Potomac River Basin Water Resources

Assessing Water Quality and Quantity in the National Capital Region Using NASA Earth Observations

 **Technical Report**

August 03, 2023

Zach Kinloch

Divina Lade

Xavier Garcia Lopez

Haydee Portillo

***Advisors:***

Sean McCartney, Science Systems and Applications, Inc., NASA Goddard Space Flight Center (Science Advisor)

***Fellow:***  
Carli Merrick (Maryland – Goddard)

# 1. Abstract

The Potomac River Basin (PRB) is responsible for providing drinking water to over 5 million residents and plays a significant role in the health of the Chesapeake Bay. Therefore, it is important to understand the relationship between water quality, landcover and the hydrological cycle within the PRB. The National Park Service (NPS) has monitored 37 streams within the National Park Units in Maryland, Virginia, West Virginia and Washington, D.C. This project aimed to help NPS better understand trends in water quality and supplement their ability to monitor changes in the National Capital Region Network (NCRN). Google Earth Engine, ArcGIS Pro, R, and Python were used for data retrieval, visualization and analysis. Earth observations included Landsat 5 TM and Landsat 8 OLI/TIRS imagery. Ancillary data included the USDA Cropland Data Layer, Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS), and soil moisture data from the Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS). Land use/land cover (LULC), Normalized Difference Vegetation Index (NDVI), precipitation and soil moisture data were all analyzed against water quality data provided by NPS at a watershed level. Land use change maps were also generated for the PRB between 2008 and 2022. The project found significant correlations between precipitation, soil moisture, NDVI, and water quality. Correlations were found between certain land use types and water quality metrics, but findings varied greatly between watersheds. These insights emphasize the imperative of strategic watershed management in preserving the integrity of key aquatic systems.

**Key Terms**

Landsat, FLDAS, CHIRPS, GEE, Water Quality, Land Use Land Cover, NDVI, Hydrology

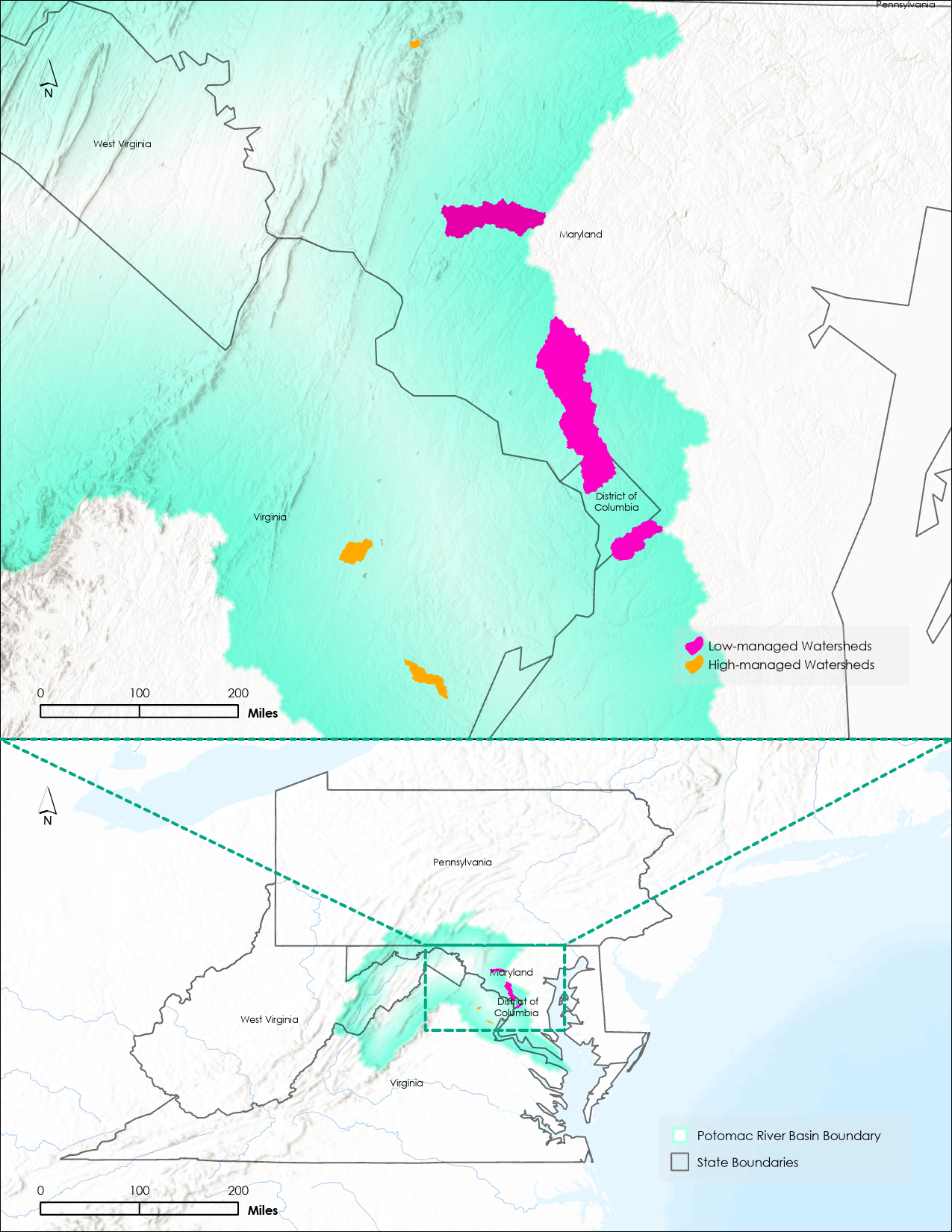
# 2. Introduction

***2.1 Background Information***

According to 2020 Census data, the Potomac River Basin (PRB) is home to more than 6 million people. Stretching across Maryland, Virginia, West Virginia, and Pennsylvania, the basin is the second largest sub-watershed in the Chesapeake Bay watershed (Interstate Commision on the Potomac River Basin [ICPRB], n.d.-b). Due to the major role the PRB plays in providing drinking water to the Washington metropolitan area, it is important to understand the relationship between water quality, quantity, and landcover within the PRB. Past studies on water and biological resource quality have revealed that the PRB, including its tributaries and the Chesapeake Bay, is in a state of continuous decline due to ongoing nutrient inputs (University of Maryland Center for Environmental Science [UMCES], 2011; Chesapeake Bay Foundation, 2012; Bricker et al., 1999, 2007). The degradation of water quality within the basin directly impacts the Chesapeake Bay, causing coastal eutrophication due to nutrient inflow from activities such as the application of fertilizers, discharges from wastewater treatment plants, and significant increase in salt application during winter (Bricker et al., 2013); Bock et al., 2018). This grim outlook has prompted the initiation of legislation aimed at minimizing nutrient inputs and restoring water quality to acceptable standards (Boesch et al., 2001).

This project analyzed changes in land use and land cover (LULC) in relation to observed alterations in water quality. Prior studies have highlighted that land cover and land use alterations can degrade water quality, mainly due to surface runoff, which increases with precipitation, thereby transporting nutrients from the surface into the rivers (Zhao et al., 2018; Nobre et al., 2020). Different land use types can significantly impact water quality, and the extent of anthropogenic land use adjacent to a water body is a primary contributor to water quality deterioration (Nobre et al., 2020). In contrast to anthropogenic land use, forested land cover has been seen to be associated with improved water quality (Broga et al., 2017). Given these impacts, this project’s LULC analysis focused primarily on forested and developed land cover. Agricultural and grassland land cover have also been seen to correlate with changes in water quality (Ferrier et al., 2001; Nielsen et al., 2012). The team analyzed grassland and agricultural land types in comparison to water quality for watersheds where these land use types made up a significant percent of total land cover.

Moreover, the amount of vegetation present in a river’s basin, indicated by Normalized Difference Vegetation Index (NDVI) values, can serve as a water quality marker. Significant relationships have been found between NDVI and water quality parameters in streams (Griffith et al., 2002). Therefore, this project analyzed changes in NDVI values over time and their correlation to water quality parameters collected by the NPS. The team examined hydrological parameters like precipitation and soil moisture content (m3/m3) to understand variations in the hydrological cycle or water quantity corresponding to NDVI parameters.



*Figure 1.* This project analyzed multi-parameter data for 6 watersheds within the Potomac River Basin spanning the period from 2007 to 2022.

Though it is crucial to understand the ramifications of anthropogenic activities within the entire PRB, this project analysis focused on six watersheds only. These watersheds were chosen based on what percent of their land falls under NPS park boundaries. *Figure 1* displays these six watersheds. Those highlighted in Purple (Rock Creek, Oxon Run, Bush Creek) lie 30% or less under NPS managed lands. Those highlighted in Orange (Young’s Branch, Blue Blazes Creek, North Fork Quantico Creek) lie 70% or more under NPS managed lands.

Data sets relating to water quality, water quantity, land cover and land use are crucial for understanding the ramifications of anthropogenic activities within the PRB. These variables underscore the complexity of water resource management and the far-reaching impacts of terrestrial activities on aquatic ecosystems. By perpetuating such research, we can provide the National Park Service (NPS) with methodologies for devising holistic strategies to mitigate adverse effects and safeguard these vital ecosystems.

***2.2 Project Partners & Objectives***

The Potomac River Basin Water Resources DEVELOP team partnered with the National Park Service (NPS) National Capital Region Network (NCRN) and the Stroud Water Research Center (SWRC). The NCRN is an inventory and monitoring program that currently monitors 37 streams within the National Park Units in Maryland, Virginia, West Virginia, and Washington D.C. As a whole, NPS works closely with local governments and organizations to preserve the natural and cultural resources of National Parks throughout the United States. SWRC is a non-profit organization whose mission is to advance knowledge and stewardship of freshwater systems through global research, education, and watershed restoration.

NPS has a particular interest in this project to understand where to better focus management actions. The main point of concern was the difference in water quality for high-managed areas compared to low-managed areas, if any. Currently, satellite data is not heavily utilized by NPS to observe water quality conditions and changes over time. The goal of this project was to use Earth observations to better enhance NPS’ ability to observe trends in water quality. The main objectives of this project were to create land use change maps, analyze the relationships between land cover change/hydrological conditions and water quality, and finally, create social media posts to communicate this research to the public.

# 3. Methodology

***3.1 Data Acquisition***

A shapefile of the Potomac River Basin was accessed from the United States Geological Survey’s (USGS) Watershed Boundary Dataset: hydrologic unit code (HUC) 8. This shapefile was buffered beyond its borders by 10 miles using ArcGIS Pro 2.9.6. This buffer ensured that the entirety of the basin was captured when downloading data. Additional shapefiles of several watersheds were provided by SWRC. These watersheds contain monitoring locations for the current 37 NCRN monitored streams. These shapefiles were used during the analysis process. The project partners use the USA Contiguous Albers Equal Area Conic USGS coordinate reference system (CRS), therefore we used this CRS as this project’s standard as well.

Water quality data for the 37 NCRN monitored streams was provided through google drive by SWRC. This dataset was provided in a CSV file format. This dataset includes water quality parameters such as specific conductance, pH, nitrogen and phosphorus nutrients, water discharge, water temperature, and dissolved oxygen. This dataset is organized by monitoring site. R scripts provided by Dr. Daniel Myers from SWRC allowed the team to extract water quality data for specific monitoring sites from this dataset.

We acquired surface soil moisture data (0–10 cm depth) from the Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) dataset via Google Earth Engine (GEE). GEE's utility in analyzing and downloading satellite data is well-established across various scientific fields (Carrasco et al., 2019; Cissell et al., 2021; Ermida et al., 2020; Prikaziuk and van der Tol, 2021). The FLDAS dataset, originally designed by McNally et al. (2017) to aid food security assessments in developing countries, provided continuous data from January 1, 2007, to December 31, 2022, for our study. A distinguishing feature of the FLDAS dataset is its incorporation of Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) precipitation data in contrast to the broader meteorological forcings employed by NLDAS. Its spatial resolution is also slightly finer at 11 km compared to NLDAS's 14 km resolution (NASA Land Data Assimilation Systems, 2023). These features, alongside its data availability in GEE, made FLDAS an optimal choice for this study.

The team also obtained the precipitation dataset using GEE. The CHIRPS Pentad Version 2.0 dataset provides monthly and yearly precipitation data. CHIRPS is a 30+ year quasi-global rainfall dataset. It incorporates 0.05° resolution satellite imagery with in-situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring. CHIRPS was created in collaboration with scientists at the USGS Earth Resources Observation and Science (EROS) Center to deliver complete, reliable, up-to-date data sets for several early warning objectives, like trend analysis and seasonal drought monitoring.

Land use/land cover (LULC) data was obtained from the USDA’s National Agricultural Statistics Service, via Google Earth Engine. The LULC product used for this analysis is the Cropland Data Layer (CDL), which provides national datasets of land classification on an annual level from 2008 through 2022, with 30-m spatial resolution. This land classification product, due to its availability on an annual level, makes it ideal to analyze changes in LULC over a relatively short period of time, such as this project’s study period. While the CDL focuses primarily on agricultural classification, classification exists for all land use types making this product useful for analyzing LULC change over time.

The USGS Landsat 8 Level 2, Collection 2, Tier 1 and USGS Landsat 5 Level 2, Collection 2, Tier 1 Surface Reflectance datasets were both accessed through the Google Earth Engine (GEE) Data Catalog. These datasets contain atmospherically corrected surface reflectance derived from data produced by the Landsat 8 Operational Land Imager (OLI)/ Thermal Infrared Sensor (TIRS) and Landsat 5 Thematic Mapper (TM) sensors, respectively. Both products consist of remotely sensed optical imagery at 30-m spatial resolution. NDVI values were derived using these 2 datasets for Dec 2006 – Feb 2023 with a 19-month gap between Dec 2011 – June 2013. This gap falls between the end of the Landsat 5 dataset and the beginning of Landsat 8 data. Landsat 7 data was not considered for this analysis because of its scan line corrector failure. This failure results in 22% less usable data (Roy et al., 2016), which would result in many data gaps withing the Potomac River Basin shapefile.

Table 1. *Overview of Earth observation data sources used in this project*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Satellite(s), Sensors, and Datasets** | **Processing Levels** | **Purpose** | **Spatial Resolution** | **Image Dates** | **Source** |
| Landsat 8, Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) | Level 2 Collection 2, Tier 1 | NDVI | 30-m | June 2013 – Feb 2023 | USGS via Google Earth Engine |
| Landsat 5, Thematic Mapper (TM) | Level 2 Collection 2, Tier 1 | NDVI | 30-m | Dec 2006 – Feb 2012 | USGS via Google Earth Engine |
| CHIRPS Pentad: Climate Hazards Group InfraRed Precipitation with Station Data (Version 2.0 Final) | Level 3 | Precipitation | 5566-m | January 2007 – December 2022 | University of Santa Barbara Climate Hazards Group via Google Earth Engine |
| FLDAS: Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System | Level 3 | Soil Moisture | 11132-m | January 2007 – December 2022 | NASA GES DISC at NASA Goddard Space Flight Center |
| USDA NASS: Cropland Data Layer (CDL) | Level 4 | Land-cover Classification | 30-m | 2008–2022 | USDA via Google Earth Engine |

***3.2 Data Processing***

*3.2.1 Water Quality Data Averaging*

Utilizing the R scripts provided by Dr. Daniel Myers, the team averaged each water quality metric for each monitoring location seasonally and annually. The seasonally averaged data was used in the soil moisture, precipitation, and NDVI analyses while the annually averaged data was used in the LULC analysis. Although this project focused primarily on six watersheds, these processed datasets and the methodology will allow SWRC and NPS to analyze remote sensing and water quality relationships within other watersheds.

To enhance and refine our data processing, we subsequently developed a Python script inspired by Dr. Myers' R scripts. We leveraged the data manipulation capabilities of Python's “pandas” library, enabling us to import the dataset from an Excel file. A logical modification we implemented was renaming the “ActivityStartDate” column to “Date” for better clarity. To ensure data integrity, we standardized every entry in the “CharacteristicName” column to string format.

A subsequent focus was on optimizing the dataset. This entailed zoning in on specific parameters like Dissolved Oxygen (DO), water temperature, and others. Leveraging functions like “isin”, we systematically filtered out non-relevant data. Each result value was converted into a numeric format, and the “Date” column was transformed into a datetime type. This enabled us to spotlight the entries from December 2006 to November 2022. We focused on entries from December 2006 to November 2022. To ensure complete data, we filled missing “MonitoringLocationName” values using their related “MonitoringLocationIdentifier” column. Our main goal was to rearrange the dataset to highlight daily averages. By carefully grouping and aggregating the data, we determined the daily average for each characteristic at every monitoring location. To avoid confusion, we replaced zero values with NaN. The “pivot\_table” function was useful for presenting daily metrics for each characteristic in a comprehensible column. We renamed the columns for better understanding and fixed any encoding problems. As a final step, the dataset was sequenced by the “Date” column, laying out a chronological narrative. Our rigorous process culminated in the consolidation of our refined data into a CSV file.

After the dataset optimization phase, our next step was to visualize the processed data. This was the first impression of our dataset, where patterns, correlations, and anomalies began to reveal themselves. We utilized several strategies and specific functions to accomplish our goals. Our process of data visualization commenced with the examination of water quality variables like Dissolved Oxygen (DO), Temperature, Specific Conductance, Salinity, and Nutrients provided by the National Park Service (NPS). We leveraged the Python data manipulation library, pandas, to handle our data. Pandas provide flexible and powerful data structures for manipulating and analyzing any kind of data. Its DataFrame structure is particularly suited for data manipulation tasks involving labeled data and heterogeneous data types.

The first step in our process was to extract location-specific data from our larger dataset. We had a predefined list of locations, "locations\_of\_interest", for which we required data. By employing a 'for' loop, we iterated over this list and used pandas' Boolean filtering capability to separate the data relevant to each location from the larger dataset. Once we had location-specific data, we split the DO and Temperature data based on the date "2018-Summer". This was a crucial step as it allowed us to interpolate these variables for the dates prior to "2018-Summer". Here, we utilized pandas' "interpolate()" function, a method which effectively fills in any missing values, preparing the dataset for further visualization.

To bring our water quality data to life visually for each location, we employed matplotlib, a Python 2D plotting library. Matplotlib produces quality figures in a variety of formats and interactive environments across platforms. Using matplotlib's plot function, we created plots for each variable. Each variable was displayed on its own y-axis, ensuring clarity and ease of reading. To further enhance these plots, we used matplotlib’s functions: "set\_title()" for plot titles, "set\_xlabel()" and "set\_ylabel()" for axis labels, and "legend()" to create a legend explaining the different lines in the plot. The combination of these elements resulted in a thorough and intuitive visualization, displaying the trends and patterns in our data.

Finally, each of these plots was saved as an SVG file in a specified directory using matplotlib's "savefig" function. SVG (Scalable Vector Graphics) is an XML-based vector graphics format that can produce high-quality, scalable, and editable images. Concurrently, each plot was displayed in the Google Colab environment using "plt.show()" for immediate visual feedback. Google Colab is a cloud-based Python development environment that offers an easy way to learn and use machine learning and supports many popular machines learning libraries, including pandas and matplotlib. See Figures C16 - C21, in the Appendix for further details.

*3.2.2 Soil Moisture*

To unlock the full potential of the FLDAS dataset, we undertook a meticulous resampling technique using Google Earth Engine (GEE). Key to our analysis was GEE's “ImageCollection” function, a versatile tool allowing users to work with the FLDAS/NOAH01 image collection previously mentioned. Filtering capabilities within GEE, particularly the “filterDate” function, refined our dataset to a precise temporal range. Our specific Area of Interest (AOI) was defined using GEE’s “ee.FeatureCollection”. This tool facilitates the integration of shapefile data, offering capabilities for visualization, analysis, and modification of raster data. A standout capability of GEE is the custom function application across datasets. The “map” function was notably invaluable. It enabled the assignment of seasons based on monthly data. By utilizing a custom function to convert months into corresponding seasons, we were able to generate an image collection that represented mean soil moisture for every season across the years of interest.

Reducers in GEE permitted the derivation of specific statistics from data. We applied “ee.Reducer.mean”, “ee.Reducer.stdDev”, and “ee.Reducer.count” for this. GEE’s “reduceRegions” function demonstrated its capacity in zonal statistics, enabling computation of values for each predefined region. For Soil Moisture (SM), the inherent spatial resolution of the FLDAS dataset was 11 km. To acquire finer information suitable for our AOI’s, we leveraged GEE’s resampling capabilities. The data was resampled to a 0.5 km grid, notably increasing the granularity. This resampling process is intrinsic to GEE's ‘reduceRegions’ function which, when applied at a scale parameter of 500 meters, inherently applies resampling to match this scale.

The results were then streamlined to encapsulate vital data such as the site name, season-year, mean value, standard deviation, and pixel count for soil moisture. This process was made efficient with the modification functionality of multiplying the count by 3, representing the three months of each season. With GEE's “Export.table.toDrive”, we conveniently export our findings directly to Google Drive as a CSV document, eliminating unnecessary columns like "system:index" and ". geo" for a cleaner output.

The CSV file generated for SM was combined with water quality and precipitation data using a Python script for subsequent analysis. The code begins by setting the paths of the input files for soil moisture, water quality data, and precipitation. An output path is also determined, which will later be used to store the merged data. Once these paths are established, the "pandas" library is employed to read these files, turning them into tables within the program. A list of specific desired locations was created. This list serves to filter the data to only include certain pertinent cells. Consequently, the water quality data undergoes a process where only the data linked to these selected locations is retained. The code then introduces a function named “reformat\_date”. This function is designed to modify the format of dates, changing entries like "Summer2023" to "2023-Summer". Once this function is defined, it's utilized on both the soil moisture and precipitation data, ensuring that dates are consistent across the datasets. The main objective of this script is to amalgamate the three datasets into a singular output document. To achieve this, it merges the tables, making sure that the data aligns based on two common columns: the location identifier and the date. The soil moisture dataset is the foundational layer, with the water quality and precipitation data joined over it.

In the world of data management, encountering missing values is common. Addressing this, when any location names are missing in the merged data, the code uses a mapping technique to populate those gaps, associating each location identifier with its appropriate name. To provide clearer insights and facilitate future analyses, the script divides the merged date column into two separate columns: one indicating the year and another the season. After the merging, reformatting, and cleaning processes, the consolidated data is saved. A concluding note emphasizes that, although the code has processed most of the data, there might still be a need for some manual tweaks, especially concerning specific columns and their values.

In the subsequent phase of our data visualization process, we shifted our attention to the "Discharge" data provided by NPS, as well as "Soil Moisture" and "Precipitation" data retrieved from remote sensing data. We processed the data for these water quantity variables in a similar manner to the approach described earlier. Just as before, we extracted the location-specific data for “Discharge” and used the pandas "interpolate()" function to fill in any missing values for dates prior to "2018-Summer". This approach ensured we had a complete dataset, ready for visualization. The visual representation of these three variables took a slightly different form than the one we used for the water quality variables. We crafted a single plot featuring multiple y-axes to accommodate these variables. "Soil Moisture" and "Discharge" were represented as line plots, giving us a clear visual on their trends over time.

"Precipitation", on the other hand, was displayed as a bar chart on a separate y-axis. We inverted this y-axis to provide a more intuitive representation of rainfall data. This method of representing the data provided a stark contrast between the variables, enabling a clear distinction among them. We ensured the plot was user friendly by personalizing it with titles, labels, and a combined legend that tied together all the different elements of the plot. Once the visualization was complete, we saved it as an SVG file for later use. The SVG format ensures high-quality, scalable, and editable images that are ideal for any kind of presentation or publication. For immediate visual feedback, we also displayed the plot directly in the Google Colab environment. Through this rigorous process of data manipulation and visualization, we were able to gain a preliminary understanding of our water quality and quantity data. This visualization set the stage for more complex analytical operations that followed. The seamless blend of data manipulation using pandas and visualization using matplotlib provided a comprehensive insight into our dataset. Refer to Figures C10 - C15, in the Appendix for a visual representation of the water quantity variables.

*3.2.3 Precipitation*

The CHIRPS Pentad dataset has a spatial resolution of 5,566 m. The precipitation data acquired was narrowed down to the six main watersheds using the AOI shape file using the “.filterBounds” function of GEE. Similar to how the soil moisture data was processed, we rescaled the precipitation raster data’s spatial resolution to 0.5 km in order to ensure a uniformed cell size. We then downloaded the data in CSV format. This file contained mean seasonal precipitation values from 2007 to 2022 along with their standard deviation for each watershed. The team also exported the data in GeoTIFF file format for each season throughout the study period. We then reprojected these files to the CRS commonly used by the project partners. As previously mentioned, "Precipitation" was represented in bars for data visualization purposes along with the other water quantity parameters such as "Soil Moisture" and "Discharge".

*3.2.4 Land Cover/Land Use Class Aggregation*

To further process the USDA CDL data, we aggregated similar LULC classes to make the data more interpretable, as well as to help better identify correlations between LULC change and water quality change. This aggregation was done using GEE via the “.remap” function. Only LULC classes that existed within the Potomac River Basin were included in this aggregation. This resulted in eight total classes from the original 200+ LULC classes. The final eight classes analyzed were: Forest, Developed, Cropland, Grassland/Pasture, Shrubland, Wetlands, Water, and Barren. Corresponding CDL classes to the new aggregated classes are available in Table A2 in Appendix A.

*3.2.5 Seasonal NDVI*

Utilizing the Landsat data, the team calculated and added NDVI bands to each image in the collection using the red and near infrared bands in the following equation:

(1)

Then we defined the seasons as follows: spring starts March 1st, summer starts June 1st, fall starts September 1st, and winter starts December 1st. Seasonal greenest pixel composites were created by using the “.qualityMosaic” function on the NDVI bands. This function, along with a custom function to split images into groups based on their dates, chooses the highest NDVI for each pixel among the group of images. Then, we calculated the average NDVI for each watershed using these seasonal greenest pixel composites. Finally, the team exported the seasonal averages for each watershed to CSV files.

For this project, four CSVs were generated from GEE; one CSV for each season included all watershed average NDVI per year. For example, the spring CSV included 2017-Spring, 2018-Spring, etc. for all 37 watersheds. The most efficient way to analyze this data against the water quality data was to combine all the seasonal NDVI data into one. The team did this manually by combining the four files into a single CSV file, organized by monitoring location, date (in Year-Season format) and NDVI value.

***3.3 Data Analysis***

*3.3.1 Precipitation and Soil Moisture Analysis Against Water Quality*

For deeper analysis following our initial time series data visualization, we considered exploring relationships between variables through scatter plots. These plots can help identify positive or negative linear trends and any potential outliers or clusters of data points. Thus, we used a Python script in Google Colab for this phase of analysis. We began by importing the necessary Python libraries. Here, "pandas" were used for data handling and analysis, "seaborn" and "matplotlib" for data visualization, and "scipy.stats" for the necessary statistical functions. To allow for direct access to datasets stored on Google Drive and to avoid repeated manual uploads, Google Drive was mounted to Google Colab using the "drive.mount()" function.

After defining the path to our data file, the CSV data file was read into a pandas DataFrame, which we named "df", using the "pd.read\_csv()" function. The use of "ISO-8859-1" encoding ensured that special characters in the data were interpreted correctly, thus maintaining the integrity of our data. Our analysis was focused on the six sub watersheds mentioned before, which were defined in the "locations\_of\_interest" variable. We then filtered our main dataset to include only data related to these locations. For each sub-watershed, we created scatter plots accompanied by trend lines to probe the relationships between several environmental variables. These variables were categorized as dependent (comprising water quality parameters and discharge) and independent (comprising Soil Moisture and Precipitation).

We created scatter plots for each pair of dependent and independent variables using the "sns.scatterplot()" function, and added a linear regression model fit using "sns.regplot()". Importantly, the "sns.regplot()" function also automatically added a 95% confidence interval for the regression line, represented as a translucent band around the line. This band indicated the range within which we could expect the true regression line to fall with 95% confidence, providing us with an understanding of the reliability of our regression model.

To quantify the strength and significance of the linear relationship between each pair of variables, we calculated the Pearson correlation and its corresponding p-value using the "pearsonr()" function from the "scipy.stats" module. After appropriately arranging and titling the graphs, we saved the generated plots as PNG files in a defined output folder using the "plt.savefig()" function, and also displayed them directly within the Colab environment using "plt.show()". Lastly, we compiled an Excel file to store the calculated Pearson correlation and p-value for each location and variable pair. Initially stored in a list, this data was then converted into a pandas DataFrame and saved to an Excel file using the "to\_excel()" function. This Excel file served as a valuable reference for understanding the statistical relationships between various environmental variables across our selected locations. See Figures C2 - C9 in the Appendix for dispersion graph and trends visualization.

*3.3.2 Land Cover/Land Use Analysis Against Water Quality*

For LULC comparison with water quality, percent land cover within each respective subbasin was analyzed with changes in water quality data. We determined the percentage land cover based on the number of pixels of each LULC class within the watersheds. To match the temporal scale of the LULC, the team averaged the water quality metrics on an annual basis. For the statistical analysis, percent LULC were used as the independent variables, and the different water quality metrics made up the dependent variables. The team used the Spearman rank sum correlation test to determine the relation between land cover types and changes in water quality. We selected this statistical test as it can provide both the strength and direction of correlations between LULC types and water quality metrics. Additionally, this test is ideal for data that is not normally distributed, which is the case for the LULC data., as determined by the Shapiro-Wilk test for normality.

Five water quality metrics were analyzed in comparison to LULC change. These metrics included salinity, specific conductance, dissolved oxygen, total nitrogen, and total phosphorus. The team analyzed LULC in comparison to salinity, specific conductance, and dissolved oxygen from 2008 to 2018, as data for these metrics was not available after 2018. We analyzed LULC in comparison to total nitrogen and phosphorus from 2017 to 2022, as nutrient data was only available for these years. For all statistical analysis regarding LULC, we excluded data from 2010, as this year had significant anomalies in LULC classification, leading to inaccurate results.

To conduct the analysis, the team utilized R 4.3.1, which outputs the correlation coefficient of the relationships, as well as a p-value to determine statistical significance. The primary LULC classes focused on were Developed and Forest, as these land use types make up the bulk of all watersheds analyzed and could contribute to significant changes in water quality. For certain watersheds, namely Young’s Branch and Bush Creek, Cropland and Grassland/Pasture classes made up a significant portion of total land cover. For these watersheds, the team also incorporated Cropland and Grassland/Pasture classes into the statistical analysis.

*3.3.3 NDVI Analysis Against Water Quality*

To investigate the relationship between water quality and NDVI, the team conducted correlation tests for each of the six watersheds. To check for normality, the team used Shapiro-Wilk tests on the variables of the NDVI and water quality datasets. The datasets for both showed mostly non-normal distribution. Because of this non-normal distribution of data, the team chose to use Spearman correlation tests when analyzing the relationship between NDVI and water quality metrics. We utilized Python through Jupyter Lab to extract data and conduct these tests. Initially, correlation tests were conducted per season, but the datasets were too sparse to observe any real trend. Because of this, we used the entire dataset for each parameter, regardless of season.

For the correlation tests, the seasonal NDVI values were set as the independent variable and the water quality metrics were set as the dependent variable. The correlation tests were conducted over the whole study period. Study periods varied between water quality metrics. For nutrient data, only data from 2017 to 2022 was used as the water quality dataset provided to us had unusable data before 2017. For the other water quality metrics, data from 2007 to 2018 was used as there was no data for the years after 2018. The previously mentioned 19-month gap in the NDVI data left room for error in this part of the analysis.

# 4. Results & Discussion

***4.1 Results***

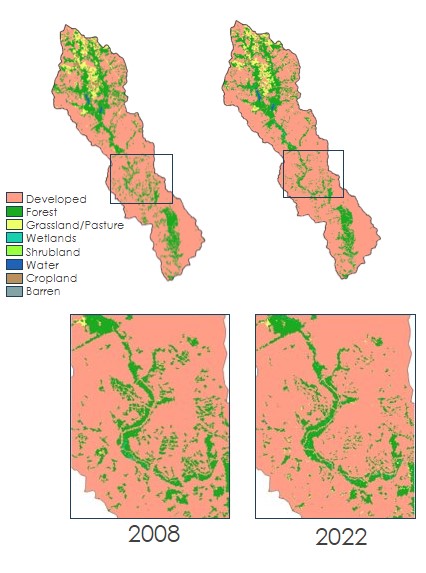
*4.1.1 Soil Moisture and Precipitation*

In the Upper Potomac River Basin (UPRB) study, numerous statistically significant correlations were found between environmental variables and water quality parameters across different locations. Soil Moisture (SM) and precipitation data were established as independent variables. The independent variables were correlated and analyzed statistically with the dependent variables including dissolved oxygen (mg/l), water temperature (°C), and discharge (cfs; Appendix C: Table C1).

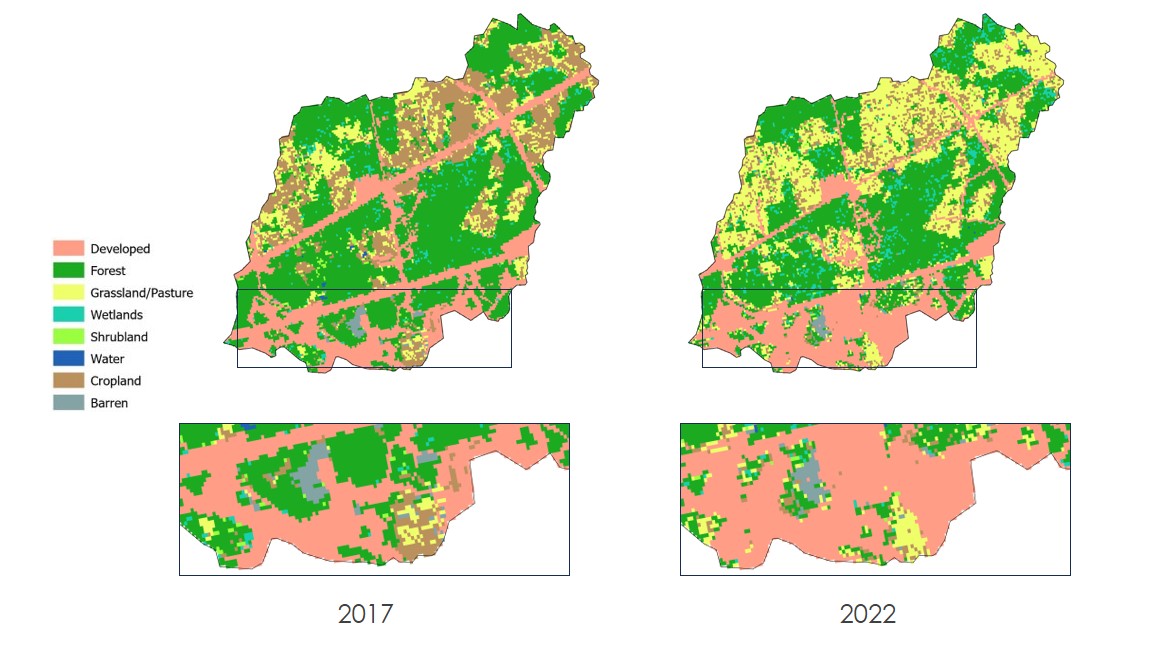
One of the high-managed watersheds, Young's Branch (Appendix C: Figure C1), shows a strong positive correlation between SM and dissolved oxygen (r = 0.63, p < 0. 05), and a moderate negative correlation with water temperature (r = -0.46, p < 0.05). Additionally, a robust positive correlation existed between SM and discharge (r = 0.72, p < 0.05). Another high-managed watershed, North Fork Quantico Creek (Appendix C: Figure C2, reveals a strong positive correlation among variables including SM, dissolved oxygen (r = 0.59, p < 0.05), and discharge (r = 0.66, p < 0.05). However, there is a moderate negative correlation between SM and water temperature (r = -0.43, p < 0.05). Lastly, for the high-managed watershed, Blue Blazes Creek (Appendix C: Figure C3) showcased a strong positive correlation between SM, and both dissolved oxygen (r = 0.68, p < 0.05) and discharge (r = 0.50, p < 0.05), while displaying a moderate negative correlation with water temperature (r = -0.50, p < 0.05).

Among the three low-managed watersheds, Bush Creek (Appendix C: Figure C4) stood out with the strongest correlations, with an extremely strong positive relationship between SM and dissolved oxygen (r = 0.80, p < 0.05), and a comparably strong negative correlation with water temperature (r = -0.77, p < 0.05). Oxon Run (Appendix C: Figure C5) had weak positive correlation overall between variables. Precipitation has a moderate positive correlation with water temperature (r = 0.41, p < 0.05), and is statistically significant. For the watershed with the largest area, Rock Creek Downstream of Dumbarton Oaks (Appendix C: Figure C6), SM exhibited a moderate positive correlation with dissolved oxygen (r = 0.47, p < 0.05) and a strong one with discharge (r = 0.65, p < 0.05), yet a moderate negative correlation with water temperature (r = -0.35, p < 0.05). On the other hand, precipitation shows a moderate positive correlation with water temperature (r = 0.41, p<0.05), a weak negative correlation with dissolved oxygen (r = -0.33, p < 0.05), and a weak positive correlation (r = 0.39, p < 0.05) with discharge water quality variable.

*4.1.2 LULC*



*Figure 2.* Map of Rock Creek watershed, focusing on the area with the highest development over the study period



*Figure 3.* Map of Young’s Branch watershed, focusing on the area with the highest deforestation over the study period

In Rock Creek, the team found a strong statistically significant correlation between percent developed land cover (Figure A1, Appendix A), which saw increases over the study period (Figure 2.), and specific conductance (ρ = 0.758, p < 0.05). Additionally, a near-statistically significant relationship was found between developed land cover and salinity in Rock Creek (ρ = 0.758, p = 0.066). In Young’s Branch, we identified a very strong negative correlation between forest land cover and total phosphorus (Figure A2, Appendix A), as there was high deforestation (Figure 3.) and increases in total phosphorus (ρ = -0.943, p < 0.05). A similar, although not statistically significant, correlation was identified in North Fork Quantico Creek (ρ = -0.829, p = 0.058). In Oxon Run, two moderate, not statistically significant, negative correlations were found between forest land cover and both salinity and specific conductance (ρ = -0.612, p = 0.066). Another near statistically significant correlation was found in Young’s Branch, where a negative correlation was identified between developed land cover and dissolved oxygen (ρ = -0.590, p = 0.073). A final notable relationship was identified in the North Fork Quantico Watershed, where the team found a negative correlation between developed land cover and salinity (ρ = -0.675, p < 0.05). All correlation coefficients and corresponding p-values are listed in Table A1 in Appendix A.

*4.1.3 NDVI*

After conducting the Spearman correlation tests between the NDVI and water quality data, the team found the correlation coefficients and p-values shown in Table B1 in Appendix B. Initial research suggested that Total Nitrogen (TN) and Total Phosphorous (TP) would have strong and significant correlation to NDVI. According to Table B1, a majority of the watersheds showed no significant correlation between TN/TP and NDVI except for Rock Creek. Rock Creek showed a negative correlation between TN and NDVI (ρ = -0.789, p < 0.05).

All watersheds showed signs of significant correlation between Temperature/DO and NDVI (p < 0.05). The high-managed watersheds showed signs of high to moderate positive correlation between temperature and NDVI (0.71 < ρ < 0.85) and moderate negative correlation between DO and NDVI (-0.70 < ρ < -0.55). For the low-managed watersheds, a moderate positive correlation was found between temperature and NDVI (0.49 < ρ < 0.71) while a moderate to low negative correlation was found between DO and NDVI (-0.65 < ρ < 0.0).

Specific Conductance was statistically significant in 3 watersheds; this water quality metric had an extremely significant correlation with NDVI for Blue Blazes Creek and Rock Creek (p < 0.001) while it had a moderately significant correlation with NDVI in Rock Creek (p<0.05). All three watersheds showed signs of moderately negative correlation (-0.60 < ρ < -0.41).

***4.2 Analysis of Results***

*4.2.1 Soil Moisture and Precipitation*

The discovered correlations suggest that SM is a pivotal determinant for various water quality parameters within the UPRB. High-managed locations like Young's Branch, North Fork Quantico Creek, and less managed Bush Creek, SM consistently corresponds with heightened levels of DO and discharge but lower water temperatures. The relationship between water quality, land cover, and the hydrological cycle can provide valuable insights into areas with cold, oxygen-rich groundwater sources feeding into water bodies. This phenomenon has been extensively studied in other water bodies, including Lake Kawaguchi and springs in Cape Cod. Recent research conducted by Yasuhara et al. (2020), Harvey and Gooseff (2015), Briggs et al. (2018) and Briggs et al. (2020) have demonstrated the presence of cold, oxygen-rich groundwater in these locations. Understanding such relationships in various water bodies, including those in the study area, may shed light on the potential influence of groundwater sources on the overall water quality and ecosystem dynamics.

The intriguing findings of this study revealed that across all locations, significant relationships were observed between SM and at least two water quality parameters (DO and temperature). However, the nature and strength of these relationships exhibited considerable variations. The impact of development activities within the watersheds may play a role in shaping these relationships, as urbanization can lead to increased surface runoff and alterations in the soil moisture balance. In specific sub-basins, such as Youngs Branch, North Fork, Oxon Run, and Rock Creek, there is a significant negative relationship between water temperature and precipitation. This suggests that urban runoff and warmer substrates may be thermally impacting these basins. A similar phenomenon was observed in the Back Creek watershed in Roanoke County, Virginia, indicating potential thermal impacts from urbanization (Krause et al. 2004).

*4.2.2 LULC Analysis Results*

For the Rock Creek watershed, developed land was found to have a strong correlation with specific conductance in the water. This finding was statistically significant and logical, as specific conductance is an indicator of dissolved solids in water, which increases with surface runoff over developed land (Zampella et al., 2007). As developed land was found to have increased by several hundred acres within the Rock Creek watershed, there is undoubtedly more developed surface for rainwater to runoff from into the creek. This could carry pollutants into the creek and increase specific conductance. Given this strong correlation, it is a likely possibility that the development within the Rock Creek watershed is a contributing factor to the increased dissolved solids that NPS has observed in their data collection. In addition to this relationship, a moderate correlation was found between water salinity and developed land. While this finding was not quite statistically significant, the relationship is still important to consider. Salinity, much like specific conductance, is an indicator of pollutants in the water. Given the strong correlation between development and specific conductance, it is important to understand that there may be a correlation between development and salinity as well, even if it cannot be concluded with 100% certainty. Considering both water quality metrics can indicate pollution, it is likely that development within the watershed has been contributing to the pollution of Rock Creek.

In the Young’s Branch watershed, a very strong negative correlation was found between percent forest land cover and TP within the stream. This watershed saw notable deforestation during the study period, while notable increases were found in the TP of Young’s Branch. This finding is logical, as heavy forest tends to absorb nutrients, thus preventing them from running off into the stream (Cheng et al., 2022). A similar, although not quite statistically significant, correlation was found in the North Fork Quantico watershed. While it cannot be confidently stated that this relationship exists, it is an interesting finding to see the same correlation is possibly present in multiple watersheds. Additionally, both of these watersheds are heavily managed by NPS. Significant, or near significant, relationships between forest land cover and phosphorus were only found in these high-managed watersheds, while they were not present in any of the low-managed watersheds. Another interesting element to these findings was the fact that in these watersheds, the team found no notable correlations with forest land cover and TN, another nutrient. Because of the strong correlation between forest and phosphorus, one would expect that a similar correlation would be found between forest and nitrogen. This was not the case, indicating that there are likely other factors influencing the nitrogen levels that were not included in the scope of this project.

In addition to the aforementioned LULC relationships, the team found several other relationships that may be present within the studied watersheds. In the Young’s Branch watershed, there was a moderate negative correlation between developed land and dissolved oxygen in the creek. In the Oxon Run watershed, the team found moderate negative correlations between forest land cover and both salinity and specific conductance. These findings align with prior research (Broga et al., 2017, Dauer et al., 2000), however they are not statistically significant. So, while the team cannot confidently say that these correlations exist, we deemed them necessary to share nonetheless, as understanding potential trends is an important part of watershed management.

The team came across one particularly interesting finding in the North Fork Quantico watershed. Here, a statistically significant negative correlation was found between developed land and salinity. This correlation is in the opposite direction of what one would expect and of the relationships found in other watersheds such as Rock Creek. This could possibly be due to anomalies within the LULC classification data. More investigation is necessary to truly understand the relationship in this watershed.

*4.2.3 NDVI Analysis Results*

Griffiths et al. (2002) prove NDVI exhibits a strong linear relationship with both nitrogen and phosphorus. The Potomac River Basin team was not able to observe this strong linear relationship. Only Rock Creek showed signs of a strong negative correlation between NDVI and TN. According to the LULC classifications, none of the watersheds had a high percentage of agricultural lands, suggesting limited inputs of nutrients from the land. Dense vegetation indicated by high NDVI values may absorb and reduce the already low water nitrogen levels.

Looking at Figures B1 and B2, no apparent trend can be observed between the nutrient data and NDVI. The spike around fall 2021 in TP can be seen in the time series for all other watersheds, leading the team to believe this may not be a typo in the water quality data. Overall, the lack of data for TN and TP hinders the ability to observe strong relationships between NDVI and nutrient data.

Strong positive correlations were observed between water temperature and NDVI while DO/NDVI correlations experienced an inverse relationship. Though photosynthetic activities can be a source of DO (Rounds et al., 2006), dense vegetation, most prominent during the summertime, can lead to warmer water temperatures, potentially intensifying microbial activity and decreasing DO levels. There was no contrast in results between high-managed areas and low-managed areas. This may be due to averaging NDVI over an entire watershed, rather than an area buffered around a water quality measuring site.

***4.3 Feasibility Assessment***

It is difficult to say with full certainty that using Earth observations as a way to understand water quality trends is feasible. In terms of statistical analysis to observe trends, soil moisture and precipitation data is a useful tool to understand water quality parameters such as discharge, however, their coarse resolution makes it difficult to visualize these trends. For LULC, the temporal limitations in the water quality data hindered our ability to observe trends or statistical significance with any of the water quality metrics. There is potential for NDVI to be used as a tool to understand trends in water quality, however, temporal limitations in the water quality data as well as the size of the study area didn’t allow us to maximize this potential.

One of the main concerns the partners wanted to address is if using Earth observations would allow them to understand where to better focus management actions. After conducting the analysis described throughout this paper, the team saw mixed or inconclusive results between the low-managed and high-managed lands. Therefore, the project partner end users may utilize the methodology employed in this project to analyze data from other watersheds within the PRB to further observe the differences between low-managed and high-managed lands.

***4.4 Future Work***

Understanding the balance between land use and water dynamics in the UPRB continues to develop. In partnership with the NPS and the Stroud Research Center, we've pinpointed key research areas for further exploration.

For a more comprehensive view of land-use patterns, analyzing larger areas is necessary. Conversely, to understand vegetation's direct influence on water quality, NDVI analysis should be applied to a small buffer around riparian corridors rather than the entire watershed. Integrating data from the Soil Moisture Active Passive satellite can enhance our understanding of soil moisture variability, considering that data has been available since 2015 and offers better temporal and spatial resolution than FLDAS.

There's a pressing need to refine the data set by addressing inconsistencies and gaps in water quality data. Enhancing this foundation will allow for clearer insights into the UPRB's hydrological processes. Additionally, data modeling, rooted in our existing findings, can project future trends, assisting in conservation efforts. Furthermore, employing methods like Principal Component Analysis (PCA) can sharpen our research focus by highlighting key influential variables. This ensures more targeted and efficient strategies. Moving forward, a collaborative approach, drawing on the expertise of both the NPS and the Stroud Research Center, is essential for the UPRB's preservation.

# 5. Conclusions

Our research within the UPRB has provided a comprehensive understanding of the interrelationships between NDVI, land use, hydrological, and water quality variables across multiple sub-watersheds. The correlations observed between soil moisture (SM), dissolved oxygen (DO), temperature, and discharge vary across different sub-watersheds, but some patterns emerged consistently.

For the sub-watersheds majorly managed by the NPS, namely Young's Branch, North Fork Quantico Creek, and Blue Blazes Creek, there's a general positive correlation between soil moisture and dissolved oxygen, and a negative correlation with temperature. This suggests that these areas, with higher NPS management, have dynamics where increased soil moisture tends to support more oxygenated waters that are cooler. Conversely, their precipitation patterns show mixed correlations, which implies that precipitation effects are more nuanced, possibly due to the nature of land use or other factors not directly related to management.

The sub-watersheds with lower NPS involvement, namely Bush Creek, Oxon Run, and Rock Creek Downstream of Dumbarton Oaks, also showed positive correlations between soil moisture and dissolved oxygen, but with some deviations. Oxon Run notably showed a slight negative correlation between soil moisture and dissolved oxygen. The reasons for these deviations warrant further investigations, especially considering the difference in management extent. While we observed robust correlations between certain parameters, our study predominantly utilized Pearson's correlation coefficient, emphasizing linear relationships. It's pivotal for upcoming research to assess data distribution to justify the use of Pearson's over Spearman's coefficient.

Despite comprehensive precipitation and SM data from 2007 to 2022, some variables had inconsistencies. Delving into alternative datasets and encompassing a broader range of variables can bolster understanding and pinpoint the role of management in these correlations.

When it comes to the LULC analysis, while statistically significant correlations were found, findings varied greatly between the different watersheds analyzed. There were many logical findings that were found to be near statistical significance, but not quite significant at the alpha = 0.05 level. The small size of the watersheds potentially played a role in the inconsistency within the findings. The smaller the watershed is, errors within the land classification will have a larger impact. To get a better grasp of land use change, it is ideal to analyze a larger area. Additionally, a longer study period allows for more change to be detected in LULC. Land cover can change quickly, but more drastic changes can be seen over longer periods of time. This project’s study period, for the LULC analysis, was only 2008–2022, limiting the amount of change in LULC that can actually be seen and limiting the data points in the analysis. All in all, there are likely relations between changes in LULC and water quality. However, given the scope of this project, we were unable to discover consistent relationships between the two. More analysis is necessary to fully understand the trends between land use and water quality in the UPRB.

The relationships found between NDVI and water quality metrics suggest that NDVI is a strong tool to use when trying to understand trends in water quality, only when datasets have extensive amounts of data. For all the watersheds studied in this project, NDVI showed signs of significant correlation with Temperature and Dissolved Oxygen and no correlation to nutrient data. Based on previous studies, NDVI has been proven to have a strong linear relationship with nutrient data. Because of this, it is important to further investigate the relationship between the change in NDVI and the change in nitrogen and phosphorous levels in the UPRB, despite the results found in this study.

This study paves the path for additional research aimed at enhancing water resource management in the UPRB. The insights offer the community a clearer comprehension of human activities' effects on local water resources, especially concerning watershed management. Recognizing the importance of effective management, as exemplified by the differences between areas with varying NPS involvement, it becomes evident that informed strategies can significantly influence water resource outcomes in the UPRB.

# 6. Acknowledgements

Maps throughout this work were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. All rights reserved.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

# 

# 7. Glossary

**CDL** – Cropland Data Layer

**CHIRPS** – Climate Hazards Group InfraRed Precipitation with Station data

**Earth observations** – Satellites and sensors that collect information about the Earth’s physical, chemical, and biological systems over space and time

**FLDAS** – Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) dataset

**GEE** – Google Earth Engine

**LULC** – Land Use/Land Cover

**NDVI** – Normalized Difference Vegetation Index

**Pearson Correlation Test** – correlation test that measures the strength of a linear association between two variables. The correlation coefficient is denoted by r. This test requires the datasets to be normally distributed.

**Raster –** a grid of pixels or points representing digital images or spatial data.

**Shapefile** – a file format to store geographic vector data

**Shapiro – Wilk** – evaluates whether a data set is normally distributed

**Spearman Correlation Test** – correlation test that measures the strength of a linear association between two variables. The correlation coefficient is denoted by ρ. This test does not require the datasets to be normally distributed.

# 8. References

Bock, A. R., Falcone, J. A., & Oelsner, G. 2. (2018). Estimates of road salt application across the conterminous United States (1992–2015). US Geological Survey data release. doi: <https://doi>. org/10.5066/P96IX385.

Bricker SB, Clement CG, Pirhalla DE, Orlando SP, Farrow, DRG (1999) National estuarine eutrophication assessment. Effects of nutrient enrichment in the Nation’s Estuaries. NOAA, National Ocean Service, Special Projects Office and National Centers for Coastal Ocean Science, Silver Spring. <http://spo.nos.noaa.gov/projects/cads/nees/Eutro_Report.pdf>

Bricker SB, Longstaff B, Dennison W, Jones A, Boicourt K, Wicks C, Woerner J (2007) Effects of nutrient enrichment in the Nation’s Estuaries: a decade of change, national estuarine eutrophication assessment update. NOAA Coastal Ocean Program Decision Analysis Series No. 26. National Centers for Coastal Ocean Science, Silver Spring, MD. 322 pp. <http://ccma.nos.noaa.gov/news/feature/Eutroupdate.html>

Bricker, S. B., Rice, K. C., & Bricker, O. P. (2014). From headwaters to coast: influence of human activities on water quality of the Potomac River Estuary. Aquatic Geochemistry, 20, 291-323.

Briggs, M. A., Harvey, J. W., Hurley, S. T., Rosenberry, D. O., Mccobb, T., Werkema, D., & Lane, J. W., Jr. (2018). Hydrogeochemical controls on brook trout spawning habitats in a coastal stream. Hydrology and Earth System Sciences, 22, 6383–6398. <https://doi.org/10.5194/hess-22-6383-2018>

Briggs, M. A., Tokranov, A. K., Hull, R. B., LeBlanc, D. R., Haynes, A. B., & Lane, J. W. (2020). Hillslope groundwater discharges provide localized stream ecosystem buffers from regional per-and polyfluoroalkyl substances contamination. Hydrological Processes, 34(10), 2281-2291.

Brogna, D., Michez, A., Jacobs, S., Dufrêne, M., Vincke, C., & Dendoncker, N. (2017). Linking forest cover to water quality: A multivariate analysis of large monitoring datasets. *Water*, *9*(3), 176. doi:10.3390/w9030176

Carrasco, L., O’Neil, A. W., Morton, R. D., & Rowland, C. S. (2019). Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. Remote Sensing, 11(3), 288.

Cheng, C., Zhang, F., Shi, J., & Kung, H.-T. (2022). What is the relationship between land use and surface water quality? A review and prospects from Remote Sensing Perspective. *Environmental Science and Pollution Research*, *29*(38), 56887–56907. doi:10.1007/s11356-022-21348-x

Chesapeake Bay Foundation (2012) State of the Bay report. <http://www.cbf.org/about-the-bay/state-of-thebay/2012-report>

Cissell, J. R., Canty, S. W., Steinberg, M. K., & Simpson, L. T. (2021). Mapping national mangrove cover for belize using google earth engine and sentinel-2 imagery. Applied Sciences, 11(9), 4258.

Dauer, D. M., Weisberg, S. B., & Ranasinghe, J. A. (2000). Relationships between benthic community condition, water quality, sediment quality, nutrient loads, and land use patterns in Chesapeake Bay. *Estuaries*, *23*(1), 80. doi:10.2307/1353227

Ermida, S. L., Soares, P., Mantas, V., Göttsche, F. M., & Trigo, I. F. (2020). Google earth engine open-source code for land surface temperature estimation from the landsat series. Remote Sensing, 12(9), 1471.

Ferrier, R. C., Edwards, A. C., Hirst, D., Littlewood, I. G., Watts, C. D., & Morris, R. (2001). Water quality of Scottish rivers: Spatial and temporal trends. *Science of The Total Environment*, *265*(1–3), 327–342. doi:10.1016/s0048-9697(00)00674-4

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell & Joel Michaelsen (2015), CHIRPS Pentad: Climate Hazards Group InfraRed Precipitation With Station Data (Version 2.0 Final). "The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes". Scientific Data 2, 150066. Accessed: 03 July 2023, [doi:10.1038/sdata.2015.66](https://doi.org/10.1038/sdata.2015.66)

Griffith, J.A., Martinko, E.A., Whistler, J.L. and Price, K.P. (2002), INTERRELATIONSHIPS AMONG LANDSCAPES, NDVI, AND STREAM WATER QUALITY IN THE U.S. CENTRAL PLAINS. Ecological Applications, 12: 1702-1718. Accessed: 13 July 2023, doi.org/10.1890/1051-0761(2002)012[1702:IALNAS]2.0.CO;2

Harvey, J. W., & Gooseff, M. N. (2015). River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. Water Resources Research, 51, 1–30. https://doi.org/10.1002/2015WR017617

Interstate Commission on the Potomac River Basin (ICPRB). (n.d.-b). Potomac Basin Facts. ICPRB. Retrieved July 13, 2023, from <https://www.potomacriver.org/potomac-basin-facts/>

Krause, C. W., Lockard, B., Newcomb, T. J., Kibler, D., Lohani, V., & Orth, D. J. (2004). Predicting influences of urban development on thermal habitat in a warm water stream. JAWRA Journal of the American Water Resources Association, 40(6), 1645-1658.

Landsat-5 TM (Thematic Mapper) image courtesy of the U.S. Geological Survey. Earth Engine Data Catalog. Retrieved June 21, 2023, from <https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LT05_C02_T1_L2>

Landsat 8, Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) image courtesy of the U.S. Geological Survey. Earth Engine Data Catalog. Retrieved June 21, 2023, from <https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2>

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., & Verdin, J. P. (2017). A land data assimilation system for sub-Saharan Africa food and water security applications. Scientific data, 4(1), 1-19.

McNally, A. NASA/GSFC/HSL (2018), 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), Accessed: 03 July 2023, [doi:10.5067/5NHC22T9375G](https://doi.org/10.5067/5NHC22T9375G)

NASA Land Data Assimilation Systems (2023, June 9). Global Land Data Assimilation Systems (FLDAS). Retrieved July 3, 2023, from <https://ldas.gsfc.nasa.gov/fldas/models/global>

Prikaziuk, E., Yang, P., & van der Tol, C. (2021). Google Earth Engine Sentinel-3 OLCI Level-1 dataset deviates from the original data: Causes and consequences. Remote sensing, 13(6), 1098.

Rounds, S. A., Wilde, F. D., & Ritz, G. F. (2006) U.S. Geological Survey, Chapter A6. Section 6.2. Dissolved Oxygen, <https://pubs.usgs.gov/twri/twri9a6/twri9a62/twri9a6_6.2_ver3.pdf>. Accessed 1 Aug. 2023.

Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote sensing of environment, Volume 185(Iss 1), 57–70. <https://doi.org/10.1016/j.rse.2015.12.024>

University of Maryland, Center for Environmental Science (UMCES), Integration & Application Network, EcoCheck (2011) Chesapeake Bay report card <http://ian.umces.edu/ecocheck/report> cards/chesapeakebay/2011/

USDA National Agricultural Statistics Service Cropland Data Layer. 2008-2022. Published crop-specific data layer. Available at <https://nassgeodata.gmu.edu/CropScape/> (accessed 20 June 2023). USDA-NASS, Washington, DC.

US Geological Survey Earth Resources Observation and Science Center. (2018). Provisional Landsat OLI Surface Reflectance [Data set]. US Geological Survey. <https://doi.org/10.5066/F78S4MZJ>

US Geological Survey Earth Resources Observation and Science Center. (2018). Provisional Landsat TM Surface Reflectance [Data set]. US Geological Survey. <https://doi.org/10.5066/F7KD1VZ9>

Yasuhara, M., Hayashi, T., Asai, K., Uchiyama, M., & Nakamura, T. (2020). Overview of the special issue: "Groundwater in Mt. Fuji (Part 2)." Journal of Geography, 129(5), 657-660.

Zampella, R. A., Procopio, N. A., Lathrop, R. G., & Dow, C. L. (2007). Relationship of land-use/land-cover patterns and surface-water quality in the Mullica River Basin. *Journal of the American Water Resources Association*, *43*(3), 594–604. doi:10.1111/j.1752-1688.2007.00045.x

# 9. Appendices

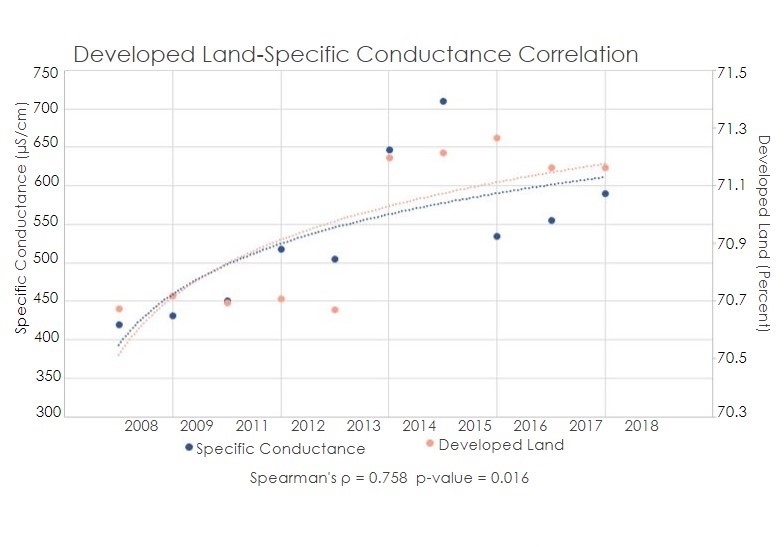
Appendix A – Analysis of Land Cover Land Use against Water Quality

Table A1. S*pearman’s correlation coefficient and p-values between the dependent variables and LULC percentage for each AOI*

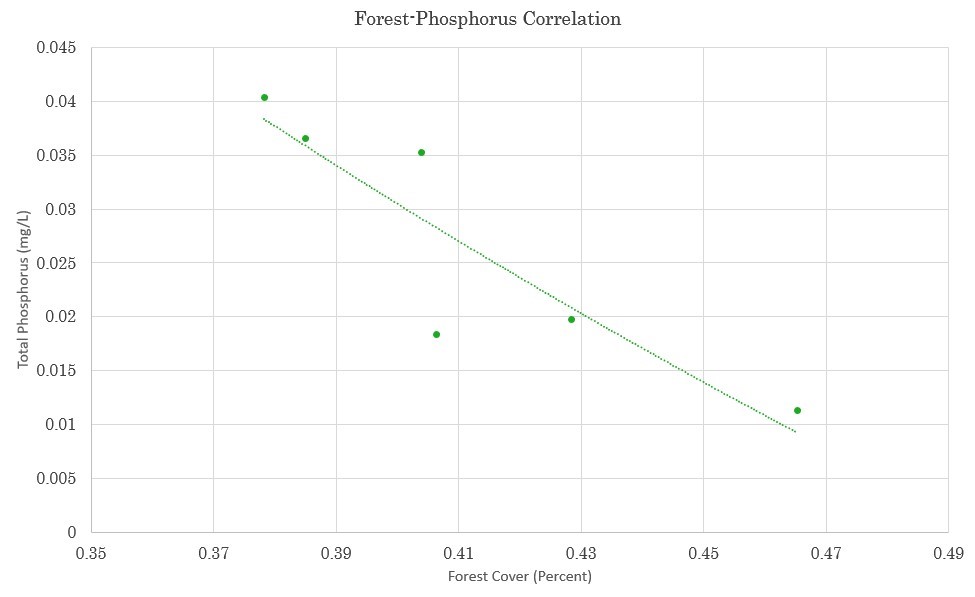
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HIGH-MANAGED WATERSHEDS** | | | | |
| **Site Name** | **Independent Variable** | **Dependent Variable** | **Spearman’s**  **Rho (ρ)** | **P-value** |
| Young’s Branch | Developed Land Cover (%) | Dissolved Oxygen (mg/l) | -0.590 | 0.073 |
| Specific Conductance (µS/cm) | -0.152 | 0.682 |
| Salinity (ppt.) | -0.479 | 0.166 |
| Total Nitrogen (mg/L) | -0.029 | 1.00 |
| Total Phosphorus (mg/L) | 0.600 | 0.242 |
| Young’s Branch | Forest Land Cover (%) | Dissolved Oxygen (mg/l) | 0.438 | 0.206 |
| Specific Conductance (µS/cm) | -0.091 | 0.811 |
| Salinity (ppt.) | 0.236 | 0.514 |
| Total Nitrogen (mg/L) | 0.486 | 0.356 |
| Total Phosphorus (mg/L) | -0.943 | \*0.017 |
| Young’s Branch | Grassland Land Cover (%) | Dissolved Oxygen (mg/l) | -0.486 | 0.154 |
| Specific Conductance (µS/cm) | 0.200 | 0.584 |
| Salinity (ppt.) | 0.018 | 0.973 |
| Total Nitrogen (mg/L) | -0.657 | 0.175 |
| Total Phosphorus (mg/L) | 0.714 | 0.136 |
| Young’s Branch | Cropland Land Cover (%) | Dissolved Oxygen (mg/l) | 0.255 | 0.477 |
| Specific Conductance (µS/cm) | -0.067 | 0.865 |
| Salinity (ppt.) | -0.273 | 0.448 |
| Total Nitrogen (mg/L) | 0.657 | 0.175 |
| Total Phosphorus (mg/L) | -0.714 | 0.136 |
| North Fork Quantico Creek | Developed Land Cover (%) | Dissolved Oxygen (mg/l) | 0.602 | 0.066 |
| Specific Conductance (µS/cm) | -0.612 | 0.066 |
| Salinity (ppt.) | -0.675 | \*0.032 |
| Total Nitrogen (mg/L) | 0.314 | 0.564 |
| Total Phosphorus (mg/L) | 0.086 | 0.919 |
| North Fork Quantico Creek | Forest Land Cover (%) | Dissolved Oxygen (mg/l) | -0.304 | 0.393 |
| Specific Conductance (µS/cm) | -0.442 | 0.204 |
| Salinity (ppt.) | -0.310 | 0.383 |
| Total Nitrogen (mg/L) | -0.371 | 0.497 |
| Total Phosphorus (mg/L) | -0.829 | 0.058 |
| Blue Blazes Creek | Developed Land Cover (%) | Dissolved Oxygen (mg/l) | 0.117 | 0.747 |
| Specific Conductance (µS/cm) | 0.078 | 0.831 |
| Salinity (ppt.) | 0.000 | 1.000 |
| Total Nitrogen (mg/L) | 0.232 | 0.658 |
| Total Phosphorus (mg/L) | 0.029 | 0.957 |
| Blue Blazes Creek | Forest Land Cover (%) | Dissolved Oxygen (mg/l) | -0.117 | 0.747 |
| Specific Conductance (µS/cm) | -0.078 | 0.831 |
| Salinity (ppt.) | 0.000 | 1.000 |
| Total Nitrogen (mg/L) | -0.232 | 0.658 |
| Total Phosphorus (mg/L) | -0.029 | 0.957 |
| **LOW-MANAGED WATERSHEDS** | | | | |
| **Site Name** | **Independent Variable** | **Dependent Variable** | **Spearman’s**  **Rho** | **P-value** |
| Bush Creek | Developed Land Cover (%) | Dissolved Oxygen (mg/l) | -0.079 | 0.838 |
| Specific Conductance (µS/cm) | 0.127 | 0.733 |
| Salinity (ppt.) | 0.224 | 0.537 |
| Total Nitrogen (mg/L) | -0.176 | 0.919 |
| Total Phosphorus (mg/L) | 0.486 | 0.356 |
| Bush Creek | Forest Land Cover (%) | Dissolved Oxygen (mg/l) | 0.115 | 0.759 |
| Specific Conductance (µS/cm) | 0.152 | 0.682 |
| Salinity (ppt.) | 0.212 | 0.560 |
| Total Nitrogen (mg/L) | -0.486 | 0.356 |
| Total Phosphorus (mg/L) | -0.486 | 0.356 |
| Bush Creek | Grassland Land Cover (%) | Dissolved Oxygen (mg/l) | -0.479 | 0.166 |
| Specific Conductance (µS/cm) | -0.539 | 0.113 |
| Salinity (ppt.) | -0.564 | 0.096 |
| Total Nitrogen (mg/L) | 0.029 | 1.000 |
| Total Phosphorus (mg/L) | 0.257 | 0.658 |
| Bush Creek | Cropland Land Cover (%) | Dissolved Oxygen (mg/l) | 0.467 | 0.178 |
| Specific Conductance (µS/cm) | 0.406 | 0.247 |
| Salinity (ppt.) | 0.333 | 0.349 |
| Total Nitrogen (mg/L) | 0.257 | 0.658 |
| Total Phosphorus (mg/L) | -0.314 | 0.564 |
| Oxon Run | Developed Land Cover (%) | Dissolved Oxygen (mg/l) | 0.470 | 0.171 |
| Specific Conductance (µS/cm) | 0.491 | 0.154 |
| Salinity (ppt.) | 0.491 | 0.154 |
| Total Nitrogen (mg/L) | -0.203 | 0.700 |
| Total Phosphorus (mg/L) | -0.63 | 0.173 |
| Oxon Run | Forest Land Cover (%) | Dissolved Oxygen (mg/l) | -0.238 | 0.508 |
| Specific Conductance (µS/cm) | -0.612 | 0.066 |
| Salinity (ppt.) | -0.612 | 0.066 |
| Total Nitrogen (mg/L) | 0.486 | 0.356 |
| Total Phosphorus (mg/L) | -0.143 | 0.803 |
| Rock Creek | Developed Land Cover (%) | Dissolved Oxygen (mg/l) | 0.358 | 0.313 |
| Specific Conductance (µS/cm) | 0.758 | \* 0.016 |
| Salinity (ppt.) | 0.612 | 0.067 |
| Total Nitrogen (mg/L) | -0.714 | 0.136 |
| Total Phosphorus (mg/L) | 0.600 | 0.242 |
| Rock Creek | Forest Land Cover (%) | Dissolved Oxygen (mg/l) | -0.297 | 0.407 |
| Specific Conductance (µS/cm) | 0.055 | 0.892 |
| Salinity (ppt.) | -0.079 | 0.838 |
| Total Nitrogen (mg/L) | 0.657 | 0.175 |
| Total Phosphorus (mg/L) | -0.429 | 0.419 |
| *\*Values indicating statistical significance between variables* | | | | |

Table A2. *Class aggregation of LULC data*

|  |  |
| --- | --- |
| **New Class** | **Original CDL Classes** |
| Developed | 121,122,123,124 |
| Forest | 70,71,141,142,143 |
| Grassland/Pasture | 58,59,176 |
| Wetlands | 190,195 |
| Shrubland | 152 |
| Water | 111 |
| Cropland | 1,2,4,5,6,10,11,12,13,21,24,25,26,27,28,29,30,31,36,37,42,43,44,46,47,53,54,61,66,67,68,69,77,205,216,219,221,222,225,226,229,236,237,240,243,254 |
| Barren | 131 |



***Figure A1.*** Correlation between developed land and specific conductance, Rock Creek

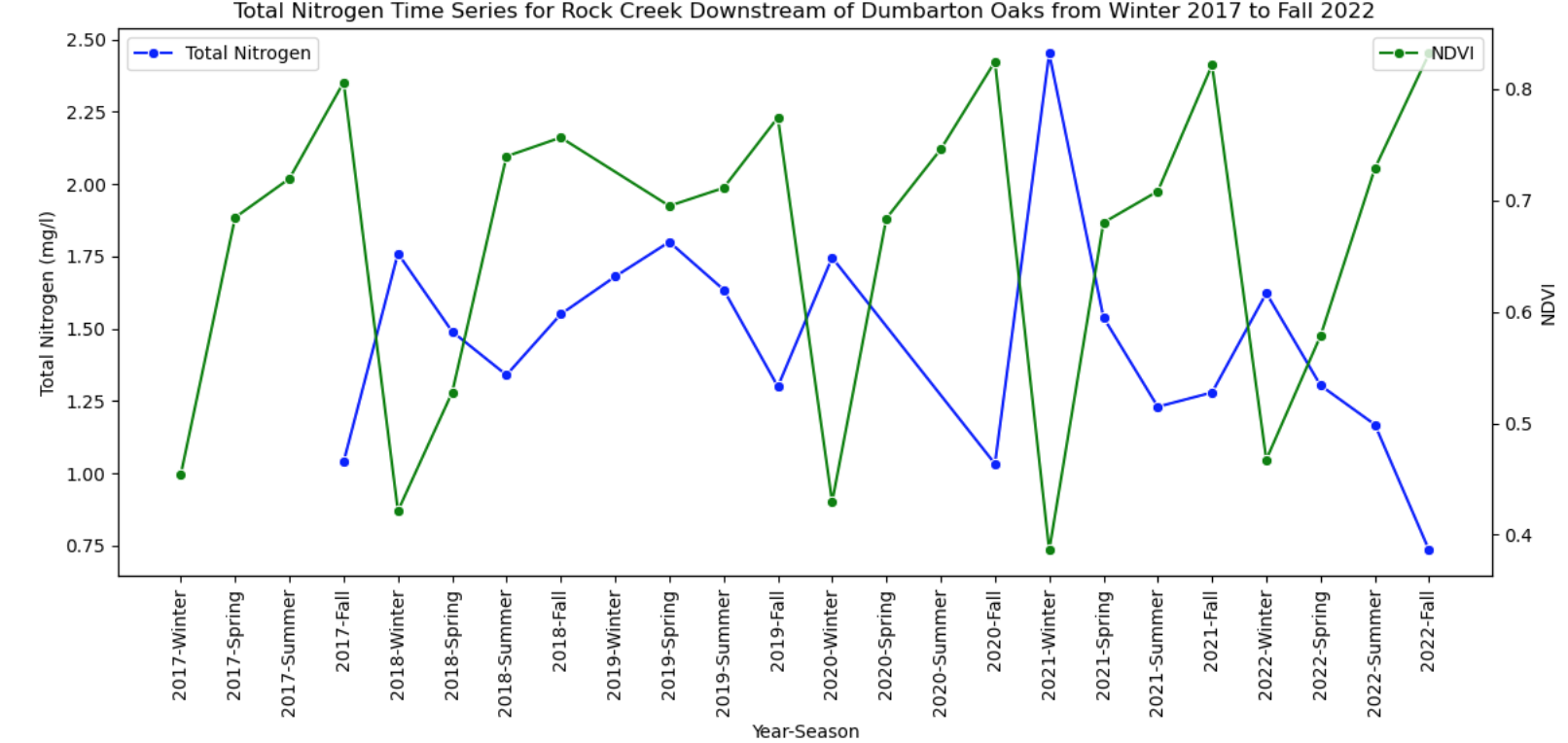


***Figure A2.***Correlation between forest land cover and total phosphorus, Young’s Branch

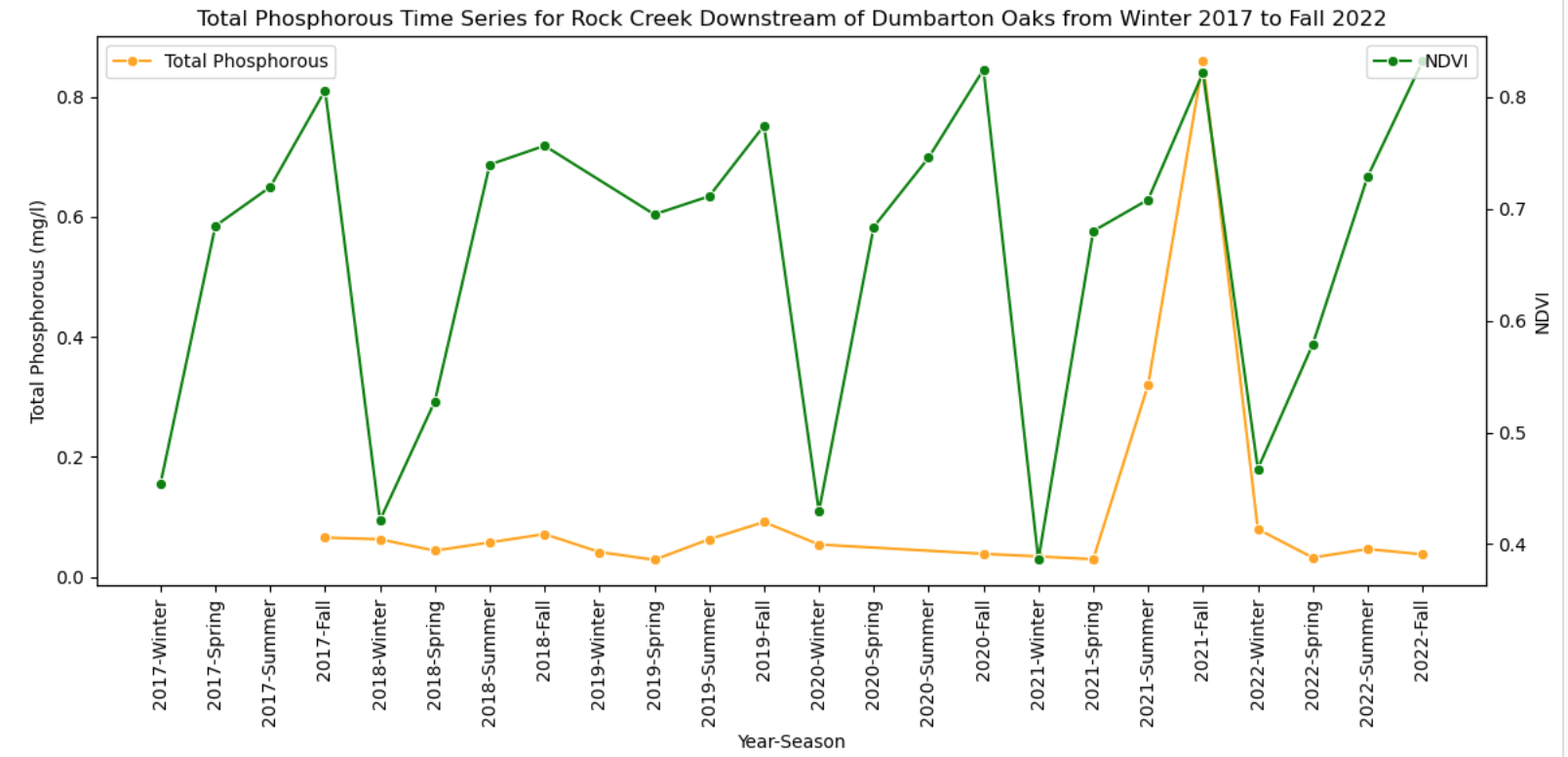
Appendix B – Analysis of NDVI against Water Quality

**Table B1.** S*pearman’s correlation coefficient and p-values between the dependent variables and NDVI for each AOI*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HIGH-MANAGED WATERSHEDS** | | | | |
| **Site Name** | **Independent Variable** | **Dependent Variable** | **Spearman’s**  **Rho (ρ)** | **P-value** |
| Young’s Branch | NDVI | Total Nitrogen (mg/L) | -0.096 | 0.705 |
| Total Phosphorus (mg/L) | -0.096 | 0.705 |
| Specific Conductance (µS/cm) | -0.014 | 0.929 |
| Temperature (°C) | 0.836 | 1.914e-11\* |
| Dissolved Oxygen (mg/l) | -0.681 | 1.834e-6\* |
| North Fork Quantico Creek | NDVI | Total Nitrogen (mg/L) | 0.233 | 0.351 |
| Total Phosphorus (mg/L) | 0.164 | 0.515 |
| Specific Conductance (µS/cm) | 0.021 | 0.894 |
| Temperature (°C) | 0.744 | 2.55e-8\* |
| Dissolved Oxygen (mg/l) | -0.611 | 2.178e-5\* |
| Blue Blazes Creek | NDVI | Total Nitrogen (mg/L) | 0.169 | 0.516 |
| Total Phosphorus (mg/L) | 0.282 | 0.273 |
| Specific Conductance (µS/cm) | -0.581 | 1.294e-4\* |
| Temperature (°C) | 0.714 | 4.793e-7\* |
| Dissolved Oxygen (mg/l) | -0.567 | 2.084e-4\* |
| **LOW-MANAGED WATERSHEDS** | | | | |
| **Site Name** | **Independent Variable** | **Dependent Variable** | **Spearman’s**  **Rho** | **P-value** |
| Bush Creek | NDVI | Total Nitrogen (mg/L) | -0.152 | 0.548 |
| Total Phosphorus (mg/L) | 0.243 | 0.332 |
| Specific Conductance (µS/cm) | -0.182 | 0.273 |
| Temperature (°C) | 0.705 | 5.343e-7\* |
| Dissolved Oxygen (mg/l) | -0.648 | 8.351e-6\* |
| Oxon Run | NDVI | Total Nitrogen (mg/L) | -0.408 | 0.093 |
| Total Phosphorus (mg/L) | 0.090 | 0.723 |
| Specific Conductance (µS/cm) | -0.405 | 0.010\* |
| Temperature (°C) | 0.492 | 1.478e-3\* |
| Dissolved Oxygen (mg/l) | -0.082 | 0.630 |
| Rock Creek Downstream of Dumbarton Oaks | NDVI | Total Nitrogen (mg/L) | -0.789 | 9.781e-5\* |
| Total Phosphorus (mg/L) | 0.251 | 0.315 |
| Specific Conductance (µS/cm) | -0.439 | 5.192e-3\* |
| Temperature (°C) | 0.506 | 1.024e-3\* |
| Dissolved Oxygen (mg/l) | -0.587 | 8.675e-5\* |
| *\*Values indicating statistical significance between variables* | | | | |



**Figure B1.** Total Nitrogen Time Series for Rock Creek

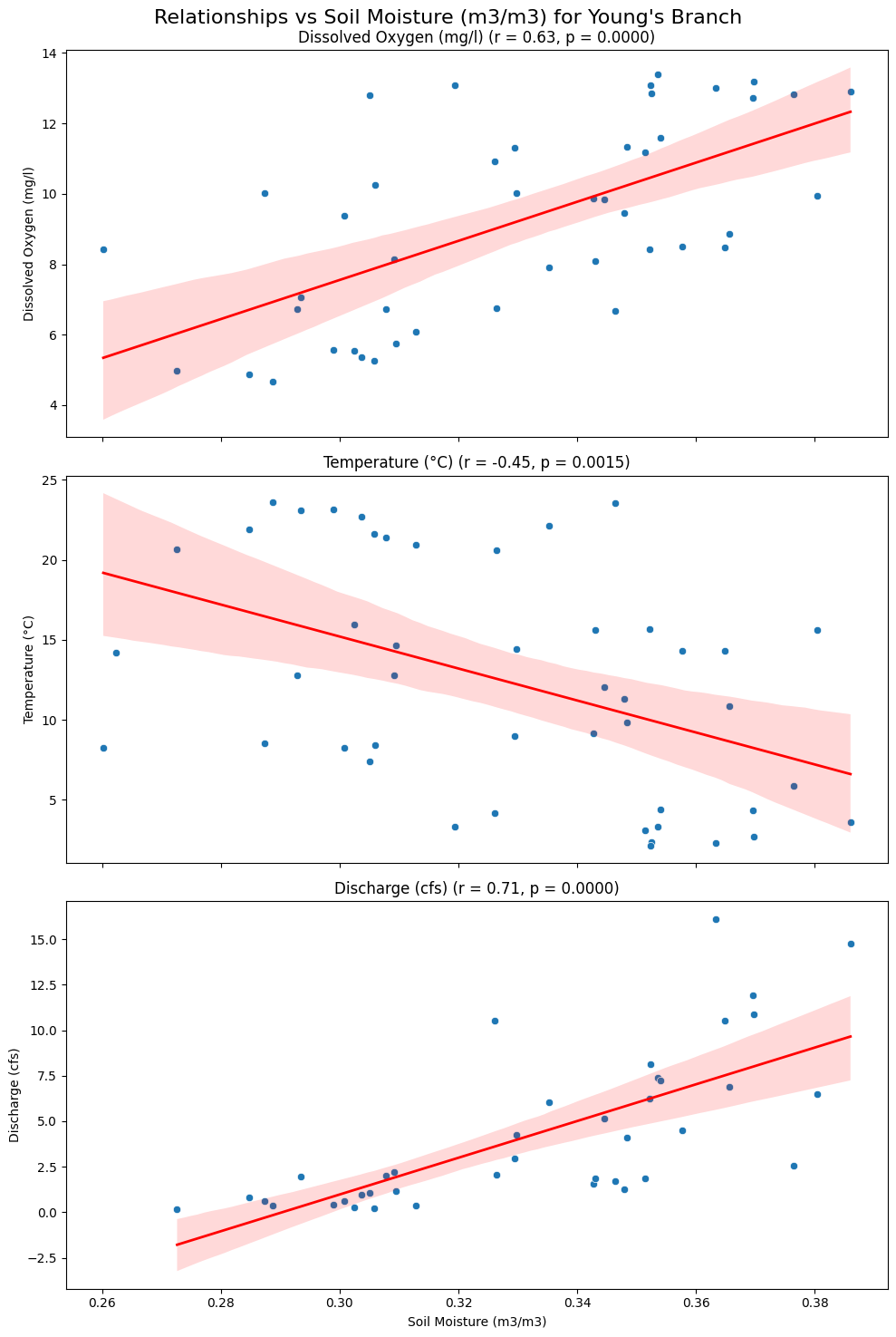


**Figure B2.** Total Phosphorous Time Series for Rock Creek

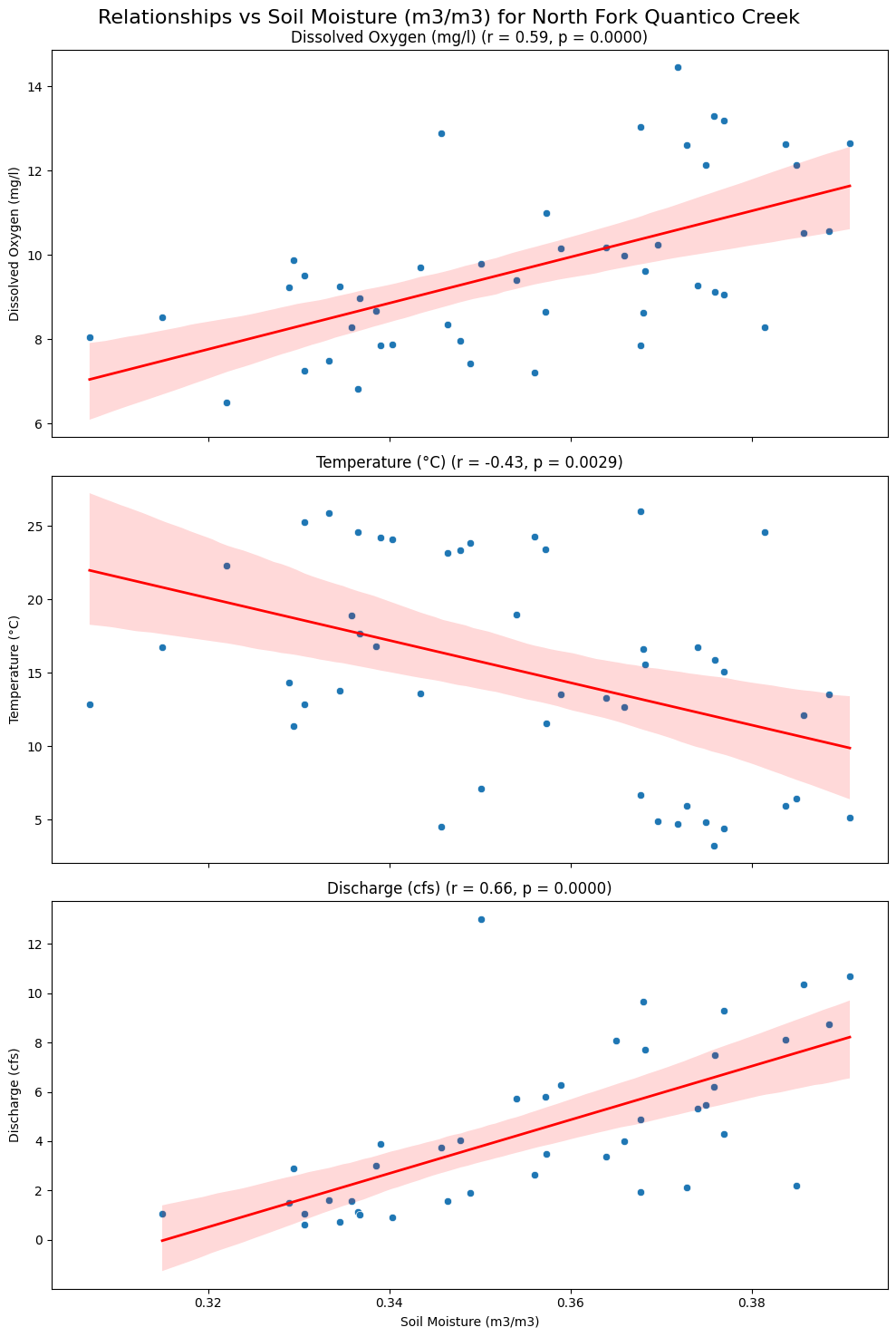
Appendix C – Analysis of Soil Moisture and Precipitation against Key Water Quality Parameters

**Table C1***. Pearson's correlation coefficient and p-values between the dependent variable and independent variables (soil moisture and precipitation) across six watershed areas.*

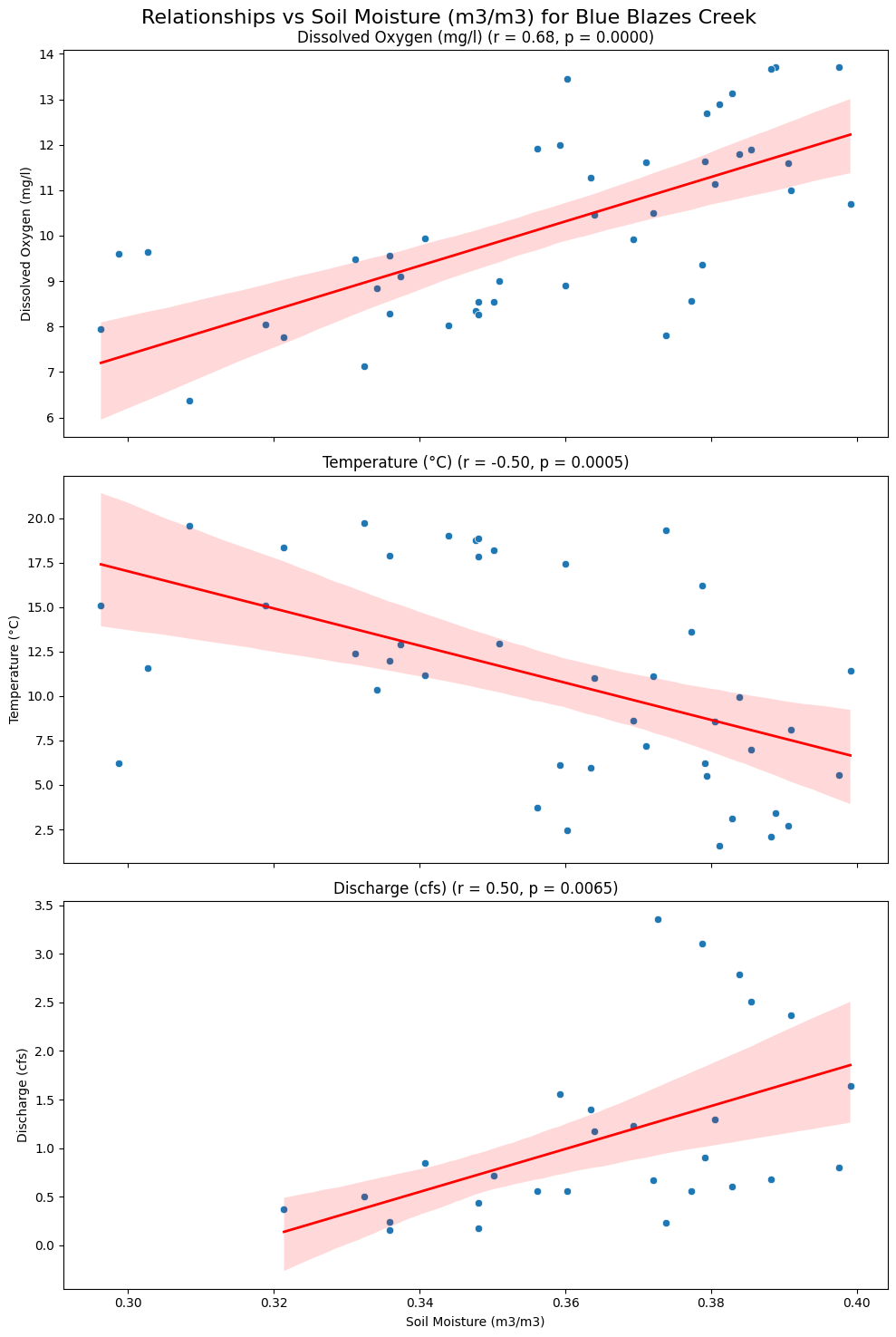
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HIGH-MANAGED WATERSHEDS** | | | | |
| **Site Name** | **Independent Variable** | **Dependent Variable** | **Pearson Correlation** | **P-value** |
| Young's Branch | Soil Moisture (m3/m3) | Dissolved Oxygen (mg/l) | 6.253047957e-1 | \*4.38340689e-6 |
| Soil Moisture (m3/m3) | Temperature (°C) | -4.547093923e-1 | \*1.499837889e-3 |
| Soil Moisture (m3/m3) | Discharge (cfs) | 7.146124828e-1 | \*2.22673294e-7 |
| Precipitation (mm) | Dissolved Oxygen (mg/l) | -1.948019526e-1 | 1.997289555e-1 |
| Precipitation (mm) | Temperature (°C) | 4.208366073e-1 | 3.588017989e-3 |
| Precipitation (mm) | Discharge (cfs) | -8.698076055e-2 | 5.935603437e-1 |
| North Fork Quantico Creek | Soil Moisture (m3/m3) | Dissolved Oxygen (mg/l) | 5.868145795e-1 | \*1.456712156e-5 |
| Soil Moisture (m3/m3) | Temperature (°C) | -4.257274577e-1 | \*2.850384136e-3 |
| Soil Moisture (m3/m3) | Discharge (cfs) | 6.624503865e-1 | \*1.763552982e-6 |
| Precipitation (mm) | Dissolved Oxygen (mg/l) | 2.488841602e-1 | 9.160737438e-2 |
| Precipitation (mm) | Temperature (°C) | 3.896979548e-1 | \*6.77601161e-3 |
| Precipitation (mm) | Discharge (cfs) | -2.360734611e-2 | \*8.820298358e-1 |
| Blue Blazes Creek | Soil Moisture (m3/m3) | Dissolved Oxygen (mg/l) | 6.836641212e-1 | \*3.138654537e-7 |
| Soil Moisture (m3/m3) | Temperature (°C) | -5.031364451e-1 | \*4.994572549e-4 |
| Soil Moisture (m3/m3) | Discharge (cfs) | 5.019183087e-1 | \*6.499770406e-3 |
| Precipitation (mm) | Dissolved Oxygen (mg/l) | 3.437815512e-2 | 8.246724482e-1 |
| Precipitation (mm) | Temperature (°C) | 1.067736935e-1 | 4.902928907e-1 |
| Precipitation (mm) | Discharge (cfs) | 4.980405502e-1 | \*6.994953537e-4 |
| **LOW-MANAGED WATERSHEDS** | | | | |
| **Site Name** | **Independent Variable** | **Dependent Variable** | **Pearson Correlation** | **P-value** |
| Bush Creek | Soil Moisture (m3/m3) | Dissolved Oxygen (mg/l) | 7.946200603e-1 | \*0 |
| Soil Moisture (m3/m3) | Temperature (°C) | -7.652949619e-1 | \*9.227127414e-10 |
| Soil Moisture (m3/m3) | Discharge (cfs) | 5.673670344e-1 | \*2.503077345e-4 |
| Precipitation (mm) | Dissolved Oxygen (mg/l) | -1.03340435e-1 | 4.993370293e-1 |
| Precipitation (mm) | Temperature (°C) | 2.716696305e-1 | 7.103512965e-1 |
| Precipitation (mm) | Discharge (cfs) | 6.835944769e-2 | 6.87674205e-1 |
| Oxon Run | Soil Moisture (m3/m3) | Dissolved Oxygen (mg/l) | -6.45749413e-2 | 6.807804681e-1 |
| Soil Moisture (m3/m3) | Temperature (°C) | 2.242515009e-1 | 1.386258548e-1 |
| Soil Moisture (m3/m3) | Discharge (cfs) | -2.881649521e-2 | 8.675026039e-1 |
| Precipitation (mm) | Dissolved Oxygen (mg/l) | -2.180997691e-1 | 1.600176443e-1 |
| Precipitation (mm) | Temperature (°C) | 4.116424139e-1 | 4.96317477e-3 |
| Precipitation (mm) | Discharge (cfs) | 5.713367181 e-3 | 9.736182761e-1 |
| Rock Creek Downstream of Dumbarton Oaks | Soil Moisture (m3/m3) | Dissolved Oxygen (mg/l) | 4.672276147e-1 | \*1.213012349e-3 |
| Soil Moisture (m3/m3) | Temperature (°C) | -3.498082291e-1 | \*1.849734896e-2 |
| Soil Moisture (m3/m3) | Discharge (cfs) | 6.543960165e-1 | \*1.49617161e-5 |
| Precipitation (mm) | Dissolved Oxygen (mg/l) | -3.309761187e-1 | \*2.636567869e-2 |
| Precipitation (mm) | Temperature (°C) | 4.140377218e-1 | \*4.692803918e-3 |
| Precipitation (mm) | Discharge (cfs) | 3.900270522e-1 | \*1.869487907e-2 |
| *\*Values indicating statistical significance between variables* | | | | |



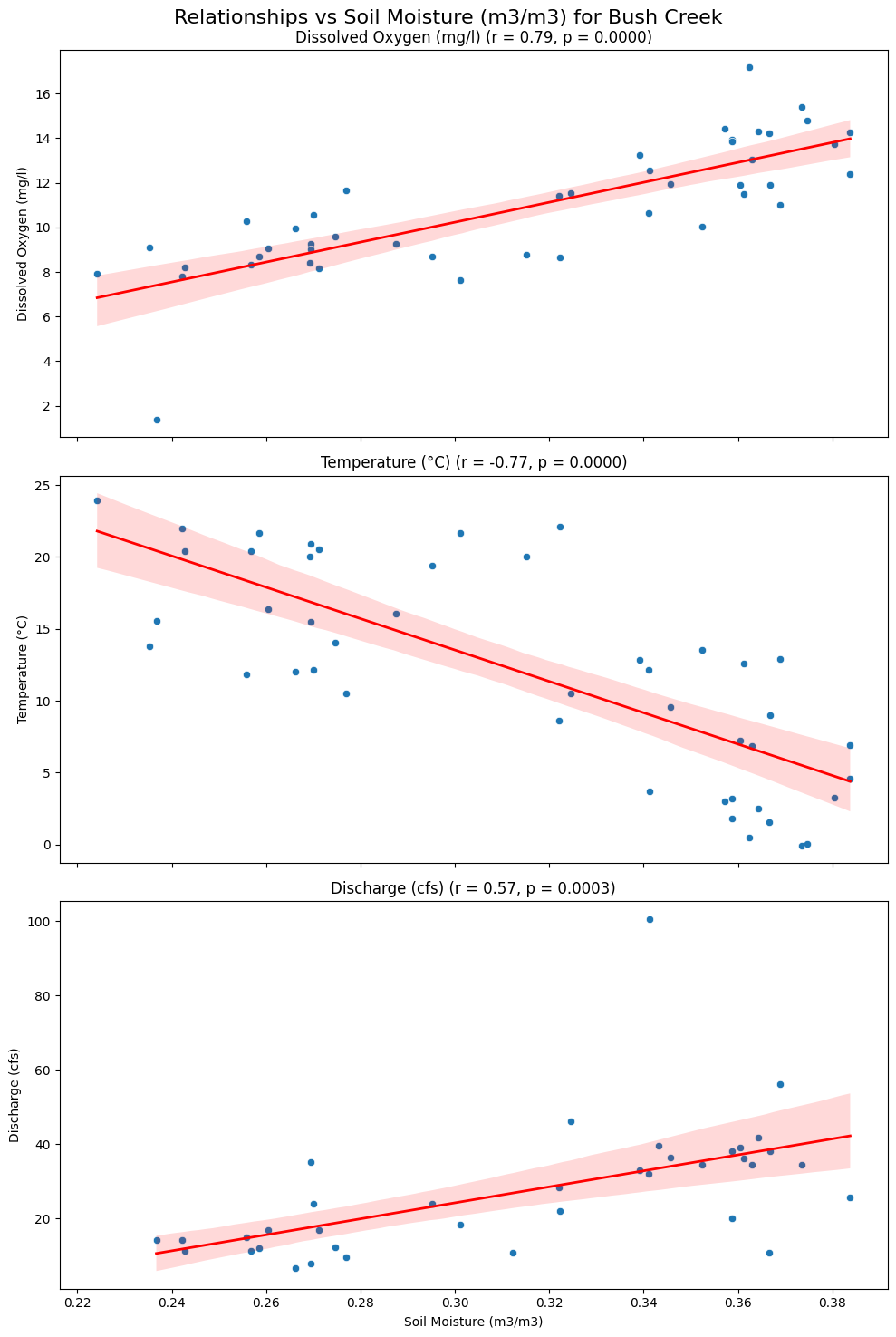
***Figure C2****.* Young’s Branch: Correlation plot between Soil Moisture and dependent variables.



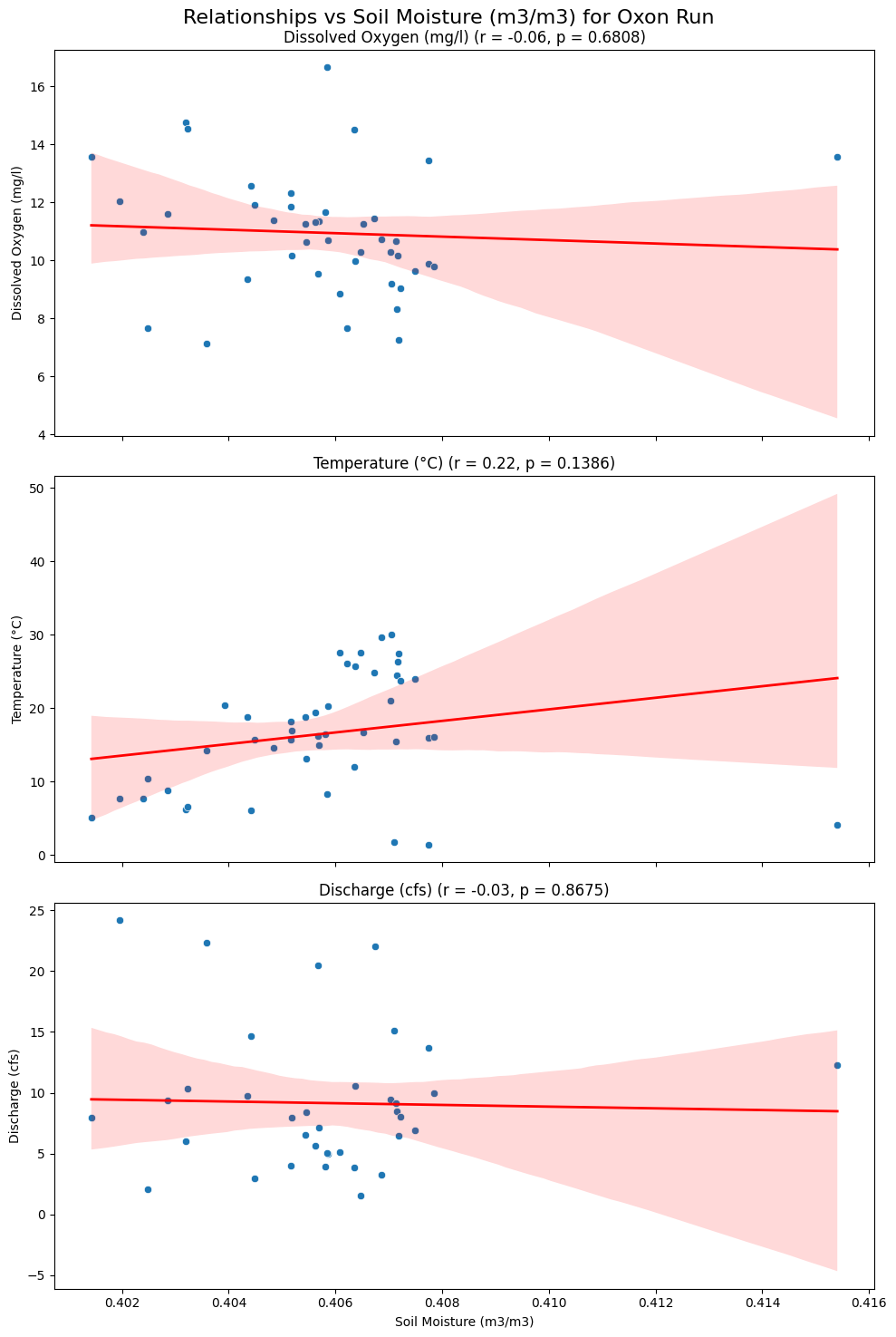
***Figure C4.*** North Fork Quantico: Correlation plot between Soil Moisture and dependent variables.



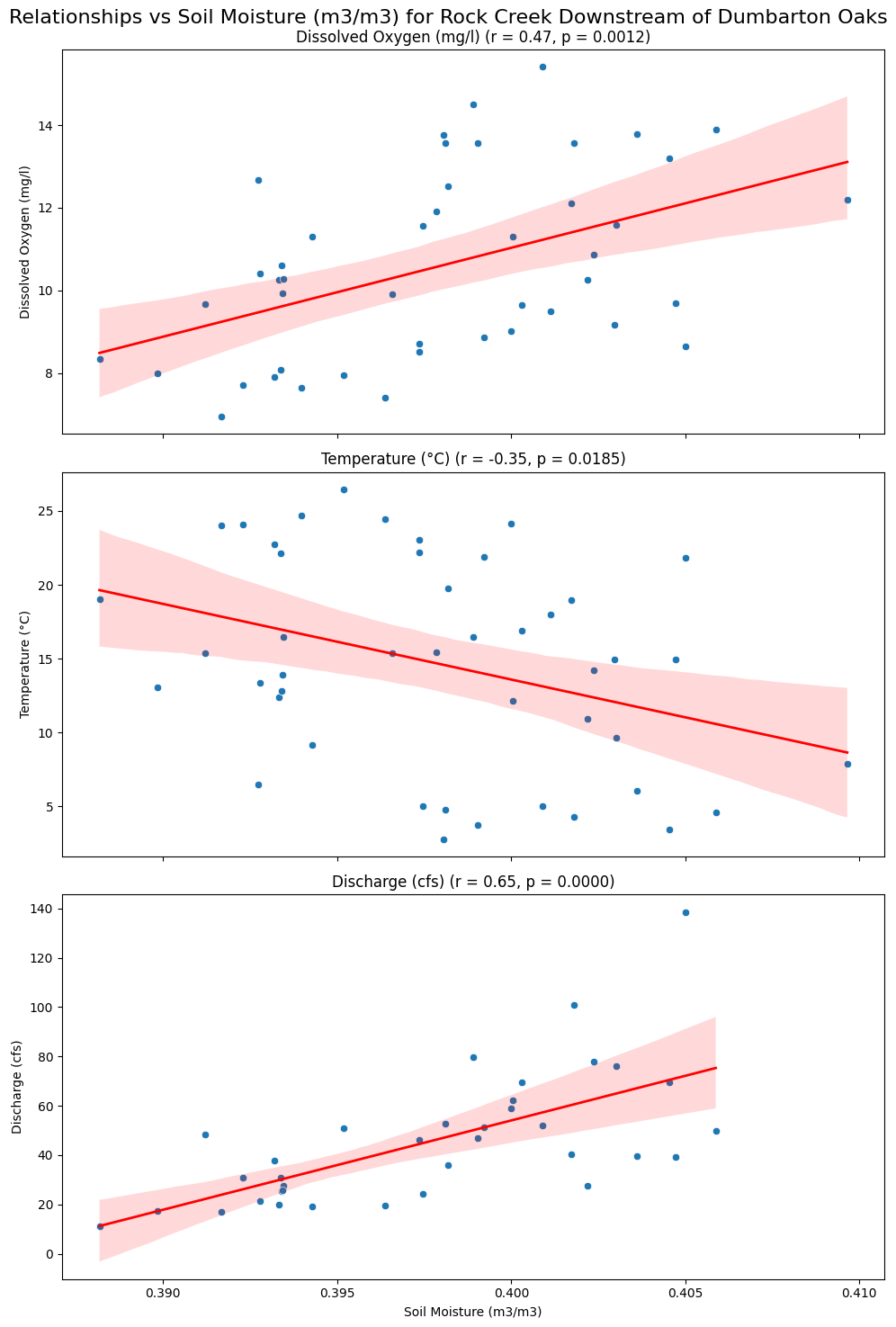
***Figure C5***Blue Blaze: Correlation plot between Soil Moisture and dependent variables.



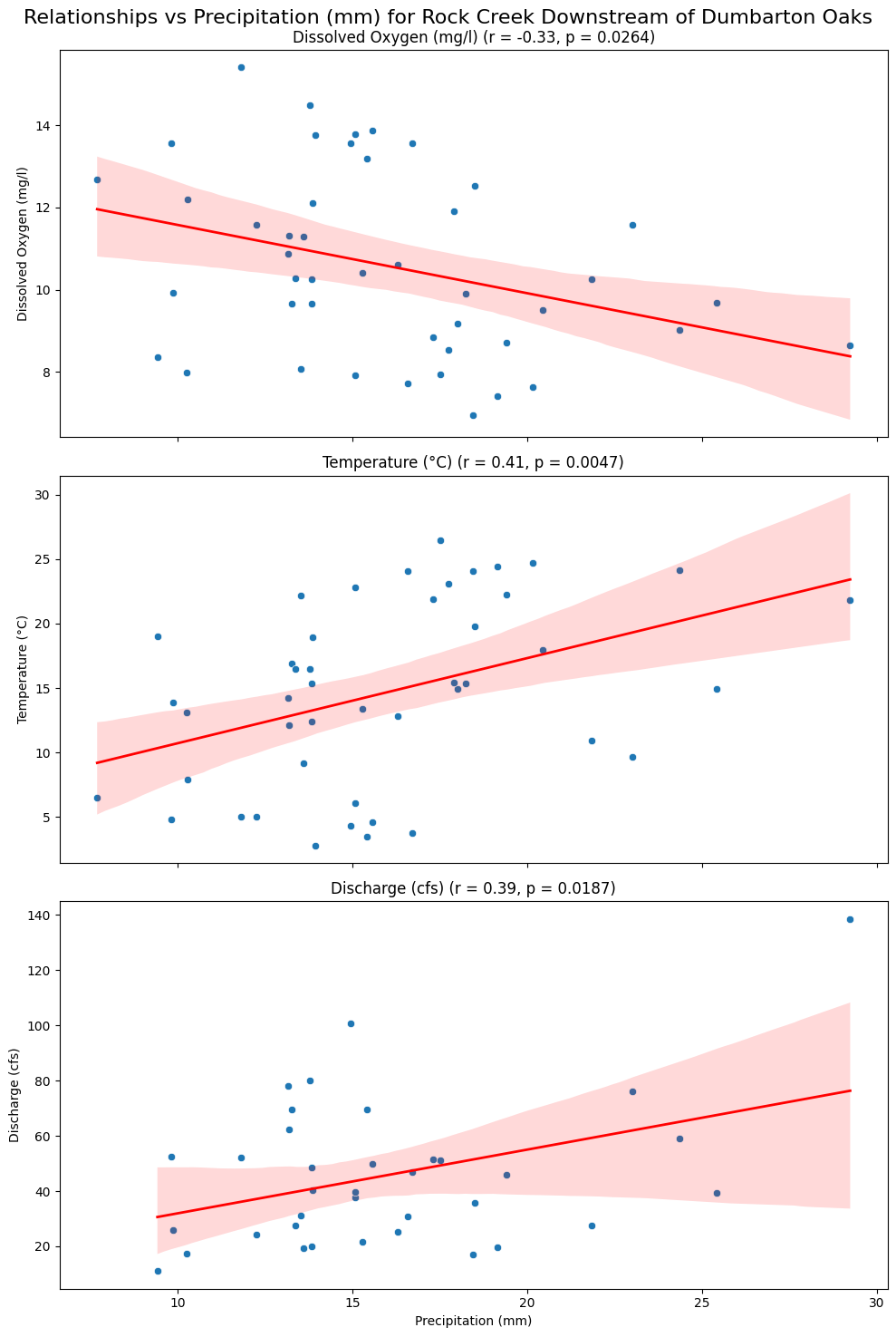
***Figure C6****.* Bush Creek: Correlation plot between Soil Moisture and dependent variables.



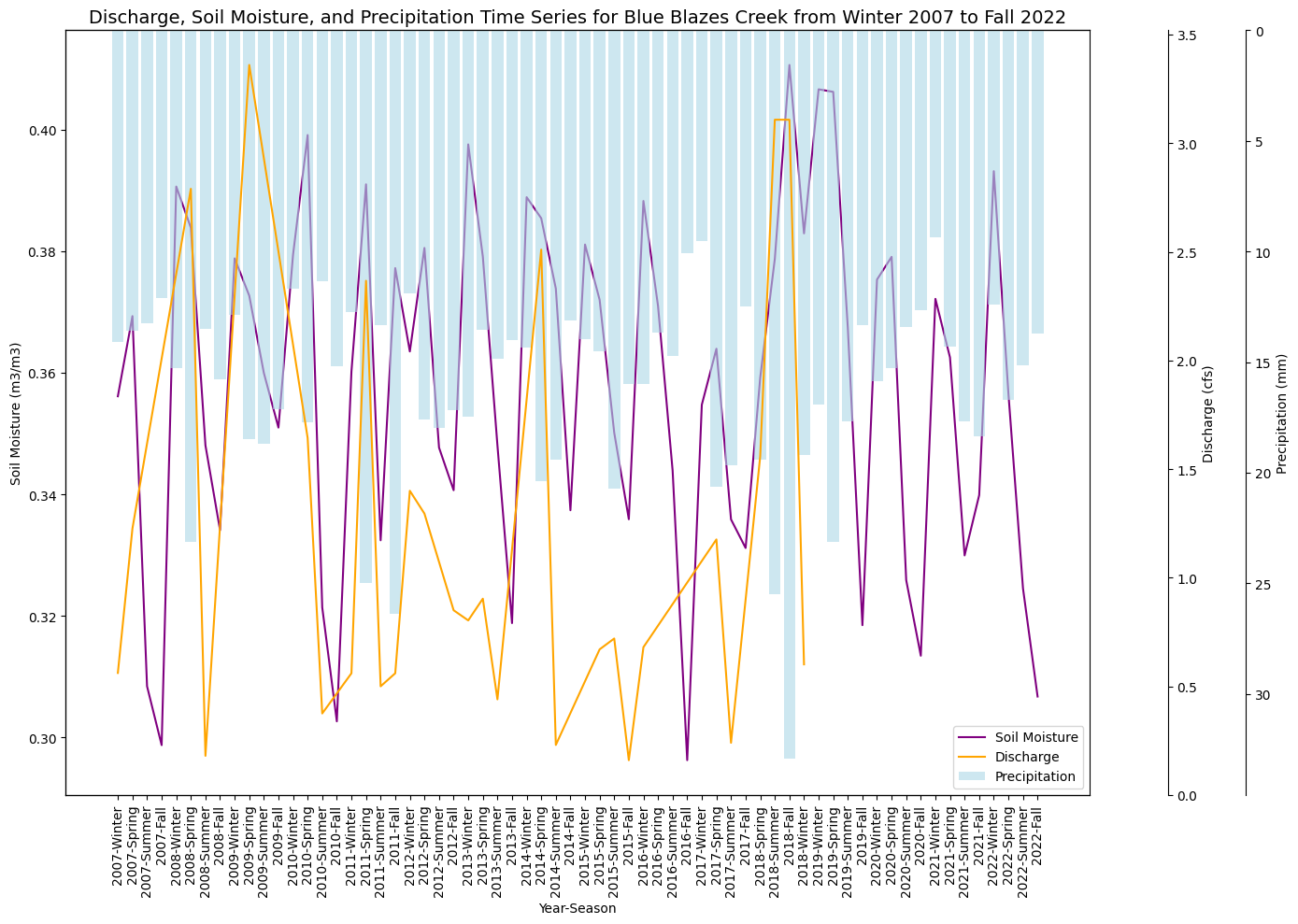
***Figure C7****.* Oxon Run: Correlation plot between Soil Moisture and dependent variables.



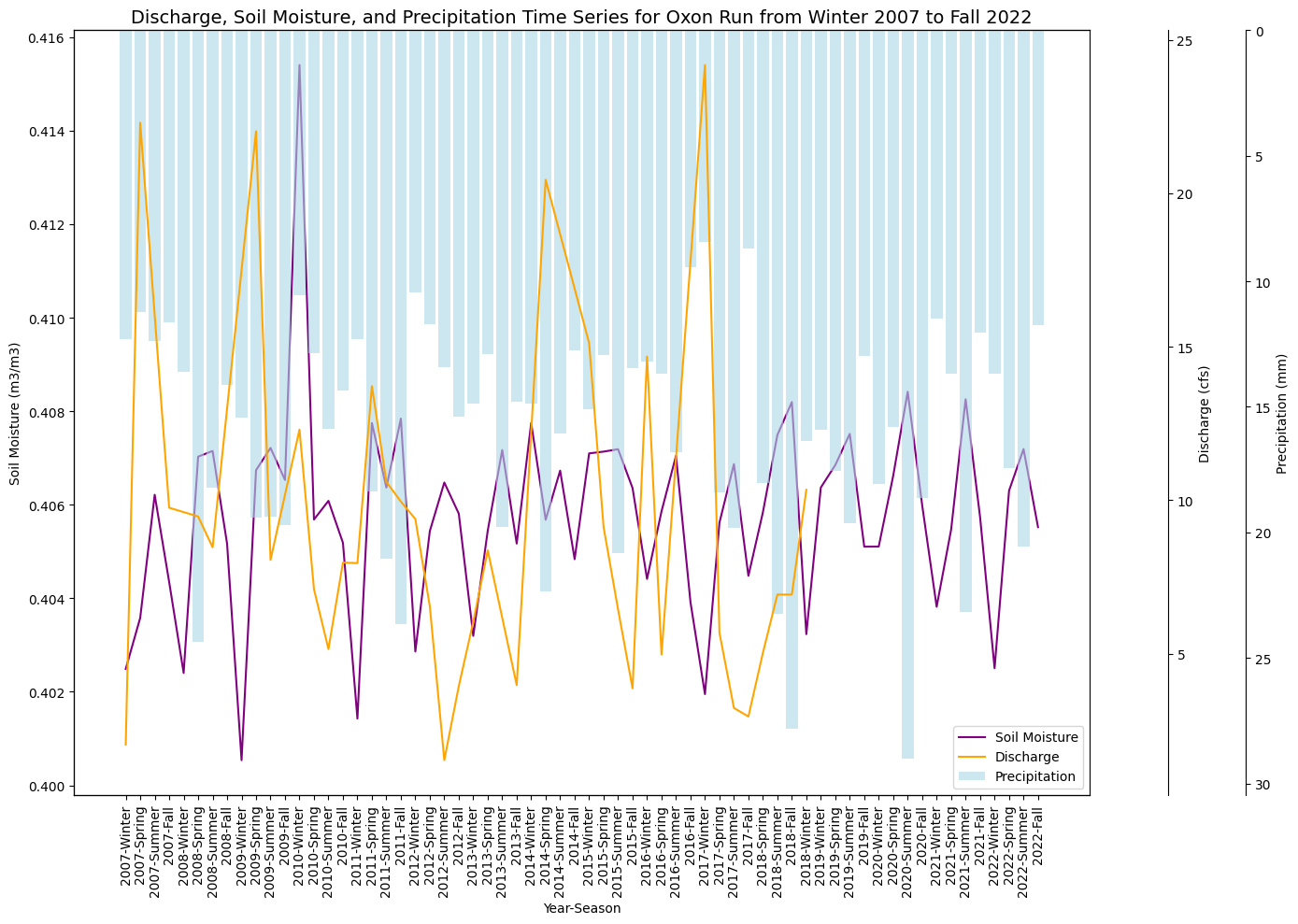
***Figure C8****.* Rock Creek: Correlation plot between Soil Moisture and dependent variables.



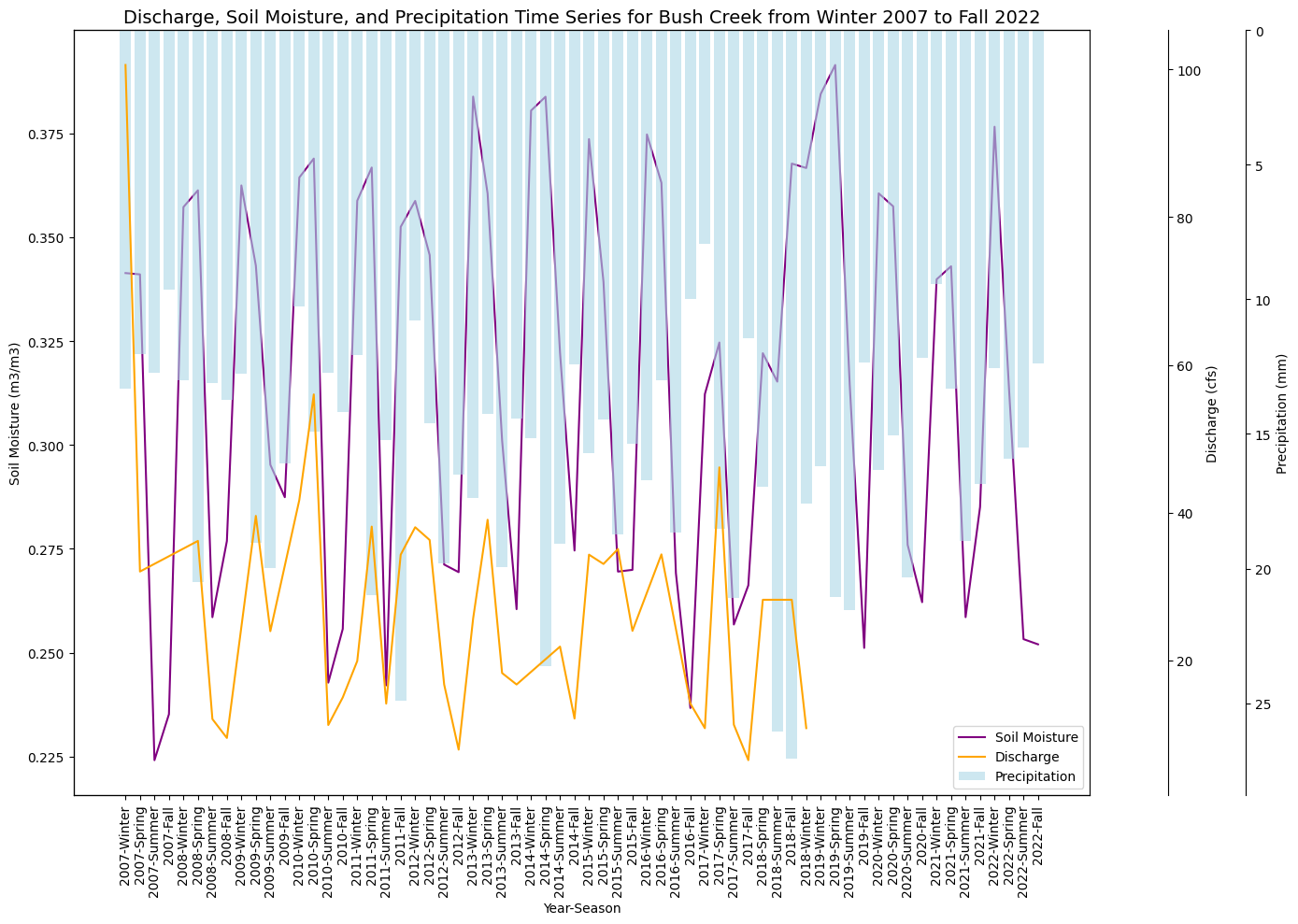
***Figure C9****.* Rock Creek: Correlation plot between Precipitation and dependent variables.



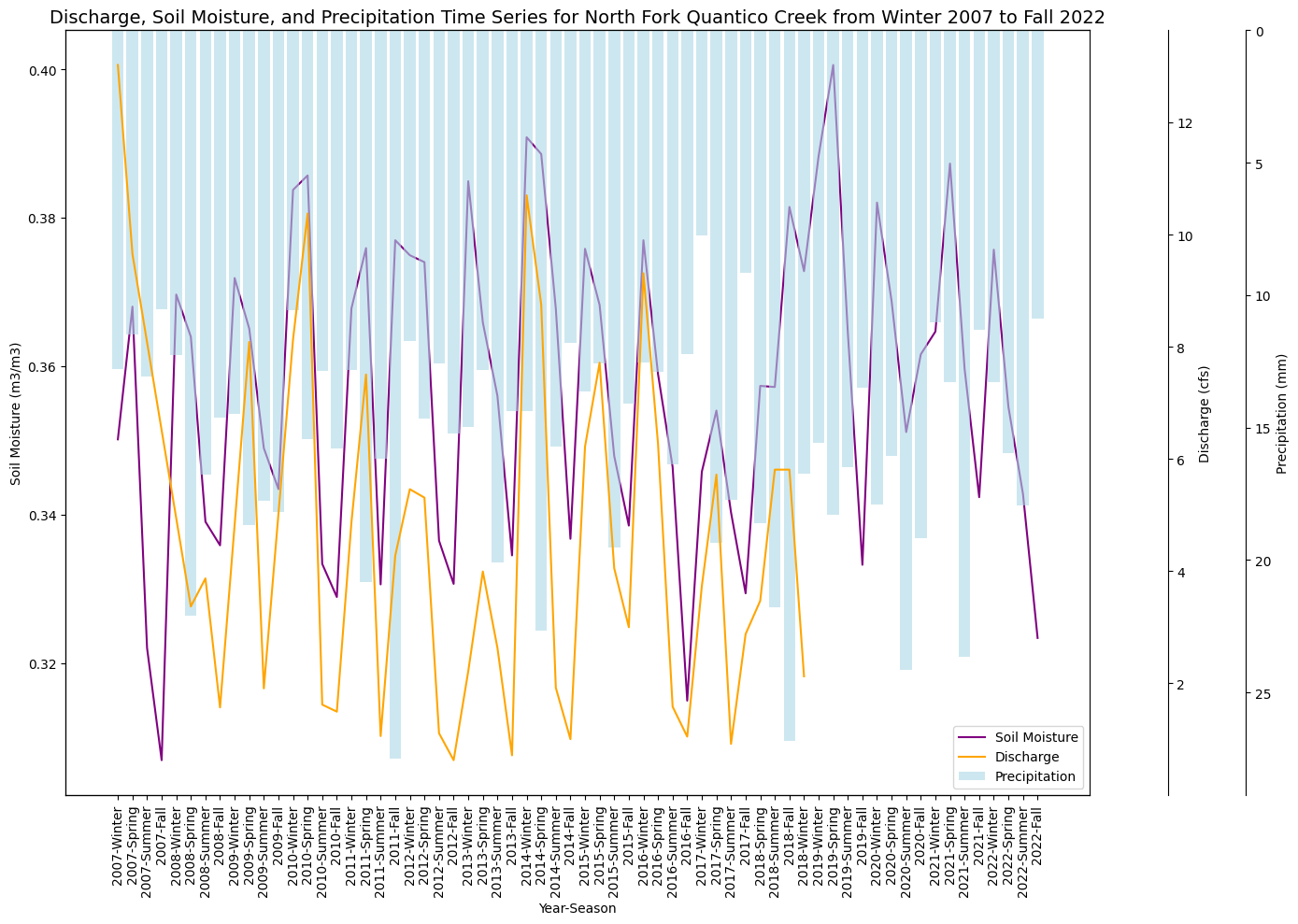
**Figure C10**. Water quantity variables time series for Blue Blazes creek.



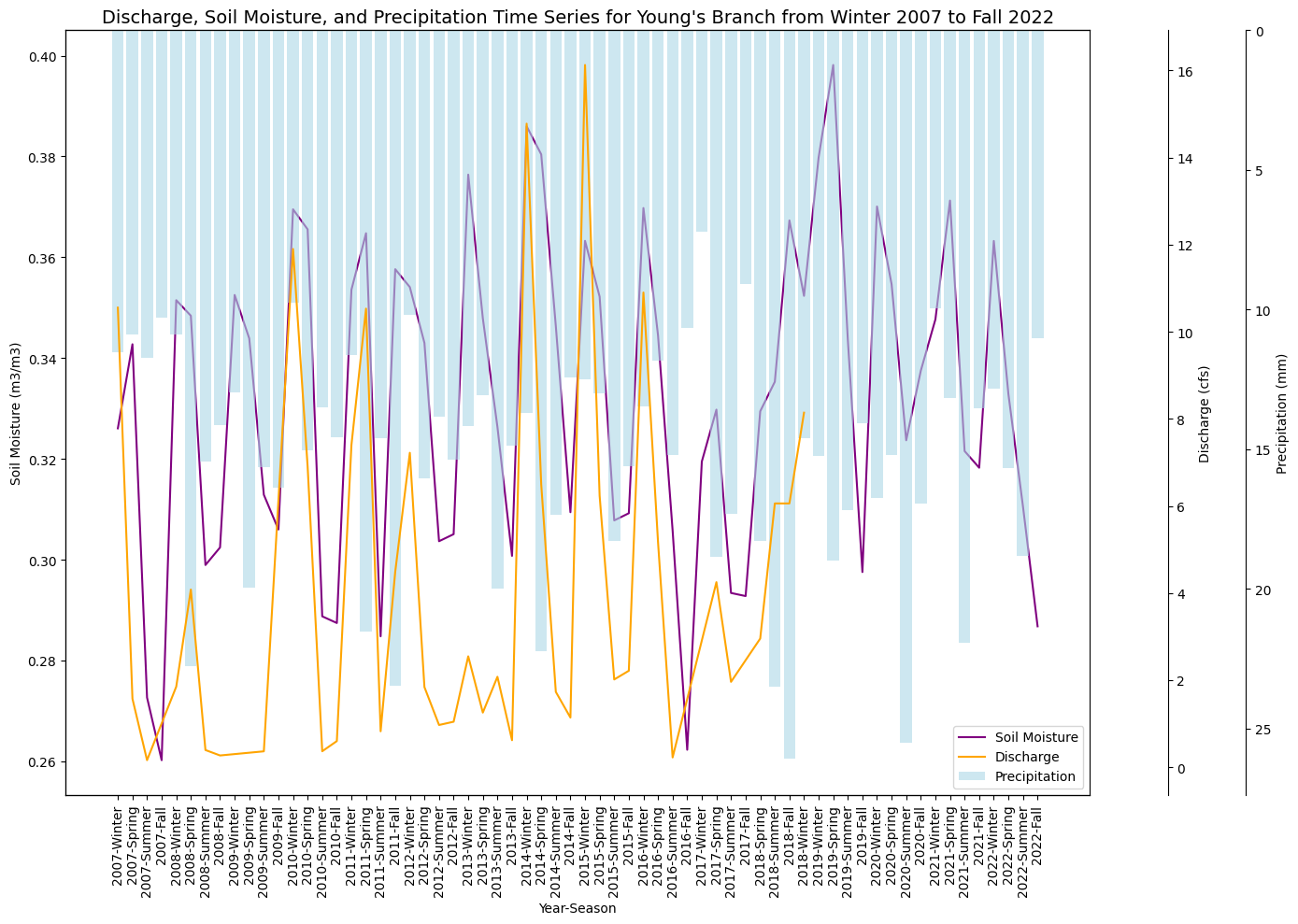
**Figure C11**. Water quantity variables time series for Oxon Run.



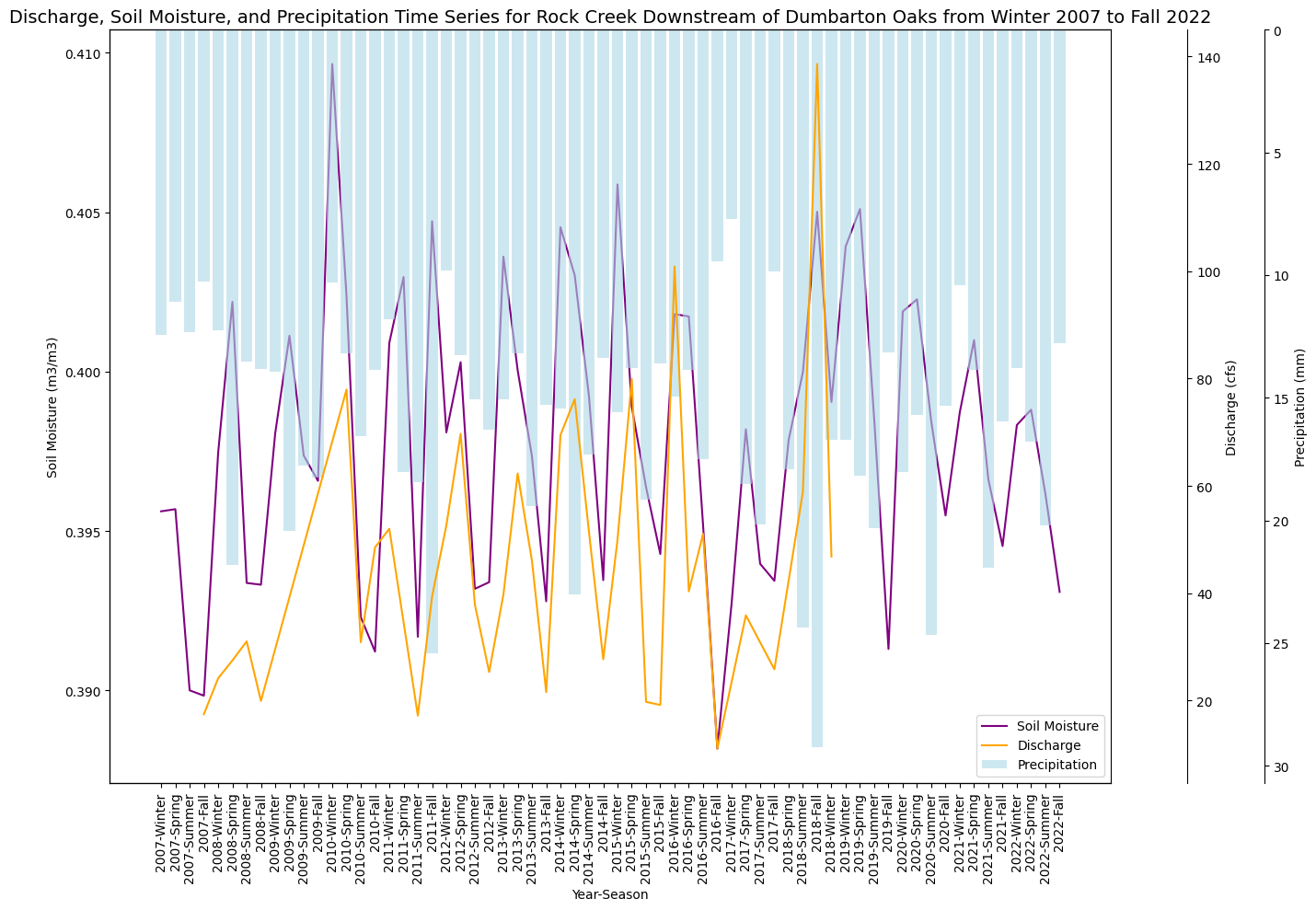
**Figure C12**. Water quantity variables time series for Bush Creek.



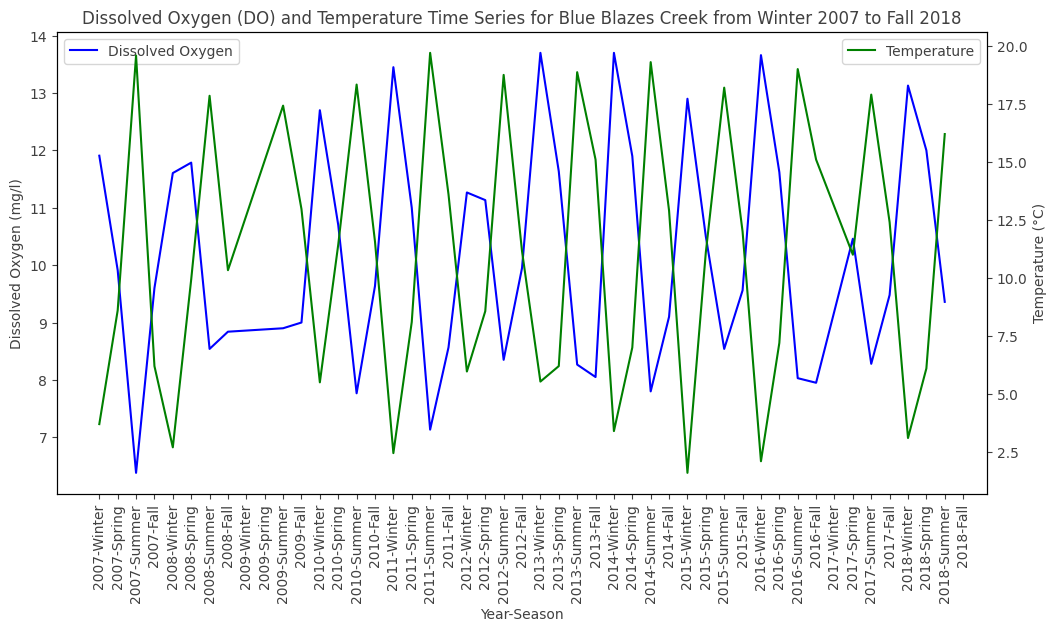
**Figure C13**. Water quantity variables time series for North Fork Quantico Creek.



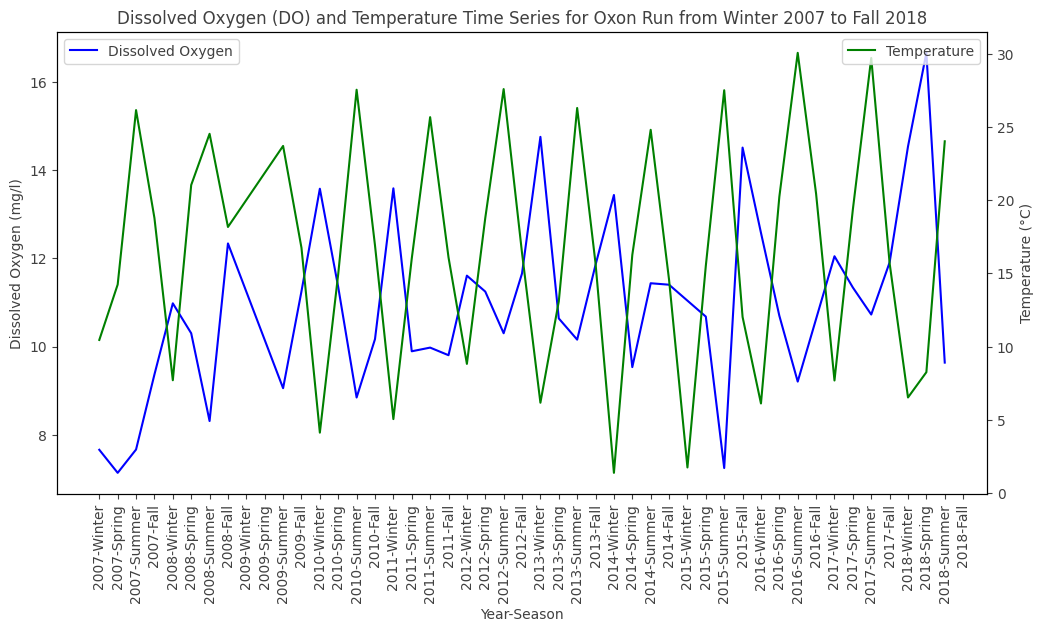
**Figure C14**. Water quantity variables time series for Young’s Branch.



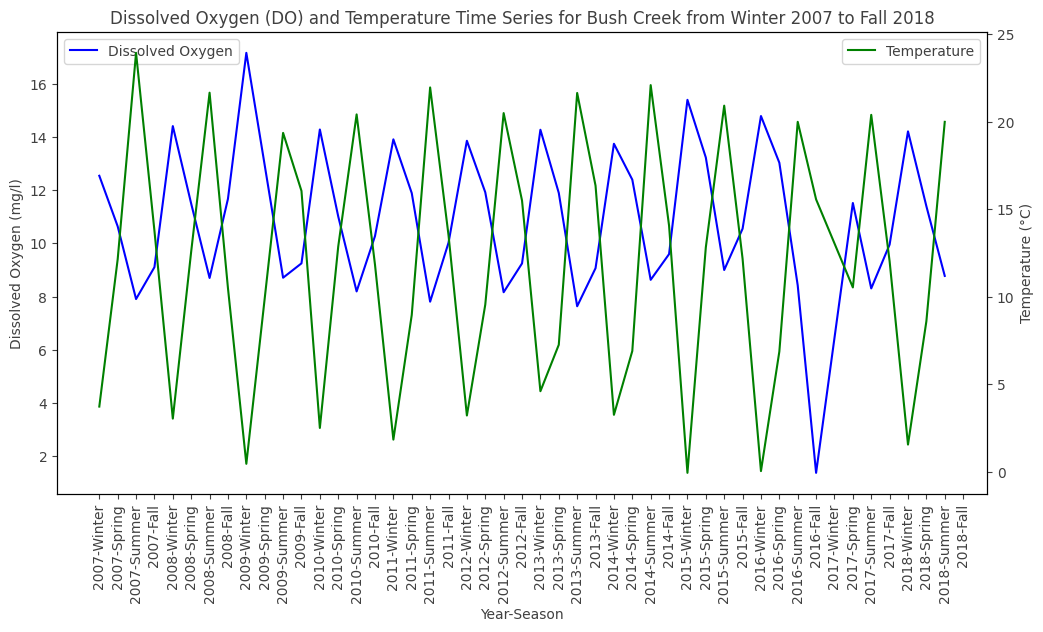
**Figure C15**. Water quantity variables time series for Rock Creek Downstream of Dumbarton Oaks.



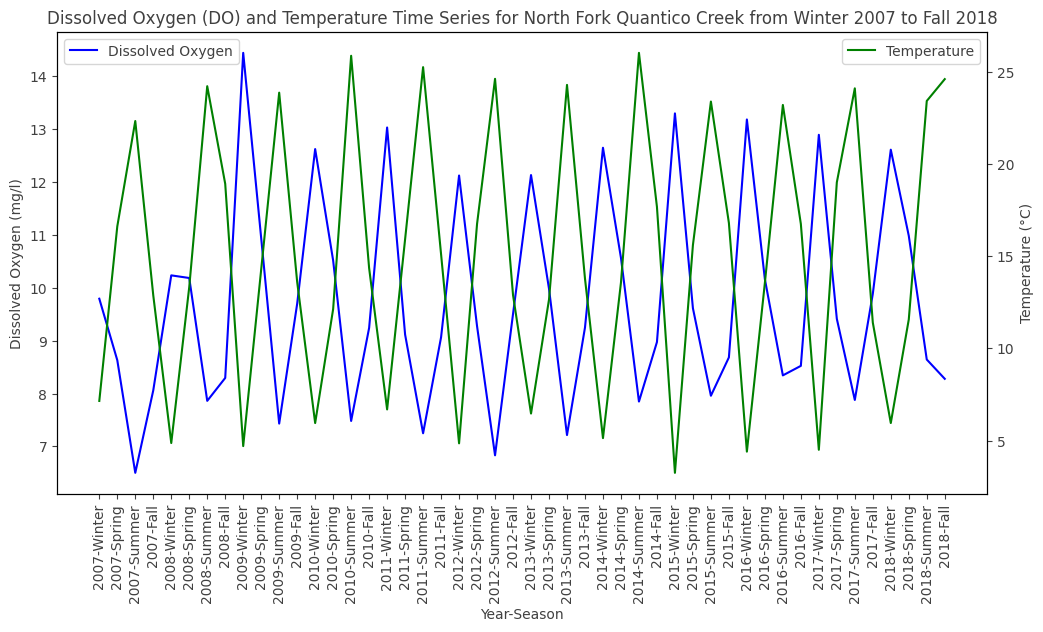
**Figure C16**. Dissolved Oxygen and Temperature time series for Blue Blazes creek.



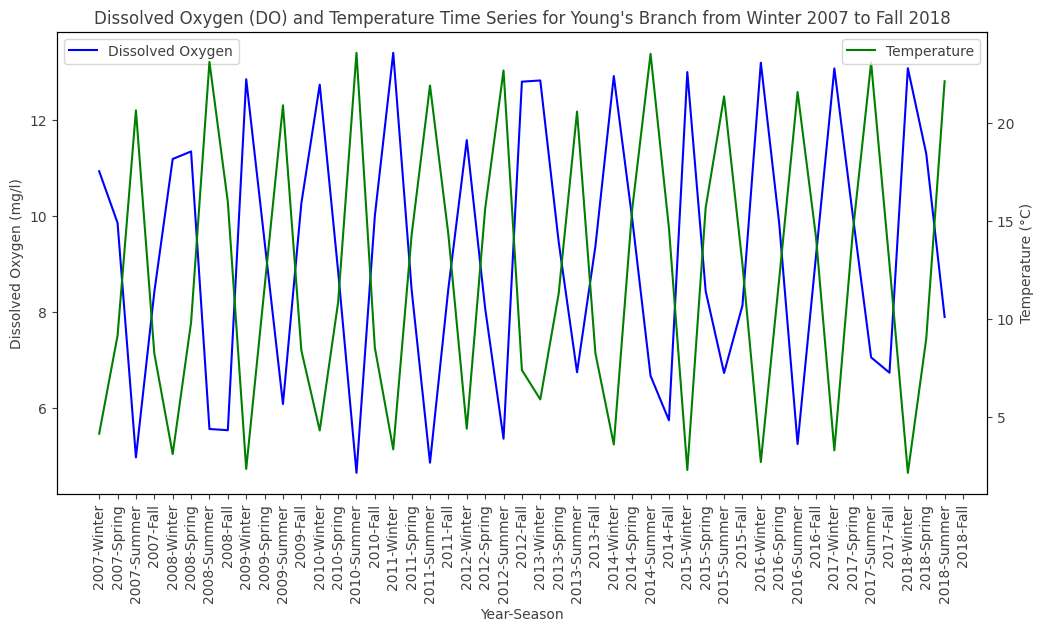
**Figure C17**. Dissolved Oxygen and Temperature time series for Oxon Run.



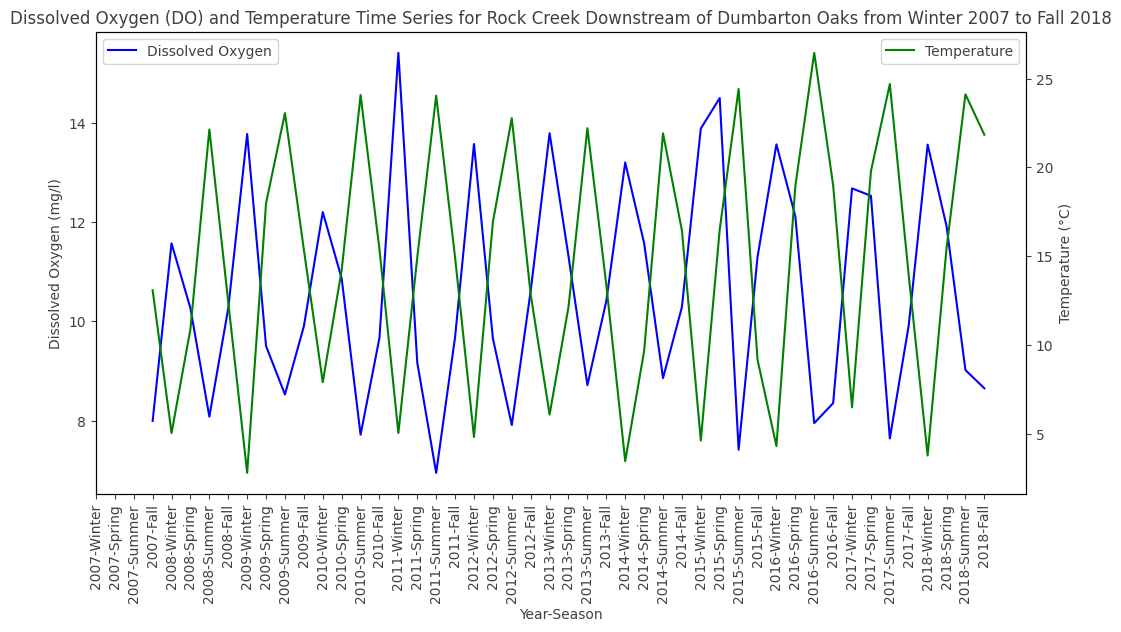
**Figure C18**. Dissolved Oxygen and Temperature time series for Bush Creek.



**Figure C19**. Dissolved Oxygen and Temperature time series for North Quantico Creek.



**Figure C20**. Dissolved Oxygen and Temperature time series for Young’s Branch.



**Figure C21**. Dissolved Oxygen and Temperature time series for Rock Creek Downstream of Dumbarton Oaks.