Article

Tracking and tracing water consumption for informed water sensitive intervention through machine learning approach

Check for updates

Abraha Tesfay Abraha ✉, Tibebu Assefa Woldeamanuel & Ephrem Gebremariam Beyene

To develop a water conscious strategy, it is critical to track and trace water from its source to the end users, understand water conservation behaviors, and identify the factors that influence water consumption. However, in developing nations, little research has been done to provide a quantitative picture of how water is consumed and transformed in urban households, as well as the water sensitive interventions needed to improve access to clean water. Hence, the main objective of the study was to determine the most significant residential water consumption variables and to predict residential water consumption in a way that can generate water consumption information for water sensitive intervention decision making using the case study of Adama city in Ethiopia. A combination of top down and bottom up data collection techniques were employed as the data collection instrument. Machine learning was integrated with spatial and socioeconomic analytic techniques to estimate daily household water consumption and identify the factors that significantly influence household water consumption. The results show that there is only "one source option" for the city's clean water supply and that different water harvesting methods are not likely to be developed. The average daily water consumption per person is 69 liters which falls below the national standard of 80 liters allocated per person per day. The result reveals that the water distribution network covers only 45% of the city master plan. About 38% of the water demand is unmet and 30% of households only receive water once every three days or fewer. This shows that the city is experiencing physical and economic water scarcity. The results demonstrated that family size, housing quality, income, number of rooms, legal status of the parcel, supply reliability, climate, and topographical features are the most important factors in predicting residential water consumption. This study further demonstrates how well supervised machine learning models, such as the Random Forest Regression algorithm, can predict the household's daily water consumption. The findings also showed that there is a need for significant improvements in water saving habits of the households. Another conclusion that can be drawn is that as long as the city's business as usual water consumption practice doesn't change, the water supply problem will worsen over time.

A sufficient supply of clean water is necessary for life and is also important for the overall wellness of ecosystems, communities, and the economy[1–4]. A healthy society is built on a reliable supply of clean water[5]. Access to clean water is a fundamental human right that must be met by everyone to satisfy their indoor and outdoor water demands[6]. The provision of clean water is widely regarded as being at the heart of all sustainable development goals[7,8]. However, the process of obtaining clean water is lagging behind the sustainable development goals, particularly in developing nations[9]. The growing water challenges emanate from various backgrounds such as relying on remote water supply systems rather than searching for flexible

Ethiopian Institute of Architecture, Building Construction and City Development, Addis Ababa University, Addis Ababa, Ethiopia.
✉e-mail: abrahagisrs@gmail.com

and alternative sources[10], the conception of water as a non-limited natural resource, and the "only one option" supply system[11], unwise water conservation practices[12], ignoring internal alternative water source harvesting following the linearity of urban water flows[13] and lack proactive water centric planning and management practices[14,15]. According to projections, by 2050, domestic water consumption will rise by 130%, and manufacturing will grow by 400%, causing a 55% increase in water demand[16].

The provision of clean water in developing nations is a major concern for several reasons. First, a lack of access to clean water is one of the main factors contributing to illness and death in developing countries[17]. Water-borne infections including cholera, typhoid, and diarrhea are major causes of death and illness in developing countries[18]. Second, women and children in many developing countries are obligated to give up prospects for employment and income sources by walking for hours each day to acquire water[19]. Thirdly, the rapid and unplanned urban growth in developing countries has exacerbated the demand for water, imposing additional stress on water infrastructure[20]. Last but not least, sectorial competition for clean water in developing nations is steadily expanding, necessitating particular water conscious strategies[21].

One of the water consumption sectors that demands emphasis is residential water consumption. Residential water consumption, often referred to as domestic water consumption, is the total amount of clean water used for indoor and outdoor activities such as drinking, cooking, bathing, washing dishes, flushing toilets, watering plants, and utilizing swimming pools[22]. There is mounting evidence that actions emphasizing wise urban residential water consumption are necessary to minimize the worsening drinking water supply problem[23–25]. As a result, there is currently a growing understanding that water sensitive intervention frameworks can be used to address the expanding urban water problems[13,26].

According to ref. 27, "water sensitive" is "a vision for urban water management that calls for shifting urban water systems away from a focus on water supply and wastewater disposal and towards more flexible systems that integrate various water sources, operate through both centralized and decentralized systems, deliver a wider range of services to communities, and integrate into urban design." Ref. 27 named three fundamental principles that guide "water sensitive city". The first principle is known as "city as water supply catchments". This recognizes the significance of diverse water sources in urban areas. This principle advocates for the adoption of appropriate water usage practices to avoid solely relying on one source of water. The second principle is known as "cities providing ecosystem services". This focuses on delivering multiple services to the ecosystem, and the third deals with the principle of "water sensitive communities". The idea of "water sensitive communities" strives to create a culture that is water mindful society.

It can be said that water sensitive intervention is a strategic tool to achieve the vision of a "water sensitive city." Ref. 13 defined "water sensitive intervention" as a means of enhancing water use efficiency and diversifying water sources through water sensitive city principles. Water sensitive intervention is explained by two ideas: supply internalization and supply diversification. Supply internalization, which lessens dependency on water obtained through a remote source, is the capacity to supply water within one's urban boundary system. Supply diversification, on the other hand, focuses on offering alternative water sources, such as surface water, groundwater, roof water harvesting, and recycling[26]. In the simplest terms, water sensitive intervention can be linked with finding a reliable, accessible, quality water, fixing huge leaks, controlling and enforcing restrictions on unauthorized connections, and looking for other alternative water sources[28,29].

To create efficient water conservation strategies and have an influence on future water sensitive planning, water authorities, and policymakers should pay special attention to where, how much, and how households consume water[23,30]. In conjunction with this premise, tracking and tracing residential water consumption is essential to address the underlying issues since it generates the baseline data needed for strategic decision support[13]. Additionally, identifying the variables that influence residential water consumption also makes it feasible to gain a deeper understanding of water

consumption[24,31]. In this regard, studies of the residential water consumption pattern and its influencing variables have been conducted. Notwithstanding the above advancements, the majority of the current state of the art frequently has certain shortcomings. For instance, refs. [23,32] claim that earlier studies had only taken into account one or a few categories of variables without integrating them all into a broader context. Significant empirical results also reported that there is an imbalance in water consumption researches between developed and developing nations. Besides, in developing nations, there is a lack of knowledge about the factors that influence residential water consumption[25,33]. Academic scholars have suggested that measuring residential water consumption and figuring out the factors that affect water consumption are crucial inputs for building a waterwise urban centers[23,34,35].

There are numerous physical, socio-demographic, and climatic variables that influence residential water consumption[23,36,37]. The factors that influence residential water consumption are complex and multifaceted, with variations based on a person's socioeconomic background, local climate, lifestyle preferences, and psychological characteristics[38–40]. The desire to learn more about the specific cultural, technological, social, and climatic factors that explain the factors that influence water consumption has grown in recent decades[23,32,41].

Ref. 42 examined the connections between urbanization and residential water consumption, using the metropolitan area of Barcelona as a case study. According to the researchers' findings, income, housing type, household size, type of outdoor services, and the plant types in the garden are statistically significant in influencing residential water consumption. Ref. 43 also examined per capita water consumption in Sirte, Libya, using the multiple linear regressions (STEPWISE) method. In their research, they found that, interestingly, the per capita water consumption decreased as family size increased and household income had no impact on water consumption.

Ref. 44 examined Johannesburg's residential water consumption trends and the adoption of water efficient technology using probit regression models. According to this research, older, male and low-income households are more likely to consume water efficiently. Ref. 45 used descriptive statistics and inferential statistics to analyze the variables affecting residential water consumption. According to their findings, socioeconomic and demographic variables have the greatest influence on residential water consumption. They found that age, gender, and households headed by women all significantly influence residential water consumption. Ref. 46 employed artificial neural networks and principal component analysis to identify factors that influence residential water consumption in a semi-arid environment. Their result indicated that residential water consumption is influenced by family size, monthly income, housing unit type, housing size and number of rooms.

Ref. 47 looked at the water consumption trend in public pipes in Ghana. They discovered that people consume less water during the rainy season, suggesting that rainwater may be used as a substitute water source. Furthermore, their research shows that the presence of streams and hand dug wells causes even higher reductions in consumption. Ref. 32 reviewed factors that influence household water consumption and concluded that the factors can be categorized under economic, socio-demographic, physical, technological, climatic, and spatial factors. They also acknowledged that determinants in all studies have not led to consistent results. For instance, ref. 25 discovered that households with big family sizes need more water even though per capita water consumption decreases as family size increases. Developed nations, however, despite having small households, consume more water[48]. Income has also ambiguous results. For example, refs. [48,49] found that income is statistically significant in water consumption. However, ref. 25 discovered that there are no appreciable differences in water consumption between high and low income groups.

According to ref. 34, educational background does not consistently affect residential water consumption. Refs. [50,51] found that the number of rooms and the housing condition has a significant impact on indoor and outdoor water consumption. Similarly, ref. 52 asserted that new houses
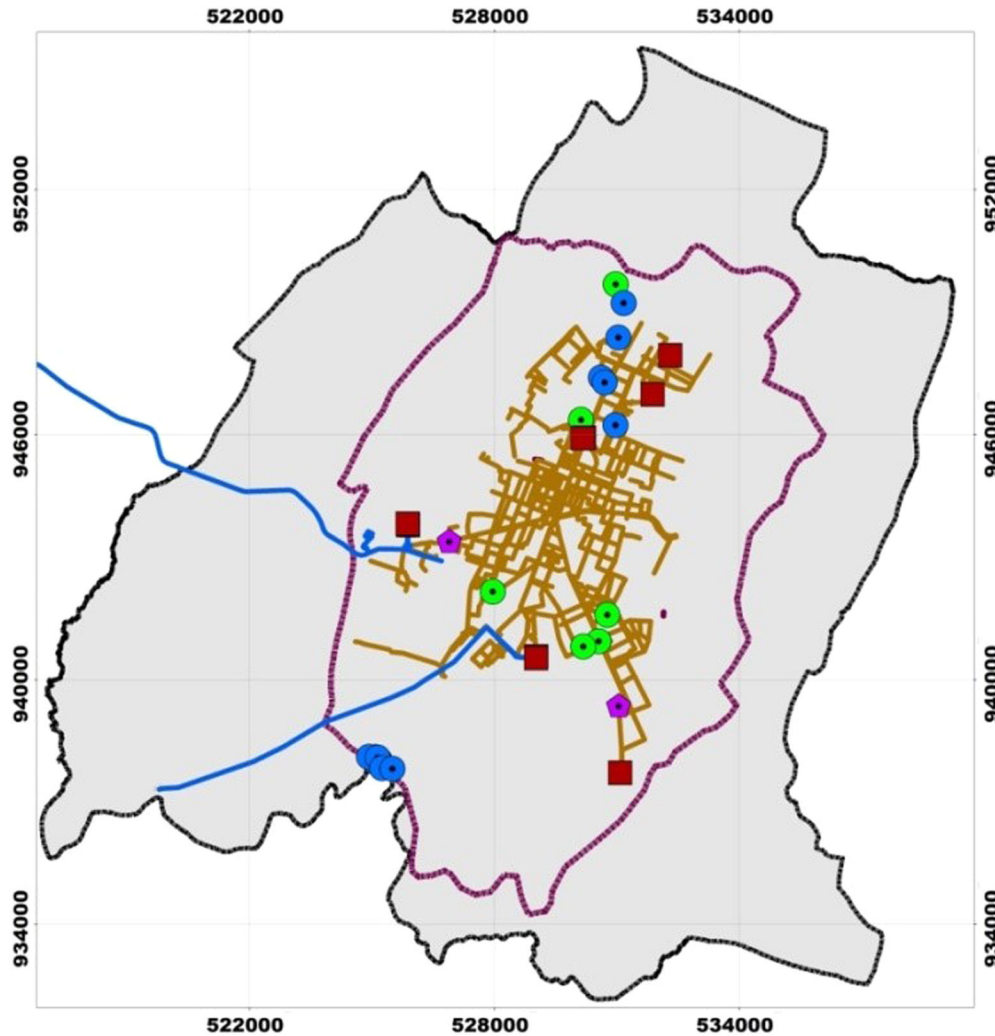
Fig. 1 | Water source locations and supply distribution network map of Adama city. The figure presents the location of water sources and distribution networks. Functional boreholes are denoted by circular green and nonfunctional boreholes are indicated by circular blue shapes. A purple pentagonal shape shows the booster pump, and a dark red rectangle represents the service reservoir. Furthermore, water supply transmission is illustrated by blue lines, and brown lines represent the distribution network. The dataset was obtained from Adama City Water Supply and Sewerage Service Enterprise (2023) and the figure is produced by the researchers using GIS application.
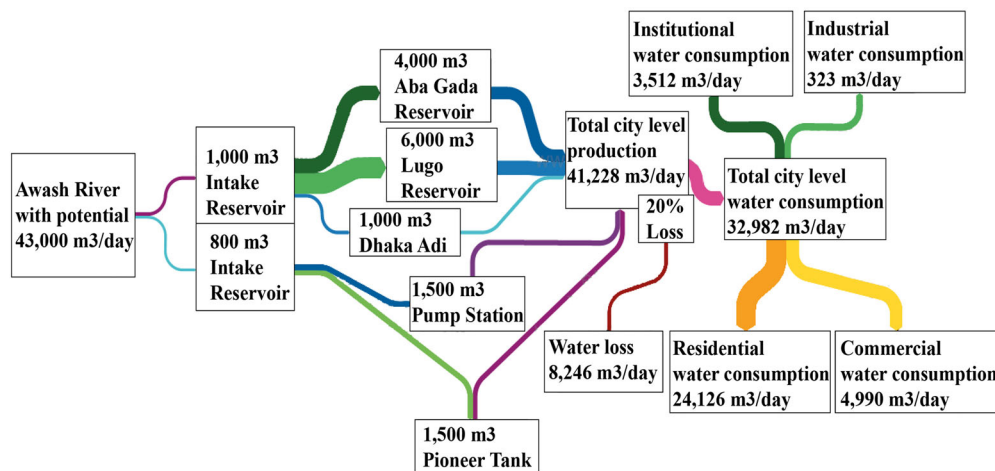

Fig. 2 | Sankey water supply flow diagram from source to service reservoirs and to customers. The figure displays the water supply flow from the Awash River source to five primary service reservoirs, each with distinct capacities (25, 500, 1000, 1500, 400, and 6000 m³), and then to customers. The dataset was obtained from Adama City Water Supply and Sewerage Service Enterprise (2023) and the figure is produced by the researchers.

Article

Table 1 | Population growth, water production, consumption, and non-revenue water (NRW) (2013-2023)

Year	Population	Production (m³ per year)	Consumption (m³ per year)	NRW (m³ per year)	Rate NRW (%)
2013	283,000	6,549,811	5,870,000	679,811	11.58
2014	295,000	6,376,766	5,101,412.8	1,275,353.2	20.00
2015	307,000	6,826,007	5,460,805.6	1,365,201.4	20.00
2016	320,000	7,572,768	6,058,214.4	1,514,553.6	20.00
2017	334,000	9,400,000	7,520,000	1,880,000	20.00
2018	348,000	12,725,867	10,180,693	2,545,174	20.00
2019	362,000	13,782,322	10,787,558	2,994,764	22.00
2020	378,000	14,085,850	11,268,680	2,817,170	20.00
2021	394,000	14,576,609	11,661,287	2,915,322	20.00
2022/23	431,442	14,842,289	11,873,832	2,968,457	20.00

Adama city Water Supply and Sewerage Service Enterprise and field survey (2023).

Table 2 | Proportion of water consumption by sector in Adama city (m3 per year)

Sector	Average Annual water consumption (m³ per year)	Proportion of water consumption (%)
Residential	8,685,708	73.15
Governmental	1,012,838	8.53
Commercial	1,796,511	15.13
NGO	49,870	0.42
Religious	53,432	0.45
Community	148,423	1.25
Industry	116,364	0.98
Others	10,686	0.09
Total	11,873,832	100

Adama city Water Supply and Sewerage Service Enterprise (2023).

with good quality consume more water. It is also reported that installing high indoor efficiency fixtures brings relative water savings between 9% and 12%[53], whereas complete appliance replacement with high water efficiency equipment might cut indoor water use by 35–50%[53]. Ref. 32 reported that climatic determinants have also a significant impact on water consumption.

In addition to identifying influential factors that determine water consumption, establishing a reliable and evidence based future water consumption prediction model is a pivotal input to establish water conscious management plan[54]. In the previous couple of decades, water consumption models have been prospered[55,56]. The model types range from straightforward linear regressions and traditional mathematical equations to more complex and data driven machine learning algorithms[57]. Water consumption models can be roughly divided into two categories called deterministic and probabilistic models[58]. Deterministic models also known as conventional models commonly depend on extrapolating previous per capita water consumption trends[59]. Conventional modeling depends on small data. One popular illustration is linear regression, which evaluates the impact of each independent variable on the dependent variable[60].

The probabilistic model predicts future outcomes by utilizing the influence of random occurrences. One of the most significant advantages of the probabilistic modeling approach is its capacity to precisely quantify the uncertainty associated with predictions[61]. One of the best applications of probabilistic modeling is machine learning[58]. In simplest terms, machine learning is used to produce predictions, and the effectiveness of the technique may be gauged by how well it generalizes to previously unlearned data[60]. There is growing evidence that several water consumption models have been growing which include linear Regression, Deep Neural Network (DNN), Machine Learning (ML) and Hybrid models. Since now, water consumption models focus on pinpointing key water consumption drivers and making water consumption predictions based on past water consumption trends[62–65].

Nowadays, machine learning and deep learning are two of the newest models being used to model nonlinear data. This is owing to the expansion of data driven research, which is revolutionizing the modeling and management of complex and large data driven systems[66,67]. Machine learning-based water consumption modeling has drawn increased attention in recent times[58]. Generally speaking, the current level of water consumption modeling is changing from linear to data driven or AI-based modeling, which mostly uses machine learning modeling and other hybrid models[68–70].
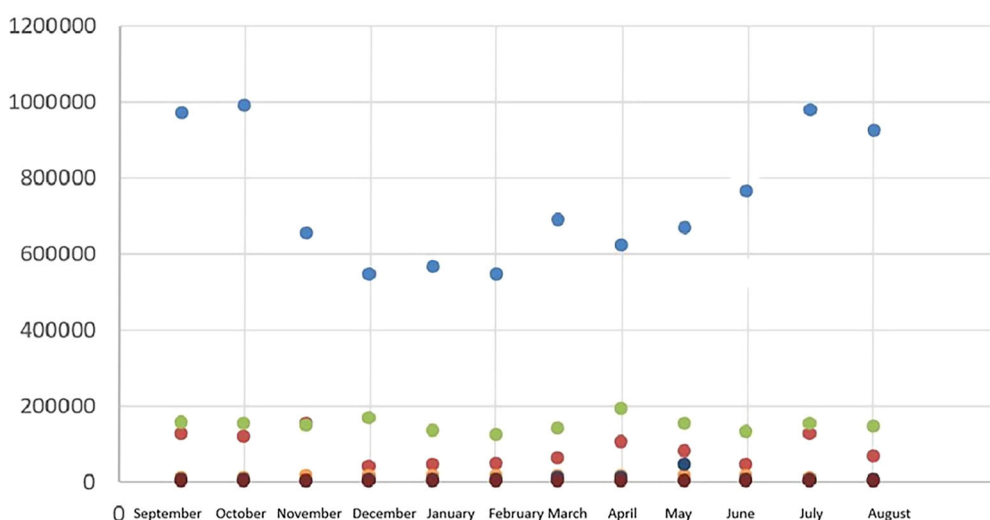

Fig. 3 | Monthly water consumption variation (m3 per month) in Adama city (2022/23). The figure presents the monthly temporal water consumption patterns. The findings reveal that residential water consumption, indicated by blue dots, exhibits notable fluctuations over time. Specifically, there is an increase in residential water consumption during the rainy season, attributed to the upsurge of water supply from the source during the rainy season. The dataset was obtained from Adama City Water Supply and Sewerage Service Enterprise (2023) and the figure is produced by the researchers.

Machine learning is a component of artificial intelligence that enables computers to predict the future by learning from their prior experiences[71]. "Machine learning" is a branch of artificial intelligence that tries to develop systems that can learn from prior knowledge, identify trends, and draw logical conclusions with little to no human input[72]. Ref. 73 discussed that there are at least six major steps for building machine learning model including defining the problem, listing the data inputs and anticipated outcomes, converting the data into a machine learning model compatible format, dividing the data into training and test sets, selecting the machine learning algorithm that best fits the problem, and evaluating the model's performance.

Machine learning algorithms fall into two major categories called unsupervised and supervised. Unsupervised learning is typically used to identify links between datasets, while supervised machine learning is used to categorize data or generate predictions[74]. Supervised machine learning can comprise both classification and regression random forest algorithms[75]. Random Forest is a well-known machine learning algorithm that makes use of supervised learning techniques. A random forest that is constructed from a decision tree algorithmic process is part of supervised machine learning[76]. Random Forest can be applied in machine learning either in the form of classification or regression algorithms[77]. One of the important characteristics of the Random Forest algorithm is its capability to handle data sets with continuous variables, as in regression, and categorical variables, as in classification[78,79].

In this study, the target variable which is household water consumption per day measured in liters is continuous variable. Hence, for this specific case applying random forest regression is the best option. Particularly, regression in the random forest is a method that makes use of an algorithm to understand the relationship between dependent and independent

variables[80]. Regression model helps to predict numerical values based on multiple data sources[74]. Random forest is al\so one of the most popular algorithms for regression problems particularly for predicting continuous variables due to its ease of use and superior accuracy[75,81]. The "R" software platform is also one of the statistical languages that support the random forest application[82].

In summary, the research motivation derives from the reason that water consumption is influenced by both location and context. Hence, there is a need for localized research to comprehend the determinants of water consumption more effectively. In Ethiopia, the existing studies have predominantly concentrated on specific communities with limited sample sizes. Moreover, a significant number of these studies have not methodologically tracked, traced, or mapped water sources in a manner that enables the generation of essential information, enabling well-informed decisions regarding water-sensitive interventions. According to empirical findings, limited researches have been done on residential water consumption in developing countries than in industrialized nations. In developing countries, there is currently limited knowledge about how to apply machine learning for water management[83–86].

Given the identified gaps, this study focuses on Adama city to address key questions: Where are the water sources, and to what extent do they support urbanization and growing water supply concerns? What factors influence residential water consumption in Adama city? How do urban neighborhoods differ in terms of water consumption patterns and reliability? Lastly, what practical water-sensitive intervention strategies can be recommended for Adama city. Hence, the study makes three significant contributions. Firstly, it provides essential baseline data on water consumption for the city. Secondly, it adds to the growing body of knowledge through a localized examination of variables affecting household water usage in developing countries. Lastly, it demonstrates the greater reliability of machine learning results over traditional linear regression models for decision support in predicting residential water consumption.

Results
Tracking and baselining city-level water consumption pattern
To understand the city's water consumption pattern, the study identified the sources of water supply, examined the amount of water distributed, and assessed the discrepancy between water supply and demand. Though historically Adama city's primary supply of drinking water has come from groundwater, the findings in this regard indicate that groundwater exploration can no longer keep up with the city's expanding water demand. Groundwater is becoming a less viable source of water due to the declining capacity of the deep wells and the poor water quality caused by rising fluoride content. According to the tracking and tracing analysis, only six of the nine boreholes are operational. The total daily production from the six working boreholes, even with the irregular flow, is about 3024 m³ per day. This implies that serious doubts exist about the well water supply's sustainability over the long run. Adama's infrastructure for providing water is far from meeting the city's expansion areas. The spatial analysis result confirms that Adama's urban area has grown 22 times in 30 years, from 13,211
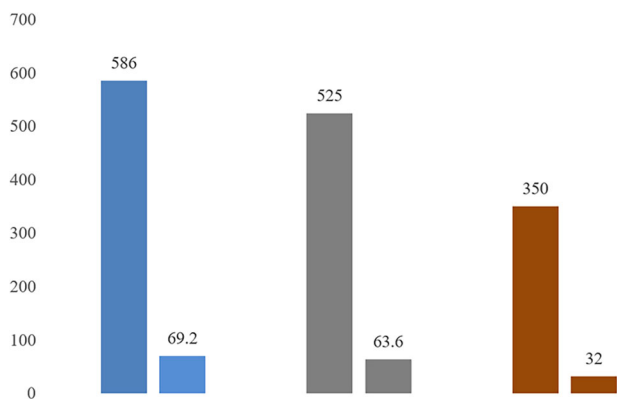

Fig. 4 | Water consumption per household and per capita per day across the sample households. In the figure, the water consumption metrics per household and per capita per day in liters are presented for the sampled households. The blue data points represent the water consumption per household at 586 liters per day, as well as the water consumption per person per day at 69.2 liters. Additionally, the median and standard deviation for both household and individual water consumption are depicted in black and dark red, respectively. The figure is produced by the researchers from household questionnaire survey (2023).

Table 3 | Water consumption per household and per capita per day across central, intermediate, and peripheral neighborhood settlements

Consumption type	Central		Intermediate		Periphery	
	Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation
Average consumption per household per day (liter)	568	358	739	283	402	251
Average consumption per capita per day (liter)	79	32	76	26	47	29

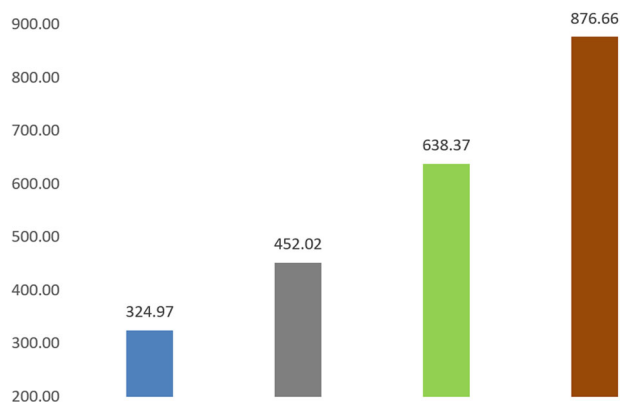Household questionnaire survey (2023).

Fig. 5 | Relationship between average household water consumption and housing conditions. In the figure, the relation between average household water consumption in litters and housing conditions is demonstrated. The varying color denote different categories: blue signifies water consumption for houses categorized as "bad," black represents "fair" conditions, green shows to "good," and dark red indicates houses in "very good" condition. The figure is produced by the researchers from household questionnaire survey (2023).
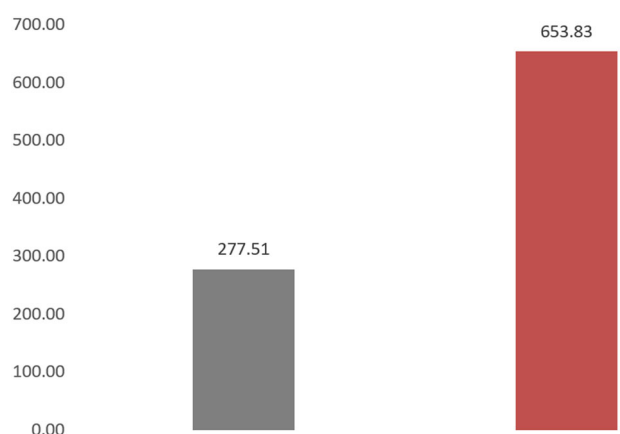

Fig. 6 | Relationship between average household water consumption and parcel legal status. In the figure, the relation between average household water consumption in litres and parcel legal status is illustrated. The black color represents the daily water consumption of informal houses, while the dark red color shows the average daily water consumption of formal house. The figure is produced by the researchers from household questionnaire survey (2023).

Table 4 | Water supply reliability per week across the sampled households

Frequency of water supply per week	Frequency	Percent
Once a week	7	1.80
Two times a week	63	15.80
Three times a week	50	12.50
Four times a week	46	11.50
Five times a week	60	15.00
Six times a week	44	11.00
Always available but with restricted flow	130	32.50
Total	400	100.0

Household questionnaire survey (2023).

Table 5 | Installing water saving devices and retrofitting old fixtures practice

Installation and retrofitting of water-saving technologies	Mean	Std. deviation
Repairing water leaks in toilets and faucets	4.59	0.81
Inspecting dripping taps and toilets	4.59	0.81
Installing water efficient faucets	1.30	0.50
Installing a low water or dual flush toilets	1.29	0.45
Installing flow restrictor in the showers	1.00	0.00
Installing water sense showerheads	1.00	0.00
Installing water sense toilets	1.00	0.00

Household questionnaire survey (2023).

Table 6 | Household water saving behavior in indoor engagements

Indoor water conservation behavior	Mean	Std. deviation
Turn off the water when shaving and washing face and hands	5.14	0.35
Turn off the water when brushing teeth	4.87	0.63
Turn off the shower while soaping	4.80	0.63
Take short showers or limit showering time to 5 min	3.75	1.11
Turn off tap when washing dishes	4.96	0.59
Plug the sink when washing dishes	4.96	0.58
Track water consumption based on the bill	4.65	0.65
Reuse or recycling wastewater	2.54	0.63

Household questionnaire survey (2023).

Table 7 | Outdoor water conservation practice

Outdoor water conservation behavior	Mean	Std. deviation
Washing car with a bucket of water rather than a running hose	3.57	0.78
Watering early in the morning or afternoon	4.95	0.63
Watering plants when necessary and based on plant type	2.47	0.90
Use alternative water resources for outdoor services	2.45	0.50
Improve soil condition with organic material that improves water retention	2.69	1.34

Household questionnaire survey (2023).

hectares in 1991 to 313,211 hectares in 2023. The primary challenges in the water distribution system include aged pipelines, sediment deposits, water leakages, and unauthorized pipeline connections. Additionally, the findings indicate that the water distribution network currently extends to only 45% of the city's master plan (Fig. 1).

Currently, the dominant and probably the only water supply source for Adama city is the Awash River. The Awash River is located 11 km from the city and has a daily output capacity of 43,000 m^3 per day. Six service reservoirs are also part of the Adama water supply system. Each of the reservoirs has a capacity of 25, 500, 1000, 1500, 400 and 6000 m^3 (Fig. 2). The effective water production from Awash River is estimated to be 40,663 m^3 per day. Meanwhile, water demand projection indicated that the estimated average daily domestic and non domestic water demand for the year 2023 is estimated to be 65,226.6 m^3 per day. This implies that the daily gap between

Table 8 | Model train result

400 samples
16 predictor
Pre-processing: scaled (19), centered (19)
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 368, 369, 367, 365, 366, 368…
Resampling results across tuning parameters:

mtry	min.node.size	RMSE (liter)	R squared	MAE (liter)
1	20	253.0211	0.5863452	187.0763
1	30	254.8815	0.5802180	188.8105
3	20	201.9555	0.6922547	140.8809
3	30	206.3987	0.6799379	144.8156
5	20	185.6864	0.7330260	127.8779
5	30	190.2526	0.7211520	132.1991
7	20	176.3966	0.7550056	121.3599
7	30	182.4313	0.7401859	126.3876
9	20	171.8452	0.7646027	117.9047
9	30	178.1024	0.7488656	123.3120
11	20	168.2627	0.7720581	115.6154
11	30	174.7499	0.7558457	121.0202
13	20	166.6818	0.7742122	114.2791
13	30	173.2456	0.7577750	119.8080

Tuning parameter "splitrule" was held constant at a value of maxstat RMSE was used to select the optimal model using the smallest value. The final values used for the model were mtry = 13, splitrule = maxstat and min.node.size = 20.

Table 10 | Model testing result

ME	_0.23
MAE (liter)	110.79
MSE (litre^2)	23071.59
RMSE (liter)	151.89
NRMSE %	48.60
PBIAS %	0.00
RSR	0.49
rSD	0.77
NSE	0.76
mNSE	0.57
rNSE	0.75
d	0.92
md	0.75
rd	0.92
cp	0.82
r	0.88
R2	0.77
bR2	0.72
KGE	0.74
VE	0.82

Table 9 | Variable importance

	Overall scale
Total house hold size	100.0000
Very good housing condition	29.2418
Total household income (monthly)	27.7228
Number of rooms	26.6620
Formal housing ownership	18.1504
Parcel area	13.1409
Households that receive water more frequently than three times per week	11.4178
Monthly mean minimum temperature	10.5959
Intermediate settlement (Location2)	10.4794
Good housing condition	10.2551
Monthly mean maximum temperature	6.2405
Aspect	5.4620
Total annual rainfall	4.0662
Digital elevation model (Dem)	3.1795
Fair housing condition	3.1569
Topographic Position Index (TPI)	1.7441
Central settlement (Location3)	0.9615
Topographic Ruggedness Index (TRI)	0.1460
Slope	0.001

water demand and supply is about 24,563 m^3 per day, or 38% of the water demand is unmet.

Many urban neighborhoods still adopt the practice of centralized point water supply and shift water supply to solve the issues associated with water supply reliability problems. Furthermore, the present

analysis found that the water supply and demand balance index (SDBI) is found to be 0.6.

The result confirms that the current supply method ignores interior alternate water harvesting and is subjected to linear water flow. This study also found that the potential amount of roof water that can be collected annually by residences with roof sizes between 09 m^2 and 270 m^2 ranges from 6 to 181.9 m^3. However, the analysis asserts that the traditional roof water collection from the current residential settlement only amounts to 1–3 m^3. Adama city's population, water production, and consumption have shown an increase year by year. In the previous three decades, Adama city grew almost 15-fold in its population, 3-fold its water production and 2-fold its water consumption. The maximum percentage of nonrevenue water (NRW) in Adama city is 20% of total water production (Table 1).

The result reveals that water consumption residential settlements consume 73.13% of the available water supply. Commerce and government sectors account for 15.33% and 8.53% of total water consumption respectively (Table 2).

Adama city's temporal monthly water consumption is distinct from other cities. Contrary to most cities, the rainy season (June, July, August, and partially September) exposed to a rise in water consumption. According to the technical head of the Adama city Water Supply and Sewerage Service Enterprise, there are root causes for such temporal water consumption variations. Initially, during rainy season the river's flow increases, allowing for adequate water production and even water supply distribution that causes to avoid water rationing. Secondly, the majority of municipal communities have adequate water during the rainy season because water supply pressure becomes normal. But during the dry season, water output decreases and the supply is rationed, resulting in less consistent water availability for each settlement (Fig. 3).

Response rate and background of the respondents

A door-to-door household questionnaire survey was used to acquire the residential water consumption characteristics of the households. A guided map was created to search the sampled households. If they were not at their house, a convenient time and appointment were made. This strategy resulted in a response rate of 100%. According to the analysis of
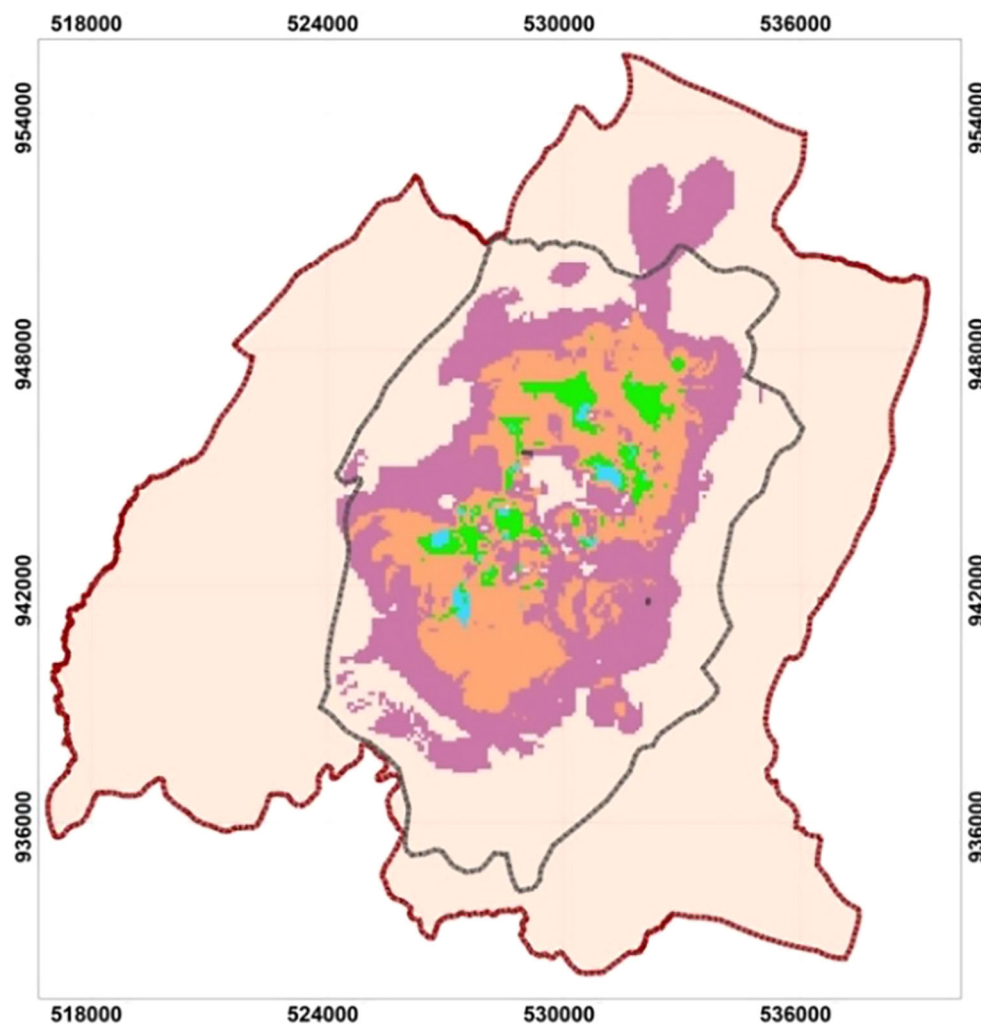
Fig. 7 | Residential water consumption prediction raster output map of Adama city. Figure 7 displays the output of a water consumption prediction raster map, generated through the integration of various factor maps using R Software. The color-coded representation includes cyan for households consuming more than 1135 liters per day, green for water consumption ranging from 909 to 1135 liters per day, orange for consumption between 682 and 909 liters, purple for consumption between 455 and 682 liters, and a peripheral area in pink, representing households with a water consumption ranging from 229 to 455 liters per day. The figure is produced by the researchers using Random Forest Regression in R Software (2023).

the household questionnaire survey, just 28% of the 400 sample households are headed by women, while 72% of the households are headed by men. 79% of respondents reported they are married, 14% are widowed, and 7% are divorced. In terms of respondents' educational backgrounds, 16% of respondents have a degree or higher, 4% diploma and certificate, 34% have finished elementary school, 28% have finished high school, 11% can read and write, and 7% are illiterate. Furthermore, 35% of household heads work for themselves, 15% work for the government, 26% work for businesses, and 16% are unemployed and the rest 8% other sectors.

Household water consumption characteristics

According to the current study, the average household's daily water consumption is estimated to be 586 liters per household. On the other hand, it is estimated that each person consumes 69.2 liters of water each day on average (Fig. 4).

The findings show significant variations in average daily water consumption among households throughout the urban settlement. Particularly, intermediate settlements exhibit higher water consumption per household compared to both central and peripheral settlements. The city center records the highest average daily water consumption per person (Table 3).

Furthermore, the study reveals a correlation between housing quality and water consumption, with high-quality houses using three times more water than lower-quality ones (Fig. 5).

Additionally, formal residential parcels show 2.4 times higher water consumption than informal ones (Fig. 6). This discrepancy is attributed to the remote locations and poor infrastructure of informal parcels, coupled with less reliable water supply in these areas compared to formal residential settlements.

The households also reported that due to the unreliable water supply, local communities are forced to purchase a decentralized water supply from truck tankers. The result also reveals that 40% of the residential area of the city receives at most onetime per week, 24% of the area of the city receives 1–2 times per week, 18% of the area of the city receives 2–3 times per week, 13% of the area of the city receives 3–4 times per week and 5% of the area of the city receives greater than five times per week. Additionally, 70% of houses get water at least four times per week, compared to 30% who get it no more than three times per week (Table 4).

Water consumption conservation practices across the sampled households

Concerning the installation of water-saving devices and the retrofitting of old fixtures, residents sometimes exhibit habits in addressing water leaks in

Table 11 | Data sources and attribute types

Type of variables	Sub attribute	Data type	Source of data
Bill reading	Household bill reading	primary	Household bill
Households socio-economic variables	Household size	Continuous	Household survey
	Household income	Continuous	Household survey
Building and parcel characteristics variables	Housing condition	1. Poor housing condition 2. Fair housing condition 3. Good housing condition 4. Very good housing condition	Household sample survey followed by observation
	Number of rooms	Continuous	Household survey
	Parcel legal status	Categorical (coded values) 1. Formal 2. Informal	Household survey
	Area of parcel	Continuous	Household survey
Distance from the center of the city variable	Central Intermediate Periphery	Categorical (coded values) 1. Periphery 2. Intermediate 3. Central	Household sample survey followed by observation
Water service reliability variable	Reliability 1 Reliability 2	Categorical (coded values) Reliability 1 refers to households that receive water no more than three times per week Reliability 2 refers to households that receive water at least three times per week	Household sample survey followed by observation
Climatic variables	Mean monthly minimum temperature	Continuous (raster)	Historical 10 years of data from 2013 to 2022 obtained from the National Metrology Institute of Ethiopia (NMIE)
	Mean monthly maximum temperature	Continuous (raster)	
	Annual total rainfall	Continuous (raster)	
Topographic variables	DEM	Continuous (raster)	
	Aspect	Continuous (raster)	
	Topographic Position Index (TPI)	Continuous (raster)	
	Topographic Ruggedness Index (TRI)	Continuous (raster)	
	Slope	Continuous (raster)	

The National Metrology Institute of Ethiopia (NMIE) and household survey (2023).

faucets and toilets. Nevertheless, installing water saving equipment with flow controls and sensors is still never for families (Table 5).

Meanwhile, the households have sometimes a habit of turning off the water when taking soap baths, cleaning their faces, and brushing their teeth. In the kitchen, there is sometimes a habit to shut off the tap while washing the dishes. Nonetheless, household members still occasionally limit their shower times to at most five minutes and hardly use recycled water as an alternative source of water (Table 6).

The household members, water their gardens frequently in the morning or afternoon when evaporation is lower. In the meantime, family members hardly ever practice using alternative water sources for outdoor service (Table 7).

Predicting residential water consumption from socioeconomic and spatial information

The "K" fold cross validation method is used in machine learning modeling to evaluate predictions. There are "K" folds or subgroups in the dataset and a new fold is used as the validation set each time the model is trained and evaluated "K" times[87,88]. Similarly, this study also employed 10-fold cross validation that was repeated 5 times. According to the model training summary, the R^2 value is 77.4%. This means that to an accuracy of 77.4%, the independent variables may predict the value of the target variable or the water consumption per household per day in liters (Table 8).

The random forest regression (RFR) approach was used to construct the feature importance plot which shows the factors that are most crucial for predicting household water consumption (water consumption per household per day in liters). Table 9 presents the variables along with their importance scores.

Table 10 shows that the model's performance (R^2 score) is 77%. Besides, based on the prediction raster map output, this study found that the minimum water consumption prediction ranges from 229 to 455 liter per household per day. The maximum water consumption prediction is 682 to 909 liter per household per day (Fig. 7).

Proposed water sensitive intervention approaches for Adama city

The first water sensitive intervention can be achieved by improving indoor and outdoor water conservation practices at the household level. Water is mostly used in households for drinking, cooking, cleaning, bathing, flushing toilets, and watering lawns and gardens. Therefore, the first step to create water wise household is to improve the household's water saving practices. Installing water saving fixtures like faucets, toilets, showers, retrofitting old fixtures, educating the entire family to adopt water saving behaviors like shutting off the faucet while preparing dishes, brushing their teeth, and washing their faces, as well as turning off the shower while using soap and taking quick showers, are some examples of this initiative. Outdoor water saving strategies that

Table 12 | Sample size proportion of residential water users at three spatial scopes

Zone	Local administrative areas	Total population	Household	Sample size Household
Central settlement	Barreechaa	29,646	6588	28
	Daddacha	32,045	7121	30
	Biiqqaa	28,215	6270	24
	Odaa	17,363	3858	16
	Gurmuu	11,147	2477	10
	Badhaatuu	18,386	4086	17
	Gada	16,954	3768	16
	Sub total	153,756	34,168	141
Intermediate Settlement	Gooroo	30,109	6691	28
	Dagaagaa	30,252	6723	28
	Irreecha	34,854	7745	32
	Angaatuu	33,209	7380	31
	Caffee	17,226	3828	16
	Migira	15,057	3346	14
	Sub total	160,707	35,713	149
Periphery Settlement	Migra (partially)	6453	1434	06
	Gaara Lugoo	31,421	6982	29
	Dhaka Adil	2538	564	3
	Dabe Solloque	7083	1574	7
	Boku Kebele	6845	1521	6
	Melka Adama	2637	586	3
	Rural gandas	59,761	13,280	56
	Sub Total	116,739	25,942	110
	Grand total	431,202	95,823	400

Adama City Structural Plan Report, 2020 and researchers' elaboration (2023).

households should implement include planting drought tolerant trees, reusing indoor water use for gardening, and watering the landscape earlier in the day when it is cooler or water evaporation is less.

The second approach can be linked with ensuring reliable water supply by expanding alternate water sources. Diversification of sources increases the availability of water for households, which is one of the best ways to guarantee a sustainable water supply. Increasing the amount of centralized surface and subsurface water supplies, collecting rainwater, and reusing or recycling wastewater are the major components of the supply diversification strategy. The third strategy can be associated with understanding the diverse water competitions and introducing the function of a fit for purpose approach. It is a well-known fact that competition for water is growing rapidly as settlements across the world become more urbanized. Adama city is no exception. Therefore, the municipality must develop an approach that balances the competition among sectors centering on a fit for purpose water supply system. The fourth strategy can be linked with striving to support the transition to a water sensitive city. City level water management should be in line with ensuring flexible systems that integrate different sources of water, operate through both centralized and decentralized systems, and provide a wider range of opportunities to communities at city level.

Discussion
Historically, Adama city has been depended on groundwater as its primary drinking water source. However, this time, groundwater is becoming a less viable source of water due to the declining well capacity

and deteriorating water quality attributed to an upsurge in fluoride content. Currently, the dominant and probably the only water supply source for Adama city is the Awash River. According to ref. [89] who looked into the city's water supply, claimed that if the Awash River supply failed, Adama city's water supply issue would get worse. Similarly, scholars report that the "one source supply option" cannot be a sustainable solution for the growing water scarcity[7,13]. The result also demonstrates that the current water consumption management ignores interior alternate water harvesting approaches and is subjected to linear water flow. Studies also report that in developing countries diversified water supply harvesting methods are not potentially developed[90,91]. Ref. [92] also suggests that implementing alternative water sources such as rainwater harvesting can significantly narrow the gap between water supply and demand.

This study specifically pinpointed socioeconomic, spatial, and climatic variables that determine residential water consumption. Accordingly, the study reveals that household income and size are the most influential from the socioeconomic variables[32,93,94]. Also reported that household size is one of the unwavering water consumption determinants. Refs. [48,49,95] found a positive relationship between income and water consumption. The result demonstrates that houses in a better quality have significant impact on water consumption. In a similar vein, ref. [96] reported that there is a significant correlation between household water consumption and housing quality. Number of rooms is also one of the significant variables that impacts water consumption in Adama city. Refs. [97,98] witnessed that water consumption increased in proportion to the number of rooms. The result found that formal parcels have a higher predictive capacity than informal ones. Ref. [99] also reported that formal parcels consume more water than the informal parcels. In terms of climate variables, minimum and maximum temperatures as well as total yearly precipitation stand out as having a major impact on Adama city's household water consumption. According to refs. [24,100], climatic variables such as maximum temperature have an impact on water consumption patterns. From the topographic variables aspect and elevation have the highest influence on water consumption from all the topographic variables. Similarly, refs. [101,102] reported that natural topography affects water supply distribution and in turn supply reliability.

Given that a sizeable portion of urban water consumption comes from residential settlements, it is crucial to fully understand the factors that influence water consumption and develop a predictable model so that urban planners and water resource managers can make more informed decisions. Machine learning is one of the most innovative approaches that address the appealing issues. Machine learning can be used to analyze water consumption trends, identify factors that influence water consumption, and develop water consumption predictive models. This allows municipalities to better anticipate water demand and adjust their supply accordingly. To sum up, this study also identifies limitations that call for additional investigation. One of the study's shortcomings is that it only used one city as a case study, which makes it difficult to draw broad conclusions or develop knowledge that can be applied generally. Thus, future research will look into other cities that aid in producing comprehensive and generalizable knowledge for making decisions.

Methods
Overview of the study approach
The current study used both Top down and Bottom up data collection approaches. The municipal water supply database served as the foundation for the Top down data collection. The database contains monthly city level water consumption records. The database was an important source of information to quantify the share of water consumption in residential, commercial, industrial, service, and other sectors. The database was also highly helpful for measuring and evaluating city level
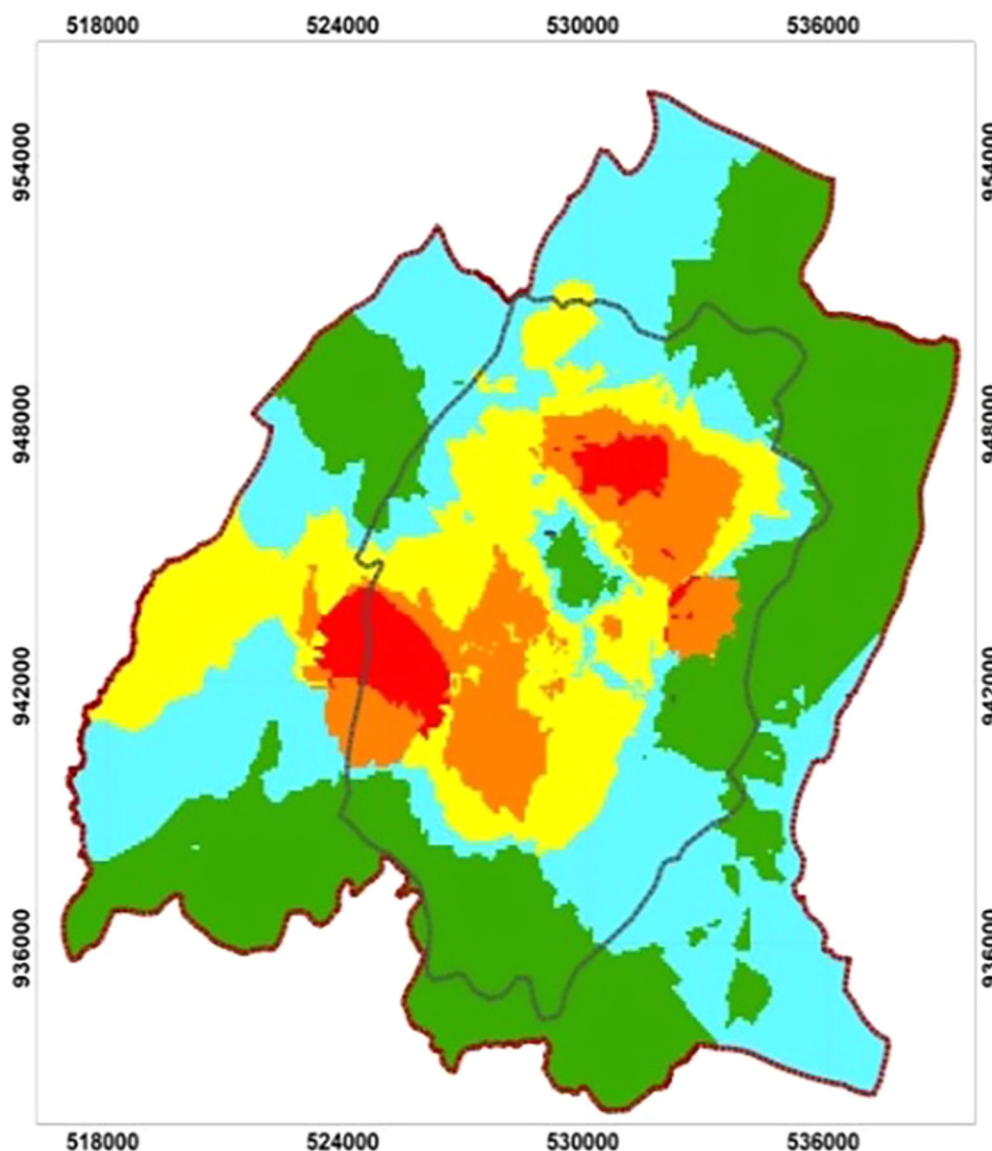
Fig. 8 | Total household income distribution map. Figure 8 shows the spatial distribution of household incomes in Birr. The colors denote varying income groups, with green representing households earning between 2000 and 4000 Birr per month, aqua for those with incomes between 4000 and 6000 Birr, yellow for the 6000−8000 Birr range, orange for 8000−10000 Birr, and dark red for households earning more than 10,000 Birr. The figure is produced by the researchers using GIS from household questionnaire survey (2023).

water production, consumption, and non-revenue water. In contrast, the main approach for getting bottom up data was a questionnaire survey. The data sources for the questionnaire survey were households. In the household survey, the monthly water consumption of each household was collected from their monthly water bill records. To measure the water conservation behavior seven point likert sale measurements was used. Machine learning was utilized in combination with socioeconomic and spatial information to identify the residential water consumption influential factors and to generate residential water consumption model.

Dependent and independent variables

The response variable in this study is the household's daily water consumption, measured in liters per household per day. In essence, this study used three classes of independent variables. The first is the socioeconomic variable, which includes household size, monthly income, building quality, parcel characteristics, the reliability of the water supply, and the location of the residence to the city center. The second variable is the topographical which comprises slope, Topographic Position Index (TPI), Topographic Ruggedness Index (TRI), and aspect, and the third is climatic variable that includes mean monthly minimum temperature, mean monthly maximum temperature, and annual total rainfall (Table 11). Finally, by combining all the three factors a single regression matrix was created. First, the socioeconomic data was converted to spatial data (point vector data) on the GIS platform by using the joining and relating principle. The joined data was used to create raster data using interpolation. This was carried out because the socioeconomic variables were collected with unique identifiers and geographic locations. The socio economic, topographic, and climatic raster data were then processed in the "R" software to produce an aggregated raster file. Lastly, the attribute value of the combined raster file was exported to Excel to create the regression matrix, which serves as the final input for the training data, cross validation, and model test.
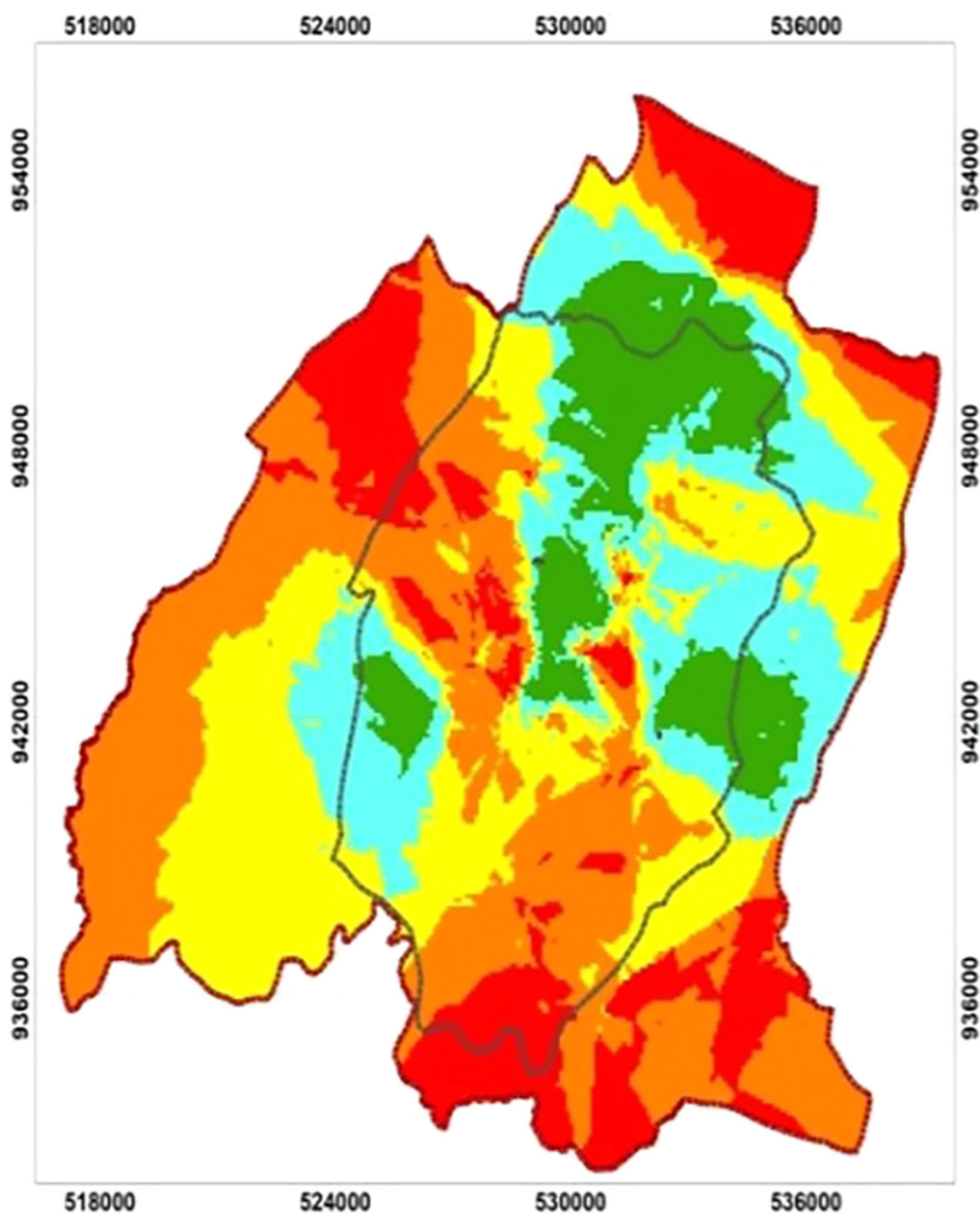
Fig. 9 | Total beneficiary per a meter connection map. In the figure, the spatial distribution of water beneficiary households is indicated. The color coded legend shows different household sizes. Green indicates to households with 4–7 members, aqua represents households with 7–9 members, yellow denotes those with 9–11 members, orange indicates households with 11–13 members, and dark red shows households exceeding 13 members. The figure is produced by the researchers using GIS from household questionnaire survey (2023).

Generating a random sample and recruiting respondents

The recruitment or selection of household respondents is influenced by the choice of the sampling frame. The sampling frame for this inquiry is the list of households that are present in the residential communities. The list of households and their parcel attributes (names, parcel area, and geographic location data) were extracted from the municipal land inventory using GIS software. To accomplish this, the database was changed to Excel format to use random number generators in Excel. After the sampling frame was created, successive parcel "id" numbers were assigned to use simple random sampling to select households from the list of sample frames, giving each item an equal chance of being chosen. To draw representative samples, first, the city settlement was classified into three settlement categories (central, intermediate, and periphery settlements) to obtain a geographical representation of the residences and to assess the spatial difference in water consumption and reliability as we move from the center to the periphery. The target population (N) in this study was estimated to be 95,823 households, which is defined as all residential water consumers in Adama city. The sample size was determined using Yamane's formula. Yamane (1967) provides a simplified formula to calculate sample sizes (Eq. 1). Based on the formula a total of 400 households where 141 households from central, 149 households from intermediate, and 110 from peripheral settlements were considered. Finally, the proportion approach was used to calculate the sample size for the core, intermediate, and periphery samples (Table 12).

$$n = \frac{N}{1 + N(e^2)} \qquad (1)$$

where "N" is population size, "n" is sample size and "e" is margin of error (5%).
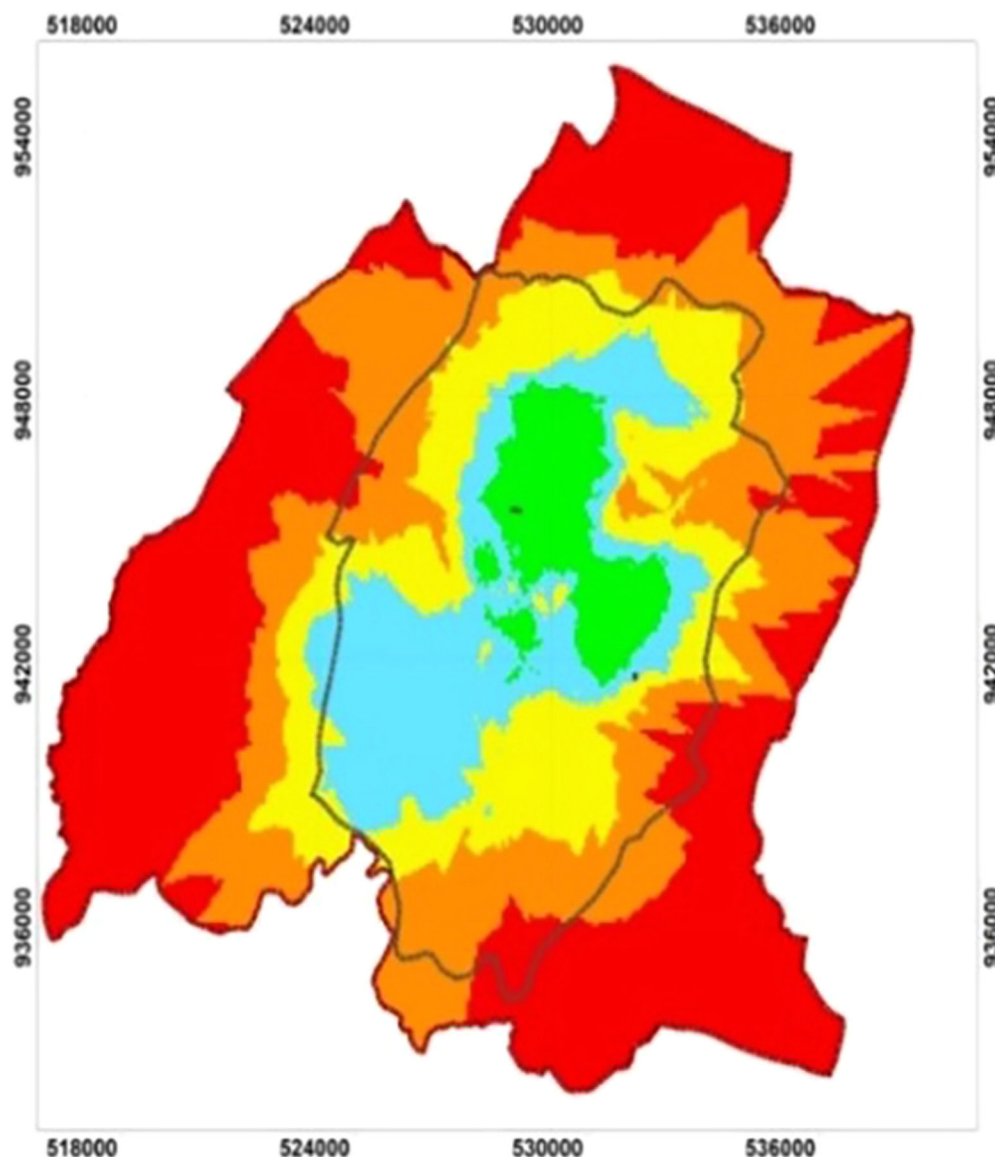
Fig. 10 | Water supply reliability distribution map. In the figure, the spatial distribution of water supply reliability is shown. The color pattern is indicative of the frequency of water availability for households. Dark red shows households receiving water once per week, orange represents those receiving water 1–2 times per week, yellow indicates households with water supply 2–3 times per week, aqua denotes 3–4 times per week, and green reflects households getting water access more than 5 times per week. The figure is produced by the researchers using GIS from household questionnaire survey (2023).

Water conservation perception and behavior measurement. A seven point Likert scale was developed to measure respondents' attitudes towards water conservation behavior about installing water saving appliances as well as indoor and outdoor water conservation behavior. The rating scales were represented as 1 never, 2 rarely, 3 occasionally, 4 sometimes, 5 frequently, 6 usually, and 7 every time.

Raster data preparation or model input raster data preparation
To predict residential water consumption a random forest regression (RFR) model was applied. The RFR model was created using 16 rasterized variables. Raster map layers containing geographical data, such as topographical and climatic characteristics, were constructed in a GIS environment as the prediction's input. A digital topographic map with a 2 m contour interval was subjected to 3D analysis to create elevation, slope, Topographic Position Index (TPI), Topographic Ruggedness Index (TRI), and aspect factor maps. To create a socio-economic raster dataset, the socioeconomic point data were also interpolated using the Kriging method. The following are sample maps that are used to generate the prediction water consumption raster map (Figs. 8–13).

Model execution procedure
In this study, machine learning was used in conjunction with the Random Forest Algorithm. The following primary steps were taken to generate the final water consumption raster map.

Preprocessing stage. The data cleaning procedure was mainly conducted to identify and correct data entry errors and outliers. The missing
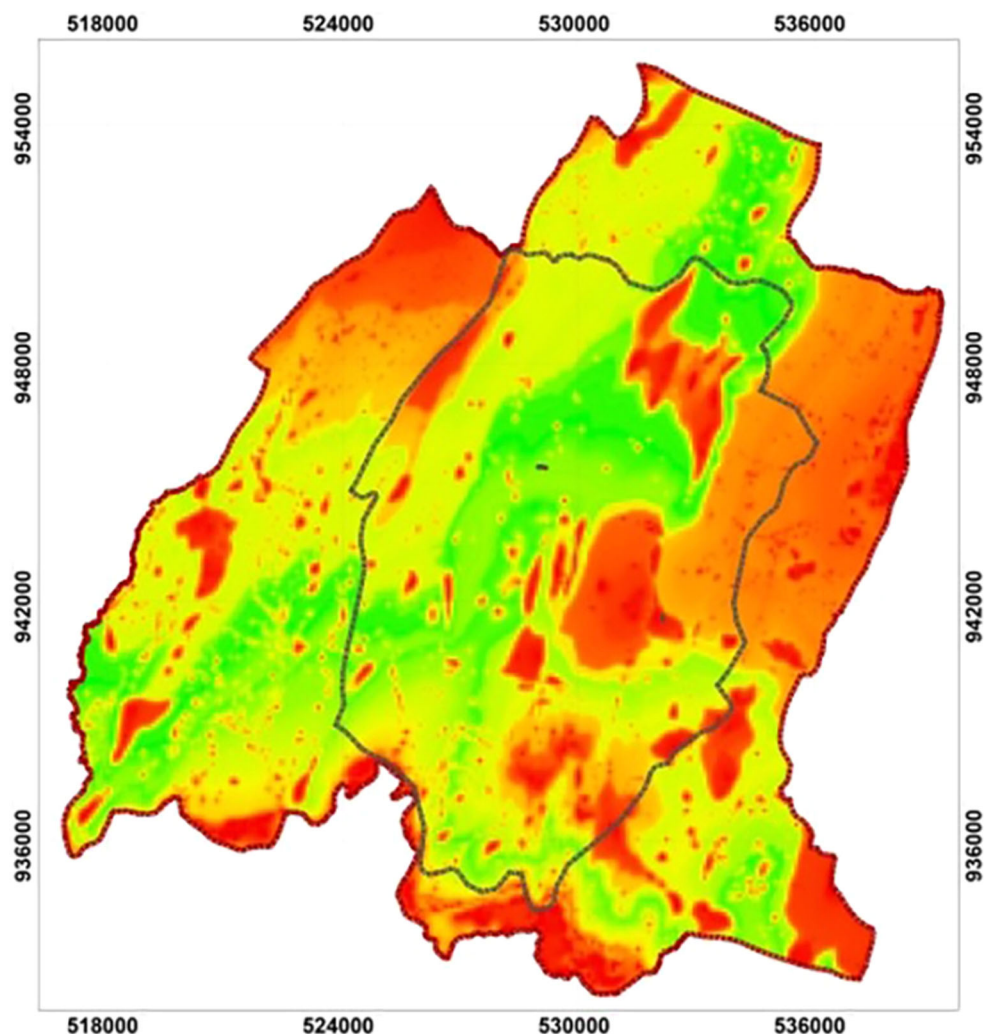
Fig. 11 | Digital elevation map of Adama city. Figure 11 presents the topographic elevation map of the Adama. The color gradient, transitioning from dark red to green, denotes the elevation variations with an increase observed from the red to green color. The dataset was obtained from National Metrology Institute of Ethiopia (NMIE) and the figure is prepared by the researchers using GIS application.

values resulting from incorrect data entry were identified by looking at the frequencies for each variable in SPSS. After looking through the number of "Valid and Missing" values derived from the frequencies result, just two missed values were discovered. To correct the missed values, the original questionnaires were tracked down using their unique ID, and the correct recording was completed. The Z score detection approach was used to find the outliers. The Z score method is one of the most used techniques in machine learning for outlier detection[103]. If a data point has a Z score greater than 3 or less than -3, it is considered as an outlier[103,104]. To obtain the Z score for an observation, the raw measurement is subtracted from the mean, and the standard deviation is divided. The formula for Z score is denoted by Z, while the mean is represented by μ and the standard deviation by σ (Eq. 2).

$$Z = \frac{X - \mu}{\sigma} \qquad (2)$$

Using the Z score outlier identification approach, two outlier data were found out of 400 samples with a Z score value greater than three. After understanding the outlier characteristics, the researchers decided to replace the outliers using the median value imputation approach rather than total removal. As the mean value is highly influenced by the outliers, it is advised to replace the outliers with the median value[105]. After correcting errors, data transformation was the next step. Data transformation is the process of converting unstructured data into a form that can be used by machine learning[106]. As part of this step, the integrated Excel data (regression matrix) was first converted to a CSV which is compatible with the "R" software.

Data splitting into train and test procedure. Building a model depends on training data whereas prediction depends on testing data[107]. In data partitioning determining the partition ratio is important[108]. Scholars use various splitting ratios such as 80:20 and 70: 30[109]. But depending the size of the data ratio is not consistent across all researchers[110]. This study also considered 90:10, which implies that 90% of the dataset passed in the training dataset and 10% in the testing dataset.

Model training procedure. Following the splitting process, cross validation was done using the random forest technique or the Bagging (Bootstrap Aggregating). In the random forest regression, a Bootstrap is constructed to get R^2, MSE, and RMSE[111].
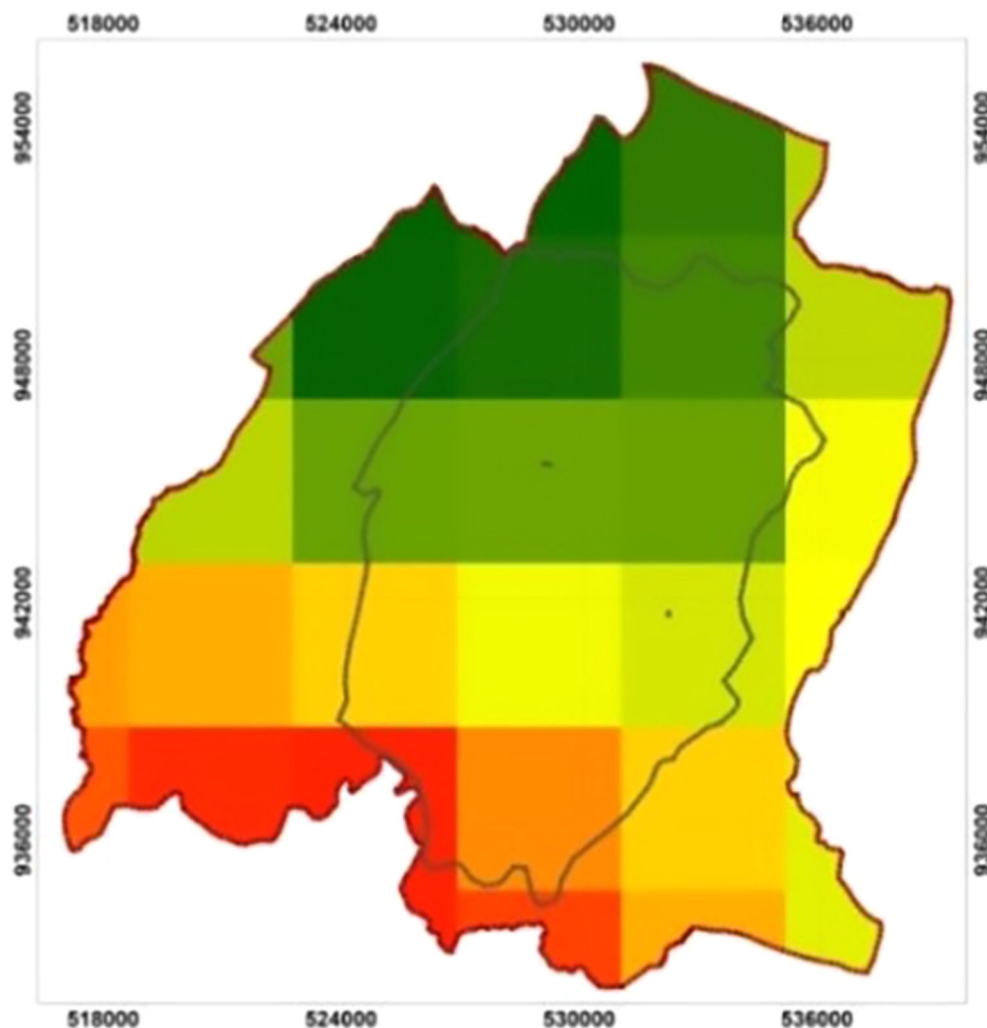
Fig. 12 | Mean monthly maximum temperature. Figure 12 presents mean monthly maximum temperature. The color gradient, transitioning from green to red, denotes the temperature variations with an increase observed from the green to red color. The dataset was obtained from National Metrology Institute of Ethiopia (NMIE) and the figure is prepared by the researchers using GIS application.

Model evaluation against the test data procedure. The hydroGOF in the R package was used to measure the goodness of fit between observed and simulated values. Besides, the model evaluation was used to check the Mean Error (me), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE), and Coefficient of Determination (R^2)

Software used
As a branch of data science, it is now widely known that machine learning primarily involves dealing with a lot of data and statistics[68–70]. The use of the "R" programming language is the most prevalent among the rising number of programming tools for machine learning applications[112]. Working with machine learning requires the "R" programme, which makes tasks simpler, quicker, and more creative[113]. To determine the variables that influence residential water consumption and to produce a raster map of predicted residential water consumption raster map, this study also used the Random Forest algorithm in "R" software. We used the "R 4.0.5" version with packages/libraries (https://www.r-project.org/), "caret" for model training and validation (especially ranger function), terra - for spatial data analysis, and "dplyr" for working on the data frame/CSV file.

Third-party material
All the socioeconomic figures, including household income, beneficiary size, housing quality, supply reliability, location, parcel legal status, and area, were gathered through household surveys and processed using geographic information system (GIS) applications by the authors. The data sets for climate and topography were sourced from the National Metrology Institute of Ethiopia (NMIE) which are legally and freely provided for academic research purposes. Then the topographical and climatic figures, such as the digital elevation, temperature, and rainfall, were prepared in geographic information system (GIS) software by the authors themselves. Municipal level water consumption data was obtained from the Adama City Water Supply and Sewerage Service Enterprise's database following proper legal procedures. Ethical considerations were a priority throughout the dataset acquisition. The authors carefully obtained necessary permissions from relevant offices, ensuring compliance with ethical and legal standards.
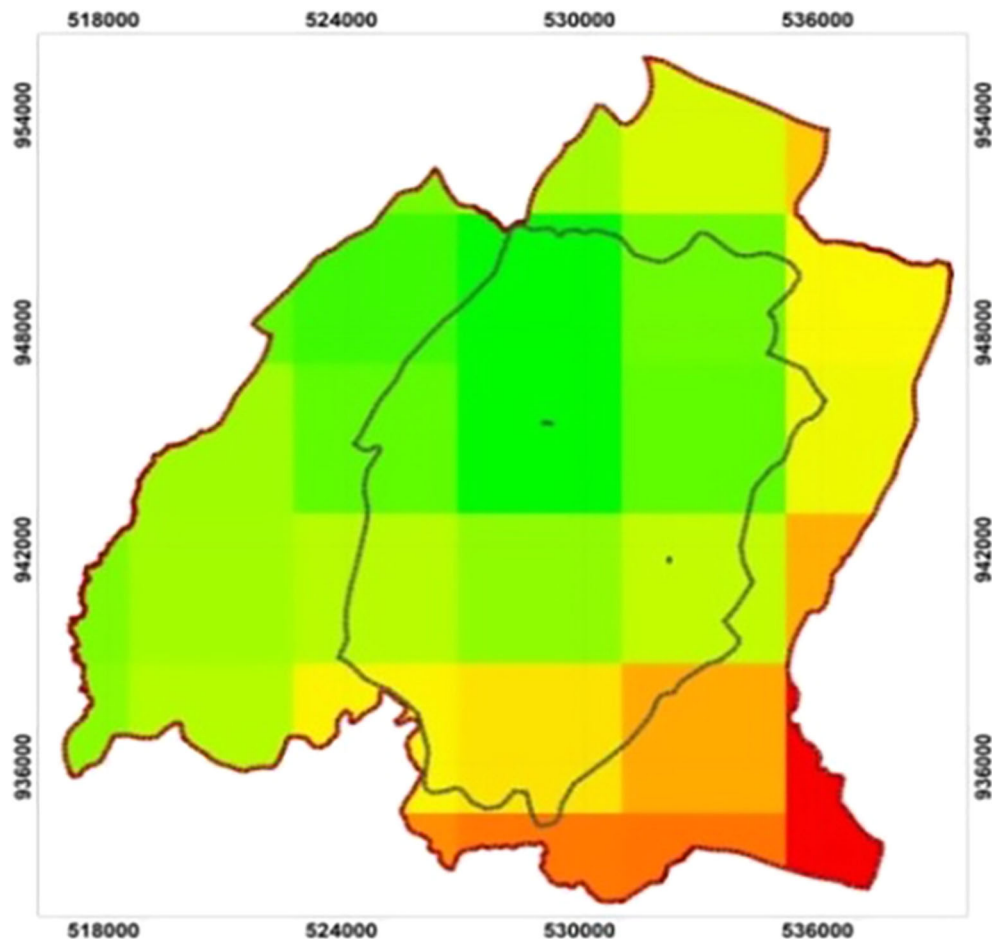
Fig. 13 | Total annual rainfall map of Adama city. Figure 13 presents total annual rainfall map of Adama city. The color gradient, transitioning from red to green, denotes the rainfall variations with an increase observed from the red to green color. The dataset was obtained from National Metrology Institute of Ethiopia (NMIE) and the figure is prepared by the researchers using GIS application.

Data availability
All relevant data are available from the authors.

Code availability
The data analysis code used in this study is available from the authors upon request.

References
1. Mishra, B. K., Kumar, P., Saraswat, C., Chakraborty, S. & Gautam, A. Water Security in a Changing Environment: Concept, Challenges and Solutions. *Water* https://doi.org/10.3390/w13040490 (2021).
2. EPA (The U.S. Environmental Protection Agency). Clean Water Rule. https://www.epa.gov/wotus and https://www.epa.gov/sites/default/files/2016-02/documents/cleanwaterrulefactsheet.pdf (2016).
3. EEA (European Enviromental Agency). Clean water is life, health, food, leisure, energ. https://www.eea.europa.eu/downloads/ed54368c51fc432ba2561a69e83a6593/1620729304/clean-water-is-life-health.pdf (2018).
4. Hunter, P. R., MacDonald, A. M. & Carter, R. C. Water supply and health. *PLoS Med.* **7**, e1000361 (2010).
5. Slekiene, J. & Mosler, H.-J. The link between mental health and safe drinking water behaviors in a vulnerable population in rural Malawi. *BMC Psychol.* **7**, 44 (2019).
6. Spijkers, Otto The sustainable human right to water as reflected in the sustainable development goals. *Utr. Law Rev.* **16**, 18–32 (2020).
7. UN-Water. Summary Progress Update 2021 – SDG 6 – water and sanitation for all. Version: July 2021, Geneva, Switzerland. https://www.unwater.org/app/uploads/2021/12/SDG-6-Summary-Progress-Update-2021_Version-July-2021a.pdf (2021).
8. UNESCO i-WSSM. Water Security and the Sustainable Development Goals. Global Water Security Issues (GWSI) Case Studies, UNESCO i-WSSM, Daejeon. https://unesco-iwssm.org/atchmnfl/download?unityAtchmnflNo=0000000691&fileSn=1 (2019).
9. Admin. The Fight for the Human Right to Water: Stories from the Frontlines. Maji Na Ufanisi (Water and Development). https://majinaufanisi.org/human-right-to-water/?gclid=Cj0KCQjw1_SkBhDwARIsANbGpFuC1GiSsoBmdiiMBnlzszUUfkrKn8vENi27jxzFZExWTn0AEIMmehYaAul9EALw_wcB
10. Luetkemeier, R. et al. Telecoupled groundwaters: new ways to investigate increasingly de-localized resources. *Water* **13**, 1–16 (2021).
11. UNICEF. Water security for all: reimagining WASH water security for all. New York, NY, 10017, USA. https://www.unicef.org/media/95241/file/water-security-for-all.pdf (2021).
12. Wang, X. C. & Luo, L. in *Water-Wise Cities and Sustainable Water Systems: Concepts, Technologies, and Applications*. (eds Xiaochang, C. W. & Guangtao, F.) (IWA, 2021).
13. Renouf, M et al. Urban metabolism for planning water sensitive city-regions: proof of concept for an urban water metabolism evaluation

framework B1.2 Milestone Report. https://watersensitivecities.org.au/wp-content/uploads/2017/12/Milestone-Report-Urban-metabolism_FINAL-1.pdf (2017).

14. Leigh, N. G. & Lee, H. Sustainable and resilient urban water systems: the role of decentralization and planning. *Sustainability* **11**, 918 (2019).

15. Liku Workalemahu Habtemariam, A. A. M., Gelaye, F. & Du, J. Water resilience in a changing urban context: Africa's challenge and pathways for action. Available: https://wrirosscities.org/sites/default/files/water-resilience-in-changing-urban-context.pdf (2021).

16. Director of Publishing for Save The Water. World water demand will increase 55% by 2050. https://savethewater.org/water-demand-to-increase-55-globally-by-2050/ (2019).

17. Government of Canada. Access to water in developing countries. https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/environmental_protection-protection_environnement/water-eau.aspx?lang=eng (2021).

18. WHO. Drinking water. https://www.who.int/news-room/fact-sheets/detail/drinking-water (2022).

19. Sommer, M., Ferron, S., Cavill, S. & House, S. Violence, gender and WASH: spurring action on a complex, under-documented and sensitive topic. *Environ. Urban.* **27**, 105–116 (2015).

20. Güneralp, B., Lwasa, S., Masundire, H., Parnell, S. & Seto, K. C. Urbanization in Africa: Challenges and opportunities for conservation. *Environ. Res. Lett.* https://doi.org/10.1088/1748-9326/aa94fe (2018).

21. United Nations. Policy analysis: aligning economic development and water policies in small island developing states. https://unctad.org/system/files/official-document/aldc2021d6_en.pdf (2022).

22. Water Resources Mission Area. Domestic Water Use. USGS. https://www.usgs.gov/mission-areas/water-resources/science/domestic-water-use (2019).

23. Cominola, A. et al. The determinants of household water consumption: a review and assessment framework for research and practice. *npj Clean Water* **6**, 11 (2023).

24. Bich-Ngoc, N., Prevedello, C., Cools, M. & Teller, J. Factors influencing residential water consumption in Wallonia, Belgium. *Util. Policy* **74**, 101281 (2022).

25. Willis, R. M., Energy, O., Stewart, R. A. & Talebpour, M. R. End use water consumption in households: impact of socio-demographic factors and efficient devices. https://doi.org/10.1016/j.jclepro.2011.08.006 (2011).

26. Musango, J. K., Currie, P. & Robinson, B. Urban metabolism for resource efficient cities: from theory to implementation. https://resourceefficientcities.org/wp-content/uploads/2017/09/Urban-Metabolism-for-Resource-Efficient-Cities.pdf (2017).

27. Wong, T. H. F. & Brown, R. R. The water sensitive city: principles for practice. *Water Sci. Technol.* **60**, 673–682 (2009).

28. Karadirek, I. E. & Aydin, M. E. in *Water and Wastewater Management: Global Problems and Measures* (eds Bahadir, M. & Haarstrick, A.) 53–65 (Springer International Publishing, 2022).

29. Tzanakakis, V. A., Paranychianakis, N. V. & Angelakis, A. N. Water supply and water scarcity. *Water* **12**, 2347 (2020).

30. Russell, S. V. & Fielding, K. S. Water demand management research: a psychological perspective. *Water Resour. Res.* **46**, 1–12 (2010).

31. Rondinel-Oviedo, D. R. & Sarmiento-Pastor, J. M. Water: consumption, usage patterns, and residential infrastructure. A comparative analysis of three regions in the Lima metropolitan area. *Water Int.* **45**, 824–846 (2020).

32. Bich-Ngoc, N. & Teller, J. A review of residential water consumption determinants. *Lect. Notes Comput. Sci.* **10964**, 685–696 (2018).

33. Chang, H., Bonnette, M. R., Stoker, P., Crow-Miller, B. & Wentz, E. Determinants of single family residential water use across scales in four western US cities. *Sci. Total Environ.* **596–597**, 451–464 (2017).

34. Bennett, C., Stewart, R. A. & Beal, C. ANN-based residential water end-use demand forecasting model ANN-based residential water end-use demand forecasting model. https://doi.org/10.1016/j.eswa.2012.08.012 (2012).

35. Makki, A. A., Stewart, R. A., Panuwatwanich, K. & Beal, C. "Better targeted urban water conservation strategies Revealing the determinants of shower water end use consumption: enabling better targeted urban water conservation strategies. https://doi.org/10.1016/j.jclepro.2011.08.007 (2011).

36. Di Mauro, A., Cominola, A., Castelletti, A. & Di Nardo, A. Urban water consumption at multiple spatial and temporal scales. a review of existing datasets. *Water* **13**, 36 (2021).

37. Jorgensen, B., Graymore, M. & O'Toole, K. Household water use behavior: an integrated model. *J. Environ. Manag.* **91**, 227–236 (2009).

38. Addo, I. B., Thoms, M. C. & Parsons, M. Household water use and conservation behavior: a meta-analysis. *Water Resour. Res.* **54**, 8381–8400 (2018).

39. Llausàs, A. & Saurí, D. A research synthesis and theoretical model of relationships between factors influencing outdoor domestic water consumption. *Soc. Nat. Resour.* **30**, 377–392 (2017).

40. Babel, M. S. & Shinde, V. R. Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok. *Water Resour. Manag.* **25**, 1653–1676 (2011).

41. Villar-Navascués, R. A. & Pérez-Morales, A. *Factors Affecting Domestic Water Consumption on the Spanish Mediterranean Coastline* (Routledge, 2018).

42. Domene, E. & Saurí, D. Urbanisation and water consumption: influencing factors in the urbanisation and water consumption: influencing factors in the metropolitan region of Barcelona. https://doi.org/10.1080/00420980600749969 (2006).

43. Alharsha, I., Memon, F. A., Farmani, R. & Hussien, W. E. A. An investigation of domestic water consumption in Sirte, Libya. *Urban Water J.* **19**, 922–944 (2022).

44. Murwirapachena, G. Understanding household water-use behaviour in the city of Johannesburg, South Africa. *Water Policy* **23**, 1266–1283 (2021).

45. Raj, K. *Factors Influencing Urban Residential Water Consumption* (Institute for Social and Economic Change, 2020).

46. Zeroual, M., Hani, A. & Boustila, A. Assessing domestic factors determining water consumption in a semi-arid area (Sedrata City) using artificial neural networks and principal component analysis. https://doi.org/10.24425/jwld.2021.137115.Abstract (2021).

47. Kulinkina, A. V. et al. Piped water consumption in Ghana: a case study of temporal and spatial patterns of clean water demand relative to alternative water sources in rural small towns. *Sci. Total Environ.* **559**, 291–301 (2016).

48. Corbella, H. & i Pujol, D. S. What lies behind domestic water use?: a review essay on the drivers of domestic water consumption. *Bol. La Asoc. Geogr. Esp.* **50**, 297–314 (2009).

49. Dalhuisen, J. M., Florax, R. J. G. M., De Groot, H. L. F. & Nijkamp, P. Price and income elasticities of residential water demand: a meta-analysis. *Land Econ.* **79**, 292–307 (2003).

50. Fox, C., McIntosh, B. S. & Jeffrey, P. Classifying households for water demand forecasting using physical property characteristics. *Land Policy* **26**, 558–568 (2009).

51. Wentz, E. & Gober, P. Determinants of small-area water consumption for the city of phoenix, Arizona. *Water Resour. Manag.* **21**, 1849–1863 (2007).

52. Peters, L. A. H. & Chang, H. Urban water demand modeling: review of concepts, methods, and organizing principles urban water demand modeling: review of concepts, methods, and organizing principles. https://doi.org/10.1029/2010WR009624 (2011).

53. Inman, D. & Jeffrey, P. Implementation effectiveness a review of residential demand-side management tool performance and influences on implementation. https://doi.org/10.1080/15730620600961288 (2006).

54. Yousefi, P., Courtice, G., Naser, G. & Mohammadi, H. Nonlinear dynamic modeling of urban water consumption using chaotic approach (case study: City of Kelowna). *Water* **12**, 3 (2020).

55. Thakur, B. K., Bal, D. P., Nurujjaman, M. & Debnath, K. Developing a model for residential water demand in the Indian Himalayan Region of Ravangla, South Sikkim, India. *Groundw. Sustain. Dev.* **21**, 100923 (2023).

56. House-peters, L. A., Beach, L., & Pratt, B. & Chang, H. Effects of urban spatial structure, sociodemographics, and climate on residential water consumption in Hillsboro, Oregon. https://doi.org/10.1111/j.1752-1688.2009.00415.x (2010).

57. Adamo, N., Al-ansari, N. & Sissakian, V. K. How dams can affect freshwater issues in the Euphrates-Tigris Basins. *J. Earth Sci. Geotech. Eng*. 10, 43–76 (2020).

58. Rezaali, M., Quilty, J. & Karimi, A. Probabilistic urban water demand forecasting using wavelet-based machine learning models. *J. Hydrol.* **600**, 126358 (2021).

59. Power, N., Volker, R. & Stark, K. Deterministic models for predicting residential water consumption. *JAWRA J. Am. Water Resour. Assoc.* **17**, 1042–1049 (2007).

60. KDnuggets. Statistical modeling vs. machine learning: what's the difference? JPT. https://jpt.spe.org/statistical-modeling-vs-machine-learning-whats-difference?gclid=CjwKCAjwjOunBhB4EiwA94JWsHdBj-8GLiI69QktE__iNycmzpbaISukne42PnJCwGJVH5tSWZ6wGRoCG_IQAvD_BwE (2019).

61. Simplilearn. What are probabilistic models in machine learning?. https://www.simplilearn.com/ and https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-are-probabilistic-models#examples_of_probabilistic_models (2013).

62. Grespan, A., Garcia, J., Brikalski, M. P., Henning, E. & Kalbusch, A. Assessment of water consumption in households using statistical analysis and regression trees. *Sustain. Cities Soc.* **87**, 104186 (2022).

63. Alshaikhli, M., Aqeel, S., Valdeolmillos, N., Fathima, F. & Choe, P. A multi-linear regression model to predict the factors affecting water consumption in Qatar. *IOP Conf. Ser. Earth Environ. Sci.* **691**, 12004 (2021).

64. Stelzl, A., Pointl, M. & Fuchs-Hanusch, D. Estimating future peak water demand with a regression model considering climate indices. *Water* **13**, 14 (2021).

65. Lawens, M. & Mutsvangwa, C. Application of multiple regression analysis in projecting the water demand for the City of Cape Town. *Water Pract. Technol.* **13**, 705–711 (2018).

66. Kim, J. et al. Development of a deep learning-based prediction model for water consumption at the household level. *Water* **14**, 9 (2022).

67. Xiang, Z., Yan, J. & Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **56**, e2019WR025326 (2020).

68. Bata, M., Carriveau, R. & Ting, D. S.-K. Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. *Smart Water* **5**, 2 (2020).

69. Ibrahim, T. Omar, Y. & Maghraby, F. A. Water Demand Forecasting Using Machine Learning and Time Series Algorithms. In *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)* 325–329 (ESCI, 2020).

70. Almobarek, M. Monitoring water consumption using machine learning. In *Proc. 11th Annual International Conference on Industrial Engineering and Operations Management* 3493–3499 (IEOM Society International, 2021).

71. Javaid, N. Machine learning fundamentals: basic theory underlying the field of machine learning. Towards Data Science. https://towardsdatascience.com/machine-learning-basics-part-1-a36d38c7916 (2018).

72. Kanade, V. What is machine learning? Definition, types, applications, and trends for 2022. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/ (2022).

73. Marr, H. IX steps to a successful machine learning model. Acerta. https://acerta.ai/blog/six-steps-to-a-successful-machine-learning-model/ (2022).

74. Delua, J. Supervised vs. unsupervised learning: what's the difference? IBM. https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning (2021).

75. Donges, N. A complete guide to the random forest algorithm. Built In. https://builtin.com/data-science/random-forest-algorithm (2021).

76. Onesmus, M. Introduction to random forest in machine learning. https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/#:~:text. In a random forest, regression of the decision trees' class (2020).

77. Chaudhary, M. Random forest algorithm - how it works & why it's so effective. https://www.turing.com/kb/random-forest-algorithm (2023).

78. Sakshi, G. Regression vs. classification in machine learning: what's the difference? https://www.springboard.com/blog/data-science/regression-vs-classification/#:~:text. The most significant difference between types of machine learning algorithms. (2021).

79. Sruthi, E. R. Understand random forest algorithms with examples. https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/ (2021).

80. Anon. 10 steps to build and optimize a ML model. DEV Community. https://dev.to/mage_ai/10-steps-to-build-and-optimize-a-ml-model-4a3h (2021).

81. Bakshi, C. Random forest regression. Medium. https://levelup.gitconnected.com/random-forest-regression-209c0f354c84 (2020).

82. Simplilearn. What is R: overview, its applications and what is R used for. https://www.simplilearn.com/what-is-r-article (2021).

83. Patel, H. What is feature engineering — importance, tools and techniques for machine learning. Towards Data Science. https://towardsdatascience.com/what-is-feature-engineering-importance-tools-and-techniques-for-machine-learning (2021).

84. Kalashak, E. *Prediction of Water Consumption Using Machine Learning*. Mater's thesis, Østfold University (2021).

85. Poornima, P. & Boyapati, S. Prediction of water consumption using machine learning algorithm. In *ICCCE 2020* 891–908 (ICCCE, 2021).

86. Aggarwal, S. & Sehgal, S. Prediction of water consumption for New York city using machine learning. In *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)* 486–490 (SPIN, 2021).

87. Pandian, S. K-fold cross validation technique and its essentials. https://www.analyticsvidhya.com/blog/2022/02/k-fold-cross-validation-technique-and-its-essentials/ (2023).

88. Nellihela, P. What is K-fold cross validation? https://towardsdatascience.com/what-is-k-fold-cross-validation-5a7bb241d82f (2022).

89. DH-Consult. *Design Review, Construction Supervision and Contract Administration of Adama Town Water Supply and Sanitation Improvement Project* (DH-Consult, 2018).

90. Kanno, G. G. et al. Estimation of rainwater harvesting potential for emergency water demand in the era of COVID-19. The case of Dilla town, Southern, Ethiopia. *Environ. Challenges* https://doi.org/10.1016/j.envc.2021.100077 (2021).

91. Taher, T. M. Quantity and quality considerations of rooftop rainwater harvesting as a substantial resource to face water supply shortages. *Int. J. Water Resour. Arid Environ*. **3**, 1–10, (2014).

92. IWA. Alternative water resources: a review of concepts, solutions and experiences. https://iwa-network.org/publications/alternative-water-resources-a-review-of-concepts-solutions-and-experiences/ (2016).

93. Bolton, E., Berglund, E. & Goyal, R. The impact of demographic factors, beliefs, and social influences on residential water consumption and implications for non-price policies in urban India. *Water* **9**, 844 (2017).

94. Ouyang, Y., Wentz, E., Ruddell, B. & Harlan, S. A multi-scale analysis of single-family residential water use in the Phoenix Metropolitan Area. *JAWRA J. Am. Water Resour. Assoc.* **50**, 448–467 (2014).

95. Hussien, W. A., Memon, F. A. & Savic, D. A. Assessing and modelling the influence of household characteristics on per capita water consumption. *Water Resour. Manag.* **30**, 2931–2955 (2016).

96. Morote, Á., Hernández, M. & Rico, A. Causes of domestic water consumption trends in the city of Alicante: exploring the links between the housing bubble, the types of housing and the socio-economic factors. *Water* https://doi.org/10.3390/w8090374 (2016).

97. Millock, K. & Nauges, C. Household adoption of water-efficient equipment: the role of socio-economic factors, environmental attitudes and policy. *Environ. Resour. Econ.* **46**, 539–565 (2010).

98. Cavanagh, S. M., Hanemann, M. & Stavins, R. N. et al. Muffled price signals: household water demand under increasing-block prices. https://ssrn.com/abstract=317924 (2002).

99. Atkins, J. F. et al. The urban water metabolism of Cape Town: towards becoming a water sensitive city. *South Afr. J. Sci.* **117**, 1–11 (2021).

100. Chang, H., & Praskievicz, S. Sensitivity of urban water consumption to weather and climate variability at multiple temporal scales: the case of Portland, Oregon. *Int. J. Geospatial Environ. Res.* **1**, 7 (2014).

101. Hazeltine, B. "CHAPTER 8 - Water Supply," in *Field Guide to Appropriate Technology* B. Hazeltine and C. Bull, Eds. San Diego: Academic Press, 2003, pp. 731–809. https://doi.org/10.1016/B978-012335185-2/50052-8.

102. Adama stuctural plan. Adama city structural plan report. (2020).

103. Prabhakaran, S. How to How to detect outliers with z-score outliers with z-score. https://www.machinelearningplus.com/machine-learning/how-to-detect-outliers-with-z-score/#google_vignette (2023).

104. Iden, W. Z-score and how it's used to determine an outlier. https://medium.com/clarusway/z-score-and-how-its-used-to-determine-an-outlier-642110f3b482 (2021).

105. Harika, B. Detecting and treating outliers. Treating the odd one out! https://www.analyticsvidhya.com/blog/2021/05/detecting-and-treating-outliers-treating-the-odd-one-out/#2e43 (2023).

106. Mayank, B. Machine learning steps: a complete guide! https://www.simplilearn.com/tutorials/machine-learning-tutorial/machine-learning-steps#machine_learning_steps (2023).

107. Bernardo, I. Data science tutorials — training a random forest in R. Medium. https://towardsdatascience.com/data-science-tutorials-training-a-random-forest-in-r-a883cc1bacd1 (2022).

108. Brownlee, J. Train-test split for evaluating machine learning algorithms. https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/ (2020).

109. Jordan, J. Evaluating a machine learning model. https://www.jeremyjordan.me/evaluating-a-machine-learning-model/ (2017).

110. Joseph, V. R. Optimal ratio for data splitting. *Stat. Anal. Data Min. ASA Data Sci. J.* **15**, 531–538 (2022).

111. Brownlee, J. A gentle introduction to the bootstrap method. https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/ (2018).

112. Shruti, M. Introduction to random forest in R. https://www.simplilearn.com/tutorials/data-science-tutorial/random-forest-in-r (2023).

113. Zulaikha, L. A comprehensive guide to random forest in R. https://www.edureka.co/blog/random-forest-classifier (2020).

Acknowledgements

Author contributions
All authors conceived of the presented idea and contributed to the study's conception and design. Data collection, analysis and interpretation of results, and manuscript preparation were performed by A.T. Meanwhile, T.A. and E.G. verified the analytical methods and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Competing interests
The authors declare no competing interests.

Consent to participate
This study was conducted after receiving ethical approval from the Ethiopian Institute of Architecture, Building Construction and City Development, Addis Ababa University. Participants were all informed about the purpose of the study, the reason for their recruitment, and the time needed to complete the questionnaire. Written informed consent was obtained from the participants before commencing data collection. They also informed that their participation is voluntary and they can terminate the consent at any time they thought it is inappropriate. All participants noticed that the information collected would be treated confidentially and analyzed without any personal identification. Besides, they were provided the contact address in case they needed additional information.

Additional information
Correspondence and requests for materials should be addressed to Abraha Tesfay Abraha.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.