SATELLITE REMOTE SENSING AND MACHINE LEARNING TO MONITOR SURFACE WATER RESOURCES IN ETHIOPIA

MATHIAS TESFAYE ABEBE

Department of Landscape Ecology and Resources Management

Justus Liebig University Giessen

Dissertation for the degree of

Doctor of Natural Sciences (Dr. rer. nat.)

submitted by Mathias Tesfaye Abebe, M.A.

Referees from the Justus Liebig University Giessen

1st Supervisor: Prof. Dr. Lutz Breuer

2nd Supervisor: Prof. Dr. Till Kleinebecker

Examiner: Prof. Dr. Jan Siemens

Examiner: Prof. Dr. Suzanne Jacobs

Chairperson: Prof. Dr. Jakob Santner

Submitted: 15.06.2025

Abstract

Surface water conservation and management require monitoring of surface water resources at national and site-specific scales. Additionally, investigating the effects of climate, land use, and land cover (LULC) on surface water resources is necessary to ensure their sustainability. Even though it is significant, there is a shortage of practical techniques and thorough evaluations for managing surface water resources and their factors in Ethiopia. In a first step, I therefore evaluate several water indices to detect water bodies at the national scale in Ethiopia with Sentinel remote sensing data. Second, machine learning methods are assessed running on the Google Earth Engine (GEE) with Landsat data at four sites in Ethiopia during 1986 to 2023 to analyze spatiotemporal surface water distribution and its trends. Finally, the effects of climate and LULC on surface water resources are evaluated through the use of machine learning and statistical models. In the first step, the findings indicate that the water index (WI) and the automatic water extraction index with shadows (AWEIsh) are the most effective for detecting water bodies, exhibiting overall accuracies of 96% and 95%, respectively. The extent of surface water covered by both indices is almost equivalent, at 82,650 km² (WI) and 86,530 km² (AWEIsh). Second, the classification accuracy achieved through gradient tree boosting (GTB), support vector machines (SVMs), and random forest (RF) is satisfying. In general, classification accuracy is greater than 90% for both producer and user accuracies, albeit GTB achieves a slightly higher accuracy. From 1986 to 2023, a long-term upward trend in the extent of water bodies is observed at the four sites, with significant interannual fluctuations. The water occurrence frequency (WOF) shows that 84–94% of the water remains in a stable condition, whereas the rest of the water area changes throughout time. Finally, extreme gradient boosting (XGBoost) is the optimal model for evaluating and interpreting the influence of climate and LULC on surface water changes. The dynamics of surface water in this study area are mainly impacted by precipitation, minimum temperature and LULC change. Overall, this dissertation presents spatiotemporal insights for monitoring water bodies in Ethiopia. The findings underscore the necessity of an integrated perspective on water policy and management, considering the interaction between climate variability, LULC changes, and surface water resources.

Zusammenfassung

Die Überwachung der Verteilung von Oberflächenwasser auf regionaler und nationaler Ebene ist für die Unterstützung der Bewirtschaftung und Erhaltung von Oberflächenwasser von entscheidender Bedeutung. Darüber hinaus ist die Analyse der Auswirkungen von Klima, Wetter, Landnutzung und Landbedeckung auf die Dynamik des Oberflächenwassers relevant. Trotz ihrer Bedeutung mangelt es in Äthiopien an operativen Methoden und umfassenden Bewertungen zur Überwachung der Oberflächenwasserressourcen und ihrer Einflussfaktoren.

In einem ersten Schritt werden in dieser Arbeit daher mehrere Wasserindizes getestet, um Oberflächenwasser auf nationaler Ebene in Äthiopien mithilfe von Sentinel-Daten zu erfassen. Die Ergebnisse zeigen, dass der Wasserindex (WI) und der automatische Wasserentnahmeindex mit Schatten (AWEIsh) mit einer Gesamtgenauigkeit von 96% bzw. 95% am genauesten für die Erkennung von Oberflächenwasser sind. Beide Indizes detektieren mit 82.650 km² (WI) und 86.530 km² (AWEIsh) nahezu die gleiche räumliche Abdeckung von Oberflächenwasser.

Zweitens werden maschinelle Lernverfahren in Google Earth Engine (GEE) unter Verwendung von Landsat-Daten für die Oberflächenwasserüberwachung an vier Standorten in Äthiopien von 1986 bis 2023 evaluiert. Obwohl Gradient Tree Boosting (GTB), Support Vector Machine (SVM) und Random Forest (RF) alle eine ausgezeichnete Klassifizierungsgenauigkeit aufweisen, mit einer Gesamt-, Produzenten- und Benutzergenauigkeit von durchweg über 90%, schneidet GTB etwas besser ab als die anderen. Von 1986 bis 2023 zeigen die Oberflächengewässerdynamiken der vier Untersuchungsgebiete einen langfristigen Aufwärtstrend mit erheblichen Schwankungen zwischen den Jahren. Die Häufigkeit des Wasserauftretens im Laufe der Zeit zeigt, dass 84–94% des Wassers permanent vorhanden sind, während sich die verbleibenden Wasserflächen im Laufe der Zeit verändern.

Schließlich werden maschinelle Lernverfahren und statistische Modelle geprüft, um die Auswirkungen des Klimas und der Landnutzung/Landbedeckung (LULC) auf die Oberflächenwasserdynamik zu bewerten. Extreme Gradient Boosting (XGBoost) ist das effektivste und interpretierbarste Modell zur Analyse der Auswirkungen von Klima- und LULC-Faktoren auf die Oberflächengewässerdynamik. Niederschlag, Mindesttemperatur und LULC-Veränderungen haben einen großen Einfluss auf die Oberflächenwasserdynamik.

Insgesamt liefert diese Arbeit detaillierte Ergebnisse und räumlich-zeitliche Informationen für die Überwachung von Oberflächengewässern in Äthiopien. Diese Studie unterstreicht auch die Notwendigkeit eines integrierten Ansatzes für die Wasserpolitik und -bewirtschaftung, der die Wechselwirkungen zwischen Klimavariabilität, LULC-Veränderungen und der Dynamik der Oberflächenwasserressourcen in Äthiopien berücksichtigt.

Table of Contents

List of tables

List of figures

Scientific contributions

Peer-reviewed original research papers

Tesfaye, M., & Breuer, L. (2024). Performance of water indices for large-scale water resources monitoring using Sentinel-2 data in Ethiopia. *Environmental Monitoring and Assessment*, *196*(5), 467. https://doi.org/10.1007/s10661-024-12630-1.

Author contribution: Mathias Tesfaye: conceptualization, methodology, formal analysis, validation and draft preparation. Lutz Breuer: conceptualization, validation, supervision, writing, review and editing.

Tesfaye, M., & Breuer, L. (2025a). Remote sensing with machine learning for multi-decadal surface water monitoring in Ethiopia. *Scientific Reports*, *15*(1), 12444. https://doi.org/10.1038/s41598-025-96955-y.

Author contribution: Mathias Tesfaye: conceptualization, methodology, formal analysis, validation and draft preparation. Lutz Breuer: conceptualization, validation, supervision, writing, review and editing.

Tesfaye, M., & Breuer, L. (2025b). Influence of climate, land use, and land cover on surface water dynamics in the Ethiopian highlands. Manuscript number: RSASE-D-25-01440. In review in *Remote Sensing Applications: Society and Environment.*

Author contribution: Mathias Tesfaye: conceptualization, methodology, formal analysis, validation and draft preparation. Lutz Breuer: conceptualization, validation, supervision, writing, review and editing.

1 Extended summary

1.1 Introduction

Water-related Sustainable Development Goals

Meeting numerous of the United Nations' Sustainable Development Goals (SDGs) relies on the availability of water and the sustainable management of water resources, particularly SDG 6 "Clean water and sanitation" and its target 6.6 "Protect and restore water-related ecosystems." These aims underscore the importance of providing a quantifiable indicator 6.6.1 "Change in the extent of water-related ecosystems over time" (Dickens et al., 2017). Surface water monitoring is crucial for preserving terrestrial ecosystems which are mainly reliant on surface water availability. SDG 15 "Life on land", emphasizes preserving, rehabilitating, and fostering the sustainable use of terrestrial ecosystems, which is vital for ensuring ecosystem conservation and preventing the degradation of ecosystems. Additionally, surface water is highly sensitive to climate variability, and its monitoring is critical for climate adaptation strategies. Changes in precipitation, increasing temperatures, and rising evapotranspiration rates affect water bodies globally. SDG 13 "Climate action" indicates the need to take rapid measures to mitigate climate change and its effects on water resources. Water-related issues have long been a concern at the United Nations and in international expert meetings. In December 2016, the United Nations General Assembly (UNGA) approved measures stating the "International Decade for Action – Water for Sustainable Development" (2018–2028). This initiative aims to support the achievement of SDG 6 and the 2030 Agenda for Sustainable Development's water-related targets through national leadership and global partnerships.

Surface water resources and their drivers

Surface water covers about 3% of global land cover and consists of lakes, rivers, streams, ponds, reservoirs, and other inland water bodies (Pekel et al., 2016). Despite its limited share, it is used for daily human needs, agriculture, hydropower and industry. Surface water not only supports the preservation of biodiversity and ecosystem services (Dudgeon et al., 2006; Vörösmarty et al., 2010) but also has a vital role in the overall hydro-climate process (Chahine, 1992). Additionally, surface water is a critical component of global biogeochemical cycles (Aufdenkampe et al., 2011) including the carbon cycle, and contributes to the emission of carbon dioxide (CO_2) (Cole et al., 2007; Raymond et al., 2013) and methane (CH_4) (Holgerson & Raymond, 2016).

Surface water resources are affected by the pressures of land use changes and climate variability (Gaines et al., 2022; Vörösmarty et al., 2010). These factors alter the occurrence and pattern of surface water, as well as its quality and the hydrological cycle, resulting in ecological and socio-economic consequences (IPCC, 2021). Climate variability impacts inland water bodies by affecting rainfall, temperature, and evapotranspiration. These variations affect flow rates in streams and lead to more severe hydroclimatic events, such as droughts and floods (Huntington, 2006; Kundzewicz et al., 2008). Furthermore, surface water resources face mounting pressure from LULC changes driven by deforestation, urban sprawl, and agricultural intensification. These transformations escalate runoff volumes, shift infiltration patterns, and degrade soils, adversely affecting aquatic quality (Foley et al., 2005). Synergistic interactions between LULC shifts and climatic variability further heighten the vulnerability of these resources. Therefore, understanding these interactions and their impacts on water resources is critical for effective management of water resources, conservation of biodiversity, and adaptation planning.

Despite existing studies showing that climate and LULC drivers influence surface water dynamics, the degree and magnitude of these impacts vary by location. For example, the impact of climate and land use on water bodies in a semi-arid area of Australia was investigated using a linear mixed-effects model. This study found that the climate is the primary hydrological driver in the northwestern part of the Murray-Darling Basin (MDB), whereas land use exerts greater control in the central MDB. (Tulbure & Broich, 2019). Likewise, a linear mixed-effects model, random forest (RF), and mixed-effects RF were used to investigate the impact of climate and human activities on inland water bodies in the United States from 2000 to 2018, indicating that anthropogenic factors have a superior effect than climate factors (Gaines et al., 2022). Additionally, statistical analyses with correlation, were conducted to examine the influence of explanatory variables on water resources. These examines that vegetation and anthropogenic factors, such as dam and reservoir expansion, are significant explanatory variables for surface water (Zeng et al., 2020a). Furthermore, vegetation dynamics were found to play a crucial role in lake fluctuation (Mulu et al., 2024). Prior studies have also used RF and statistical analysis to identify the impact of explanatory variables on water body distribution and occurrence. Therefore, they underscore the importance of local and regional investigations into how inland water bodies are impacted by climate and LULC drivers using various modeling methods in the study area.

Surface water resources in Ethiopia

Ethiopia, the water tower of Africa, consists of 12 major river basins and several lakes spanning diverse climates and LULC types. However, this potential has not been fully utilized or converted into growth because of inadequate management and technological limitations in the water sector (Berhanu et al., 2014). Ethiopia's water-related ecologies encounter various issues, including interannual climate variability, altering water cycles (Seyoum et al., 2015), and escalating sedimentation (Gadissa et al., 2018; Haregeweyn et al., 2012). Despite being widely recognized as having abundant water bodies, which is disproportionately accessible in Ethiopia. Insufficient water bodies are available in the eastern part of Ethiopia, whereas ample surface water resources are prevalent in the western part. The northeastern part of Ethiopia has a diverse climate and substantial LULC transformation (Tadesse et al., 2025), which has a significant impact on surface water resources. Moreover, there is extensive flooding, especially in the plains of the Awash River Basin (Legass et al., 2025a). This has resulted in reduced crop yields, depleted nutrients in the soil, and severe erosion (Desta et al., 2024). However, solid evidence of long-term surface water dynamics and a thorough understanding of climate and LULC factors on surface water dynamics are lacking. This underscores the need to monitor and manage surface water resources, as well as to address their intricate interplay with climate variability and LULC changes in Ethiopia.

Several studies have examined surface water resources in Ethiopia. For example, changes in surface water in the Central Rift Valley were analyzed using NDWI from 1973 to 2014. This analysis revealed minimal changes in Lakes Langano and Shala, as well as a substantial reduction in Lake Abjata (Sisay, 2016). In the same way, surface water monitoring using NDWI from 1984 to 2017 exhibited a decrease in Lakes Chelekleka, Kiroftu, and others (Sathianarayanan, 2018). AWEI's performance in detecting surface water was superior to MNDWI's and NDWI's, and Lake Ziway's surface water area has shrunk annually from 2009 to 2018 (Asfaw et al., 2020). Additionally, the surface water dynamics of Ethiopia's major lakes have been examined using RF, revealing variability with increasing trends in Lakes Ziway, Hawassa, and Abaya and decreasing trends in Lakes Abiyata, Langano, Shalla, Chamo, and Hayq from 1985 to 2023 (Getaneh et al., 2024). To the best of my knowledge, a comprehensive evaluation of various modeling approaches for

long-term monitoring at the national level and in diverse locations in Ethiopia remains unexplored. Given its varied landscapes and environments, Ethiopia requires surface water monitoring approaches tailored to specific sites. Such approaches would improve remote sensing applications for water resources and promote sustainable and management and conservation at the regional level.

Cloud-based remote sensing and machine learning

Traditional remote sensing approaches often face challenges regarding computing power, scalability, and data storage. Whereas, recently, cloud-based remote sensing with Google Earth Engine (GEE) leverages disseminated computing to enable the efficient analysis of massive amounts of satellite images (e.g., Landsat, Sentinel, Moderate Resolution Orbiting Spectroradiometer (MODIS)) (Gorelick et al., 2017). This high-performance cloud computing platform enables monitoring natural resources, particularly temporally and spatially extensive land features, including surface water bodies. Thus, the democratization of satellite data through cloud computing platforms allows for the effective assessment and monitoring of water resources at global (Pekel et al., 2016; Pickens et al., 2020) as well as regional and local scales (Mueller et al., 2016; Tesfaye & Breuer, 2024, 2025a; Tulbure et al., 2016). The Landsat satellite series serves as a critical earth observation resource, enabling diverse applications including tracking long-term hydrological variations, identifying environmental patterns, flood mapping, and drought assessment to facilitate evidence-based policy formulation (Wulder et al., 2019). Recently, the Copernicus program by the European Space Agency (ESA) has provided open-access, high-resolution Sentinel images as part of its earth observation initiative.

Machine learning is a key component of data-driven artificial intelligence (AI). It enables systems to learn patterns, relationships directly, thereby contributing to the decision-making processes of scientists and policymakers. The development of learning algorithms, vast datasets, and low-cost computation has allowed machine learning to progress in a wide range of applications (Jordan & Mitchell, 2015). Deep learning is an advanced approach in solving problems by discovering intricate structures within large amount of data (LeCun et al., 2015). The performance of machine learning and deep learning models enhances through training data iteratively. Machine learning employs multiple algorithmic approaches, such as decision tree classifiers and support vector machine (SVM), while deep learning specifically implements artificial neural network architectures with layered

structures. The performance of these models depends on the quality, quantity, and variety of the data on which they are trained. However, the reliance on data-driven machine learning and deep learning models presents challenges, including biases in training datasets that can lead to unfairness and adversely affect overall AI systems (Mehrabi et al., 2019). Methods such as cross-validation, regularization, and fairness-aware algorithms are vital to address such unfairness and ensure robust AI systems (Barocas et al., 2023).

Integrating machine learning and cloud-based remote sensing further improves the capabilities of surface water detection and monitoring and LULC classification. The cloud computing platform allows data from various sources, including climate and LULC to be integrated. This enables the development of robust models that can be used to analyze hydrological systems and their responses to the various factors that affect surface water availability (Gaines et al., 2022). In addition to its technical importance, cloud-based remote sensing promotes collaboration, reproducibility, scalability, and knowledge sharing within the scientific community. Open-access cloud computing platforms like GEE allow us to share knowledge and visualizations, and provide access to numerous datasets, ensuring transparency and consistency in various applications, including surface water assessment and monitoring (Gorelick et al., 2017). Thus, integrating various approaches and datasets with cloud computing supports sustainable water resource management and ensures mutual benefits, especially in regions with limited data, such as Ethiopia and its neighboring countries.

Reviews of the several studies on surface water detection and monitoring (e.g., Acharya et al., 2018; Mueller et al., 2016; Tulbure et al., 2016; Vanderhoof et al., 2023) show that RF and water index methods have mostly been applied using Landsat data over space and time. Despite advancement in existing studies, water indices performance varies by location, as does the selection of thresholds. Thus, this dissertation assesses the effectiveness of a number water indices at the large-scale, such as, the normalized difference water index (NDWI) (McFeeters, 1996), the modified normalized difference water index (MNDWI) (Xu, 2006), land surface water index (LSWI) (Xiao et al., 2002), automatic water extraction index with shadow (AWEIsh) and with no shadow (AWEInsh) (Feyisa et al., 2014a), Water Index (WI) (Fisher et al., 2016) and the sentinel water index (SWI) (Jiang et al., 2021). This assessment helps identifying a preferable approach to monitoring large-scale water resources in Ethiopia. The selected indices are examined alongside the Joint Research Center's (JRC) global surface water dataset, which offers extensive spatio-temporal

information on inland water bodies (Pekel et al., 2016). However, there are some data gaps in the JRC dataset, and it lacks specific spatio-temporal detection and analysis. Moreover, an inclusive evaluation of various machine learning modeling approaches (Gradient Tree Boost (GTB), SVM, and RF) across different geographic locations for long-term surface water detection and monitoring remains uncovered. Both shortcomings underscore the need to investigate multi-source and multiple approaches to surface water monitoring at national and site-specific locations.

Therefore, this dissertation aims to bridge the knowledge and skill gaps in surface water monitoring at a national scale and in site-specific locations of Ethiopia using multi-source satellite data from Sentinel and Landsat, as well as multiple approaches ranging from water indices to machine learning. It also provides a surface water information through a cloud computing system. The study also incorporates an assessment of multiple modeling approaches, including extreme gradient boosting (XGBoost), artificial neural networks (ANNs), random forest (RF), support vector regression (SVR), and multiple linear regression (MLR). These methods are employed to examine climate and LULC influences on surface water systems while conducting spatiotemporal evaluations of LULC changes. Insights from this dissertation will improve the understanding of water resource management and its relationship to climate and LULC in cloud-based remote sensing with machine learning approaches.

1.2 Objectives

The main aim of this dissertation is:

> *To assess the use of remote sensing and machine learning approaches to monitor surface water distribution and the climate and LULC drivers in Ethiopia.*

This can be subdivided into three objectives, which will be addressed in separate chapters:

1. To assess the performance of water indices and identify the optimum indices for detecting and monitoring water bodies with Sentinel data in a large-scale assessment (Chapter 2);
2. To analyze spatiotemporal surface water monitoring and its trends from 1986 to 2023 using remote sensing and machine learning (Chapter 3); and

3. To assess the impact of climate and LULC drivers on surface water resources using data-driven models, as well as to analyze the spatiotemporal LULC change (Chapter 4).

1.3 Materials and Methods

1.3.1 Description of the study area

Ethiopia has a diverse topography, ranging from the Afar Depression, which is 116 meters below sea level, to the Ras Dashen Mountains in the Simien Range, which are about 4,600 meters above sea level. This variability has led to the development of different agro-ecological zones, climates, and soil types throughout the country. LULC is characterized by its varied landscape, climatic variations, and socio-economic activities. These influences hydrological cycles and ecosystem dynamics. Ethiopia's agro-ecological zones are categorized into six groups: arid, semi-arid, sub-moist, moist, sub-humid, and humid. These zones are further delineated by Ethiopia’s traditional climate classification system, which identifies five zones based on elevation and temperature: "Wurich" (cold to moist), "Dega" (cool to humid), "Weynadega" (cool sub-humid), "Kola" (warm semi-arid), and "Berha" (hot arid). Ethiopia's climate exhibits significant variability, with mean annual temperatures ranging from 5°C in the highlands to 40°C in the lowlands (Gebrechorkos et al., 2023). Rainfall patterns are also characterized by pronounced spatial and temporal variability. Annual rainfall ranges from approximately 100 mm per year in the arid northeastern lowlands to over 2,500 mm per year in the humid southwestern highlands. The country experiences two main rainy seasons: the Belg from March to May, characterized by light rainfall, and the Kiremt from June to September, the main rainy season.

Ethiopia's water bodies are divided among 12 main river basins and numerous lakes (Figure 1.1). The Abay (Blue Nile), Baro-Akobo, Tekeze, and Mereb basins — which are sub-basins of the larger Nile Basin — together account for approximately 33% of the country's land area. These basins cover the northern, central, and western areas of Ethiopia (FAO, 2016). About 90% of the total water bodies in Ethiopia are found in the Abay, Tekeze, Baro-Akobo, and Omo-Gibe basins (Smola & Schölkopf, 2004). Conversely, there are scarce perennial water bodies in eastern Ethiopia, particularly below an elevation of 1,500 m above sea level. Additionally, most of the Aysha, Dinakle, and Ogaden basins are dried up and do not drain consistently. (Berhanu et al., 2014).

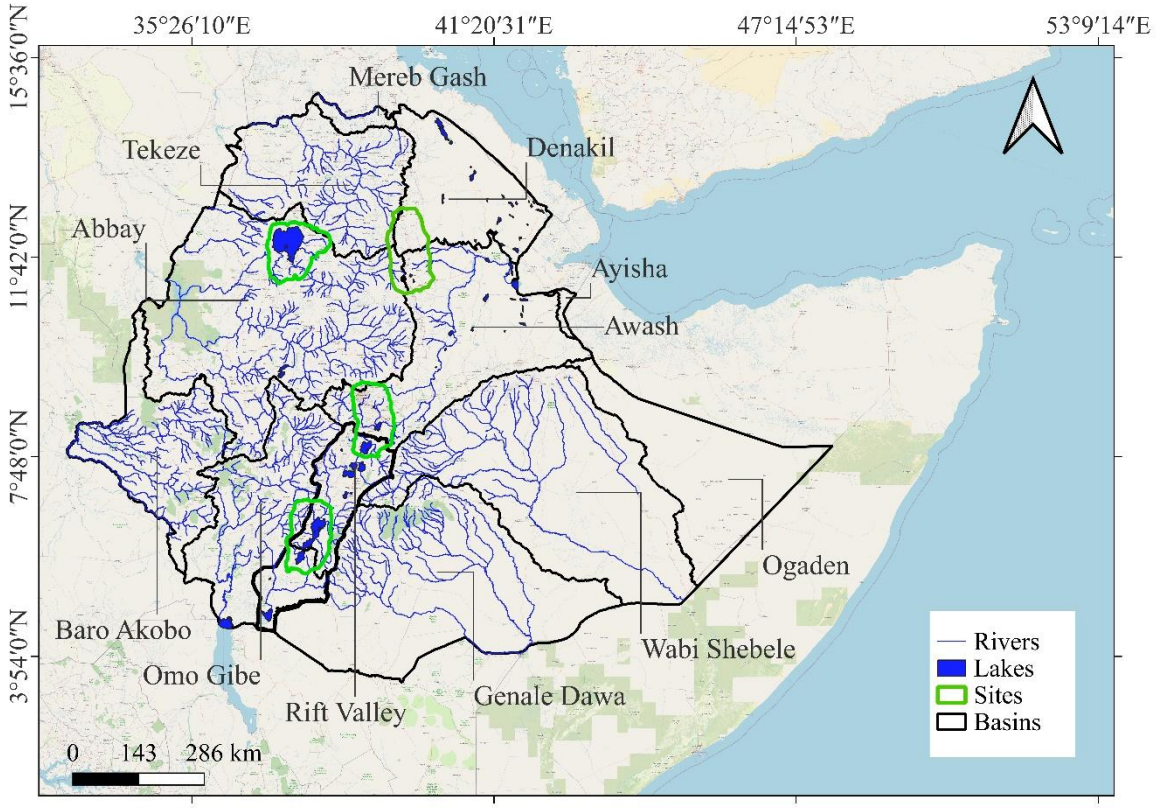


Figure 1.1 Study sites with the prominent inland water bodies of Ethiopia (Tesfaye & Breuer, 2025a)

1.3.2 Data acquisition and preprocessing

To detect surface water on a national scale using water indices, Sentinel-2 imagery is obtained with the Sentinel-2 Surface Reflectance product (COPERNICUS/S2_SR). The images, provided by the ESA Copernicus Earth Observation Program, offer multispectral data in the visible, near-infrared, and shortwave infrared spectral bands — a total of 13 bands. The images are filtered based on the specified time range and refined spatially by study area. Then, clouds are masked out using a custom cloud/shadow mask. Finally, the median value for each pixel across all filtered images is calculated using the median function. This results in a single composite image that effectively aggregates Sentinel-2 observations over the specified temporal and spatial extent.

To conduct long-term spatiotemporal surface water monitoring and LULC classification, I obtain Tier 1 Landsat 5 (from 1986 to 1998), Landsat 7 (from 1999 to 2012), Landsat 8 (from 2013 to 2021) and Landsat 9 (from 2022 to 2023) images from the United States Geological Survey (USGS) via GEE data access. I leverage a cloud computing system for

preprocessing satellite images because it can manage massive amounts of remotely sensed data and support long-term time series analysis (Gorelick et al., 2017). To address scan line anomalies in post-2003 Landsat 7 imagery, the gap-filling methodology outlined in the USGS L7 Phase-2 Protocol (Scaramuzza et al., 2004) is applied. Subsequently, annual composite images for each study site are generated by merging multiple Landsat scenes through median value reduction (Z. Zhang et al., 2021). To improve data quality, the CFmask algorithm (Zhu & Woodcock, 2012a) is applied for cloud and shadow masking. Climate and remote sensing data from the GEE system were used annually from 2000 to 2020 to examine the impact of climate and LULC on surface water dynamics. Precipitation and temperature (minimum and maximum) data were acquired from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS-v2) and the Climate Hazards Center Infrared Temperature with Stations (CHIRTS), respectively. The Terra Vegetation Index (MOD13Q1 V6.1) and the actual evapotranspiration (MOD16A2GF V6.1) were obtained from the National Aeronautics and Space Administration (NASA) via GEE.

1.3.3 Water indices and machine learning

Several water indices are calculated in the GEE computing system for large-scale surface water detection with the normalized difference water index (NDWI) (McFeeters, 1996), the modified normalized difference water index (MNDWI) (Xu, 2006), the land surface water index (LSWI) (Xiao et al., 2002), the automatic water extraction index with shadow (AWEIsh) and without shadow (AWEInsh) (Feyisa et al., 2014a), the water index (WI) (Fisher et al., 2016) and the Sentinel water index (SWI) (Jiang et al., 2021). However, the default threshold value does not provide the highest water detection accuracy (Feyisa et al., 2014). Thus, optimal thresholds are applied to distinguish between water and land by considering errors of commission and omission. The optimal threshold is selected through an iterative trial-and-error process (Acharya et al., 2018).

Machine learning approaches are leveraged for long-term surface water monitoring at four specific sites and to calculate the extent of surface water. Sample points are utilized to train machine learning algorithms to detect and monitor surface water by analyzing long-term Landsat image. From 1986 to 2023, training points are obtained with a stratified random sampling technique from water and non-water areas in satellite imagery. As in previous studies, the points are gathered manually via visual interpretation (Getaneh et al., 2024; Tulbure et al., 2016; Vanderhoof et al., 2023). The samples consider the type, size, color, and pattern of surface water (rivers, streams, ponds, reservoirs, lakes, and wetlands) and

non-water bodies, excluding these features. Water samples are taken from all sites in the Addis-Ziway, Hayq-Hashenge, Tana, and Abaya-Chamo areas, with 1,086, 1,124, 1,218, and 1,036 pixels, respectively. The number of non-water pixels from the same sites is 1,091, 1,228, 1,025, and 1,038, respectively. Therefore, each machine learning model is trained using 4,464 water pixels and 4,382 non-water pixels.

In the Google Colab platform, the GEE Python application programming interface (API) is used to retrieve, process, and analyze Landsat imagery. GTB, RF, and SVM are frequently employed and show effectiveness as machine learning models for assessing surface water distribution over space and time (Schmitt, 2020; Tulbure et al., 2016; Vanderhoof et al., 2023). Friedman, (2002) states that GTB is an ensemble learning model that uses gradient descent optimization to construct a predictive model by reducing the loss function. It improves a model iteratively using the addition of successive trees to rectify errors. This iterative process differentiates GTB from other ensemble learning algorithms, such as random forest (RF), where trees are constructed separately. RF is a nonparametric algorithm capable of handling nonlinear and large datasets. It builds fully grown trees in parallel using bootstrap sampling, and classification is done by averaging the outputs of all trees for out-of-bag observations (Breiman, 2001). SVM identifies one or more hyperplanes in a high-dimensional feature space that maximize the margin between different classes of data points (Cortes & Vapnik, 1995). The machine learning models are fine-tuned for optimal performance with a grid search and tenfold cross-validation.

Five models — XGBoost, ANN, SVR, RF, and MLR — are tested to assess the impact of climate and LULC on surface water changes. The GEE Python API is used to process, validate, and analyze the model impact evaluation in the Google Colab platform. Shapley additive explanation (SHAP) values provide a quantitative measure of the impact of each variable on the model, enabling us to determine the significance of the variable and its relationships to the model (Lundberg & Lee, 2017).

XGBoost is a scalable tree boosting system that utilizes important features and algorithmic optimizations (Chen & Guestrin, 2016). The justified weighted quantile sketch technique allows for the use of weights in estimated tree learning. Distributed and simultaneous processing allows for rapid learning, leading to faster model evaluation. An ANN represents a layered computational architecture of interconnected nodes that extract patterns from training data through iterative processing. Its standard structure comprises three components: (1) an input layer receiving raw data, (2) hidden layers performing weighted

transformations, and (3) an output layer generating predictions. The system employs forward/backward propagation algorithms with loss minimization to refine its parameters through multiple training cycles, enabling effective handling of complex datasets. SVR adapts SVM principles for continuous variable prediction, constructing optimal regression functions between input features and target outputs (Smola & Schölkopf, 2004). RF is an ensemble learning model that develops a collection of decision trees while training and provides the average regression prediction of the distinct trees. MLR is a common statistical approach to modeling the association between dependent and independent variables. Hyperparameter tuning is undertaken in machine learning with grid search and 5-fold cross-validation to optimize and improve model accuracy.

1.3.4 Spatio-temporal analysis

This study examines long-term changes in surface water extent by analyzing a 38-year Landsat dataset (1986–2023). The Mann-Kendall test, known for its resilience to outliers, detects monotonic trends in the time series (Mann, 1945; Yue et al., 2002). Statistical significance is determined using p-values, while Kendall's tau coefficient quantifies the association between temporal trends and variations in water extent (Ahamed & Bolten, 2017; Yue et al., 2002). To estimate the magnitude of change, Sen's slope method is applied (Sen, 1968), and the Pettitt test identifies abrupt shifts in the dataset (Pettitt, 1979). Water Occurrence Frequency (WOF) is computed for each pixel, indicating the proportion of observations classified as water (Pekel et al., 2016; Yang et al., 2020). Pixels with WOF exceeding 60% are categorized as permanent water bodies (Tariq & Qin, 2023), distinguishing them from ephemeral or seasonal water. All computations are executed in Python within the Google Colab platform.

To classify and analyze LULC dynamics from 2000 to 2020, training sample points are collected from satellite images and prior reference maps through visual interpretation. The LULC classes include water, cropland, grassland, shrubland, built-up, and forestland. A total of 1,864, 1,908, and 1,982 samples are collected for 2000, 2010 and 2020, respectively. Landsat imagery is processed through analytical procedures implemented via Python scripts that integrate the GEE API with Geemap's mapping capabilities running in the Colab environment. A gradient boosted model is chosen because it has consistently proven to be more effective in classifying images (Hastie et al., 2009). To improve classification

accuracy, a grid search and 5-fold cross-validation are employed to optimize the gradient boosting classifier.

1.3.5 Model validation and accuracy assessment

Quantitative and qualitative accuracy assessments are performed for the water index-based surface water detection. For the first objective, a total of 4,680 sample points are collected for the validation. Specifically, 2,340 samples are obtained from water bodies, and the remaining 2,340 samples cover non-water bodies with a stratified random sampling method. For the second objective, the machine learning-based surface water monitoring, a total of 4,464 water pixels and 4,382 non-water pixels are used to validate each model and the JRC dataset at the specific sites.

To assess LULC classification accuracy for the third objective, a total of 1,864, 1,908, and 1,982 samples are also collected for the years 2000, 2010, and 2020, respectively. Classification results for surface water detection and LULC are evaluated using user's accuracy, producer's accuracy, and overall accuracy. User's accuracy indicates the degree to which the classified map aligns with real-world conditions (Story & Congalton, 1986). Producer's accuracy assesses the probability that a reference sample is correctly classified (Congalton, 1991), and overall accuracy provides a comprehensive measure of classification effectiveness among all classes. These metrics are widely used in remote sensing image analysis (Hamunyela et al., 2022; Pickens et al., 2020; Tulbure et al., 2016) and are essential for validating classification results (Foody, 2002). Moreover, for the third objective, the RMSE, dr, MBE, and MPE are calculated to validate the XGBoost, ANN, SVR, RF, and MLR models for assessing the influence of climate and LULC on surface water dynamics. The RMSE is a widely used statistical metric for model assessment in climate research (Chai & Draxler, 2014). The Willmott index, on the other hand, evaluates multiple facets of model effectiveness (Willmott et al., 2012). The MBE and MPE represent a model's tendency to overestimate or underestimate values. The GEE Python API is used for all model validation and accuracy assessment.

1.4 Main results

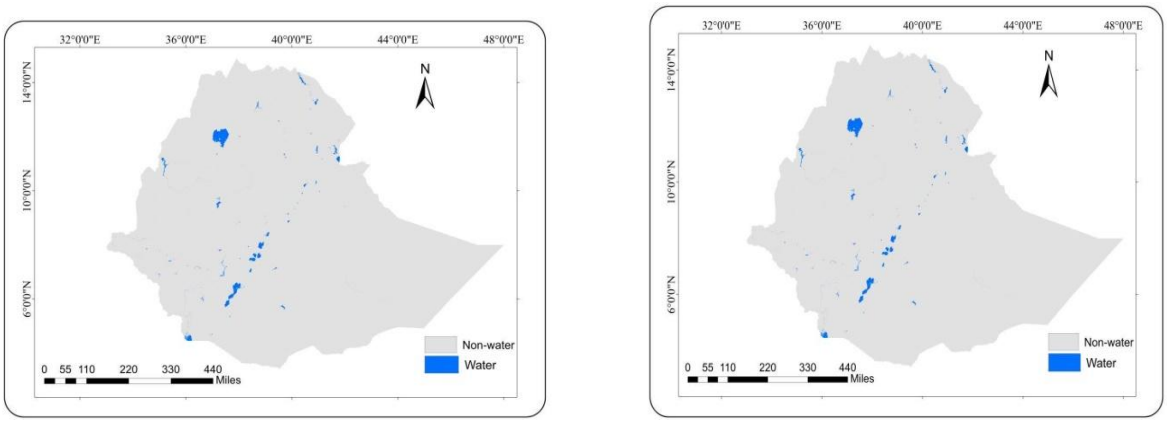
1.4.1 Work package 1: Performance of water indices

The optimal thresholds for distinguishing water pixels from non-water pixels are 0.06 for NDWI, 0.14 for MNDWI, 0 for AWEIsh, 0.04 for AWEInsh, 0.02 for WI, 0.03 for SWI, and 0.12 for LSWI. The WI and AWEIsh are the most optimal for visually assessing surface water detection in the large-scale water monitoring experiment conducted using Sentinel-2 imagery in the GEE computing platform (Figure 1.2). This is because they are better at detecting surface water in areas with vegetation and urban areas. SWI and NDWI perform worse because of misclassifications and unclassified water pixels, respectively. AWEInsh is less effective at identifying rivers and small water bodies of water, and LSWI performs even worse at accurately detecting water bodies.

Satisfactory classification accuracy of over 85% has been achieved by all water indices except the LSWI in terms of producer, user, and overall accuracies (Table 1.1). Water producer's accuracy and non-water producer's accuracy range from 32% to 96% and 84 to 97%, respectively. WI achieves the highest accuracy with an overall accuracy of 96%. LSWI obtains the lowest accuracy, with an overall accuracy of 78%. This study shows that the LSWI performs worse in terms of user, producer, and overall accuracies. Overall, WI and AWEIsh are the most accurate water indices in terms of accuracy indicators and visual inspection in the large-scale assessment of Ethiopia, and the surface water extent is estimated at 82,650, and 86,530 km^2, respectively.

Table 1.1 Heat map of accuracy assessment of water indices (Tesfaye & Breuer, 2024)

	Producer accuracy Water (%)	Error of omission	Producer accuracy Non-water (%)	Error of omission	User accuracy Water (%)	Erro of commission	User accuracy Non-water (%)	Error of commission	Overall accuracy (%)
NDWI	85	15	92	8	90	10	92	8	90
MNDWI	92	8	93	7	93	7	93	7	92
AWEInsh	94	6	95	5	93	7	95	5	94
AWEIsh	95	5	96	4	94	6	97	3	95
WI	96	4	97	3	96	4	96	4	96
SWI	91	9	94	6	93	7	94	6	93
LSWI	32	68	84	16	36	66	81	19	78



WI AWEIsh

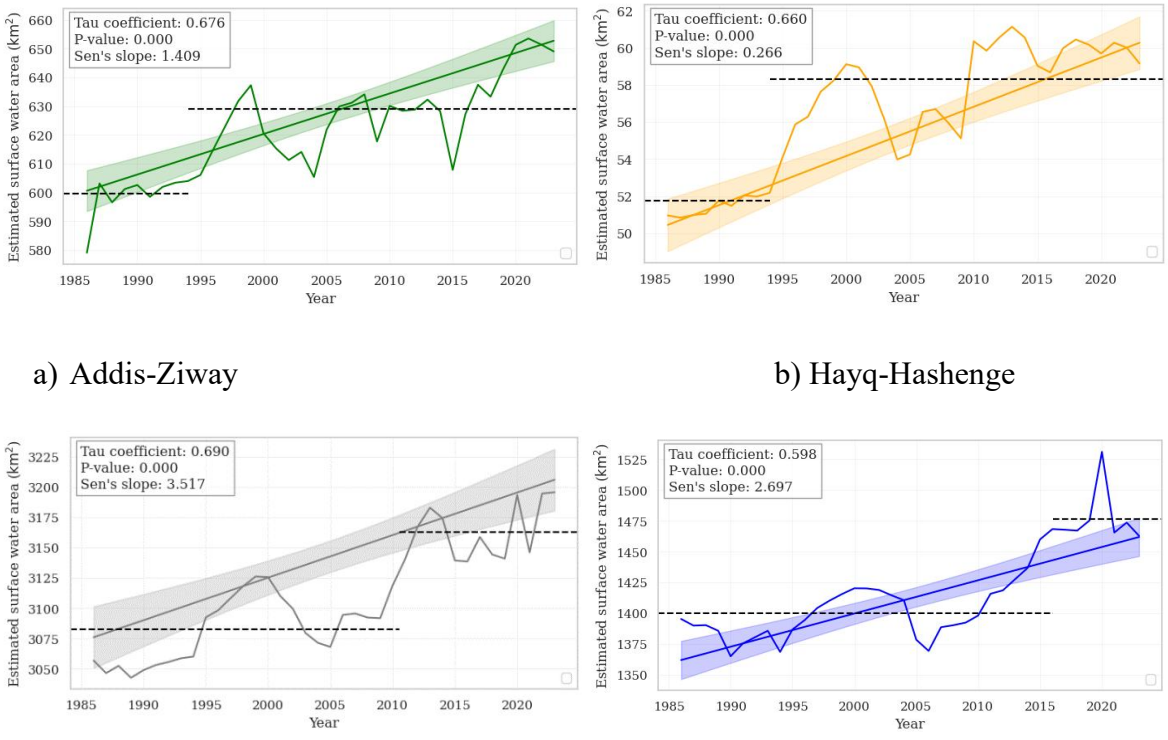
Figure1.2 Ethiopia surface water distribution with most effective water indices (Tesfaye & Breuer, 2024)

1.4.2 Work package 2: Multi-decadal surface water monitoring

Machine learning techniques, including GTB, RF, and SVM, along with the JRC dataset, achieve high-performance surface water detection in remote sensing, with classification accuracies exceeding 90% across producer's, user's, and overall accuracy metrics. The overall accuracy of RF and SVM ranges between 92% and 97%, whereas GTB delivers even higher precision, with results between 94% and 98% from 1986 to 2023. Consequently, the GTB model is selected for surface water monitoring at four sites from 1986 to 2023 due to its superior performance. Across all sites, producer's accuracy, user's accuracy, and overall accuracy range from 91% to 98%, indicating a consistently high level of accuracy. The highest overall agreement between the reference data and the model results is observed at the Addis-Ziway site, and the lowest is observed at the Abaya-Chamo site.

The estimated surface water area and the JRC dataset show strong agreement, as demonstrated by R^2 values surpassing 0.9 and RMSPE values lower than 1% at all sites. Furthermore, the range of the RMSE and MAE values is between 0.54-9.11, and 0.42-8.09, respectively, thereby indicating that the absolute deviations are minimal. Despite significant year-to-year fluctuations, a substantial increase in surface water extent is evident across all sites over nearly four decades (Figure 1.3). A notable increase in surface water is observed at Addis-Ziway, with a robust Kendall tau value of 0.676 ($p = 0.001$) and an annual expansion rate of 1.409 $km^2\ yr^{-1}$ (Sen's slope). Since 2000, the average water area has increased by 5% compared to pre-2000 levels. Likewise, the Hayq-Hashenge site depicts a

gradual increment in surface water area at a rate of 0.266 km² km² yr^{-1}. Both sites show a noticeable trend shift around 1995. A marked shift in trend is evident at both sites beginning around 1995. Hayq-Hashenge notably experienced a 13% increase in surface water area within six years, reaching approximately 58.3 km². The Tana site's average surface area increased by about 2.6% after 2010. In contrast, the Abaya-Chamo site showed a more significant expansion of 5.4% after 2015, indicating a more substantial rise at the latter location. Nevertheless, changes before the early 2000s were less noticeable because of their relatively large size compared to the significant shifts observed after 2010 at the Tana site and after 2016 at the Abaya-Chamo site. Overall, these findings highlight clear spatiotemporal heterogeneity in surface water dynamics, characterized by an upward trend across the study areas.



a) Addis-Ziway b) Hayq-Hashenge

c) Tana d) Abaya-Chamo

Figure 1.3 Long-term trends of surface water dynamics with Tau coefficient, Sen's slope, and p-values. The dashed line indicates the mean surface water coverage pre- and post-change year. (Tesfaye & Breuer, 2025a)

To evaluate the long-term occurrence of water, WOF is calculated over 38 years and classified into five categories ranging from 0% (no water) to 100% (permanent water).

Across all study sites, most surface water falls within the 61–100% range (Table 1.2). Lake Tana exhibits the greatest stability, with 94% of its area falling within this range. It is followed by Abaya-Chamo, with 91.6%. In contrast, smaller surface waters (Addis-Ziway and Hayq-Hashenge) show less stability, with approximately 84% of their areas falling within the highest WOF class. These sites also exhibit a much higher proportion of dynamic pixels (WOF 1–40%), reflecting frequent wet-dry transitions. Lower WOF values are mainly distributed along the periphery of more stable water bodies and indicate the seasonal or temporary presence of water, such as on floodplains. Overall, high WOF values represent permanent water bodies, while lower values reflect seasonal variability.

Table 1.2 Surface water coverage in total and percentage by water occurrence frequency (WOF) category between 1986 and 2023 (Tesfaye & Breuer, 2025a)

	Surface water area (km^2) per WOF				
	Area [km^2]	WOF1-20%	WOF21-40%	WOF41-60%	WOF61-100%
Hayk-Hashenge	63.8	5.0	3.1	2.1	53.6
Addis-Zeway	659.2	62.4	19.7	25.9	551.2
Abaya-Chamo	1525.0	83.4	23.7	21.6	1396.3
Tana	3218.4	125.7	46.9	30.6	3015.3
	Percent of surface water area (%) per WOF class				
	Area [km^2]	WOF1-20%	WOF21-40%	WOF41-60%	WOF61-100%
Hayk-Hashenge	63.8	7.8	4.9	3.3	84.0
Addis-Zeway	659.2	9.5	3.0	3.9	83.6
Abaya-Chamo	1525.0	5.5	1.6	1.4	91.6
Tana	3218.4	3.9	1.5	1.0	93.7

1.4.3 **Work package 3: Effects of climate and LULC drivers on surface water dynamics**

To quantify climate and LULC effects on water surface dynamics, five modeling approaches - ANN, XGBoost, SVR, RF, and MLR are evaluated through statistical measures of RMSE, dr, MBE, and MPE. XGBoost is the optimal model, demonstrating the smallest RMSE (0.7) and the greatest dr (0.9), and its MBE (0.1) and MPE (0.3%) are close to zero, indicating minimal bias and high accuracy. The XGBoost model incorporates

rainfall data, temperature (minimum and maximum), vegetation indices, and actual evapotranspiration rates as explanatory variables to analyze their impact on changes in bodies of water between 2000 and 2020. The SHAP summary plot reveal that precipitation is the most influential driver, followed by minimum temperature and vegetation (Figure 1.4). While maximum temperature and actual evapotranspiration show relatively less influence on surface water dynamics in our study area.

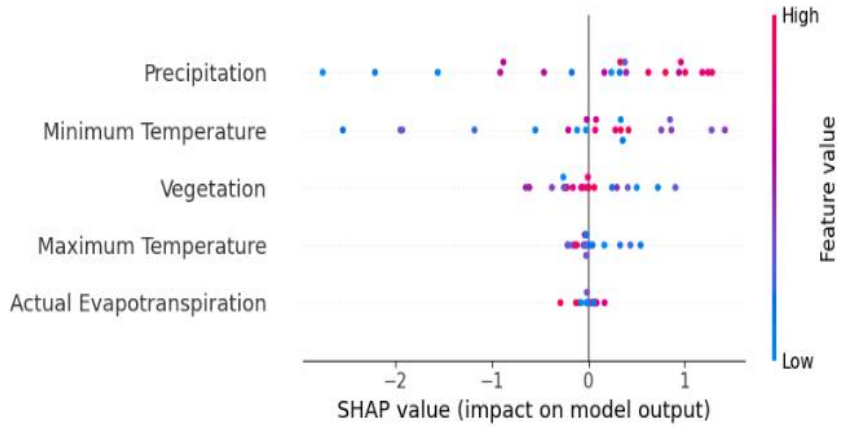


Figure 1.4 XGBoost model with Shapely Additive exPlanation (SHAP) values shows the associations between the predictor variables (Tesfaye & Breuer, 2025b)

The remote sensing-based LULC classification shows satisfactory classification accuracy of over 80% in terms of producer, user, and overall accuracies. This provides reliable evidence for the analysis of LULC dynamics. During the 2000-2020 period, substantial variations in lLULC transformations are observed across space and time. Although shrubland maintained its position as the predominant cover type during the entire timeframe, it shows a significant decline, from 38.3% in 2000 to 30.7% in 2020 (Table 1.3). Likewise, grassland consistently declined from 35.9% in 2000 to 27.1% in 2020. In contrast, built-up areas and cropland exhibit consistent expansion, rising from 6.0% to 13.1% and 7.3% to 16.1%, respectively. Forestland also depicts a significant increase from 12.2% to 12.6%, while water bodies show a relatively slight increment from 55.0 km² (0.40%) to 59.7 km² (0.43%) over the entire years.

Table 1.3 Decadal LULC proportion and dynamics during 2000 and 2020 (Tesfaye & Breuer, 2025b)

LULC type	2000		2010		2020		2000-2020		
	km^2	%	km^2	%	km^2	%	km^2	%	% yr^{-1}
Water	55.0	0.40	58.7	0.42	59.7	0.43	4.7	8.5	0.4
Built-up	830.6	6.0	1413.0	10.2	1821.7	13.1	991.1	119.3	6.0
Forestland	1694.8	12.2	1873.3	13.5	1748.5	12.6	53.7	3.2	0.2
Shrubland	5310.2	38.3	4756.3	34.3	4258.7	30.7	-1051.5	-19.8	-1.0
Cropland	1007.0	7.3	1309.2	9.4	2229.0	16.1	1222.0	121.4	6.1
Grassland	4972.5	35.9	4459.5	32.2	3752.5	27.1	-1220.0	-24.5	-1.2

From 2000 to 2020, shrubland and grassland shrink annually by 1.0% and 1.2%, corresponding to total reductions of 19.8% and 24.5%, respectively. Conversely, built-up areas, forests, cropland, and water bodies expand at annual rates of 6.0%, 0.2%, 6.1%, and 0.4%, for total increases of 119.3%, 3.2%, 121.3%, and 8.5%, respectively. Figure 1.5 shows the transformation of surface water and its conversion into and out of other LULC categories. From 2000 to 2010, 0.76 km^2 of cropland converted into water, while 0.04 km^2 of water converted into shrubland. Between 2010 and 2020, 1.04 km^2 of grassland converted into water, whereas 0.44 km^2 of water converted into cropland. Despite being minimal, the conversion of water into other LULC and vice versa is still relevant for expanding surface water coverage.

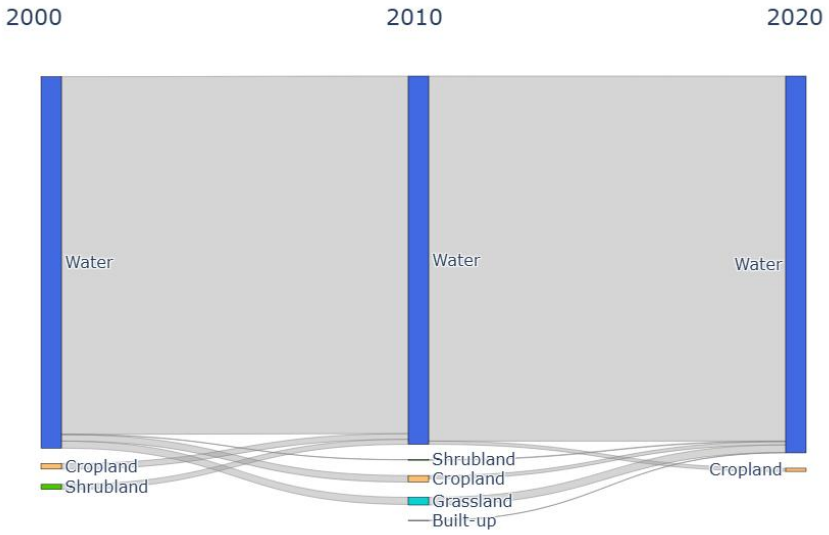


Figure 1.5 Water-related LULC conversion from 2000 to 2020 (Tesfaye & Breuer, 2025b)

2 Performance of water indices for large-scale water resources monitoring using Sentinel-2 data in Ethiopia

This chapter is published in *Environmental Monitoring and Assessment* as:

Tesfaye, M., & Breuer, L. (2024). Performance of water indices for large-scale water resources monitoring using Sentinel-2 data in Ethiopia. *Environmental Monitoring and Assessment*, *196*(5), 467. https://doi.org/10.1007/s10661-024-12630-1.

Abstract

Evaluating the performance of water indices and water-related ecosystems is crucial for Ethiopia. This is due to limited information on the availability and distribution of water resources at the country scale, despite its critical role in sustainable water management, biodiversity conservation, and ecosystem resilience. The objective of this study is to evaluate the performance of seven water indices and select the best-performing indices for detecting surface water at country scale. Sentinel-2 data from December 1, 2021, to November 30, 2022, were used for the evaluation and processed using the Google Earth Engine. The indices were evaluated using qualitative visual inspection and quantitative accuracy indicators of overall accuracy, producer's accuracy, and user's accuracy. Results showed that the water index (WI) and automatic water extraction index with shadow (AWEIsh) were the most accurate ones to extract surface water. For the latter, WI and AWEIsh obtained an overall accuracy of 96% and 95%, respectively. Both indices had approximately the same spatial coverage of surface water with 82,650 km^2 (WI) and 86,530 km^2 (AWEIsh) for the whole of Ethiopia. The results provide a valuable insight into the extent of surface water bodies, which is essential for water resource planners and decision makers. Such data can also play a role in monitoring the country's reservoirs, which are important for the country's energy and economic development. These results suggest that by applying the best-performing indices, better monitoring and management of water resources would be possible to achieve the Sustain able Development Goal 6 at the regional level.

Keywords Copernicus data · Google Earth Engine · Remote sensing · SDG 6 · Surface water

2.1 Introduction

Land surface water is one of the most decisive natural resources substantially changing in spatiotemporal terms, particularly because of land use and land cover dynamics as well as climate change. Land surface water encompasses streams, rivers, ponds, reservoirs, lakes, wetlands, and other inland water bodies (Pekel et al., 2016; Vandas et al., 2002). In spite of its limited extent, particularly in semi-arid and arid regions of the globe, land surface water is a resource for numerous human uses (e.g., drinking water supply, sanitation, and hygiene), ensures irrigation agriculture, or is used for hydropower production and industrial use. In addition, it maintains and supports biodiversity and provides essential and diverse ecosystem services (Brauman, 2015; Dudgeon et al., 2006; Zedler & Kercher, 2005). Surface water resources also play a vital role in the climate system and the hydrological cycle (Chahine, 1992; Tranvik et al., 2009). However, water-related ecosystems are fragile and vulnerable to anthropogenic impacts and climate change (Vörösmarty et al., 2000). The biodiversity of water-related ecosystems continues to deteriorate at an alarming rate (Collen et al., 2014). Associated with climate change are hydrological extremes such as flooding or droughts, and emerging water-related diseases, both leading to increasing losses of lives. Therefore, timely monitoring of regional-scale surface water resources is critical for policy and decision-making processes for its sustainable use and management (Giardino et al., 2010; Morss et al., 2005).

Global initiatives and policy frameworks like the SDGs and the Aichi Biodiversity Targets under the Convention on Biological Diversity (CBD) aim to ensure sustainable development of water resources, to reduce human impact, and prevent the loss of biodiversity (CBD, 2010; Griggs, 2013). Especially, the SDG 6 “Clean water and sanitation” and its target 6.6 “Protect and restore water-related ecosystems” emphasize the need to quantify its indicator 6.6.1 “Change in the extent of water-related ecosystems over time” (Dickens et al., 2017). Surface water monitoring is crucial for sustainable development, biodiversity conservation, ecosystem resilience, and people’s livelihoods. For this reason, improved monitoring of surface water resources on a large scale is needed, for example, by utilizing the latest advances of the Google Earth Engine with Sentinel-2 data for effective water management and decision-making processes. The implementation of water indices in this cloud computing system for improved water monitoring and management is in line with the broader goal of achieving SDG 6, thus contributing to both policy interventions and sustainable development efforts.

Thus, having the aforementioned potential conflicts in mind, this study contributed to quantifying the spatial distribution of the extent of water-related ecosystems, thereby ensuring its sustainable management in Ethiopia. Surface water plays a pivotal role, constituting more than 50% of the world's water utilized for agricultural, domestic, and industrial purposes (OECD, 2015). This substantial reliance is largely attributed to lakes and reservoirs. Monitoring reservoirs is particularly important for providing early warning of drought, water, and food insecurity (Donchyts et al., 2022). Furthermore, effective reservoir monitoring serves as a deterrent to transboundary water conflicts and promotes international collaboration and also providing timely information by analyzing surface water dynamics as demonstrated for Turkey (Donchyts et al., 2022). The insights provided by surface water monitoring is needed for informed decision-making in environmental, agricultural, and urban water use, as for example shown for semi-arid Australia where it plays a crucial role in shaping water policy and management strategies (Tulbure et al., 2016). Overall, the integrated monitoring of surface water resources emerges as a support in addressing water challenges on a global scale, contributing not only to resource sustainability but also fostering international cooperation and informed governance.

Ethiopia has significant surface water resources with 12 major river basins, one lake basin and three dry basins (Figure 1). A strong rainfall gradient separates the central and western highlands with abundant annual rainfall of up to 1200 mm from the arid south east, east, and northeast, which receive 200 mm and below (Berhanu et al., 2014). Therefore, surface water resources are comparatively less available in the eastern part of Ethiopia (especially in the Awash basin) while the western part with the Abay (Blue Nile) river basin has large water resources. In contrast to the importance of water resources for human well-being and the country's ecosystems, monitoring systems for the availability of surface water and hydrological flows are in poor condition and sometimes unreliable or malfunctioning (Dile et al., 2018). A country-wide continuous assessment of a water resource indicator as requested as part of SDG6 is not in place.

In contrast to in situ measurements, satellite remote sensing is an effective and efficient tool for monitoring and mapping surface water distributions. It allows covering a wide range of spatial and temporal scales because of its accessibility, repeatability, geospatial consistency, and global coverage (Fisher et al., 2016; Mueller et al., 2016). Satellite remote sensing plays a vital role in monitoring inland sur face waters, particularly lakes and reservoirs. In the recent past, accurate surface water detection methods have been developed for a wide range

of environmental conditions (Feyisa et al., 2014b; Fisher et al., 2016), with particular benefit for methods and data fusion techniques using machine learning methods (Schmitt, 2020). Recent advances in remote sensing technology, characterized by high spatial, temporal, and spectral resolution, have contributed substantially to the monitoring of inland water bodies, with emphasis on their accurate detection (McCarthy et al., 2017), even though monitoring lakes and reservoirs is still a challenge for classification algorithms due to their characteristics. This challenge is for example shown for Lake Mead (USA), Nova Ponte reservoirs (Brazil), and Lake Williston (Canada) because of its temporal variability, various narrow ends, dynamic shape, and missing values by cloud or ice cover (Khandelwal et al., 2017). Therefore, this study tested surface water detection methods for the study area of Ethiopia using high spatial and temporal resolution in a cloud computing platform.

Google Earth Engine is such a cloud computing platform for global-scale analyses. It provides access to a range of satellite imagery, particularly Landsat and Sentinel-2 products and facilitates high-performance computing for social and environmental analysis, including water monitoring (Gorelick et al., 2017). The launch of the Sentinel program by the European Space Agency (ESA) as part of the Copernicus program was a breakthrough moment as it provided access to free, high-resolution images on a large scale for the first time. The band characteristics and spectral response methods of Sentinel-2 data are helpful to detect surface water bodies from the background (Jiang et al., 2021). Thus, in this study, Google Earth Engine was employed to process the potential of water index methods and selected the best-performing indices for surface water detection using multi-spectral satellite imageries of Sentinel-2 data.

Surface water can be detected using multi-spectral satellite imageries on the base of the significantly lower infrared reflectance of water compared to other land cover types. Hence, based on the peculiarities of the Near-Infrared/Short-Wave Infrared (NIR/SWIR) domain, numerous approaches have been developed for extracting surface water from remote sensing imageries.

The water indices method is a common classification method using multi-bands (Fisher et al., 2016). It is easy to use and quick to calculate (Ryu et al., 2002). Further indices include the normalized difference water index (NDWI) (McFeeters, 1996), the modified normalized difference water index (MNDWI) (Xu, 2006), and the land surface water index (LSWI) (Xiao et al., 2002). All these indices have been widely used in Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper (ETM+) imagery analyses. They are easy

to compute and rely only on two input bands. Feyisa et al. (2014) developed the automatic water extraction with no shadow (AWEInsh) using four bands (green, NIR, SWIR1, and SWIR2) (Feyisa et al., 2014b). The AWEIsh additionally utilizes the blue band. AWEIsh is designed for shadows from mountains, buildings, and clouds. The sentinel water index (SWI) is computed using red-edge1 and SWIR1 bands of Sentinel-2 (Jiang et al., 2021). The water indices methods have the benefits of robustness, quick detection, and high accuracy in large scale surface water detection (Feyisa et al., 2014b; Li et al., 2013; X. Wang et al., 2020). Each water index was designed and previously tested using Landsat imagery, except for the SWI. Therefore, this study evaluated the performance of each water index method using Sentinel-2 data in Ethiopia.

A few studies have been carried out to monitor surface water resources in Ethiopia using water indices with Landsat imagery. For example, NDWI was used for surface water extraction and change detection in the Central Rift Valley region of Ethiopia, specifically Abjata, Shala, and Langano Lakes (Sisay, 2016). Another study demonstrated the use of NDWI to monitor surface water in Chelekleka, Crafty, Coke, and other lakes (Sathianarayanan, 2018). An evaluation of the performance of three water indices was carried out at Lake Zeway, which confirmed that the AWEI performed better than the MNDWI and NDWI in detecting surface water bodies (Asfaw et al., 2020). However, Ethiopia faces challenges related to limited spatio-temporal information on water resources at the national level. This lack of information is particularly problematic given its importance for effective water management and decision-making. It is also a key challenge for achieving the SDG 6 indicators. Ethiopia's diverse landscapes require the selection of appropriate water detection and monitoring techniques tailored to local conditions. Such an analysis would also increase the knowledge of remote sensing applications for water resources in different geographical settings. The performance of a wide range of water index methods has not been tested on the country level of Ethiopia, and a comparative performance analysis for the more recent Sentinel-2 data is missing.

Choosing the best index for large-scale surface water detection is difficult due to inconsistent results obtained from various indices and unstable threshold values to differentiate water from non-water, which is changing with location and scene (Ji et al., 2009). Therefore, this study sets out to identify optimal thresholds for large-scale assessment. In Ethiopia, surface water detection is challenging at a large scale due to the labeled water and non-water feature datasets being limited. Hence, the objective of this

work was to demonstrate the potential of water index methods and to select the best-performing indices for detecting surface water using high-resolution and multi-temporal Sentinel-2 data at the country scale. The indices were calculated using the Google Earth Engine. The spatial distribution of water resources in Ethiopia and monitor surface water resources in relation to the fulfillment of SDG 6 are investigated.

2.2 Materials and methods

2.2.1 Description of the study area

Ethiopia's diverse topography ranges from the depths of the northeastern Afar Depression at 116 m below sea level to the towering heights of the Ras Dashin Mountains at about 4600 m above sea level. This has created a diverse of agro-ecological zones, climates, and soil compositions. The agro-ecological zones are categorized into six major groups (MoA, 2000), which include arid, semi-arid, sub-moist, moist, sub humid, and humid zones. Ethiopia's climate is traditionally divided into five distinct zones, based on altitude and temperature: "Wurich" (cold to moist), "Dega" (cool to humid), "Weynadega" (cool sub humid), "Kola" (warm semi-arid), and "Berha" (hot arid). Average temperature variations are significant, ranging from 5 °C in the highlands to about 40 °C in the lowlands (Gebrechorkos et al., 2023). Rainfall patterns show considerable spatial and temporal variation. It ranges from 100 mm/yr in the arid northeastern lowlands to 2500 mm/year in the abundant southwestern highlands. The country experiences two main rainy seasons: the "Belg" from March to May, which is characterized by light rainfall, and the "Kiremt" from June to September, which is the main rainy season.

Geologically, Ethiopia's landscape is shaped by three main structural units, i.e., Precambrian basement, Paleozoic and Mesozoic sediments, and Cenozoic volcanites and sediments (Hurni et al., 2007). These geological formations contribute to the diverse topography and soil characteristics that are observed throughout the country. Soil composition adds another layer of diversity to Ethiopia's agricultural landscape. More than half of the country's arable land is covered by nitosols (23%), cambisols (19%), and vertisols (18%) (Dubale, 2001). Nitosols and cambisols dominate the highlands west of the Rift Valley and the Afar region, whereas xerosols and yermosols dominate the northern and southern regions, respectively. Agriculture is the backbone of the Ethiopian economy, with crop production being the main economic activity. It contributes about 40% of the country's gross domestic product (Jimma et al., 2024).

Ethiopia is endowed with substantial surface water resources and 12 major river basins. (Figure 2.1). A majority of Ethiopia's rivers experience seasonal variation and approximately 70% of the total runoff occurs between June and September (FAO, 2016). Four of Ethiopia's river basins, namely Abbay or Blue Nile, Baro-Akobo, Tekeze, and Mereb which are parts of Nile basin, cover 33% of the country and drain the northern, central, and western parts (FAO, 2016). In the eastern part of Ethiopia, surface water resources are limited since almost no perennial rivers are found below 1500 m a.s.l. Three of the main basins (Aysha, Dinakle, and Ogaden) are mainly dry with no permanent discharge (Berhanu et al., 2014). Ethiopia's 12 major lakes cover around 7300 km^2. Lake Tana is the largest lake in the Abay Basin. It is the main water source of the Abay River. Most other lakes are saline and located in the Rift Valley.

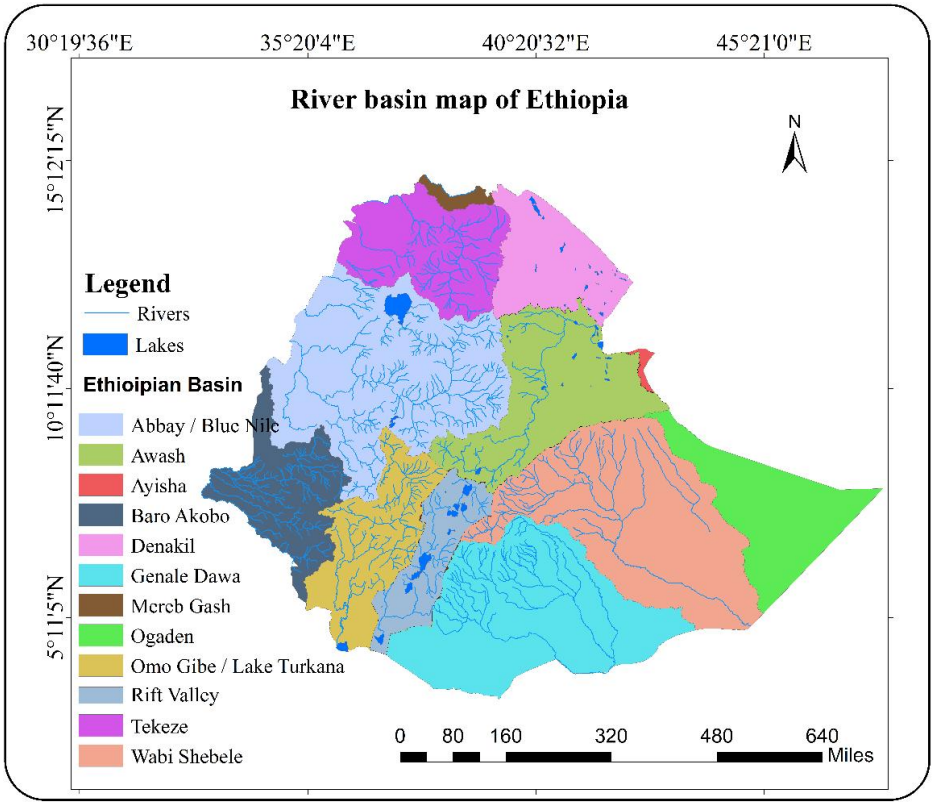


Figure 2.1 Main water bodies with basins of Ethiopia

2.2.2 Sentinel-2 data and pre-processing

Sentinel-2 data from December 1, 2021, to November 30, 2022, were used. Images were acquired by the Copernicus program in the Earth Observation program of the European Union, which provides multi-spectral data in the visible, near-infrared, and shortwave

infrared parts of the spectrum, a total of 13 bands. The tiles cover 100 by 100 km^2 with a spatial resolution of 10 m (bands blue (B2), green (B3), Red (B4), and NIR (B8), 20 m (red edge 1 (B5), red edge 2 (B6), red edge 3 (B7), red edge 4 (B8a), SWIR 1 (B11), and SWIR 2 (B12) and 60 m (aerosols (B1), water vapor (B9), and cirrus (B10)) for one of the most widely available Level-1C standard product. The temporal resolution or the revisit frequency of each individual Sentinel-2 satellite is 10 days and the combined constellation revisit is 5 days.

The Sentinel-2 imagery was collected using the Sentinel-2 Surface Reflectance product COPERNI CUS/S2_SR. However, Sentinel-2 imagery covers all of Ethiopia only in December, March, and January among all months. The images were then filtered based on the specified time range using the filterDate function. The dataset was further refined spatially by filtering based on a designated region of interest using filterBounds (roi). Subsequently, images with a cloudy pixel percentage exceeding 20% are excluded using ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 20). Clouds were then masked out using a custom cloud masking function through the cloud/shadow mask function in Google Earth Engine. Processing was performed in all spectral bands and the cloud mask detected cloud-free and cloudy pixels, including both dense clouds and cirrus clouds. Thus, the processing steps combining cloud and cirrus masking and atmospheric correction function in the Google Earth Engine contributed to cloud-free pixels and atmospheric correction of Sentinel-2 imagery. Lastly, the median value for each pixel across all filtered images was calculated using the median() function, resulting in a single composite image that effectively aggregates Sentinel-2 observations over the specified temporal and spatial extent.

2.2.3 Surface water detection

Seven water indices (Table 2.1) were calculated in the Google Earth Engine computing system. All indices have originally been developed using Landsat images except for the SWI. Various mechanisms have been used to obtain optimal thresholds for distinguishing water from land. We applied the optimum threshold value for the automated water extraction index by considering the error of commission and omission. The threshold ranges between − 1 and + 1, and a default threshold of zero does not represent the highest water detection accuracy (Feyisa et al., 2014). The selection of the optimum threshold involves an iterative trial and error process (Acharya et al., 2018). At the beginning, an initial threshold of default value zero for the water index was set and applied to the dataset. The classification accuracy was then assessed by incrementally adjusting the threshold until the

highest overall accuracy values were achieved. This approach ensured optimal differentiation between water and land.

Table 2. 1 Water index methods.

Index	Index name	Source	Equation
NDWI	Normalized Difference Water Index	McFeeters (1996)	$\frac{(Green - NIR)}{(Green + NIR)}$
MNDWI	Modified Normalized Difference Water Index	Xu, (2006)	$\frac{(Green - SWIR1)}{(Green + SWIR1)}$
AWEInsh	Automated Water Extraction Index non shadow	Feyisa et al., (2014)	(4*(Green-SWIR1))-(0.25*NIR+(2.75*SWIR2))
AWEIsh	Automated Water Extraction Index shadow	Feyisa et al., (2014)	(Blue+2.5*Green-1.5*(NIR+SWIR1)-0.25*SWIR2)
WI	Water Index	Fisher et al., (2016)	(1.7204+171*Green+3*Red-70*NIR-45*SWIR1-71*SWIR2)
SWI	Sentinel Water Index	(Jiang et al., 2021)	$\frac{(RedEdge1 - SWIR1)}{(RedEdge1 + SWIR1)}$
LSWI	Land Surface Water Index	Xiao et al., (2002)	$\frac{(NIR - SWIR1)}{(NIR + SWIR1)}$

2.2.4 Evaluation of water indices

Accuracy assessments were carried out to determine whether or not the results of surface water extraction are acceptable. In this study, qualitative (visual) and quantitative assessments were undertaken. In the visual assessment, the magnitude of continuousness and the smoothness of the boundary of water bodies were assessed and cross-checked with reference data. During classification, 80% of the samples were used for calibration and training, while the remaining 20% were kept for validation (Elith et al., 2011). For validation, sample points were collected from different locations and considered seasons from December 1, 2021, to November 30, 2022, using georeferenced samples from Sentinel-2 and coincident high-resolution imagery and OpenStreetMap data in Google Earth Engine.

For a coverage of < 4000 km^2 per land cover and less than 12 classes of land cover, Lillesand et al. (2008) recommend at least 50 samples for each map class. However, sample point sizes vary in the literature. For example, to assess accuracy by visual interpretation using Sentinel-2 data, Xia et al. (2019) selected 1000 sample points for water and non-water in the Huai River Basin and Wang et al. (2020) collected 1500 points each in the Hetao Plain. For surface water extent estimation in France using 2 years of Sentinel-2 data, a total of 4800 sample points were evaluated by visual interpretation in the Google Earth Engine platform (Yang et al., 2020). Hence, in this study, we collected a total of 4680 sample

points for the validation as a whole. Specifically, 2340 samples were taken from water bodies. The remaining 2340 samples included non-water bodies from a wide range of land cover types such as agricultural land, forest, grassland, built-up areas, and bare land. These sample points were collected throughout the entire study area using a stratified random sampling approach. Water bodies were additionally stratified by size and type to ensure that the assessment was representative of the diversity of water bodies present in the study area. Spatial stratification also improves the assessment of the classification accuracy of remote sensing data (Dong et al., 2022). Spatial autocorrelation examines how pixels that are close together are more similar than those that are far apart (Karasiak et al., 2022), resulting in falsely high precision metrics (Meyer et al., 2019; Roberts et al., 2017). Thus, features were separated by a minimum distance of 500 m to reduce spatial autocorrelation (Cabra-Rivas et al., 2016). This minimum distance for separating features ensures that they are spatially independent, reducing the risk of biased accuracy results and providing a more objective assessment of algorithm's performance. The quantitative assessment was carried out using the 4680 feature sample points. Indicators of evaluation include producer accuracy, user accuracy, and overall accuracy.

Confusion matrices, the common method of describing the accuracy of the classification (Lillesand et al., 2008), were used to compare the reference data and the corresponding classification outputs on a category-by-category basis. User accuracy is computed by dividing the number of correctly classified pixels in each category by the total number of pixels that are classified in that category (the row total), which is known as the specificity or true negative rate, and the complement of the commission error (Lillesand et al., 2008). Whereas the producer accuracy is obtained by dividing the number of correctly classified pixels in each category (on the major diagonal) by the number of training set pixels used for that category (the column total), which is known as the sensitivity or true positive rate, and the complement of the omission error. The user's accuracy, producer's accuracy, overall accuracy, error of omission and error of commission were computed including the confusion matrix for the classification in Google Earth Engine computing. Finally, based on the aforementioned indicators of accuracy, the better-performing indices were selected and then used to quantify the spatial distribution of surface water resources in Ethiopia.

2.3 Results

2.3.1 Surface water detection

The results of water detection using seven indices for specific locations and the whole of Ethiopia were analyzed by visual inspection (Figurs 2.2, 2.3, and 2.4). The optimum threshold values that separate water from non-water pixels were selected for NDWI (0.06), MNDWI (0.14), AWEIsh (0), AWEInsh (0.04), WI (0.02), SWI (0.03) & LSWI (0.12). WI and AWEIsh were relatively better at detecting surface water where the surrounding areas are vegetated areas and urban areas (Figure 2.2, #1 and #3). SWI was less good as there was some misclassification as water bodies (Figure2. 2, #1 and #2). NDWI was relatively less effective than MNDWI because of some unclassified water pixels (Figure 2.2, #1 and #3). In this case, AWEInsh only detects large water bodies, and LSWI was also unable to detect water bodies correctly. In flat areas, WI and AWEIsh were more effective to detect rivers compared to other water indices (Figure 2.3). MNDWI was also superior in these flat and less vegetated areas, while NDWI was less able to map the rivers. SWI detected the river but misclassified non-water pixels as water pixels (Figure 2.4). AWEInsh and NDWI were unable to identify rivers and small water bodies. Therefore, in this large-scale water detection experiment using Sentinel-2 data in Google Earth Engine, the WI and AWEIsh were most effective for surface water detection from a visual inspection point of view.

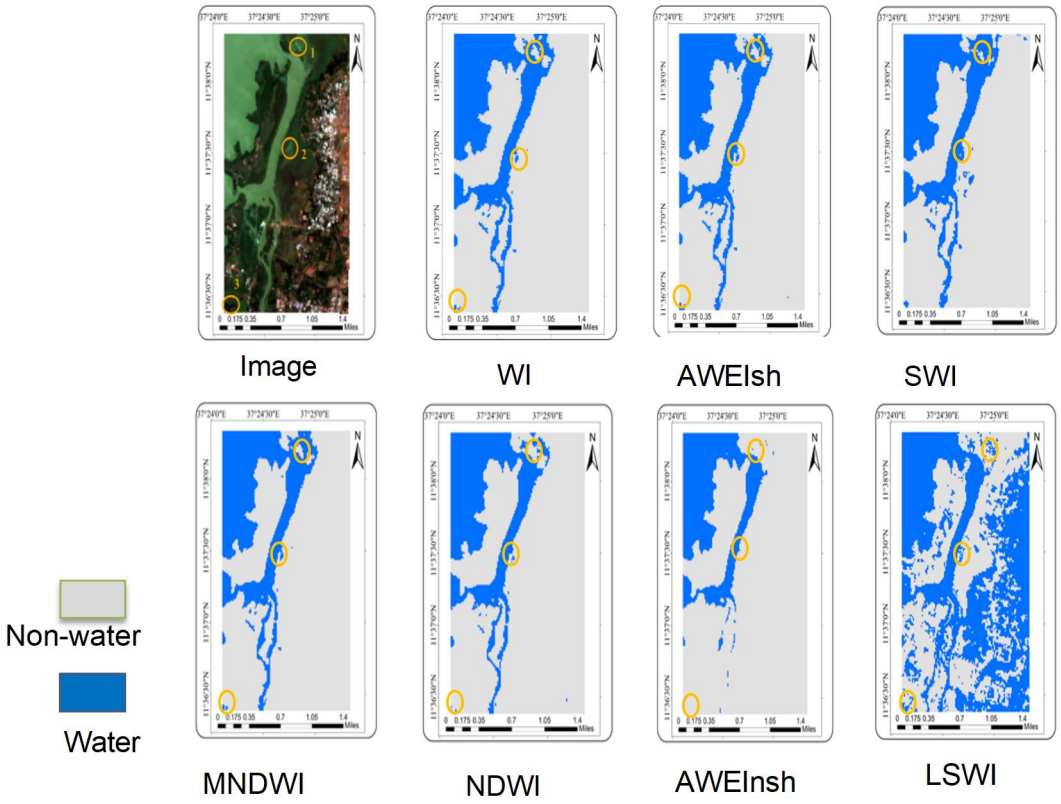


Figure 2.2 Specific map section for visual assessment and comparison of water indices outputs for part of the lake Tana area in Bahir Dar

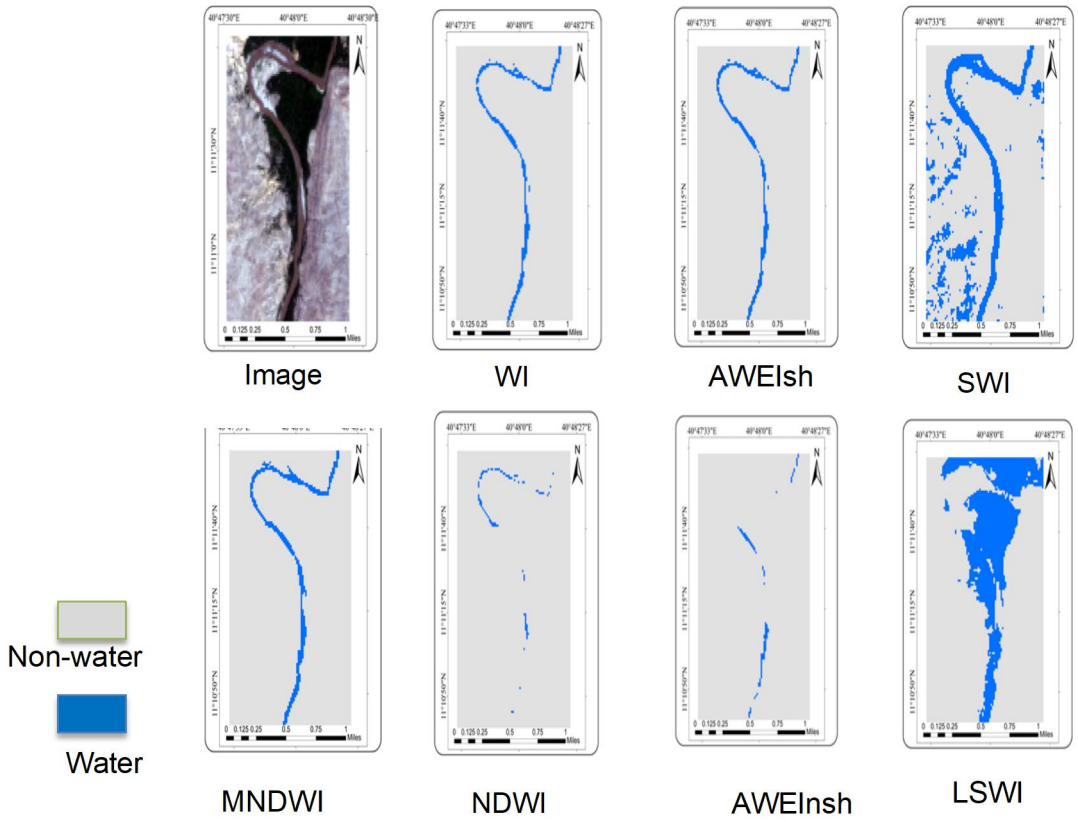


Figure 2.3 Specific map section for visual assessment and comparison of water indices outputs for the Awash river reach in Adaytu, Afar

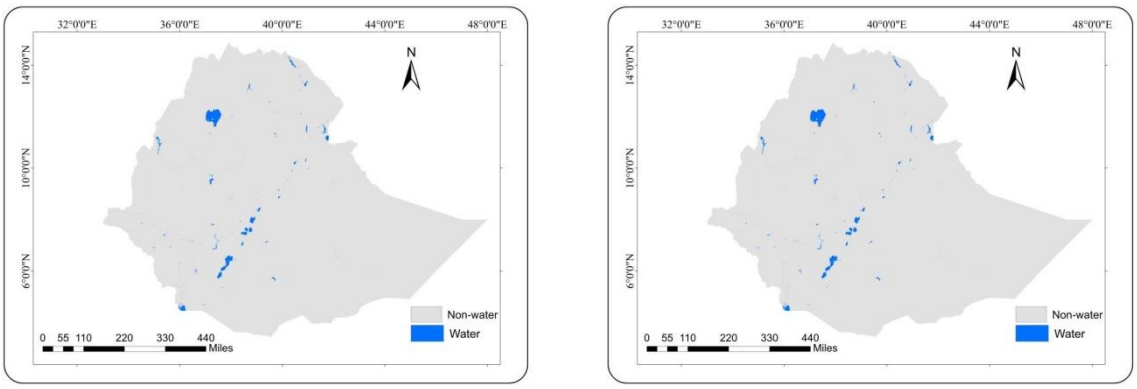


Figure 2.4 Ethiopia surface water coverage map using best-performing water indices

2.3.2 Performance evaluation of water indices

The water producer's accuracy ranged from 32 to 96%, whereas non-water producer's accuracy ranged from 84 to 97%. All indices, except LSWI, had a satisfying classification accuracy > 85% with regard to the overall accuracy, producer's accuracy, and user's accuracy (Table 2). The WI achieved the highest accuracy with an overall accuracy of 96%. All other indicators, except for the LSWI, had only slightly worse performance criteria in terms of user's and producer's accuracy for water, as well as overall accuracy. The difference in accuracy between the LSWI and the other indices was striking. This index was the least-performing one with an overall accuracy of 78%. Overall, the AWEIsh and particularly WI outperform the other indices regarding accuracy indicator performances depicted by the heat map (Table 2.2), whereby it must be said that the differences between six of the seven indices were in part only marginal. Together with the information from the visual assessment, we concluded that in this study, the WI and AWEIsh were the most accurate water indices using the high and multi-temporal resolution of Sentinel-2 data for surface water detection in Ethiopia.

Table 2.2 Heat map of classification accuracies of water indices

	Producer accuracy Water (%)	Error of omission	Producer accuracy Non-water (%)	Error of omission	User accuracy Water (%)	Error of commission	User accuracy Non-water (%)	Error of commission	Overall accuracy (%)
NDWI	85	15	92	8	90	10	92	8	90
MNDWI	92	8	93	7	93	7	93	7	92
AWEInsh	94	6	95	5	93	7	95	5	94
AWEIsh	95	5	96	4	94	6	97	3	95
WI	96	4	97	3	96	4	96	4	96
SWI	91	9	94	6	93	7	94	6	93
LSWI	32	68	84	16	36	66	81	19	78

Finally, the surface water coverage was calculated using the seven indices across Ethiopia. WI, AWEIsh, and MNDWI indices extracted more or less similar surface water coverages of 82,650, 86,530, and 88,160 km^2, respectively. Slightly deviating was the coverage estimated by NDWI (79,650 km^2). Completely different were the remaining coverages derived by the AWEInsh (51,800 km^2), SWI (111,500 km^2), and LSWI (207,520 km^2).

2.4 Discussion

The special spectral properties of water are utilized when detecting surface water and non-water surfaces. Water absorbs light above 700 nm, particularly in the NIR and MIR bands. Thus, water bodies can be identified based on their unique absorption characteristics by analyzing the reflectance spectrum of different surfaces. This analysis goes beyond color information and is based on how water behaves in the electromagnetic spectrum. In the visual assessment, the results showed that the WI and AWEIsh indices had a superior accuracy in surface water detection. They were more effective to detect surface water in urban and vegetated areas, despite that the detection in such areas is commonly difficult due to shadows (Feyisa et al., 2014). Shadows in urban areas are often misclassified as water bodies, as they have similar low reflectivity characteristics to water bodies (Liu et al., 2022). Fisher et al. (2016) stated that WI and AWEIsh performed best, whereas the MNDW and AWEInsh performed less accurately, and NDWI the least. In this study, MNDWI was more effective than NDWI, AWEInsh and SWI in detecting rivers and small water bodies. The MNDWI is an improved version of the NDWI in that it uses the SWIR band instead of the NIR band that is used in the NDWI to normalize the water and vegetation indices (Xu, 2006). NDWI was less sensitive to small water bodies and rivers than MNDWI and was also affected by noise from urban and vegetated areas. This is due to the spectral signature response of water bodies, which is less sensitive in the NIR band of the NDWI than in the MIR band of the MNDWI (Xu, 2006). In contrast, the MIR band in MNDWI can enhance the contrast between water and surrounding land, making it more effective at detecting smaller water features that may be missed by NDWI. In addition, the MIR band used in MNDWI is less sensitive to vegetation and less affected by noise from urban areas than the NIR band used in NDWI. As a result, the NDWI was unable to detect part of the Awash River and small water bodies in the Lake Tana area (Figurs 2 and 3). The contrast between water and land is acceptable for MNDWI, although it is less accurate than the WI and AWEI in detecting small water bodies or small streams (Liu et al., 2022). SWI was better than NDWI, AWEInsh, and LSWI for the detection of the same features. However, it was challenged by the elimination of shadow noise from surrounding non-water features. In general, SWI is better than NDWI at detecting wide river channels (Jiang et al., 2021).

The results showed that the WI and AWEIsh were the most accurate water indices. A similar finding was observed in four case studies in Switzerland, Ethiopia, South Africa, and New Zealand, where AWEIsh achieved water user's and producer's accuracies of 96–

99% and 91–99%, respectively (Feyisa et al., 2014). Similarly, AWEInsh was also detected in a shadow-free image from Denmark with water user's and producer's accuracy of 98% and 92%, respectively (Feyisa et al., 2014). AWEI with Google Earth Engine is a quick and robust method for surface water monitoring (Nguyen et al., 2019). Overall accuracies of WI (98%) were also outperforming other indices (MNDWI 97%, NDWI 95%) in image analyses from eastern Australia (Fisher et al., 2016). WI achieved an overall accuracy of 96% and AWEIsh also performed well with an overall accuracy of 96% (Liu et al., 2022). In this study, LSWI performed worse due to the limitation of the index in interpreting Sentinel-2 data in this study area. It was also heavily influenced by the background noise of non-water features in the study area.

Overall, the performance evaluation criteria in this study are common and acceptable for evaluating the results of surface water detection using water indices (Feyisa et al., 2014; Fisher et al., 2016). The results present were even slightly higher than those of previous works, which range from 90 to 96% in overall accuracy, except for LSWI. This is perhaps due to the use of relatively higher spatial and spectral resolution of the Sentinel-2 data with Google Earth Engine processing. In addition, Sentinel-2's band characteristics and spectral response methods are likely more effective than Landsat imagery at detecting water bodies from the background (Jiang et al., 2021).

In this study, the better-performing WI and AWEIsh predicted 82,650 and 86,530 km^2 of surface waters in Ethiopia, respectively. A slightly higher coverage of 91,056 km^2 was obtained by extracting the Global Surface Water dataset which was developed by the European Commission's Joint Research Centre (Pekel et al., 2016) (Figure 2.5). Also, on a smaller domain, our assessment proofed good results. Lake Tana and Zeway surface water coverages were estimated to 3095 and 408 km^2, similar to results extracted from the Global Surface Water dataset with 3132 km^2 and 434 km^2, respectively (Pekel et al., 2016). In other studies, the Lake Tana surface water area was estimated to 3041 km^2 (Ayele & Atlabachew, 2021) and that of Lake Zeway to 418 km^2 between October 2010 and February 2016 (Asfaw et al., 2020).

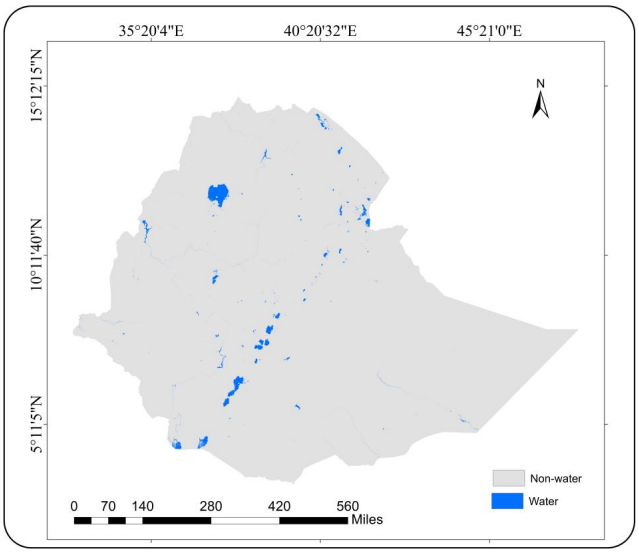


Figure 2.5 Extracted map of surface water from GSW dataset

The best-performing indices, WI and AWEIsh, were efficient and replicable in Sentinel-2 data using Google Earth Engine computing. Even though they are specifically designed for Landsat sensors (Feyisa et al., 2014; Fisher et al., 2016), they are also effective in producing high-quality maps for large-scale surface water detection using high-resolution Sentinel-2 imagery. Consequently, these methods are useful for policymakers, experts and stakeholders in water resources monitoring, also in the light of missing monitoring efforts of SDG6 for Ethiopia.

Ethiopia's man-made reservoirs have become critical to the country's energy and economic development, particularly through hydropower projects like the Grand Ethiopian Renaissance Dam (GERD). The natural lakes of Ethiopia, such as Lake Tana and Lake Zeway, are freshwater lakes and have a considerable amount of outflow. They have both ecological features and cultural significance. They support a variety of ecosystems, supporting aquatic life and contributing to biodiversity. In addition, natural lakes but also wetlands provide ecosystem services such as water filtration and climate regulation. However, changing land use and land cover due to population pressure, water scarcity due to contamination, soil erosion, over grazed land, and alien species are major problems of African lakes (Singh et al., 2006). In the northern and western areas of the Lake Ziway, high expansion of local and commercial agricultural irrigation is exhibited (Asfaw et al., 2020). Understanding and mitigating anthropogenic and environmental impacts will be critical to achieving a harmonious balance between development and conservation as Ethiopia

navigates the complex interaction of man-made and natural water bodies. Water indices per se cannot distinguish between man-made (e.g., constructed wetlands, river diversions, reservoirs) and natural water surfaces. But water indices provide excellent tools for monitoring spatial changes in water coverage. Together with information of long-term land use, water indices could also contribute in the separation of man-made versus natural changes in water surface areas. This would require a sufficiently long time series of land use maps. However, such information is currently not available for Ethiopia.

2.5 Conclusion

This paper demonstrated the potential of water index methods and selected the best-performing indices for large-scale surface water detection using Sentinel-2 data with Google Earth Engine in Ethiopia. Sentinel-2 data processed within Google Earth Engine has immense potential to provide large-scale surface water detection with a very high efficiency, accuracy, and temporal frequency, thereby supporting water resource monitoring and management. The results showed that WI and AWEIsh were the best performers in terms of accuracy indicators for overall accuracy, producer's accuracy, and user's accuracy, ranging from 0.94 to 0.97. Applying both indices, the WI and AWEI extracted surface water areas of 82,650 and 86,530 km^2 respectively. The results confirmed that WI and AWEIsh indices using Sentinel-2 data provide reliable assessments of surface water coverage. Utilizing such indices could substantially improve the monitoring of the country's reservoirs, which have become critical to the country's energy and economic development. These best-performing indices could play a critical role in surface water monitoring for water resource planners and decision-makers. They provide accurate, timely, and reliable spatial information to support informed decision-making, planning and development, risk management, and policy formulation related to surface water resources. This would be all the more the case if high-temporal resolution sequences of remote sensing images were available. The application of water indices could also successfully contribute to the achievement of SDG 6 at the regional level, as they are useful for guiding conservation efforts and the sustainable use and management of surface water resources.

However, any approach of surface water detection that only considers remote sensed imagery has difficulties in making future projections of surface water coverage, as images are always from the past. For future projections, for example, if information is needed on climate or land use change effects on surface water resources, dynamic variables that change over time should complement surface water detection. This likely includes hydro-

meteorological variables from weather forecasts or climate projections as well as dynamic land use features of land use projections. Such information should be fused with satellite imagery to enable future projections. Machine learning models with data fusion techniques, which have recently shown promising results for accurate surface water detection over a wide range of environmental conditions, are seen as having particular potential for this purpose.

3 Remote sensing with machine learning for multi-decadal surface water monitoring in Ethiopia

This chapter is published in *Scientific Reports* as:

Tesfaye, M., & Breuer, L. (2025a). Remote sensing with machine learning for multi-decadal surface water monitoring in Ethiopia. *Scientific Reports*, *15*(1), 12444. https://doi.org/10.1038/s41598-025-96955-y

Abstract

Monitoring the temporal evolution of surface water distribution is crucial to support surface water management and conservation, and could also effectively contribute to the achievement of Sustainable Development Goal 6 (SDG 6) 'Clean Water and Sanitation' at the regional level. Despite its importance, there is a lack of an operational method for determining surface water extent that also shows the interannual variability in Ethiopia. We assess Gradient Tree Boosting (GTB), Support Vector Machines (SVM), and Random Forest (RF) running on the Google Earth Engine (GEE) using Landsat for surface water monitoring at four sites in Ethiopia from 1986 to 2023. The results show that GTB, RF, and SVM have excellent classification accuracies, with overall, producer, and user accuracies consistently above 90%. GTB slightly outperforms the other two machine learning methods. The estimated water cover for our study sites shows a high degree of agreement with a benchmark dataset from the Joint Research Center (JRC), as indicated by coefficient of determination (R2) > 0.9 and root mean square percentage error (RMSPE) < 1%. The surface water dynamics of the four study sites depict a long-term increasing trend from 1986 to 2023, characterized by notable inter-annual variability. We identify the locations of this variability by analyzing the frequency of water occurrence over time and find that 84–94% are permanent water bodies, with the remaining water surface area changing over time. Mann–Kendall trend analysis does not confirm a general pattern over time for the four sites, suggesting that local site characteristics, water management and anthropogenic impacts are superimposed on the likely effects of climate change. Therefore, our results provide spatiotemporal information for surface water monitoring to support water resource management and policy in Ethiopia. This could also effectively contribute to the sustainable use and achievement of SDG 6 at the regional level.

Keywords Surface water dynamics, SDG 6, Remote sensing, Google Earth Engine, Machine learning

3.1 Introduction

Surface water is an important natural resource that is subject to considerable spatial and temporal dynamics. It comprises a range of inland water bodies, including rivers, streams, ponds, reservoirs, lakes and wetlands (Pekel et al., 2016; Vandas et al., 2002). They are a critical components of water resources for humans and terrestrial ecosystems, covering approximately 3% of the world's land area (Zou et al., 2018). In addition to maintaining biodiversity and providing essential and diverse ecosystem services (Brauman, 2015; Dudgeon et al., 2006; Zedler & Kercher, 2005), surface waters play a crucial role in the climate system and the water cycle (Chahine, 1992; Tranvik et al., 2009). However, the biodiversity of these ecosystems is rapidly declining (Collen et al., 2014), exacerbated by climate-induced hydrological extremes such as floods or droughts and the emergence of water-related diseases, both of which contribute to escalating loss of life. Therefore, timely monitoring of surface water dynamics is crucial for informed policy and decision making processes to ensure its sustainable use and management (Giardino et al., 2010; Morss et al., 2005).

Ethiopia has substantial surface water resources, comprising eight river basins, one lake basin, and three arid basins (Figure 1). A notable rainfall gradient separates the central and western highlands, which receive abundant annual rainfall of up to 1200 mm, from the arid southeastern, eastern, and northeastern regions, which receive 200 mm and less (Berhanu et al., 2014). As a result, surface water resources are less abundant in the eastern part of Ethiopia, particularly in the Awash Basin, while the western part, including the Abay (Blue Nile) River Basin, has substantial water resources. Despite the critical importance of water resources to both human well-being and the country's ecosystems, monitoring systems for surface water availability and hydrological flows are in a poor state and have sometimes proven to be unreliable or dysfunctional (Dile et al., 2018). A comprehensive assessment of a long-term indicator of surface water resources, as required by SDG 6, is currently lacking.

Although Ethiopia is recognized as having abundant water resources, this potential is not fully exploited and translated into basin planning and development due to technical constraints and inefficient management in the water sector (Berhanu et al., 2014). Flooding is particularly prevalent in the lowlands of parts of the Awash river basin in Ethiopia (Legass et al., 2025a) and has resulted in crop damage, loss of soil nutrients and erosion (Desta et al., 2024). Water-related ecosystems in Ethiopia are experiencing various

challenges including inter-annual climate variability and changing water regimes (Seyoum et al., 2015) or increasing sedimentation (Gadissa et al., 2018; Haregeweyn et al., 2012). Thus, timely monitoring of surface water resources and detection of spatiotemporal changes is critical for sustainable use and management of the country's water resources.

Satellite remote sensing provides the ability to continuously monitor surface water at multiple spatio-temporal scales (Halabisky et al., 2016; Pekel et al., 2016; Tulbure et al., 2016). This can fill in the gaps in poorly mapped areas. Landsat imagery is the primary data source for long term surface water monitoring due to its consistency, free access, and long-time availability (Hansen et al., 2014; Pekel et al., 2016).

Recent advances in remote sensing technology, characterized by high spatial, temporal and spectral resolution, have contributed substantially to the monitoring of surface waters, with a focus on their accurate detection (McCarthy et al., 2017). However, monitoring lakes and reservoirs and providing accurate spatio-temporal information on the dynamics of surface water extent is still a challenge for classification algorithms due to their characteristics. This challenge is demonstrated for example for Lake Mead (USA), Nova Ponte reservoirs (Brazil) and Lake Williston (Canada) due to their temporal variability, different narrow ends, dynamic shape, and missing values due to cloud or ice cover (Khandelwal et al., 2017). Therefore, this study explores the potential of machine learning to detect and monitor surface water bodies using Landsat data in a cloud computing platform.

Large-scale surface water monitoring is a challenge in processing large amounts of remotely sensed imagery (Wang et al., 2018). In recent years, Google Earth Engine (GEE) has become a cloud computing platform for analyzing large amounts of data, including global datasets. It provides access to a range of satellite imagery, particularly Landsat and Sentinel products, and facilitates high-performance computing for social and environmental analysis, including water monitoring (Gorelick et al., 2017). Thus, in our study, we use GEE to process and analyze surface water information from Landsat imagery using machine learning.

Water indices based on a multi-band classification method are often used in remote sensing to detect surface water (Fisher et al., 2016). Water indices are quick and easy to calculate (Ryu et al., 2002). However, the spectral signatures of water bodies vary in space and time, making it challenging to select a common threshold for distinguishing water bodies from non water bodies (Tulbure et al., 2016). These problems can be overcome by using machine

learning algorithms such as Gradient Tree Boost (GTB), Random Forest (RF) and Support Vector Machine (SVM).

Several studies have explored the detection and dynamics of surface water, leveraging advanced machine learning and remote sensing techniques. For instance, a random forest approach was applied to almost four decades of satellite data to analyze surface water variability in a semi-arid region of Australia (Tulbure et al., 2016). In the same way, spatio-temporal variations in surface water from 1985 to 2020 have been investigated using machine learning algorithms applied to time-series remote sensing datasets (Tariq & Qin, 2023). In the United States, a gradient-boosted regression model has been employed to distinguish between open and vegetated water bodies, providing enhanced classification accuracy (Vanderhoof et al., 2023). Australia has also been the focus of large-scale, multi-year surface water mapping efforts, offering high-resolution spatio-temporal data critical for regional planning and water resource management (Mueller et al., 2016). Moreover, the hydrological dynamics of Ethiopia's major lakes have been assessed using RF and modified normalized water index, further contributing to our understanding of surface water fluctuations in diverse environments (Getaneh et al., 2024). However, surface water monitoring and trends, including lakes surrounding rivers, ponds, reservoirs, and wetlands, has not been addressed. Despite the advancements, existing studies primarily focus on individual models applied to specific geographic regions. To the best of our knowledge, a comprehensive assessment of multiple modeling approaches across diverse geographical locations for surface water detection and long-term monitoring remains unexplored.

The JRC dataset provides surface water data for the entire globe1. However, for a regional application like Ethiopia, there are some limitations in the JRC data, such as data gaps and lack of detailed detection and analysis of local spatio-temporal variability in the operational application using machine learning with GEE. Nevertheless, the JRC data set allows to benchmarking new methods in surface water detection, which can then be used for gap filling or further regional analysis. Therefore, this study aims to (i) evaluate a number of machine learning methods (i.e., GTB, SVM, RF), (ii) compare the results with the JRC dataset for time series surface water monitoring, and (iii) statistically analyze inter-annual and long-term trends of surface water coverage in Ethiopia.

3.2 Materials and methods

3.2.1 Description of study area

Ethiopia has significant surface water resources and 12 major river basins (Figure 3.1). Four of Ethiopia's river basins, namely Abay or Blue Nile, Baro-Akobo, Tekeze, and Mereb, which are part of the Nile Basin, cover 33% of the country and drain the northern, central, and western parts (FAO, 2016). The four river basins of Abay (Blue Nile), Tekeze, Baro Akobo, and Omo Gibe contain about 90% of Ethiopia's water sources (Mola Shako & Ping, 2021). In the eastern part of Ethiopia, surface water resources are limited as there are almost no perennial rivers found below 1,500 m a.s.l. Three of the major basins (Aysha, Dinakle, and Ogaden) are mainly dry and have no permanent runoff (Berhanu et al., 2014). Ethiopia's 12 major lakes cover an area of about 7300 km^2. Lake Tana (in the Abay Basin) is the largest and is the main source of water for the Abay River. Most of the other lakes are saline and located in the Rift Valley. In this study, four sites are selected due to the diverse availability of water resources including lakes, rivers, reservoirs and wetlands and their complex interactions with human activities across different geographical locations and basins of Ethiopia (Figure 3.1). The first site is the Addis-Ziway site which includes Addis Ababa and its surroundings, Bishoftu lakes, Ziway and Koka reservoirs. The Hayq-Hashenge site is the second site covering the northeastern part of Ethiopia including Hayk, Hardibo and Hashenge. Thirdly, the Tana site covers mainly Lake Tana. Finally, the Abaya Chamo site includes Arba-Minch and the lakes of Abaya and Chamo.

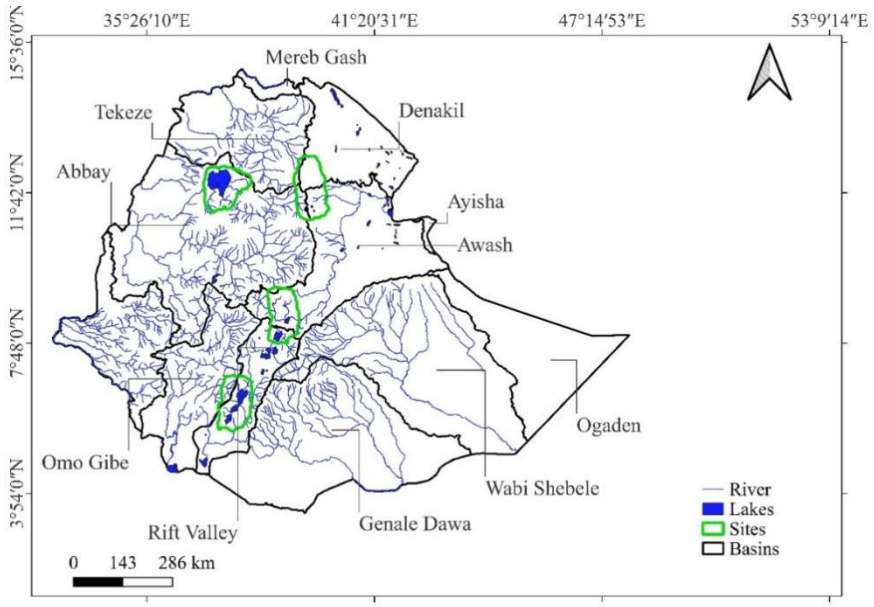


Figure 3.1 Location of the study areas with the main water bodies of Ethiopia

3.2.2 Data acquisition and preprocessing

We utilize Tier 1 Landsat time series imagery retrieved from the United States Geological Survey (USGS) data provider, obtained through GEE's data access from 1986 to 2023. We select Landsat imagery for its continuous access at 30 m spatial resolution over time and a 16-day revisit time. However, the Landsat sensor does not detect smaller water bodies due to its spatial resolution (Pekel et al., 2016; Pickens et al., 2020) and might miss observations resulting in an underestimation of surface water extent. In this study, Landsat imagery includes the Thematic Mapper (TM) sensor for Landsat 5 covers from 1986 to 1998; the Enhanced Thematic Mapper Plus (ETM +) for Landsat 7 between 1999 and 2012 and the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) for Landsat 8 and OLI-2 and TIRS 2 for Landsat 9 covers from 2013 to 2021 and 2022 to 2023, respectively. For each time step, the feature vector consists of multiple spectral bands, including quality assurance (QA) bands to providing important information for surface water detection and improving classification accuracy. The spectral bands included in the dataset cover various portions of the electromagnetic spectrum, supporting accurate distinguishing of water bodies from other land cover types. The blue band covers wavelength ranges from 0.45 to 0.52 μm in Band 1 for Landsat 5/7 and 0.45–0.51 μm in Band 2 for Landsat 8/9. The green band captures wavelengths from 0.52–0.60 μm in Band 2 for Landsat 5/7 and 0.53–0.59 μm in Band 3 for Landsat 8/9. The red band represents Band 3 (0.63–0.69 μm) for Landsat 5/7 and Band 4 (0.64–0.67 μm) for Landsat 8/9. The near-infrared (NIR) band covers wavelengths from 0.77 to 0.90 μm in Band 5 for Landsat 5/7 and 0.85–0.88 μm in Band 5 for Landsat 8/9. The shortwave infrared 1 (SWIR1) band represents Band 6 (1.55–1.75 μm) for Landsat 5/7 and Band 6 (1.57–1.65 μm) for Landsat 8/9. The Short Wave Infrared 2 (SWIR2) band covers wavelengths from 2.09 to 2.35 μm in band 7 for Landsat 5/7 and 2.11–2.29 μm in Band 7 for Landsat 8/9. In addition to spectral bands, QA bands are included into the feature vector to enable cloud and shadow masking, thereby enhancing classification accuracy.

The standard Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm is used to produce surface reflectance data for Landsat 5 and Landsat 7 (Masek et al., 2006; Schmidt et al., 2013). The Landsat surface reflectance code (LaSRC) algorithm is used to produce surface reflectance data for Landsat 8 (Vermote et al., 2016). Landsat 9 surface reflectance is also produced using the LaSRC algorithm. These products are radiometrically and geometrically corrected. We use the GEE cloud computing platform to

preprocess Landsat time series imagery. We chose GEE for its ability to handle large remotely sensed data and long-term time series analysis (Gorelick et al., 2017). The USGS L7 Phase-2 Gap Filling Protocol (Scaramuzza et al., 2004) is employed to correct scan line gaps for Landsat 7 imagery acquired after 2003. We combine the multiple Landsat images into a single annual composite per site using the GEE median reducer function (Z. Zhang et al., 2021). The data are preprocessed to include blue, green, red, near-infrared, shortwave infrared 1, and shortwave infrared 2 spectral bands, which are primarily used for water detection (Wulder et al., 2016), as well as quality assurance (QA) bands for cloud and shadow masking. The CFmask algorithm (Zhu and Woodcock, 2012) is used for cloud and shadow masking to ensure data quality. This process identifies pixels affected by clouds or shadows and excludes them from further analysis using the QA_PIXEL band.

3.2.3 Machine learning for surface water detection

Machine learning algorithms are employed to detect, and quantify the extent of surface water, like to other studies (Chignell et al., 2018; Hamunyela et al., 2022; Tulbure et al., 2016; Vanderhoof et al., 2023). Training samples are used to train the machine learning for surface water detection and area estimation using time series Landsat imagery, and validation samples are used to assess their accuracy. These sample points are collected at four sites using a stratified random sampling approach from water and non-water areas, respectively. Stratified random sampling approach is mainly used in remote sensing based classification because smaller classes can be adequately represented regardless of the size of the features and their spatial extent (Gao, 2009). Spatial stratification also improves the classification accuracy of remote sensing (Dong et al., 2022). For example, stratified random sampling involves dividing the study area into distinct strata of water and non-water. The water and non-water training sample points are collected from the satellite images from 1986 to 2023. These samples are collected manually through visual interpretation similar to previous studies (Getaneh et al., 2024; Tulbure et al., 2016; Vanderhoof et al., 2023) considering the type, size, color, and pattern of inland water bodies including rivers, streams, ponds, reservoirs, lakes, and wetlands (Pekel et al., 2016; Pickens et al., 2020) and non-water bodies defined except these features. Then, random samples are drawn independently from each layer to ensure balanced and accurate representation of the areas and to train the model. Water pixels are collected from the sites Addis-Ziway, Hayq-Hashenge, Tana, and Abaya-Chamo, with a number of 1086, 1124, 1218, and 1036, respectively. The non-water

pixels from the same sites are 1091, 1228, 1025, and 1038, respectively. Thus, a total of 4464 water pixels and 4382 non-water pixels are used to train each model.

We use the GEE Python application programming interface (API) and the geemap package to access, process, and analyze the Landsat images in the Google Colab environment. We select GTB, RF, and SVM as widely used and efficient tools for remotely sensing surface water detection across space and time. For example, GTB was used to detect changes in open water and vegetated water at 12 different sites from 2017 to 2021 (Vanderhoof et al., 2023). Similarly, RF has been used to assess surface water dynamics in semi-arid regions at sub-continental scale from 1986 to 2011, where it demonstrated its robustness and efficiency in handling remote sensing data (Tulbure et al., 2016). In addition, SVM has been shown to be effective in quantitative and qualitative assessment of the detection of large inland water bodies (Schmitt, 2020). Each machine learning method is briefly described below:

(1) GTB is an ensemble learning method that builds a predictive model in the form of an ensemble of weak learners into a stronger ensemble using gradient descent optimization to minimize the loss function (Friedman, 2002). This is done by successively refining a model by adding new trees to correct errors. This sequential nature distinguishes GTB from other ensemble methods such as RF, where trees are built independently. In GTB, each new tree is trained to predict the residuals (errors) of the previous combined ensemble. The ensemble is then updated by adding a weighted version of the new tree. This iterative process continues until a predefined number of trees is reached.

The estimation function *f(x)* maps x to y while considering for the joint distribution of all (y,x) values to model the relationship between a random output (or response) variable y, and a set of random input (or explanatory) variables x = {× 1,…, xn} (Friedman, 2002). Given a training dataset.

$\{\mathrm{yi}, \mathrm{xi}\}_{i=1}^{N}$ with known values of y and x , *X*= {*x*1 , *x*2 ,…,*xn* } is the features of Landsat data (spectral bands) while *Y*={*y*1 ,*y*2 ,…,*yn* } is the labels (water = 1, no-water = 0).

The algorithm starts by initializing the model $F_0(x)$ that minimizes the loss function $\Psi(yi,\gamma)$ over all training samples (*xi,yi*) in Eq. (1).

$$F_0(x) = arg\ min\gamma \sum_{i-1}^{N} \Psi(yi, \gamma) \quad (1)$$

The algorithm builds M trees sequentially, where each tree corrects the errors of the previous ones. Thus, it iterates for m= 1 to M (Number of Trees). For each training sample i, the pseudo-residual $\tilde{Y}im$ is calculated , which is the negative gradient of the loss function Ψ regarding to the current model $Fm-1$ (x) Eq. (2).

$$\tilde{Y}im = -\left[\frac{\partial \Psi(yi,F(xi))}{\partial F(xi)}\right]_{F(x)=Fm-1(x)} ,i=1, N \quad (2)$$

In the next step a decision tree with L terminal nodes (leaves) is fit to the pseudo-residuals $\tilde{Y}im$ using the input features xi Eq. (3) .

$$\{Rlm\}_1^L = L - terminal\ node\ tree\{\tilde{Y}im, xi\}_1^N \quad (3)$$

For each leaf region Rlm , the optimal value γlm that minimizes the loss function is calculated when added to the current model $Fm-1$ (x) Eq. (4).

$$\gamma lm = arg\ min \sum_{xi \in Rlm} \Psi(yi, Fm-1(xi) + \gamma) \quad (4)$$

Following, the model is updated by adding the predictions of the new tree, weighted by a learning rate ν (shrinkage parameter) Eq. (5).

$$F_m(x) = F_{m-1}(x) + \nu \cdot \gamma lm 1(x \in Rlm) \quad (5)$$

The number of decision trees depends on the complexity of the model. A larger number of decision trees can result in a more accurate but more computationally intensive model. The shrinkage parameter controls the contribution of each tree to the model and helps prevent overfitting, while the maximum node parameter limits the size of individual trees to avoid overly complex structures. This combination of parameters aims to strike a balance between model accuracy and computational efficiency. This makes it suitable for surface water detection in GEE. We conduct hyperparameter tuning for the GTB classifier using a grid search with tenfold cross-validation, evaluating combinations of the decision trees (100, 300, 500), learning rate (0.01, 0.05 and 0.2), maximum leaf nodes (5, 10 and 20), sampling rate (0.6, 0.7 and 1.0), and loss (least square and least absolute deviation). The algorithms arrive at the final hyperparameter values of 500 for decision trees, 0.05 for learning rate, 20 for maximum leaf nodes, 1.0 for sampling rate and least square loss function.

(2) RF is a non-parametric method that can deal with non-linear and large data sets. The RF builds the trees in parallel procedures that are fully grown, and each tree is used to predict the missing observations that do not occur in a bootstrap sample (Breiman, 2001). The average of all predictions outputs is used to calculate the predicted class of an out-of-bag observation (Breiman, 2001). We undertake hyperparameter tuning by grid search with tenfold cross validation, evaluating combinations of the decision trees (100, 300, 500), minimum leaf population (1, 5 and 10), bag fraction (0.6, 0,8 and 1.0) and maximum leaf nodes (5, 10 and 20). The algorithms arrive at the optimal hyperparameter values of 100 for decision trees, 1 for minimum leaf population, 0.6 for bag fraction, 5 for maximum leaf nodes.

RF is an ensemble classification model which mainly uses trees structured classifiers $\{h(x,\Theta k),\ k = 1,\ldots\}$, where the $\{\Theta k\}$ are independent and identically distributed random vectors and x is an input (Breiman, 2001; Hastie et al., 2009). Each tree h($x,\Theta k$) contributes a single vote for the most frequently occurring class at input x and the final classification is assigned based on the majority vote across all trees in Eq. (6) (Hastie et al., 2009). For classification, $\hat{C}b(x)$ be the class prediction of the b-th tree in the random forest. Then, the final random forest classification is given by:

$$\hat{\mathrm{C}}\mathrm{b}(\mathrm{x}) = \text{majority vote } \{\hat{\mathrm{C}}b(x)\}_{b=1}^{B} \tag{6}$$

where B is the total number of trees in the forest.

(3) The SVM finds a hyperplane or set of hyperplanes in infinite dimensional space that uniquely classifies the data points and finds the optimal separating hyperplanes (Rodrigues & De La Riva, 2014). The balance between achieving a smooth decision boundary and correctly classifying training points is controlled by the regularization parameter cost. The parameter gamma is also relevant to the flexibility of the decision boundary. The choice of the kernel type is important which includes linear, polynomial and radial basis function. Therefore, we choose the polynomial kernel among linear and radial basis function based on the higher classification accuracy results in distinguishing water from non-water pixels. The degree parameter is specific to the polynomial kernel. The hyperparameters of the polynomial kernel allow fine-tuning of the decision boundary using a grid search with tenfold cross validation, evaluating the combination of gamma values (0.0001, 0.001, 0.01, and 0.1), cost values (0.1, 1, 10, 100, 1000), and degree (2, 3, and 4). The algorithms arrive

at the optimal values of 0.0001, 0.1, and 2 for gamma, cost, and degree, respectively, to ensure that the model achieves high accuracy.

The optimal hyperplane algorithm consists of the labeled training patterns (*Y1, X1*), *.....,(Yl, Xl), Yi ∈ {-1, 1},* which is linearly separable when there is a vector w and a scalar b satisfying the inequalities for all items of the training dataset in Eqs. (7), (8) (Cortes & Vapnik, 1995; Huang et al., 2002).

$$w \cdot xi \quad + b \geq 1 \quad if\, yi \quad = 1 \tag{7}$$

$$w \cdot xi \quad + b \leq -1 \quad if\, yi \quad = -1 \tag{8}$$

The above equation can be written, combining the inequalities as Eq. (9).

$$yi \quad (w \cdot xi \quad + b) \geq 1,\ i = 1, \ldots, l \tag{9}$$

The standard optimization method is used and the Lagrangian is constructed in Eq. (10)

$$L(w, b, \Lambda) = \frac{1}{2} w \cdot w - \sum_{i=1}^{l} \alpha i [yi(xi \quad \cdot w + b) - 1], \tag{10}$$

where $\Lambda T = (\alpha 1, \ldots., \alpha l)$ is the vector of Lagrange multipliers that incorporates the constraints. The decision rule separating the two categories is written as follows Eq. (11)

$$f(x) = sign\ (\sum_{support\ vector} yi\, \alpha^0 (xi\ 'x) - b^0) \tag{11}$$

where α^0 and b^0 are the optimized Lagrange multipliers and bias term, respectively. As mentioned before, we use a polynomial kernel given in Eq. (12).

$$K(x\ 1\ ,\ x\ 2\ ) = ((x1\ 'x2 + 1)^p \tag{12}$$

where p is the degree of the polynomial kernel. Thus, this extends the decision rule to use the kernel function for non-linear classification according to Eq. (13).

$$f(x) = sign\ (\sum_{support\ vector} yi\, \alpha^0 K(xi\ 'x) - b^0) \tag{13}$$

3.2.4 Accuracy assessments

Accuracy assessments are carried out to determine whether or not the results of surface water detection and area estimation are acceptable. Validation samples are used to assess the accuracy of the three machine learning methods and the JRC dataset at our specific sites. Confusion matrices, the common method for describing classification accuracy (Lillesand et al., 2008), are used to compare the reference data and the corresponding classification

results on a category-by-category basis. User's accuracy, also referred to as reliability, measures the likelihood that a sample classified on the map accurately represents that class in reality, thereby assessing how well the map reflects the actual conditions on the ground (Story & Congalton, 1986). Another metric, referred to as producer's accuracy, shows the likelihood that a reference pixel is accurately categorized and evaluates how effectively a particular area is able to be categorized (Congalton, 1991). Overall accuracy gives a general measure of how well the classifier performs across all categories. These metrics are applied in many remote sensing image classification (Hamunyela et al., 2022; Pickens et al., 2020; Tulbure et al., 2016) and are convenient for the interpretation of the classification (Foody, 2002).

We calculate the user's accuracy, producer's accuracy, and overall accuracy to assess the results of the surface water detection and area estimation. The user's accuracy (UA) is calculated by dividing the number of correctly classified pixels in each category by the total number of pixels classified in that category (the row total). This is known as the specificity or true negative rate, and is the complement of the commission error (Lillesand et al., 2008). The producer's accuracy (PA) is obtained by dividing the number of correctly classified pixels in each category (on the main diagonal) by the number of training set pixels used for that category (the column sum). This is known as the sensitivity, or true positive rate, and the complement of the omission error. The overall accuracy (OA) is a measure of how well an algorithm performs compared to reference the training data. It is computed as the percentage of correctly classified pixels versus total pixels. tenfold cross-validation is used to evaluate the performance of the model (Fushiki, 2011).

3.2.5 Assessing surface water area estimates with the JRC dataset

GEE is utilized for data processing, analysis, and visualization, leveraging its cloud computing capabilities to retrieve the JRC dataset. The JRC dataset provides continuous surface water data over the globe, although there are data gaps from 1988 to 1993, 1996–1997 and 2021–2023 in our study area during our study period from 1986 to 2023. Hence, we identify and exclude such data gaps through visual inspection of each site to assess and compare our estimated surface water area with the Joint JRC Global Surface Water Data (GSWD). A linear regression model, along with its metrics of R^2, root mean square error (RMSE), mean absolute error (MAE), and RMSPE, is used to assess the agreement between the JRC dataset and our surface water area estimates.

3.2.6 Spatio-temporal surface water dynamics and trends

We construct a time series product of surface water extent using continuous Landsat imagery for 38 consecutive years from 1986 to 2023. We use the Mann–Kendall test to analyze the long-term monotonic (increasing or decreasing) trends in various hydroclimate time series (Mann, 1945). This test is robust and less sensitive to outliers, which enhances its applicability in the analysis of noisy data (Yue & Wang, 2004). We assess the statistical significance of trends using the p-value and evaluate the strength of the relationship between the time trend and changes in the surface water area using Kendall's Tau coefficient (Hamed, 2008; Yue et al., 2002). We also use Sen's slope estimator to estimate the extent of the change in surface water area over time (Sen, 1968). The Pettit test is applied to identify points of change or abrupt shifts in a time series of hydroclimatic datasets (Pettitt, 1979). The percentage change in surface water area is calculated by taking the difference between the post-change and pre-change areas, dividing that difference by the pre-change area, and then multiplying the result by 100 to express the change as a percentage. This provides a quantitative assessment of the relative changes in surface water area over time. We further calculate the water occurrence frequency (WOF) for each site by determining the percentage of time during the study period that each pixel is classified as water using Eq. (1) (Pekel et al., 2016; Yang et al., 2020). The WOF classes are used to distinguish permanent water bodies from those with seasonal or temporary surface water cover. In particular, areas with a WOF value of > 60% are classified as permanent water bodies, indicating that the presence of surface water is permanent (Tariq & Qin, 2023). Field observations and verification are undertaken to ensure the spatial distribution and pattern of surface water. Python programming in the Google Colab environment is used to implement all tests and calculate WOF as:

$$WOF\ (\%) = \left(\frac{\sum wi^t}{T}\right) \cdot 100 \tag{14}$$

where $\sum wi^t$ denotes the total count of water presence (pixels with a value of 1) for a specific pixel *i*, across all time steps *t*, and *T* denotes the total number of time steps (years) considered in the calculation.

3.3 Results

3.3.1 Machine learning for surface water detection

The remote sensing-based surface water detection of the Addis Ziway site using machine learning methods GTB, RF and SVM as well as results of the JRC data set depicts satisfactory classification accuracies of > 90% in terms of producer's accuracy, user's accuracy, and overall accuracy (Figure 3.2). The GTB achieves slightly better accuracies with an overall accuracy of 94–98% from 1986 to 2023. However, RF and SVM also have high overall accuracies of 93% to 97% and 92% to 96%, respectively. Thus, these results highlight the effectiveness of all three machine learning models for surface water detection.

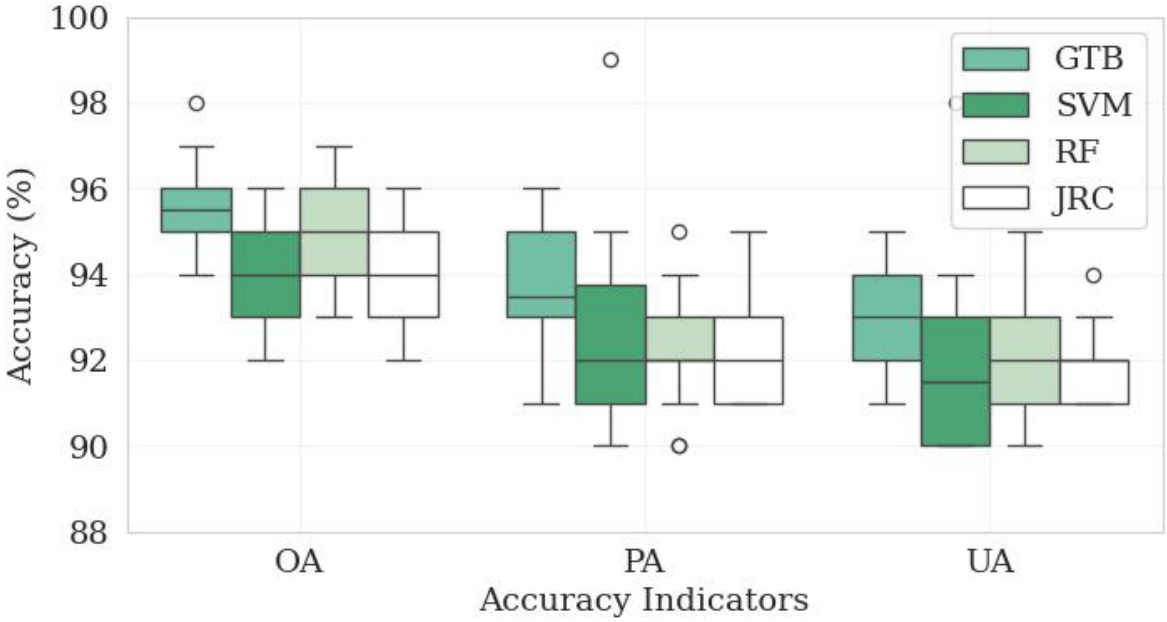


Figure 3.2 Accuracy assessment of surface water estimation using the machine learning methods random forest (RF), gradient tree boosting (GTB) and support vector machine (SVM) along with results obtained from the Joint Research Center (JRC) dataset in the Addis-Ziway site. OA overall accuracy, PA producer accuracy, UA user accuracy

Due to its best performance, the GTB model is selected for further estimation of the surface water area for all four sites (Figure 3. 3) and subsequent analysis. At all of the sites, the three accuracy indicators range between 91 and 98% and thus at a very high and comparable level. The highest overall agreement between training data and the machine learning methods is found for Addis-Ziway and the lowest values for Abaya-Chamo. We attribute this to the high image quality and low cloudiness at the Addis-Ziway site. In addition, the separation of water from non-water bodies at the Abaya-Chamo site was more

difficult, as water bodies often overlap with vegetation areas at this site. In general, the overall accuracies of the GTB method for all locations and all time periods is well over 90%.

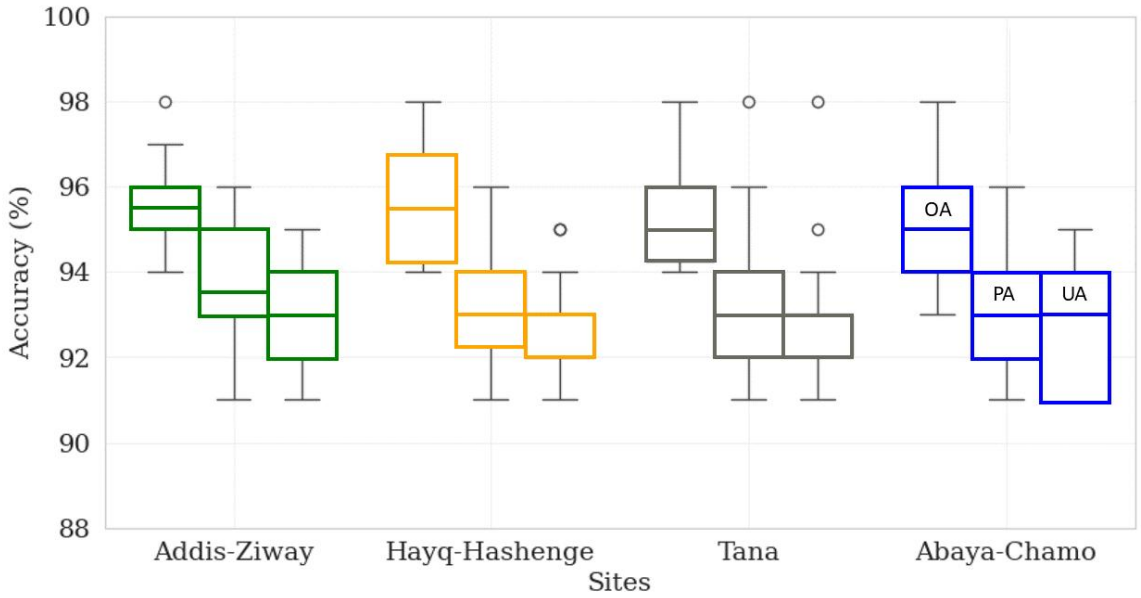
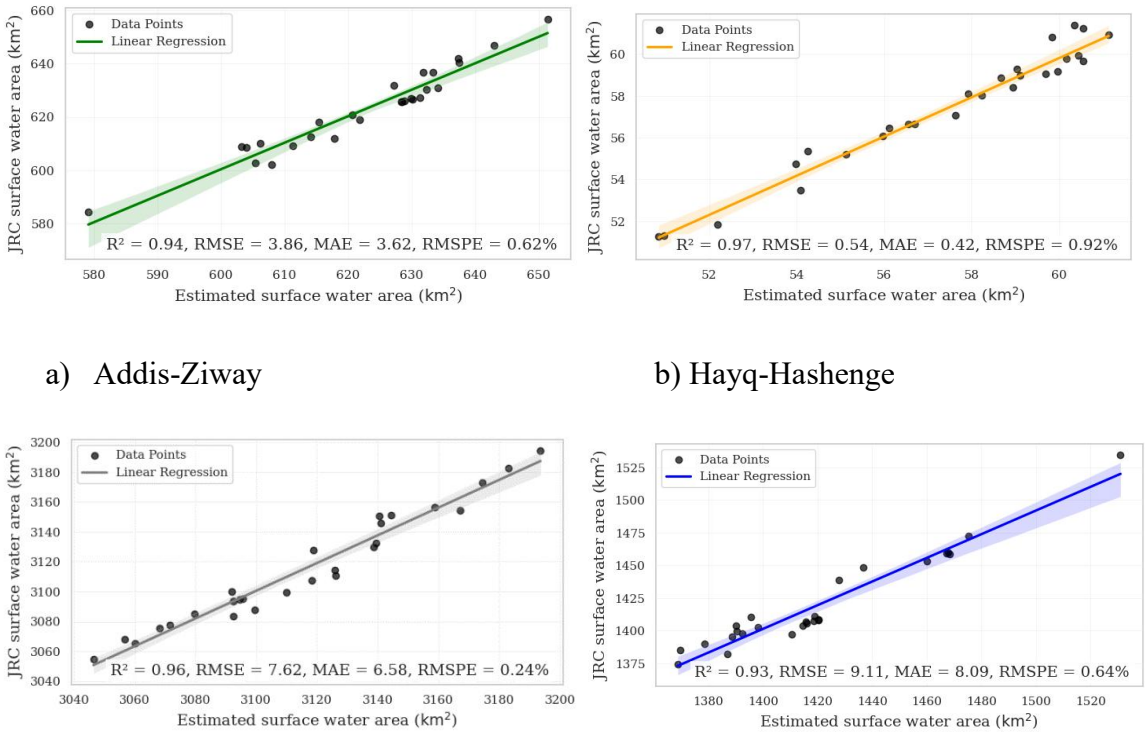


Figure 3.3 Accuracy assessment of the best performing machine learning method gradient tree boosting (GTB) at the four sites with overall accuracy (OA), producer accuracy (PA) and user accuracy (UA)

3.3.2 Assessing surface water area estimates with the JRC dataset

The comparison between the estimated surface water area and the JRC dataset reveals a good agreement and the reliability of the machine learning models. This is evident by R^2 values above 0.9 and RMSPE values below 1% for all sites (Figure 3. 4). The RMSE and MAE values range from 0.54 to 9.11 and 0.42 to 8.09, respectively. Hence, the absolute deviations between the machine learning estimated and JRC-derived surface water areas are low.



a) Addis-Ziway

b) Hayq-Hashenge

c) Tana

d) Abaya-Chamo

Figure 3.4 Correlation analyses and linear regression for comparison of the gradient tree boosting (GTB) estimated surface water area and the JRC dataset for the four sites

In addition to the statistical agreement, Figure 3.5 displays the time series for the estimated surface water area. As can be seen, the temporal development is well captured by the GTB algorithm. We do not see a general difference in the performance of GTB in small (Hayq-Hashenge), medium (Addis-Ziway) or large scale (Abaya-Chamo, Tana) test sites. But what we observe is a significant increase in surface water extent over the almost four decades for all locations, despite some major fluctuations between the years. For all four areas, we find an interim decrease until 2005 after an expansion of the water areas until the beginning of 2000. From 2005 on, the expansion of the water surface generally increases again. It is noticeable that there are subsequently increasing differences in the dynamics of the four areas, which in turn can be attributed to more regional trends in precipitation and evaporation. For example, in Abaya-Chamo there is no decrease or stagnation of the water surface area in 2009 as is the case elsewhere. The peak in 2020 in the same site is also striking. In Hayq-Hashenge, the dramatic increase of 9.51% from 2009 to 2010 is also more pronounced than in any other region. A Mann–Kendall test for monotony of trends is carried out to analyse the temporal trends further and to check for significant changes.

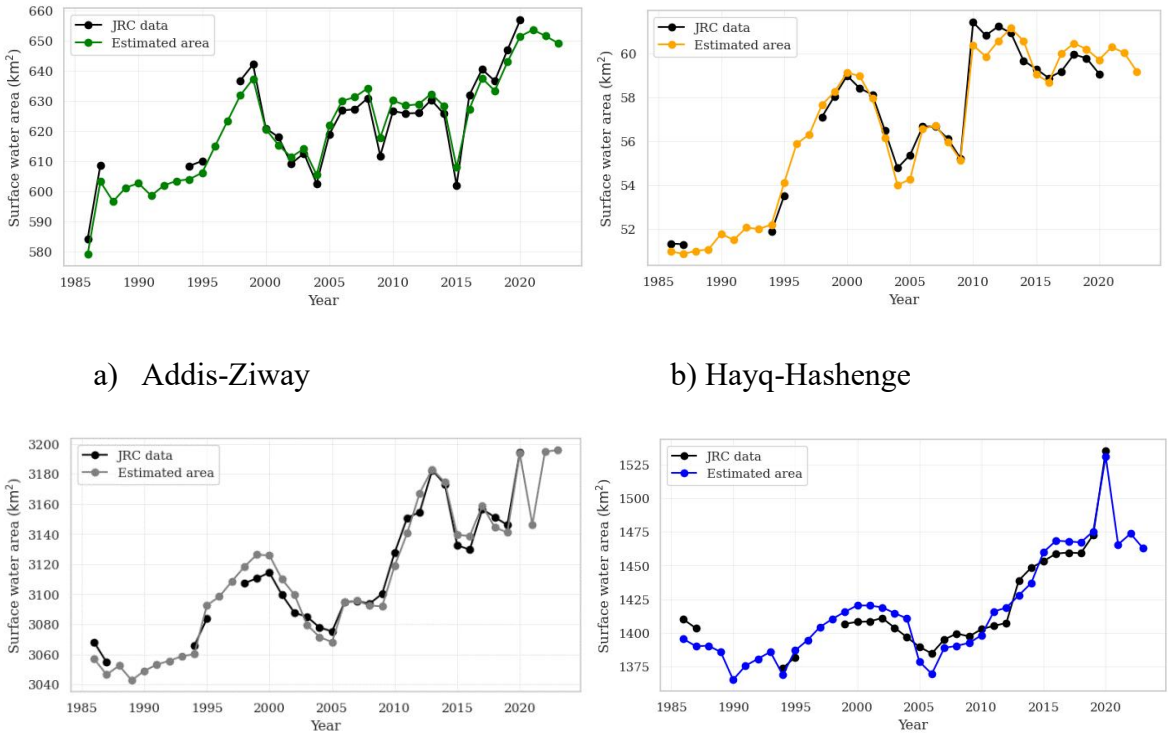
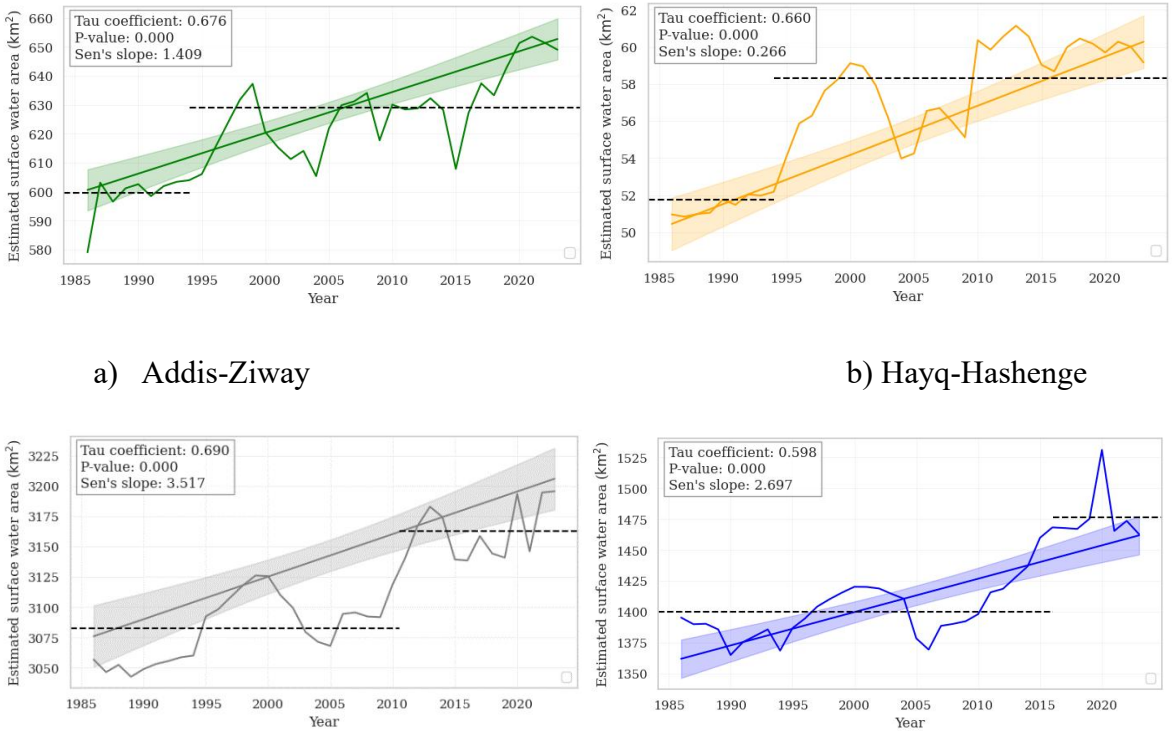


Figure 3.5 Comparison of time series of surface water area estimation based on the gradient tree boosting (GTB) method and the JRC dataset for all four sites. Note the data gaps of surface water extend in the JRC dataset from 1988-1993, 1996-1997 and 2021-2023

3.3.3 Spatio-temporal surface water dynamics and trends

The long-term surface water dynamics at the four study sites show an increasing trend from 1986 to 2023, with notable inter-annual variability (Figure 3.6). At Addis-Ziway, a significant increasing trend is observed with a Kendall Tau coefficient of 0.676 ($p = 0.000$) and a Sen's slope of 1.409 $km^2\ y^{-1}$. The average estimated surface water area increased by 5% after 2000 compared to before 2000 period. In the same way, the Hayq-Hashenge site shows a gradual increase with a rate of change of 0.266 $km^2\ y^{-1}$. For both sites, a change in the trend is apparent at 1995. In Hayq-Hashenge the surface water area increases within 6 years by around 13% up to 58.3 km^2. The average estimated surface water area shows an increase of about 2.6% after 2010 at the Tana site and a more significant increase of 5.4% after 2015 at Abaya-Chamo, highlighting a more pronounced expansion at the latter site. However, due to the size of the sites, changes before the turn of the millennium are not as pronounced as the changes from 2010 (Tana) and 2016 (Abaya-Chamo) onwards.

Therefore, these results underscore significant spatio-temporal variations in surface water dynamics with general increasing trends.



a) Addis-Ziway b) Hayq-Hashenge

c) Tana d) Abaya-Chamo

Figure 3.6 Analysis of long-term trends in surface water area. Tau fitted line with Kendall's Tau coefficient, Sen's slope, and p-values. The dotted black line shows the average surface water area before and after the year of change

3.3.4 Surface water distribution and pattern

For the assessment of water occurrence over time we calculate the water occurrence frequency WOF and classify it into five classes of no water (0%) to full water coverage (100%) during the 38 years of observations. For the majority, a WOF 61–100% contributes to the majority of the surface water distribution across all sites (Figure 3.7 and Table 3.1). The highest frequency is found for Tana with almost 94% followed by Abaya-Chamo with 91.6%. The two smaller sites have less stable surface area coverages with around 84% at WOF 61–100%. These sites show more dynamic pixels that often change between wet and dry conditions and which are less stable over time (WOF1-20% and 21–40%). As can be seen in Figure 7, the WOFs show significant spatial variation across all sites. Lower WOF1-60% are spatially distributed and occur predominantly along the edges of areas with higher WOF61-100%, displaying a scattered spatial pattern. Overall, the high frequency of water

occurrence across all sites represents the permanent water bodies, while the remaining proportions (1–60%) represent seasonal and temporary water bodies, including floodplains.

Table 3.1 Total (and percentage) surface water distribution by water occurrence frequency (WOF) classes from 1986 to 2023

	Surface water area (km^2) per WOF				
	Area [km²]	WOF1-20%	WOF21-40%	WOF41-60%	WOF61-100%
Hayk-Hashenge	63.8	5.0	3.1	2.1	53.6
Addis-Zeway	659.2	62.4	19.7	25.9	551.2
Abaya-Chamo	1525.0	83.4	23.7	21.6	1396.3
Tana	3218.4	125.7	46.9	30.6	3015.3
	Percent of surface water area (%) per WOF class				
	Area [km²]	WOF1-20%	WOF21-40%	WOF41-60%	WOF61-100%
Hayk-Hashenge	63.8	7.8	4.9	3.3	84.0
Addis-Zeway	659.2	9.5	3.0	3.9	83.6
Abaya-Chamo	1525.0	5.5	1.6	1.4	91.6
Tana	3218.4	3.9	1.5	1.0	93.7

3.4 Discussion

3.4.1 Machine learning for surface water monitoring

This study demonstrates the effectiveness of machine learning algorithms (GTB, RF, SVM) in surface water detection and monitoring. It is observed that all methods have a satisfactory classification accuracy of > 90% in terms of overall accuracy, producer's accuracy and users's accuracy. The robustness of the GTB is slightly higher, as evidenced by the overall accuracy rates of 94% to 98% across all sites. The effectiveness of GTB is probably due to its ensemble learning approach, which builds decision trees sequentially, each correcting the errors of the previous ones. This approach is particularly useful for detecting complex spectral patterns associated with water bodies. For example, Hastie et al., 2009 found that gradient boosting methods often outperformed RF in image classification. GTB achieved an average overall accuracy of 93.3% for mapping open and vegetated water bodies in the United States from 2017 to 2021(Vanderhoof et al., 2023). However, we found that RF also

has a commendable overall accuracy of over 90%. Similarly, an overall accuracy of 99% and user's and producer's accuracies of 96% and 87%, respectively, were observed for RF in the Murray-Darling Basin using Landsat data in Australia from 1986 to 2011(Tulbure et al., 2016). Furthermore, high user's (91.5 ± 2.5%) and producer's (91.1 ± 6%) accuracies were achieved in the Cuvelai-Etosha Basin in Southern Africa from 1990 to 2021 (Hamunyela et al., 2022). Thus, machine learning methods are effective for surface water detection, provide valuable spatio-temporal information for surface water area estimation, and offer excellent tools for long-term monitoring of different geographical locations in view of SDG 6.

3.4.2 Comparison of surface water area estimates with the JRC dataset

Our estimated surface water area depicts a good agreement with the JRC dataset within the same aligned years, similar to other studies (Getaneh et al., 2024). Despite this commonality and its good accuracy values, the JRC dataset has certain limitations at the local scale that make it less suitable to fully capture the surface water extent and interannual variability in our study areas from 1986 to 2023. This is due to the fact that the JRC method utilizes globally tuned models (Hamunyela et al., 2022). The GTB model, as confirmed by its high accuracy across all study sites, can effectively compensate for the data gaps in the JRC dataset. By providing a consistent and credible dataset, the model can support detailed analyses of interannual variability, trends, and drivers of surface water dynamics. This is crucial for understanding the implications of climate change, human activities, and natural variability on water resources. In conclusion, the GTB model effectively overcomes the limitations of the JRC dataset and provides robust and consistent spatiotemporal information for surface water monitoring across different geographic locations and time periods.

The strong agreement between our estimated surface water areas and the JRC dataset is evidenced by a high coefficient of determination ($R^2 > 0.9$) and low RMSPE < 1% across all study sites. This underscores the reliability of the estimated surface water areas, and confirms the applicability of machine learning methods as was previously reported by RF and the detection lake areas of Abiyata, Chamo, Ziway, Hawassa, and Abaya. Overall, we consider the JRC dataset to be a very useful reference for testing and developing alternative ways of estimating the spatio-temporal dynamics of surface water areas in data-poor regions. Despite the occasional missing values in the JRC dataset, the temporal patterns of our estimated surface water area and the JRC datasets remain consistent. This consistency

suggests that our machine learning models can effectively compensate for data gaps in the JRC dataset and provide a consistent dataset of surface water monitoring for different geographical locations over time. Therefore, our approach is applicable and operational for different geographical regions and long-term monitoring of surface water in support of water resource management and conservation using publicly available data sources of GEE.

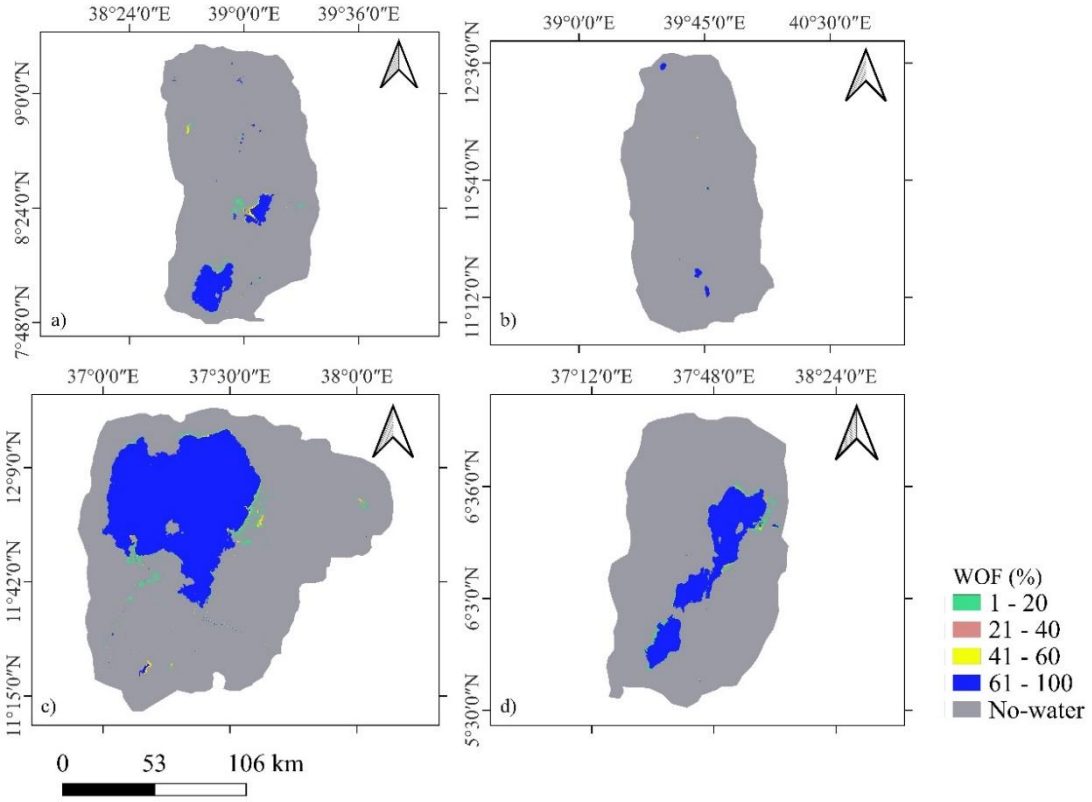


Figure 3.7 The Water Occurrence Frequency (WOF) from 1986 to 2023 in the four sites of (a) Addis-Zeway, (b) Hayk-Hashenge, (c) Tana and (d) Abaya-Chamo

3.4.3 Surface water dynamics and trends

The surface water dynamics of our study sites show a long-term increasing trend from 1986 to 2023, characterized by considerable inter-annual variability. The expansion of surface water at the Addis-Ziway site is influenced by anthropogenic factors, such as the Koka reservoir and associated dam infrastructure. In addition, the surface area of several lakes in the region, including Hora and Bishoftu, showed an increasing trend between 1984 and 2017 (Sathianarayanan, 2018). Lake Ziway also showed a substantial long-term increase in surface area between 1985 and 2023 (Getaneh et al., 2024), although other studies have

reported a consistent trend over earlier periods (Belete et al., 2015; Getnet et al., 2014). At the Hayq-Hashenge site, the surface water showed a gradual increase with a Sen's slope of 0.266 km^2 yr^{-1}, with a notable shift in 1995 attributed to the construction of a dam near the town of Hara. However, the coverage of Lakes Hayq and Hardibo showed a declining trend (Demlie et al., 2007; Getaneh et al., 2024). For the Tana site, the average surface water area increased, with a notable increase after 2010. Prior studies have identified trends in the 1980s and early 1990s that show steady increases, preceded by significant variation about every five years (Getaneh et al., 2024). At the Abaya-Chamo site, the statistically significant increasing trend was also observed by (Mulu et al., 2024), who reported surface water expansion between 2000 and 2022. In this study, we found that the construction of small and large dams and reservoirs across the study sites has contributed to an increase in surface water extent, along with the expansion of some natural lakes.

Dynamic changes in surface water extent is typical for many parts of the world. The permanent surface water extent in North America, for example, has increased by about 17,000 km^2 from 1984 to 2015 (Pekel et al., 2016), highlighting significant hydrological changes across the continent. Similarly, in Turkey, a long-term trend of increasing surface water extent has been observed due to the construction of new reservoirs and the development of surface water bodies (Donchyts et al., 2022). Thus, dams and reservoirs have been identified as the main drivers of surface water increase worldwide (Pekel et al., 2016). In addition, surface water dynamics show considerable spatio-temporal variability in certain regions. For example, at the subcontinental scale in the semi-arid regions of Australia, high inter- and intra annual variability in surface water extent has been observed between 1986 and 2011 (Tulbure et al., 2016). Similarly, annual sample based estimates in the Cuvelai-Etosha Basin (CEB) of southern Africa show significant interannual fluctuations in surface water extent ranging from as low as 520.8 ± 375.7 km^2 to as high as 12,372.3 ± 1,154.7 km^2 between 1990 and 2021 (Hamunyela et al., 2022). These spatio-temporal variability of surface water plays a key role in shaping water availability in many regions.

The results of the Mann–Kendall trend analysis do not show a general consistent temporal pattern in surface water dynamics across the four study sites. This indicates that local site characteristics, water management practices, and anthropogenic impacts play a significant role in determining surface water trends, potentially superimposing the broader impact of climate change. This underscores the importance in considering local factors influencing

hydrological systems, especially in regions with diverse climatic and anthropogenic influences (Kazemi Garajeh et al., 2024; Tulbure & Broich, 2019). A one-size-fits-all method of water resource management may not be adequate to address the specific challenges of each study site. Instead, tailored site-specific water management strategies that take into account local hydrological, climatic, and socioeconomic conditions are needed to ensure sustainable water resource management (Ramsar Convention on Wetlands, 2018).

The predominance of high-frequency water occurrence across all study sites indicates the stability and persistence of surface water in these areas, reflecting the presence of permanent water bodies such as lakes, rivers, and reservoirs. These water bodies are maintained by consistent hydrological inputs, including upstream inflows and groundwater recharge (Downing et al., 2006; Lehner & Döll, 2004). Such areas are particularly crucial to the impacts of climate change and also vulnerable to human activities such as water withdrawals, land use changes, and dam construction, which can alter their spatial extent and hydrological regimes.

Surface waters with low to moderate water occurrence frequencies contribute less to the overall surface water distribution. These areas are characterized by the intermittent or seasonal presence of water, often associated with temporary water bodies, floodplains or wetlands. Such water bodies are highly dynamic and sensitive to both, climatic variability (e.g., droughts, floods) and anthropogenic activities. In particular, these areas are often affected by land use changes, such as the draining of wetlands for agriculture, which is particularly evident at the Hayq-Hashenge and Abaya-Chamo sites. Therefore, the surface water distributions and patterns associated with water frequency highlight the need for site-specific water management and conservation strategies that align with SDG6 monitoring efforts at the regional level.

3.5 Conclusion

In this study, we provide and quantify spatio-temporal information for monitoring surface water extent using machine learning at four sites in Ethiopia from 1986 to 2023. GEE allows for the processing time series of surface water dynamics and monitoring using Landsat imagery with machine learning. It therefore has a great potential for monitoring surface water expansion also in remote areas. Our results confirm that machine learning using Landsat data produces reliable results for surface water monitoring and provides spatio-temporal information to support surface water management and water policy in Ethiopia.

The results of this study may have important implications for policy makers and water resource planners in Ethiopia, particularly in the context of sustainable water management and achieving SDG 6 at the regional level. The observed long-term increasing trend in surface water extent, together with high inter-annual variability, underscores the need for adaptive water management strategies. Policy makers should prioritize the development of Integrated Water Resource Management (IWRM) strategies, as local site characteristics, and anthropogenic influences affect surface water dynamics and trends. In addition, Ethiopia's water policy should incorporate real time surface water monitoring using machine learning-based remote sensing technologies and decision making to promote water conservation.

However, any approach to the dynamics of surface water extent that relies solely on remotely sensed imagery and historical data sets has difficulties in making fully-fledged future projections of surface water extent, as imagery and datasets are captured and collected from the past. Thus, for future projections, dynamic variables that change over time could complement the future spatio-temporal information on surface water extent. For example, hydro-meteorological variables from weather and climate projections or land use features from land use and land cover projections should enable such projections into the future. Therefore, future research needs to undertake large-scale assessments to provide a comprehensive national-level analysis of surface water dynamics and trends, taking into account the impacts of climate and land use change. This could be achieved by establishing an automated and operational system using GEE to provide near real-time spatio-temporal information on surface water extent to decision makers.

4 Influence of climate, land use, and land cover on surface water dynamics in the Ethiopian highlands

This chapter is submitted to *Remote Sensing Applications: Society and Environment* as:

Tesfaye, M., & Breuer, L. (2025b). Influence of climate, land use, and land cover on surface water dynamics in the Ethiopian highlands. Manuscript number: RSASE-D-25-01440.

Abstract
Analyzing the influence of climate drivers and change in land use and land cover (LULC) on surface water dynamics is crucial for the sustainable management and conservation of water. Despite its importance, there is a lack of comprehensive assessments particularly for low-income economies in the Global South. Our study evaluates extreme gradient boosting (XGBoost), artificial neural network (ANN), random forest (RF), support vector regression (SVR), and multi-linear regression (MLR) models to detecting changes in surface water dynamics in the highlands of Ethiopia from 2000 to 2020. XGBoost performs best in terms of the root mean square error (RMSE), Willmott's agreement index (dr), mean bias error (MBE), and mean percentage error (MPE), and is the most interpretable of all the models. We show that precipitation and to a lesser extent minimum temperatures and vegetation cover are the main drivers of surface water dynamics during the study period. To further clarify the role of vegetation change, we are investigating decadal LULC changes. It appears that a large proportion of the water areas are stable, with changes occurring in the peripheral areas in terms of grassland, shrubland and cropland. The results of our study highlight the necessity of integrated water policy and management approaches to address the intricate interplay between climate, LULC, and water resource in Ethiopia.

Keywords Water resources, climate, LULC, remote sensing, Google Earth Engine, machine learning

4.1 Introduction

Surface water is a crucial natural resource on our planet, playing a key role in the climate and hydrological system (Chahine, 1992; Tranvik et al., 2009). In addition, it preserves and sustains biodiversity and contributes to the provision of vital and diversifying ecosystem services (Brauman, 2015; Dudgeon et al., 2006; Zedler and Kercher, 2005). However, surface water resources are subject to significant spatiotemporal dynamics driven by short-term weather changes or changes in climate, land use and land cover (LULC) (Gaines et al., 2022; Tulbure and Broich, 2019; Vörösmarty et al., 2000). These changes potentially have the capacity to exacerbate water scarcity and intensify tensions, for instance between agricultural and urban areas (Flörke et al., 2018). Surface water resources are affected by climate variability through variations in temperature, rainfall, and the rate of evapotranspiration. These changes impact streamflow and increase the severity of hydrological disasters, including droughts and severe flooding (Huntington, 2006; Kundzewicz et al., 2008). Likewise, changes in LULC driven by urbanization, agriculture, and deforestation/reforestation also put substantial pressure on surface water resources. LULC conversion exacerbates surface runoff, infiltration rates, and sedimentation, leading to adverse impacts on water quality and heightened hydrological variability (Foley et al., 2005). The interaction between climate and LULC changes exacerbates the susceptibility of surface water resources.

Moreover, associated with climate change are hydrological extremes, for example floods and droughts, resulting in rising loss of life in Ethiopia. The lowlands within the Awash River basin in Ethiopia, for example, experience frequent flooding (Legass et al., 2025), which has led to crop failure, nutrient loss from the soil, and soil degradation (Desta et al., 2024). Droughts have also been observed several times in the past in Ethiopia (Haile et al., 2022). Overall, Ethiopia's water-related ecosystems encounter various challenges, including interannual climate variability and shifts in water distribution (Seyoum et al., 2015). Therefore, it is crucial to improve our understanding of the factors that influence water bodies in order to make decisions and ensure sustainable water resources management in Ethiopia.

Cloud-based remote sensing has transformed the processing of earth observation data by providing a scalable, high-performance computing platform without the need for powerful local computing resources. Google Earth Engine (GEE) allows users to process and analyze vast amounts of remote sensing and geospatial data in a cloud computing environment

(Gorelick et al., 2017). The capability to offer near real-time monitoring is a key benefit of cloud-based remote sensing. This capability is crucial for early disaster response and water resource management. Near-real-time National Aeronautics and Space Administration (NASA) imagery, for instance, provides rapid updates to track flood inundation dynamics and enables the effective allocation of emergency resources (Ahamed and Bolten, 2017). The Landsat archive, in particular, is now a vital source of earth observation data for numerous applications. These include monitoring long-term surface water dynamics, trend analysis, flood detection, and drought monitoring and enable timely, data-driven decision-making (Wulder et al., 2016). GEE is capable of carrying out time-series analysis, machine learning analysis , and real-time monitoring of LULC, climate, and hydrology at global to regional scales (Amani et al., 2020; Tamiminia et al., 2020).

The impact of climate and LULC variables on surface water dynamics has been investigated by numerous studies, with machine learning and statistical analysis. For example, a linear mixed-effects model was utilised for the analysis of the impact of climate and land use on the dynamics of surface water extent in a semi-arid region of Australia, exhibiting that climate influences more in the northwestern Australian Murray-Darling Basin (MDB), while land use effect primarily in the central MDB (Tulbure and Broich, 2019). In the same way, the influences of climate and human activities on surface water resources in the United States from 2000 to 2018 have been examined with a linear mixed-effects model, random forest (RF) and mixed-effects RF models, depicting that anthropogenic drivers have a higher impact than climate drivers (Gaines et al., 2022). Moreover, the impact of drivers on surface water dynamics was conducted using statistical analysis, including correlation analysis, depicting that vegetation and human activities related to dam and reservoir construction are vital drivers (Zeng et al., 2020), and vegetation change also impacts the surface water bodies (Mulu et al., 2024).

Despite the advancements of existing studies, a comprehensive assessment of multi-modeling approaches, including the XGBoost, artificial neural networks (ANN), support vector regression (SVR), RF, and multiple linear regression (MLR) for model impact assessment of climate and LULC drivers on surface water dynamics lacking. Our study not only follows such a multi-model approach, but also provides a detailed analysis of the spatio-temporal LULC dynamics in the context of its contribution to surface water dynamics. Particularly in Ethiopia, to the best of our knowledge, such data-driven models for analyzing the effects of climate and LULC drivers on surface water dynamics remain

unexplored. Therefore, this study aims to (i) evaluate five machine learning and statistical approaches (i.e., XGBoost, ANN, SVR, RF, and MLR) and identify the most effective one for surface water detection; (ii) assess the effect of climate variables on surface water dynamics; and (iii) investigate the spatiotemporal LULC dynamics in the highlands of Ethiopia.

4.2 Materials and Methods

4.2.1 Descriptions of study area

Ethiopia has ample surface water resources with 12 major river basins (Figure 4.1a). Of these, the Abay (Blue Nile), Baro-Akobo, Tekeze, and Mereb Gash river basins, account for 33% of Ethiopia's surface area and drain the northern, central, and western parts of the country. About 90% of Ethiopia's water sources are found in the Abay (Blue Nile), Tekeze, Baro-Akobo and Omo-Gibe river basins (Smola and Schölkopf, 2004). In eastern Ethiopia, the availability of surface water resources is rather scarce, as perennial rivers are almost lacking below 1,500 m amsl. In addition, the Aysha, Dinakle and Ogaden basins are mostly dry and do not permanently drain (Berhanu et al., 2014).

The study area is located on the peripheries of the highlands of eastern Ethiopia and the western cliff of the Afar lowlands in the Great Rift Valley region. It is characterized by uneven terrain with elevations ranging from 822 m to 3,879 m amsl (Figure 4.1b). Hardibo and Hayq are highland lakes found in our study area with a closed drainage system. From 2000 to 2020, the mean daily lowest and highest temperatures are 13.6°C and 25.0°C, respectively (Table 4.1). The mean annual rainfall is 876 mm, with a minimum of 576 mm and a maximum of 1073 mm. The actual evapotranspiration ranges from a minimum of 280 mm to a maximum of 464 mm. On average, the vegetation cover is around 42.5% from 2000 to 2020.

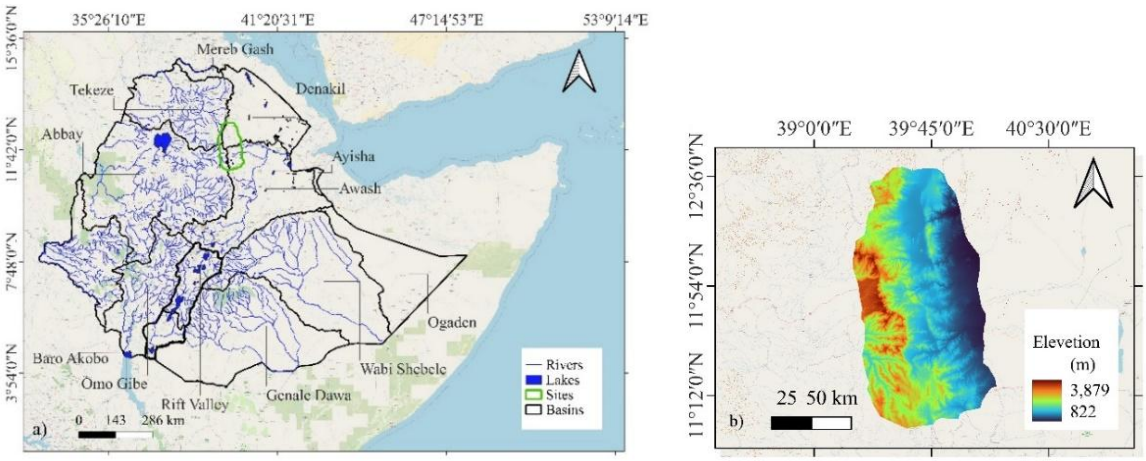


Figure 4.1 Location of the study area with the main water bodies of Ethiopia

Table 4.1 Meteorological characteristics and vegetation cover of the study area. SD=standard deviation

Variable	Mean	SD	Min	25th perc	Median	75th perc	Max	Source
Precipitation (mm)	876.1	112.2	575.6	794.0	889.1	959.2	1073.2	CHIRPS-v2
Min. temperature (oC)	13.6	0.4	12.8	13.3	13.6	13.9	14.3	CHIRTS-ERA5
Max. temperature (oC)	25.0	0.5	23.5	24.8	25.1	25.4	26.1	CHIRTS-ERA5
Actual evapotranspiration (mm)	359.8	44.4	280.1	332.9	358.1	388.4	463.9	NASA MOD16A2GF V6.1
Vegetation cover (%)	42.5	9.9	23.8	36.6	42.3	50.1	57.9	NASA MOD13Q1 V6.1

4.2.2 Data acquisition and preprocessing

We utilize climate and remote sensing data from 2000 to 2020 provided within the GEE to analyze the effects of climate and LULC drivers on surface water dynamics. Further remote sensing data are obtained from Tesfaye and Breuer (2025). We gathered the rainfall from the Climate Hazards Group Infrared Precipitation with Station data (CHIRPS-v2) from GEE at a spatial resolution of 0.05° (Funk et al., 2015). This dataset is selected due to the high performance observed in our region (Ayehu et al., 2017; Gashaw et al., 2023). Minimum and maximum temperature data were obtained from the Climate Hazards Center Infrared Temperature with Stations version 5 of the European Centre for Medium-Range Weather Forecasts Re-Analysis (CHIRTS-ERA5) (https://data.chc.ucsb.edu/experimental/CHIRTS-ERA5/). This dataset is based on the ERA5 reanalysis, integrating high-resolution (0.05°) CHIRTS temperature data (Verdin et al., 2020) and is appropriate for our region (Mamuye et al., 2024; Steinkopf and Engelbrecht, 2022). We access Terra vegetation index products (MOD13Q1 V6.1) and actual evapotranspiration (MOD16A2GF V6.1) of the Moderate

Resolution Imaging Spectroradiometer (MODIS) dataset using GEE from the National Aeronautics and Space Agency (NASA) at 250 m and 500 m spatial resolution, respectively. This vegetation dataset is used due to its high temporal resolution and is effective for modeling temporal dynamics of vegetation (Maynard et al., 2016). The actual evapotranspiration dataset also has a high temporal resolution (8 days), is available since 2000, and incorporates inputs of daily meteorological reanalysis data.

We obtain Tier 1 Landsat images from the United States Geological Survey (USGS) for 2000, 2010, and 2020 to analyze LULC dynamics. Landsat imagery is chosen because it is consistently accessible at a spatial resolution of 30 m and 16-day temporal resolution. For the years 2000 and 2010, Landsat 7 equipped with Enhanced Thematic Mapper Plus (ETM+) sensors was employed. In 2020, data from Landsat 8's Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) were also used. Surface reflectance for Landsat 7 was generated using the standard Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al., 2006; Schmidt et al., 2013), while the Landsat Surface Reflectance Code (LaSRC) was applied to derive surface reflectance for Landsat 8 (Vermote et al., 2016).We use the GEE median reducer function to merge several Landsat images into a single annual composite (Zhang et al., 2021). The preprocessed data includes blue, green, red, near-infrared, shortwave infrared 1, and shortwave infrared 2 bands. The CFmask algorithm (Zhu and Woodcock, 2012) is applied to identify and eliminate pixels containing clouds and shadows from the image using the QA_PIXEL band, thereby improving data quality.

4.2.3 Model descriptions

We evaluate five machine learning and statistical models, XGBoost, ANN, SVR, RF, and MLR, to assess the impact of drivers on surface water dynamics. We calculate Shapely Additive exPlanation (SHAP) values to analyze the drivers' importance and relationship for each model (Lundberg and Lee, 2017). The SHAP algorithm uses statistical analysis and its values quantify the influence of the drivers in the models. The SHAP feature importance plot indicates the contribution of the drivers to the model output, and the summary plot provides the distribution of the drivers' impact on the output and their relationship. The RMSE is utilized as a standard statistical evaluation metric of model effectiveness, often utilized in climate studies (Chai and Draxler, 2014). The Willmott index dr measures model accuracy (Willmott et al., 2012). The mean bias error (MBE) and mean percentage error

(MPE) evaluate the bias of a model in overestimating or underestimating values. The GEE Python application programming interface (API) is utilised to carry out the analysis and validation of the models within the Google Colab platform. Each of the five models is briefly described below:

1) XGBoost is a scalable tree boosting system that is widely used to solve real-world problems due to its effective scalability (Chen and Guestrin, 2016). It uses vital features and algorithmic optimizations. A novel tree-learning algorithm is designed to also apply it to constrain the data. The use of weights in predicted tree-learning is enabled by a validated weighted sketch technique. The combination of parallel and dispersed processing enables quicker learning and consequently more rapid model assessment. We optimize the XGBoost model and reduce overfitting by tuning its hyperparameters with a grid search and 5-fold cross-validation. We evaluate the decision trees encompassing 100, 300, and 500, with a maximum depth of 3, 6, and 9, a learning rate of 0.01, 0.1, and 0.2, and subsampling ratios of 0.8 and 1.0 for data and features, as well as sample bytes of 0.8 and 1.0, and a gamma of 0, 0.1, and 0.2. The tuning process resulted in an optimal hyperparameter setup comprising 500 for decision trees, 6 for maximum depth, 0.01 for the learning rate, 0.8 for subsample ratio, 0.8 for sample bite, and 0 for gamma value. For a fixed form q(x) in Ij = {i|q(xi) = j} the optimum weight Wj* of leaf j is calculated by eq. (1), the respective optimum values by eq. (2) (Chen and Guestrin, 2016).

$$Wj* = \frac{\sum_{i\in 1j} gi}{\sum_{i\in 1j} hi+\lambda} \tag{1}$$

$$L(t)\ (q) = -\frac{1}{2}\sum_{j=1}^{T} \frac{(\sum_{i\in Ij} gi)^2}{\sum_{i\in 1j} hi+\lambda} + \gamma T \tag{2}$$

XGBoost utilizes the exact greedy algorithm to obtain the optimum split as defined in eq. (3) (Chen and Guestrin, 2016). IL and IR are the left and right sets of nodes resulting from the partition, with I = IL ∪ IR. The loss reduction is used to find the split and is given by the following equation.

$$Lsplit = \frac{1}{2}\left[\frac{(\sum_{i\in IL} gi)^2}{\sum_{i\in 1j} hi+\lambda} + \frac{(\sum_{i\in IR} gi)^2}{\sum_{i\in 1j} hi+\lambda} - \frac{(\sum_{i\in I} gi)^2}{\sum_{i\in 1} hi+\lambda}\right] - \gamma \tag{3}$$

where gi represents the gradient for the ith data point, hi is the second-order gradient of the loss function for instance I, λ is a regularization parameter, T is the total number of nodes or

leaves in the decision tree, and γ is the parameter that regulates the magnitude of the model's complexity.

2) ANN is a computational model based on the structure of the human brain. It comprises interlinked artificial neural nodes that learn to recognise patterns through training. An ANN typically consists of input, hidden, and output layers. The input accepts incoming data and forwards it to the subsequent layer, with individual neurons in the layer aligning with specific attributes in the dataset. A hidden layer contains neurons that transfer the input data to the output layer and perform computations using weighted connections. ANNs analyze complex data patterns through forward propagation, loss optimization, as well as back propagation, and go through multiple training to improve accuracy. Parameters influence the accuracy, robustness and effectiveness of the model (Asiltürk and Çunkaş, 2011). We undertake hyperparameter tuning using a grid search with 5-fold cross-validation to enhance the accuracy. We assess the set of the learning rate (0.001, 0.01, 0.1), neurons (32, 64, 128), batch size (8, 16, 32), and epochs (50, 100) and obtain a hyperparameter set with 0.01 learning rate, 64 neurons, 8 batch size and 100 epochs. Overall, an ANN can be mathematically expressed according to eq. (4).

$$Yj = f(\sum_{i=1}^{n} w_{ij} x_i + b_j) \quad (4)$$

where Yj is the output of neuron j, xi is the i-th input feature, wij is the weight connecting the i-th input neuron to the j-th neuron, and bj is the bias term for neuron j.

3) SVR is a supervised learning algorithm that extends support vector machines (SVM) for regression purposes (Smola & Schölkopf, 2004). SVR determines a function that describes the relationship between the input features and a continuous target variable. SVR works by doing a nonlinear mapping of the input data to a higher-dimensional feature space and performing a linear regression process in the feature space. We perform hyperparameter tuning to improve model accuracy using kernel, C, and gamma parameters. We evaluate the combinations using a grid search with 5-fold cross-validation of kernel (linear and radial basis function), C (0.1, 1, 10, 20), and gamma (scale, auto, 0.01, 0.1, 1). We then obtain the final hyperparameter values of the rbf kernel, and 0.1 and 0.01 for the C and gamma values,

respectively. The SVR function is expressed in eq. (5) in a series of input features $\{xi, yi\}_i^n$, where yi denotes the target value, n is the data size, and xi is input vector.

$$f(x) = w \cdot \phi(xi\ \ ) + b \tag{5}$$

where w is a weight vector, ϕ(xi) represents a nonlinear map that shows the data vectors to a high-dimensional feature space and b is the bias.

The regression function can be finally expressed in eq. (6) after minimizing the regularized risk function, Ɛ-insensitive loss function and Lagrangian function that can solve the constrained optimization problem.

$$f(x) = \sum_{i=1}^{n} (\partial_{i^*} - \partial\ _{i^*}) K\left(x_i, x_j\right) + b \tag{6}$$

where $K\left(x_i, x_j\right)$ is the kernel function and $\partial\ and\ \partial\ _{i^*}$ are the Lagrangian factors.

4) RF is an ensemble learning model, which involves integrating various decision trees to improve accuracy and minimise overfitting. It builds the decision trees in concurrent methods that are fully trained, and each tree is utilized to predict the missing out-of-bag occurrences of an observation in a bootstrap sample (Breiman, 2001). RF works by creating a number of decision trees during the training process and making a classification decision based on the majority vote of the different trees. We conduct an optimisation of the hyperparameters using a grid search and 5-fold cross-validation to improve accuracy. The hyperparameters are a combination of the decision tree values of 100, 200 and 300, the maximum depth values of none, 10, 30, 50 and 100, the minimum sample split values of 2, 5 and 10, the minimum sample leaf values of 1, 2 and 4 and the maximum features values of auto, sqrt and log2. The optimal hyperparameter values, as determined by the algorithms, correspond to the setup of 100 decision trees with a maximum depth of 50, a minimum sample split of 10 and a minimum sample leaf of 2, as well as a $\log_2$ value for the maximum features.

RF is a classifier based on decision trees {h(x,Θk), k=1,…}, where the {Θk} are random vectors that are distributed independently and identically, and x is an input data (Breiman, 2001; Hastie, Tibshirani and Friedman, 2008). RF can be mathematically expressed as a regression to make a prediction at a new point x as given in eq. (7) (Hastie, Tibshirani and Friedman, 2008).

$$f_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} Tb(X) \quad (7)$$

with build tree Tb from the bootstrapped data for b=1 to B.

5) MLR is a widely applied statistical technique for carrying out predictions. It is used to model the association between a dependent and multiple independent factors. MLR offers interpretable coefficients for prediction and inference, though it requires checking for multicollinearity. It can be represented mathematically as given in eq. (8).

$$Y = \beta 0 + \beta 1 \quad X1 + \beta 2 \quad X2 + \ldots + \beta n \quad Xn + \epsilon \quad (8)$$

where Y is the dependent (target) variable, X1 , X2 ,..., Xn are the independent factors, β_0 is the constant, and β_1, β_2, ..., βn are the slope coefficients. The error term is denoted by ε.

4.2.4 LULC classification and change detection

We collect reference sample points from satellite imagery and previous reference maps using visual interpretation for training and validation of each LULC type in the study area (Table 4.2). A total of n=5,754 sample points are collected for the years 2000 (n=1,864), 2010 (n=1,908) and 2020 (n=1,982), respectively. Again, we leverage the GEE Python API to process the Landsat data in Google Colab. We employ a gradient boosted model to classify LULC for the years 2000, 2010, and 2020. The gradient boosting approach involves constructing an ensemble of weak learners to create a more accurate model. This is achieved by minimizing the loss function using gradient descent (Friedman, 2002). It has shown improved performance in image classification (Hastie et al., 2009). The amount of decision trees changes based on the degree of model complexity. A larger number of decision trees leads to a more correct, albeit more computationally intensive model. The shrinkage parameter minimizes overfitting by controlling the influence of each tree on the model, while the maximum node level restricts how complicated the patterns are. This collection of variables is intended to strike a delicate balance between model accuracy and computational effectiveness.

Hyperparameter tuning is carried out for the gradient boosting classifier with a grid search and 5-fold cross-validation. The tuning involves assessing combinations of trees set to 100, 300, and 500, learning rates set to 0.01, 0.05 and 0.2, sampling rates set to 0.6, 0.7 and 1.0, maximum leaf nodes set to 5, 10 and 20 and loss functions set to least squares and least

absolute deviation. The results exhibit that the optimum hyperparameter set is 500 for decision trees, 0.05 for learning rate, 1.0 for sampling rate, 20 for maximum leaf nodes and the least squares loss function.

Table 4.2 Description of LULC types

LULC type	Descriptions
Water	Lakes, rivers, streams, intermittent ponds, wetlands, reservoirs
Cropland	Land used for crop production, including land under preparation
Grassland	Grass, interspersed with small shrubs, including and seasonally grassed areas
Shrubland	Shrubs, bushes
Built-up	Buildings, roads, other infrastructure
Forestland	Natural forests, reforestation

Accuracy assessments are undertaken using the 5,754 reference sample points to evaluate whether or not the results of LULC classification are acceptable. We apply producer, user and overall accuracies because they are suitable for interpreting the LULC classification (Foody, 2002). User accuracy (UA) measures the reliability of a classified sample on a map in representing the actual LULC type, and thus how accurate the map is in representing the actual situation in the field (Story and Congalton, 1986). The UA is attained by taking the ratio of the correctly classified pixels in each class to the sum of all the pixels in that category (the raw sum) (Lillesand et al., 2008). The other metric, producer's accuracy (PA), determines the likelihood of an accurate classification of a given pixel and assesses how well a specific area can be classified (Congalton, 1991). The PA is determined by calculating the proportion of pixels in each class that have been correctly classified, and then dividing this value by the total number of pixels in the corresponding training group (Lillesand et al., 2008). Overall accuracy (OA) is a general metric that measures how well the classifier works for all LULC categories. It is expressed as a ratio of the number of correctly classified pixels to the entire pixel set. 10-fold cross-validation is undertaken to determine the effectiveness of the classifier (Fushiki, 2011) and field verification was undertaken.

After classification, LULC change estimations are carried out, showing the transition from 2000 to 2010 further to 2020. The percentage changes in the LULC are calculated using eq. (9). The LULC change rate is then calculated using eq. (10).

$$\text{LULC change [\%]} = \frac{(A2\text{-}A1)}{(A1)} * 100 \quad (9)$$

$$r = \frac{LULC\ change\ [\%]}{\Delta t} \quad (10)$$

where r represents the annual change rate in %; Δt is the time interval during LULC change; and A1 and A2 are the initial and final LULC areas, respectively.

4.3 Results

4.3.1 Model performance for assessing drivers on surface water dynamics

We assess the performance of the ANN, XGBoost, SVR, RF, and MLR models using the four statistical metrics of RMSE, dr, MBE, and MPE to analyze the impact of drivers on surface water dynamics from 2000 to 2020. XGBoost has high performance value with the lowest RMSE (0.7), highest dr (0.9), and its MBE (0.1) and MPE (0.3%) are close to zero, showing minimal bias and high accuracy (Table 4.3). The SVR also performs well in terms of dr (0.8), MBE (0.3) and MPE (-0.5), though its RMSE value is lower than RF. Whereas, ANN has relatively poor performance, with the highest RMSE (3.2), lowest dr (0.4), and the MBE (1.9) and the MPE (3.6) shows the significant overestimation. RF and MLR show moderate performance with some bias. Therefore, XGBoost is the best performing model and is selected to be applied in this study to analyze the effect of climate drivers on surface water dynamics and changes in LULC.

Table 4.3 Statistical performance of models. RMSE=root mean square error, dr=Willmott's agreement index, MBE=mean bias error, MPE=mean percentage error

Model	Statistical metrics			
	RMSE	dr	MBE	MPE
ANN	3.2	0.4	1.9	3.6
XGBoost	0.7	0.9	0.1	0.3
SVR	1.4	0.8	0.3	-0.5
RF	1.1	0.4	0.8	0.8
MLR	1.9	0.6	-1.2	2.1

4.3.2 Climatic drivers of surface water dynamics

We utilize precipitation, minimum temperature, vegetation, maximum temperature, and actual evapotranspiration as input drivers for the XGBoost model to assess its effect on

surface water dynamics across our study area from 2000 to 2020. The SHAP feature importance describes the contributions of drivers to surface water dynamics. Precipitation is found to be the most influential predictor exhibiting the highest mean SHAP value (average SHAP values 0.91). Minimum temperature (0.75) and vegetation (0.30) also show significant contributions. Whereas, maximum temperature (0.13) and actual evapotranspiration (0.07) have relatively less direct influence on surface water dynamics in our study area.

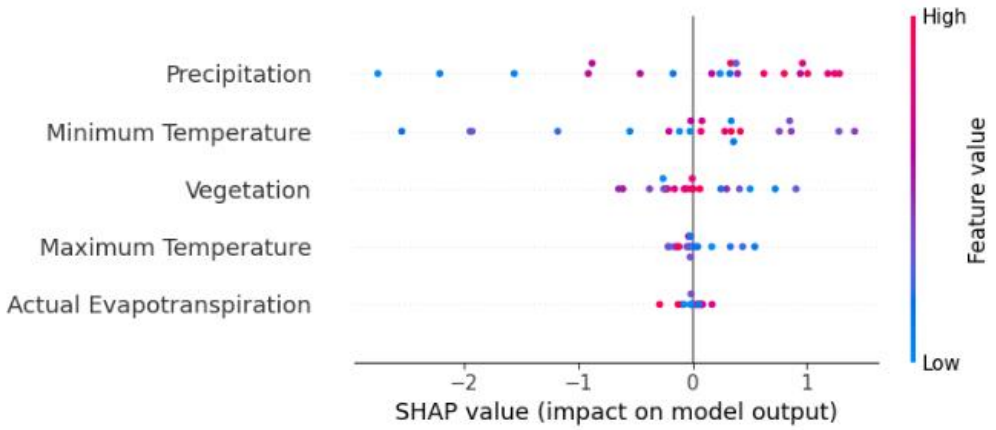


Figure 4.2 Feature importance of drivers contributing to surface water dynamics using the XGBoost model. Positive and negative Shapely Additive exPlanation (SHAP) values represent the associations among the variables.

Besides the variable importance, Figure 4.2 shows the SHAP summary, which demonstrates the effects of each variable on the XGBoost model's output for surface water dynamics. Of the five explanatory features, precipitation has the strongest influence, as can be seen from the broader distribution of SHAP values along the x-axis. Higher precipitation values (red dots) are associated with a positive effect on surface water dynamics, whereas lower precipitation values (blue dots) contribute negatively. Minimum temperature and vegetation also show considerable importance for the model predictions with their SHAP values. In contrast, maximum temperature and actual evapotranspiration have a relatively low effect among the variables considered, as shown by the small distribution of SHAP values close to zero. Thus, the results confirm that precipitation is the dominant driver in determining surface water dynamics, while minimum temperature and vegetation also play a significant role.

4.3.3 LULC classification accuracy

The remote sensing based classified LULC maps using the XBoost model have satisfied producer, user, and overall accuracies of >80 (Table 4.4). The results of producer's accuracy in 2000 was highest for water with 96% and lowest for cropland with 83%. In 2010, the producer's accuracy ranges from a minimum of 85% (cropland) to a maximum accuracy of 97% (water), while in 2020 the respective ranges are from 92% (built-up) to 98% (water). The results of user's accuracy in this study shows that water is relatively correctly classified with 94%, 96% and 97% in 2000, 2010 and 2020, respectively. While the lowest user's accuracy is observed for cropland in 2000 and 2020 (81% and 92%, respectively) and for shrubland in 2010 (83%). The overall accuracies are 90%, 92%, and 95% for the years 2000, 2010, and 2020, respectively. Therefore, this study shows that the classified LULC maps from remotely sensed imagery have improved their classification accuracy over time and provide reliable evidence for the analysis of LULC dynamics.

Table 4.4 Heat map and values of the accuracy assessment of the LULC classification. OA=overall accuracy, PA= producer's accuracy, UA= user's accuracy

LULC type	2000		2010		2020	
	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
Water	96	94	97	96	98	97
Cropland	83	81	85	87	93	92
Grassland	88	84	87	92	95	93
Shrubland	90	84	89	83	94	95
Built up	92	90	94	91	92	95
Forests	92	94	96	95	97	95
OA (%)	90		92		95	

4.3.4 LULC dynamics and rates of change

The LULC from 2000 to 2020 reveals significant spatial and temporal dynamics across different LULC types (Figure 4.3). Shrubland is the dominant LULC type throughout the study period, though it has been decreasing significantly since 2000 from 5,310 km² (38.3%) to 4258.7 km² (30.7%) in 2020 (Table 4.5). Likewise, grassland experiences a consistent reduction from 35.9% in 2000 to 27.1% in 2020. Their combined LULC coverage dropped from almost 75% in 2000 to less than 60% in 2020, showing decrease rates of around 1% y^{-1}. In contrast, built-up and cropland rise from 6.0% to 13.1.% and 7.3% to 16.1%, representing the highest change rates in LULC with a round 6% yr^{-1}. Together they represent the increasing counterpart of LULC starting with 13% coverage in 2000,

spreading to virtually 30% in 2020. Forest shows a minor expansion from 12.2% (2000) to 12.6% (2020) depicting a small, but positive change rate of 0.2% yr^{-1}. Water bodies exhibit a similar small annual change rate of 0.4% yr^{-1}, with an increase coverage from 55.0 km^2 (0.40%) to 59.7 km^2 (0.43%) between 2000 and 2020. Despite their overall small share in total LULC we find a notable increase of nearly 10% over time.

Table 4.5 Decadal LULC coverage and changes from 2000 to 2020

LULC type	2000		2010		2020		2000-2020		
	km^2	%	km^2	%	km^2	%	km^2	%	% yr^{-1}
Water	55.0	0.40	58.7	0.42	59.7	0.43	4.7	8.5	0.4
Built-up	830.6	6.0	1413.0	10.2	1821.7	13.1	991.1	119.3	6.0
Forestland	1694.8	12.2	1873.3	13.5	1748.5	12.6	53.7	3.2	0.2
Shrubland	5310.2	38.3	4756.3	34.3	4258.7	30.7	-1051.5	-19.8	-1.0
Cropland	1007.0	7.3	1309.2	9.4	2229.0	16.1	1222.0	121.4	6.1
Grassland	4972.5	35.9	4459.5	32.2	3752.5	27.1	-1220.0	-24.5	-1.2

Notable transformations in landscape composition are evident Figure 4.3. In 2000, forestland and shrubland dominate the western portion, while grassland covers much of the eastern region. Cropland is scattered but not extensive, and built-up areas are minimal. Water bodies appear sparse and relatively stable throughout the years. By 2010, cropland has noticeably expanded, particularly in the central and northern regions, often replacing shrubland and forestland. This suggests increasing agricultural activities, likely due to population pressure and demand for food. Built-up areas also increase slightly, indicating growing urbanization. By 2020, the land cover pattern becomes more fragmented and forms a distinctive striped or banded distribution, especially in the transition zones between different ecosystems. This pattern suggests intensified land use, possibly related to expansion into previously undisturbed areas. Forestland and shrubland continue to decline, while cropland dominates large stretches of the landscape. Built-up areas become more prominent in 2020, with small, dispersed patches mainly concentrated in the central and northern parts, indicating ongoing but spatially limited urban growth. The clear demarcation in LULC types and the striped pattern indicate more systematic, perhaps policy-driven, land use changes—such as organized agricultural schemes or resettlement programs. Surprisingly, along with the anthropogenic transformation of land use towards cropland and built-up areas at the expense of natural vegetation, we do not see a decline in surface water area, but rather a slight increase.

Figure 4.4 depicts the changes in surface water coverage and the transition from and into other LULC types. Between 2000 and 2010, 0.68 km² of shrubland and 0.76 km² of cropland turned into water bodies. From 2010 to 2020, 0.44 km² of cropland and 1.04 km² of grassland converted into water bodies. In contrast, 0.04 km² and 0.88 km² of water bodies transformed into shrubland and cropland, respectively, between 2000 and 2010. Between 2010 and 2020, 0.44 km² of water bodies turned into cropland. Thus, the transformation of water into other LULC and vice versa is small, but overall, we find a continues increase in surface water coverage, despite that intensive land uses types such as croplands or built-up areas could indicate a stronger demand of surface water for irrigation or human consumption.

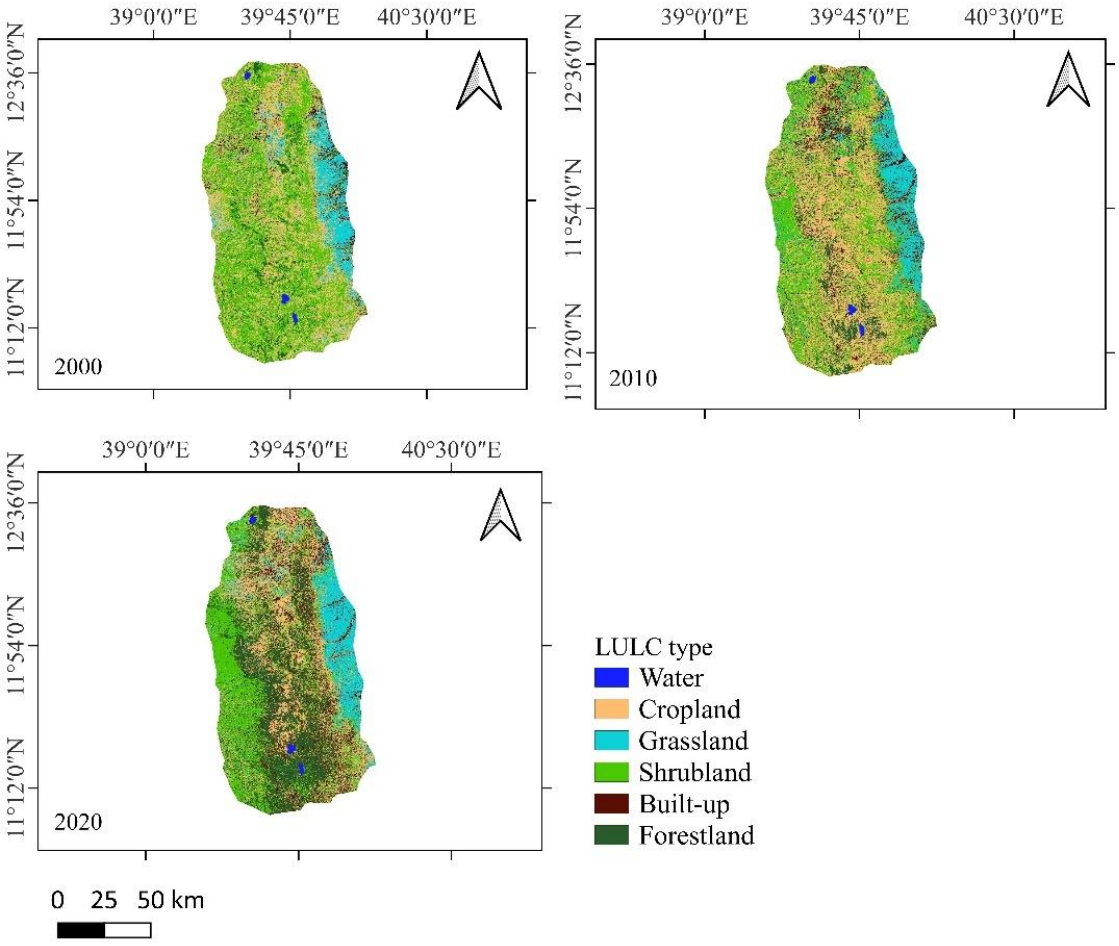


Figure 4.3 LULC classification for 2000, 2010 and 2020

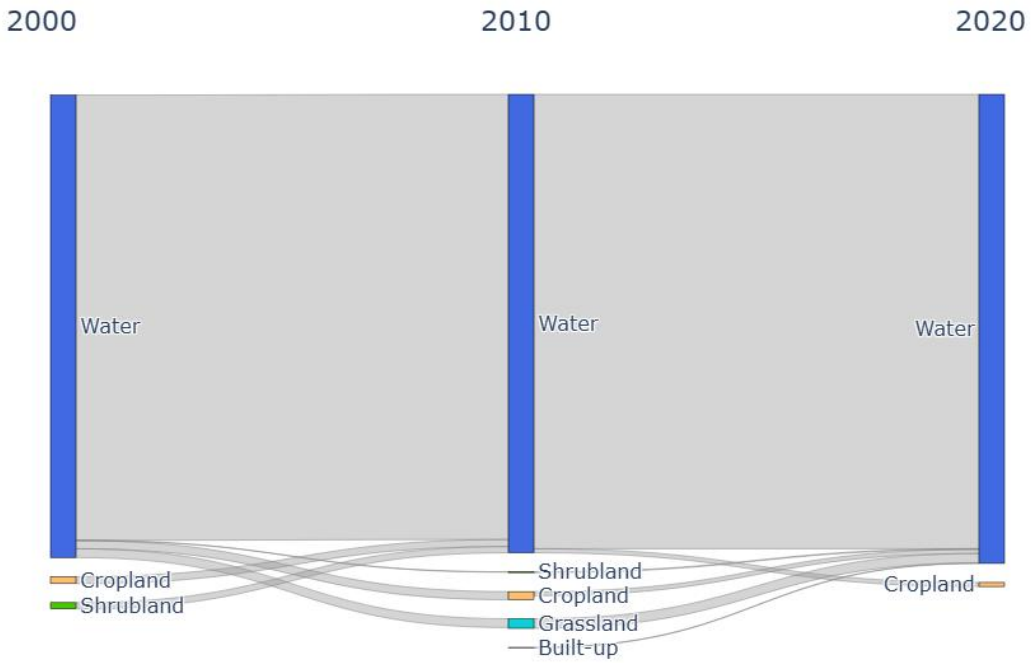


Figure 4.4 Decadal water-related LULC transition from 2000 to 2020

4.4 Discussion

4.4.1 Machine learning for impact assessment on surface water

This study leverages five models to assess the impact of climate drivers and LULC changes on surface water dynamics, differs from other studies that have modeled using linear mixed-effects model (Tulbure & Broich, 2019), mixed-effects RF (Gaines et al., 2022), and statistical analysis with correlation (Mulu et al., 2024b; Zeng et al., 2020b). Our study demonstrates the effectiveness of XGBoost among ANN, SVR, RF, and MLR models as evidenced by statistical metrics with a minimum RMSE of 0.7, maximum dr of 0.9, and low bias in MBE and MPE. These results highlight that XGBoost provides high accuracy while minimizing overestimation and underestimation errors, making it a reliable model. XGBoost's superior performance is probably attributed to its ensemble learning technique, which uses gradient boosting to rectify errors in iterative steps, thereby achieving greater accuracy. This approach is particularly important for determining the relationships between drivers and surface water dynamics, especially for interpreting the impact on model output in terms of SHAP values. On the other hand, ANN exhibits low performance, probably attributed to ANN's reliance on large datasets and relatively lower interpretability than XGBoost, particularly for analyzing the impacts of drivers on surface water dynamics in our study.

Our results support by the broader perspective in predicting climate data in Germany that the XGBoost has high performance in its ability to predict and is well-suited for interpreting its results from this complex model with SHAP (Ghafarian et al., 2022). In another way, the mixed effect RF model has a good performance in its ability to assess surface water dynamics using climate and anthropogenic drivers in the United States from 2000 to 2018 (Gaines et al., 2022). Therefore, SHAP's ability to provide detailed insight into model predictions supports the application of XGBoost to hydrological studies, providing a robust and interpretable framework, particularly for analyzing the effects of climate and LULC drivers on surface water dynamics.

4.4.2 Climate drivers and LULC changes

We explore the impact of climate variables and LULC changes explicitly on surface water dynamics, whereas most studies address the effects of climate and anthropogenic drivers on water resources more generally (e.g., Gaines et al., 2022; Mulu et al., 2024; Tulbure and Broich, 2019; Zeng et al., 2020). We find that climate drivers contributing to surface water dynamics are dominant, while the spatiotemporal LULC transformation plays a minor, but still relevant role in the expansion of water bodies in our study area. This is probably attributed to Ethiopia's strong dependence on rainfall when it comes to surface water distribution (Berhanu et al., 2014). The significant expansion of cropland and built-up areas is due to strong human pressure in Ethiopia (Abebe et al., 2022). Thus, the LULC change analysis helps to understand the overall LULC transformation and process in our study area, in addition to climate effects.

Our results are consistent with hydrological studies emphasizing that the climate drivers influence surface water dynamics (Gaines et al., 2022; Tulbure & Broich, 2019; G. Zhang et al., 2019). In addition to this, likewise, vegetation cover is an important driver in determining surface water changes (Zeng et al., 2020b). Evapotranspiration is also likely to indirectly influence surface water dynamics, though its effect is not as pronounced as precipitation in our study area. In the same way, evapotranspiration has been observed to be a less important climate driver of surface water dynamics in a prior study (Heimhuber et al., 2016). A better understanding of the drivers of surface water dynamics allows decision makers to develop adapted water resources management plans. In our study here it also reflects the susceptibility of future water resources availability to climate change impacts (Chanie, 2024).

Cropland and forests have expanded into the areas surrounding built-up areas. The built-up area themselves increased as well. This is due to villagization and settlement programs concentrated near infrastructure, as well as human activities that contribute to cropland expansion (Abebe et al., 2022). In contrast, shrublands and grasslands have shrunk and transformed into the aforementioned LULC types over time. Somewhat similar results were reported by Tadesse et al. (2025) who found that cropland, forest, and built-up expanded, while shrubland, grazing land, and water bodies have decreased from 2001 to 2021. On the contrary, some studies have observed a reduction in forestland (Habte et al., 2021; Muche et al., 2023), whereas our study shows a marginal increase. The reason might be the introduction of plantations on former grassland from 2000 to 2010, which we considered in our LULC change investigation, and which was partly compensated by forest losses to shrubland, built-up areas, cropland, and again grassland. Furthermore, the reduction of grassland and shrubland and the expansion of cropland and built-up areas over time have been reported in other parts of Ethiopia, consistent with our results, for example, in the Birr River watershed of the Abay Basin (Malede et al., 2023), in the Abune Yosef mountain range, in the highlands of Ethiopia (Gebrehiwot et al., 2021) and also in western Ethiopia (Betru et al., 2019). Therefore, the expansion of cropland, forest, and built-up areas at the expense of shrubland and grassland plays a substantial role in overall LULC transformation, impacting the future use of surface water, which needs to be considered in water resources management and conservation.

The LULC change analysis from 2000 to 2020 reveals that a substantial LULC transition has been observed in all LULC categories during this time. Despite water having the smallest share in our study area, it has expanded slightly over time. This can be attributed to the LULC transformation process in our study area, especially the conversion of shrubland to surface water from 2000 to 2010 and the dynamic changes of cropland, grassland and surface water between 2000 and 2020 as shown in Figure 4. This result is in line with a previous study, whereby surface water in the same study area depicted a gradual increase from 1986 to 2023 (Tesfaye & Breuer, 2025a).

4.5 Conclusion

This study demonstrates the potential of the XGBoost model for assessing the effects of drivers on surface water dynamics. The application of XGBoost with SHAP provides a detailed insight and interpretable results for the effects of each driver on the model

assessment. Our study provides evidence that precipitation and minimum temperature play a pivotal role in surface water dynamics, followed by a lesser but still significant contribution of vegetation changes. To further identify which LULC types contribute to changes of surface water coverage, we analysed decadal satellite imageries and identified losses of cropland, shrubland and grassland converted into surface water and, vice versa, losses of surface water area to cropland and grassland in our study area.

The results of our study underscore the need for integrated policy and management approaches to address the complex interplay between LULC changes, climate variability, and surface water resource dynamics in Ethiopia. Water policy and management strategies should be put in place to ensure sustainable water resources management. Integrated water resources management should be prioritized, considering both climate variability and LULC changes. In addition, sustainable land use planning is critical in protecting against the inappropriate conversion of land use and in promoting the necessity of sustainable land use policies that prioritise environmental conservation alongside development. Thus, understanding these wide-range implications is vital for policy makers, land managers, and stakeholders to achieve sustainable development and water use in Ethiopia.

Nevertheless, any method of modelling the effect of variables on surface water dynamics that depends entirely on past climate data and remotely sensed imagery is limited in achieving future projections of surface water extent. Therefore, future surface water projections require climate and LULC scenarios to enable long-term predictions.

5 General discussion and conclusion

Surface water monitoring

Surface water monitoring plays a fundamental role in maintaining and preserving aquatic and water-related ecosystems,. Effective monitoring not only supports ecosystem sustainability but also provides critical information for the sustainable management and utilization of water resources at both national and local scales. Given Ethiopia's diverse hydro-climatic and topographic conditions, reliable monitoring approaches are essential to capture the spatial and temporal dynamics of surface water bodies.

Various methodological approaches have been employed for remote sensing-based surface water monitoring, primarily categorized into spectral water index and machine learning-based approaches. Among these, spectral water indices have been widely adopted due to their computational efficiency and independence from extensive ground truth data. Evaluating the performance of these indices is essential, especially for large-scale surface water detection, where reliance on training samples is often impractical.

In this regard, the WI and AWEIsh have demonstrated high effectiveness in identifying water bodies across diverse Ethiopian landscapes (Tesfay and Breuer, 2024). These indices estimated national surface water extents of approximately 82,650 km² and 86,530 km², respectively. Despite challenges in distinguishing water from spectrally similar features such as urban surfaces or dense vegetation due to comparable low reflectance characteristics (Liu et al., 2022), both WI and AWEIsh exhibit superior performance in such complex environments. Likewise, Fisher et al. (2016) confirmed that WI and AWEIsh outperform other commonly used indices, including the MNDWI, AWEInsh, and NDWI.

In addition to the water index approach, machine learning algorithms (GTB, RF, SVM) have proven highly effective for long-term and site-specific surface water monitoring when supported by accurately collected training samples (Tesfay and Breuer, 2025a). These algorithms consistently achieved classification accuracies surpassing 90% in terms of overall, producer's and user's accuracy though GTB outperformed the others slightly. The surface water extents derived from GTB closely align with estimates from the benchmark JRC Global Surface Water dataset, though GTB addresses some of the limitations of the JRC product by using locally tuned models rather than globally generalized model (Hamunyela et al., 2022).

Long-term temporal analyses from 1986 to 2023 indicated a significant increase in surface water extent across all sites, accompanied by substantial inter-annual fluctuations (Tesfaye and Breuer, 2025a). These variations are largely driven by anthropogenic activities, such as dam and reservoir construction and the expansion of natural lakes (Sathianarayanan, 2018; Getaneh et al., 2024). Understanding the specific drivers influencing water body dynamics is thus essential, particularly the complex interactions between climate variability and LULC changes (Kazemi Garajeh et al., 2024; Tulbure & Broich, 2019).

The frequency analysis of water occurrence further highlights distinct hydrological behaviors. Areas where high water occurrence frequency represent stable water bodies, while lower ccurrence areas show seasonal or intermittent water bodies that are highly sensitive to climatic and human-induced pressures. These temporary water bodies exhibit strong fluctuations due to factors such as droughts, floods, irrigation expansion, and land degradation.

In conclusion, a comparative evaluation of approaches to detecting surface water reveals that spectral indices, particularly WI and AWEIsh, are reliable for identifying and quantifying surface water on a large scale in Ethiopia's various geographical areas via Sentinel-2 imagery on the GEE system. The operational use of these indices can substantially enhance national-scale water monitoring efforts, thereby supporting sustainable water resource management and contributing to Ethiopia’s broader energy, agricultural, and economic development goals. Moreover, the integration of machine learning notably the GTB algorithm applied to Landsat time series, offering valuable long-term spatio-temporal insights into water dynamics. Such information is crucial for developing data-driven water management policies and for making evidence-based decisions about how to manage water resources sustainably in Ethiopia.

Impact of climate and LULC on surface water dynamics

This study shows that the XGBoost model is more effective than other machine learning and statistical models, including ANN, SVR, RF, and MLR, in assessing the influence of climatic and LULC drivers on surface water dynamics (Tesfaye and Breuer, 2025b). The superior performance of XGBoost is further supported by its interpretability through SHAP values, which provide detailed insights into the contribution of each predictor variable to model outcomes (Ghafarian et al., 2022).

Among the explanatory variables, climatic factors particularly rainfall emerge as the dominant drivers of surface water variability. This finding aligns with broader hydrological studies that underscore the critical role of precipitation in regulating the extent and dynamics of surface water (Gaines et al., 2022; Tulbure & Broich, 2019; Zhang et al., 2019). Furthermore, minimum temperature exhibits a considerable influence, primarily through its impact on evaporation rates and vegetation growth, which indirectly modulate surface water availability. Unlike previous studies, this study quantifies and evaluates the contribution of spatio-temporal LULC transformations to surface water changes. Although the relative contribution of LULC dynamics is smaller compared to that of climatic factors, it still plays a notable role in determining surface water resources, especially within the Ethiopian context. The increase in agricultural land, forested areas and urban settlements at the expense of shrubland and grassland is causing significant changes LULC, which are transforming regional landscapes. These changes probably impact the hydrological cycle, altering surface runoff, infiltration, and evapotranspiration patterns. Consequently, they influence the sustainability and spatial distribution of surface water bodies. Therefore, integrating LULC transitions into long-term water resource management and conservation planning is essential for ensuring ecological resilience and sustainable water availability in Ethiopia.

Implications of findings and further research

A comparative analysis of various remote sensing water indices using Sentinel data over Ethiopia reveals valuable information about the suitability of each index for large-scale surface water monitoring. WI and AWEIsh are particularly effective, as they demonstrate consistent and accurate performance in detecting surface water bodies across diverse landscapes, including vegetated and urban areas, as well as flat terrains. These results suggest that these indices are robust in the face of diverse land cover conditions in Ethiopia and similar landscapes and environmental conditions. Selecting the most effective water indices is crucial for accurately estimating surface water coverage in large-scale assessments or at the national level. Conversely, using less effective indices can lead to significant over- or underestimations of water availability and misguide water management and policy interventions. Given Ethiopia's reliance on water resources for agriculture, hydropower, and ecosystems, prioritizing the selection of the best-performing water indices in national water resource monitoring and management programs is essential. Integrating Sentinel data with higher-resolution imagery could improve the detection of small water

bodies and enhance surface water monitoring in general. However, little is known about the fusion of satellite imagery data for surface water detection (Bioresita et al., 2019; Schmitt, 2020), or its use in combination with higher-resolution imagery for large-scale assessments.

The high classification accuracy of machine learning, particularly GTB, demonstrates its effectiveness and reliability in detecting surface water using remotely sensed data. This suggests its potential for long-term surface water monitoring, particularly in regions with diverse water body sizes and intricate land features. The substantial increase in surface water across all four sites indicates a regional shift in the hydrological cycle. This has crucial implications for disaster risk and water resource management. The expansion of surface water provides opportunities to improve water storage and irrigation planning. Furthermore, expanding surface water is crucial for well-being and preserving ecosystems, ensuring a water supply for agricultural and industrial use (O'Reilly et al., 2015). However, it also raises concerns about increased flooding risk (Alsdorf et al., 2007). Thus, effective water resource management policies need to balance competing interests by incorporating cloud-based remote sensing and machine learning–based surface water monitoring and management strategies.

Spatial heterogeneity in WOF highlights the dynamic nature of water bodies and indicates their susceptibility to climate variability and human-caused changes. Regions with lower WOF stability, such as Hayq-Hashenge and Addis-Ziway, require tailored, site-specific conservation measures. This highlights the importance of developing adaptive water management strategies and policies. Inter-annual variability also highlights the need for real-time monitoring to address unforeseen hydrological changes and the need for dynamic water governance frameworks. Thus, cloud-based remote sensing with machine learning provides high-accuracy, near-real-time data to support these frameworks. These systems enable proactive management and support early warning systems for water scarcity and flooding events. Despite these advancements, combining high-resolution, multi-sensor data with subpixel analysis could reduce classification uncertainties and improve edge detection in areas with mixed water and vegetation.

Applying the XGBoost model demonstrates its effectiveness in modeling the impact of drivers on surface water dynamics. This suggests that XGBoost, a type of ensemble machine learning, is effective in handling complex, nonlinear relationships and interactions

among variables which is critical for the multifaceted nature of hydrological modeling. SHAP analysis enhances the interpretability of the XGBoost model by providing clear insight into how drivers influence surface water dynamics. SHAP analysis reveals that precipitation is the most influential driver of surface water dynamics, indicating its critical role in regulating hydrological changes. Observed fluctuations in surface water dynamics have led to altered hydrological cycles and frequent extreme weather events (IPCC, 2021). In regions where precipitation variability is high, water resource management strategies need to prioritize climate adaptation measures. Minimum temperature and vegetation cover also contribute significantly. This implies that climate variability and changes in land cover substantially influence surface water availability. Time-series analyses have revealed that areas experiencing climate change exhibit dynamic surface water patterns (Donchyts et al., 2016).

The expansion of built-up and cropland areas, coupled with the decline of shrubland and grassland, suggests increased anthropogenic pressure and agricultural intensification. Mitigating potential adverse hydrological effects may require sustainable land-use planning, reforestation efforts, and ensuring water-efficient agricultural activities.

In conclusion, these results highlight the necessity of holistic resource management approaches that balance ecosystem preservation with anthropogenic influences. Since climate patterns and land use changes affect water resources, future modeling efforts should account for anticipated climate and land use transformations. These scenarios can be derived from downscaled general circulation models (GCMs) and combined with socioeconomic land use predictions to develop a more comprehensive outlook. The machine learning based methods developed in this study provide a sound basis for the analysis of climate and land use projections to inform sciene, stakeholders and society about future surface water area and water resources availability in Ethiopia.

References

Abebe, M. T., Degefu, M. A., Assen, M., & Legass, A. (2022). Dynamics of land use/land cover: Implications on environmental resources and human livelihoods in the Middle Awash Valley of Ethiopia. *Environmental Monitoring and Assessment*, *194*(11), 833. https://doi.org/10.1007/s10661-022-10498-7

Acharya, T. D., Subedi, A., & Lee, D. H. (2018). Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. *Sensors*, *18*(8), 2580. https://doi.org/10.3390/s18082580

Ahamed, A., & Bolten, J. D. (2017). A MODIS-based automated flood monitoring system for southeast asia. *International Journal of Applied Earth Observation and Geoinformation*, *61*, 104–117. https://doi.org/10.1016/j.jag.2017.05.006

Alsdorf, D. E., Rodríguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, *45*(2), 2006RG000197. https://doi.org/10.1029/2006RG000197

Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., Moghaddam, S. H. A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., & Brisco, B. (2020). Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5326–5350. https://doi.org/10.1109/JSTARS.2020.3021052

Asfaw, W., Haile, A. T., & Rientjes, T. (2020). Combining multisource satellite data to estimate storage variation of a lake in the Rift Valley Basin, Ethiopia. *International Journal of Applied Earth Observation and Geoinformation*, *89*, 102095. https://doi.org/10.1016/j.jag.2020.102095

Asiltürk, İ., & Çunkaş, M. (2011). Modeling and prediction of surface roughness in turning operations using artificial neural network and multiple regression method. *Expert Systems with Applications*, *38*(5), 5826–5832. https://doi.org/10.1016/j.eswa.2010.11.041

Ayehu, G. T., Tadesse, T., Gessesse, B., & Dinku, T. (2017). *Validation of new satellite rainfall products over the Upper Blue Nile Basin, Ethiopia*. Others (Wind, Precipitation, Temperature, etc.)/Remote Sensing/Validation and Intercomparisons. https://doi.org/10.5194/amt-2017-294

Ayele, H. S., & Atlabachew, M. (2021). Review of characterization, factors, impacts, and solutions of Lake eutrophication: Lesson for lake Tana, Ethiopia. *Environmental Science and Pollution Research*, *28*(12), 14233–14252. https://doi.org/10.1007/s11356-020-12081-4

Belete, M., Diekkrüger, B., & Roehrig, J. (2015). Characterization of Water Level Variability of the Main Ethiopian Rift Valley Lakes. *Hydrology*, *3*(1), 1. https://doi.org/10.3390/hydrology3010001

Berhanu, B., Seleshi, Y., & Melesse, A. M. (2014). Surface Water and Groundwater Resources of Ethiopia: Potentials and Challenges of Water Resources Development. In A. M. Melesse, W. Abtew, & S. G. Setegn (Eds.), *Nile River Basin* (pp. 97–117). Springer International Publishing. https://doi.org/10.1007/978-3-319-02720-3_6

Betru, T., Tolera, M., Sahle, K., & Kassa, H. (2019). Trends and drivers of land use/land cover change in Western Ethiopia. *Applied Geography*, *104*, 83–93. https://doi.org/10.1016/j.apgeog.2019.02.007

Brauman, K. A. (2015). Hydrologic ecosystem services: Linking ecohydrologic processes to human well-being in water research and watershed management. *WIREs Water*, *2*(4), 345–358. https://doi.org/10.1002/wat2.1081

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cabra-Rivas, I., Saldaña, A., Castro-Díez, P., & Gallien, L. (2016). A multi-scale approach to identify invasion drivers and invaders' future dynamics. *Biological Invasions*, *18*(2), 411–426. https://doi.org/10.1007/s10530-015-1015-z

CBD. (2010). *Convention on Biological Diversity for Aichi Biodiversity Targets*. http://www.cbd.int/sp/targets/

Chahine, M. T. (1992). The hydrological cycle and its influence on climate. *Nature*, *359*(6394), 373–380. https://doi.org/10.1038/359373a0

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chanie, K. M. (2024). Hydro-meteorological response to climate change impact in Ethiopia: A review. *Journal of Water and Climate Change*, *15*(4), 1922–1932. https://doi.org/10.2166/wcc.2024.711

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chignell, S. M., Luizza, M. W., Skach, S., Young, N. E., & Evangelista, P. H. (2018). An integrative modeling approach to mapping wetlands and riparian areas in a heterogeneous Rocky Mountain watershed. *Remote Sensing in Ecology and Conservation*, *4*(2), 150–165. https://doi.org/10.1002/rse2.63

Collen, B., Whitton, F., Dyer, E. E., Baillie, J. E. M., Cumberlidge, N., Darwall, W. R. T., Pollock, C., Richman, N. I., Soulsby, A., & Böhm, M. (2014). Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography*, *23*(1), 40–51. https://doi.org/10.1111/geb.12096

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

Demlie, M., Ayenew, T., & Wohnlich, S. (2007). Comprehensive hydrological and hydrogeological study of topographically closed lakes in highland Ethiopia: The case of Hayq and Ardibo. *Journal of Hydrology*, *339*(3–4), 145–158. https://doi.org/10.1016/j.jhydrol.2007.03.012

Desta, G., Legesse, G., Ahmed, M. I., Muluneh, A., & Birhanu, B. (2024). Assessing flood risks and exploring opportunities for flood-based farming in the dry lowlands of Ethiopia. *Frontiers in Sustainable Food Systems*, *8*, 1348094. https://doi.org/10.3389/fsufs.2024.1348094

Dickens, C., Rebelo, L.-M., & Nhamo, L. (2017). *Guidelines and indicators for Target 6.6 of the SDGs: Change in the extent of water related ecosystems over time. Report by the International Water Management Institute. CGIAR Research Program on Water, Land and Ecosystems (WLE). 44p.* (Report by the International Water Management Institute. CGIAR Research Program on Water, Land and Ecosystems (WLE)., p. 44p).

Dile, Y. T., Tekleab, S., Ayana, E. K., Gebrehiwot, S. G., Worqlul, A. W., Bayabil, H. K., Yimam, Y. T., Tilahun, S. A., Daggupati, P., Karlberg, L., & Srinivasan, R. (2018). Advances in water resources research in the Upper Blue Nile basin and the way forward: A review. *Journal of Hydrology*, *560*, 407–423. https://doi.org/10.1016/j.jhydrol.2018.03.042

Donchyts, G., Baart, F., Winsemius, H., Gorelick, N., Kwadijk, J., & Van De Giesen, N. (2016). Earth's surface water change over the past 30 years. *Nature Climate Change*, *6*(9), 810–813. https://doi.org/10.1038/nclimate3111

Donchyts, G., Winsemius, H., Baart, F., Dahm, R., Schellekens, J., Gorelick, N., Iceland, C., & Schmeier, S. (2022). High-resolution surface water dynamics in Earth's small and medium-sized reservoirs. *Scientific Reports*, *12*(1), 13776. https://doi.org/10.1038/s41598-022-17074-6

Dong, S., Guo, H., Chen, Z., Pan, Y., & Gao, B. (2022). Spatial Stratification Method for the Sampling Design of LULC Classification Accuracy Assessment: A Case Study in Beijing, China. *Remote Sensing*, *14*(4), 865. https://doi.org/10.3390/rs14040865

Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G., McDowell, W. H., Kortelainen, P., Caraco, N. F., Melack, J. M., & Middelburg, J. J. (2006). The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography*, *51*(5), 2388–2397. https://doi.org/10.4319/lo.2006.51.5.2388

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A., Soto, D., Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, *81*(2), 163–182. https://doi.org/10.1017/S1464793105006950

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists: Statistical explanation of MaxEnt. *Diversity and Distributions*, *17*(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

FAO. (2016). AQUASTAT Country Profile – Ethiopia. *Food and Agriculture Organization of the United Nations (FAO). Rome, Italy*.

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014a). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, *140*, 23–35. https://doi.org/10.1016/j.rse.2013.08.029

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014b). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, *140*, 23–35. https://doi.org/10.1016/j.rse.2013.08.029

Fisher, A., Flood, N., & Danaher, T. (2016). Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sensing of Environment*, *175*, 167–182. https://doi.org/10.1016/j.rse.2015.12.055

Flörke, M., Schneider, C., & McDonald, R. I. (2018). Water competition between cities and agriculture driven by climate change and urban growth. *Nature Sustainability*, *1*(1), 51–58. https://doi.org/10.1038/s41893-017-0006-8

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N., & Snyder, P. K. (2005). Global Consequences of Land Use. *Science*, *309*(5734), 570–574. https://doi.org/10.1126/science.1111772

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*(1), 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 150066. https://doi.org/10.1038/sdata.2015.66

Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*, *21*(2), 137–146. https://doi.org/10.1007/s11222-009-9153-8

Gadissa, T., Nyadawa, M., Behulu, F., & Mutua, B. (2018). The Effect of Climate Change on Loss of Lake Volume: Case of Sedimentation in Central Rift Valley Basin, Ethiopia. *Hydrology*, *5*(4), 67. https://doi.org/10.3390/hydrology5040067

Gaines, M. D., Tulbure, M. G., & Perin, V. (2022). Effects of Climate and Anthropogenic Drivers on Surface Water Area in the Southeastern United States. *Water Resources Research*, *58*(3), e2021WR031484. https://doi.org/10.1029/2021WR031484

Gao, J. (2009). *Digital analysis of remotely sensed imagery*. McGraw-Hill.

Gashaw, T., W. Worqlul, A., Lakew, H., Teferi Taye, M., Seid, A., & Haileslassie, A. (2023). Evaluations of satellite/reanalysis rainfall and temperature products in the Bale Eco-Region (Southern Ethiopia) to enhance the quality of input data for hydro-

climate studies. *Remote Sensing Applications: Society and Environment*, *31*, 100994. https://doi.org/10.1016/j.rsase.2023.100994

Gebrechorkos, S. H., Taye, M. T., Birhanu, B., Solomon, D., & Demissie, T. (2023). Future Changes in Climate and Hydroclimate Extremes in East Africa. *Earth's Future*, *11*(2), e2022EF003011. https://doi.org/10.1029/2022EF003011

Gebrehiwot, K., Teferi, E., Woldu, Z., Fekadu, M., Desalegn, T., & Demissew, S. (2021). Dynamics and drivers of land cover change in the Afroalpine vegetation belt: Abune Yosef mountain range, Northern Ethiopia. *Environment, Development and Sustainability*, *23*(7), 10679–10701. https://doi.org/10.1007/s10668-020-01079-0

Getaneh, Y., Abera, W., Abegaz, A., & Tamene, L. (2024). Surface water area dynamics of the major lakes of Ethiopia (1985–2023): A spatio-temporal analysis. *International Journal of Applied Earth Observation and Geoinformation*, *132*, 104007. https://doi.org/10.1016/j.jag.2024.104007

Getnet, M., Hengsdijk, H., & Van Ittersum, M. (2014). Disentangling the impacts of climate change, land use change and irrigation on the Central Rift Valley water system of Ethiopia. *Agricultural Water Management*, *137*, 104–115. https://doi.org/10.1016/j.agwat.2014.02.014

Ghafarian, F., Wieland, R., Lüttschwager, D., & Nendel, C. (2022). Application of extreme gradient boosting and Shapley Additive explanations to predict temperature regimes inside forests from standard open-field meteorological data. *Environmental Modelling & Software*, *156*, 105466. https://doi.org/10.1016/j.envsoft.2022.105466

Giardino, C., Bresciani, M., Villa, P., & Martinelli, A. (2010). Application of Remote Sensing in Water Resource Management: The Case Study of Lake Trasimeno, Italy. *Water Resources Management*, *24*(14), 3885–3899. https://doi.org/10.1007/s11269-010-9639-3

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Griggs, D. (2013). Sustainable development goals for people and planet. *Nature*, *,495,*(1–3).

Habte, D. G., Belliethathan, S., & Ayenew, T. (2021). Evaluation of the status of land use/land cover change using remote sensing and GIS in Jewha Watershed, Northeastern Ethiopia. *SN Applied Sciences*, *3*(4), 501. https://doi.org/10.1007/s42452-021-04498-4

Haile, B. T., Bekitie, K. T., Zeleke, T. T., Ayal, D. Y., Feyisa, G. L., & Anose, F. A. (2022). Drought Analysis Using Standardized Evapotranspiration and Aridity Index at Bilate Watershed: Sub-Basins of Ethiopian Rift Valley. *The Scientific World Journal*, *2022*, 1–14. https://doi.org/10.1155/2022/1181198

Halabisky, M., Moskal, L. M., Gillespie, A., & Hannam, M. (2016). Reconstructing semi-arid wetland surface water dynamics through spectral mixture analysis of a time series of Landsat satellite images (1984–2011). *Remote Sensing of Environment*, *177*, 171–183. https://doi.org/10.1016/j.rse.2016.02.040

Hamed, K. H. (2008). Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. *Journal of Hydrology*, *349*(3–4), 350–363. https://doi.org/10.1016/j.jhydrol.2007.11.009

Hamunyela, E., Hipondoka, M., Persendt, F., Sevelia Nghiyalwa, H., Thomas, C., & Matengu, K. (2022). Spatio-temporal characterization of surface water dynamics with Landsat in endorheic Cuvelai-Etosha Basin (1990–2021). *ISPRS Journal of Photogrammetry and Remote Sensing*, *191*, 68–84. https://doi.org/10.1016/j.isprsjprs.2022.07.007

Hansen, M. C., Egorov, A., Potapov, P. V., Stehman, S. V., Tyukavina, A., Turubanova, S. A., Roy, D. P., Goetz, S. J., Loveland, T. R., Ju, J., Kommareddy, A., Kovalskyy, V., Forsyth, C., & Bents, T. (2014). Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD). *Remote Sensing of Environment*, *140*, 466–484. https://doi.org/10.1016/j.rse.2013.08.014

Haregeweyn, N., Melesse, B., Tsunekawa, A., Tsubo, M., Meshesha, D., & Balana, B. B. (2012). Reservoir sedimentation and its mitigating strategies: A case study of Angereb reservoir (NW Ethiopia). *Journal of Soils and Sediments*, *12*(2), 291–305. https://doi.org/10.1007/s11368-011-0447-z

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (New York: springer., Vol. 2).

Heimhuber, V., Tulbure, M. G., & Broich, M. (2016). Modeling 25 years of spatio-temporal surface water and inundation dynamics on large river basin scale using time series of Earth observation data. *Hydrology and Earth System Sciences*, *20*(6), 2227–2250. https://doi.org/10.5194/hess-20-2227-2016

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*(4), 725–749. https://doi.org/10.1080/01431160110040323

Huntington, T. G. (2006). Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology*, *319*(1–4), 83–95. https://doi.org/10.1016/j.jhydrol.2005.07.003

Hurni, H., Bantider, A., Herweg, K. G., Portner, B., & Veit, H. (2007). *Landscape transformation and sustainable development in Ethiopia: Background information for a study tour through Ethiopia*.

IPCC. (2021). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Ji, L., Zhang, L., & Wylie, B. (2009). Analysis of Dynamic Thresholds for the Normalized Difference Water Index. *Photogrammetric Engineering & Remote Sensing*, *75*(11), 1307–1317. https://doi.org/10.14358/PERS.75.11.1307

Jiang, W., Ni, Y., Pang, Z., Li, X., Ju, H., He, G., Lv, J., Yang, K., Fu, J., & Qin, X. (2021). An Effective Water Body Extraction Method with New Water Index for Sentinel-2 Imagery. *Water*, *13*(12), 1647. https://doi.org/10.3390/w13121647

Jimma, T., Chemura, A., Spillane, C., Demissie, T., Abera, W., Ture, K., Terefe, T., Solomon, D., & Gleixner, S. (2024). Coupled Impacts of Soil Acidification and Climate Change on Future Crop Suitability in Ethiopia. *Sustainability*, *16*(4), 1468. https://doi.org/10.3390/su16041468

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Karasiak, N., Dejoux, J.-F., Monteil, C., & Sheeren, D. (2022). Spatial dependence between training and test sets: Another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*, *111*(7), 2715–2740. https://doi.org/10.1007/s10994-021-05972-1

Kazemi Garajeh, M., Haji, F., Tohidfar, M., Sadeqi, A., Ahmadi, R., & Kariminejad, N. (2024). Spatiotemporal monitoring of climate change impacts on water resources using an integrated approach of remote sensing and Google Earth Engine. *Scientific Reports*, *14*(1), 5469. https://doi.org/10.1038/s41598-024-56160-9

Khandelwal, A., Karpatne, A., Marlier, M. E., Kim, J., Lettenmaier, D. P., & Kumar, V. (2017). An approach for global monitoring of surface water extent variations in reservoirs using MODIS data. *Remote Sensing of Environment*, *202*, 113–128. https://doi.org/10.1016/j.rse.2017.05.039

Kundzewicz, Z. W., Mata, L. J., Arnell, N. W., Döll, P., Jimenez, B., Miller, K., Oki, T., Şen, Z., & Shiklomanov, I. (2008). The implications of projected climate change for freshwater resources and their management. *Hydrological Sciences Journal*, *53*(1), 3–10. https://doi.org/10.1623/hysj.53.1.3

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Legass, A. M., Alamirew, T., Gebrehiwot, S. G., & Bernhofen, M. V. (2025a). Comprehensive Flood Vulnerability Analysis and Mapping for the Awash River Basin, Ethiopia. *Natural Hazards Research*, S2666592125000125. https://doi.org/10.1016/j.nhres.2025.01.012

Legass, A. M., Alamirew, T., Gebrehiwot, S. G., & Bernhofen, M. V. (2025b). Comprehensive Flood Vulnerability Analysis and Mapping for the Awash River Basin, Ethiopia. *Natural Hazards Research*, S2666592125000125. https://doi.org/10.1016/j.nhres.2025.01.012

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*(1–4), 1–22. https://doi.org/10.1016/j.jhydrol.2004.03.028

Li, W., Du, Z., Ling, F., Zhou, D., Wang, H., Gui, Y., Sun, B., & Zhang, X. (2013). A Comparison of Land Surface Water Mapping Using the Normalized Difference Water Index from TM, ETM+ and ALI. *Remote Sensing*, *5*(11), 5530–5549. https://doi.org/10.3390/rs5115530

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2008). *Remote Sensing and Image Interpretation* (6th Edition). John Wiley & Sons, Hoboken.

Liu, H., Hu, H., Liu, X., Jiang, H., Liu, W., & Yin, X. (2022). A Comparison of Different Water Indices and Band Downscaling Methods for Water Bodies Mapping from Sentinel-2 Imagery at 10-M Resolution. *Water*, *14*(17), 2696. https://doi.org/10.3390/w14172696

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1705.07874

Malede, D. A., Alamirew, T., Kosgie, J. R., & Andualem, T. G. (2023). Analysis of land use/land cover change trends over Birr River Watershed, Abbay Basin, Ethiopia. *Environmental and Sustainability Indicators*, *17*, 100222. https://doi.org/10.1016/j.indic.2022.100222

Mamuye, M., Gallemore, C., Jespersen, K., Kasongi, N., & Berecha, G. (2024). Changing rainfall and temperature trends and variability at different Spatiotemporal scales threaten coffee production in certain elevations. *Environmental Challenges*, *15*, 100950. https://doi.org/10.1016/j.envc.2024.100950

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, *13*(3), 245. https://doi.org/10.2307/1907187

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., Gao, F., Kutler, J., & Lim, T.-K. (2006). A Landsat Surface Reflectance Dataset for North America, 1990–2000. *IEEE Geoscience and Remote Sensing Letters*, *3*(1), 68–72. https://doi.org/10.1109/LGRS.2005.857030

Maynard, J. J., Karl, J. W., & Browning, D. M. (2016). Effect of spatial image support in detecting long-term vegetation change from satellite time-series. *Landscape Ecology*, *31*(9), 2045–2062. https://doi.org/10.1007/s10980-016-0381-y

McCarthy, M. J., Colna, K. E., El-Mezayen, M. M., Laureano-Rosario, A. E., Méndez-Lázaro, P., Otis, D. B., Toro-Farmer, G., Vega-Rodriguez, M., & Muller-Karger, F. E. (2017). Satellite Remote Sensing for Coastal Management: A Review of Successful Applications. *Environmental Management*, *60*(2), 323–339. https://doi.org/10.1007/s00267-017-0880-x

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, *411*, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

MoA. (2000). *Agro-Ecological Zonation's of Ethiopia. Addis Ababa, Ethiopia.*

Mola Shako, G., & Ping, F. (2021). Ethiopia Water Resources Quantitative Potentials, Management Issues and Challenges. A literature Review Article Paper. *International Journal of Scientific and Research Publications (IJSRP)*, *11*(2), 67–74. https://doi.org/10.29322/IJSRP.11.02.2021.p11008

Morss, R. E., Wilhelmi, O. V., Downton, M. W., & Gruntfest, E. (2005). Flood Risk, Uncertainty, and Scientific Information for Decision Making: Lessons from an Interdisciplinary Project. *Bulletin of the American Meteorological Society*, *86*(11), 1593–1602. https://doi.org/10.1175/BAMS-86-11-1593

Muche, M., Yemata, G., Molla, E., Adnew, W., & Muasya, A. M. (2023). Land use and land cover changes and their impact on ecosystem service values in the north-eastern highlands of Ethiopia. *PLOS ONE*, *18*(9), e0289962. https://doi.org/10.1371/journal.pone.0289962

Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., Lymburner, L., McIntyre, A., Tan, P., Curnow, S., & Ip, A. (2016). Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sensing of Environment*, *174*, 341–352. https://doi.org/10.1016/j.rse.2015.11.003

Mulu, B. A., Zimale, F. A., & Kebede, M. G. (2024). Remote Sensing-Based Long-Term Assessment of Water Dynamics and Influencing Factors in Abaya and Chamo Lakes, East African Rift Valley, Ethiopia. *Air, Soil and Water Research*, *17*, 11786221241299932. https://doi.org/10.1177/11786221241299932

Nguyen, U. N. T., Pham, L. T. H., & Dang, T. D. (2019). An automatic water detection approach using Landsat 8 OLI and Google Earth Engine cloud computing to map lakes and reservoirs in New Zealand. *Environmental Monitoring and Assessment*, *191*(4), 235. https://doi.org/10.1007/s10661-019-7355-x

OECD. (2015). *Drying Wells, Rising Stakes: Towards Sustainable Agricultural Groundwater Use*. OECD. https://doi.org/10.1787/9789264238701-en

O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., Schneider, P., Lenters, J. D., McIntyre, P. B., Kraemer, B. M., Weyhenmeyer, G. A., Straile, D., Dong, B., Adrian, R., Allan, M. G., Anneville, O., Arvola, L., Austin, J., Bailey, J. L., … Zhang, G. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, *42*(24). https://doi.org/10.1002/2015GL066235

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics*, *28*(2), 126. https://doi.org/10.2307/2346729

Pickens, A. H., Hansen, M. C., Hancher, M., Stehman, S. V., Tyukavina, A., Potapov, P., Marroquin, B., & Sherani, Z. (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. https://doi.org/10.1016/j.rse.2020.111792

Puyravaud, J.-P. (2003). Standardizing the calculation of the annual rate of deforestation. *Forest Ecology and Management*, *177*(1–3), 593–596. https://doi.org/10.1016/S0378-1127(02)00335-3

Ramsar Convention on Wetlands. (2018). *Global Wetland Outlook: State of the World's Wetlands and their Services to People.Gland, Switzerland:* Ramsar Convention Secretariat.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rodrigues, M., & De La Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software*, *57*, 192–201. https://doi.org/10.1016/j.envsoft.2014.03.003

Ryu, J., Won, J., & Min, K. (2002). Waterline extraction from Landsat TM data in a tidal flatA case study in Gomso Bay, Korea. *Remote Sensing of Environment*, *83*(3), 442–456. https://doi.org/10.1016/S0034-4257(02)00059-7

Sathianarayanan, M. (2018). Assessment of surface water dynamics using multiple water indices around Adama woreda, Ethiopia. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *IV–5*, 181–188. https://doi.org/10.5194/isprs-annals-IV-5-181-2018

Scaramuzza, P., Micijevic, E., & Chander, G. (2004). *SLC-off Gap-Filled Products Gap-Fill Algorithm Methodology Phase 2 Gap-Fill Algorithm*. US Geological Survey Earth Resources Observation and Science (EROS) Center.

Schmidt, G., Jenkerson, C. B., Masek, J., Vermote, F., & Gao, F. (2013). *Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description*.

Schmitt, M. (2020). Potential of Large-Scale Inland Water Body Mapping from Sentinel-1/2 Data on the Example of Bavaria's Lakes and Rivers. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *88*(3–4), 271–289. https://doi.org/10.1007/s41064-020-00111-2

Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Seyoum, W. M., Milewski, A. M., & Durham, M. C. (2015). Understanding the relative impacts of natural processes and human activities on the hydrology of the Central Rift Valley lakes, East Africa: Assessment of Natural and Human Impacts on the Hydrology of Lakes. *Hydrological Processes*, *29*(19), 4312–4324. https://doi.org/10.1002/hyp.10490

Singh, A., Diop, S., & M'mayi, P. (2006). *AFRICA'S LAKES. United Nations Environment Programme*.

Sisay, A. (2016). Remote Sensing Based Water Surface Extraction and Change Detection in the Central Rift Valley Region of Ethiopia. *American Journal of Geographic Information System*, *5*(2).

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Steinkopf, J., & Engelbrecht, F. (2022). Verification of ERA5 and ERA-Interim precipitation over Africa at intra-annual and interannual timescales. *Atmospheric Research*, *280*, 106427. https://doi.org/10.1016/j.atmosres.2022.106427

Story, M., & Congalton, R. (1986). Accuracy assessment: A user's perspective,. *Photogramm. Eng. Remote Sens*, *52*(3), :397-399.

Tadesse, S., Mekuriaw, A., & Assen, M. (2025). Spatiotemporal dynamics of land surface phenology and its response to climate change in the upper Gelana watershed, northeastern highlands of Ethiopia. *Environmental and Sustainability Indicators*, *25*, 100574. https://doi.org/10.1016/j.indic.2024.100574

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *164*, 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001

Tariq, A., & Qin, S. (2023). Spatio-temporal variation in surface water in Punjab, Pakistan from 1985 to 2020 using machine-learning methods with time-series remote sensing data and driving factors. *Agricultural Water Management*, *280*, 108228. https://doi.org/10.1016/j.agwat.2023.108228

Tesfaye, M., & Breuer, L. (2024). Performance of water indices for large-scale water resources monitoring using Sentinel-2 data in Ethiopia. *Environmental Monitoring and Assessment*, *196*(5), 467. https://doi.org/10.1007/s10661-024-12630-1

Tesfaye, M., & Breuer, L. (2025a). Remote sensing with machine learning for multi-decadal surface water monitoring in Ethiopia. *Scientific Reports*, *15*(1), 12444. https://doi.org/10.1038/s41598-025-96955-y

Tesfaye, M., & Breuer, L. (2025b). Influence of climate, land use, and land cover on surface water dynamics in the Ethiopian highlands. *Remote Sensing Applications: Society and Environment*, Submitted.

Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., Dillon, P., Finlay, K., Fortino, K., Knoll, L. B., Kortelainen, P. L., Kutser, T., Larsen, Soren., Laurion, I., Leech, D. M., McCallister, S. L., McKnight, D. M., Melack, J. M., Overholt, E., … Weyhenmeyer, G. A. (2009). Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology and Oceanography*, *54*(6part2), 2298–2314. https://doi.org/10.4319/lo.2009.54.6_part_2.2298

Tulbure, M. G., & Broich, M. (2019). Spatiotemporal patterns and effects of climate and land use on surface water extent dynamics in a dryland region with three decades of Landsat satellite data. *Science of The Total Environment*, *658*, 1574–1585. https://doi.org/10.1016/j.scitotenv.2018.11.390

Tulbure, M. G., Broich, M., Stehman, S. V., & Kommareddy, A. (2016). Surface water extent dynamics from three decades of seasonally continuous Landsat time series at subcontinental scale in a semi-arid region. *Remote Sensing of Environment*, *178*, 142–157. https://doi.org/10.1016/j.rse.2016.02.034

Vandas, S. J., Winter, T. C., & Battaglin, W. A. (2002). *Water and the environment*. American Geological Institute.

Vanderhoof, M. K., Alexander, L., Christensen, J., Solvik, K., Nieuwlandt, P., & Sagehorn, M. (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. https://doi.org/10.1016/j.rse.2023.113498

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, *185*, 46–56. https://doi.org/10.1016/j.rse.2016.04.008

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global Water Resources: Vulnerability from Climate Change and Population Growth. *Science*, *289*(5477), 284–288. https://doi.org/10.1126/science.289.5477.284

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*(7315), 555–561. https://doi.org/10.1038/nature09440

Wang, C., Jia, M., Chen, N., & Wang, W. (2018). Long-Term Surface Water Dynamics Analysis Based on Landsat Imagery and the Google Earth Engine Platform: A Case Study in the Middle Yangtze River Basin. *Remote Sensing*, *10*(10), 1635. https://doi.org/10.3390/rs10101635

Wang, X., Xiao, X., Zou, Z., Chen, B., Ma, J., Dong, J., Doughty, R. B., Zhong, Q., Qin, Y., Dai, S., Li, X., Zhao, B., & Li, B. (2020). Tracking annual changes of coastal tidal flats in China during 1986–2016 through analyses of Landsat images with Google Earth Engine. *Remote Sensing of Environment*, *238*, 110987. https://doi.org/10.1016/j.rse.2018.11.030

Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, *32*(13), 2088–2094. https://doi.org/10.1002/joc.2419

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., … Zhu, Z. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, *225*, 127–147. https://doi.org/10.1016/j.rse.2019.02.015

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., & Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, *185*, 271–283. https://doi.org/10.1016/j.rse.2015.11.032

Xia, H., Zhao, J., Qin, Y., Yang, J., Cui, Y., Song, H., Ma, L., Jin, N., & Meng, Q. (2019). Changes in Water Surface Area during 1989–2017 in the Huai River Basin using Landsat Data and Google Earth Engine. *Remote Sensing*, *11*(15), 1824. https://doi.org/10.3390/rs11151824

Xiao, X., Boles, S., Liu, J., Zhuang, D., & Liu, M. (2002). Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 VEGETATION sensor data. *Remote Sensing of Environment*, *82*(2–3), 335–348. https://doi.org/10.1016/S0034-4257(02)00051-2

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

Yang, X., Qin, Q., Yésou, H., Ledauphin, T., Koehl, M., Grussenmeyer, P., & Zhu, Z. (2020). Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data. *Remote Sensing of Environment*, *244*, 111803. https://doi.org/10.1016/j.rse.2020.111803

Yue, S., Pilon, P., Phinney, B., & Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, *16*(9), 1807–1829. https://doi.org/10.1002/hyp.1095

Yue, S., & Wang, C. (2004). The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. *Water Resources Management*, *18*(3), 201–218. https://doi.org/10.1023/B:WARM.0000043140.61082.60

Zedler, J. B., & Kercher, S. (2005). Wetland Resources: Status, Trends, Ecosystem Services, and Restorability. *Annual Review of Environment and Resources*, *30*(1), 39–74. https://doi.org/10.1146/annurev.energy.30.050504.144248

Zeng, Y., Yang, X., Fang, N., & Shi, Z. (2020a). Large-scale afforestation significantly increases permanent surface water in China's vegetation restoration regions. *Agricultural and Forest Meteorology*, *290*, 108001. https://doi.org/10.1016/j.agrformet.2020.108001

Zeng, Y., Yang, X., Fang, N., & Shi, Z. (2020b). Large-scale afforestation significantly increases permanent surface water in China's vegetation restoration regions. *Agricultural and Forest Meteorology*, *290*, 108001. https://doi.org/10.1016/j.agrformet.2020.108001

Zhang, G., Yao, T., Chen, W., Zheng, G., Shum, C. K., Yang, K., Piao, S., Sheng, Y., Yi, S., Li, J., O'Reilly, C. M., Qi, S., Shen, S. S. P., Zhang, H., & Jia, Y. (2019). Regional differences of lake evolution across China during 1960s–2015 and its natural and anthropogenic causes. *Remote Sensing of Environment*, *221*, 386–404. https://doi.org/10.1016/j.rse.2018.11.038

Zhang, Z., Wei, M., Pu, D., He, G., Wang, G., & Long, T. (2021). Assessment of Annual Composite Images Obtained by Google Earth Engine for Urban Areas Mapping

Using Random Forest. *Remote Sensing*, *13*(4), 748. https://doi.org/10.3390/rs13040748

Zhu, Z., & Woodcock, C. E. (2012a). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, *118*, 83–94. https://doi.org/10.1016/j.rse.2011.10.028

Zhu, Z., & Woodcock, C. E. (2012b). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, *118*, 83–94. https://doi.org/10.1016/j.rse.2011.10.028

Zou, Z., Xiao, X., Dong, J., Qin, Y., Doughty, R. B., Menarguez, M. A., Zhang, G., & Wang, J. (2018). Divergent trends of open-surface water body area in the contiguous United States from 1984 to 2016. *Proceedings of the National Academy of Sciences*, *115*(15), 3810–3815. https://doi.org/10.1073/pnas.1719275115

Acknowledgements

This research is generously funded by the German Academic Exchange Service (DAAD) scholarship program. The support of the Department of Landscape Ecology and Resources Management (ILR), Justus Liebig University Giessen is also greatly appreciated. My sincere thanks go to my supervisor and advisor, Prof. Dr. Lutz Breuer, who provided all the support that I needed and gave me sufficient space and time to explore my research interest. I would like to thank my colleagues at Addis Ababa University for their support in engaging me in research projects and giving me study leave to pursue my education in Germany. I am grateful for the kind support and inspiration of my parents, siblings, and closest family members.

Declaration

I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and cited all text passages that are derived verbatim from or are based on the content of published work of others, and all information relating to verbal communications. I consent to the use of an anti-plagiarism software to check my thesis. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University Giessen "Satzung der Justus-Liebig-Universität Giessen zur Sicherung guter wissenschaftlicher Praxis" in carrying out the investigations described in the dissertation.

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten.

_____

Mathias Tesfaye Abebe

Giessen, 15.06.2025