects in developed valley reaches.

Fig. 9 Residual binning trend analysis of driving factors.

(a) Precipitation: $R^2 = 0.000$, $P = 0.988$.

(b) Air temperature: $R^2 = 0.105$, $P = 0.239$.

(c) RH: $R^2 = 0.168$, $P = 0.129$.

(d) LST: $R^2 = 0.353$, $P = 0.019$.

(e) NDVI: $R^2 = 0.631$, $P < 0.001$.

(f) Elevation: $R^2 = 0.043$, $P = 0.456$.

(g) Slope: $R^2 = 0.261$, $P = 0.051$.

(h) PD: $R^2 = 0.203$, $P = 0.092$.

(i) NTL: $R^2 = 0.524$, $P = 0.003$.

(j) ISF: $R^2 = 0.693$, $P = 0.020$.

(k) CF: $R^2 = 0.502$, $P = 0.033$.

(l) HAI: $R^2 = 0.105$, $P = 0.332$.

Fig. 9 Residual binning trend analysis of driving factors. (a), precipitation; (b), air temperature; (c), RH; (d), LST; (e), NDVI; (f), elevation; (g), slope; (h), PD; (i), NTL; (j), ISF; (k), CF; (l), HAI.

Fig. 10 Temporal trends of q-values for each driving factor

3.6.2 Interaction mechanisms of driving factors

The 2001 GeoDetector interaction heatmap shows that two-factor interactions generally had stronger explanatory power than single factors (Fig. 11). All tested factor pairs show nonlinear enhancement, with interaction q-values exceeding the corresponding single-factor q-values. For example, the single-factor q-values of NDVI and slope are 0.25 and 0.07, respectively, whereas their interaction value reaches 0.79. This result indicates that the mapped surface water pattern in the Huangshui River Basin is associated with coupled environmental gradients rather than additive single-factor effects.

Fig. 11 Interaction detector heatmap of driving factors for surface water in the study area in 2001. Diagonal values represent single-factor q-values and off-diagonal values represent two-factor interaction q-values.

The strongest interactions involve terrain and surface biophysical variables. $NDVI \cap slope$ ($q = 0.79$) and $NDVI \cap elevation$ ($q = 0.69$) have the highest values in the interaction matrix. These results show that elevation and slope structure the spatial setting of surface water together with vegetation conditions and land surface thermal patterns. This coupling is consistent with the mountain-valley setting, where runoff concentration, storage potential, vegetation cover, and hydrothermal conditions vary strongly along topographic gradients.

Topography also strengthens the explanatory power of human-activity proxies. The single-factor q-values of PD and ISF are low (0.05 and 0.10, respectively), but their interactions with elevation increase the q-values to 0.43 and 0.50, respectively. This pattern indicates that PD and ISF are most informative under specific topographic conditions, especially in low-elevation valley reaches where settlement, infrastructure, and mapped surface water are spatially concentrated.

4 Discussion

4.1 Scale-dependent anthropogenic impacts: natural vs. urban river reaches

Our Random Forest attribution results should be interpreted within the specific regional context of the Xining-Haidong Corridor rather than as evidence of a generic urban-water relationship. In the Huangshui River Basin, urban land, cropland, transport corridors, and persistent open-water are all concentrated on the same low-slope valley floor. Consequently, ISF and NTL partly act as proxies for corridor occupation within the limited geomorphic space where both settlement and stable surface water can persist. This co-location helps explain why ISF shows only a weak positive association with mapped water and why NTL adds little independent explanatory power once correlated predictors are taken into account (Debeer and Strobl, 2020; Kaneko, 2022; Wies et al., 2023). This pattern is also consistent with studies showing that urban water fluxes, impervious surfaces, and managed channel environments are often coupled in developed river corridors (Best, 2019; Oswald et al., 2023; Banner et al., 2024). In the Hehuang Valley, recent vegetation change has been strongly influenced by human activities; settlement expansion and land development have been compressed into narrow valley-bottom areas, and cropland restructuring has further altered ecological conditions in the Huangshui River Basin (Wei et al., 2021; Li and Xin, 2024; Xu et al., 2024).

At the reach scale, however, the weak positive ISF-water association should be interpreted cautiously because the temporal pattern of ISF did not show a consistent increase. A more plausible explanation is that impervious surfaces are spatially co-located with persistent valley-bottom water bodies and managed water environments, including embanked channels, localized impoundments, storage waters, and other regulated open-water features. This interpretation is consistent with recent urban-hydrology syntheses indicating that impervious surfaces, channel confinement, and managed water bodies often occur together in developed corridors rather than as isolated processes (Best, 2019; Oswald et al., 2023; Banner et al., 2024). Recent work on the Huangshui River shows that human activities already exert a measurable influence on runoff and ecological flow, while studies of plateau urban rivers further indicate that rubber-dam operation and similar forms of regulation can reorganize local water persistence and habitat conditions (Mao et al., 2021; Liu et al., 2024). In this sense, ISF and NTL are best understood as spatial indicators of urban occupation and managed water regulation, rather than as direct evidence of a temporal expansion effect of the impervious-surface indicator or spontaneous surface water expansion. This interpretation also aligns with broader evidence from the Yellow River Basin that impervious surfaces are important ecological signals in intensively developed corridors (Zhang et al., 2023).

4.2 Topography as a spatial regulator: deconstructing the spatial mismatch between rainfall and water accumulation

The negative spatial association between annual precipitation and mapped surface water does not mean that rainfall is unimportant in the Huangshui River Basin. Rather, it indicates that, in a

mountain-valley corridor, areas receiving precipitation and areas where open-water persists occupy different parts of the landscape. Rainfall is concentrated over higher and steeper terrain, whereas persistent open-water is mainly distributed along the main river, wider tributaries, reservoirs, and relatively flat valley bottoms. This pattern is consistent with a typical mountain-routing system, but in the Huangshui River Basin, it is conditioned by a much narrower and more intensively used valley system than that of many plateau interior basins (Viviroli et al., 2007, 2020; Immerzeel et al., 2020).

This spatial mismatch is further reinforced by the regional structure of the Hehuang Valley. The same low-elevation belt that promotes runoff concentration also contains the densest settlements, cultivated land, transport infrastructure, and river engineering works in eastern Qinghai Province. Recent basin-scale studies further indicate that ecological flow, ecohydrological regime, and agricultural restructuring in the Huangshui system have already been shaped by intensive human occupation and regulation (Wei et al., 2021; Fan et al., 2024;

Li and Xin, 2024; Liu et al., 2024; Xu et al., 2024). As a result, the mapped surface water pattern reflects not only where water is generated, but also where flow is stored, diverted, constrained, or maintained through channel regulation and valley-floor use.

Plateau-scale studies show that warming and wetting have generally supported surface water expansion in many parts of the Qinghai-Xizang Plateau. However, the northeastern corridor should not be interpreted as a simple local rainfall-to-water system (Konapala et al., 2020; Ran et al., 2023; Sui et al., 2023; Xiao et al., 2024). In the Huangshui River Basin, the climatic background affects runoff supply and interannual wetness, while topography controls where water can accumulate and where human use is physically feasible. These two controls are further reshaped by urban occupation and river regulation within the corridor.

For this reason, topography is better understood here as a spatial regulator rather than as a static background variable. It constrains the feasible locations of both surface water and human activity, and it amplifies human influence by concentrating development and regulation within the same valley-floor belt where open-water is observable (Viviroli et al., 2007, 2020; Veldkamp et al., 2017). This interpretation is consistent with the stronger interaction effects between elevation and human variables in the GeoDetector results: human-activity variables alone explain little, but its explanatory power rises sharply once the topographic setting is taken into account (Wang et al., 2010, 2016).

4.3 Engineering regulation as an explanatory context

To assess the role of dams and sluices, we compiled georeferenced records of barriers and reservoirs from the Global Dam Watch database (Lehner et al., 2024) and overlaid them with the riparian corridor and representative-year water maps. Candidate structures were retained only when they had a clear hydrological association with the corridor. Specifically, they either occur within the defined riparian corridor or are located on hydrologically connected upstream reaches whose regulated discharge enters the Huangshui River network. This criterion follows recent global studies of river connectivity and barrier inventories, which emphasize hydrological connectivity rather than straight-line distance alone (Grill et al., 2019; Lehner et al., 2024). The latter category therefore includes upstream storage facilities and inter-basin inflows that can physically modulate river stage within the study area, including key upstream reservoirs and the broader Qinghai Lake drainage context, rather than structures selected by an arbitrary distance buffer.

Several persistent valley-floor water patches coincide with barrier locations or nearby reservoir polygons. This spatial correspondence suggests that regulated impoundment may contribute to local water persistence, in addition to climatic and topographic controls. Because the analysis retained only barriers with a plausible hydrological connection to the corridor, this interpretation is based on structures that could reasonably influence the observed water distribution in the Xining-Haidong Corridor, rather than on a purely distance-based inventory. This interpretation is also consistent

with previous studies showing that dams, reservoirs, and smaller river barriers can modify upstream water levels, promote local ponding, regulate flow, and alter river connectivity, including studies of rubber-dam operation in plateau urban rivers of the same region (Best, 2019; Grill et al., 2019; Mulligan et al., 2020; Mao et al., 2021; Zhang and Gu, 2023).

However, these facilities were used only as spatial explanatory context. A continuous annual database of dam or sluice operation was not available for 2000–2024, and the candidate records do not document annual operational intensity. This limitation is consistent with the current status of large barrier inventories, which provide robust information on structural locations and attributes but do not yet offer basin-wide and facility-level annual operation records (Mulligan et al., 2020; Zhang and Gu, 2023; Lehner et al., 2024). Engineering regulation was therefore not introduced as a yearly predictor in the main model.

4.4 Methodological assumptions, uncertainties, and perspectives

The GEE-Random Forest-GeoDetector framework provides an effective approach for reconstructing long-term water dynamics, but it identifies association patterns rather than strict causal relationships. This distinction is particularly important in the Huangshui River Basin where topography, land use, urban expansion, and river regulation are closely coupled (Breiman, 2001; Wang et al., 2010; Wang et al., 2016; Tyralis et al., 2019).

The 30 m optical mapping framework is most reliable for the Huangshui River main stem, wider tributaries, reservoirs, and other medium-to-large open-water patches. By contrast, small wetlands, narrow irrigation channels, seasonal ponds, and other micro-scale water features were likely underestimated or omitted (Dong et al., 2022; Vanderhoof et al., 2023). The reported annual series and representative-year maps were derived from the same three-expert ensemble voting scheme applied to Landsat imagery and are internally consistent across 2000–2024. The JRC Yearly History product was used only as an independent temporal-consistency reference over 2000–2021, and did not serve as the source of Figure 3 or the reported annual area series (Pekel et al., 2016). Nevertheless, the post-2021 period coincides with an increase in valid observations associated with Landsat 9, which extended the Landsat record and improved mission redundancy and observation continuity (Masek et al., 2020). This increase may enhance the detection of ephemeral open-water. Therefore, the interannual signal in the most recent years should be interpreted as mapped open-water extent, rather than as a direct measure of runoff or water-resource abundance. The conclusions of this study are therefore most robust for medium-to-large open-water dynamics within the corridor.

Engineering regulation represents another important source of uncertainty. Because no consistent annual dataset of dam and sluice operation was available, the influence of regulation could be interpreted in a spatial context but could not be directly quantified within the annual driver-attribution framework (Mulligan et al., 2020; Zhang and Gu, 2023; Lehner et al., 2024).

5 Conclusions

By applying a three-expert ensemble voting scheme to Landsat surface reflectance, this study reconstructed mapped open-water dynamics in the Huangshui River Basin within the Xining-Haidong Corridor from 2000 to 2024. It further examined the underlying controls through trend analysis, a RESTREND-inspired Random Forest attribution framework, and GeoDetector. Mapped open-water showed a significant long-term increase, and both the manual accuracy assessment and the JRC overlapping-window comparison supported the reliability of the extraction trend. Topography remained the dominant basin-scale control on the spatial persistence of open-water. Within the corridor, persistent open-water in the most developed reaches was spatially associated with concentrated valley-floor land use and stronger population and NTL signals, and was further influenced by dam- and reservoir-related regulation. The impervious-surface indicator showed a different temporal pattern and should therefore be interpreted cautiously as a spatial co-location proxy rather than direct evidence of continuous expansion. Together, these findings support the interpretation of this arid mountain corridor as a coupled natural-human water system rather than a purely climate-driven basin. These conclusions should be considered within the methodological limits of the analysis. The results are most reliable for medium-to-large open-water bodies, whereas small wetlands and narrow channels were likely undetected. The most recent magnitudes also require caution because Landsat 9 increased observation availability and recent warm seasons were wetter than usual. Future work should incorporate sub-seasonal mapping, higher-resolution imagery, and continuous annual records of dam and sluice operation. These additions would improve the quantitative attribution of engineering regulation in the Huangshui River Basin and provide stronger support for water-resource management and ecological restoration on the northeastern Qinghai-Xizang Plateau.

Conflict of interest

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

Acknowledgements

The authors acknowledge Google Earth Engine (GEE) for providing cloud-based geospatial processing support and the GitHub open-source community for publicly available resources that supported data processing and figure preparation.

Author contributions

Conceptualization: CHENG Ruixuan; Methodology: CHENG Ruixuan; Formal analysis: CHENG Ruixuan, YIN Sijiang; Data curation: YIN Sijiang; Software: YIN Sijiang; Visualization: CHENG Ruixuan; Validation: QIN Ke; Supervision: QIN Ke; Writing - original draft preparation: CHENG Ruixuan; Writing - review and editing: QIN Ke. All authors approved the manuscript.

References

Banner J L, Black B A, Tremaine D M. 2024. Positive unintended consequences of urbanization for climate-resilience of stream ecosystems. npj Urban Sustainability, 4: 16, doi: 10.1038/s42949-024-00144-1.

Best J. 2019. Anthropogenic stresses on the world' s big rivers. Nature Geoscience, 12(1): 7–21.

Breiman L. 2001. Random forests. Machine Learning, 45: 5–32.

Debeer D, Strobl C. 2020. Conditional permutation importance revisited. BMC Bioinformatics, 21: 307, doi: 10.1186/s12859-020-03622-2.

Dong Y T, Fan L B, Zhao J, et al. 2022. Mapping of small water bodies with integrated spatial information for time series images of optical remote sensing. Journal of Hydrology, 614(Part B): 128580, doi: 10.1016/j.jhydrol.2022.128580.

Evans J, Geerken R. 2004. Discrimination between climate and human-induced dryland degradation. Journal of Arid Environments, 57(4): 535–554.

Fan L J, Liu L X, Hu J, et al. 2024. A long-term evaluation of the ecohydrological regime in a semiarid basin: A case study of the Huangshui River in the Yellow River Basin, China. Hydrology, 11(10): 168, doi: 10.3390/hydrology11100168.

Feyisa G L, Meilby H, Fensholt R, et al. 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sensing of Environment, 140: 23–35.

Gorelick N, Hancher M, Dixon M, et al. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202: 18–27.

Grill G, Lehner B, Thieme M, et al. 2019. Mapping the world' s free-flowing rivers. Nature, 569(7755): 215–221.

Heritage G, Entwistle N. 2020. Impacts of river engineering on river channel behaviour: Implications for managing downstream flood risk. Water, 12(5): 1355, doi: 10.3390/w12051355.

Immerzeel W W, Lutz A F, Andrade M, et al. 2020. Importance and vulnerability of the world' s water towers. Nature, 577(7790):

364–369.

Jones J W. 2019. Improved automated detection of subpixel-scale inundation –Revised Dynamic Surface Water Extent (DSWE) partial surface water tests. Remote Sensing, 11(4): 374, doi: 10.3390/rs11040374.

Kaneko H. 2022. Cross-validated permutation feature importance considering correlation between features. Analytical Science Advances, 3(9–10): 278–287.

Konapala G, Mishra A K, Wada Y, et al. 2020. Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. Nature Communications, 11: 3044, doi: 10.1038/s41467-020-16757-w.

Lehner B, Beames P, Mulligan M, et al. 2024. The Global Dam Watch database of river barrier and reservoir information for large-scale applications. Scientific Data, 11: 1069, doi: 10.1038/s41597-024-03752-9.

Li R, Zhu G F, Lu S Y, et al. 2023. Effects of urbanization on the water cycle in the Shiyang River Basin: Based on a stable isotope method. Hydrology and Earth System Sciences, 27(24): 4437–4452.

Li X Y, Xin Z B. 2024. Land-use composition, distribution patterns, and influencing factors of villages in the Hehuang Valley, Qinghai, China, based on UAV photogrammetry. Remote Sensing, 16(12): 2213, doi: 10.3390/rs16122213.

Liu L X, Fan L J, Hu J, et al. 2024. Human activities impacts on runoff and ecological flow in the Huangshui River of the Yellow River Basin, China. Water, 16(16): 2331, doi: 10.3390/w16162331.

Mann H B. 1945. Nonparametric tests against trend. Econometrica, 13(3): 245–259.

Mao X F, Wei X Y, Engel B, et al. 2021. Biological response to 5 years of operations of cascade rubber dams in a plateau urban river, China. River Research and Applications, 37(8): 1201–1211.

Masek J G, Wulder M A, Markham B, et al. 2020. Landsat 9: Empowering open science and applications through continuity. Remote Sensing of Environment, 248: 111968, doi: 10.1016/j.rse.2020.111968.

Mulligan M, van Soesbergen A, Sáenz L. 2020. GOODD, a global dataset of more than 38,000 georeferenced dams. Scientific Data, 7: 31, doi: 10.1038/s41597-020-0362-5.

Oswald C J, Kelleher C, Ledford S, et al. 2023. Integrating urban water fluxes and moving beyond impervious surface cover: A review. Journal of Hydrology, 618: 129188, doi: 10.1016/j.jhydrol.2023.129188.

Pekel J F, Cottam A, Gorelick N, et al. 2016. High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633): 418–422.

Pickens A H, Hansen M C, Hancher M, et al. 2020. Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full

Landsat time-series. Remote Sensing of Environment, 243: 111792, doi: 10.1016/j.rse.2020.111792.

Ran Q W, Aires F, Ciais P, et al. 2023. The status and influencing factors of surface water dynamics on the Qinghai-Tibet Plateau during 2000–2020. IEEE Transactions on Geoscience and Remote Sensing, 61: 4200114, doi: 10.1109/TGRS.2022.3231552.

Sen P K. 1968. Estimates of the regression coefficient based on Kendall' s Tau. Journal of the American Statistical Association, 63(324): 1379–1389.

Sui T B, Ye C M, Tang R, et al. 2023. Spatio-temporal change monitoring for surface water on the Qinghai-Tibet Plateau from 1990 to 2020 using remote sensing. Frontiers in Earth Science, 11: 1124425, doi: 10.3389/feart.2023.1124425.

Tyralis H, Papacharalampous G A, Langousis A. 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water, 11(5): 910, doi: 10.3390/w11050910.

Vanderhoof M K, Alexander L, Christensen J, et al. 2023. High-frequency time series comparison of Sentinel-1 and Sentinel-2 satellites for mapping open and vegetated water across the United States (2017–2021). Remote Sensing of Environment, 288: 113498, doi: 10.1016/j.rse.2023.113498.

Veldkamp T I E, Wada Y, Aerts J C J H, et al. 2017. Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century. Nature Communications, 8: 15697, doi: 10.1038/ncomms15697.

Viviroli D, Dürr H H, Messerli B, et al. 2007. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water Resources Research, 43(7): W07447, doi: 10.1029/2006WR005653.

Viviroli D, Kummu M, Meybeck M, et al. 2020. Increasing dependence of lowland populations on mountain water resources. Nature Sustainability, 3(11): 917–928.

Wang J F, Li X H, Christakos G, et al. 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. International Journal of Geographical Information Science, 24(1): 107–127.

Wang J F, Zhang T L, Fu B J. 2016. A measure of spatial stratified heterogeneity. Ecological Indicators, 67: 250–256.

Wei H, Lu C H, Liu Y Q. 2021. Farmland changes and their ecological impact in the Huangshui River Basin. Land, 10(10): 1082,

doi: 10.3390/land10101082.

Wessels K J, Prince S D, Malherbe J, et al. 2007. Can human-induced land degradation be distinguished from the effects of rainfall variability? A case study in South Africa. Journal of Arid Environments, 68(2): 271–297.

Wies C, Miltenberger R, Grieser G, et al. 2023. Exploring the variable importance in random forests under correlations: A general concept applied to donor organ quality in post-transplant survival. BMC Medical Research Methodology, 23: 209, doi: 10.1186/s12874-023-02023-2.

Xiao Z, Ding M J, Li L H, et al. 2024. Divergent changes of surface water and its climatic drivers in the headwater region of the Three Rivers on the Qinghai-Tibet Plateau. Ecological Indicators, 158: 111615, doi: 10.1016/j.ecolind.2024.111615.

Xu B, Mao X F, Li X Y, et al. 2024. Anthropogenic activities dominated the spatial and temporal changes of normalized difference vegetation index (NDVI) in the Hehuang Valley in the northeastern Qinghai Province between 2000 and 2020. Frontiers in Environmental Science, 12: 1384032, doi: 10.3389/fenvs.2024.1384032.

Xu H Q. 2006. Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing, 27(14): 3025–3033.

Zhang A T, Gu V X. 2023. Global Dam Tracker: A database of more than 35,000 dams with location, catchment, and attribute information. Scientific Data, 10: 111, doi: 10.1038/s41597-023-02008-2.

Zhang J, Du J Q, Fang S F, et al. 2023. Dynamic changes, spatiotemporal differences, and ecological effects of impervious surfaces in the Yellow River Basin, 1986–2020. Remote Sensing, 15(1): 268, doi: 10.3390/rs15010268.

Note: Figure translations are in progress. See original paper for figures.