**Quantifying divided watershed governance and its correlation with water problem burden.**

CP 6596 Summer 2025

Prepared for Dr Sivakumar

Submitted 7/25/2025

Submitted toward the completion of the requirements of MS-GIST

**Acknowledgements**

I would like to acknowledge the Muscogee Creek people, whose confederacy had, prior to the treaty of Indian springs, stewarded the land now occupied by the Georgia Institute of Technology. I also would like to acknowledge the horrific suffering brought upon the people of the Muscogee Creek by the state of Georgia.

I would also like to acknowledge, first and foremost, Dr Robert Kirkman for challenging me to think about scientific and engineering issues in their ethical and human context. I would not be the thinker and professional I am today without him, and what he helped me to see in the works of Aristotle and Wendell Berry. As I was taught by Dr Kirkman, when asked which political and social system could accurately capture and protect the value of land, Wendell Berry answered “the kingdom of God.” His remark, although somewhat sarcastic, captured his belief that stewardship of land, and all of the soils, waters, and ecologies that comprise it, is a matter of cultural values, not of governance. Berry further argued that science was not the path out of our cultural and agricultural crisis, now popularly referred to as “the climate crisis,” because science was what created this crisis. Instead of attempting to “study, fund, and organize” our way out of the ecological catastrophes we have created with the products of science, Berry implores us instead to embrace a culture and agriculture centered around community as well as a deeper connection with nature. It is Berry’s very contempt for the knowledge economy of academia that inspired me to write this paper studying the relationship between the organization of watershed management entities as the water stewardship failures we see all around us.

Lastly, I would like to thank the entire staff of Trees Atlanta for teaching me to know and love the Georgia Piedmont and its forests.

**Abstract**

Watershed management is often conducted by municipal or county agencies whose jurisdiction ends at the border of the city or county, and not at the border of the watershed. Bio-regionalist thinkers have hypothesized that this represents a substantial boundary to effective watershed governance as it present challenges for agencies to incentivize water resource stewardship, penalize egregious pollution, and enact adequate flood and drought control measures throughout their watershed. Georgia watersheds were examined and scored according to various metrics measuring the number of local governments responsible for their management, the length of local government boundaries running through them, and the proportion of their control by a single local government. Area weighted averages were created for flood risk, drought severity, and water quality in each watershed. Statistical analysis revealed metrics for divided watershed governance bore no strong association with flood risk in much of the state but were moderately negatively associated with flood risk in parts of the state such as Augusta, and slightly positively associated in other parts of the state, such as in Savannah. Analysis revealed very little association between divided watershed governance and drought severity or water quality. This study reveals the limited and variable influence of local government boundaries on watershed issues.

**Table of contents**

|  |  |  |
| --- | --- | --- |
| Section number | Section description | Page number |
| 1 | Introduction |  |
| 2 | Literature Review |  |
| 3 | Methodology |  |
| 4 | Results and Analysis |  |
| 4.1 |  |  |
| 5 | Discussion |  |
| 6 | Conclusion and recommendations |  |
| 7 | References |  |
| 8 | List of figures |  |
| 9 | List of charts |  |
| 10 | List of tables |  |
| 11 | Appendix A: Code & Data |  |

**Introduction**

Watersheds gather water to a common outlet. I therefore believe that the inhabitants of watersheds often have more water-related problems in common that the inhabitants of counties or cities do. This is reflected in the way residents of the Proctor Creek watershed share problems such as sewage, sediment, and industrial pollutants in their rivers that their neighbors in other watersheds but the same municipalities do not. Water can carry a host of pollutants and hazards from E-Coli to heavy metals to raw sewage and that does so ignoring property lines and political boundaries. Therefore, it follows that if a pollutant is improperly handled a watershed, the pollutant will be the shared burden of the watershed. It also follows that if flooding or drought is poorly controlled in the watershed, it will be the shared burden of the watershed. These problems, which we might call “water problems,” therefore seem much less the function of the actions of individuals, cities, or states, but of the community of a particular watershed.

Therefore, it would appear the way we manage watersheds is fundamentally wrong-headed. Watersheds are ordinarily managed by county, state, and municipal governments. For example, the portions of Chattahoochee watershed that lie within the city of Atlanta, are management by the department of Watershed Management. However, on the other side of the Chattahoochee River, this department has no jurisdiction. In fact, the Proctor Creek-Chattahoochee River watershed unit has no less than 4 different municipal and county agencies managing parts of the land comprising it. It is dissected by over 15 kilometers of municipal boundaries and county lines, despite being only 61.63 square kilometers large. This is what I call a watershed with divided governance. Its inhabitants must redress their grievances over their poisoned creeks with four different political entities, despite not being a constituent of all four. This has provoked some, known as bio-regionalists, to argue that we must rethink the shape of our governments such that citizens are always deemed constituents of the land that their livelihood relies on.

The intention of this paper is to test the hypothesis that there is a correlation between divided watershed governance or “watershed dividedness”, and drought, flooding, and water pollution or “water problem burden” in the state of Georgia. This will first involve operationalizing watershed dividedness, which will output a map of watersheds with dividedness encoded. It will then involve operationalizing the water problem burden, which will output a map of watersheds with water problem burden encoded. Lastly, it will involve running and refining regressions to determine if there is a statistically significant correlation between dividends and burden, and if such correlation explains a notable portion of the variance in burden. This results in the following research questions:

1. In the state of Georgia, which watersheds are most divided by political boundaries?
2. In the state of Georgia, do more divided watersheds suffer greater incidence of flooding?
3. In the state of Georgia, do more divided watersheds suffer greater incidence of drought?
4. In the state of Georgia, do more divided watersheds suffer greater incidence of pollution?
5. In the state of Georgia, do more divided watersheds suffer greater incidence of water problem burden?

I hypothesize that watersheds in the northeast of the state, and watersheds in urban areas, will be more divided by political boundaries. I hypothesize that these more divided watersheds will suffer greater incidence of all water problem burdens including flooding, drought, and pollution.

**Literature Review**

**Fires, droughts, and floods are localized, mitigated, or influenced by watersheds**

Climate related disasters such as fires, droughts, and floods occur differently in different watersheds, suggesting the need for different adaptation and mitigation strategies. Different watersheds recover differently according to an analysis of remotely sensed data studying Western US watersheds, with various watersheds returning to pre-disaster flows and runoff ratios differently than others (Saxe, Hogue, & Hay, 2018). As per an analysis of 2275 watersheds, differing watersheds, particularly the plant life they contain, also recover differently from droughts, (Xue, Wang, Xiao, & Helman, 2020). For example, drier watersheds appear to contain vegetation more resilient to drought. Furthermore, properties of watersheds such as recharge seasonality, storage-discharge relationship, and sensitivity of baseflow storage control the intensity of droughts (Apurv & Cai, 2020). Flood risk and existing flood regulation such as artificial water bodies and runoff prevention practices also vary widely between watersheds, as found by an analysis of long term flood records (Mogollón, Villamagna, & Frimpong, 2016). Floods and droughts are also subject to watershed specific factors such as land use change as they are to global meteorological factors. (Whitfield, 2012), meaning that updated development guidelines may be just as important as global atmospheric changes resulting in increased or decreased rainfall to a watershed’s outcomes.

**Pollution and erosion are localized to watersheds**

Anthropogenic disturbances such as to water quality and quantity, as well as the subsequent benefits and harms to people and wildlife occur differently across varying watersheds. Watersheds are the fundamental unit of monitoring and managing non point source pollution such as nitrate and sediment runoff (Yuan, Sinshaw, & Forshay, 2020). Likewise, soil erosion is a result of watershed level factors such as landscape types (Zhang, 2017).

**Water resource management between municipalities**

Municipal boundaries also prompt conflict over resources that the boundary divides, particularly water resources. In the American west, where irrigation has long been vital for the success of cropland (United States Congress, 1890), there arise numerous water conflicts prompted by the discontinuity between political boundaries and the boundaries of watersheds (Powell, 1890).

**General impact of municipal boundaries**

People usually work, live, and recreate in the same municipality with peers who share that municipality. Municipal boundaries discourage people from working, shopping, and engaging in other economic activity across the boundaries, as shown by a spatial interaction analysis (Tordoir, Raan, & Poorthuis, 2023). Likewise, and consequently, municipal boundaries that divide governance of a watershed present challenges to water governance requiring work to circumvent (Gordon & Jones, 2000). For example, enforcing controls to runoff quality is not effective in avoiding hazardous runoff into a watershed if such controls can be made to apply to the whole watershed.

**Municipal boundaries and their impact on water stewardship**

Studies of individual watersheds have yielded conflicting results as to whether watershed management is best when municipal boundaries match watershed limits. If political boundaries cannot be redrawn along the boundaries of watersheds, the geographic limits of water resource management agencies and other water governance initiatives would be best set using the limits of watersheds (Kauffman, 2002). Using watershed boundaries to define a watershed management area has improved accountability, participation, and empowerment in water governance in Ontario (Davidson & Loë, 2014). In this case, defining a single, specific agency with responsibility for a specific problem aided the public in identifying and protesting lapses in responsibility. Furthermore, watersheds are an ideal unit of analysis for understanding the intersection between social and environmental problems (Hill, Collins, & Vidon, 2018). On the other hand, divided governance of watersheds can help involve beneficiaries of a watershed in financing the governance of watersheds (Patterson, Hughes, Barnes, & Berahzer, 2012), and may also serve to better represent the diverse interests of inhabitants of watersheds by serving to better give each interest its own management authority (Blomquist & Schlager, 2005). The EPA offers guidance for managing the interests of different governments, landowners, and residents in a particular watershed using stakeholder working groups, but these groups are not mandatory to achieve federal funding (Tetra Tech Inc., 2015).

**Connecting divided watershed governance and poor water stewardship**

Given that previous research where the unit of analysis is an individual watershed has provided conflicting answers to whether divided water governance is better or worse for solving water governance problems, a study of numerous watersheds might be the best way to answer the question. It does not appear that anyone has studied this question outside the context of a specific watershed or set or related watersheds. Therefore, to answer the question of whether divided water governance impedes good water stewardship, and therefore creates water problems, a multi-watershed comparative analysis is called for.

**Data**

The study area of this study is: the state of Georgia and it is 59,425 square miles. The population of the study area is roughly 11.2 million.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Layer | Source | Date/Year | Data Type | Data Format | Use |
| tl\_2020\_13\_county20 | US Census | 2020 | Polygon vector feature layer | Shapefile | “Independent Variables” - Municipal Boundaries |
| Cities\_2019\_TIGER | GIO Data Hub | 2019 | Polygon vector feature layer | Shapefile | “Independent Variables” - Municipal Boundaries |
| WBD | USGS Data Downloader | 2025 | Line and polygon vector feature layers | File Geodatabase OR gpkg OR shapefile | “Independent Variables” – Watershed Boundaries |
| PSDI | NCEI (NOAA) | 2025 | polygon vector feature layers | File Geodatabase OR gpkg OR shapefile | “Dependent variable” – Drought |
| Flood Risk | NRI (FEMA) | 2025 | polygon vector feature layers | File Geodatabase OR gpkg OR shapefile | “Dependent variable” – Flooding |
| Elevation | USGS NED | 2025 | Polygon vector feature layers | File Geodatabase OR gpkg OR shapefile | “Control variable” – Elevation |
| Wildfires | Georgia Forestry Commission | 2024 | Point vector feature layers | File Geodatabase OR gpkg OR shapefile | “Control variable” – Wildfire |
| Drinking Water Non Compliance, Wastewater discharge | EJScreen | 2025 | Polygon vector feature layer | File geodatabase OR shapefile | “Dependent variable” – Water Quality |
| Population | EJScreen | 2025 | Polygon vector feature layer | File geodatabase OR shapefile | “Control variables” - Population |

**Methods**

**Operationalizing divided water governance**

The first task for this project was to operationalize divided water governance. However, divided water governance has no existing index. Therefore, I implemented four methods of quantifying watershed dividedness, which would allow me to capture the different ways in which watersheds are divided.

All quantification methods use a consolidated county (US Census, 2020) and city (GIO Data Hub, 2019) data referred to as the “municipalities” layer. This was created by first changing the CRS on the cities layer to match the counties layer. Then the cities layer was dissolved using the st\_union function. I then clipped the counties layer using this dissolved incorporations layer using the st\_difference functrion. Lastly, I combined the clipped counties layer and the original cities layer into the final municipalities layer using rbind. Following this, a linestring layer named boundary\_lines is created for municipal boundaries using st\_boundary and st\_cast.

All four methods are calculated in the same loop which iterates through all watersheds in my WBD HU12 dataset (USGS Data Downloader, 2025). The loop first runs a conditional to determine if the watershed intersects with any municipalities using st\_intersects. Keep in mind that prior to the loop, all watersheds that do not intersect with a Georgia municipality should have been filtered out using an st\_filter with an st\_intersects predicate. If this condition is met, calculations for methods #2 and #3 begin. If it is not met, values for all scores are given as zero, however again this edge case should not be reached.

Calculating methods #2 and #3 starts by using st\_intersection between the boundary\_lines layer and the current watershed the loop is studying (which is name ws). The product of this operation is named lines\_in\_ws and is casted to a linestring. A conditional is then run which checks the number of rows in the lines\_in\_ws layer. If the number is greater than zero, calculations for methods #2 and #3 can be computed, whereas if not, a zero is given for both of these scores. The calculations for method #2 involve creating a lines\_union using st\_union on lines\_in\_ws, before summing the length of this union using st\_length and sum. This length is calculated in kilometers, which means the value must be divided by 1000 to convert from meters.

To then calculate the score for method #3 a centroid is calculated the ws using st\_centroid. Then, a layer named points is created by sampling one point per meter on the lines\_in\_ws using st\_segmentize, as well as a st\_cast to convert the lines into points. The distances between each sampled point and the centroid are then calculated using st\_distance. Following this, the maximum distance is calculated sampling the boundary of the watershed’s geometry and taking the maximum distance between those sampled points and the watershed centroid using st\_sample, st\_boundary, and st\_distance. Weights are then calculated by adding distances divided by max distance to 0.5 using st\_distance and as.numeric. These weights are then summed and converted into weighted kilometers by dividing by 1000 and using sum.

Following this conditional, a list of municipalities in the ws is created using st\_intersection. An intersection count is created using nrow, giving us our value for method #1. If this value is greater than 1, we can calculate method #4. If not, method #4s score is set to 1. To calculate method #4s score.

Following all the calculation of all of these methods of quantifying watershed dividedness, a figure is also generated for the score normalized by the land area of the watershed. The methods are then combined into a “watershed dividedness score” which combines the difference between the weighted boundary length and the total boundary length normalized by area, the intersection count divided by half of its maximum, and the dominance ratio divided in half.

**Operationalizing water problem burden**

To operationalize my first water problem variable, pollution, I needed to take my drinking water compliance and wastewater discharge data from (EJScreen, 2025) and convert it from the block group level to the watershed level using weighted averages. I did this by taking the intersection between my EJScreen data and my watersheds data using st\_intersection. I then calculated an area\_overlap field using st\_area. Then, I calculated an area-weighted average using a group\_by, and a summraise that created Dwater\_mean from the sum of Dwater multiplied by area\_overlap, divided by the sum of area\_overlap. I did the same for wastewater discharge, population, and poverty. Lastly, I dropped the geometry and joined this with my results table of watershed dividedness, such that the table now holds a Dwater\_mean variable.

To operationalize my second water problem variable, drought, I used NOAA’s NCEI PDSI (Palmar Drought Severity Index) at the county level (NOAA, 2025). A low PSDI indicates a dry area, whereas a high PSDI indicates a wet area. Ideally, PSDI scores are moderate, as extreme drought or extreme rain are both undesirable. I took this csv data and left\_joined it with my GA counties shapefile (United States Census Bureau, 2020) to get 1901-2000 PDSI means by county. I then updated the CRS on this data to match my watersheds data in UTM Zone 16N. I then took an intersection between this new drought\_counties data and my watersheds shapefile before calculating the area overlap using st\_area. I then used a group\_by and summarize to calculate a weighted drought mean before using st\_drop\_geometry and left\_join to bundle these area weighted averages into my watershed\_statistics data frame.

To operationalize my third water problem variables, flooding, I used FEMA’s National Risk Index shapefile (FEMA, n.d.). I first clipped it to Georgia using filter. I also updated the CRS to UTM Zone 16N before taking an intersection between the floods by census tract data and my watersheds data. I then used a group\_by and summarize to calculate a weighted flood risk mean before using st\_drop\_geometry and left\_join to add these area weighted flood risk averages into my watershed\_statistics data frame.

With these three variables operationalizing water problem burden in place, I attempted to combine the three into a “water problem burden score” by taking the sum of drinking water noncompliance, wastewater discharge dived by one hundred, flood risk mean, and the absolute value of the PDSI multiplied by one hundred.

**Regression diagnostics and control variables**

With all four methods of quantifying watershed dividedness and all three forms of water problem burden operationalized, I began to work on a more refined regression. This involved operationalizing several control variables, as well as adjusting several independent variables to ensure no OLS assumptions are violated.

To operationalize my first control variable, wildfire impact, I had to take my wildfire point data from (GFCGIS) and reprojecting it to UTM Zone 16N. Following this, I converted from USFS Fire Sizes encoded as letters from A to G into equivalent numeric values using mutate. I then joined this wildfire data with my watershed data using st\_join before aggregating by watershed using st\_drop\_geometry, group\_by, and summarize. I then added this fire summary to my watershed statistics data frame using a left\_join.

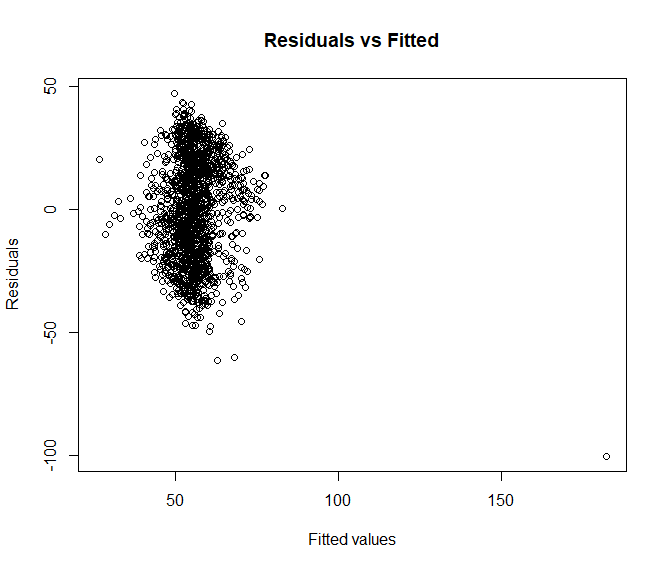
To run diagnostics effectively, I had to choose a dependent variable and associated model. I chose to focus my work on the regression that showed the most promise, which at this point was my flood risk model as it had the most significant variables, and the highest R squared. I therefore reran my flood model with the new control variable: wildfire, as well as drought as an additional control variable.

Previously, I had run each regression with all four dividedness methods as independent variables. However, methods #2 and #3 (total boundary length and weighted boundary length) are somewhat multicollinear with a VIF of 17.2 and 16.7. In view of this, I decided to try several different methods combined methods #2 and 3# that maximized R squared but avoided multicollinearity. Running a regression with solely method #2 (0.07547) or #3 (0.07526) produced an R squared around 0.75. Taking the difference between methods #2 and #3 and normalize it by method #3 to capture a “boundary centrality score” produced an R squared of 0.07909. Values were between -1 and 1, with a mean near zero, with a very tolerable vif of 1.21. A high boundary centrality score indicates the weighted boundary length is greater than the total boundary length, meaning more of the boundaries are closer to the watershed center. A low boundary centrality score indicates the boundaries are closer to the edge of the watershed, or possibly almost exactly at the border of the watershed. In cases where both method #2 and method #3 output a zero, the centrality score was set at -1 to indicate the boundaries are outside the watershed.

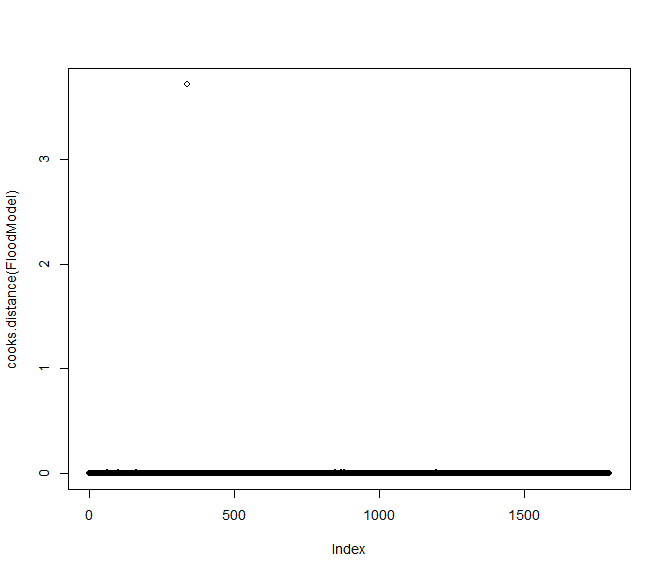
I also tried taking the difference of methods #2 and #3 and normalizing by watershed area, which produced an R squared of 0.08094 and a very tolerable vif of 1.213229. Values for this index were between -0.5 and +0.5, and zero values for methods #2 and #3 did not produce NAs. I then also tried taking dividing method #3 by method #2 to create a boundary centrality ratio, which produced an R squared of 0.07531 and a very tolerable vif of 1.004751. Values for this ratio were between 0 and 200, with NAs replaced by 0. I then further tried taking the difference between methods #2 and #3 without normalizing for anything which produced an r squared of 0.08474 and a very tolerable vif of 1.356482. Values for this figure were between -100 and 50, and zero values for methods #2 and #3 did not produce NAs. Out of curiosity at this point, I tried taking the sum of methods #2 and #3 without normalizing, which produced an r squared of 0.07503 and a vif of 1.658382. Values for this figure were between 0 and 450, and zero values for methods #2 and #3 did not produce NAs. Normalizing this by area produced an R squared of 0.07527 and a vif of 1.611785 with values ranging between 0 and 200 and zero values for methods #2 and #3 did not produce NAs.

With these results in hand, I decided to produce with the unnormalized difference between methods #2 and #3. Therefore, I was able to solve our multicollinearity problem and explain more variance by combining methods 2 and 3 into an actual score than by combining the two methods.

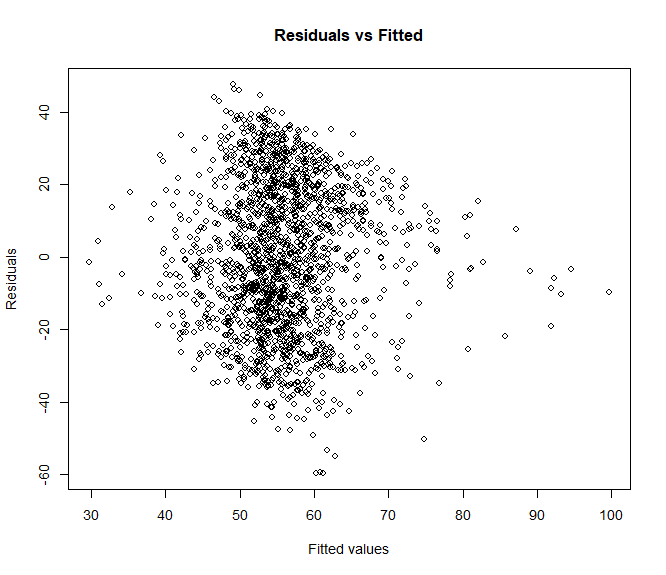
With multicollinearity handled, I now needed to test for heteroskedasticity. I started by running a Breusch-Pagan Test which came back very significant with a p value less than 2.2e-16. To look into this issue further I plotted the residuals against the fitted values which looked like this:



The model appears mostly homoscedastic, with the residuals and the fitted values in one mostly uncorrelated cluster. However, there is one significant outlier with a residual of -100 and a fitted value of 175. This outlier turns out to be the Okefenokee swamp, which makes a great deal of sense. The Okefenokee is completely like any other watershed in Georgia. Its governance is largely managed not by a municipality but by the National Parks Service. Furthermore, flooding is normal in the Okefenokee as it is a wetland: it is supposed to be partially inundated. I therefore checked its cook’s distance out of the suspicion that it is an egregious outlier.



The Okefenokee has a cooks distance over 3, well beyond a reasonable threshold of 0.5 Therefore, I removed the Okefenokee. This improved the p value of by Breusch-Pagan test to 9.405e-05 and changed the residuals plot to the below.

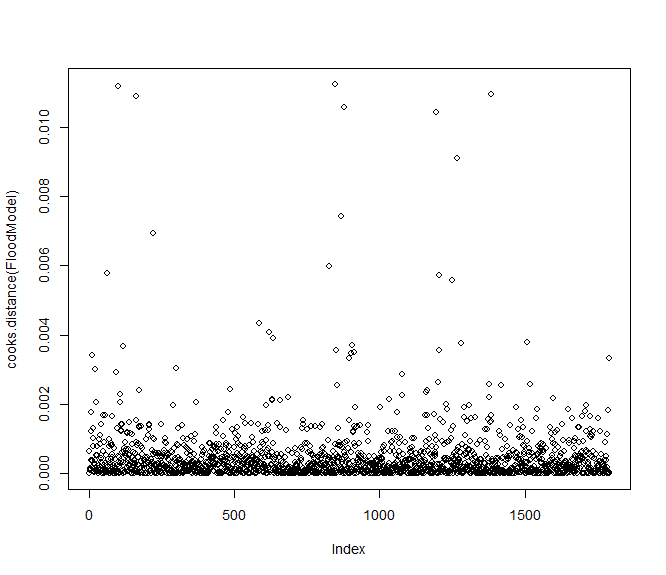


In my opinion, this plot shows a mostly homoscedastic model, with some heteroskedasticity. When we attempt to fit a line to this plot, the line is completely flat.

A graph of a graph showing a red line

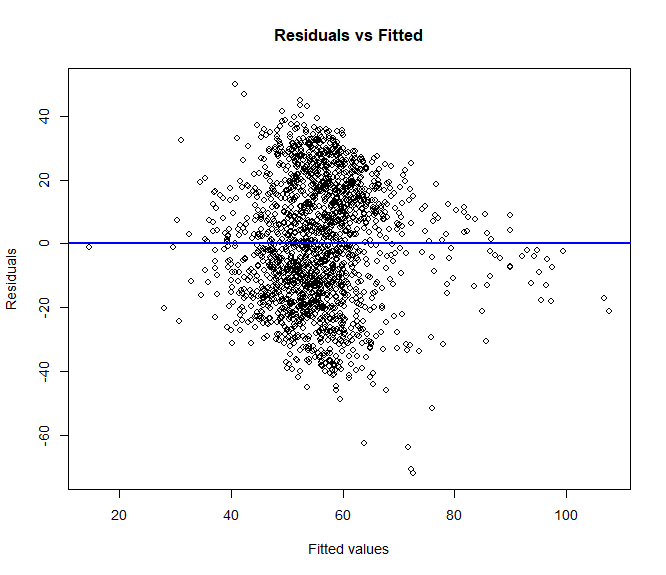
AI-generated content may be incorrect.

Removing the Okefenokee also ensured no rows in the data frame had a cook’s distance greater than 0.5, as the highest cooks distance was now below 0.2.

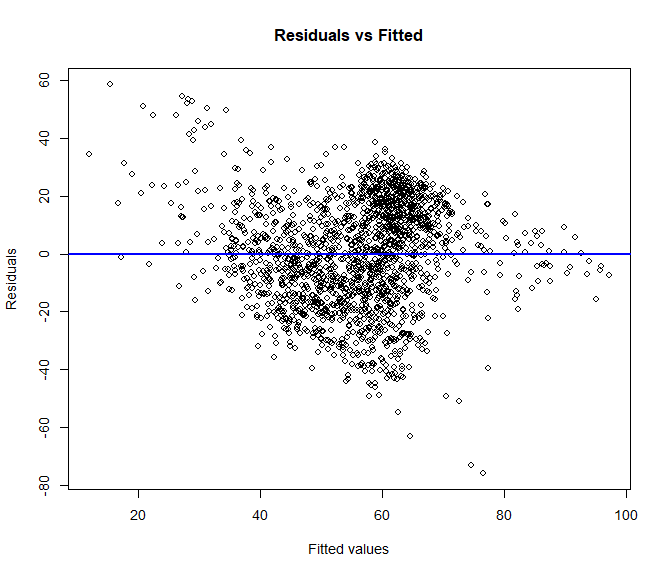


This suggested to me that the heteroskedasticityis a result of group level variance, a specific predictor, or some other model misspecification. Before I began log-transforming of the dependent variable, a WLS model, or imploring more into my standard errors, I attempting to incorporate more control variables. I planned to incorporate variables approximating low lying regions, which I planned on modelling by distance to sea level or elevation. I also planned on specifying an urban/rural parameter and a water body parameter. However, I realized that a weighted average block group population and county water area would be easier than these parameters but would still capture the urban/rural and water body parameters. I aggregated county water area by adding to the code to convert county drought data to watershed units, using a similar group by, summarize, drop geometry, and left join that I have been using. Likewise, I aggregated tract populations by adding to the code to convert EJScreen drinking water noncompliance data to watershed units, again using group by, summarize, drop geometry, and left join on the total population field.

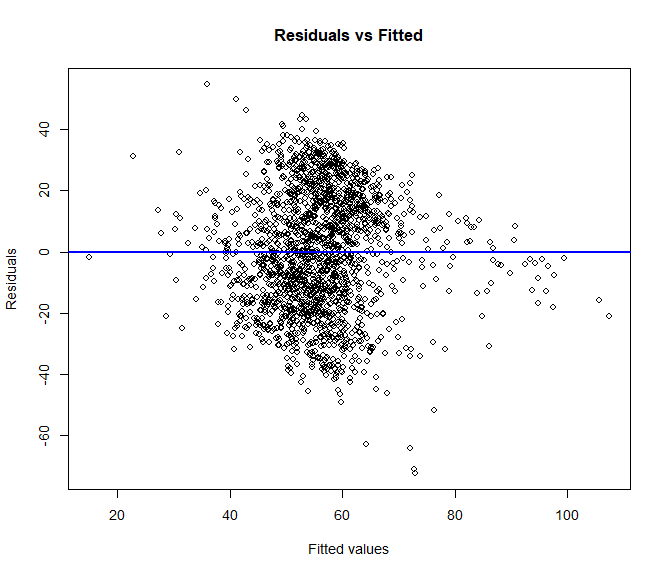
After adding these two variables into the model, and removing dominance ratio due to its p value of 0.68669, I was left with a model with a R squared of 0.1694, all significant variables, VIFs all below 2, and a residuals plot depicted below.



However, my Breusch-Pagan test still came back as significant with a p value of 1.11e-06, indicating there may be a case for adding additional control variables. I tried the model with drought and wildfire, however this actually increased the significance of the BP test and resulted in a more heteroskedastic residuals plot. Therefore, I added elevation as a control variable using get\_elev. While elevation was significant and improved r squared, it resulted in the below residual plot.



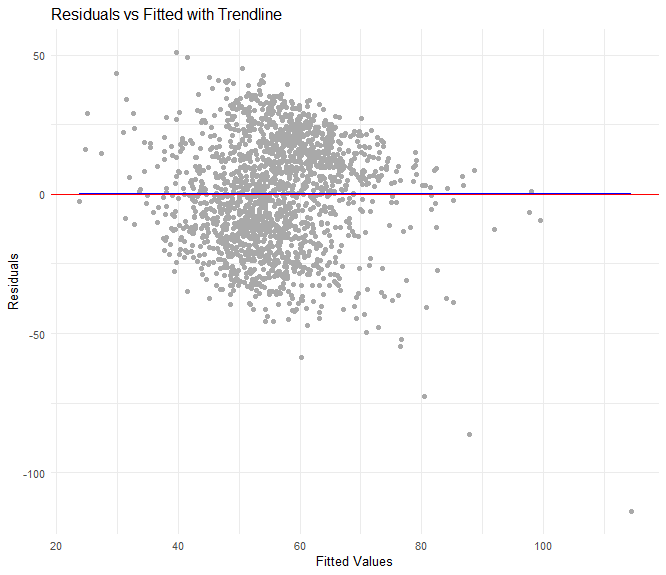
As well as a slightly more significant BP test result. Therefore, while drought, wildfire, and elevation may eventually make their way into the final model, they will be omitted for now. Wastewater discharge on the other hand improved the BP result, the shape of the residuals plot (below), and the R squared value, and therefore will be included.



To improve heteroskedasticity as well as the model overall, I turned my attention to logarithmic variables. I tested for skew for the six variables left in the model, yielding the following results.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | intersection\_count | boundary\_centrality | pop\_mean | AWATER\_mean | wastewater\_discharge\_mean | "area\_km2" |
| Skew | 1.278608 | -2.335407 | 1.205616 | 5.692835 | 9.518383 | 2.644639 |

Given that a skew greater than 1 indicates log transformation, I decided to log transform all variables with a skew greater than 1. However, this resulted in a markedly more significant BP test result (p value of 2.2e-16) and a far less ideal residuals plot, depicted below, so I reverted these changes and only log transformed wastewater discharge.



Running the model to predict flood risk with intersection count, boundary centrality, population mean, water area, and total watershed area resulted in a BP test that was still significant (p-value = 0.0002065), however less so with a BP of 24.113 on five degrees of freedom.

At this point, I realized that while carefully identifying variables that needed to be logarithmic and specifying interaction terms may help somewhat, the most likely candidate for significant OLS assumption violation was spatial auto correlation. Therefore, I ran a geographically weighted regression.

**Geographically weighted regression:**

My initial geographically weighted regression attempted to model my water problem burden score using my watershed dividedness score, and area weighted averages of population, water area, and overall area. However, this presented a multicollinearity problem and only resulted in one significant variable at the 5% significance level: population mean (p = 0.0495). The multiple R-squared was also abysmally poor, at 0.003585. Overall, the model was not significant with a p-value of 0.1698.

I therefore attempted to run a geographically weighted regression with the same parameters as the OLS regression I had diagnosed and fine-tuned previously, namely attempting to model area weighted mean flood risk using intersection count, boundary centrality, population mean, water area mean, with the log of wastewater discharge and the watershed area removed for theoretical reasons and due to multicollinearity concerns. This second model’s Breusch pagan test came in as significant (3.521e-5), with a BP of 25.77 on four degrees of freedom.

This made me suspect that the omission of land area for theoretical multicollinearity reasons may have been unwise. Therefore, I reran the GWR with land area included. However, this resulted in a similar residuals plot, with an even more significant Breusch pagan (p-value = 5.829e-08) rest result of 42.022. Therefore, I decided to continue with the previous (second) GWR model.

The final geographically weighted regression used a bisquare kernal function with an adaptive bandwidth (number of nearest neighbours) of 27 where regression points were the same locations as observations are used and the distance metric used was Euclidean. The model had 1792 data points, each of which was a watershed from my data frame of 1801 watersheds in Georgia. The effective number of parameters was 830.9209 and the effective degrees of freedom was 961.0791. The geographically weighted model’s diagnostic AICc was 13357.77, it’s AIC was 11918.83, and it’s BIC was 14414.91. The model’s diagnostic residual sum of squares was 56145.45, its R-square value was 0.9375629, and its adjusted R-square value was 0.8835253.

**Results and Analysis**

**Watershed dividedness analysis results**

The results of dividedness analysis are complex and multi-faceted, so put them in context, I have compared the results to my reference case of Proctor Creek, a watershed which suffers intense water problem burden (EPA, 2025).

My first method of quantifying watershed dividedness tallies the number of separate municipalities in a watershed to create an “intersection count.” For example: Proctor Creek-Chattahoochee River is governed by Cobb County, the City of Atlanta, Smyrna, and South Fulton. The average intersection count was 2.713, and when normalized for land area in kilometers squared it came down to 0.036. This makes the reference case, Proctor Creek, above average both in terms of intersection count (4) and normalized intersection count (0.065).

A map of a state with red and black dots

AI-generated content may be incorrect.

My second method of quantifying watershed dividedness measured the total length of political boundaries within the watershed. Every county line and city limit here is counted as a “political boundary.” The average Georgia watershed has 15km of municipal boundaries running through it, or 0.1759 once normalized by land area. This makes Proctor Creek only slightly above average, with 15.5km of boundaries running through it, or 0.252 once normalized by land area.

A map of a state with red and black squares

AI-generated content may be incorrect.

My third method of quantifying watershed dividedness measures total length of political boundaries is counted, but weights boundaries closer to the center of the watershed higher than boundaries further from the center of the watershed. Each municipality has an average of 12.2km of weighted boundary length, a figure without a real unit. Once normalized for land area, the average weighted boundary length was 0.146km/km squared. Proctor Creek came in at 12.9km of weighted boundary length, or 0.209 km/km squared, reflecting that the political boundaries dividing this watershed strike right through the center, where the Chattahoochee river divides Fulton and Cobb county.

A map of a state with red squares and black dots

AI-generated content may be incorrect.

My fourth method of quantifying watershed dividedness compares the ratio between the land area of the largest municipality in the watershed and the land area held by other municipalities. For watersheds with only one municipality, the ratio is set at 1 to reflect the total authority of that government to manage that watershed. The average ratio is 0.410 where the maximum score is 1 and the minimum score is -1. Proctor creek scores at 0.068, reflecting that the City of Atlanta controls roughly half of the watershed.

A map of a state with red and black dots

AI-generated content may be incorrect.

The final method of quantifying divided water governance combines the previous four methods into one score. This results in a score with a mean of 0.629, a maximum of 1.59, and a minimum of 0.2482. Proctor Creek scores 0.658, an above average score.

A map of the state of georgia

AI-generated content may be incorrect.

**Water problem burden analysis results**

The average Georgia watershed had a drinking water noncompliance mean of 1.1667, with the highest being 74.7, and the lowest being 0. Interestingly, the 1st quartile had a value of 0, indicated over 25% of Georgia watersheds have no instances of drinking water noncompliance. Proctor creek was among these, as it draws its water from upstream of the bulk of the Chattahoochee river’s pollution.

A map of water in different colors

AI-generated content may be incorrect.

Proctor creek’s wastewater discharge mean was 715 however, whereas the average watershed had a wastewater discharge of 2102, with the minimum being zero and the maximum being 188,500, indicating that while proctor creek has problems with wastewater discharge, such problems pale in comparison to other Georgia watersheds. Interestingly, the 1st quartile for wastewater discharge was 0.795, indicating there are several watersheds in Georgia with much lower figures for wastewater discharge.

A map of a state with red and black colors

AI-generated content may be incorrect.

With this in hand, I was able to calculate a preliminary regression model for drinking water noncompliance with all four methods of dividedness, as well as watershed area, as independent variables. The only variable with a P value below 0.05 was intersection count, which had a p value of 0.00174, and a negative correlation coefficient. This preliminary regression suggested that having multiple municipalities responsible for a watershed might actually be associated with improved drinking water compliance.

The average Georgia watershed had a flood risk mean of 55.7, with the lowest being 0.5613 and the highest being 99. Several coastal watersheds did not have flood risk means, and therefore they are omitted from this analysis. Proctor Creek had a flood risk mean of 19.86, suggesting it does not experience as severe flooding as the average Georgia watershed, likely because it is in the piedmont and not the coastal plain, where flooding is relatively lower.

A map of a flooded area

AI-generated content may be incorrect.

On conducting a preliminary regression, all variables except dominance\_ratio were significant with p values well below 0.001. Intersection count and weighted\_boundary length appeared to be negatively correlated with flood risk, whereas total boundary length appeared to be positively correlated with flood risk.

The average watershed had a PSDI of -0.2111, with values ranging from -0.38 to 0.03. Proctor Creek had a PSDI of -0.242, suggesting the watershed has been somewhat dry over the previous century, but not significantly more dry than the mean Georgia county.

A map of a state with different colored areas

AI-generated content may be incorrect.

On conducting a preliminary regression, the only variable that was significant was intersection\_count with a p value of 0.0973 which appeared to negatively correlate with drought mean.

The preliminary regressions suggested that none of the watershed dividedness scores were a significant predictor of drought, wastewater discharge, or wildfire, and that they did not explain almost any of the variance in drinking water noncompliance. They did however explain a respectable portion of the variance in flood risk.

On combining the three dependent variables into a single score, the average Georgia watershed had a water problem burden score of 99.142, with scores ranging between 3.561 and 1951.050. Proctor creek had a score of 51.18591, indicating a watershed water problem burden in the first quartile (64.238). The log transform of this score had a minimum of 1.27, a maximum of 7.56, and a mean of 4.412. The map of water problem burden reveals odvious spatial trends. Curiously, the Chattahoochee river watersheds appear to have much more model water problem burden scores than those belonging to other rivers in the coastal plain, the ridge and valley, and the southern Appalachians. A map of water problems

AI-generated content may be incorrect.

**Regression results**

For the global model, all variables are significant at every significance level, with the following coefficients, estimates, standard errors, t values, and P values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Coefficients: | Estimate | Std. Error | t value | Pr(>|t|) |
| Intercept | 7.397e+01 | 1.659e+00 | 44.603 | < 2e-16 |
| Intersection count | -3.057 | 3.942e-01 | -7.755 | 1.48e-14 |
| Boundary centrality | -5.617e-01 | 9.677e-02 | -5.805 | 7.62e-09 |
| Population area weighted mean | -9.993e-03 | 1.006e-03 | -9.934 | < 2e-16 |
| Estimated water area proportion | 6.956e-08 | 8.077e-09 | 8.612 | < 2e-16 |

The residuals for this model range between -73.399 and 51.395, with a median of -0.103. The residual standard error was 20.8 on 1787 degrees of freedom, with a multiple R-squared of 0.1406 and an adjusted R-squared of 0.1387. The model’s F-statistic was 73.09 on 4 and 1787 DF, and the model’s p-value was significant at < 2.2e-16.

Additionally, the residual sum of squares was 772794.6, with a Sigma(hat) of 20.77809, an AIC of 15968.97, an AICc of 15969.02, and a BIC of 14254.86. The VIF values were all well below 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | intersection\_count | boundary\_centrality | pop\_mean | AWATER\_mean |
| VIF | 1.190443 | 1.187903 | 1.026592 | 1.027081 |

The plot of the model’s residuals was very similar to previous models. It showed slight heteroskedasticity, but the trendline was flat.

A graph with a line and dots

AI-generated content may be incorrect.

The global model’s Breusch pagan test came in as significant (3.521e-5), with a BP of 25.77 on four degrees of freedom.

However, the geographically weighted model shows markedly different ranges for coefficient estimates (shown below).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| coefficient estimates | Min. | 1st Qu. | Median | 3rd Qu. | Max. |
| Intercept | -56.752 | 28.686 | 52.951 | 74.491 | 156.707 |
| Intersection count | -13.750 | -1.6847 | -0.009 | 1.467 | 13.470 |
| Boundary centrality | -5.787 | -0.607 | -0.224 | 0.108 | 4.186 |
| Area weighted average population | -0.0432 | -0.006 | 0.001 | 0.010 | 0.0636 |
| Area water percentage | 0.0000 | 0.0000 | 0.0000 | 0.000 | 0.000 |

The map of the coefficient estimate for water area percentage show some noteworthy spatial trends, however the range for the coefficient values is so miniscule, it makes the spread equally miniscule.

A map of a country

AI-generated content may be incorrect.

Spatial trends regarding coefficient estimates for boundary centrality are much more interesting. The appear to show that boundary centrality has a positive influence in southern Appalachia, parts of Atlanta & Macon, but not in Savannah, Colombus, or the ridge and valley.

A map of a country

AI-generated content may be incorrect.

For intersection count, values appear far more negative in Savannah and the border between the coastal plain and the piedmont, however spatial trends are harder to discern.

A map of a country

AI-generated content may be incorrect.

Lastly, the coefficient for population appears to have a positive influence in Macon, Savannah, Georgia, and Agusta, but not in Colombus.

A map of a large area with different colored squares

AI-generated content may be incorrect.

Local R squared values showed that the local model of the feature was almost universally explanatory of the majority of the variance, and that the main area where its explanatory power is reduced were in the ridge and valley and in the Augusta area.

Overall, it is clear that all variables have markedly different coefficient estimates dependent on location, confirming the hypothesis that spatial autocorrelation was a factor in this model.

Local R squared values showed that the local model of the feature was almost universally explanatory of the majority of the variance, and that the main area where its explanatory power is reduced were in the ridge and valley and in the Augusta area.

A green and white map

AI-generated content may be incorrect.

Furthermore, the map of the residuals showed no spatial trends I could identify, suggesting that the geographically weighted regression had accounted for any spatial auto-correlation that may have been present in the OLS model.

A map of a country

AI-generated content may be incorrect.

**5: Discussion**

My attempt to model watershed problem burden, drought risk, and drinking water noncompliance as a function of watershed dividedness did not result in a significant model with a robust R squared. My attempts to model flood risk as a function of dividedness, while they did result in a significant model albeit with a modest R squared, did not result in a solidly homoscedastic model. It is my belief that this indicates model misspecification, specifically spatial auto correlation.

Even if the global models were homoscedastic, the correlation between divided watershed governance and flood risk is extremely modest and on average opposite to what I hypothesized. Watersheds with divided governance appear to have very slightly lower flood risk, although the dividedness of the watershed accounts for very little of the flood risk of a particular watershed. This may be a result of the financial and representation advantages afforded by divided governance discussed earlier (Patterson, Hughes, Barnes, & Berahzer, 2012), (Blomquist & Schlager, 2005).

The geographically weighted model proved much more promising than its corollary global model. Its r squared was very strong, however the bulk of this strong r squared appeared to be a function of local r squared, and not of global r squared. Every coefficient estimate was highly spatially variable, suggesting that location was a far more powerful predictor than any or all of the variables specified in the model. In other words, I believe that in addition to proving null hypothesis #2. Divided watershed governance is not associated with flood risk in the state of Georgia. Divided watershed governance is however associated with higher flood risk in some municipalities according to the geographically weighted regression.

The variables used in the global model, and their associated estimates are not the best linear unbiased estimates. Therefore, divided watershed governance, in addition to being a poor global and local predictor of general water problem burden, is not a globally reliable predictor of flood risk. This would match the findings of my literature review, which showed that divided watershed governance has functioned as an aid to water resource stewardship in some watersheds, and as an impediment in other watersheds.

This may be a result of the hypothesis that water problem burdens and their causes, often thought to be the poor stewardship of land, are less a function of the political systems that govern a watershed, and more a function of the citizenry of a watershed. By poor stewardship of land, I mainly refer to agricultural practices such as reliance on tilling and nitrogen fertilizer, as well as damning and irrigation, heavy use of impervious surfaces, and improper waste handling. These practices may happen regardless of the governance structure of the watershed and the assignment of its constituents because the constituency is indifferent, ignorant, or powerless. Alternatively, one could hypothesize that water problem burdens are the result of the geographic features of a watershed, not the cultural geography of a watershed.

**6: Conclusions and recommendations**

I can conclude that divided watershed governance as I have quantified it is not a reliable global predictor of flood risk, drought, or water quality. In other words, based on my attempts to model it, there appears to be no reliable global statistical relationship between these burdens of water problems and the divided watershed governance. However, there exist watersheds where water problem burden and divided watershed governance are significantly positively correlated. There also exist watersheds where the water problem burden and divided watershed governance are negatively correlated. The local model explains the lions’ share of the variance in water problem burden, with one notable exception: the Augusta area, where a greater incidence of watersheds split in half by political boundaries appears to correlate with increased flood risk.

Future work may attempt to verify or falsify my demonstration of the global null hypothesis. This may involve specifying a different global model, either by quantifying watershed dividedness differently, quantifying water problem burdens differently, or by including alternative control variables. One interesting method of operationalizing the variables differently would be to conduct sensitivity testing by conducting the comparison on a different scale.

Future work may attempt to examine the cultural values, government watershed management practices, and watershed terrain conditions in those watersheds where I have reported a positive or negative coefficient between watershed dividedness and water problem burden. My suspicion is that watersheds where divided governance is positively correlated with water problem burden are ones in which citizens are engaged in favor of water resource stewardship, but governments are not, and that the opposite is true where the correlation is negative. However, it could equally be the result of other factors mentioned or not mentioned.

Additionally, future work may attempt to explain the lack of a relationship between divided watershed governance and water problem burden in select watersheds where I have shown there is no correlation. My suspicion is that the lack of a relationship is a result of the lack of a relationship between divided watershed governance and poor water stewardship. Unified watershed governance may not be associated with better water stewardship if the organization of the polity is irrelevant to the actions of its local government if the constituency is indifferent towards, ignorant of, or powerless to affect the role of its local government in creating water problems. One alternative explanation is that water stewardship and water problem burden are unrelated. The link between poor land & water stewardship and water problem burden however appears to be sound both at the global (Curry, 2005) and regional scale (Ma, 2020).

Assuming that the first explanation is true, which I have not proven but do suspect there is an additional direction for future work that I intend to follow. Namely, that each of us aim to be knowledgeable about, engaged with, and effectual in the protection of our local soil, our local ecologies, and our local community, regardless of their governance.

**7: References**

Apurv, T., & Cai, X. (2020). Drought propagation in contiguous U.S. watersheds; a process-based understanding of the role of climate and watershed properties. *Water Resources Research, 56*(9). Retrieved February 10, 2025, from https://galileo-gatech.primo.exlibrisgroup.com/permalink/01GALI\_GIT/dm3uqa/cdi\_proquest\_journals\_2448840311

Blomquist, W., & Schlager, E. (2005). Political Pitfalls of Integrated Watershed Management. *Society & Natural Resources, 18*(2), 101-117. doi:10.1080/08941920590894435

Curry, J. (2005). Changes in Tropical Cyclone Number, Duration, and Intensity in a Warming Environment. *Science, 309*(5742), 1844-1846. Retrieved July 22, 2025, from https://www.science.org/doi/10.1126/science.1116448

Davidson, S. L., & Loë, R. C. (2014). Watershed Governance: Transcending Boundaries. *Water Alternatives, 7*(2), 367-387. Retrieved February 10, 2025, from https://www.water-alternatives.org/index.php/volume7/v7issue2/251-a7-2-5/file

EPA. (2025, February 14). *Urban Waters and the Proctor Creek Watershed/Atlanta (Georgia)* . Retrieved from EPA: https://www.epa.gov/urbanwaterspartners/urban-waters-and-proctor-creek-watershedatlanta-georgia

FEMA. (n.d.). *National Risk Index*. Retrieved July 1, 2025, from Data Resources: https://hazards.fema.gov/nri/data-resources#shpDownload

Georgia Geospatial Information Office. (2019). *Cities 2019 TIGER*. Retrieved from GIO Data Hub: https://data-hub.gio.georgia.gov/datasets/bbb826ed88914118831ab37be0bf608c\_0/explore?location=32.946966%2C-83.042343%2C7.79

GFCGIS, a. (n.d.). GA State Wildfires. *Feature Service Feature Class*. Retrieved from https://services2.arcgis.com/iXA1dC6ldRMKRwra/arcgis/rest/services/GA\_State\_Wildfires/FeatureServer

Gordon, S. I., & Jones, A. L. (2000). From plan to practice: Implementing watershed-based strategies into local, state, and federal policy. *Enviornmental Toxicology and Chemistry, 19*(4), 1136-1142. doi:10.1002/etc.5620190445

Hill, D. T., Collins, M. B., & Vidon, E. S. (2018). The environment and environmental justice: Linking the biophysical and the social using watershed boundaries. *Applied Geography, 95*, 54-60. doi:10.1016/j.apgeog.2018.04.007

Kauffman, G. J. (2002). What if... the United States of America were based on watersheds? *Water Policy, 4*(1), 57-68. doi:10.1016/S1366-7017(02)00019-3

Ma, T. (2020, Jan 3). China’s improving inland surface water quality since 2003. *Science Advances, 6*(1). Retrieved July 22, 2025, from https://www.science.org/doi/10.1126/sciadv.aau3798

Mogollón, B., Villamagna, A. M., & Frimpong, E. A. (2016). Mapping technological and biophysical capacities of watersheds to regulate floods. *Ecological Indicators*(61), 483-499. doi:10.1016/j.ecolind.2015.09.049

NOAA. (2025, June 9). *Climte at a Glance County Mapping*. Retrieved from National Centers for Enviornmental Information: https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/mapping/9/pdsi/202505/1/mean

Patterson, L. A., Hughes, J., Barnes, G., & Berahzer, S. I. (2012). A Question of Boundaries: The Importance of “Revenuesheds” for Watershed Protection. *Journal of the American Water Resources Association*, 838-848. doi:10.1111/j.1752-1688.2012.00655.x

Powell, J. W. (1890). *Eleventh Annual report of the Director of the United States Geological Survey.* USGS. Washington: USGS. doi:10.3133/ar11\_1

Public Enviornmental Data Partners. (2025, February 14). *Environmental Justice Screening and Mapping Tool (Version 2.3)*. Retrieved from EJScreen: https://pedp-ejscreen.azurewebsites.net/

Saxe, S., Hogue, T. S., & Hay, L. (2018). Characterization and evaluation of controls on post-fire streamflow response across western US watersheds. *Hydrology and Earth System Sciences*, 1221-1237. doi:10.5194/hess-22-1221-2018

Tetra Tech Inc. (2015, November). *Engaging and Involving Stakeholders in your watershed*. Retrieved July 20, 2025, from EPA: https://www.epa.gov/sites/default/files/2015-11/documents/stakeholderguide.pdf

Tordoir, P. P., Raan, A. F., & Poorthuis, A. (2023). Effects of municipal boundaries measured by combining urban scaling and spatial interaction. *Journal of the Royal Society, 20*(198). doi:10.1098/rsif.2022.0775

United States Census Bureau. (2020). *2020 TIGER/Line Shapefiles*. Retrieved from Census Tiger Line Shapefiles: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2020.html#list-tab-790442341

United States Congress. (1890). *Report of the Special Committee of the United States Senate on the Irrigation and Reclamation of Arid Lands* (Vols. 1-2). Washington, DC, United States of America: Government Print Office. Retrieved February 10, 2025, from https://ir.library.oregonstate.edu/concern/defaults/4b29bb711

United States Geologic Survey. (2023, December 27). *Watershed Boundary Dataset Georgia*. Retrieved 3 24, 2025, from USGS Data Download Tool: https://prd-tnm.s3.amazonaws.com/StagedProducts/Hydrography/NHD/State/Shape/NHD\_H\_Georgia\_State\_Shape.zip

Whitfield, P. (2012). Floods in future climates: a review. *Journal of Flood Risk Management, 5*(4), 336-365. doi:10.1111/j.1753-318X.2012.01150.x

Xue, B., Wang, G., Xiao, J., & Helman, D. (2020). Global convergence but regional disparity in the hydrological resilience of ecosystems and watersheds to drought. *Journal of Hydrology, 591*. doi:10.1016/j.jhydrol.2020.125589

Yuan, L., Sinshaw, T., & Forshay, K. J. (2020). Review of Watershed-Scale Water Quality and Nonpoint Source Pollution Models. *Geosciences, 10*(1). doi:10.3390/geosciences10010025

Zhang, S. (2017). The influence of changes in land use and landscape patterns on soil erosion in a watershed. *574*, 34-45. doi:10.1016/j.scitotenv.2016.09.024

**8: List of figures**

To be completed

**9: List of charts**

To be completed

**10: List of tables**

To be completed

**11: Appendix A**

Code for this project is available on github at <insert link>. Data produced by this project is available on ArcGIS online at <insert link>.