MANAGING WATER RESOURCES IN KENYA'S UPPER TANA RIVER BASIN

Student Team: Sonia Aronson, Ahmed Elauri, Clara Hu, Andrew Johnson, Liam Maxwell, Adriana Rangel, Nikki Zhang

Advisor: *Edith de Guzman*

A SENIOR PRACTICUM PROJECT WITH THE UCLA INSTITUTE OF THE ENVIRONMENT AND SUSTAINABILITY

Table of Contents

Abstract

The Tana River provides half of Kenya's power and nearly all of Nairobi's water supply. Kenya's population has grown rapidly in the last half-century, with more growth expected in the years to come. Preserving the health of the Upper Tana River watershed is critical to help Kenya meet the demands of an increasing population especially as climate variability increases. The Nature Conservancy established the Upper Tana-Nairobi Water Fund to protect and restore the quality and supply of water in the Upper Tana River basin. The Water Fund brings stakeholders in the Upper Tana Basin together to enact interventions, or sustainable agricultural practices that improve water flow and quality for users downstream. We used streamflow, land cover and precipitation data to analyze the impact of The Nature Conservancy's interventions. We have investigated the efficacy of currently employed interventions in improving water availability in the watershed. While current interventions are effective, we have made multiple recommendations for the Water Fund moving forward and brought to light the current limitations in evaluating the watershed.

1. Introduction

According to estimates from the World Health Organization, over half the world's population will live in water-stressed areas by 2025 (Garthwaite, 2019). Climate change, population growth, and pollution are depleting water sources across the globe, threatening the livelihoods of millions of people. One particularly high-risk area is Kenya, whose population has quintupled since the 1970s, growing from 11 million to 52 million, with the United Nations estimating that by 2050, it will reach 95 million (Anderson, 2021). Nairobi, Kenya's capital, is growing at an annual rate of 2.8%, from half a million residents in 1971 to nearly 4.5 million today (Anderson, 2021).Water is in short supply, with a 2013 survey reporting that over 60% of Nairobi's residents are water insecure (TNC, 2015). A steady and healthy water supply provides valuable ecosystem services and is integral to support Kenya's growing population.

The Tana River supplies 95% of Nairobi's water and half of the country's energy (TNC, 2019). It is Kenya's longest river, flowing more than 1500 kilometers from the Aberdare Mountains in the north before terminating in a delta at the Indian Ocean (Langat et al., 2017). The entire Tana River Basin covers ~95,000 km^2 and is home to millions of wildlife species as well local communities (TNC, 2015). Since the 1970s, the number of small subsistence farms in the upper basin has skyrocketed (TNC, 2015). The dual effects of climate change and rapid population growth are leading to increased sediment erosion in the Upper Tana River Basin, reducing the capacity of reservoirs, and increasing water treatment costs (TNC, 2019). Water scarcity is predicted to disproportionately affect communities that have least contributed to climate change, such as the millions of farmers and fishermen who rely on the Tana River and its tributaries for survival and economic prosperity. Taking action now to prevent further damage and help these communities preserve their water resources is critical to ensure the region's water security for the future. To this end, The Nature Conservancy created a water fund, an organization that brings together public and private stakeholders to address threats to water security at their source through targeted, long-term investments in watershed conservation and management activities (TNC, 2015). In 2015, The Nature Conservancy established the Upper Tana-Nairobi Water Fund to conserve the watershed's health and protect Kenya's water supply (TNC, 2019). The Water Fund aims to improve water quality and quantity for all stakeholders, including farmers, Kenya's hydropower generators, and the city of Nairobi. The Nature Conservancy partnered with IBM to improve water security in this at-risk region and evaluate the effectiveness of the Water Fund.

In this project, we developed an open water data repository to help stakeholders in the Tana River region better understand and access available resources and analyzed the impact of TNC's interventions in the region using IBM's data analysis software. We gathered and analyzed publicly available water quality data, including turbidity, and flow rate, and socioeconomic characteristics to explore relationships between environmental and social factors that may affect water availability in the area.

2. Background

2.1 The Upper Tana River Basin

2.1.1 Physical Characteristics

Covering roughly 100,000 square kilometers in southeastern Kenya, the Tana River Basin is Kenya's largest watershed. Its rivers originate from two high elevation focal points, Mount Kenya and the Aberdare Mountains, flowing over 1,000 kilometers before terminating in a large delta at Ungwana Bay in the Indian Ocean (Langat et. al, 2017). The Tana River's main tributaries are the Chania, Thika, Sagana, Thiba, and Mutonga (Langat et. al, 2017). These rivers form sub-basins within the larger river basin. Some of the tributary rivers are perennial while others are dry for part of the year (Knoop, 2012). The Tana River Basin can be split into three separate sub-regions: an upland forested area with more rainfall and a drier, less mountainous mid-basin that leads into the lower catchment, and the Tana Delta (Baker et al. 2015).

Mountain area soils are mainly volcanic ash soils, while soils in lower elevations are derived from metamorphic rocks, resulting in fertile clay and poorer leached clay (Dijkshoorn et al. 2010). Lower slopes are mainly cultivated or forested while high elevation mountain areas are moorlands with U-shaped and shallower valleys created by glaciations (Veldkamp et al. 2011).

2.1.2 Precipitation

Over 80% of the Tana River area is classified as arid and semi-arid (Water Resources Authority, n.d.). The area's average annual precipitation is 679 millimeters but varies significantly both spatially and temporally (Knoop et. al, 2012). For example, average annual precipitation in the highlands is roughly 2,200 millimeters, but only 370 millimeters in the lower delta (Langat et. al, 2017). Temporally, the basin follows a bimodal distribution; it has two wet and two dry seasons. The first wet season spans April to early June, and the second falls between November and December. Approximately 92% of the basin's rainfall occurs during the wet seasons. The remainder of the year, mid-December to March and June to October, are considered dry seasons (Langat et. al, 2017).

Excessive flooding is a problem during the wet seasons, especially in low-lying areas. Flooding picks up large amounts of soil, leading to increased erosion and decreased water quality. This results in longer, more expensive purification in water treatment facilities, which can serve as a limiting factor to the availability of clean water (Knoop et. al, 2012). Flooding can also increase the risk of mosquito-borne diseases. A study by Otieno et al. (2019) showed that flooding resulted in higher risks of disease outbreak. The increase in stagnant water provides better conditions for mosquito larvae that ultimately mature to infect people and livestock with deadly diseases such as the Rift Valley fever. While increases in sediment accumulation due to greater water retention and surface runoff is the main concern during the wet seasons, finding water sources for domestic and agricultural use during the dry seasons becomes the challenge because often, natural and stored water are insufficient for Kenya's growing population during dry spells.

2.1.3 Land Use and Land Cover

Agriculture constitutes a critical land use for this region. In the higher, wetter regions of the Tana River Basin, cash crop agriculture is common and consists primarily of tea, coffee, and maize (Knoop et. al, 2012; Langat et. al, 2017). In the mid and low elevations, irrigated crops such as flowers, fruit, and vegetables as well as livestock are produced (Huinink & Droogers 2015; Water Resources Authority, n.d.). The most popular method of farming in the Tana River Basin is flood recession farming, which relies on natural irrigation and fertilization from regular floods. An estimated 115,000 farmers in the basin rely on this method to grow crops (Leauthaud et al., 2013). Flood recession farming has a very high net return to energy expenditure because the only human inputs are land and labor as farmers rely upon rainfall and flooding instead of irrigation (Saarnak 2003). Rainfed subsistence agriculture now constitutes over 60% of the land use in the Upper Tana Basin (Hunink et al., 2013). Currently 64,425 hectares of the basin are under irrigation, and an estimated 292,100 additional hectares are slated for irrigation at the lower Tana by the year 2030 (Langat et. al, 2017).

2.1.4 Hydropower

Approximately 50% of Kenya's total energy needs are met by hydropower production, with much of that attributed to the five dams on the Tana River (Baker et al., 2015). These five dams, called the "Seven Forks Project," include the Kindaruma, Kiambere, Kamburu, Gitaru and Masinga. The dams regulate the river, providing water for irrigated crops as well as 75% of Kenya's total hydropower output (TNC, 2015). The Masinga dam is the largest and most significant in terms of regulating the river. An issue the dams face is that high sediment loads upstream can be abrasive to hydropower generating plants, causing energy losses and increased maintenance costs (Baker et al., 2015). While the dams provide water and power, the reservoirs created can increase sediment loads downstream, increasing operation costs and causing reductions in peak flows downstream (Kauffman, 2007; Leauthaud and Duvail, 2013).

2.2 Climate Change and Land Use Change in the Tana River Region

2.2.1 Climate Change and Climate Variability

The IPCC projects the global median temperature to rise by between 1.4 and 5.5°C by the end of this century, with temperatures in Africa rising the fastest, projected to increase by more than 2°C by 2050 and more than 4°C by 2100 (Adhikari et al., 2015). Along with temperature changes, median precipitation is expected to change between -2 and +20% (Adhikari et al., 2015). Changes in temperature and precipitation will likely alter water resource availability in the area.

Variability in weather and precipitation is expected to lead to intensifications of flood and drought frequencies and magnitudes, but the extent of change is uncertain (Adhikari et al., 2015). Much of Kenya is already considered water-scarce and will likely continue to be although precipitation is actually projected to increase by at least 10% in the country by 2100 (Adhikari et al., 2015). The IPCC projects precipitation will increase by 5-20% and decrease by 5-10% during wet and dry seasons respectively (Mango et al., 2011). Both average annual precipitation and drought frequency are expected to increase (Adhikari et al, 2015), and periods of severe drought are predicted to alternate with periods of heavy rain, leading to greater variability (Nakaegawa and Wachana, 2012). Historically dry and wet months of the year are also predicted to shift, and anecdotal evidence suggests such shifts may already be occurring (Gathagu, 2021). Seasonal shifts are expected to be a source of confusion and difficulty for farmers.

2.2.2. Predicted Effects of Climate Change

Variability in precipitation due to climate change may cause issues in water availability and increase demand for irrigation water. In addition, rising temperatures will increase saturation vapor pressure in air as well as evaporative demand, leading to increased water and heat stress (Adhikari et al., 2015). Water and heat stress both limit the land available for growing. In Kenya, around 75% of the population relies on agriculture, but only 20% of the land is currently arable, with the percentage of arable land projected to decrease in the future (Cheruto et al., 2016).

The dual impacts of climate change and deforestation in East Africa are expected to increase the amount of total suspended solids in the Tana River and its tributaries. The Tana River Basin may become more vulnerable due to increased peak flows and hillslope erosion that also cause land degradation (Mango et al., 2011). Water resources are also expected to face degradation from anthropogenic pollution, such as from agrochemicals, and from siltation due to increased soil erosion caused by irrigation, cultivation, and overgrazing (Mango et al., 2011). In Africa, groundwater is a main source of irrigation and drinking water (MacDonald et al., 2012), but information on groundwater sources and storage in the region is lacking.

The effects of climate change are significant threats to agriculture and soil in East Africa. Although precipitation is projected to increase, seasonal variation may limit water availability. Climate change may lead to higher susceptibility to poverty due to reductions in crop yields and food availability. Water availability and soil moisture are two primary limiting factors on agricultural productivity in the area and crucial in determining climate change effects on agriculture. The region is extremely vulnerable during dry spells and the dry seasons due to a lack of irrigation. With increased climate variability, much of the land may become less suitable for growing crops (Adhikari et al., 2015). The growing season length is projected to decrease after the middle of the century as heat and water stress limit soil suitable for crops (Adhikari et al., 2015). In the higher-altitude, mountainous regions of Kenya, temperature is the limiting factor, so conditions may become more suitable for agriculture as temperatures rise. However, the situation is unfavorable in the lowlands, where warmer temperatures may increase water stress, leading to decreased soil suitability (Adhikari et al., 2015). When periods of increased rainfall follow periods of drought and dry soil conditions, soil erosion is expected to accelerate, shrinking areas suitable for agriculture (De Pauw and Ramasamy, 2020). In more arid areas, erosion and removal of topsoil contribute to diminished soil moisture storage capacity, leaving the land even more vulnerable to soil loss. Soil moisture can be measured on site using a soil moisture sensor or estimated using remote sensing.

Deep percolation may occur more frequently in arid areas since rainy seasons generally coincide with lower temperatures and potential evaporation (Okello et al., 2020). Under the "best case" emissions scenario, RCP 2.6, where global CO_2 emissions have already peaked and will decline to zero by 2080, average deep soil percolation is projected to increase by 14% and average soil water content by 1%. Under the "business as usual" RCP 8.5 emissions scenario, deep percolation is projected to increase by 10% and average soil water content is projected to decrease by 2% by the end of the century (Okello et al., 2020). Precipitation increases may impact soil moisture and water storage positively in the first half of the century, but overall impacts of increased warming for future generations are uncertain. A 2010 assessment of historical climate trends indicates a threat to crop surplus regions in central Kenya and the Upper Tana River Basin (Funk et al.). Extreme climate events, such as floods and droughts, already cost the Kenyan economy 2.4% of GDP (KES 16 billion) per year due to crop and livestock losses, reduced hydropower generation, and declines in industrial production (Mogaka et al., 2006).

2.2.3 Land Use Change

Population Growth

While much of Kenya's population is currently rural and reliant on agriculture, the rate of urbanization is increasing rapidly. Kenya's current population is around 54 million, and the annual growth rate is expected to become 4.3% by 2025 (Wangai et al., 2019), one of the highest in the world. By 2030, 33% of the population is projected to live in urban areas. By 2050, the urbanization rate is estimated to reach 3.8%, aided by the development of tarmac roads that the government is planning to build (Wangai et al., 2019). Urban land cover is predicted to increase by 590% in Africa by 2030; it is projected to increase by a whopping 1900% in East Africa, mostly around the northern part of Lake Victoria in Kenya (Seto et al., 2012). Africa as a whole is expected to have the world's highest urbanization rate, primarily occurring in regions that were previously mostly unaffected by urban development. The mountainous region of the Upper Tana River Basin is a hotspot for population growth (Baker et al., 2015). Since 1970, in the forested headwaters region around Mount Kenya human settlement has increased by ~60%.

Impacts of Population Growth on Land Use

Kenya's natural land cover includes wetlands, grasslands, savannahs, forests and deserts. The greatest proportion of land in Kenya is devoted to agriculture and is expected to increase with population growth. High demand for arable land endangers forests as they are cleared for agriculture and settlements, with portions of land already set aside for industrialization and urbanization (Mango et al., 2011). Water, air, and land resources are expected to deteriorate due to forest degradation and deforestation, along with harmful land use and agricultural practices such as overgrazing and poor irrigation techniques (Seto et al., 2012). Rapid population growth has implications for natural resources. Forests and wetter zones have high risks of degradation due to higher population pressure, and the Tana River region is no exception.

The growing population in the Upper Tana River Basin is leading to increased land degradation, as land shortages result in smaller plot sizes and more intensive agricultural practices (Tanui, 2006). The forests at the headwaters of the Tana River are important natural infrastructure features because they control the amount and timing of flows within the catchment, but deforestation to support smallholder farms in the region is increasing rapidly (Baker et al., 2015). Unsustainable farming practices send sediment into the river, resulting in higher costs for water treatment, lower water levels, and lower hydropower output. A survey found that 77% of inhabitants in the Thikia-Chania catchment said erosion occurs on their farms and 79% had observed a deterioration in water quality in rainy seasons over the last five years (Gathagu et al. 2018). Water security will only become more challenging as climate change brings increasingly unpredictable rainfall and the country's population continues to climb (TNC, 2019).

2.3 Water Funds

2.3.1 What is a Water Fund?

A water fund is an organization that assembles public and private stakeholders and non-governmental organizations (NGOs) to establish programs for improving water quality and economic activity in a water stressed area. Water quality is maintained by preserving ecosystem services upstream to prevent contamination or reduced flows downstream (Joslin & Jepson, 2018). Each water fund has objectives specific to its location. Most focus on improving water quality and quantity and/or local ecological health (Joslin & Jepson, 2018; Ozment et al., 2016).

Upstream, water funds work to prevent pollution and environmental degradation from agricultural practices and industry to maximize ecological services, such as increased pollinators, healthier crops, reduced wildfire risk, and thriving aquatic populations, across the entire basin (Apse et al., 2016). Improved water quality upstream should reduce water treatment and importation costs, leading to healthier, more economically secure communities across the river basin (Cheatham, 2020). Water fund activities downstream attempt to maintain water quality and availability for towns and cities in lower parts of a watershed. For many cities with burgeoning population growth, such as Nairobi, this is critical for public health (Abell et al., 2017).

Water funds function as funding and organizing mechanisms for watershed protection (Bremer et al, 2015). An annual budget, usually ranging around US$2-4 million, is used for watershed projects, including monitoring, permitting, and collecting data (Bremer et al., 2015). Funding also comes from the water fund's programs. As profits and economic gains enter downstream communities, they are expected to give back to the water fund. The money is reinvested into the watershed through projects for maintaining healthy ecological services upstream (Ozment et al., 2016). Eventually, water funds are expected to become economically independent through endowment from community programs, but they initially require support from other groups, including NGOs, governments, utilities, and corporations (Zyla et al., 2018).Communication is an essential element of a successful water fund. A watershed is a shared natural resource that provides many valuable ecosystem services, so working together is important even when there are conflicting interests among stakeholders (Ozment et al., 2016).

Figure 1: Area covered by the Upper Tana-Nairobi Water Fund

Image courtesy of The Nature Conservancy

2.3.2. The Upper Tana-Nairobi Water Fund (UTNWF)

The Nature Conservancy currently has over 30 Water Funds around the world that have each raised $10 million or more for water conservation (TNC, 2019). The Upper Tana-Nairobi Water Fund (UTNWF), shown in Figure 1, was established in 2015 with the primary goal of improving farming practices in the watershed (TNC, 2019). It has since been followed by the launch of the South African Greater Cape Town Water Fund in 2018. In establishing the UTNWF, TNC followed their five-step plan to implement and maintain a successful Water

Fund (Zyla et al., 2018), including doing a variety of physiographic, hydrologic and socio-economic studies on the Upper Tana Region. The project area of the Tana River Water Fund covers 2.4 million acres (977,936 hectares) and focuses on water and soil conservation strategies in three of the main sub- catchments of the Upper Tana River watershed: Thika-Chania, Maragua and Sagana-Gura (Hess et al. 2018).

Socio-Economic Findings

In 2017, TNC carried out a survey of 1,002 randomly selected households in the Tana River region to gather socio-economic information (Hess et al., 2018). Farming was the leading occupation for household heads (70%). Of the 97% of households with agricultural land, over half said their land bordered a river or stream (Hess et al., 2018). Overall, 63% had heard of the term climate change, and 91% had perceived changes in weather patterns since they were young. More than half of those who had witnessed changed weather patterns said they altered their agricultural practices, mostly shifting to more drought-resistant, fast-maturing crops or seed.

Agriculture accounts for approximately 25% of Kenya's gross domestic product (GDP). The detrimental effects of climate change, which are expected to worsen, have already been documented in the region. Through changes in regional temperature, precipitation frequency and precipitation intensity, key agricultural crops will be affected, which is why this project aims to understand and predict how climate will change. Our projects are geared to help farmers in the region prepare for the coming years.

Interventions

In our context, interventions refer to agricultural conservation practices that can reduce suspended sediments in waterways and increase water flow during dry seasons (TNC, 2015). Examples include vegetation buffer zones, agroforestry, terracing of steep farmland, grass buffer strips in farmlands, reforestation, and mitigation of erosion from dirt roads (TNC, 2015). Some practices have already been implemented in the Tana River region with positive results.

Evaluating the progress of TNC intervention work in the region is important, as it will allow key stakeholders to identify successful aspects of interventions and make informed decisions about proposed next steps. Each year, the TNC and its various committees and key stakeholders publish an annual progress report, which highlights its accomplishments for each working year.

This report provides a qualitative overview of some achievements detailed in the progress reports, including social and hydrological benefits. Specifically, social benefits can be defined as outcomes from the watershed which resulted in positive improvements which are not directly related to improved hydrological outcomes. Hydrological benefits can be defined as improved hydrological metrics and outcomes which contribute to the success of the water fund's overarching goal of improving water-resource management and conservation efforts

Moreover, this report includes a quantitative analysis between priority micro watersheds to understand the relationship between intervention and control zones. The overall objective of the UTNWF is to make the Upper Tana River basin well-conserved by improving water quality in the whole basin, maintaining regular water flow rates throughout the year, enhancing ecosystem services such as increasing biodiversity and ensuring food security, and overall improving the overall quality of life for local communities. To achieve this objective, UTNWF has done numerous interventions. And we categorized the interventions into two types: hydrological benefits and social benefits.

Table 1. Descriptions of Hydrologic Interventions in the Upper Tana-Nairobi Water Fund

Hydrological Interventions	**Description**
Rainwater Harvesting Water Pans	Water pans, rainwater harvesting tools, are one of the techniques to promote climate-resilient agricultural production systems to provide increased water supply for agricultural purposes during the dry season. Farmers are able to use water pans to store water during the dry seasons for irrigation to maintain high crop yield.
Low-head Drip Irrigation Equipment	The equipment enables the use of the harvested water from the water pans more efficiently, improving water efficiency during dry seasons.
Agroforestry	Agroforestry is a land use management system in which trees are grown around farmland. Promoting agroforestry at the landscape level is done to increase carbon stock in the Upper Tana catchment.
Riparian Zones, Wetlands Protection, and Restoration	In order to assess crops and livestock within the identified wetlands, establishing a baseline map and a wetland biodiversity map in the Upper Tana. The findings will inform potential sustainable interventions in the wetland.
Rural Road Improvement	In order to reduce the negative impacts of runoff from rural roads and to decrease sediment loads in the runoff, UTNWF is working with the Kenya Rural Roads Authority to integrate rainwater harvesting from road runoff.
River Gauging Monitoring Stations	The stations are used to monitor the improvements in water quality and water quantity as well as the water levels and river flows of all rivers. Prior to establishment of the water fund, manual collection of water data was conducted twice a day. Currently, 28 river stations transmit data to Ndakaini Dam officials every two hours
WHO Turbidity Standards Reached	Turbidity levels are less than 5 NTU.
Social Interventions	**Description**
Mobile Phone SMS Platform	The platform aims to reach farmers in the Upper Tana watershed with conservation messages. These messages aim to modify how farmers manage soil erosion on their lands by sending out conservation tips

	through SMS. Overall, enrollment into the platform is projected to increase, and 29,398 farmers are applying soil conservation and water-saving methods.
IT and Scientific Support	Knowledge management and learning systems are implemented. Outreach conducted through surveys, training, and mobile phone usage.
Rainforest Alliance certification	The Rainforest Alliance certification program trains farmers to produce products using methods that support sustainability. Over 8,500 coffee farmers have achieved Rainforest Alliance certification, while 70,000 farmers are in the process of obtaining certification

Hunink et al. (2013) found that natural areas in higher mountain areas contributed very little sediment (3%) compared to cultivated and grazed areas, and that the most sediment is produced on steeply sloped agricultural areas. Areas where coffee is grown produced the most sediment while areas growing cash crops like tea and maize generally eroded at similar rates. The researchers simulated impacts of implementing contour strips and check dams to reduce erosion and reservoir sedimentation. The simulations showed that if these measures were implemented in all areas where coffee, maize and tea are grown, total sediment inflow to the two main downstream reservoirs could be reduced by 47%. The analysis found that if the local farmers' unions, the Water Resources User Associations (WRUAs), were successfully involved in implementing and maintaining erosion control measures, sediment yield to reservoirs could be reduced by up to around 25% and reservoir life expectancy would increase by the same amount.

UTNWF Economic Return on Investment

The Nature Conservancy estimates that by 2025, water funds can be capable of providing 70 million people with water security, improving livelihoods for 150,000 rural community members, and protecting 2 million hectares of freshwater in Latin America, Asia, and Africa (Cheatham, 2020). Scientists have found that for every dollar invested in conservation strategies in the Tana River watershed two dollars will be saved in costs for correcting impacts on water supply and energy production (TNC, 2019). Overall, a US$10 million investment in Water Fund interventions is expected to return US$21.5 million in economic benefits over the 30-year timeframe, with many stakeholders expected to profit (TNC, 2015).

2.4 Stakeholders

2.4.1 Introduction and Definitions of Stakeholders

Stakeholders are defined as people who utilize and manage the natural resources (Buanes et al., 2004). Stakeholders of a water fund can be very diverse, varying from policymakers and governments to small-scale farmers and local communities. For a water fund to be successful in the long run, all stakeholders must benefit. Upstream communities are incentivized to work on projects that keep ecological services healthy (Joslin and Jepsin, 2018), and downstream communities save money through reducing importation and treatment costs and other benefits, such as reduced healthcare costs (Abell et al., 2017). In a secure water fund, the watershed health, which includes habitats, local economies, and human health, improves. Generally, stakeholders who benefit the most are expected to contribute the most, although benefits are often widely distributed throughout the community (Ozment et al., 2016). The relationship in contributions between the water fund and stakeholders is illustrated in Figure 2.

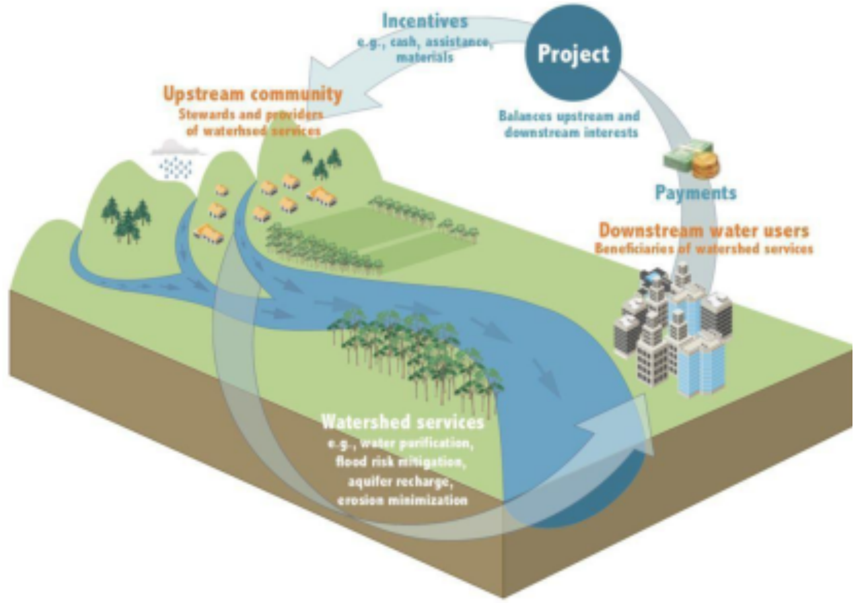


Figure 2: Transfer of money among different members of a water fund (Zyla et al. 2018)

2.4.2 Upper Tana-Nairobi Water Fund Stakeholders

Local communities, the Kenyan government, and public institutions all affect and are affected by the Tana River; they are all considered stakeholders of the Upper Tana-Nairobi Water Fund. They are categorized into two groups— formal and informal stakeholders. Formal stakeholders, such as The Nature Conservancy, are organized or institutionalized, while informal stakeholders, such as local fishermen, are not or are insufficiently organized (Beukering, 2015). TNC identified four main stakeholder groups with different interests for the UTNWF: municipal water suppliers, agricultural producers, hydropower operators, and government agencies (TNC, 2015).

Agricultural Producers

Agriculture and fishing are major components of Kenya's economy. In rural areas, people depend on farming and fishing as a direct food source and income (Zwarts, 2006). In Tana River County, 86% of people are farmers or fishermen. Farmers in the lower catchment particularly rely on consistent flooding since many use the flood recession method of subsistence agriculture (Knoop et al., 2012). Farmers are most interested in improving production yield, and any project or regulation changing the flood regime will directly impact soil fertility and crop production. For fishermen in the Tana River Basin, fishing occurs in both the Tana River itself and in inundation zones (Leauthaud et al., 2013). Inundation zones provide a nutrient-rich environment that can serve as a natural nursery for fish growing. Based on a study conducted by the Institute for Environmental Studies in the Tana River Basin, flow rate and production yield in fisheries have a positive association. Water level and discharge are limiting factors for fishing (Beukering, 2015), so any intervention disturbing river discharge and water levels would affect fisheries as well as farming. Agricultural yield is a function of cumulative action; sediment accumulates over time leading to a concave-down curve for yield increase over time (TNC, 2015).

Water Suppliers

The Nairobi City Water and Sewerage Company (NCWSC) is the major water and sewerage service provider for Nairobi. Reducing sediment concentrations to NCWSC has multiple benefits, including avoided flocculant costs, avoided electricity costs, and avoided loss of revenue from lost water (TNC, 2015). The Upper Tana-Nairobi Water Fund is predicted to result in approximately US$250,000 of cost savings a year for NCWSC stemming from avoided filtration, lowered energy consumption, reduced sludge disposal costs, and fewer shutdown days (TNC, 2015).

Energy Suppliers

The Kenya Electricity Generating Company (KenGen) is the leading electric power generation company in Kenya and is solely responsible for hydropower generation in the Tana River Basin (Baker et al, 2015). KenGen's benefits from conservation interventions include increased water yield, avoided service interruptions, avoided dredging costs for small upstream dams, and improved storage capacity. The UTNWF is predicted to increase KenGen's annual revenue by over US$600,000 as a result of increased power generation and avoided shutdowns and spillages.

Government Agencies

Key government agencies for this region include Kenya's Ministry of Agriculture, Livestock and Fisheries (MALF), the Ministry of Energy & Petroleum (MEP), and the Ministry of Health. According to the MALF's Chief Engineer, "Kenya is currently too dependent on the import of food," and sustainable natural resource management would ensure food security, improving the livelihood of Kenyans (KIPPRA, 2007). The MALF is devoted to improving food security through commercializing Kenya's agricultural production system. The MEP is interested in programs and legislation that increase energy supply and decrease energy tariffs in Kenya. For the Tana River, the MEP focuses on the construction and implementation of existing and future dams that benefit communities with hydroelectricity (Moe, 2012). Along with water availability, the Ministry of Health is interested in health risks and safe drinking water from the Tana River. The Ministry of Health's Chief Officer stated that "public facilities such as medical practices are [often] not sufficient for the number of people living there" (Beukering, 2015). Without enough medical facilities and water treatment systems, higher risks of waterborne diseases and health issues arise. The Ministry of Health is thus invested in the promise of provision of ample clean water. The UTNWF is predicted to improve water quality and decrease waterborne pathogens for more than half a million people (TNC, 2015).

2.5 Hydrologic Modeling

Hydrologic models predict how climate change and events like hurricanes and droughts will impact water sources. Models use river basin and watershed characteristics like size, shape, and water quantity and quality to sdetermine a water source's depth, flow rate, and area coverage.

Hydrologic models can be simplified into a water budget, an equation that accounts for the flows of water into and out of the system (Wang, 2014). For a drainage basin, water inflow is equal to the outflow, and a water budget expression breaks down the inflow and outflow to individual components based on changes in water storage in the atmosphere, land surface, and subsurface. These components account for precipitation, infiltration, evapotranspiration, baseflow, and human water demands. Hydrologists have invented a multitude of detailed techniques and technologies to measure and evaluate the individual components of a water budget in order to create an accurate simulation of water movement in watersheds or river basins. Several datasets are required as inputs for any computer-based hydrological analysis program, including drainage basin characteristics such as size, shape, slope, land use, soil type, surface infiltration rate, and storage; stream channel characteristics such as the geographic configuration of streams in the basin; and meteorological characteristics such as precipitation type, rate and amount (MDOT, 2006). In many hydrologic models, the watershed is split into sub-basins. These are then split into hydrologic response units (HRUs) based on similarities in land-cover, soil, and management (Antonopoulos et al., 2015). After the characteristics are added to the model, the programs use equations for each process in the hydrologic cycle to simulate water distribution, flow, and quantity in the study area.

2.5.1 SWAT Case Studies Modeling of the Tana River Basin

A common tool for hydrologic analysis is the Soil and Water Assessment Tool (SWAT), which is used to assess soil erosion and water-management practices in watersheds and large river basins (Arnold et al., 1998). SWAT represents components of the hydrological cycle, including rainfall, snow-cover and snowmelt, interception storage, surface runoff, infiltration, evaporation, lateral flow, percolation, pond and reservoir water balances, shallow and deep aquifers, and channel routing. SWAT was used by TNC to inform conservation practices in the Tana River Basin during the Water Fund's deployment. SWAT was also used to understand the impacts of contour farming on water and sediment yield in the Thikia-Chania catchment (Gathagu et al., 2018) and quantify soil erosion and reservoir sedimentation rates for targeted interventions (Hunink et al., 2013). TNC (2015) used SWAT to assess biophysical impacts of erosion and flow rates in the Upper Tana after implementing sustainable farming interventions.

Originally, an updated SWAT model was planned to evaluate TNC interventions for this project. The model would examine water quantity and sediment transportation change over time in the micro watersheds of interest. However, constructing and calibrating an accurate model can take years. Given the time restraints for this project and the fact that TNC has a dedicated team for SWAT modelling, the strategy for evaluating interventions shifted to statistical analyses.

2.5.2 Conceptual Model of Tana River Basin

The Upper Tana Nairobi River Basin has multiple inputs, outputs, and uses. A proper picture of the Tana and its uses is foundational to understanding and building upon any efforts to improve conditions in the basin. In order to gain an accurate understanding of the moving parts of the Tana's water uses, our team developed a conceptual model (Figure 3) which visually illustrates these varied components. The conceptual model provides a large-scale overview of the status of the Upper Tana River Water Fund and is designed to allow external parties to easily understand the dynamics of the Upper Tana River Basin and where the TNC interventions are put into place. It should be noted that the model is not geographically based, meaning that the flow of the graph does not dictate that its entities are upstream or downstream of one another.

In addition to the conceptual model, we also created a Feedback Loop Diagram of the Upper Tana River Basin to demonstrate how variables in the basin are connected to each other (Appendix A).

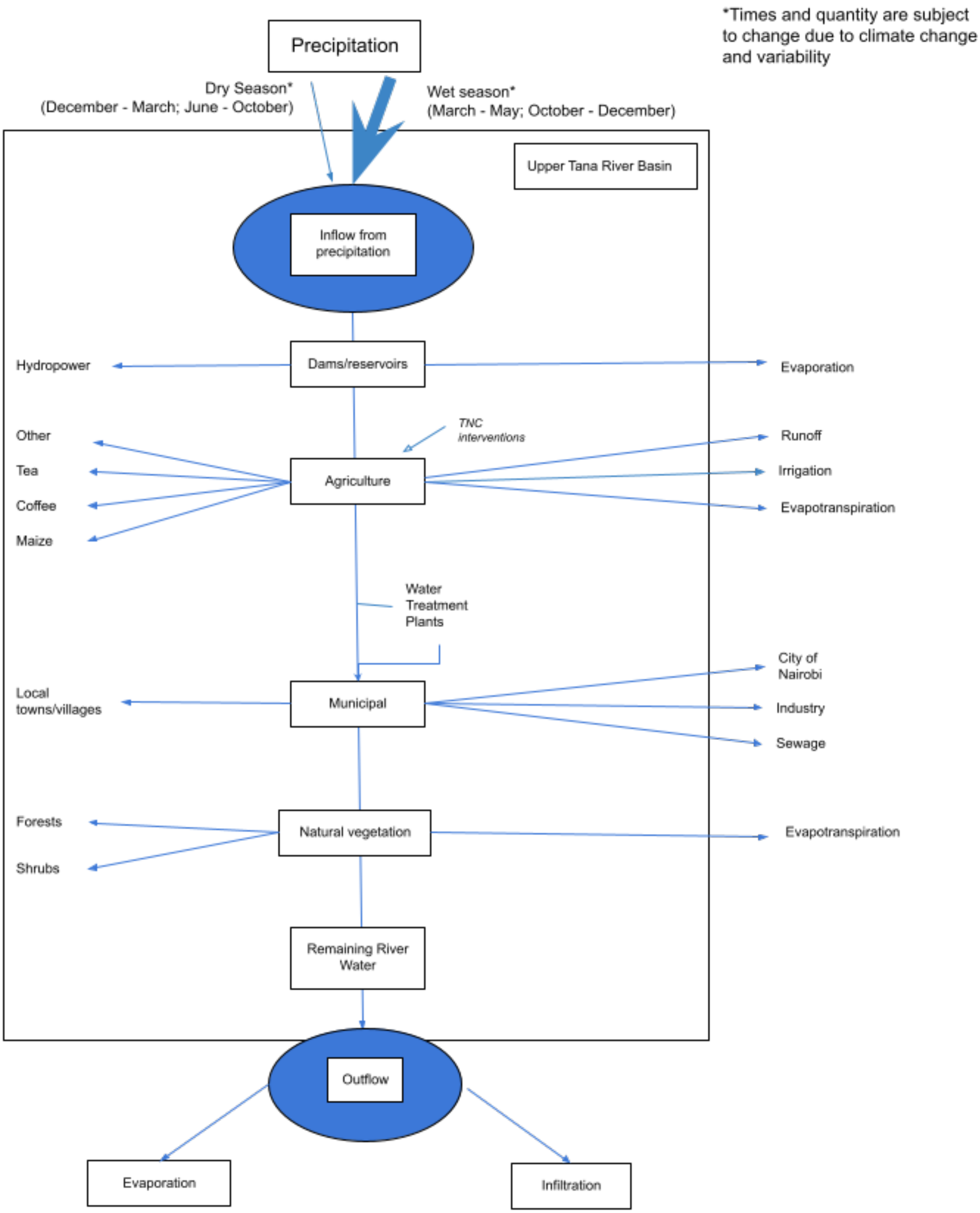


Figure 3: Conceptual Model of the Upper Tana River Basin

3. Research Objectives

We created three research questions to guide our project and best help TNC evaluate the success of the Water Fund to date. The research questions stated below focus on the objectives of understanding land use and cover and its effects on water supply, the effects of climate variability, and evaluating the progress of TNC interventions up to date. The results and findings of answering these questions can have beneficial implications for the region's hydrologic, social, and economic livelihoods.

Research Questions:

1. **How do land use and land cover affect water quantity, water quality, and soil erosion in Kenya's Upper Tana River Basin?**
2. **How has and will climate variability affect seasonal water distribution?**
3. **What are the social and hydrologic impacts of the Water Fund, and how effective are they?**

4. Methods

4.1 Approach

We used open source data as well as data provided by IBM and TNC to answer our research questions. We were granted access from the TNC to proprietary data accumulated from stream gauges and weather stations from the Water Fund. The majority of our analysis centered around analyzing water level[m], discharge[m3/s], total suspended solids[mg/l], and turbidity [NTU] from stream data. Water Level is defined as the total height of the river subtracted by the elevation. In this way water level is a measurement of river height at a specific location, with no association with a datum. Discharge, or flow rate, is the volume of water passing through a specified cross-section of a river for a given unit of time. Total suspended solids, or TSS is the total weight of solid particles suspended in a sample of water. Turbidity is the cloudiness of water due to suspended particles. We used statistical modeling, mapping software and other tools to create our deliverables. Our deliverables consist of a final report, an ArcGIS story map, a data dictionary, a spreadsheet of the literature sources we encountered about the Tana River Basin and the code used to conduct our statistical analyses. The present effort is part of a long-term project that will likely continue after our research. To make our analysis transparent, all of the code used will be available for IBM/TNC on IBM's CloudPak software for future use.

4.2 Background Research and Literature Review

We conducted a literature review of past hydrological studies and projects to understand the historical background and present state of hydrology in the Tana River Basin. We focused on understanding data-driven management of water resources, climate variability, and land use and land cover change within the Upper Tana River Basin. We reviewed physiographic and socio-economic studies done by The Nature Conservancy that laid the groundwork for the Water Fund. Combining the reports from our client with the team's previous literature review gave us a holistic understanding of what has and has not been studied in the Tana River Basin and allowed us to identify areas of study that could benefit from further analysis for our team to focus on.

4.3 Data

To complete data analysis the team used data on multiple variables, including water quality and flow data, meteorological data, and satellite imagery. Data was either publicly available or provided by TNC.

4.3.1 Weather Data

We used IBM's Weather Channel API and The European Union's Copernicus System implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) to obtain data on precipitation, temperature, relative humidity, wind speed, wind direction and dewpoint.

Meteorological data can be obtained from The Weather Channel's core and History on Demand API, provided by IBM. The data are provided hourly for a specified latitude and longitude with a spatial resolution of 0.01° (~1.11 km) from 2015 to 2020. However, the API rejects a number of calls exceeding ~250, so the team used data from specific locations described by latitude and longitude points.

Once the team selected rainfall flux data from 1980 to 1985 from The European Union's Copernicus System the files were downloaded as NetCDF (Network Computer Document Format) files then converted to CSVs. The data are averaged over (0.5 lat 0.5 long) blocks and given in rain flux which is easily converted to milliliters of precipitation.

4.3.2 Ground Truth Precipitation Data

To obtain the amount of precipitation that falls in a watershed, hydrologists place rain gauges, instruments that measure rainfall, at regular points throughout a catchment area (Salas et al., 2014). The Nature Conservancy provided rain gauge data for 32 different stations throughout the Upper Tana Basin.Gauged data are generally more precise in capturing absolute daily rainfall amounts while satellite rainfall estimates are more accurate in capturing spatial rainfall patterns (Antonopoulos et al., 2015).Therefore, TNC's gauge data have been used for measuring intervention effectiveness in the micro watersheds and the satellite data are being used over a much larger area and time frame.

4.3.3 Climate Change Data

The Intergovernmental Panel on Climate Change (IPCC) provides data on climate change that is available for download on climate change through for predictions from Coupled Model Intercomparison Project Phase 5 (CMIP5) (Kenya Data, n.d.). The IPCC Data Distribution Centre provides information and predictions from four emission scenarios as NetCDF files. The team did not end up using CMIP data since CMIP5 data is out of date and CMIP6, which is more recent, has not been downscaled for the Tana River region. The code to analyze the CMIP5 data is available on the IBM Cloud Pak, though we did not use it in our final analysis.

4.3.4 Satellite Imagery

Land cover data was extracted from three global land cover datasets: the MODIS Land Cover Type (MCD12Q1 version 6), Copernicus Global Land Cover Collection-3 (CGLS Col-3), and the Africover dataset derived from Sentinel-2 imagery. The temporal and spatial resolutions for each dataset vary, with the MODIS dataset available for the longest period of time and Sentinel-2 for the shortest. The MODIS imagery has the coarsest spatial resolution, which means it provides the least amount of land cover detail. It is available in 500 x 500 meter resolution, the Copernicus imagery is available in 100 x100 meter resolution, and the Sentinel-2 imagery is available in 20 x 20 meter resolution. The land cover from each satellite dataset was sized to the Tana River region and two smaller study sites within the region using Google Earth Engine and then imported to ArcMap. From there the number of pixels assigned to each of the land classes was determined and the total area for each class for each year from 2001-2019 was calculated in Excel. The MODIS imagery is classified using the IGBP Class-Type-1 classification scheme, while the Copernicus imagery was pre-classified using a discrete classification method.

Data from NASA's Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (LDAS) Noah Land Surface Model L4 was used to obtain a variety of land surface parameters including soil moisture, rainfall flux, surface temperature, and evapotranspiration. The FLDAS Noah Land Surface Model L4 dataset contains a series of land surface parameters simulated from the Noah 3.6.1 model at monthly 0.10 degree resolution from January 1982 to present. The datasets were downloaded as NetCDF files for each month and analyzed using R.

Water Quantity

Discharge volume, flow velocity, turbulence, and depth all impact water quality in rivers (Kuusisto, 1996). Due to a wide variety of limitations, discharge data are not available for a large area of the watershed including micro watersheds that were of interest in this study. However, water level data acquired from HOBO water level loggers is much more readily available both in area (water level data are readily available for the vast majority of watersheds), and duration (water level data date back to at least 2015 for the majority of watersheds). Water level and discharge are closely related, and individual measurements can be transformed to a rating curve that forecasts future stream flows (USGS Stream Gaging Basics).

IBM has already coded a way to relate water level and discharge with a power-law relationship. The specific formula is as follows:

$$Q = C(h - a)$$

where Q is discharge, h is elevation, a is effective depth of zero flow, and C are empirical constants calculated using a linear regression model. While the formula above results in an exponential relationship, when plotted in a log-log space (both x- and y- axes are logarithmic scales) the relationship is linear. Due to the limited amount of reliable water flow rate data, the relationship was derived using data from only one of the stream gauges. The relationship has yet to be applied to the majority of micro watersheds where only water level data are available.

For the purpose of trying to identify broad trends, and due to the lack of discharge data, water level data acquired from the TNC was used to represent water quantity for the statistical analysis of the change in water quantity over time as well as change in water quantity within various micro watersheds.

Water Quality

Several parameters can be used to measure water quality, including chemical characteristics like nutrient concentration as well as physical characteristics like temperature and pH. Instruments for measuring water quality in the field include pH meters, thermometers and EC meters, which measure conductivity and total dissolved solids (TDS) in water. Measurements of dissolved oxygen (DO) indicate if the water is a good environment for aquatic life and are taken with a probe and meter, reported in milligrams of gas per liter of water (Fredette, 2018).

The TNC acquired water quality data with the Hanna Portable Multi-parameter probe, which acts as an accumulation of some of the devices mentioned above. The device measured pH, oxidation reduction potential (ORP), Conductivity, DO, Turbidity, temperature, TDS, Salinity and GPS coordinates of the sampling site. Limited data is available from this device (TNC 2016).There is more abundant data related to total suspended solids (TSS) and turbidity, which can be used to represent water quality. Turbidity is the cloudiness of water due to suspended particles and is measured by shining a light through a sample and recording the particle concentration. It is closely related to TSS, the total weight of solid particles suspended in a sample of water. TSS levels were calculated from river water samples using the gravimetric method. This procedure involves filtering the solids from the water sample through a 47mm glass fiber filter, drying them, and weighing them to determine the total non-filterable residue (TNR) of the sample reported as mg/L (TNC 2016). Turbidity was calculated from TSS concentration in a similar way that water discharge was calculated from water level, by applying a best fit line using a regression model.

4.4 Data Analysis

We used IBM Cloud Pak to analyze data. The IBM Cloud Pak stores data in a cloud and has built-in notebooks that allow both Python and R scripts for more involved data manipulation. Since the scripts and data are stored on IBM servers and not an individual machine, coding and data visualization was a collaborative effort where anyone with access to the dashboard can view and edit code. Furthermore, IBM Cloud Pak has the capacity for automation, which is beneficial for the continual live updating of data as it is collected. Data that can be continuously updated includes daily rain gauge data, remote-sensing raw data, technical reports and published literature. We added relevant open access data (CMIP6, Weather Channel, & ECMWF) to the IBM Cloud Pak, which already contains the vast majority of TNC's data, and wrote scripts to process the data and answer the research questions.

4.4.1 Research Question 1: How do land use and land cover affect water quantity, water quality, and soil erosion in Kenya's Upper Tana River Basin?

Due to increased population growth in the Upper Tana Region, TNC reported a trend over the past five decades of arable land being converted from forest to smallholder farms. We researched how changes in climate and land cover have affected soil quality and agriculture.

Land Cover Analysis

To gauge how land cover and land use has affected water quality and quantity in the Upper Tana Basin, land cover maps were produced using Google Earth Engine and analyzed in ArcGIS. We planned to use Landsat imagery to create annual land cover maps for 1980-2021, but this proved time consuming and prone to error due to the difficulty of the classifier algorithm in ArcGIS to distinguish fallow agricultural land from barren land and dense agricultural crops from forest, grasslands, and savanna. Several datasets with pre-determined land cover are available for direct download from NASA's EarthData and Earth Explorer portals. For this analysis, we used imagery from the Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 data product, which provides global land cover types at yearly intervals from 2001 to 2019 derived from six different classification schemes.

For the purpose of this analysis, a portion of the MODIS Terra+Aqua Combined Land Cover that fit the Tana River region was exported using Google Earth Engine and then the total pixels for each land cover class for each year from 2001 to 2019 were identified in ArcMap. The MODIS Land Cover dataset produces images at a 463 x 463 meter resolution. Using code written in Google Earth Engine (Appendix B) this was extrapolated to 30 x 30 meter resolution, which does not affect the total area covered by each class in the final analysis but makes the maps produced appear less pixelated. Initially results were obtained for the years 2001-2013 using the MODIS MCD12Q1 Version 51 product, but NASA replaced this version in 2014 with MCD12Q1 Version 6 due to processing errors in classifying broadleaf and cereal crop classes in the LAI/fPAR Type3. Version 6 land cover product is thus used in this report. Land cover for the same region was also for the Copernicus Global Land Cover dataset (CGLS) obtained for the years 2015 to 2019 using Google Earth Engine. This dataset is produced at 100 x 100 meter resolution. Additionally, land cover over the same region from the European Space Agency Sentinel-2 satellite was obtained for 2016 at 20 x 20 meter resolution. The total area for each land cover class was calculated in Excel using the formula:

$$total\ area\ (square\ km) = (\#\ pixels * 30m * 30m)/1000$$

All of the land cover datasets used different formulas to identify land cover, leading to differences in the final results for each land cover type from each satellite dataset. The code to extract the annual land cover from each dataset for the Tana River region can be found in Appendix B.

Soil Moisture and Surface Runoff Analysis

To improve our understanding of how the changes in land cover and land use affected water quality and quantity and soil erosion in the area, the land cover maps we developed were compared to other parameters, including surface runoff and soil moisture. Data from NASA FLDAS was used for this. Soil moisture data for depths of 0-10cm, 10-40cm, and 40-100cm were extracted and mapped for the region. The map for soil was then compared to the land cover maps to see how soil moisture patterns compared to the land cover of the area. This gives an indication of how land cover change may affect water retention and availability in soil, which also affects agricultural production. Similarly, data for surface runoff was extracted and mapped for the region before being compared to the land cover maps. Surface runoff is a major cause of erosion, which can also decrease water quality in the area. To better see how surface runoff and therefore erosion may be changing in the region, the surface runoff data was averaged for each year for the entire region and then plotted against time to see if there was any trend over time.

4.4.2 Research Question 2: How has and will climate variability affect seasonal water distribution?

Understanding climate variability and shifts in wet and dry seasons in the region is necessary because anecdotes from people on the ground indicate that seasonal rains are becoming less predictable than they have been historically. Being able to understand the variability in the pattern of seasonal rains and how it may be changing is crucial for effective farming and water management and preparation for a range of possible outcomes, especially as the population in the Upper Tana continues to grow.

Analysis of CMIP

To answer this research question, we began by using CMIP5 data for the Upper Tana River Basin. While we have access to more than 12 ensembles, due to the nature of Cloud Pak uploading nearly 100 files would have been time consuming and used up our data allowance. Additionally, UCLA Box Security never approved our application that would have allowed us to build an alternate data pipeline. Due to these limitations, we did our preliminary examination focusing on the Beijing Climate Center Climate System Model (BCC-CSM2). The data from this ensemble was used to compare the 1970-2000 average millimeters of rainfall per day to 2010-2040 average millimeters of rainfall per day. The code is available in Cloud Pak and specific variables can be adjusted (Local and Global Maps in Appendix C).

CMIP5 data were released and processed between 2010-2014. This makes the data and predictions quite old, especially considering the anecdotal report the project team received from TNC regarding drastic shifts observed in the past five years. As mentioned, CMIP6 has major anomalies that we cannot fix in our time frame. The team did perform a comparison between the two ensembles but ultimately based on advice from expert reviews decided to answer this question using other means (Comparison available in Appendix C).

The Weather Channel and ECMWF

Other means for answering this question included IBM's The Weather Channel API and the EU's Copernicus Satellite's Rain Flux data provided by ECMWF. These systems provide trends in precipitation over time for individual sub basins. We have focused our analysis on four 55.5 km^2 plots Equating to half a degree of latitude by half a degree of longitude labeled by Block. Block A Latitude (0-0.5) Longitude (36.5-37) containing most of Aberdare National Park and Nyeri. Block B Latitude (0-0.5) Longitude (37-37.5) has a large portion of Mount Kenya National Park. Block C Latitude (0.5-1) Longitude (36.5-37) including most of the Thika Chania watershed and four high priority sub-watershed stations. Block D Latitude (0.5-1) Longitude (37-37.5) includes Masiga Reservoir. Figure 4 shows the Blocks over the Water Fund's region.

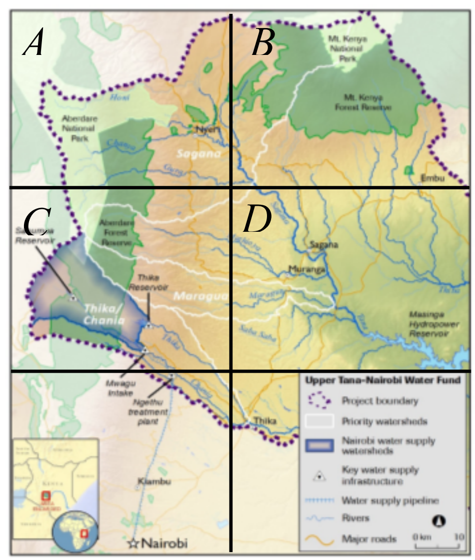


Figure 4: Map of the four land blocks the Upper Tana Water Fund has been divided for analysis

ECMWF data was downloaded and transformed from NetCDF to CSV. The Weather Channel API provided data in CSV format so the team selected four evenly distributed points within these 55.5 Km^2 plots and averaged them creating a data frame that could be compared with the ECMWF file.The team then graphed precipitation on a monthly and daily basis with the wet and dry seasons easily distinguishable. To determine if there has been a meaningful change in season timing a two-sample t test was performed.

Correlation between Rainfall and Soil Moisture

To better understand the effects of climate change and climate variability on agriculture in the area, precipitation was compared with soil moisture at different depths. Soil moisture data at different depths and rainfall flux data (data for soil moisture and rainfall flux both from NASA FLDAS) were plotted against each other to see if there was a correlation between the two parameters. The same was done for surface temperature and soil moisture. Both surface temperature and precipitation are important factors of climate change and climate variability. In order to get a better picture of how both parameters may affect soil moisture and agriculture in the region, various models were tested to find the best fit. Different combinations of the rainfall flux variable, the surface temperature variable, and the interaction between rainfall flux and surface temperature were modeled against soil moisture.

$$SoilMoisture = \beta_0 + \beta_1 RainfallFlux$$
$$SoilMoisture = \beta_0 + \beta_1 SurfaceTemperature$$
$$SoilMoisture = \beta_0 + \beta_1 RainfallFlux + \beta_2 SurfaceTemperature$$
$$SoilMoisture = \beta_0 + \beta_1 RainfallFlux + \beta_2 SurfaceTemperature + \beta_3 RainfallFlux * SurfaceTemperature$$
$$SoilMoisture = \beta_0 + \beta_1 SurfaceTemperature + \beta_2 RainfallFlux * SurfaceTemperature$$
$$SoilMoisture = \beta_0 + \beta_1 RainfallFlux + \beta_2 RainfallFlux * SurfaceTemperature$$

The Bayesian information criterion (BIC) for each model was calculated in R, and the model that best fit the data with the lowest BIC was selected.

4.4.3 Research Question 3: What are the social and hydrologic impacts of the Water Fund, and how effective are they?

We analyzed how interventions implemented by the TNC have improved water quality and water conservation within specific regions of the watershed and extrapolated the results to the entire watershed. The first step was to identify key interventions the TNC has implemented, as well as their specific geographic locations within the watersheds (Section 2.3.2). Due to the large amounts of interventions made, limited data on the specific locations to the interventions, and confidentiality concerns due to the fact that many interventions are on private property, the locations of interventions are generalized and only known at the micro watershed level. As a result it was not possible to map the specific coordinates/areas of interventions and conduct a site-specific analysis. Instead, data from micro watersheds that were known to have TNC interventions were compared to data from similarly sized micro watersheds with no interventions that served as a control. Two sets of micro watersheds, shown in the figure below, with high priority were chosen due to their relative proximity to one another: the Githambara watershed (intervention watershed) and the Karurumo watershed (control).

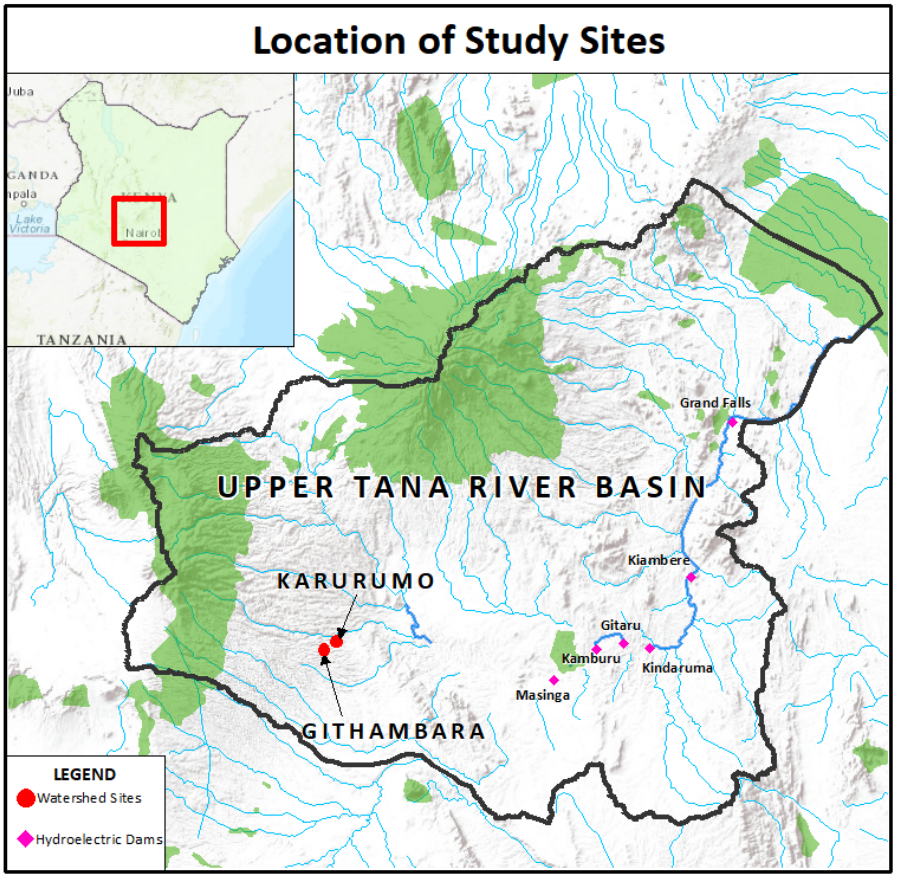


Figure 5: Location of study sites within the Upper Tana River Basin

To analyze water quantity differences, water levels from stream gauges within each respective watershed were compared between the two watersheds. The coordinates of the Githambara stream gauge are located at (-0.80988, 37.07202), and the Karurumo stream gauge is located at (-0.79624, 37.08675). The data was grouped into different time periods for analysis, including yearly and seasonal. For the seasonal comparisons water level data ranging from 2015-2019 was split into a rainy (includes data from both bimodal rainy periods) and dry season.

T-tests

Firstly, all the data was cleaned by excluding the outliers. Outliers were calculated as any data points that were smaller than Q1 - 1.5*IQR and larger than Q3 + 1.5*IQR. Before analyzing the data statistically, graphs were plotted to visualize the general trend of both water quantity change and water quality change. Specifically, heatmaps were used to analyze water level change and scatter plots were used to visualize the general trend of TSS change and turbidity change. Next, to run a t-test, the data were zgrouped yearly and annual average changes were calculated. The statistical hypothesis test t-test was run to see if there was a significant difference between the year 2015 and the year 2019. The time periods were chosen as indicators of the beginning and the end of the water fund project, and the data availability for each year was taken into considerations as well. Additionally, another set of micro watersheds was chosen and analyzed in order to have a more accurate statement on the effectiveness of TNC interventions. The micro watersheds were Mbogiti (intervention watershed) and Thika-valley (control watershed). The coordinates of the Mbogiti stream gauge are located at (-0.83977, 36.82812), and the Thika-Valley stream gauge is located at (-0.79176, 36.80548).

Analysis of Variance (ANOVA)

An ANOVA test was run to test if there was a statistically significant difference between the Gitahmabara rainy seasons, Gitahmabara dry season, Karurumo rainy seasons, and Karurumo dry season datasets. Statistical significance is defined by a p-value less than 0.05. A p-value less than 0.05 allows for the rejection of the null hypothesis, which for this study was that the difference in mean water level between two micro watersheds was equal to zero. In other words, the null hypothesis was that there is no difference in water level between two separate datasets. Since the ANOVA test only confirms that there is or is not a significant difference between at least two of the datasets, but does not confirm specifically which datasets are significantly different, individual t-tests were run between each combination of datasets.

Multivariate Regression

Next, continuous variables including precipitation, sum of accumulated weekly precipitation, wind speed, and temperature from the weather API, as well as categorical variables including the location of each watershed (Githambara or Karurumo) were plotted against water level to see if there were any noticeable trends that could be explained with a regression model. For any given day the sum of accumulated weekly precipitation, or "clumped" precipitation, is defined as the sum of the precipitation rate for the current day along with the precipitation rate of the previous six days. This value was considered in order to account for the fact that it takes time for precipitation to ultimately runoff into a river. A multitude of factors including storm intensity, slope, elevation, distance from river, soil type, etc. all contribute to the runoff time. With no accurate way to quantify this time, one week was chosen as an educated guess. It should be noted that only dates/times where water level and meteorological data were available were considered for this analysis. While interpolation was considered, due to the large gaps in stream data (weeks to months) the accuracy of the interpolation would be questionable. Data from the weather API was used due to the lack of locally collected data in the regions of interest. Since the weather API contains data in multiple geographic coordinates surrounding the region, all data points within a two-degree range in both latitude and longitude from the exact location of the stream gauges were considered. These data points were then averaged so that there was only one data point for each time. Lastly variables deemed to have a significant effect were put into a multi-variable linear regression model. The purpose of the linear regression model was to quantify an estimate on how much a change in the independent variable (i.e. precipitation and location of watershed) had on the dependent variable (water level).

5. Results

5.1 Research Question 1: Land Cover Change Over Time

Land cover data from the MODIS, Sentinel, and Copernicus datasets was obtained for various years in the period between 2000-2019. The three datasets largely agreed on the amount of area covered by forest, built-up urban land, and water, but varied widely when classifying grasslands, agricultural/cultivated lands, and savannas. This is likely due to the difficulty of distinguishing the difference between sparsely vegetated areas from satellite imagery alone.

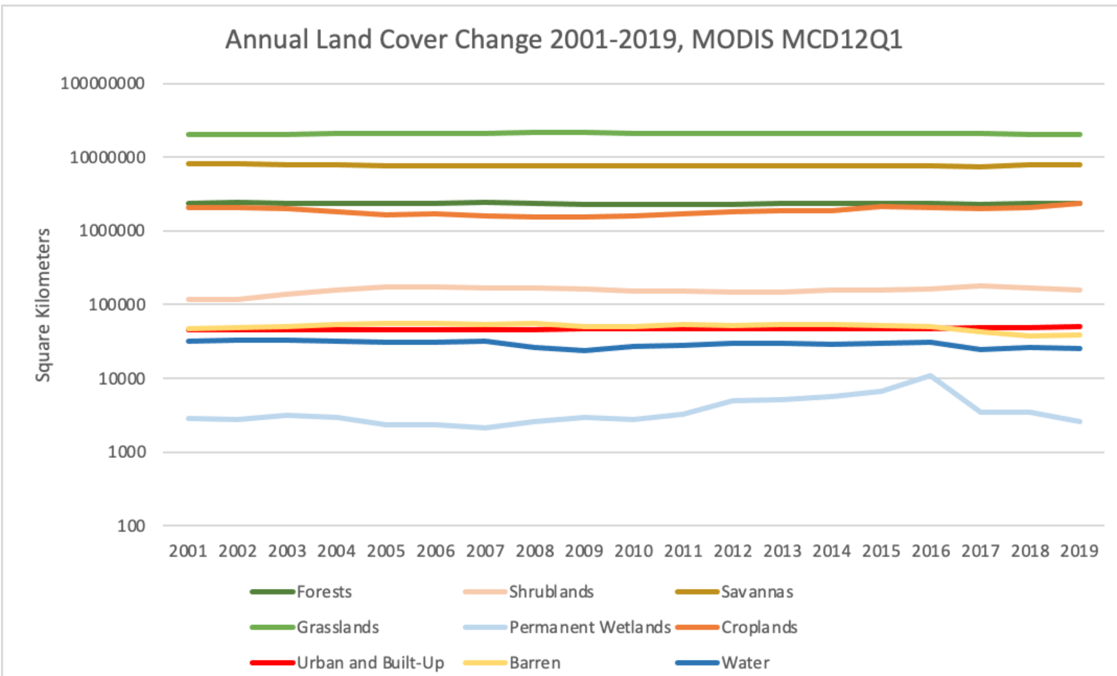


Figure 6: Annual land cover change from 2001to 2019 based on the MODIS MCD12Q1 classification algorithm.

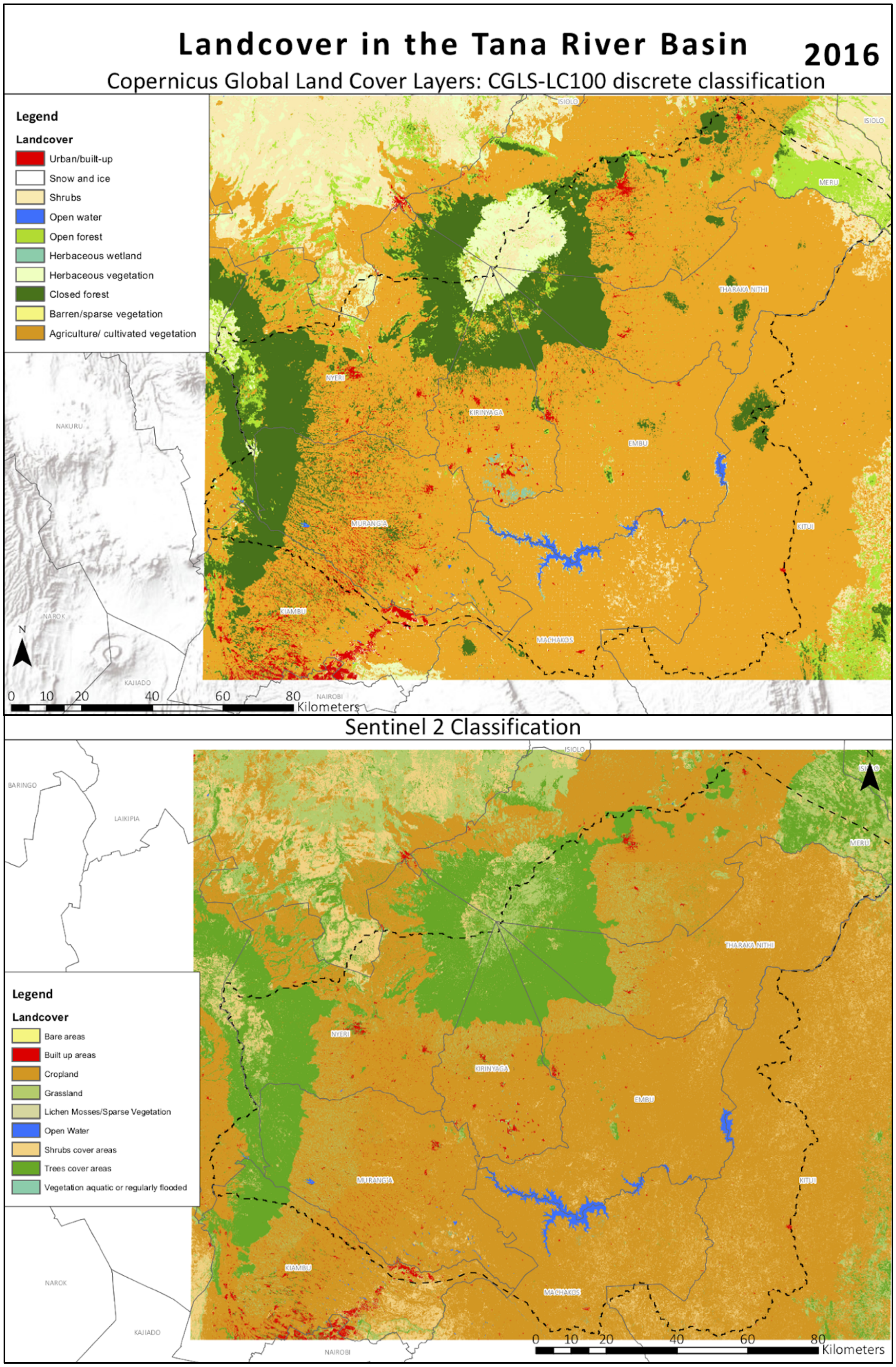


Figure 7: Landcover in the Upper Tana in 2016 from two different land cover datasets, MODIS MCD12Q1 version 6 and the Sentinel 2 Africover daatset

Appendix B has graphs that show more detailed land cover change. The area covered by grassland is the largest in the MODIS land cover classification. Land cover does not change drastically over the time period from 2001 to 2019. Forest and cropland cover about the same amount of area.

A more detailed look at land cover for 2015-2019 in the intervention and control sites examined later in this paper is shown below.

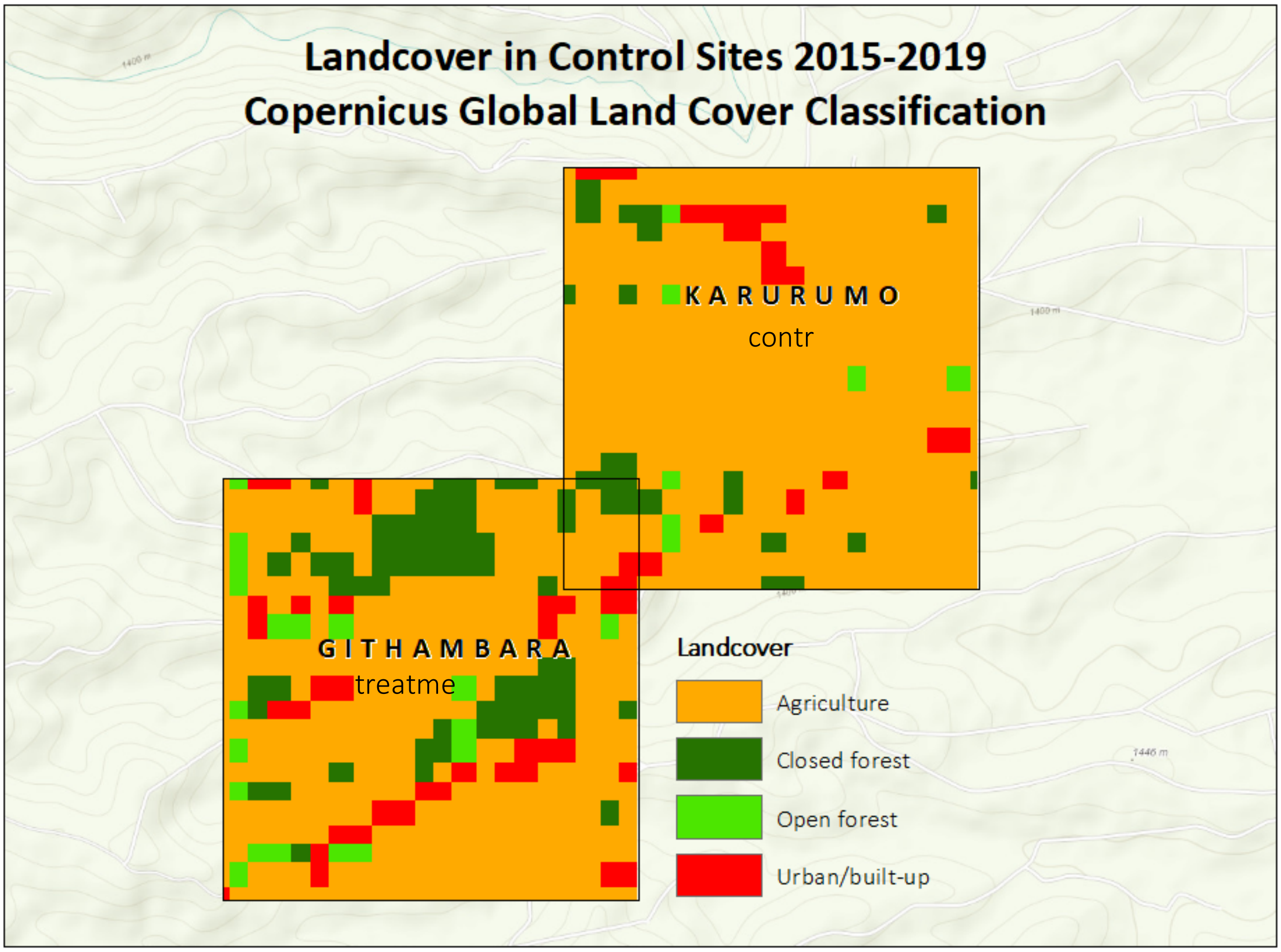


Figure 8: Land cover for 2015-2019 in Karurumo (control) and Githambara (treatment) from the Copernicus Gobal Land Cover Classification

In addition to land cover, we investigated other climatic and environmental factors such as soil moisture that can help provide some insight on how land cover affects land surface parameters and water availability and quality in general for the region. The following figures show soil moisture for the Upper Tana River Basin at different depths (0 to 10 cm, 10 to 40 cm, and 40 to 100 cm) for March 2021 at a resolution of 0.1°x 0.1°.

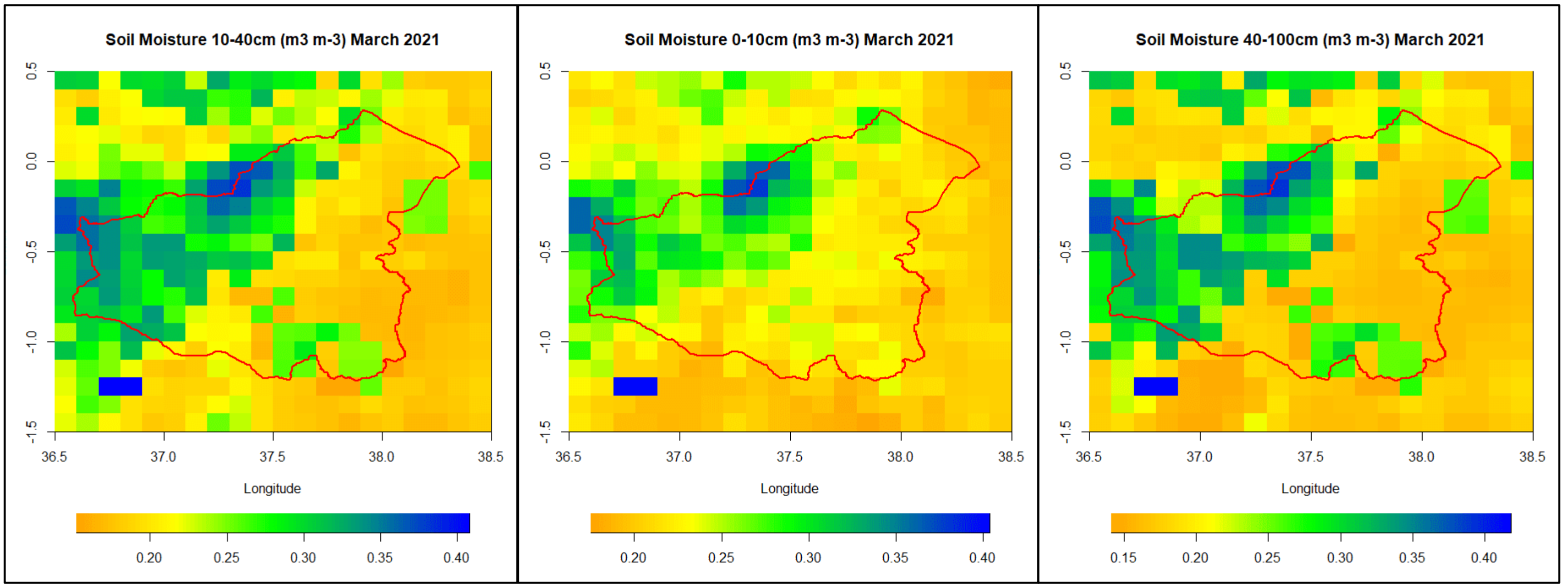


Figure 9: Soil Moisture in the Upper Tana River Basin at 0-10cm, 10-40cm, and 40-100cm depths from NASA FLDAS

When the figures for soil moisture are compared to the land cover map of the Upper Tana for the same month, March 2021 (Appendix B), we can see that for all of the depths analyzed, areas with more forest cover have a higher soil moisture content. Additionally, soil moisture seems to moderately increase at greater depths.

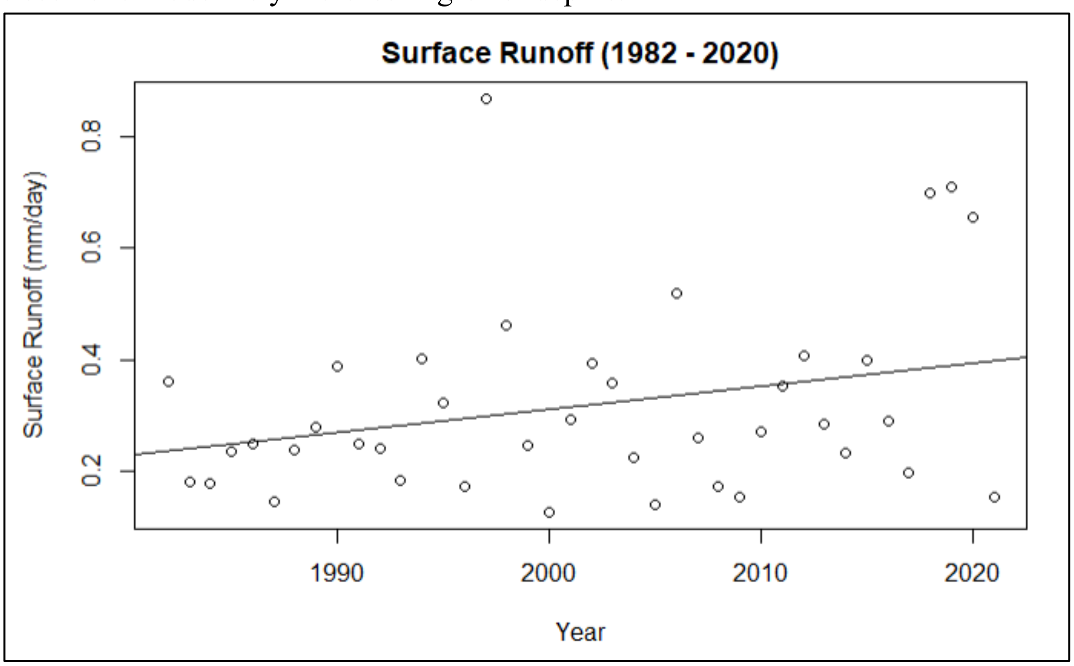


Figure 10: Average annual surface runoff (mm/day) from 1982 to 2020

Comparing land cover with surface runoff can provide insight on erosion and water quality in the region. When looking at the general trend of surface runoff over time (Figure 10), surface runoff seems to be increasing, indicating there may be an increase in erosion leading to worsening water quality. However, variation in the amount of runoff each month also seems to have increased, which could be related to factors other than land cover change, such as the increased variability in precipitation observed in Research Question 2.

The following figures depict the distribution of average runoff for the periods 2000-2005 and 2015-2020.

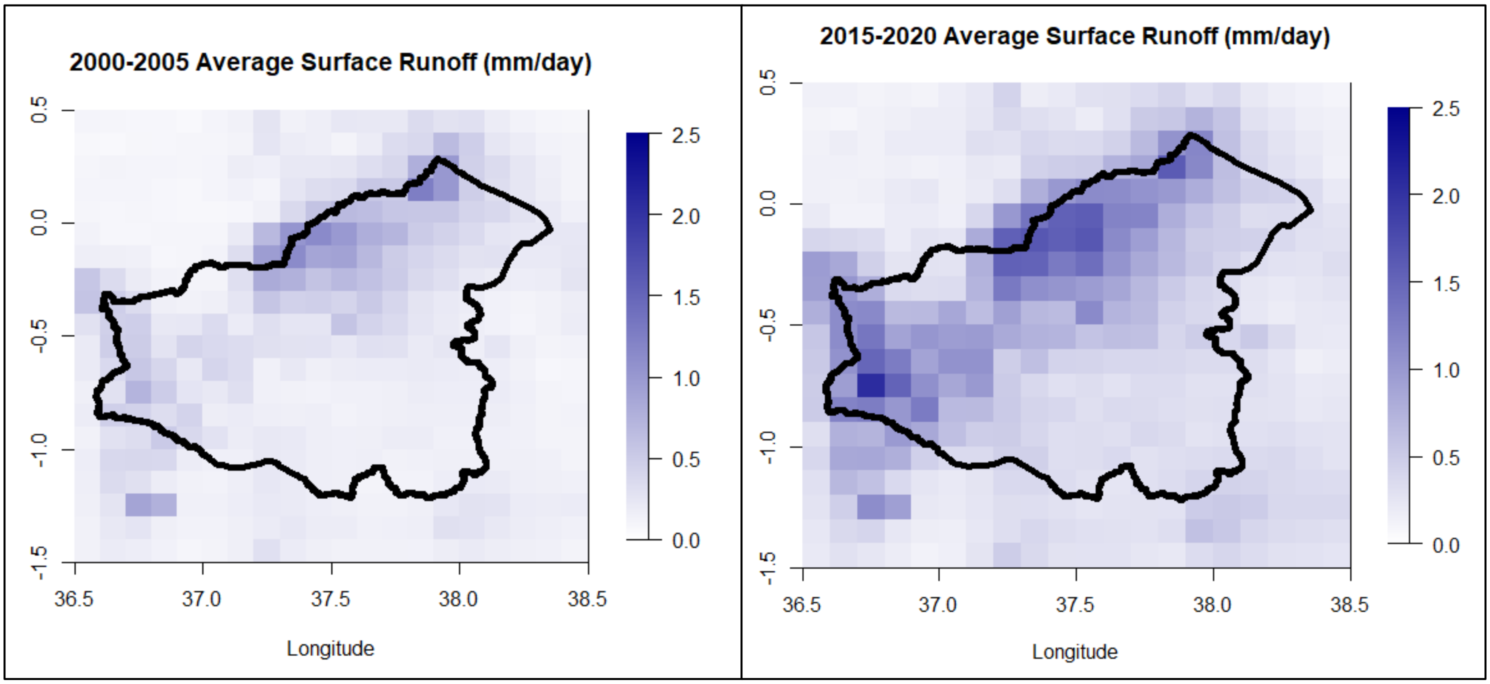
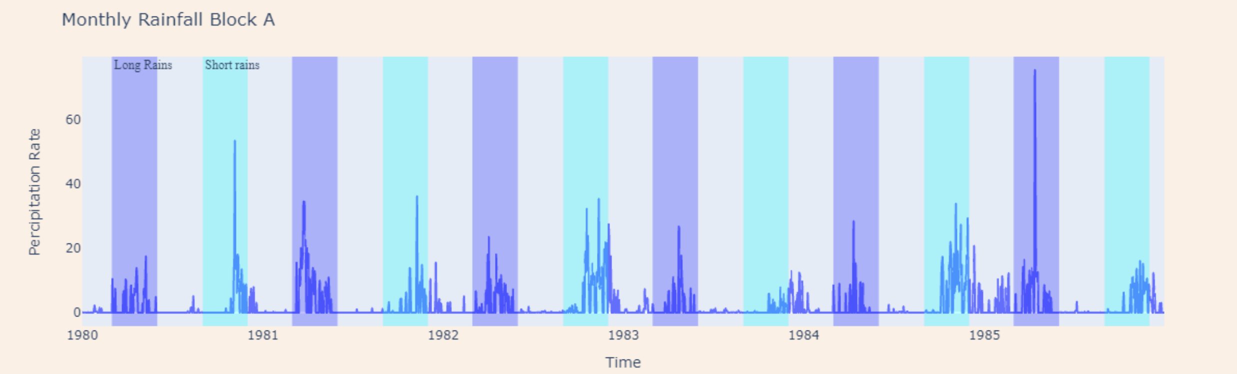


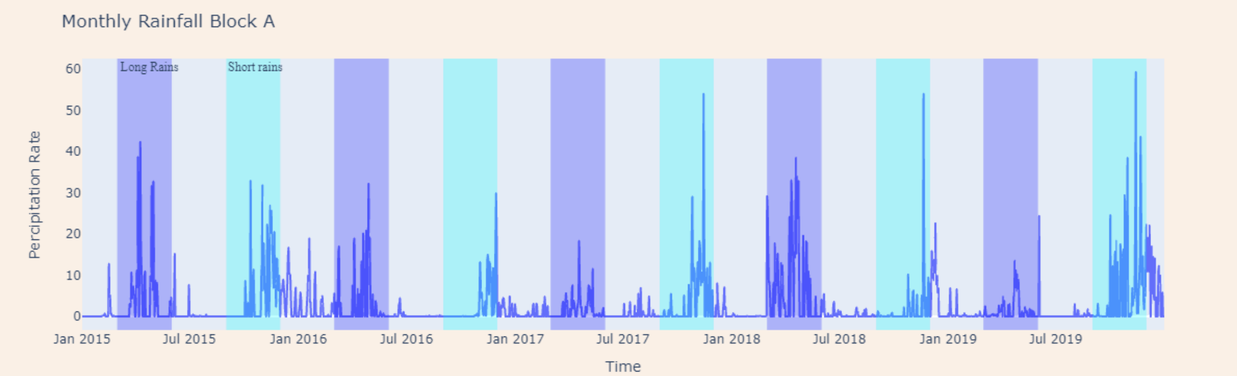
Figure 11:Average surface runoff (mm/day) for 2000-2005 and 2015-2020

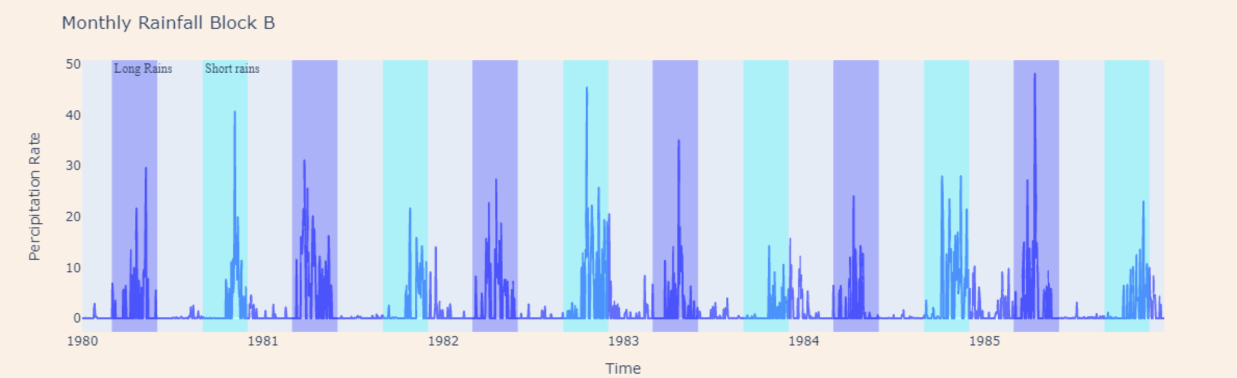
The figures show how runoff is greatest in areas with high elevation (Appendix C), which makes sense as these areas get higher amounts of precipitation and water naturally flows from areas with higher elevation to lower elevation. The figures also show that overall, there was an increase in runoff during the 2015-2020 period compared to the 2000-2005 period. The increase in runoff may not be related to land cover change because as indicated earlier, land cover did not change by a significant margin. It may instead be due to other factors such as changes in patterns in precipitation described in the following section.

5.2 Research Question 2: Effects of Climate Variability

The following images are daily precipitation levels from the four land blocks in the Upper Tana River basin: Block A (0°N-0.5°N and 36.5°E-37°E); Block B (0°N-0.5°N and 37°E-37.5°E); Block C (0.5°N-1°N and 36.5°E-37°E); Block D (0.5°N-1°N and 37°E-37.5°E). Map of the land blocks is shown in Section 4.4.2


Monthly Rainfall Block A
Long Rains
Short rains
Percipitation Rate
0
20
40
60
1980
1981
1982
1983
1984
1985
Time


Monthly Rainfall Block A
Long Rains
Short rains
Percipitation Rate
0
10
20
30
40
50
60
Jan 2015
Jul 2015
Jan 2016
Jul 2016
Jan 2017
Jul 2017
Jan 2018
Jul 2018
Jan 2019
Jul 2019
Time


Monthly Rainfall Block B
Long Rains
Short rains
Percipitation Rate
0
10
20
30
40
50
1980
1981
1982
1983
1984
1985
Time

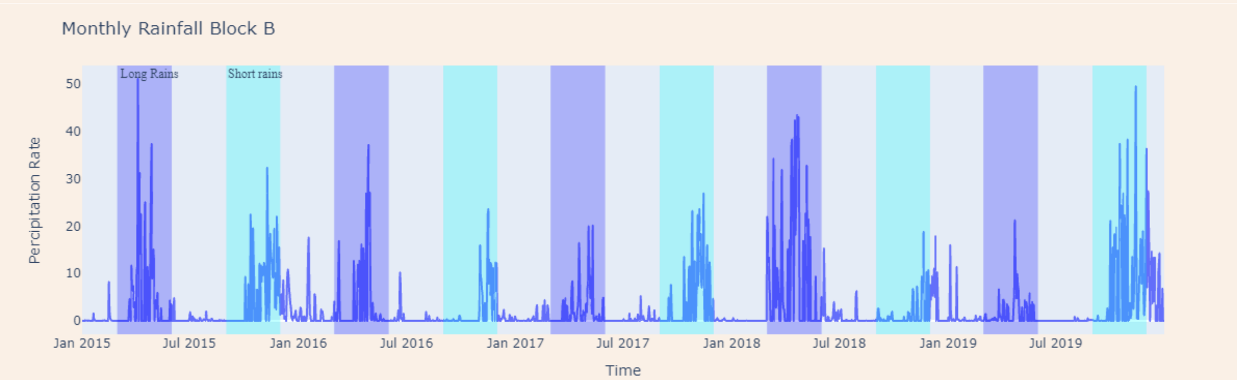

Monthly Rainfall Block B
Long Rains
Short rains
Percipitation Rate
0
10
20
30
40
50
Jan 2015
Jul 2015
Jan 2016
Jul 2016
Jan 2017
Jul 2017
Jan 2018
Jul 2018
Jan 2019
Jul 2019
Time

Figure 12: Daily precipitation levels for four land blocks in the Upper Tana River Basin for 1989-1985 and 2015-2019

Kenya has two wet seasons called the long rains (March - May) and short rains (October - December) The graphs show the long rains in a dark blue and the short rains in cyan. The measurements of precipitation do change on the y-axis but remain relatively constant in similar blocks, especially for the 1980-1985 period. While the dry and rainy seasons for the 2015-2019 period did not necessarily shift from the 1980-1985 period, the graphs suggest that the magnitude of rain events during the wet seasons have increased, which may be attributed to global climate change which is hypothesized to not affect the amount of rainfall but rather increase the severity of storms in shorter periods of times.

Block	T-Value	P-Value	Conclusion
Overall Precipitation:			
1980-1984 vs. 2015-2019			
A	7.11	1.19E-12	Significant difference
B	8.82	1.15E-18	Significant difference
C	4.11	4.04E-05	Significant difference
D	8.04	8.95E-16	Significant difference
Wet Season Precipitation (long rains): Months 3,4 &5:			
1980-1984 vs. 2015-2019			
A	4.26	2.07e-05	Significant difference
B	6.19	6.24E-10	Significant difference
C	1.5	0.13	no Significant difference
D	4.22	2.43E-05	Significant difference
Dry Season Precipitation: Months 6, 7, 8, 9, 10, 1 & 2:			
1980-1984 vs. 2015-2019			
A	-11.25	2.61E-29	Significant difference
B	-10.25	1.28E-24	Significant difference
C	-7.05	1.79E-12	Significant difference
D	-10.54	6.32E-26	Significant difference

Table 2: T-tests Comparing ECMWF precipitation data from two distinct five-year time periods

From the analysis using t-tests, we can conclude that there is a significant difference in precipitation levels measured by the ECMWF system for the five-year periods of 1980-1984 and 2015-2019 because the p-value remained below 0.05 allowing the rejection of the null hypothesis that these systems are the same. This was found in the data with the exception of the wet Season in Block C which had a p-*v*alue of 0.13 indicating no *s*ignificant difference.

More graphs for exploring climate variability including monthly precipitation and precipitation from 2015-2019 for the four blocks using data from The Weather Channel can be found in Appendix D.

To analyze the effects of climate variability on agriculture in the region, we looked at the relationship between precipitation and soil moisture, which is an important factor for agricultural productivity in the region. The following figures show soil moisture and rainfall flux mapped individually for East Africa (Upper Tana River Basin outlined in red). The data to create these maps are from NASA, and the month January 1982 is used as an example.

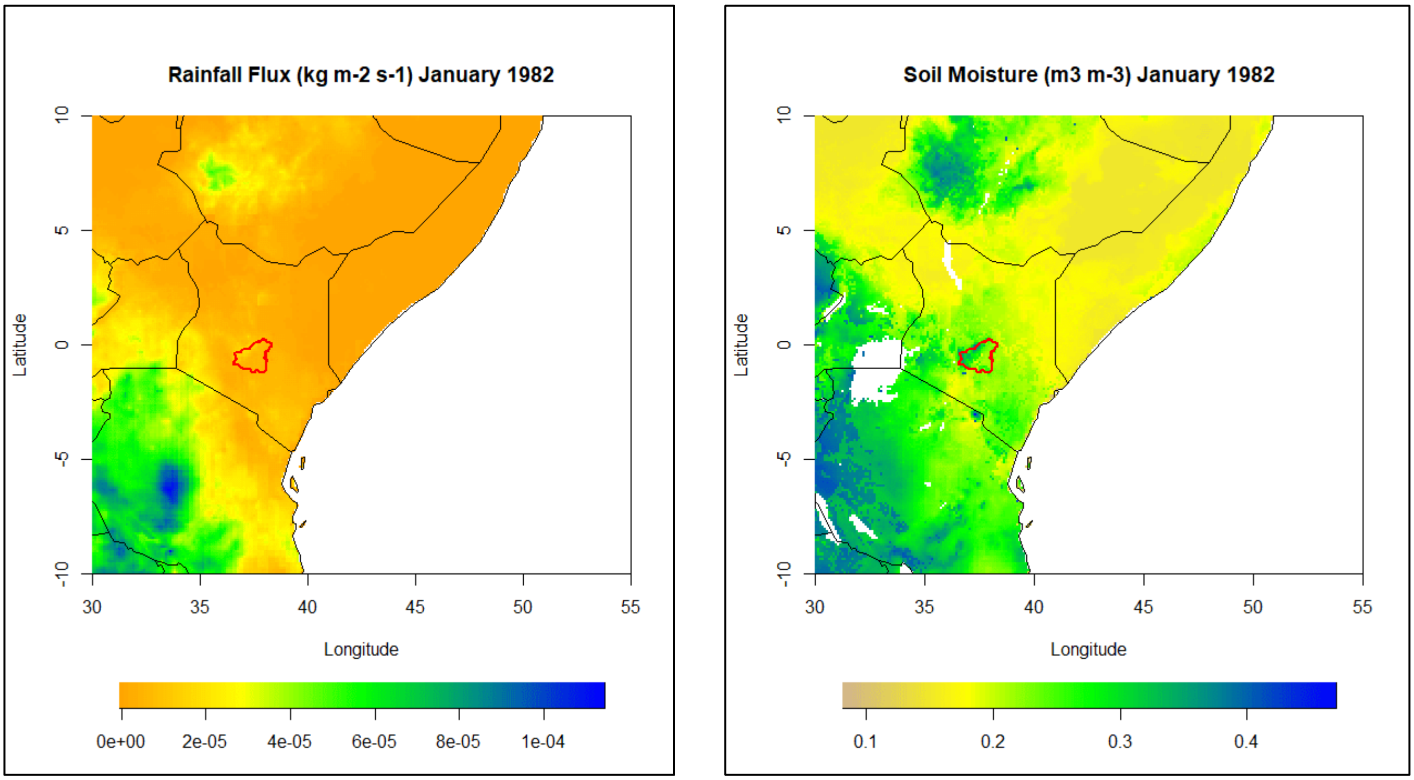


Figure 13: Rainfall Flux and Soil Moisture in East Africa for January 1982 (Upper Tana River Basin outlined in red)

When plotting precipitation rate and soil moisture over time on a single plot, the fluctuations in soil moisture and precipitation seem to be related (Figure 14).

We plotted soil moisture against precipitation from 2015 to 2020 to get a better understanding of their relationship (Figure 15) using data from NASA and the Weather Channel for the entire region. A moderate positive correlation (r = 0.6874) was found between the two parameters, meaning that increased precipitation was correlated with higher soil moisture content, which is an intuitive result.

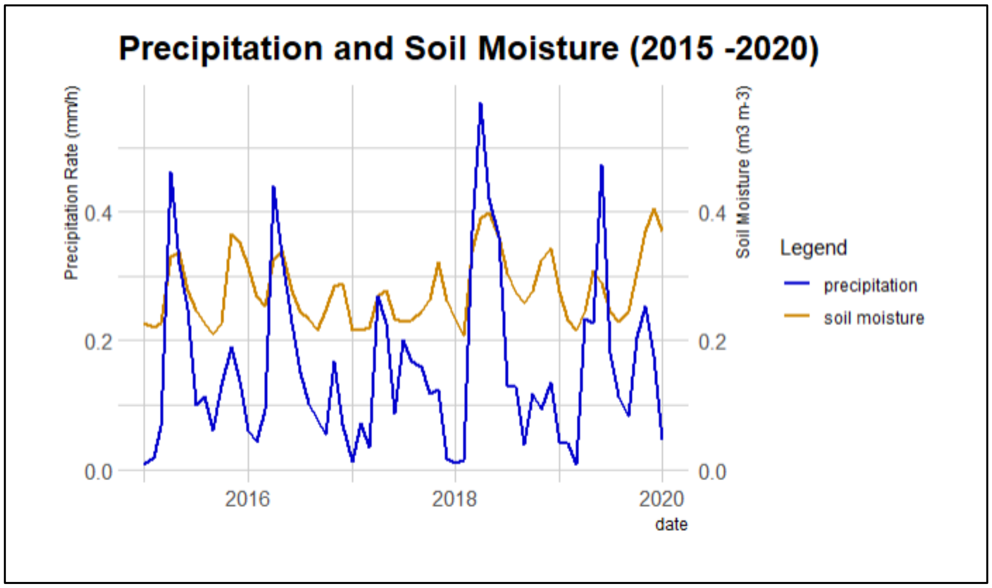


Figure 14: Monthly precipitation rate and soil moisture in the Upper Tana River Basin from 2015 to 2020

*S*oil moisture *was plotted* against *the* precipitation from 2015 to 2020 to get a better understanding of their relationship (Figure 15) using data from NASA and the Weather Channel for the entire region. A moderate positive correlation (r = 0.6874) was found between the two parameters, meaning that increased precipitation was correlated with higher soil moisture content, which is an intuitive result.

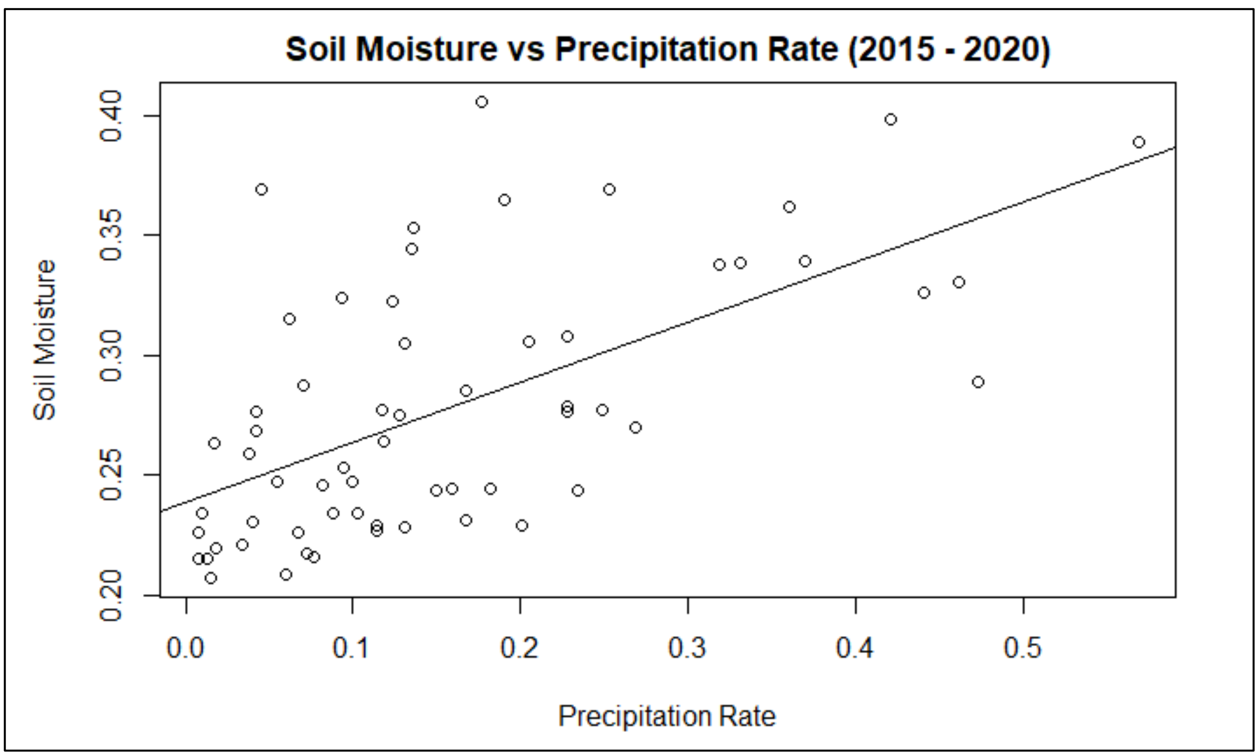


Figure 15: Monthly Average Soil Moisture vs. Precipitation Rate (2015 -2020)

Exploring the correlation between precipitation and soil moisture in the Upper Tana River Basin can provide insight on how changes in precipitation from climate variability may affect soil moisture and in turn agricultural production. NASA data for soil moisture and rainfall flux at a resolution of 0.1°x 0.1° from January 1982 to March 2021 was used to create heat maps of soil moisture and rainfall correlation (R-value) at depths of 0 to 10 cm, 10 to 40 cm, and 40 to 100 cm for the Upper Tana River Basin, which is outlined in white (Figure 16).

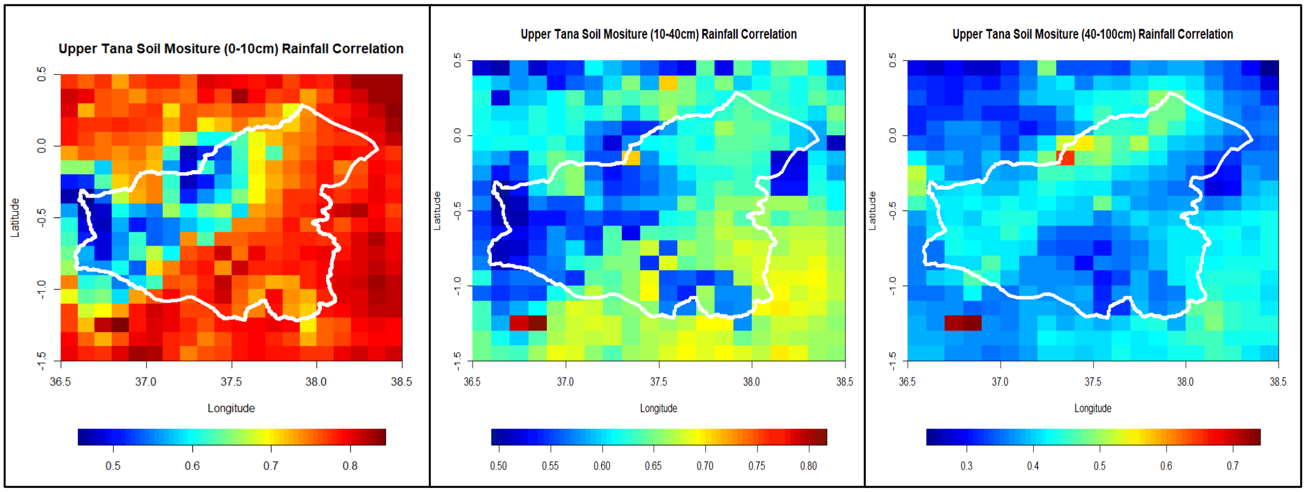


Figure 16: Soil Moisture and Rainfall Correlation (r) in the Upper Tana River Basin at 0-10cm, 10-40cm, and 40-100cm depths

The figures show that at greater depths, the correlation between soil moisture and precipitation decreases. Additionally, when compared to land cover maps from the previous section, we can see that the forested areas at higher elevation have lower correlations between soil moisture and rainfall. This may be because a greater proportion of precipitation is intercepted by forest vegetation before hitting the land surface or because forests at higher elevations typically have lower temperatures, more frequent rainfall, and less evaporation, so individual rain events have less of an effect on soil moisture. Additionally, this may be related to variances in soil type (Appendix C), but that relationship would need to be explored further before drawing any conclusions.

In addition to precipitation, surface temperature is also an important factor to consider for climate variability and climate change. We looked at surface temperature to see how it would affect soil moisture and agriculture in the region. From the heat map in Figure 17, we can see that the correlation between soil moisture and surface temperature is very weak and insignificant in most parts of the region, although in the areas with higher elevation, there is a moderate negative correlation.

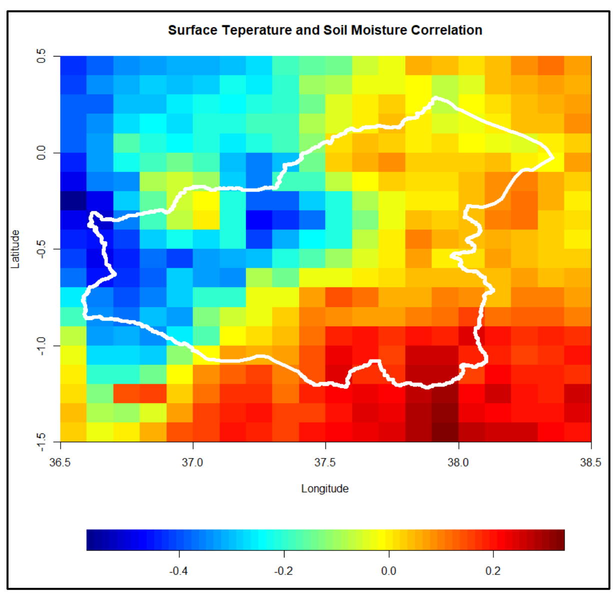


Figure 17: Surface temperature and surface moisture correlation

To take a closer look at soil moisture (m^3/m^3) with both surface temperature (K) and rainfall flux (kg/m^2s), we tested various models and found that the most appropriate regression model included rainfall flux as a parameter as well as the interaction between rainfall flux and surface temperature, but not surface temperature individually. The regression equation is:
SurfaceMoisture = 0.2174 + 0.0006311RainfallFlux - 0.0211RainfallFlux*SurfaceTemperature

From the model, we can see that greater amounts of precipitation will increase soil moisture, but effects of precipitation will vary depending on temperature. In addition, higher temperatures will lower the amount that precipitation increases soil moisture.

5.3 Research Question 3: Efficacy of Interventions

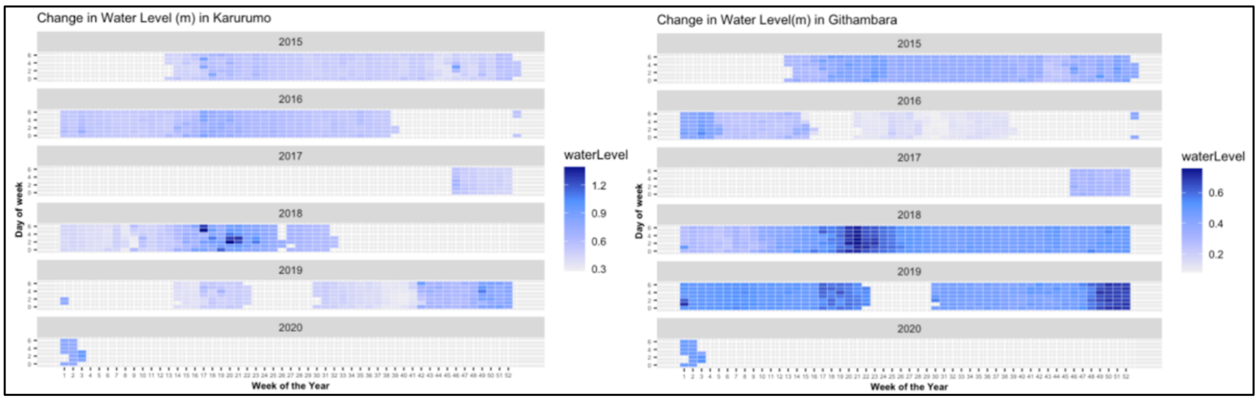


Figure 18: Changes in water level in Karurumo (C) and Githambara (T) from 2015 to 2020

The heatmaps *in Figure 18* show the general trend of water level change in Githambara (T) and Karurumo (C) from 2015 to 2020. The water level in Githambara displays an obvious increasing trend over time while there is only a slight change in water level in Karurumo.

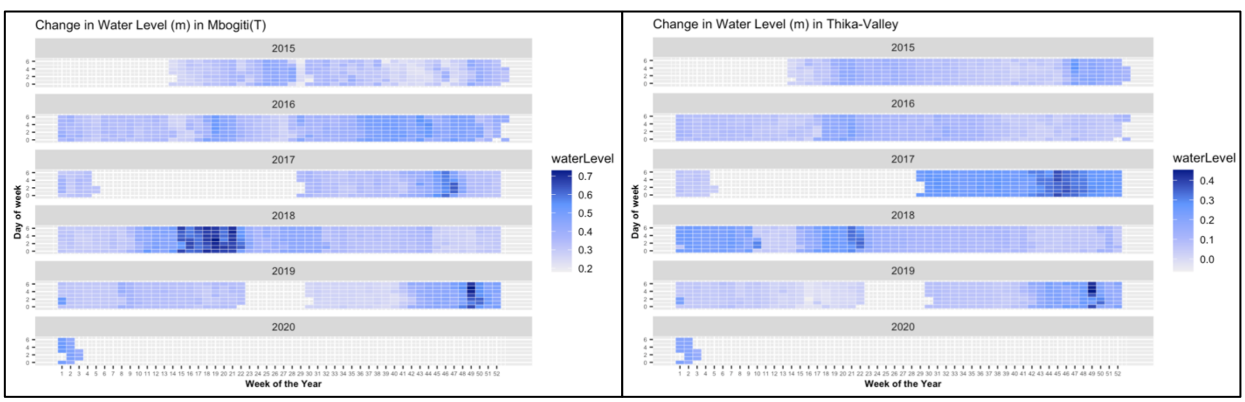


Figure 19: Changes in water level in Thika-Valley (C) and Mbogiti (T) from 2015 to 2020

Figure 19 shows the general trend of water level change in Mbogiti (T) and Thika-Valley (C) from 2015 to 2020. It is difficult to observe any obvious patterns from looking at the overall trend of water level change in both watersheds. Further analysis in t-test is needed to analyze the difference in years.

watershed	p-value	conclusion
water level		
Githambara	< 2.2 * 10-16	significant difference
Karurumo	0.09436	no significant difference
Mbogiti	4.181 * 10-6	significant difference
Thika-Valley	0.05585	no significant difference
total Suspended Solids (TSS)		
Mbogiti	0.02094	significant difference
Thika-Valley	0.8328	no significant difference
turbidity		
Mbogiti	0.0154	significant difference
Thika-Valley	0.1886	no significant difference

Table 3: Results for t-tests analyzing water quality and quantity in control and treatment sites

The table above summarizes the p-value for variables within each watershed including water level, TSS, and turbidity from 2015 to 2019. The t-test was used to analyze water level change, while the Mann-Whitney test was used to analyze TSS and turbidity change. The Mann-Whitney test was preferred due to its higher accuracy with datasets containing a small sample size. The results showed that for all variables, the p-values in treatment watersheds are smaller than 0.05, which means there is a significant difference between 2015 and 2019. And all p-values in control watersheds are larger than 0.05, indicating the difference between 2015 and 2019 is not significant. By comparing the two sets of watersheds, it can be concluded that treatment sites have had a significant change in water quantity and quality, shown in the graphs below as an increase in water level and decrease in TSS and turbidity, whereas the control sites have had no obvious change.

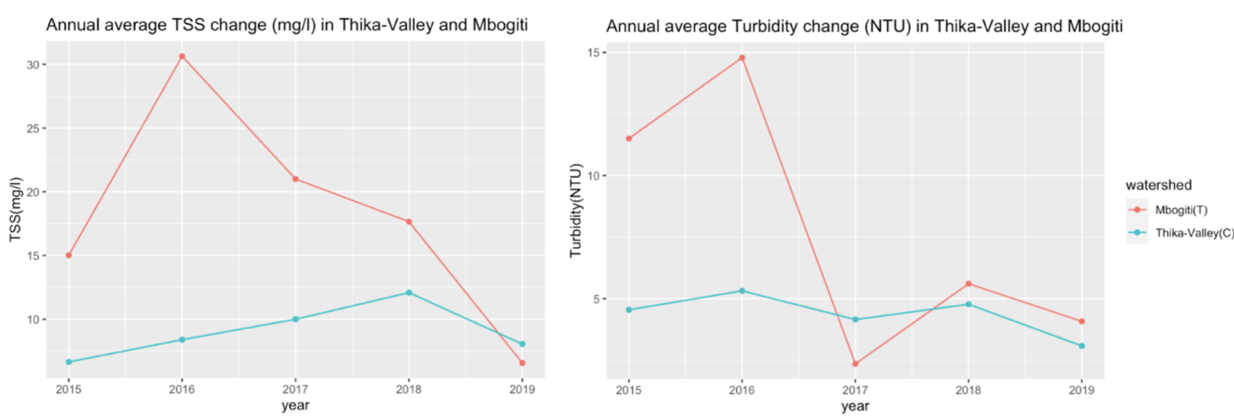


Figure 20: Changes in annual average TSS and Turbidity for Thika-Valley (C) and Mbogiti (T) from 2015 to 2019

The graphs above show annual changes in TSS and turbidity from 2015 to 2019 for Mbogiti (intervention watershed) and Thika-Valley (control watershed). Mbogiti started off with higher TSS and turbidity values. Over the five year time span, TSS and turbidity values showed a significant decrease, which is also confirmed by analyzing with t-test, while both values in Thika-Valley remained the same.

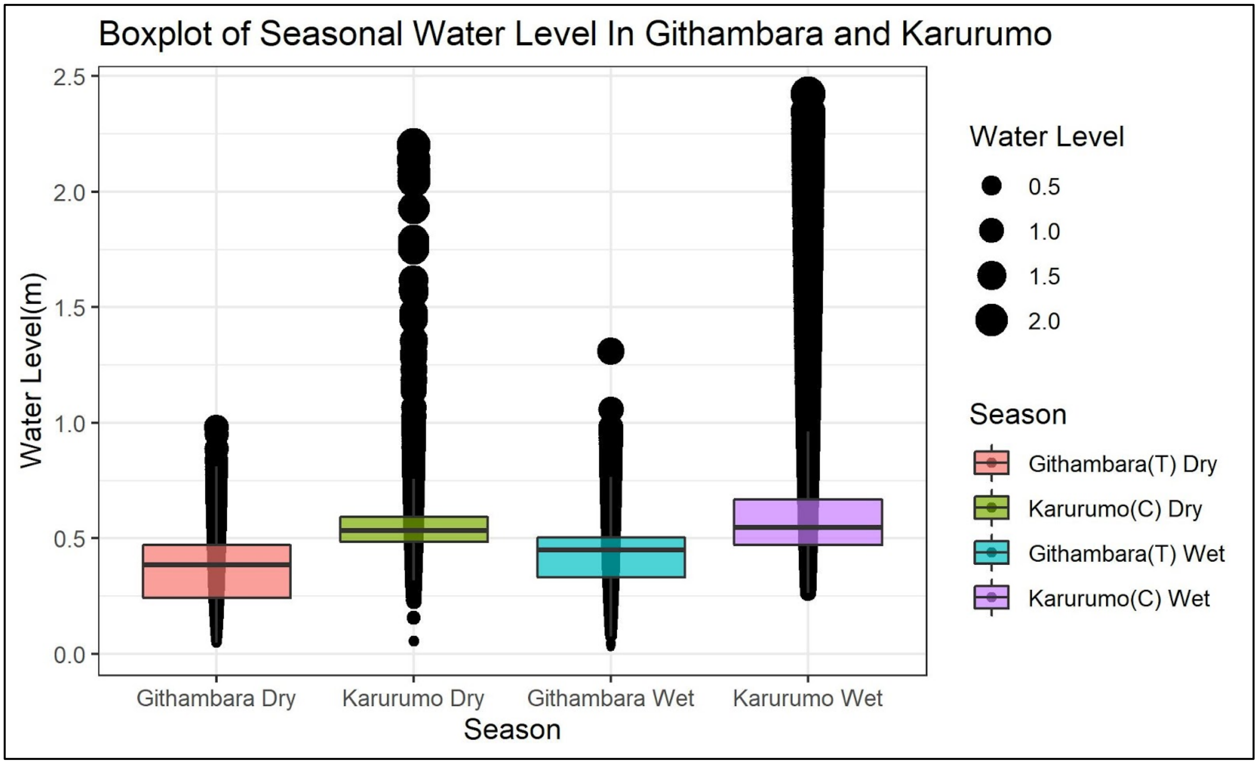


Figure 21: Boxplot of seasonal water level in Githambara and Karurumo

The boxplot above compares the distribution of water level data for Githambara and Karurumo for rainy and dry seasons. It was used to visualize the data before running an ANOVA test to notice any initial trends. It should be noted that there are few discharge values higher than 2.5 m but for visualization purposes the y-limit of the graph is set at 2.5m to better showcase the majority of the data. From the graph it can be seen that variation in each micro watershed has a greater effect on water level than variation in season. Additionally, it can be seen that Karurmo has higher average water levels than Githambara in both the rainy and dry seasons.

The results of the ANOVA test are presented above. The null hypothesis was that the difference in mean between every group (rainy-season Githam, dry-season Githam, rainy-season Karurumo, and dry-season Karurumo) was zero. In other words the mean discharge for all four groups were the same. An extremely low p-value ($p<.05$) allows us to reject the null hypothesis and conclude that there is a statistically significant difference between location and season on discharge. Additionally, an extremely high F-value confirms that the variance between groups is

ANOVA RESULTS		
	F-value	p-value
Value	13521	<2.2*10-16

Table 4: Results of ANOVA analyzing water level data for Githambara and Karurumo for wet and dry seasons.

significantly higher than variance within groups. This further confirms our conclusion to reject the null hypothesis. Logically this makes sense since the rainy season and dry season should yield significantly different discharge rates. The next step was to conduct individual t-tests to find specifically what variables are significantly different (Appendix E).

Individual T-test Results		
	p-value	conclusion
Githambara/Karurumo Wet Season	<2.2*10-16	significant difference
Githambara/Karurumo Dry Season	<2.2*10-17	significant difference
Wet/Dry Season Githambara	<2.2*10-18	significant difference
Wet/Dry Season Karurumo	<2.2*10-19	significant difference

Table 5: Results of individual t-tests analyzing water level in Githambara and Karurumo for wet and dry seasons

The results of the individual t-tests are presented above. An extremely low p-value ($p<0.05$) allows us to neglect the null hypothesis and confirm that the difference in mean water level between all four variables was statistically significant. While knowing that location and precipitation affect water level was important, the next step was to quantify how much they affect water level. A multi-variable linear regression model was used to quantify this. Before running the model an assortment of graphs and figures were developed to visualize the data in

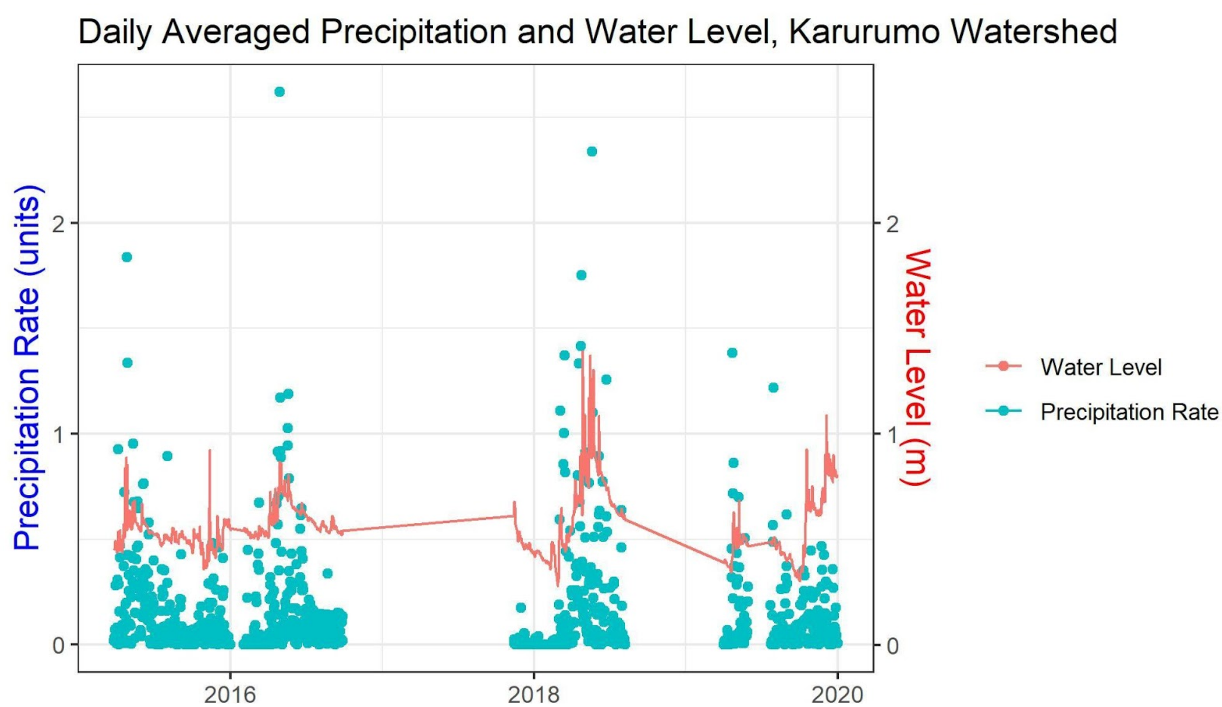


Figure 22: Daily averaged precipitation plotted against water level in Karurumo (C) from 2015 to 2020

order to confirm the accuracy of the linear regression model as well as better interpret the results of the model.

The graph above displays daily averaged precipitation plotted against water level data in Karurumo. Precipitation data was taken from various coordinates surrounding the watershed; these values were averaged based on location. The hourly location averaged data was then averaged again based on the day to limit the amount of data points and make the graph easier to read. From this graph it appears that extreme weather events (high precipitation storm events) have an effect on water level. It should also be noted that rain events fluctuate yearly as well as seasonally as evidenced by the significantly higher amount of daily averaged rainfall events above 0.5 mm/year from 2018-2019 as compared to other years. Lastly, it should be noted that there is a gap in data for the majority of 2017.

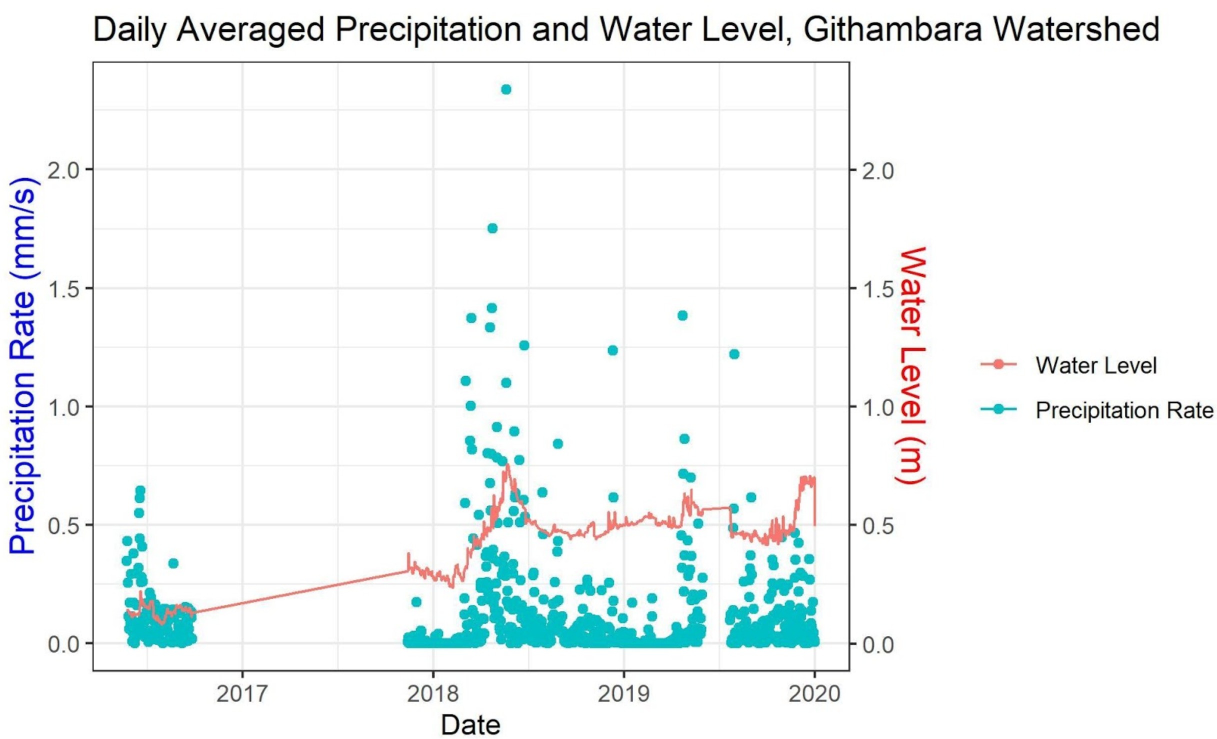


Figure 23: Daily averaged precipitation plotted against water level in Githambara (T) from 2015 to 2020

The graph above displays daily averaged precipitation plotted against water level data in Githambara. From this graph it appears that extreme weather events (high precipitation storm events) have an effect on water level in Githambara as well as Karurumo. Similar to Karurumo, Githambara is missing data for the majority of 2017.

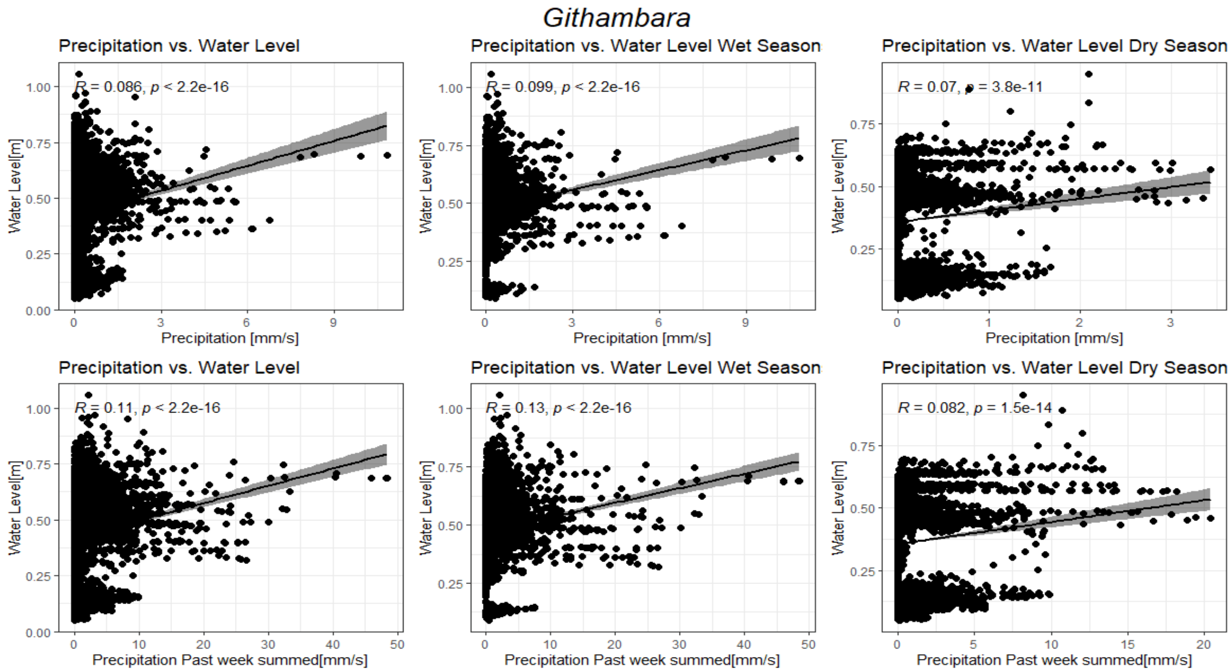


Figure 24: Precipitation vs. Water Level for Githambara and the weekly "clumped" precipitation rate vs. water level for Githambara (Treatment)

The collection of graphs above show the Precipitation vs. Water Level for Githambara and the weekly "clumped" precipitation rate vs. water level for Githambara. While the majori of data is clustered towards the left side of the x-axis (majority of days saw 0 mm/hr of rain), an

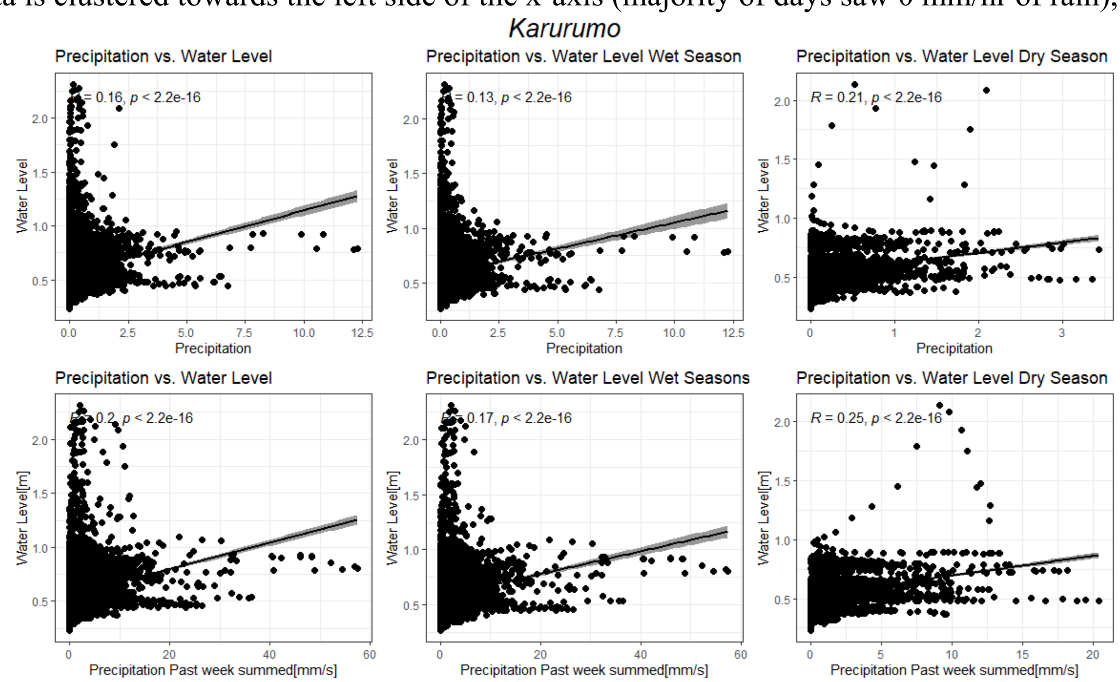


Figure 25: Precipitation vs. Water Level for Karurumo and the weekly "clumped" precipitation rate vs. water level for Karurumo (Control)

extremely low p-value ($p < 0.05$) in each graph allows us to reject the null hypothesis and confirm that the relationship between the two variables is significant. Every figure showed a slight positive correlation between precipitation and water level. Additionally, for every season, the rain clumped correlation was slightly higher, and p-value equal to or lower than the regular precipitation value, and thus was chosen to be used in the linear regression model. In Githambara, it appears that there is a slightly higher correlation between water level and precipitation rate in the wet season when compared to the dry season.

Similarly for Karurumo, every figure showed a slight positive correlation between precipitation and water level, and for every season, the rain clumped correlation was slightly higher, and p-value equal to or lower than the regular precipitation value (Figure 25). In Karurumo it appears that there is a slightly higher correlation between water level and precipitation rate in the dry season when compared to the wet season. This is the opposite result that was observed in Githambara.

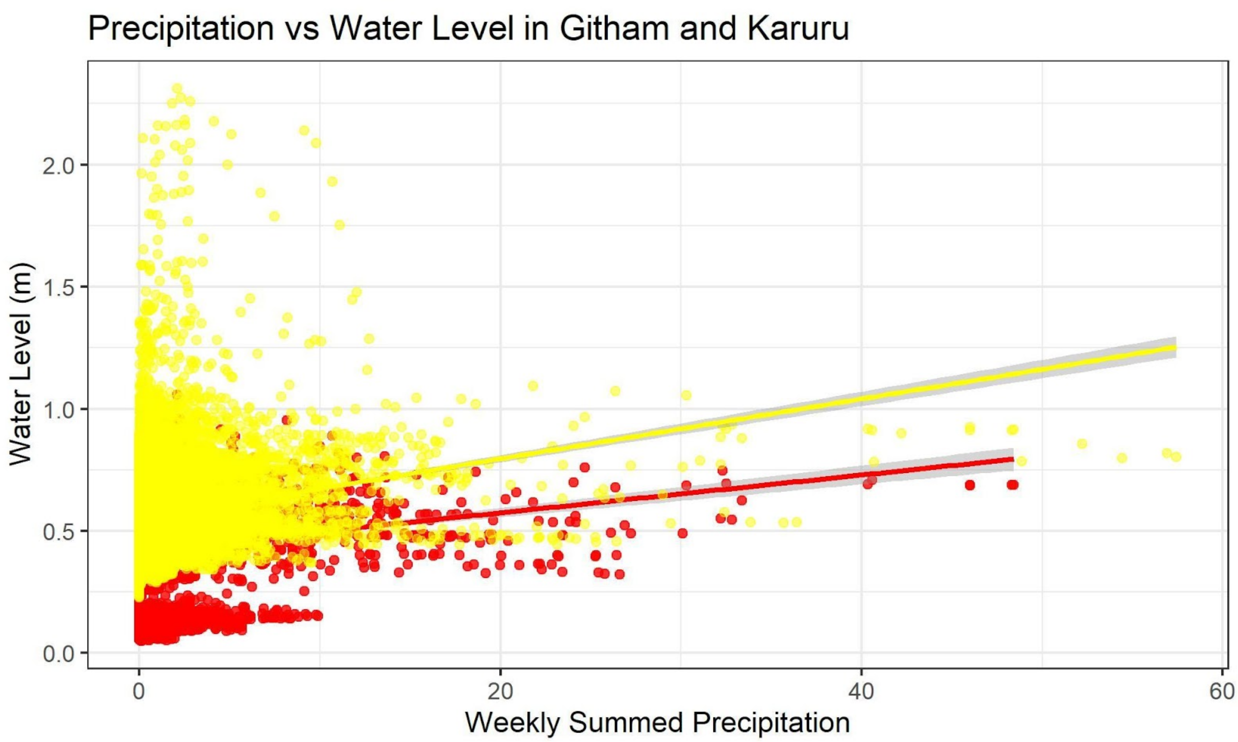


Figure 26: Weekly "clumped" Precipitation vs. Water Level for Githambara (red) and Karurumo (yellow)

The figure above takes the weekly "clumped" Precipitation vs. Water Level for Githambara (red) and Karurumo (yellow) and places them on top of one another. From this graph it was reaffirmed that on average Karurumo has higher water levels than Githambara. It can also be seen that Karurumo's water level increased at a higher rate with increase of precipitation.

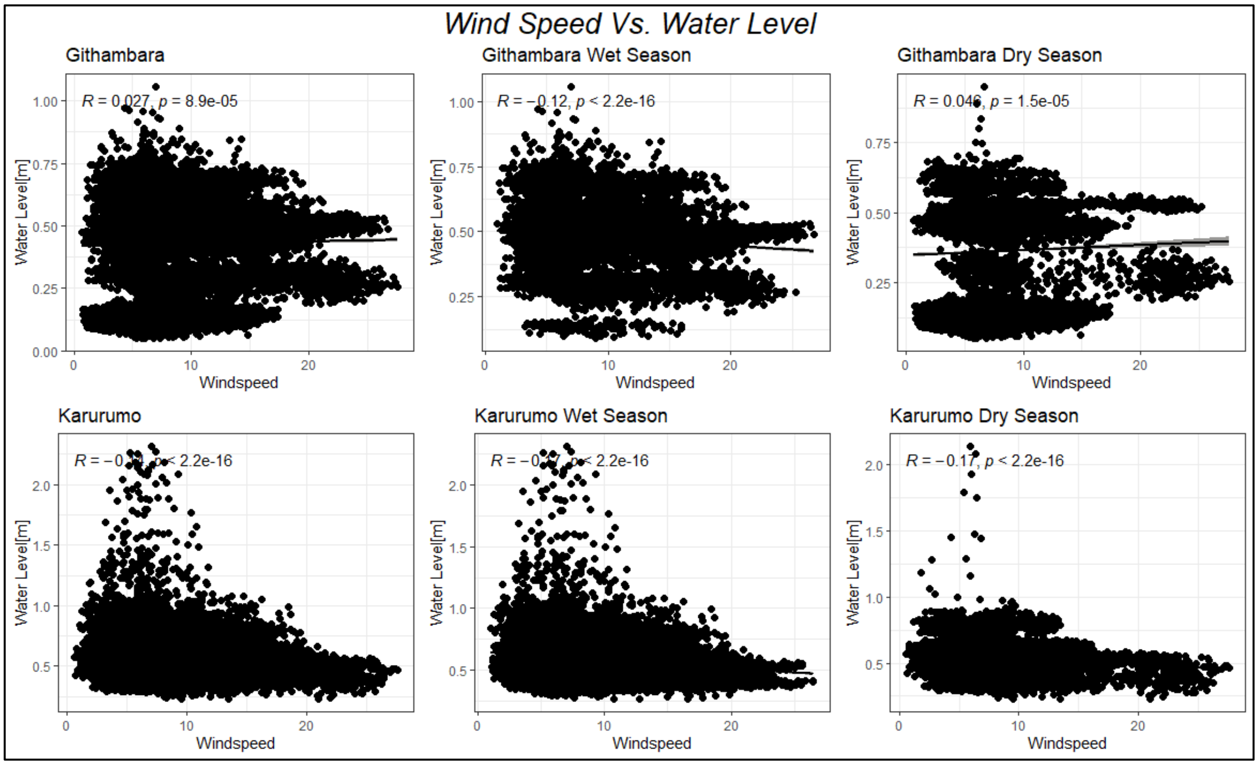


Figure 27: Wind Speed vs. Water Level for Karurumo (Control) and Githambara (Treatment) in wet and dry seasons

The collection of graphs above show water level plotted against wind speed in both Githambara and Karurumo for wet and dry seasons. Wind speed was considered a variable of interest due to its effect on evaporation. Typically, higher wind speeds contribute to higher evaporation rates. While the correlation values for Karurumo were in the same order of magnitude as the precipitation vs water level graphs, Githambara yielded values an order of magnitude lower. Additionally, a visual inspection of the Githambara and Karurumo graphs shows water level values spaced evenly across the wind speed axis suggesting that wind speed has no effect on water level. Due to the extremely low correlation values in Githambara wind speed was not considered for the linear regression model.

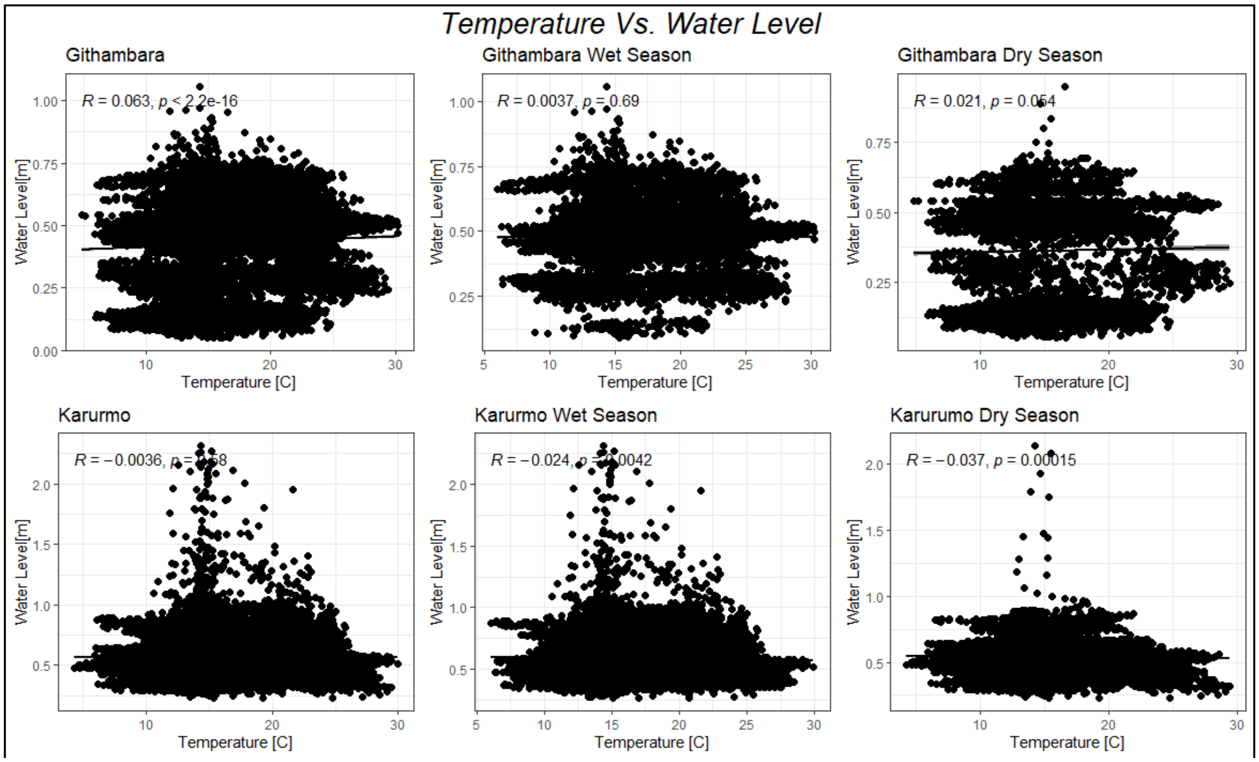


Figure 28: Temperature vs. Water Level for Karurumo (Control) and Githambara (Treatment) during wet and dry seasons

The collection of graphs above show water level plotted against temperature in both Githambara and Karurumo for wet and dry seasons. Temperature was considered a variable of interest because it plays a key role in thermal expansion (slight increase in volume of water due to increased separation of water molecules) and evaporation. The distribution of the data for every graph yields very low correlation values, which is evident from a visual inspection of the data. Also, the wet season data for Githambara wet season data yielded a p-value > 0.05 meaning the null hypothesis cannot be rejected and the difference in the two variables is not significant. For these reasons temperature was not considered in the linear regression model.

Multi-variable Linear Regression Results				
	Estimate	Std. Error	t-value	p-value
y-intercept	0.5535424	0.0010411	531.71	<2.2*10-16
Precipitation (clumped)	0.0104793	0.0003076	34.07	<2.2*10-16
Watershed	-0.1350478	0.0014662	-92.11	<2.2*10-16
Multiple R-squared	0.1834			
Adjusted R-squared	0.1834			

Table 6: Results from multivariable linear regression of water level with watershed and precipitation

The result of correlation testing resulted in a linear model considering clumped precipitation and location of watershed. The formula of the model is as follows:

$$Water\ Level = \beta_0 + \beta_1 Precipitation + \beta_2 Watershed$$

where β_0 is the y intercept and β_1 and β_2 are the weights given to precipitation and watershed respectively. Since watershed is a categorical variable (Karurumo or Githambara) dummy variables were assigned where Githambara = 1 and Karurumo = 0.

The linear regression model yielded statistically significant p-values for both precipitation and watershed, meaning it is reasonable to conclude that the slope of the interaction is not 0, thus both variables have an effect on water level. The result of the linear regression yields the following formula:

$$Water\ Level\ = 0.5535424\ +\ 0.0104793 Precipitation - 0.1350478 Watershed$$

A visual analysis of the boxplots showed that on average Karurumo had higher water levels than Githambara. The weight of the β_2 quantifies this difference in average water level as 0.1350478m. In other words the average water level is 0.1350478m higher in Karurumo when compared to Githambara. (the magnitude of this value is all that is important as the sign switches based on the classification of the dummy variable i.e. which variable was assigned a 1 or 0).

Clumped precipitation has an extremely small contribution to determining water level with a weight value of roughly 0.01. This suggests that water level remains relatively constant even with large increases in precipitation. Additionally, it suggests precipitation alone does not accurately account for changes in water level and other variables should be considered in future studies.

Previous graphs displaying the change in precipitation and water level suggested a higher correlation between water level and precipitation than predicted from the linear regression model. The linear model combined data from every year into two seasonal variables. However, as seen in previous graphs precipitation and water level varied significantly from year to year. Additionally, the model considered the combination of water level and precipitation data from both watersheds in its analysis, which may have contributed to errors in the final value of β_1. However, as seen in previous graphs precipitation and water level varied significantly from year to year. These two oversimplifications may help explain why the multiple R squared and adjusted R squared values which assess the linearity of the model were both very low with values of 0.1834. This means that the linear regression model does a poor job estimating a linear relationship between water level and precipitation and watershed location. For this reason the model should not be used for any quantitative purposes and should only be considered for observing broader trends in the data.

6. Discussion

6.1 Implications

Comparison of land cover change over time in each of the separate land cover datasets (MODIS and Copernicus) did not reveal large changes in land cover over the period 2001-2019 but this was likely due to the large spatial resolution size of the pixels. Additionally, the three different land cover datasets differed in the classification of grassland, shrubs, and cropland. Classifications of these groups are difficult to do using only remote satellite data and require validation with *in situ* measurements. Multiple land cover datasets were used to gain a comprehensive understanding of land cover in the region, as the datasets each use different algorithms to classify pixels. The Copernicus and Sentinel-2 land cover datasets are the most similar (see Appendix B) while the MODIS land cover dataset classified much of the area that is likely smallholder agriculture as grassland. The emerging narrative is that the increasing population in the region is leading to forest being converted to smallholder farms (TNC 2015); however this trend did not emerge in analysis of the three land cover datasets, and future projects should endeavor to procure higher quality data to substantiate this finding.

Analysis of ECMWF data concluded that there is a significant difference between precipitation levels between 1980-1984 and 2015-2019. While the rainfall data did not yield any noticeable seasonal changes, the severity of storm events within the rainy season seem to have increased over time, which could be related to global climate change. This finding helps to provide quantitative evidence to back up the anecdotal observations from locals about changes in precipitation.

Analyzing soil moisture with precipitation and surface temperature distributions found a positive correlation between soil moisture and precipitation, especially in areas with less forest cover, as well as a moderately negative correlation between soil moisture and temperature. There seems to be greater soil moisture retention in areas with higher forest cover, especially at greater depths. This may be important to consider for local farmers in the area when deciding how to treat soil and what agricultural techniques to use. In addition, increases in surface runoff were found over time, which could lead to greater erosion and thus worse water quality in the area.

Climate change and population growth are both predicted to have effects on water flow and quality in the region. The IPCC projects that precipitation will become more variable with increases during wet seasons and decreases during dry seasons while temperature is projected to increase for all seasons for the region (Mango et al., 2011). Because precipitation is positively correlated with soil moisture, with more variable precipitation, variability in soil moisture may also increase, leading to difficulties for agriculture especially in a region reliant on flood recession farming. During rainy seasons, increased rain intensity and floods may bring about too much water and moisture, while during dry seasons, increased droughts and higher temperatures may lead to a lack of water and soil that is too dry for crops. This amplifies the need for appropriate interventions and water management strategies in the region.

By comparing the change in water quantity and water quality for two intervention sites and control sites, data showed that there has been an increase in water level and a decrease in TSS and turbidity in both intervention sites.

From the statistical analysis comparing Githambara and Karurumo it is reasonable to conclude that the location of the site has the greatest influence on water level when compared to meteorological factors. Additionally, water levels in Karurumo were higher on average than those in Githambara. Karurumo water levels appeared to be more sensitive to increases in precipitation when compared to Githambara. In other words as rainfall increases Githambara water levels had a lower increase than Karurumo. This finding is important when considering that one of the goals of TNC interventions was to limit sediment runoff into streams during the rainy seasons. It is important to consider that comparing water level on the microshed level has inherent error. There are an assortment of factors that can contribute to variances in water level besides TNC interventions including slopes, elevation, land cover, human interventions etc. For this reason, while we can reasonably assume from our results that the TNC interventions have improved water quantity and water quality significantly during the project time, more detailed analysis is required to determine the total effect of the Water Fund's interventions.

6.2 Limitations

The team completed this project working entirely remotely and separately due to the COVID-19 pandemic, which caused many limitations due to the nature of not being on the ground in Kenya. Rather than being able to experience conditions first hand in the Upper Tana Basin, the team has relied on literature, data provided by the IBM and TNC, and open access data. There was very little on site data collection before the TNC initiated the water fund, and as a result the majority of the data was collected very recently (2015 or after). This made it difficult to analyze the state of the watershed before the TNC intervened, and as a result historical baseline records were largely ignored when analyzing the TNC's interventions. The data provided by TNC and IBM primarily consists of meteorological data and stream characteristics, and there are many temporal gaps and irregularities within the data. The timing of meteorologic and hydrologic data collection by ground instruments in the watershed is sporadic, with long, irregular intervals between data collection periods. That prevented us from accessing details about some of the micro watersheds and hindered some of our analysis and evaluation on the efficacy of those interventions put in place by TNC. Furthermore, much of TNC's data on specific interventions and their geographic coordinates were unavailable, or classified. This made understanding the interventions in place difficult, and required creativity when deciding how to assess intervention efficacy.

Much of the data used that was not from IBM or TNC, such as data for soil moisture, runoff, and landcover, was from models developed from satellite imagery. Satellite data is most effective when models can be calibrated with on site measurements. Specifically, it is difficult to classify what is grassland, savanna and agricultural land based on satellite imagery alone, since from the image a pixel of cultivated land may easily be mistaken for grassland or savanna. For example, a land cover map can be corrected by physically examining whether the classifications of the model actually represent what they intend too. Being limited to relying solely on the satellite data with no calibration invites higher probability of errors. The analyses we have performed could be strengthened in the future by verifying the satellite data with on the ground measurements in the Upper Tana River Basin.

In addition, the team faced limitations due to time constraints and complexities of certain models, which prevented us from having an even more comprehensive analysis. These limitations include the difficulties of downscaling climate projections from CMIP, which could have provided more insight on climate change and climate variability in the region, as well as the inability to use SWAT in the time we were given for the project which may have allowed us to develop a more thorough hydrologic assessment of the region.

6.3 Recommendations

More consistent data collection on variables related to water quality, including TSS and turbidity can strengthen the analysis of water conditions and the interventions put in place by TNC. To help with future projects on the Water Fund, creating a repository of relevant shape files that will help consultants trim maps to specific regions would be beneficial. Expanding the range of land surface data collected and verifying satellite data with on site measurements would also help and make future analysis easier. Through this project, the team recorded code of tests performed to analyze data such as comparing water level with precipitation and the presence of interventions. Other models, such as the rating discharge curve were already made before we joined the project. Applying these analyses to all the other micro watersheds in the Upper Tana River Basin would help to develop a more thorough view of conditions in the region.

Due to time constraints there were a number of analyses the team wanted to pursue but were unable to complete. For example, assessing the validity of the weather station's rain data when compared to on the ground measurements would help justify the usage of the data in areas where local collection was lacking. Additionally, considering more variables such as different land cover type percentages in Githambara and Karurumo may be a good way to improve the linear correlation for the regression model mentioned previously in the discussion section. This would likely require on the ground land cover observations as the pixels derived from available satellite datasets are not small enough to provide information at the individual watershed scale. While soil moisture, land cover, and runoff were all considered in this analysis important characteristics such as soil type, and sloper were largely left out. In order to gain a more holistic understanding of soil erosion these factors should be taken into account. The Food and Agriculture Association (FAO) has an open source platform on their website that allows for the download of shapefiles containing soil type data and metadata for the entire globe. Similarly, The International Soil Reference and Information Centre (ISRIC) has a database with data and metadata on soil type specific to Kenya (KENSOTER). Additionally, slope values can be extracted from a Digital Elevation Model (DEM). An open source DEM which is available for Kenya with 30 m resolution is available courtesy of The Regional Centre for Mapping of Resources for Development (RCMRD).

Based on our results and evaluation of the efficacy of the interventions at the micro watershed scale, we would recommend the Water Fund to continue to expand its deployment of interventions. Further analysis of the interventions can be conducted by applying the analytical techniques we used to more local areas could help evaluate water conditions and the effect of interventions with more granularity. To analyze location specific results of interventions, the exact locations of the interventions are necessary. For a more in depth study of the interventions it would be beneficial for specific data collection devices such as stream gauges to be implemented both downstream of the intervention (dependent variable), and upstream of the intervention (control). This would allow more local case studies to be conducted. Our project has laid the groundwork for future research in the Tana River Basin by creating a comprehensive literature repository of past studies in the region and access to code in the IBM CloudPak. This code can also act as a starting point for analysis of future data that will be collected.

Because the interventions put in place by TNC are meant to improve water quality and quantity for the benefit of local farms and communities, it is also important to include the voices of local people when studying the effects of the interventions. Communicating directly with local farmers in the Upper Tana River Basin and listening to their experiences with water and agriculture before and after the interventions were put in place would be highly beneficial. First hand accounts of experiences with the interventions could provide valuable insight on whether or not the interventions are actually effective in the community, and if the results from analysis of data match with the communities experiences. This could be completed through surveys or interviews. In addition, because the interventions are meant to benefit local communities, it may be beneficial to look beyond the effect that the interventions have on water resources to other social factors such as financial gain for the local farmers or improved quality of life.

6.4 Deliverables

We stored all code and data used in a repository to document our work. We used the IBM Cloud Pak database to store this information, creating a water data platform that ensured our project's replicability. The code repository will be uploaded onto a GitHub page. Furthermore, we have developed this final report, recounting all our research findings to be presented to our clients at the completion of the project. We have also created visual communication products in the form of multiple ArcGIS StoryMaps, one for each research question, to present our data and findings in a way suitable not only to our clients but also to local stakeholders, including farmers and government representatives. The table below describes each of the deliverables we have created.

Table 7. Description of Created Deliverables

Deliverable	**Description**
Hydrological Assessment	• Use water flow data and perform analysis to determine change in water availability of the basin based on micro-watersheds, comparing control sites and experimental sites. • Use discharge to compare the basin's hydrological characteristics before and after TNC intervention • Use soil moisture to analyze hydrological characteristics of specific sub basins. A comparison will be done between areas TNC has and has not intervened • Create models and graphs describing the findings, and statistical significance **Found in Section 5.3** **Found in Appendix E**
Conceptual Model	• A flow-chart type model for overall understanding of the watershed • Providing inputs and outputs within the watershed • Goes hand-in-hand with our Data Dictionary for getting those unfamiliar with the Upper Tana River Basin up to speed. **Found in Section 2.5.2**
IBM Cloud Pak for Data (Code Repository)	• IBM Cloud Pak for Data will enable integration of data and assets created • Upload relevant open source and open access data (IPCC, TNC Box, literature review) to IBM Cloud Pak to create a centralized online platform with information on water funds • Use remote sensing (LandSat, Sentinel 2 &3) and machine learning to fill data gaps in TNC Box and DHIS2 client database • Python/R scripts to record methodology and meaningful insights for future reference • Easily updatable dashboards that can be displayed publicly (TNC or UCLA website) • The code is compiled onto GitHub to ease sharing capabilities, and provide resources for future work **Found in Section 5.3** **Found in Appendix D**

Data Dictionary	• A collection of sources and datasets used to understand the Upper Tana River Basin, and used to develop our code and geographical products • Dataset information includes: Datapack name, dates of relevance, uses, and file type. • This will provide any future groups who build on our work a clear pathway to the data we had available. • Transparency with our data also encourages replicability of our work, and the verification of our findings regarding the efficacy of TNC interventions, and implications and recommendations for the Upper Tana River Water Fund moving forward.
Final Report	• Research findings/trends and comparison with outside sources • Describe the implications, future recommendations, and limitations of the project. • Visualizations of quantitative/qualitative findings through graphics and maps • Annotated bibliography of literature sources relevant to watershed management in the Tana River region to provide continuity to the Water Fund. This will include reports by TNC and studies by other researchers of how climate change and land use change have affected hydrology in the past and how they are predicted to affect it in the future
Communication Products	• ArcGIS StoryMaps to provide simple to use, interactive web-based storytelling products presenting data & findings for our clients. • Presentation of current status, impact of interventions, future projections **Found in Appendix F** • Shiny web application, an interactive website, presents data and findings of research question 3 • The shiny application consists of a server. R script and a UI. R script. Clients can easily modify the website content by changing the R script. **Found in Appendix G**
Land cover maps	• Analysis of satellite data to determine change in land characteristics of the basin based on land use change over time • Annual land cover maps for 2001-2019 using the MODIS MCD12Q1 Version 6 IGBP land cover classification scheme, 463x463 m resolution resampled to 30x30 m resolution • Sentinel-2 MSI land cover map for 2016 • Copernicus GCLS version 3 land cover maps for 2015-2019 • Copernicus CGLS land cover maps for a 100 meter buffer area surrounding the Githambara and Karurumo watersheds **Found in Section 5.1** **Found in Appendix B**
Climate variability projections	• Clear figures representing the changes in precipitation in recent years • Projections on future precipitation and climate change, providing perspective on what farmers are to expect in the near future. • Created using The Weather Channel data and spatio-temporal temperature and precipitation data **Found in Section 5.2** **Found in Appendix D**

A key goal of our project is creating communication products conveying project data and results. We worked with our clients to determine the most effective platform for communicating results to the stakeholders, with the overall goal of prioritizing the needs of the client and the communities residing in the Upper Tana River Basin. Our clients emphasized creating an easily digestible narrative that informs key stakeholders and public audiences within the Water Fund boundaries of our research's implications. We utilized ArcGIS Online's StoryMap feature to create interactive websites (Appendix F) to inform stakeholders of river catchment conditions and build a fuller picture of socio-ecological interactions in the watershed. More specifically, each research question is presented in separate story maps, and outlines our team's methodology and the overall significance of the research questions and objectives.

7. Conclusion

The Upper Tana Nairobi Water Fund established by the Nature Conservancy to ensure water security in the coming decades. After analyzing climate data, hydrological data, soil moisture data and remote sensing imagery, we concluded that the Water Fund should continue encouraging farmers in the basin to use sustainable farming techniques like agroforestry and crop terracing. Additionally, more data should be collected to better understand changes in land cover, precipitation and soil moisture. Methods are documented and replicable. Despite impediments and limitations during our project, we have used multiple methods of statistical analysis and remote sensing to determine that TNC interventions are effective at securing water availability in the Upper Tana River Basin.

We have recommended certain actions for TNC, and for the Upper Tana River Water Fund to adopt and implement moving forward. These recommendations are expected to further improve water fund efficacy, and to streamline further analysis. We have compiled all of our work into multiple deliverables, each serving a unique purpose.

8. Acknowledgements

This project benefited from the valuable contributions of our partners and industry experts. These individuals offered themselves as resources to our team during various stages of the project, including during the project proposal and brainstorming, the data procurement and development of methods, the analysis of data, and with review of final products and recommendations.

Dr. Erik Porse, Ph.D. - Sacramento State Office of Water Programs and UCLA IoES
Dr. Travis Longcore, Ph.D. - UCLA IoES
Lee Alexanderson, PE, ENV SP - County of Los Angeles Department of Public Works
Dr. Mike Antos, Ph.D., ENV SP - Stantec Consulting and UC Irvine Anthropology
Melissa Guerrero and Joe Mulligan - Kounkuey Design Initiative
Dr. Sanjay Mohanty, Ph.D. & Renan Valenca - UCLA Samueli School of Engineering
Scott Gruber - UCLA IoES
Noah Garrison J.D. - UCLA IoES
Dr. Tracy Baker, Ph.D. - The Nature Conservancy

Bibliography

Abell, R. et al. (2017). Beyond the Source: The Environmental, Economic and Community Benefits of Source Water Protection. The Nature Conservancy, Arlington, VA, USA.

Adhikari, U., Nejadhashemi, A. P., & Woznicki, S. A. (2015). Climate change and eastern Africa: a review of impact on major crops. *Food and Energy Security*, *4*(2), 110-132.

Anderson, J. (2021). A Kenyan Ecologist's Crusade to Save Her Country's Wildlife. The New Yorker. Available at: https://www.newyorker.com/magazine/2021/02/01/a-kenyan-ecologists-crusade-to-save- her-countrys-wildlife

Antonaropoulos, P. & Karavokyris, G. (2010). Upper Tana Physiographic Final Report, Water Resources Management Authority, Kenya.

Apse, C. et al. (2016). Sub-Saharan Africa's Urban Water Blueprint: Securing Water Through Water Funds and Other Investments in Ecological Infrastructure. The Nature Conservancy: Nairobi, Kenya.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development 1. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73-89.

Baker, T., Kiptala, J., Olaka, L., Oates, N., Hussain, A. & McCartney, M. (2015). Baseline Review and Ecosystem Services Assessment of the Tana River Basin, Kenya. Colombo, Sri Lanka: International Water Management Institute (IWMI). 107p. (IWMI Working Paper 165). doi: 10.5337/2015.223.

Beukering, P., Moel, H., Botzen, W., Eiselin, M., Kamau, P., Lange, K., Maanen, E.V., Moğol, S., Mulwa, R., Otieno, P., Overgaauw, N., Papyrakis, E., Wasonga, V., & Weert, F.V. (2015). The Economics of Ecosystem Services of the Tana River Basin - Assessment of the impact of large infrastructural interventions.

Billgren, C. & Holmén, H. (2008). Approaching reality: Comparing stakeholder analysis and cultural theory in the context of natural resource management. Land Use Policy, 25(4), 550–562.

Bremer, L. et al. (2016). Bridging Theory and Practice for Hydrological Monitoring in Water Funds (p. 98). Latin American Water Funds Partnership.

Bryant, B. (2015). Ecosystem Services Assessment and Valuation of Proposed Investments for the Upper Tana-Nairobi Water Fund. A Technical Appendix to the Upper Tana-Nairobi Water Fund Business Case, The Natural Capital Project.

Buanes, A., Jentoft, S., Karlsen, G.R., Maurstad, A. & Søreng, S. (2004). In whose interest? An exploratory analysis of stakeholders in Norwegian coastal zone planning. Ocean & Coastal Management, 47, 207–223.

Cheatham, C. (2020). Investing in Nature to Improve Water Security (Water Funds). The Nature Conservancy.

Cheruto, M. C., Kauti, M. K., Kisangau, D. P., & Kariuki, P. C. (2016). Assessment of land use and land cover change using GIS and remote sensing techniques: A case study of Makueni County, Kenya.

De Pauw, E., & Ramasamy, S. (2020). Rapid detection of stressed agricultural environments in Africa under climatic change 2000–2050 using agricultural resource indices and a hotspot mapping approach. Weather and Climate Extremes, 27, 100211.

Dijkshoorn, J.A. et al. (2010). Soil and terrain database for the Upper Tana, Kenya. Wageningen: ISRIC – World Soil Information, Green Water Credits Report 11.

Fredette, Tiffany. (2018) "Guide to Environmental Water Quality Testing." *Hanna Instruments Blog*, https://blog.hannainst.com/guide-to-environmental-water-quality-testing

Gathagu, John. (2021). *UTNWF* [Powerpoint Presentation]. IBM/TNC Practicum Meeting.

Gathagu, J., Mourad, K. and Sang, J. (2018). Effectiveness of Contour Farming and Filter Strips on Ecosystem Services. *Water*, 10(10), p.1312.

Gangodagamage, C. (2001). Hydrological Modeling Using Remote Sensing and GIS *Centre for Remote Imaging, Sensing and Processing (CRISP), National University of Singapore; Singapore Institute of Surveyors and Valuers (SISV); Asian Association on Remote Sensing (AARS).* https://crisp.nus.edu.sg/~acrs2001/pdf/075GANGO.pdf

Hess, S. & Leisher, C. (2018). Report on the 2017 Upper Tana Nairobi Water Fund Baseline Household Survey. The Nature Conservancy.

Hoff, H., Noel, S., Droogers, P., & Dent, D. L. (2007). *Water use and demand in the Tana Basin: analysis using the Water Evaluation and Planning tool (WEAP)* (No. 4). ISRIC-World Soil Information.

Hunink, J. E., Niadas, I. A., Antonaropoulos, P., Droogers, P. & de Vente, J. (2013): Targeting of intervention areas to reduce reservoir sedimentation in the Tana catchment (Kenya) using SWAT, Hydrological Sciences Journal, DOI:10.1080/02626667.2013.774090.

Hunink J.E. & Droogers P. (2015). Impact Assessment of Investment Portfolios for Business Case Development of the Nairobi Water Fund in the Upper Tana River, Kenya.

Joslin, A. J., & Jepson, W. E. (2018). Territory and authority of water fund payments for ecosystem services in Ecuador's Andes. *Geoforum*, *91*, 10–20. https://doi.org/10.1016/j.geoforum.2018.02.016

KIPPRA (2007). The Kenya Agricultural Sector Data Compendium: Volume Two [II]: Crop Production. Report in collaboration with the Ministry of Agriculture, December 2007.

Knoop, L., Sambalino, F. & van Steenbergen, F. (2012). Securing Water and Land in the Tana Basin: a resource book for water managers and practitioners. Wageningen.

Kuusisto, E. (1996). Water Quality Monitoring - A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes. *United Nations Environment Programme and the World Health Organization*.

Ladson, A. (2018). Visualizing Hydrologic Data. *Waterway Ecosystems Research Group, School of Ecosystem and Forest Science, University of Melbourne Vic 3010.* https://drive.google.com/file/d/1_-tHuEgR5cUClnrRUT8UFGEKs0HSu3Il/view

Langat P. K., Kumar, Lalit, Koech, Richard (2017). Temporal Variability and Trends of Rainfall and Streamflow in Tana River Basin, Kenya. Sustainability 9, 1963.

Leauthaud, C., Belaud, G., Duvail, S., Moussa, R., Grunberger, O., & Albergel, J. (2013). Characterizing floods in the poorly gauged wetlands of the Tana River Delta, Kenya, using a water balance model and satellite data.

Leauthaud, C., Duvail, S., Hamerlynck, O., Paul, J.-L., Cochet, H., Nyunja, J., Albergel, J., Grünberger, O. (2013). Floods and livelihoods: The impact of changing water resources on wetland agro-ecological production systems in the Tana River Delta, Kenya. Global Environmental Change, 23, 252–263.

MacDonald, A. M., Bonsor, H. C., Dochartaigh, B. É. Ó., & Taylor, R. G. (2012). Quantitative maps of groundwater resources in Africa. *Environmental Research Letters*, *7*(2), 024009.

Mango, L. M., Melesse, A. M., McClain, M. E., Gann, D., & Setegn, S. G. (2011). Land use and climate change impacts on the hydrology of the upper Mara River Basin, Kenya: results of a modeling study to support better resource management. *Hydrology and Earth System Sciences*, *15*(7), 2245-2258.

McNally, Amy. (2018). NASA/GSFC/HSL, FLDAS Noah Land Surface Model L4 Global Monthly 0.1 x 0.1 degree (MERRA-2 and CHIRPS), Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), 10.5067/5NHC22T9375G

MDOT Stormwater Drainage Manual. (2006). Chapter 3: Hydrology. *University of Michigan.*

PDF, online. https://www.michigan.gov/documents/MDOT_MS4_Chap_91716_7._03_Drainage_Manual.pdf

MoE (2012). Electricity Sub-Sector: Medium Term Plan (2012-2016). Kenya Ministry of Energy. Unpublished draft for circulation.

Natural Resources Conservation Service (NRCS). (n.d.). *Hydrology Training Series Module 106 Peak Discharge Study Guide* . https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1082994.pdf.

ODA (2005). Guidance note on how to do stakeholder analysis of aid projects and programmes.

Overseas Development Administration, Social Development Department, /www.euforic.org/gb/stake1.htmS.

Okello, C., Greggio, N., Giambastiani, B. M. S., Wambiji, N., Nzeve, J., & Antonellini, M. (2020). Modelling Projected Changes in Soil Water Budget in Coastal Kenya under Different Long-Term Climate Change Scenarios. *Water*, *12*(9), 2455.

Otieno, O.M., Abdillahi, H. S. , Wambui, E. M. , Kiprono, K. S. (2019). FLOOD IMPACT-BASED FORECASTING FOR EARLY WARNING AND EARLY ACTION IN TANA RIVER BASIN, KENYA. ISPRS. XLII-3/W8

Ozment, S. et al, & U.S. Endowment for Forestry & Communities. (2016). Protecting drinking water at the source: Lessons from United States watershed investment programs.WorldResourcesInstitute. https://www.wri.org/publication/protecting-drinking-water-source

Röling, N.G., Niels, G. & Wagemakers, M.A.E. (1998). A new practice: facilitating sustainable agriculture. In Röling, N.G. & Wagemakers, M.A.E., Facilitating Sustainable Agriculture: Participatory learning and adaptive management in times of environmental uncertainty (pp. 3–22). Cambridge University Press.

Salas, J., Govindaraju, R., Anderson, M., Arabi, M., Francés, F., Suarez, W., Lavado, W., & Green, T. (2014). Introduction to Hydrology. 10.1007/978-1-62703-595-8_1.

Saarnak, N. (2003). Flood recession agriculture in the Senegal River Valley, Geografisk Tidsskrift-Danish Journal of Geography, 103:1, 99-113, DOI: 10.1080/00167223.2003.10649483

TNC (2015). Upper Tana-Nairobi Water Fund Business Case. Version 2. The Nature Conservancy: Nairobi, Kenya.

TNC (2019). Nairobi Water Fund Annual Report.

United States Geological Survey (USGS). What is Hydrology? https://www.ecosia.org/search?q=usgs&addon=chrome&addonversion=3.3.0&method=t opbar

Veldkamp, A. et al. (2011). Mount Kenya volcanic activity and the Late Cenozoic landscape reorganisation in the upper Tana fluvial system. Geomorphology, 145–146, 19–31.

Wambua, R. M. Estimating Rainfall Intensity-Duration-Frequency (Idf) Curves For A Tropical River Basin.

Wang, S., Huang, J., Li, J., Rivera, A., McKenney, D. W., & Sheffield, J. (2014). Assessment of water budget for sixteen large drainage basins in Canada. *Journal of Hydrology*, *512*,1-15.

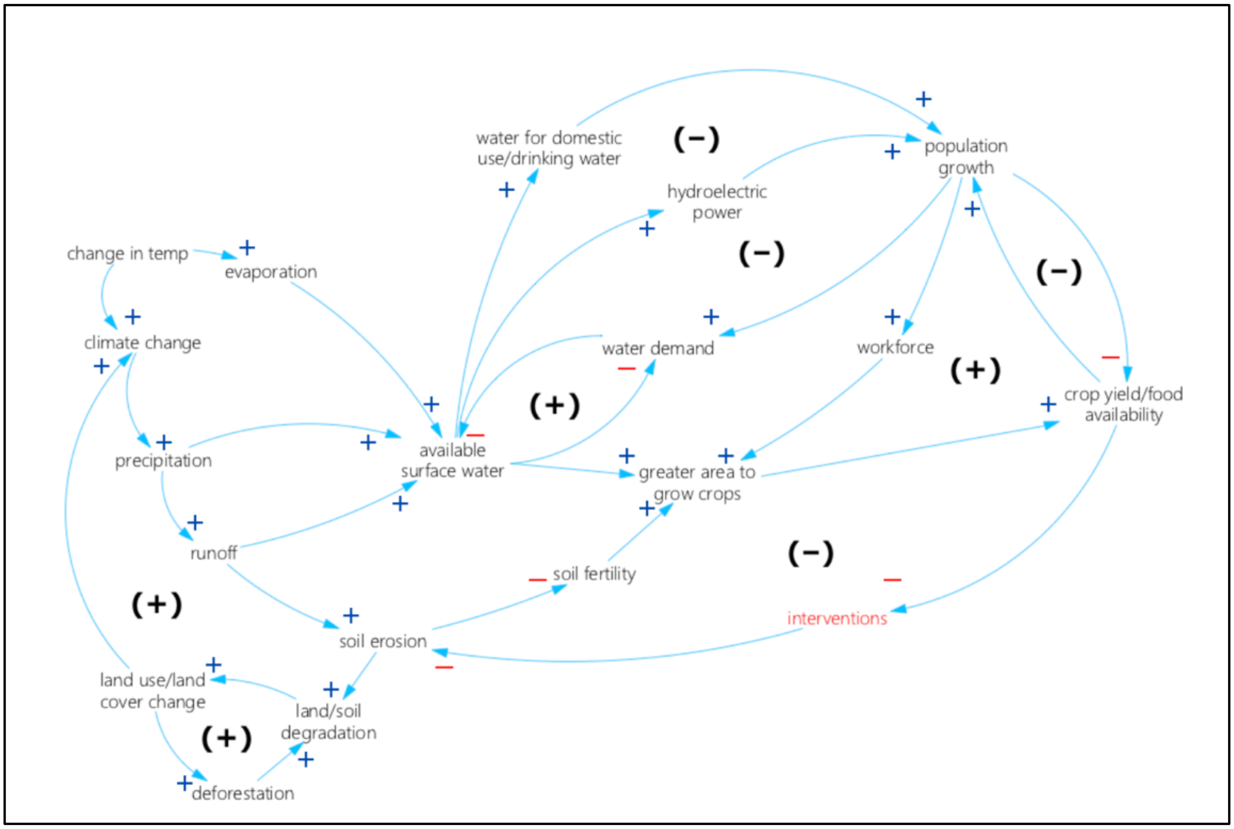
Water Resources Authority Kenya. Tana Basin Area (TBA). https://wra.go.ke/tana-basin-area-tba/

Wetlands International Kenya. Valuing the Tana River Water Resources Policy Brief. https://africa.wetlands.org/wp-content/uploads/sites/4/dlm_uploads/2017/08/Tana-Delta- TEEB-Policy-Brief.pdf

Zwarts, L., van Beukering, P.J.H., Koné, B., Wymenga, E. & Taylor, D. (2006). The Economic and Ecological Effects of Water Management Choices in the Upper Niger River: Development of Decision Support Methods. Water Resources Development, 22(1), 135–156.

Zyla, C., et al. (2018). *Water Funds: Field Guide 2018* (Water Funds, p. 83). The Nature Conservancy.

Appendix A: Feedback Loop Diagram for Tana River Basin

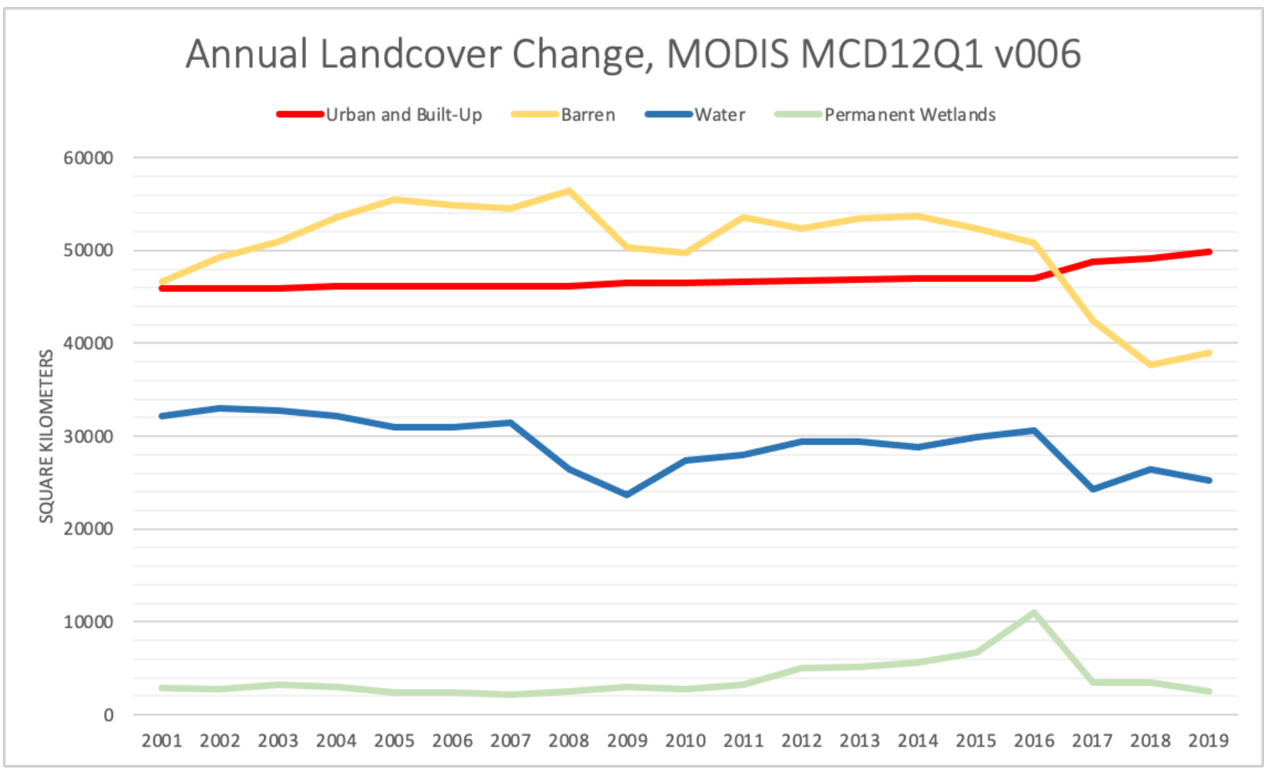


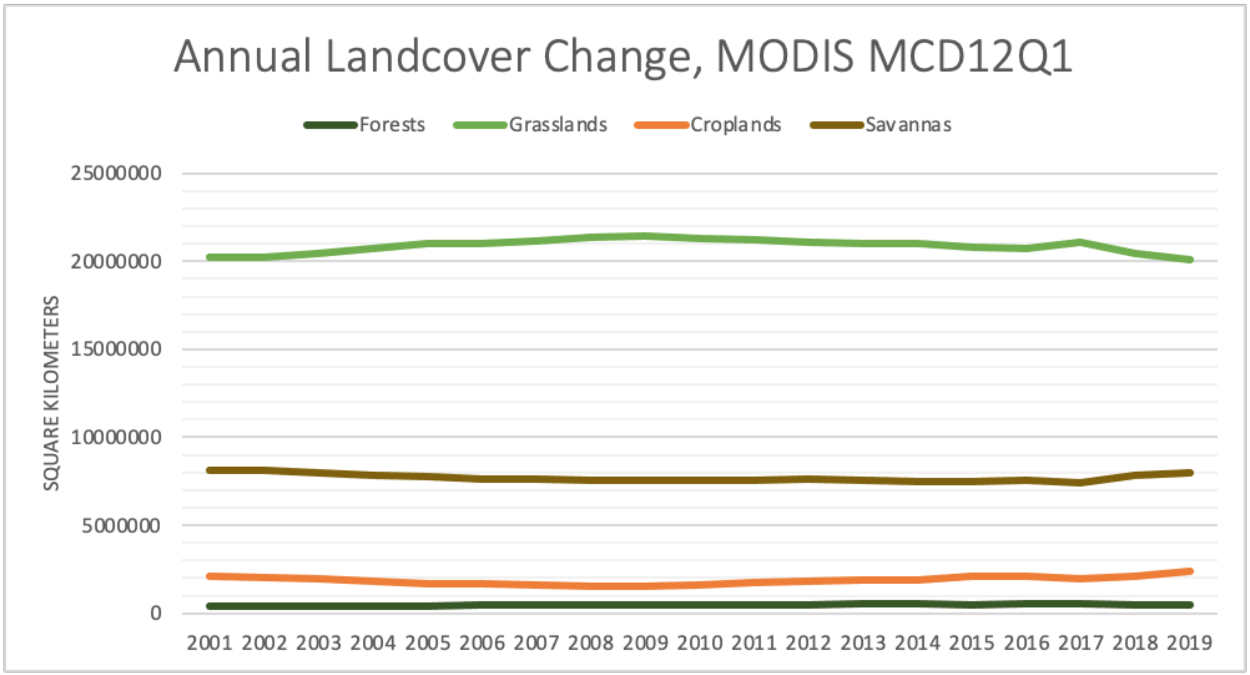
Appendix B: Code for Creating Land Cover Maps and Additional Analysis

Google Earth Engine Code for extracting land cover data:
Copernicus Global Land Cover CGLS-LC100 collection 3:
https://code.earthengine.google.com/41363504870e4855456194de05aeaaf1

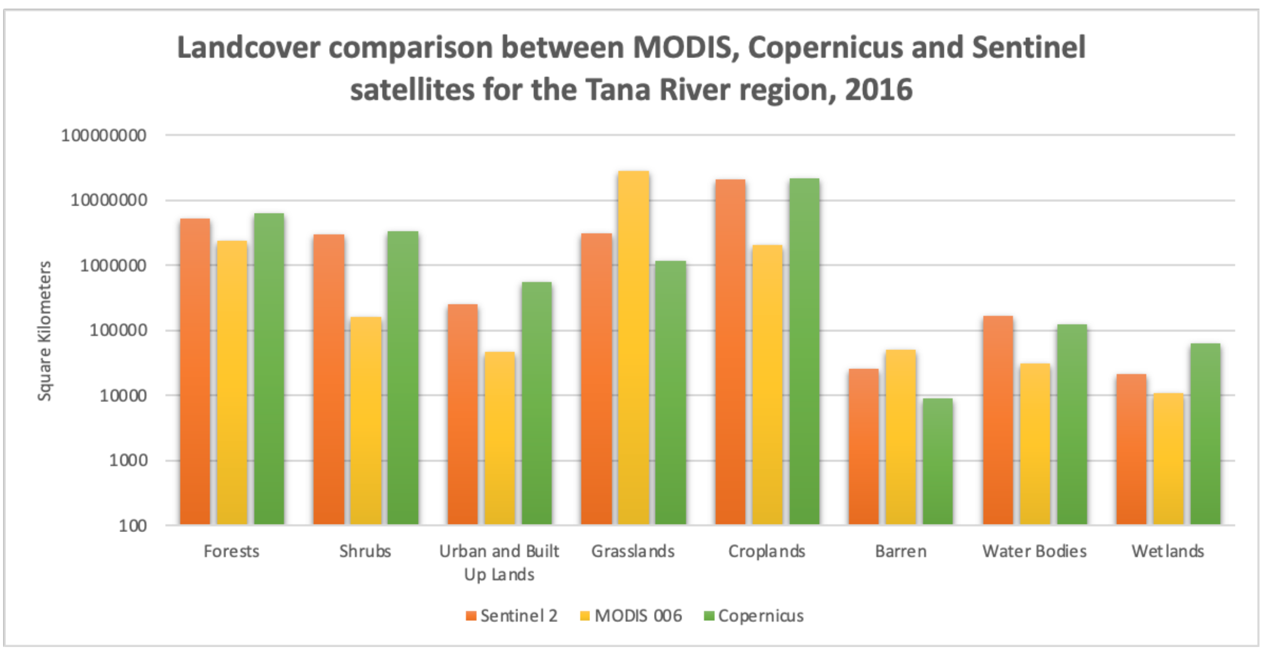
MCD12Q1.006 MODIS Land Cover Type Yearly Global 500m:
https://code.earthengine.google.com/0f1fd40fd24cbb252e11726da2d87351

Code for creating buffers around latitude and longitude points to extract land cover data for control site regions: https://code.earthengine.google.com/b4eca963b6242f3898c5d2bdac1444c6



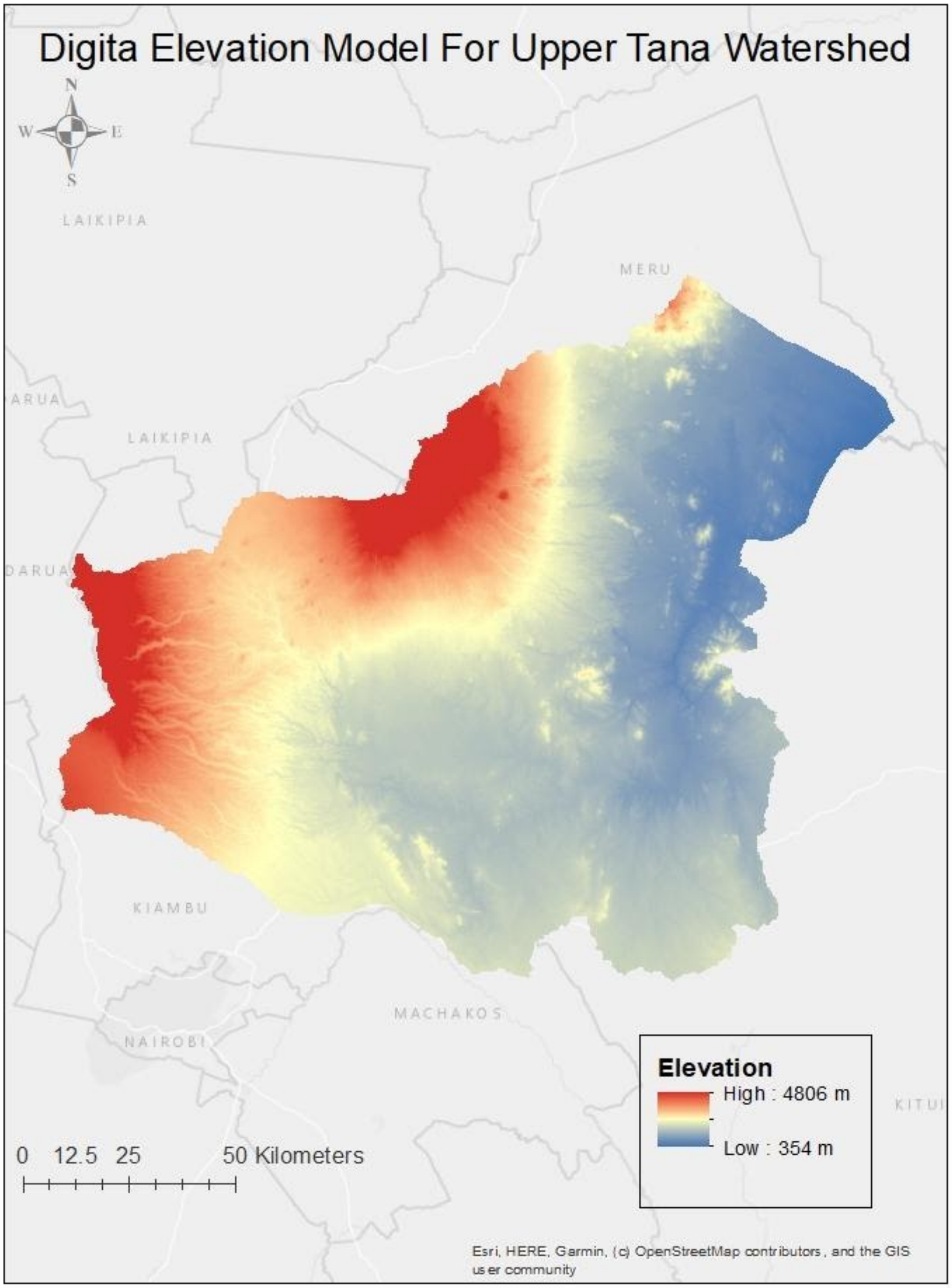


Annual Landcover Change, MODIS MCD12Q1
Forests
Grasslands
Croplands
Savannas
25000000
20000000
15000000
10000000
5000000
0
SQUARE KILOMETERS
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019

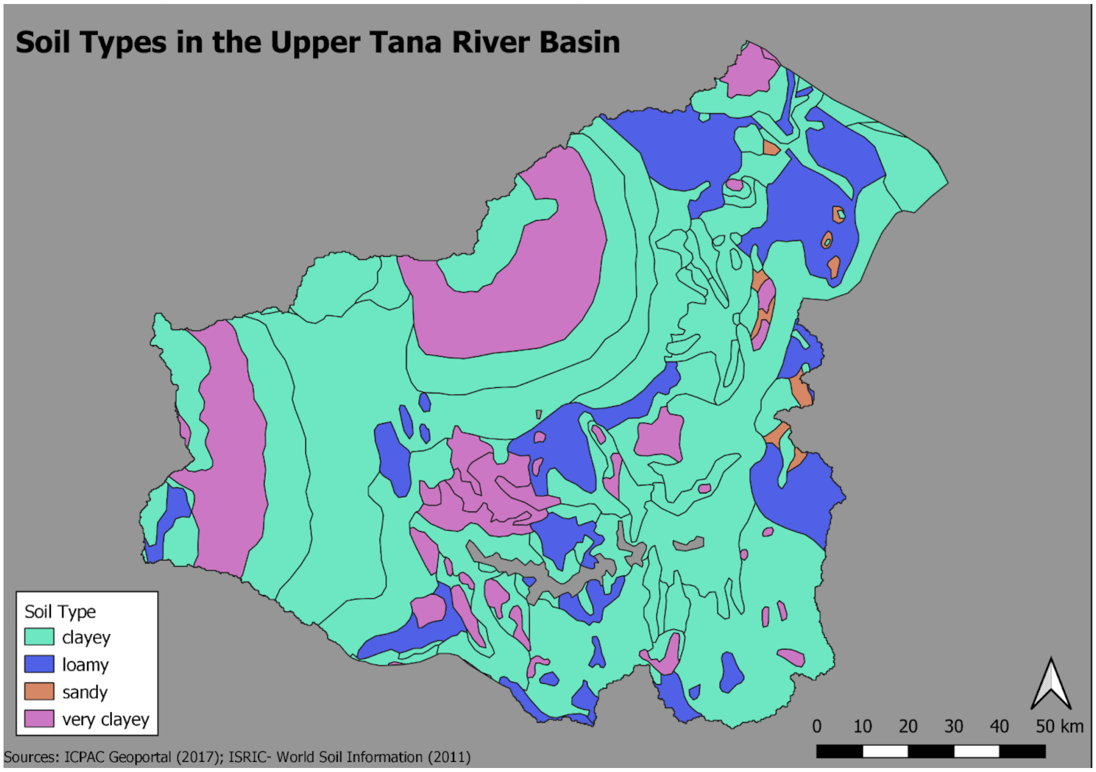


Landcover comparison between MODIS, Copernicus and Sentinel satellites for the Tana River region, 2016
100000000
10000000
1000000
100000
10000
1000
100
Square Kilometers
Forests
Shrubs
Urban and Built Up Lands
Grasslands
Croplands
Barren
Water Bodies
Wetlands
Sentinel 2
MODIS 006
Copernicus

Appendix C: Physical Characteristics of the Upper Tana River Basin

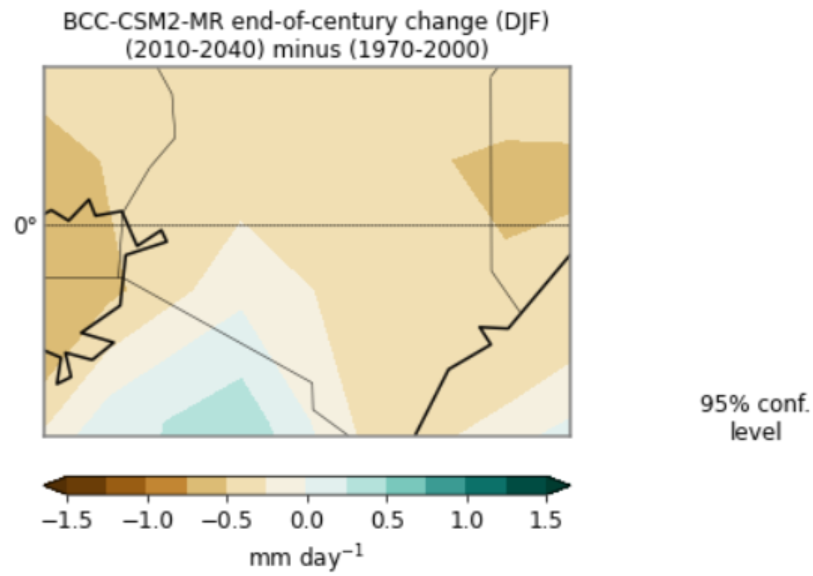


Elevation Model for the Upper Tana Watershed



Soil types in the Upper Tana River Basin

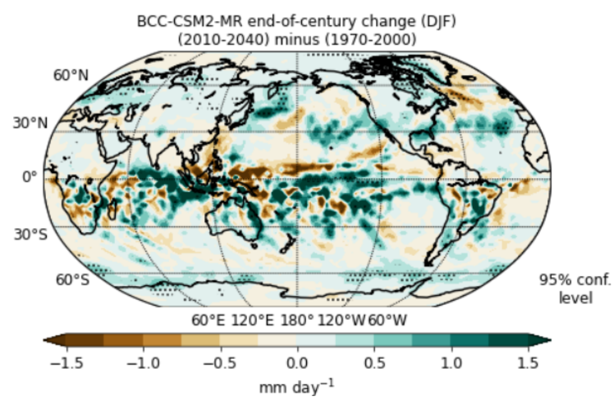
Appendix D: CMIP Analysis and Further Graphics Examining Climate Variability

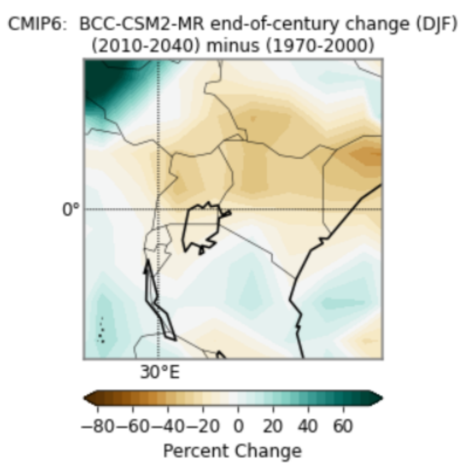


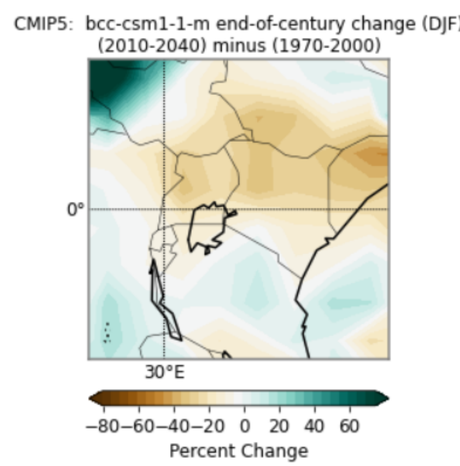
Kenyan image of CMIP 5 precipitation change between 1970-2000 and 2010-2040, focused on the months of December, January and February with 95% confidence level. Based on this graphic the Upper tana is projected to have very little precipitation change but may see some reduction in the northern section

Global End of Century Change for the months of December, January, and February.

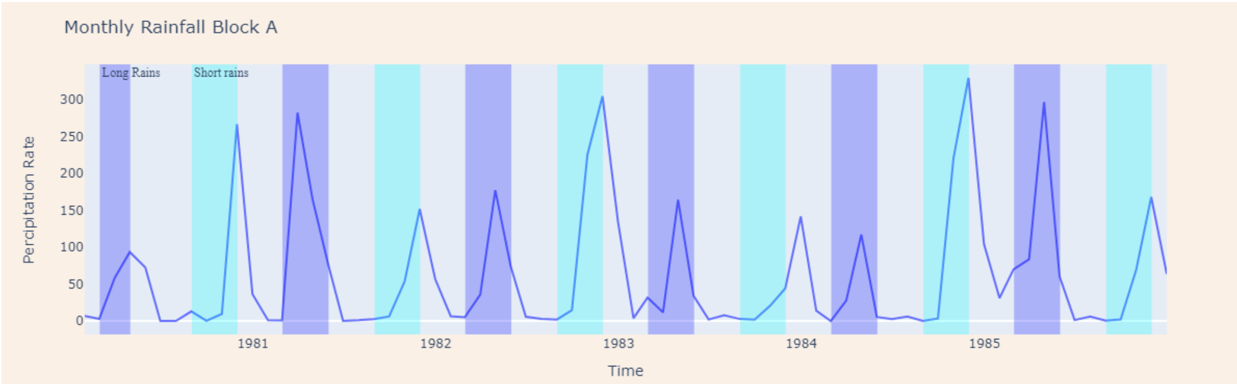
This compares the CMIP 5 and CMIP 6 data focusing on percent change between the 30 year blocks of 1970-2000 and 2010-2040 in the months of December, January and February with 95% confidence level. The goal is to provide the viewer with a qualitative understanding of the differences between the two CMIP data sets.

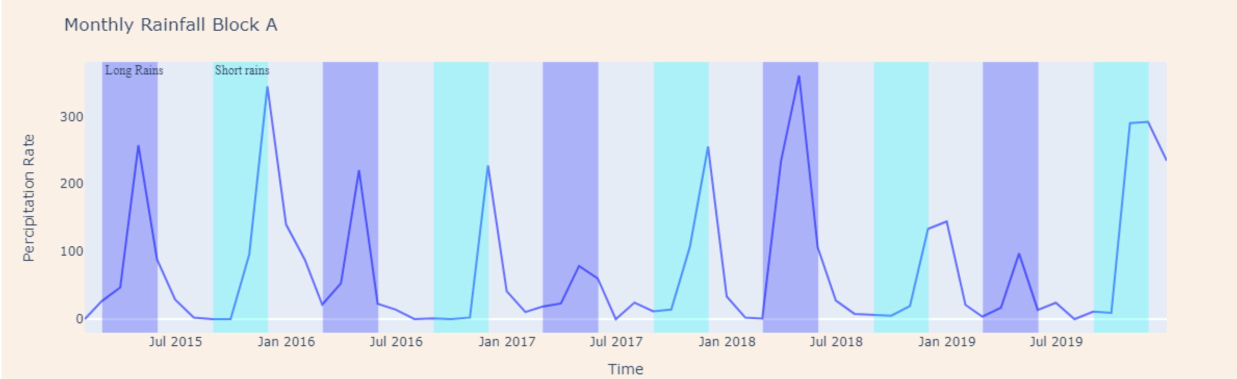


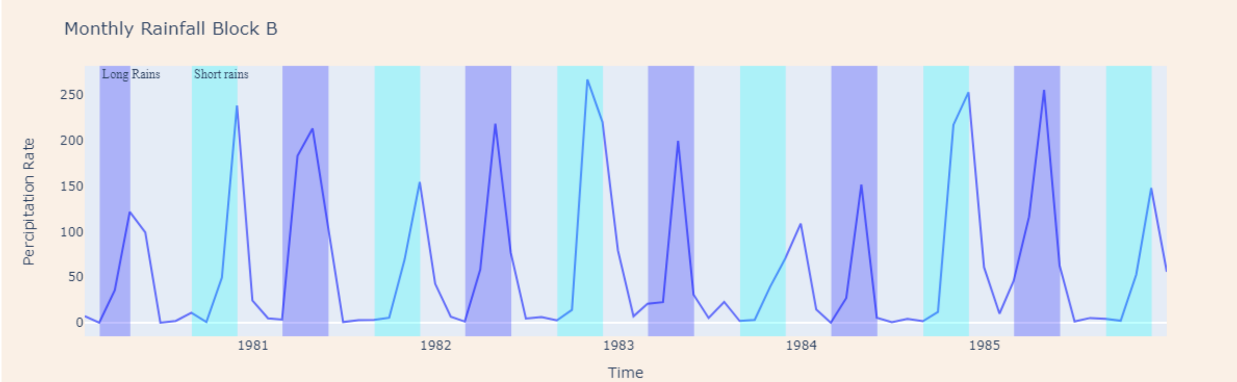


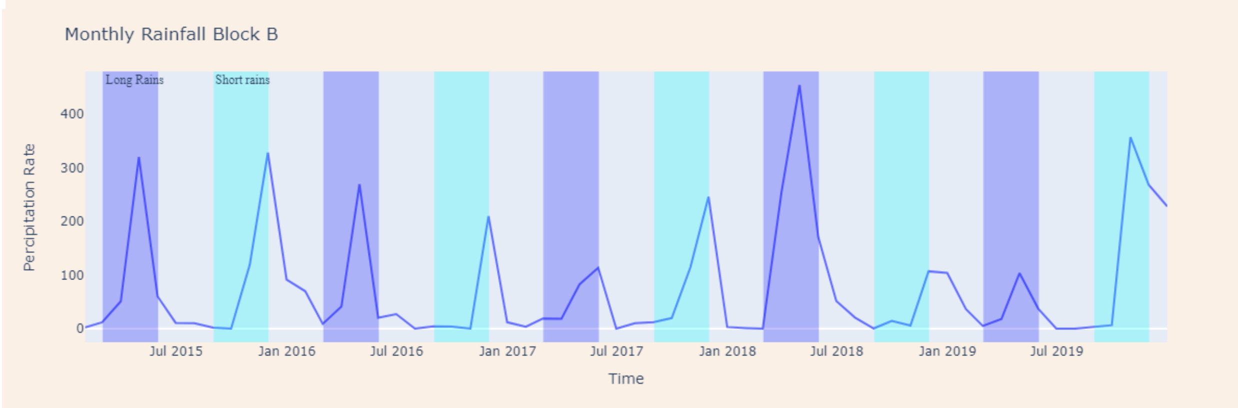


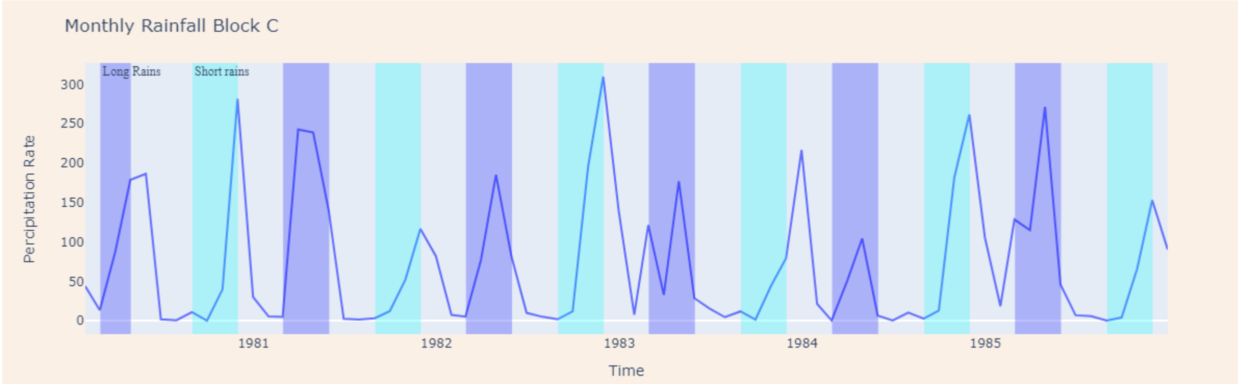
ECMWF Monthly Precipitation Graphs 2015-2019 and 1980-1985

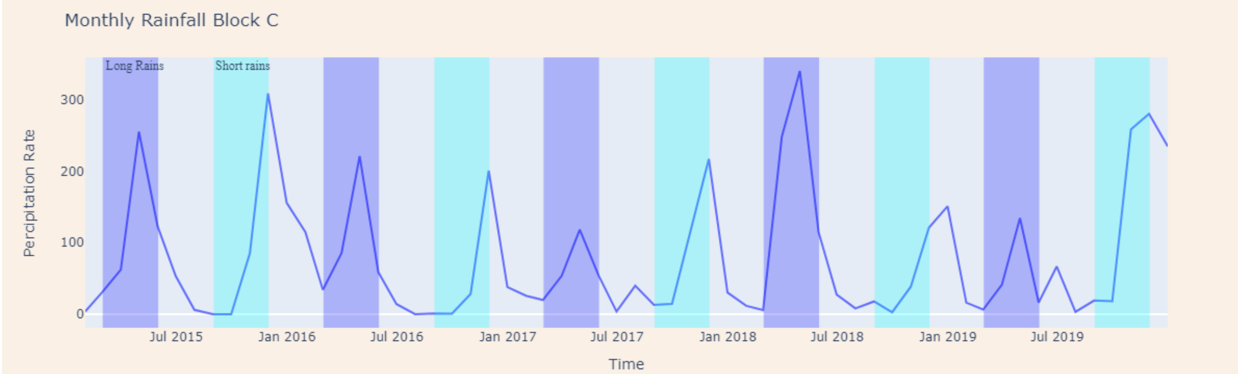


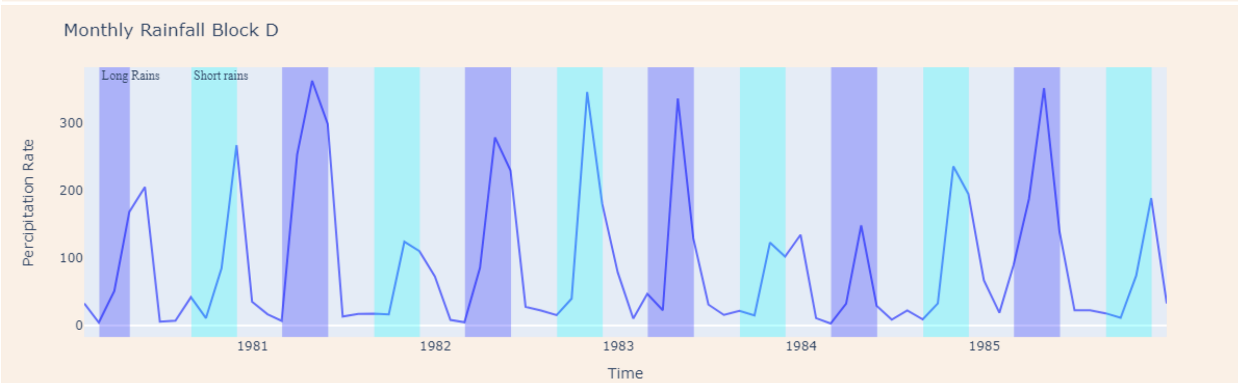


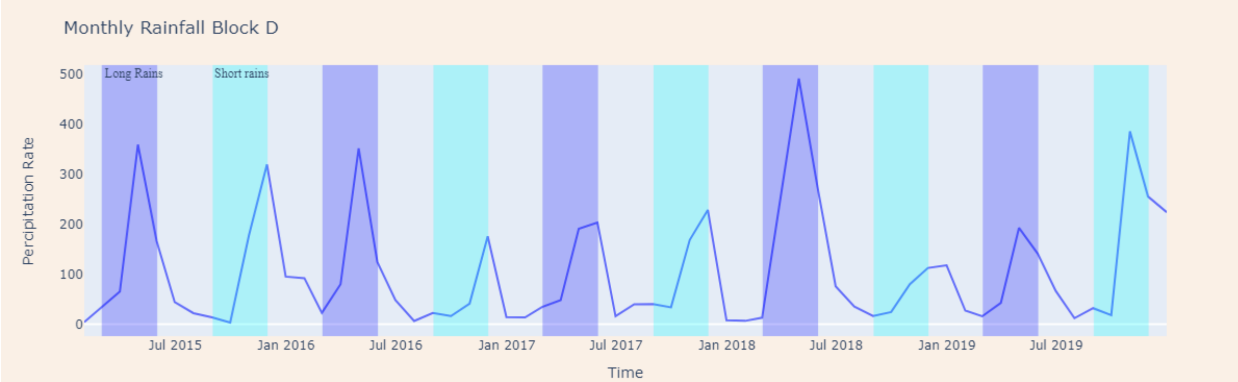




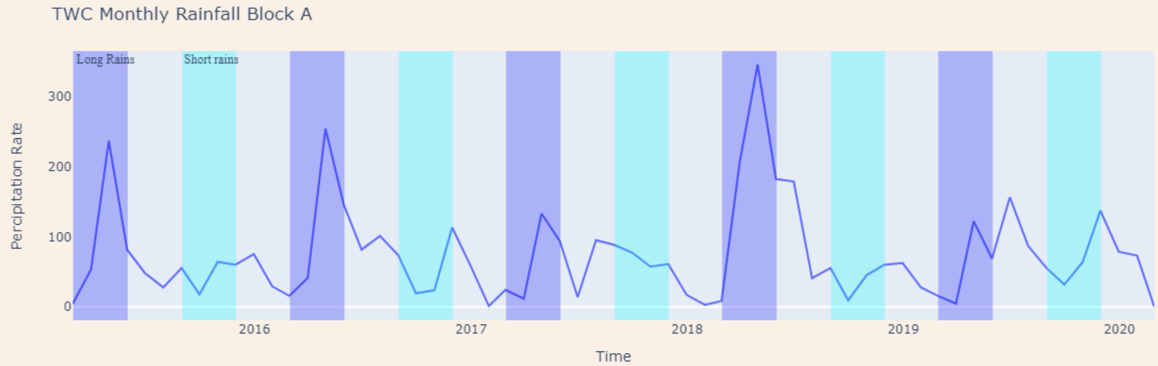

Monthly Rainfall Block C
Long Rains
Short rains
Percipitation Rate
300
250
200
150
100
50
0
1981
1982
1983
1984
1985
Time

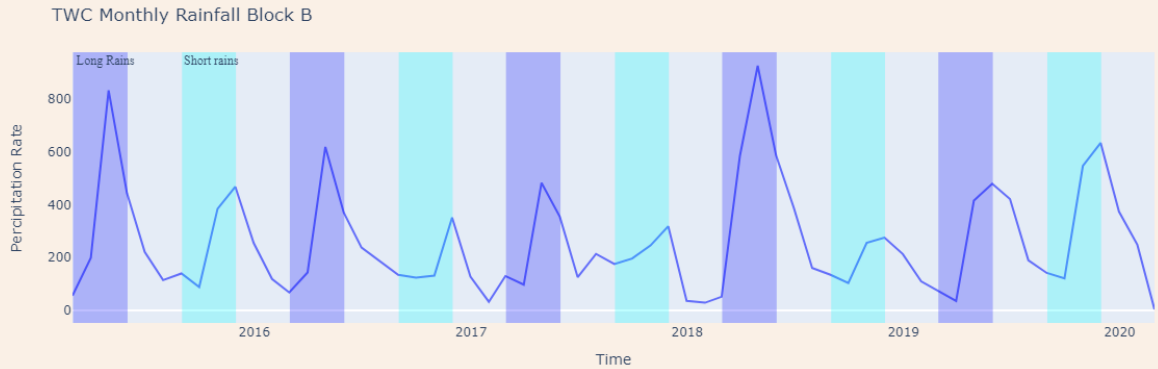

Monthly Rainfall Block C
Long Rains
Short rains
Percipitation Rate
300
200
100
0
Jul 2015
Jan 2016
Jul 2016
Jan 2017
Jul 2017
Jan 2018
Jul 2018
Jan 2019
Jul 2019
Time

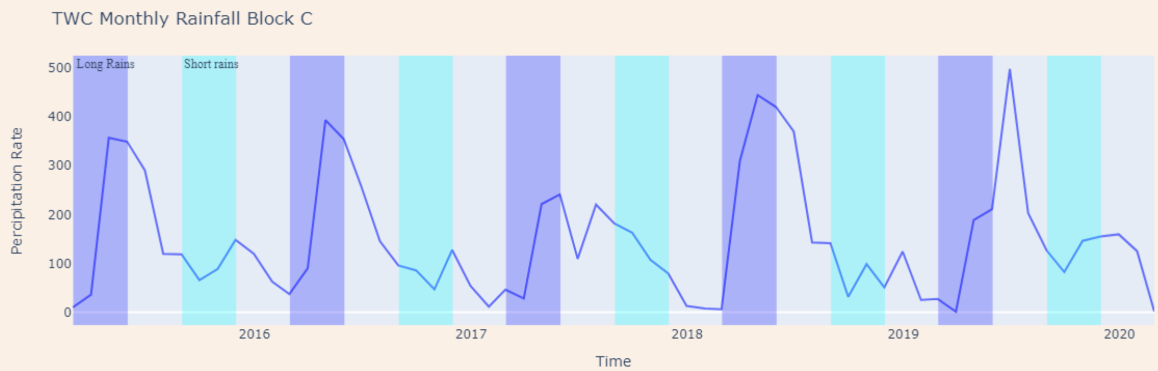

Monthly Rainfall Block D
Long Rains
Short rains
Percipitation Rate
300
200
100
0
1981
1982
1983
1984
1985
Time

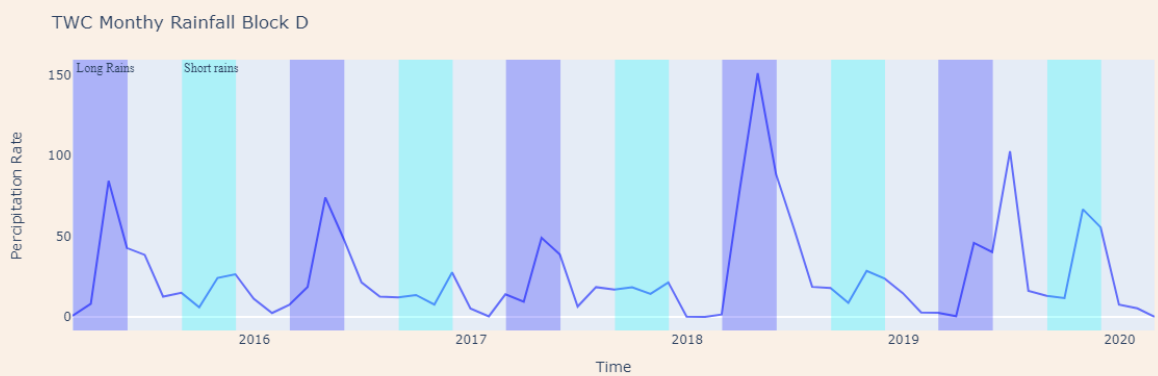

Monthly Rainfall Block D
Long Rains
Short rains
Percipitation Rate
500
400
300
200
100
0
Jul 2015
Jan 2016
Jul 2016
Jan 2017
Jul 2017
Jan 2018
Jul 2018
Jan 2019
Jul 2019
Time

The Weather Channel 2015-2020 Monthly Precipitation Graphs

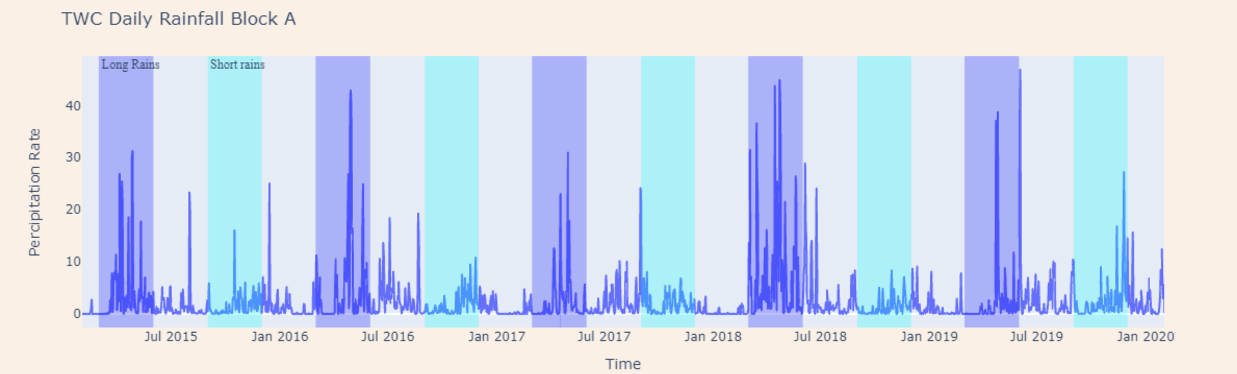


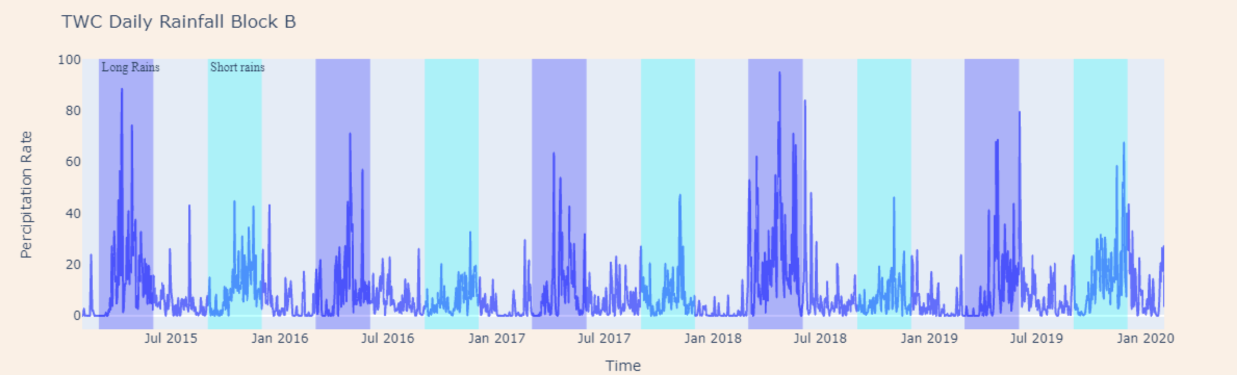


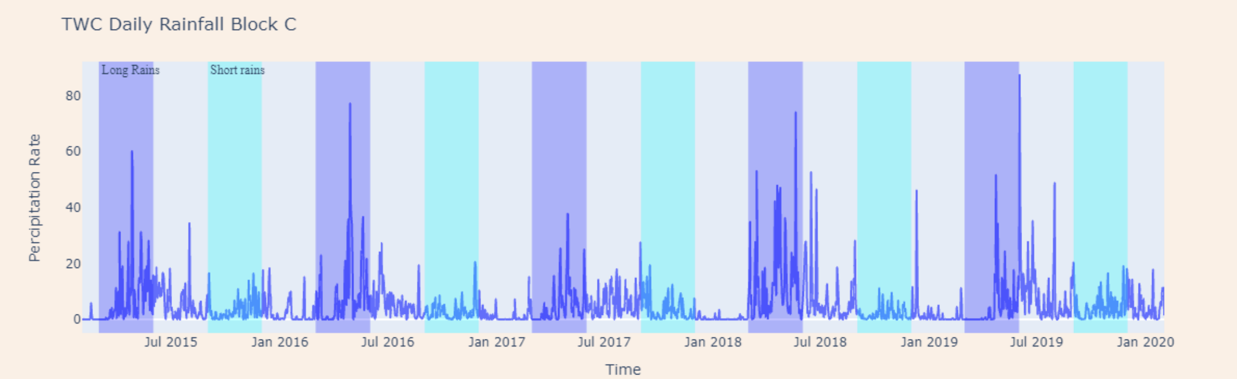


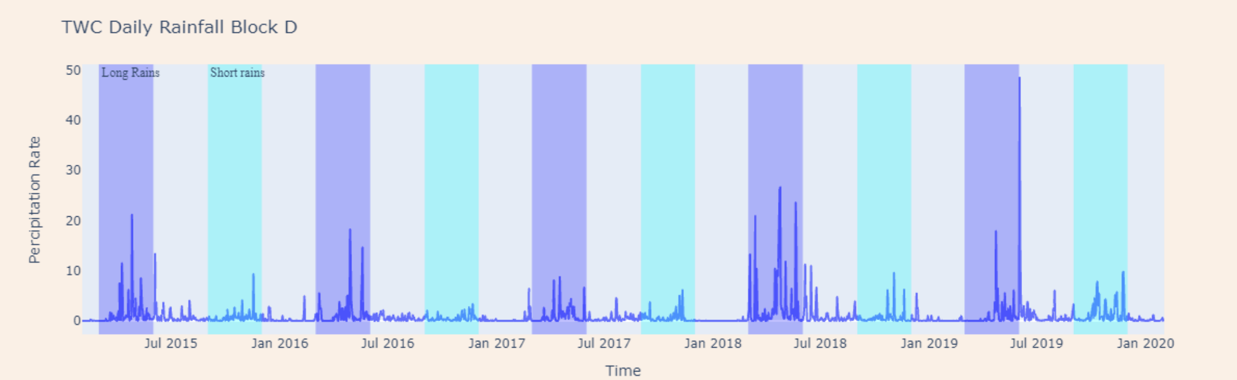


The Weather Channel 2015-2020









Appendix E: Statistical Tests Analyzing Water Level in Githambara and Karurumo

```
##########################################################################################
> # ANOVA Test
>
> Githam_Karuru_ANOVA <- a .... [TRUNCATED]

> summary(Githam_Karuru_ANOVA)
                Df Sum Sq Mean Sq F value Pr(>F)
Season           3  840.2  280.06   13521 <2e-16 ***
Residuals   110473 2288.2    0.02
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##########################################################################################


##########################################################################################
> # Individual t-tests
##########################################################################################
>
> # t-test Rainy Season between .... [TRUNCATED]

> Rainy_Ttest <- t.test(Discharge ~ Season, data = Rainy_season, mu = 0, conf = .95, var.eq = FALSE, paired =
FALSE)

> Rainy_Ttest

        Welch Two Sample t-test

data:  Discharge by Season
t = -124.61, df = 52036, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1550640 -0.1502615
sample estimates:
mean in group Rainy_Season_Githam mean in group Rainy_Season_Karuru
                0.4350048                         0.5876675

##########################################################################################

##########################################################################################
> # t-test dry Season between Githam and Karuru
>
> #create appropriate data table
> Dry_season <- filter(Githam_Karuru_stacked, str_length(Githam_K .... [TRUNCATED]

> Dry_Ttest <- t.test(Discharge ~ Season, data = Dry_season, mu = 0, conf = .95, var.eq = FALSE, paired =
FALSE)

> Dry_Ttest
```

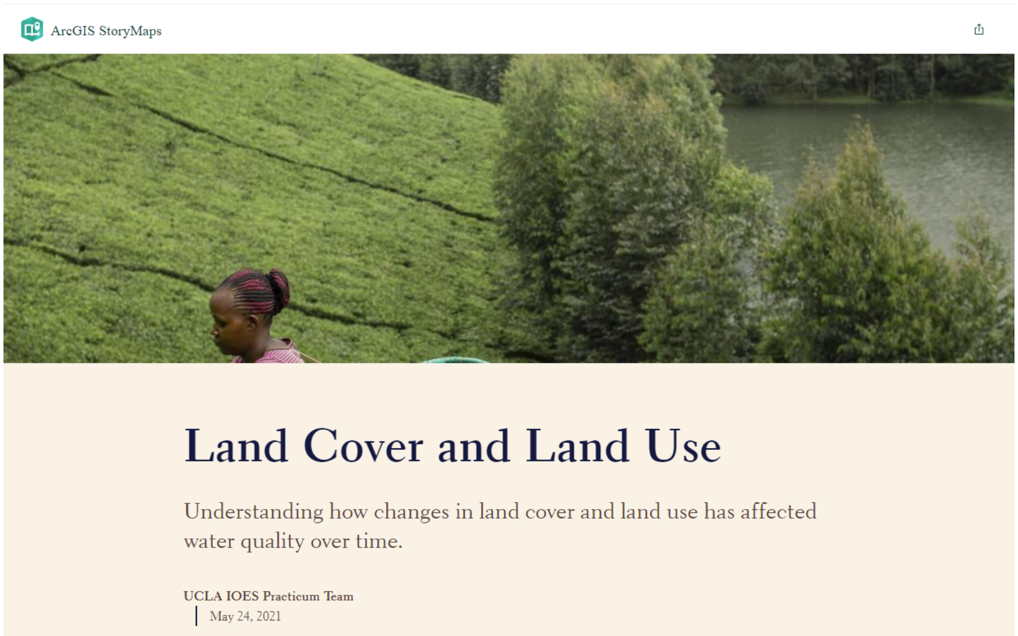
```
        Welch Two Sample t-test

data:  Discharge by Season
t = -146.09, df = 44837, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1804953 -0.1757163
sample estimates:
mean in group Dry_Season_Githam mean in group Dry_Season_Karuru
                0.3606085                       0.5387143
#################################################################################################

#################################################################################################

> # t-test Githam between Rainy and Dry seasons
>
> Githam_Ttest <- t.test(Discharge ~ Season, data = Final_Data_Githam, mu = 0, conf = .95, var.eq  ....
[TRUNCATED]

> Githam_Ttest

        Welch Two Sample t-test

data:  Discharge by Season
t = -64.449, df = 47360, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.07665876 -0.07213373
sample estimates:
 mean in group Dry_Season_Githam mean in group Rainy_Season_Githam
                0.3606085                         0.4350048

#################################################################################################

#################################################################################################

> # t-test Githam between Rainy and Dry seasons
>
> Karuru_Ttest <- t.test(Discharge ~ Season, data = Final_Data_Karuru, mu = 0, conf = .95, var.eq  ....
[TRUNCATED]

> Karuru_Ttest

        Welch Two Sample t-test

data:  Discharge by Season
t = -38.055, df = 49720, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.05147446 -0.04643186
sample estimates:
 mean in group Dry_Season_Karuru mean in group Rainy_Season_Karuru

                0.5387143                         0.5876675
```

```
###########################################################################################

##########################################################################################
#  Multi-variable linear Regression Results
##########################################################################################
Residuals:
    Min       1Q   Median       3Q      Max
-0.39899 -0.09357  0.00946  0.07951  1.73753

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.5535424  0.0010411  531.71   <2e-16 ***
Rain_clump   0.0104793  0.0003076   34.07   <2e-16 ***
Marker_num  -0.1350478  0.0014662  -92.11   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1538 on 44401 degrees of freedom
Multiple R-squared:  0.1834,      Adjusted R-squared:  0.1834
F-statistic:  4988 on 2 and 44401 DF,  p-value: < 2.2e-16
```

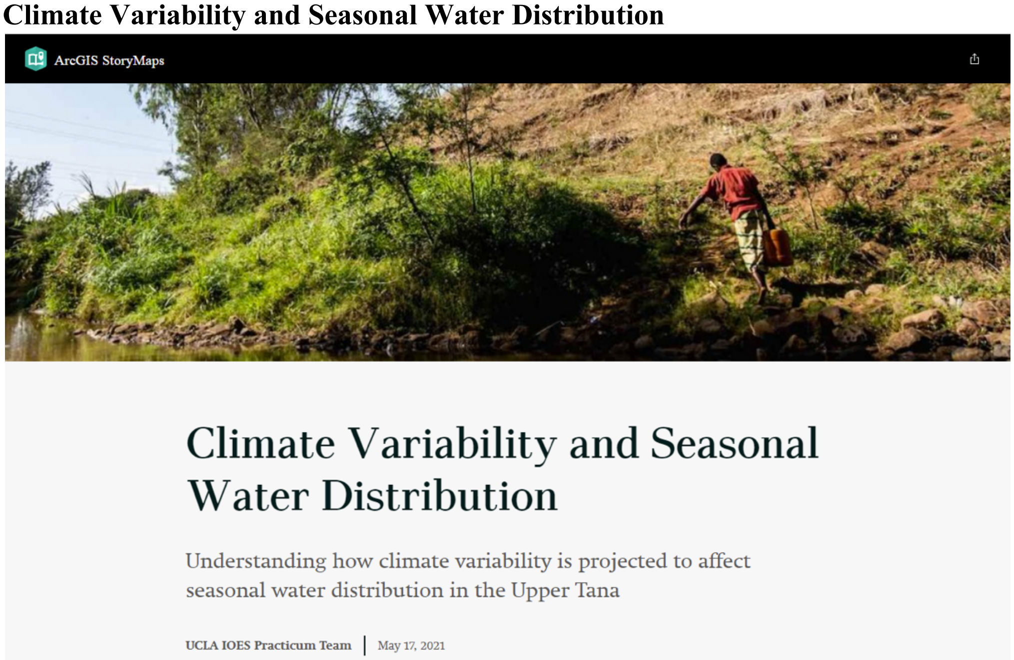
Appendix F: Link to Story Maps

Link to Collection of Story Maps: https://arcg.is/1yDrPy

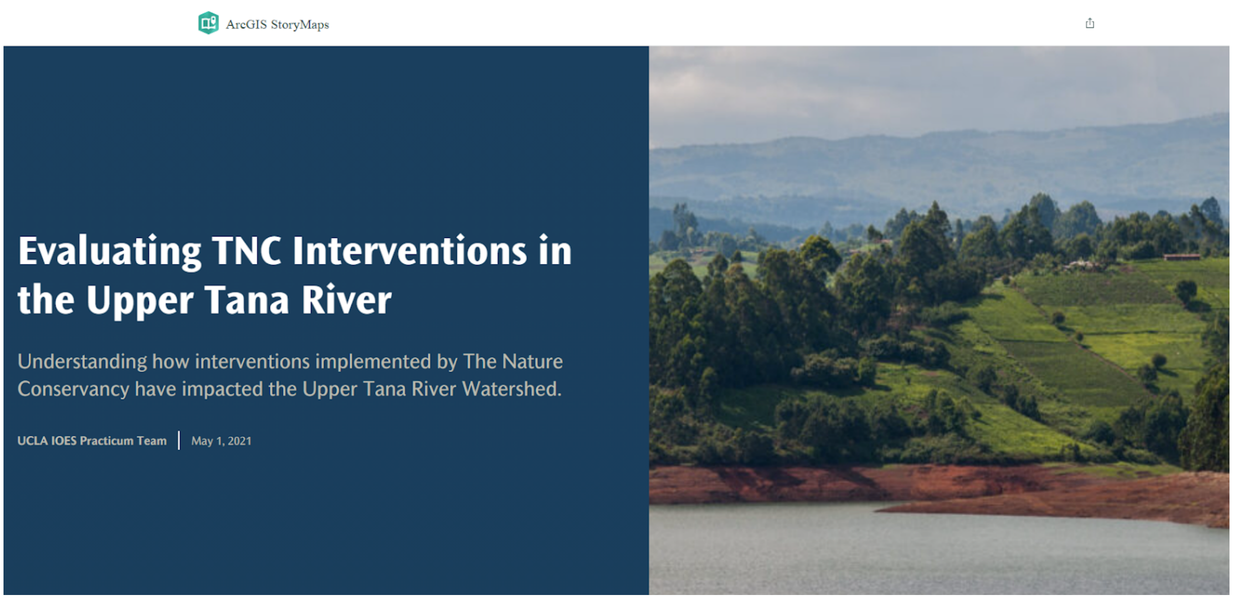
Land Cover and Land Use



Climate Variability and Seasonal Water Distribution



Evaluating TNC Interventions in the Upper Tana River Basin



Appendix G: Link to Shiny Web Application

https://ucla.shinyapps.io/ibmtnc_watershed/